



**УЧЕБНИК**

**В.В. Власов**

# **ЭПИДЕМИОЛОГИЯ**

**3-е издание, переработанное  
и дополненное**



**ИЗДАТЕЛЬСКАЯ ГРУППА  
«ГЭОТАР-Медиа»**



616.9  
B 58D



**В.В. Власов**

# ЭПИДЕМИОЛОГИЯ

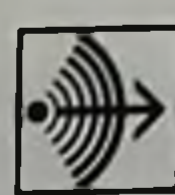
3-е издание, переработанное  
и дополненное

Министерство науки и высшего образования РФ

Рекомендовано ФГБОУ ДПО «Российская медицинская академия  
непрерывного профессионального образования» в качестве  
учебного пособия для студентов медицинских вузов

Протокол № 13 от 28 сентября 2020 года

Москва  
ИЗДАТЕЛЬСКАЯ ГРУППА  
«ГЭОТАР-Медиа»  
2021



SamDTU  
axborot-resurs markazi  
16234



УДК 51.9(075.8)  
ББК 616-036.22я73-1  
В58

01-УЧБ-3682

**Автор:**

*Власов Василий Викторович* — доктор медицинских наук, заслуженный профессор Национального исследовательского университета «Высшая школа экономики».

**Рецензенты:**

*Оганов Рафаэль Гегамович* — доктор медицинских наук, академик РАН, организатор и первый директор Государственного научно-исследовательского центра профилактической медицины;

*Реброва Ольга Юрьевна* — доктор медицинских наук, профессор кафедры медицинской кибернетики и информатики Российского национального исследовательского медицинского университета им. Н.И. Пирогова, вице-президент Общества специалистов доказательной медицины.

**Власов, В. В.**

В58 Эпидемиология : учебник / В. В. Власов. — 3-е изд., перераб. и доп. — Москва : ГЭОТАР-Медиа, 2021. — 496 с. : ил. — DOI: 10.33029/9704-6189-1-EPI-2021-1-496.  
ISBN 978-5-9704-6189-1

Учебник освещает фундаментальные аспекты современной эпидемиологии и биостатистики: проблемы причинности, измерений и вычислительного анализа, основные структуры исследований в медицине. Детально рассмотрены ключевые для медицины вопросы оценки диагностических и лечебных вмешательств, выявления и оценки последствий вредных воздействий. Изложены аспекты использования эпидемиологических методов в общественном здоровье и управлении здравоохранением.

Издание предназначено студентам медицинских вузов.

УДК 51.9(075.8)  
ББК 616-036.22я73-1

*Права на данное издание принадлежат ООО Издательская группа «ГЭОТАР-Медиа». Воспроизведение и распространение в каком бы то ни было виде части или целого издания не могут быть осуществлены без письменного разрешения ООО Издательская группа «ГЭОТАР-Медиа».*

© Власов В.В., 2021  
© ООО Издательская группа «ГЭОТАР-Медиа», 2021  
© ООО Издательская группа «ГЭОТАР-Медиа»,  
оформление, 2021

ISBN 978-5-9704-6189-1



# ОГЛАВЛЕНИЕ

Предисловие . . . . .	8
Список сокращений и условных обозначений . . . . .	10
Введение . . . . .	11
<b>ГЛАВА 1. Распространенность болезней . . . . .</b>	<b>19</b>
1.1. Причинность и методы ее изучения . . . . .	19
1.1.1. Общие сведения о причинности . . . . .	20
1.1.2. Причинность в медицине . . . . .	24
1.1.3. Признаки причинного характера связи . . . . .	27
1.1.4. Факторы риска . . . . .	33
1.2. Методология научного исследования . . . . .	39
1.2.1. Данные, факты и знания . . . . .	41
1.2.2. Структуры исследований . . . . .	42
Контрольные вопросы и задачи к разделам 1.1, 1.2 . . . . .	49
1.3. Описательные исследования . . . . .	50
1.3.1. Общие свойства описательных исследований . . . . .	51
1.3.2. Наблюдение за заболеваемостью населения . . . . .	51
1.3.3. Вероятность . . . . .	56
1.3.4. Преваленс . . . . .	61
1.3.5. Инциденс . . . . .	64
1.3.6. Показатели здоровья населения . . . . .	67
1.3.7. Показатели наглядности . . . . .	70
1.3.8. Стандартизованные показатели . . . . .	71
Контрольные вопросы и задачи к разделу 1.3 . . . . .	76
1.4. Выборочные исследования . . . . .	77
1.4.1. Репрезентативность оценки . . . . .	77
1.4.2. Случайная выборка . . . . .	78
1.4.3. Типичный и стандартный случаи . . . . .	79
1.4.4. Точность оценки . . . . .	81
1.4.5. Статистическая мощность . . . . .	82
Контрольные вопросы и задачи к разделу 1.4 . . . . .	83
1.5. Регистрация и подготовка данных в исследовании . . . . .	84
1.5.1. Типы переменных в исследовании . . . . .	84
1.5.2. Производные переменные . . . . .	87
1.5.3. Другие типы данных . . . . .	88
1.5.4. Измерения . . . . .	90



1.5.5. Документальные источники данных . . . . .	99
1.5.6. Формирование базы данных исследования . . . . .	101
1.6. Оценка распределения величин . . . . .	107
1.6.1. Оценка центральной тенденции . . . . .	107
1.6.2. Оценка вариабельности. . . . .	115
1.6.3. Нормальное распределение . . . . .	116
1.6.4. F-критерий . . . . .	121
1.6.5. Точность оценки средней . . . . .	122
1.6.6. Логнормальное распределение . . . . .	124
1.6.7. Биномиальное распределение . . . . .	125
1.6.8. Распределение Пуассона. . . . .	127
Контрольные вопросы и задачи к разделам 1.5, 1.6 . . . . .	129
1.7. Графическое представление данных . . . . .	129
1.7.1. Виды графиков . . . . .	129
1.7.2. Шкалы и их трансформация. . . . .	136
1.8. Статистический анализ данных . . . . .	137
1.8.1. Формирование гипотезы. . . . .	138
1.8.2. Проверка гипотез . . . . .	139
1.8.3. Статистическая мощность . . . . .	143
1.8.4. Выбор статистического критерия . . . . .	146
Контрольные вопросы и задачи к разделам 1.7, 1.8 . . . . .	147
<b>ГЛАВА 2. Выявление болезней и их причин. . . . .</b>	<b>149</b>
2.1. Классификация болезней и диагнозы . . . . .	149
2.1.1. Определение болезни . . . . .	149
2.1.2. Международная классификация болезней . . . . .	151
2.1.3. Виды диагнозов. . . . .	153
2.2. Клинический и эпидемиологический диагноз . . . . .	155
2.2.1. Референтный диагноз . . . . .	158
2.2.2. Оценка согласия . . . . .	159
2.2.3. Сопоставление с референтным диагнозом . . . . .	162
2.2.4. Операционные характеристики и преваленс. . . . .	165
2.2.5. Тест с количественным результатом . . . . .	171
2.2.6. Тест с качественным результатом . . . . .	174
2.2.7. Норма и референтные величины. . . . .	178
2.2.8. Место диагностических тестов в ведении больного. . . . .	179
2.2.9. Стандартное определение случая . . . . .	181
2.2.10. Прогноз . . . . .	182
Контрольные вопросы и задачи к разделам 2.1 и 2.2 . . . . .	186



2.3. Поперечные исследования . . . . .	188
2.4. Изменения заболеваемости во времени . . . . .	190
2.4.1. Вспышка заболеваний . . . . .	191
2.4.2. Применение критерия хи-квадрат ( $\chi^2$ ) и точного критерия Фишера . . . . .	193
2.4.3. Последствия внешнего воздействия . . . . .	197
2.4.4. Многолетние изменения смертности . . . . .	206
2.4.5. Критерий знаков, критерий Вилкоксона (W) и критерий Манна—Уитни (U) . . . . .	207
2.4.6. Временные ряды и прогнозирование . . . . .	214
2.4.7. Циклические изменения . . . . .	219
Контрольные вопросы и задачи к разделам 2.3 и 2.4 . . . . .	225
2.5. Различия заболеваемости в пространстве . . . . .	226
2.5.1. Географический анализ . . . . .	228
2.5.2. Применение <i>t</i> -критерия . . . . .	233
2.5.3. Сравнение частот . . . . .	235
2.5.4. Сравнение средних . . . . .	236
2.5.5. Дисперсионный анализ . . . . .	238
2.5.6. Сравнение множества групп . . . . .	241
Контрольные вопросы и задачи к разделу 2.5 . . . . .	242
2.6. Вероятность заболевания и смерти в течение жизни человека . . . . .	243
2.6.1. Демографические данные о смертности . . . . .	243
2.6.2. Анализ выживания . . . . .	250
2.6.3. Влияние болезней на смертность . . . . .	260
Контрольные вопросы и задачи к разделу 2.6 . . . . .	261
2.7. Когортные исследования . . . . .	262
2.7.1. Принцип когортного анализа . . . . .	263
2.7.2. Проспективное когортное исследование . . . . .	265
2.7.3. Дополнительный риск . . . . .	272
2.7.4. Вмешивающиеся факторы . . . . .	277
2.7.5. Когортное исследование по ретроспективным данным . . . . .	281
Контрольные вопросы и задачи к разделу 2.7 . . . . .	282
2.8. Исследования типа сравнения с контролем . . . . .	283
2.8.1. Методика исследования типа сравнения с контролем . . . . .	286
2.8.2. Измерение связи . . . . .	293
Контрольные вопросы и задачи к разделу 2.8 . . . . .	316



<b>ГЛАВА 3. Экспериментальные исследования</b> . . . . .	318
3.1. Цель и классификация клинических испытаний . . . . .	321
3.2. Виды вмешательств . . . . .	323
3.3. Объект эксперимента . . . . .	324
3.3.1. Формирования выборки . . . . .	325
3.3.2. Распределение и применение вмешательств . . . . .	329
3.3.3. Характеристика вмешательства . . . . .	339
3.4. Исходы . . . . .	345
3.4.1. Оценка исходов . . . . .	345
3.4.2. Качество жизни . . . . .	347
3.4.3. Прекращение испытания . . . . .	349
Контрольные вопросы и задачи к разделам 3.1–3.4 . . . . .	351
3.5. Анализ результатов испытания . . . . .	352
3.5.1. Анализ в подгруппах . . . . .	355
3.5.2. Статистическая и клиническая значимость результатов . . . . .	356
3.5.3. Ограничения на испытание вмешательств . . . . .	362
3.5.4. Изучение побочных эффектов вмешательства . . . . .	364
3.5.5. Приложение результатов испытаний к отдельному пациенту . . . . .	366
Контрольные вопросы и задачи к разделу 3.5 . . . . .	367
 <b>ГЛАВА 4. Специальные разделы эпидемиологии</b> . . . . .	369
4.1. Эпидемиология инфекционных болезней . . . . .	369
4.2. Эпидемиология хронических болезней . . . . .	377
4.3. Экологическая эпидемиология . . . . .	379
4.4. Генетическая эпидемиология . . . . .	383
4.5. Социальная эпидемиология . . . . .	389
Контрольные вопросы и задачи к главе 4 . . . . .	393
 <b>ГЛАВА 5. Эпидемиология и общественное здоровье</b> . . . . .	395
5.1. Обобщение результатов отдельных исследований . . . . .	396
5.1.1. Систематический обзор . . . . .	396
5.1.2. Метаанализ . . . . .	399
5.1.3. Комплексная оценка медицинских вмешательств . . . . .	404
5.1.4. Анализ принятия решений . . . . .	411
5.2. Оценка деятельности медицинских учреждений и программ . . . . .	413
5.2.1. Качество медицинской помощи . . . . .	416
5.2.2. Изучение исходов . . . . .	418
Контрольные вопросы и задачи к разделам 5.1 и 5.2 . . . . .	421



5.3. Обеспечение готовности общества к природным и социальным катастрофам . . . . .	423
5.4. Скрининг . . . . .	425
5.4.1. Теория скрининга . . . . .	427
5.4.2. Методологические проблемы скрининга . . . . .	431
5.4.3. Оценка эффективности скрининга . . . . .	436
5.5. Эпидемиология в формировании социальной политики . . . . .	442
5.5.1. Профилактика . . . . .	442
5.5.2. Оценка и модификация рисков . . . . .	444
Контрольные вопросы и задачи к разделам 5.3 и 5.4 . . . . .	447
<b>ГЛАВА 6. Этические проблемы эпидемиологии . . . . .</b>	<b>449</b>
6.1. Обязательства перед участниками исследования . . . . .	450
6.2. Обязательства перед обществом . . . . .	452
Контрольные вопросы и задачи к главе 6 . . . . .	457
Ответы на контрольные вопросы и задачи . . . . .	459
Словарь основных эпидемиологических терминов . . . . .	465
Список литературы . . . . .	472
Приложения . . . . .	474
Приложение 1. Величины $t$ -распределения при 5–200 степеней свободы для использования в вычислении доверительных интервалов . . . . .	474
Приложение 2. Критические значения $F$ -критерия . . . . .	475
Приложение 3. Оптимальные статистические критерии для сравнения групп при разных типах данных и разной структуре исследования . . . . .	476
Приложение 4. Оптимальные статистические критерии для сопоставления разных типов данных в медицинском исследовании . . . . .	478
Приложение 5. Выбор непараметрического статистического критерия для анализа данных . . . . .	479
Приложение 6. Распределение $\chi^2$ . . . . .	480
Приложение 7. Критические значения для критерия знаков при уровне значимости $P = 0,05$ . . . . .	481
Приложение 8. Критические значения критерия знаковых рангов Вилкоксона . . . . .	482
Приложение 9. Критические значения критерия Манна–Уитни $U$ при уровне значимости 0,05 . . . . .	483
Предметный указатель . . . . .	484



## ПРЕДИСЛОВИЕ

Эпидемиология является базовой медицинской дисциплиной, сформировавшейся в современном виде в 1960-е годы и с этого времени представленной в программах подготовки специалистов в медицинских вузах страны и за рубежом в университетах на степень как доктора философии, так и доктора медицины. Эпидемиология развилась на основе методов и проблем эпидемиологии инфекционных болезней и гигиены. Она изучает распространение всех болезней и состояний, их проявления и происхождение.

Настоящий учебник написан в соответствии с международным опытом преподавания эпидемиологии. Его содержание определяется тем, что эпидемиология является пропедевтической дисциплиной медицины и содержит основы понимания медицинской реальности, эффективности вмешательств и методов исследований в медицине. Эпидемиология как основная специальная дисциплина изучается на факультетах общественного здоровья (медико-профилактических, санитарно-гигиенических).

Прогресс эпидемиологии был особенно стремительным с 1950-х годов в связи с развитием сложных и эффективных методов медицинских исследований, разработанных с участием эпидемиологов, — прежде всего популяционных исследований сердечно-сосудистых болезней и рака. В 1970-е годы в основном сформировались словарь современной эпидемиологии и совокупность ее методов, которые быстро развиваются и поныне. Эти заслуги принадлежат поколению эпидемиологов, которые в нашей стране и на Западе почти все были врачами.

В 1980-е годы сформировалось движение доказательной медицины (*evidence based medicine*), которая стала синонимом медицинской практики, основанной на сведениях, полученных в научных исследованиях высокого качества. Безусловно, для врачей наиболее важны аспекты эпидемиологии, имеющие прямое отношение к медицинской практике и медицинским исследованиям. Эти разделы называют клинической эпидемиологией. В настоящем учебнике клиническая эпидемиология и доказательная медицина не выделяются, но соответствующие разделы изложены с учетом клинической проблематики.

Поскольку эпидемиологический метод применяется к анализу данных о совокупностях различающихся между собой людей и к сопоставлению сведений, полученных в разное время и в различных условиях,



он реализуется только при использовании статистических методов. Статистические методы необходимы, чтобы из совокупности отдельных наблюдений и измерений получить обобщенное знание. Эти методы излагаются в учебнике в объеме, необходимом для понимания основных методов исследований в эпидемиологии и общественном здоровье, и в форме, учитывающей возможности использования компьютерных программ. Поэтому приведенные способы расчетов и примеры не предназначены для детальной проработки изучающими эпидемиологию, но могут быть полезны начинающим исследователям в понимании принципов расчетов.

Полезным дополнением к этому учебнику является Эпидемиологический словарь, изданный в 2009 г. [1].

Я признателен многим коллегам, рецензировавшим предыдущие издания, прежде всего Н.И. Брико, М.С. Гельфанду, В.П. Леонову, С.Л. Плавинскому, О.Ю. Ребровой. Моя особая благодарность Б.А. Ревичу, написавшему для этого издания раздел «Экологическая эпидемиология».

*В.В. Власов, заслуженный профессор  
Национального исследовательского университета  
«Высшая школа экономики»*



## СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

- — торговое название лекарственного средства и/или фармацевтическая субстанция
- АД — артериальное давление
- ВИЧ — вирус иммунодефицита человека
- ВОЗ — Всемирная организация здравоохранения
- ДИ — доверительный интервал
- ИСК — исследования типа сравнения с контролем
- КИ — клиническое испытание
- ЛС — лекарственное средство
- МКБ — Международная классификация болезней
- ОИМ — острый инфаркт миокарда
- ОР — относительный риск
- ОШ — отношение шансов
- РКИ — рандомизированное контролируемое испытание
- РМЖ — рак молочной железы
- СКО — среднее квадратическое отклонение
- СПИД — синдром приобретенного иммунодефицита
- ТКФ — точный критерий Фишера
- ЧПЛП — число пациентов, подвергаемых лечению, на один предотвращенный вредный исход
- ЭКГ — электрокардиограмма
- LR (likelihood ratio) — отношение правдоподобия



## ВВЕДЕНИЕ

Эпидемиология возникла как наука об эпидемиях. Под эпидемиями понималось примерно то, что и сейчас, — заболевание множества людей в относительно короткие сроки. Будучи частью медицины, эпидемиология отличается от медицинской практики подходом к проблеме. Врач рассматривает особенности больного и помогает ему. Эпидемиолог изучает различия и общие свойства больных, чтобы помочь большим группам людей. Соответственно, диагноз эпидемиолога отличается от диагноза врача. Наука и практика медицины и эпидемиологии постоянно обогащали друг друга. В рамках медицины и эпидемиологии возникли идеи, которые составили затем основу для понимания инфекционных болезней и дали толчок развитию микробиологии.

Экзогенное и эндогенное в патологии — основные подходы к пониманию болезней. Понимание болезней как экзогенных ведет к поиску внешнего начала. Такими были медицинские и эпидемиологические теории, связывавшие заболевания с «миазмами», особенностями местности, другими внешними причинами. Это направление послужило основой для выделения общих для заболевающих людей внешних агентов, таких как микроорганизмы, химические вещества, условия труда. Поскольку под очевидным действием одинаковых условий заболевают обычно не все люди, всегда остается место для внимания к различиям людей — «конституционального» подхода. На основе изучения различий людей были получены знания об иммунитете и, в более широком смысле, об индивидуальной устойчивости к внешним воздействиям.

При выделении новых болезней всегда встает вопрос об их происхождении. Одни болезни были распознаны как экзогенные, инфекционные очень давно (малярия), другие — позднее. Например, специфическое поражение нервной системы при сифилисе долго считали следствием склонности к путешествиям, а язвенную болезнь до 1980-х годов рассматривали как полиэтиологическую и психосоматическую болезнь. Происхождение третьих до сих пор остается неясным, и они рассматриваются как эндогенные (шизофрения, «эссенциальная» артериальная гипертензия). Современное понимание возникновения и течения болезней опирается на представление о взаимодействии внешних и внутренних факторов. Соответственно, в эпидемиологи-



ческом исследовании<sup>1</sup> ставится вопрос не только о главной причине, но и об условиях, изменяющих взаимодействие организма с внешним агентом. В результате не только достигается более полное понимание процесса, но и разрабатываются способы, которые помогут изменить течение болезни или распространение инфекции, воздействуя на условия, а не только на причину.

**Эпидемиологический анализ явлений.** Многие гипотезы о происхождении болезней, в том числе инфекционных, исходили из того, что их истинная причина неизвестна и может быть неизвестна еще долго. Обычно это означало сознательную помощь отдельному пациенту или обществу на основе имеющихся знаний, которые не обязательно правильны. В социальной и медицинской практике есть множество общепринятых вмешательств<sup>2</sup>, полезность которых никто не проверял. В изучении распространения инфекционных болезней ученые исторически длительное время развивали теоретические представления о гипотетическом начале, вызывающем болезнь, не зная о существовании микроорганизмов.

Примером анализа заболеваний на основе доступных признаков является изучение их распространенности по территории и появление заболеваний во времени. Этот анализ традиционно практиковался в медицине, но получил широкое признание с начала XIX в. Материалы такого анализа были, в частности, использованы для проверки предположения о том, что холера в Лондоне связана с прокладкой канализации. Сначала удалось отклонить связь холеры со структурой городской канализации. Позднее Дж. Сноу с помощью карты (рис. 1) показал концентрацию случаев холеры вокруг нескольких водоразборных колонок, подающих воду из грязной Темзы. Карта не решила проблему сама, но подтверждала предположения Сноу и его предшественников о связи заболеваний с источником воды.

---

<sup>1</sup> Исследованием (научным) называют деятельность, направленную на получение нового общественно важного знания (вариант познавательной деятельности). Отличается от другой познавательной активности исторически сложившимся строгим научным методом. Иногда словом «исследования» обозначают лабораторные и другие инструментальные тесты, выполняемые в медицинских целях.

<sup>2</sup> В медицине вмешательством называют любое действие, предпринимаемое с целями профилактики, диагностики или лечения, например пункцию, антибиотикотерапию, обучение пациентов методам измерения концентрации глюкозы в крови.



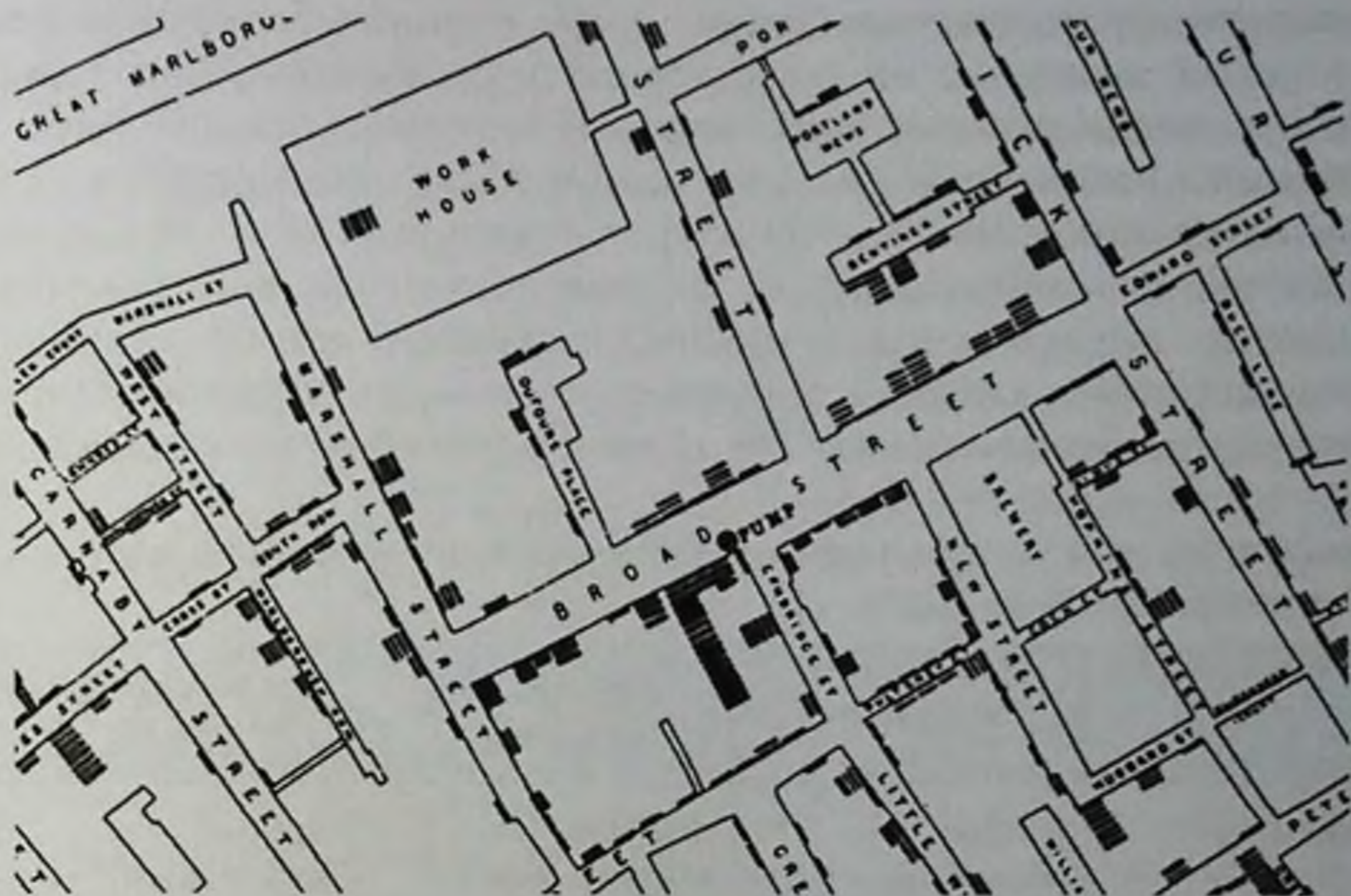


Рис. 1. Схема района Лондона с нанесенными на ней водоразборными колонками (Pump) и случаями смерти от холеры. Отдельные случаи обозначены штрихами и образуют столбики по сторонам улиц соответственно домам. Карта использовалась в анализе Дж. Сноу (Snow J. On the Mode of Communication of Cholera, 2<sup>nd</sup> Ed., John Churchill, New Burlington Street, London, England, 1855)

**Смена теорий в эпидемиологии.** Научный метод имеет особенности в естественных, социальных и точных науках, но в целом ему свойственно стремление к контролю над влиянием привходящих факторов на результат исследования, к опоре на эмпирические (опытные) данные и к получению результатов, имеющих общее значение (не только для отдельного эксперимента). Научный метод предполагает, что реальность существует независимо от того, как ее воспринимает человек, а природные явления развиваются по устойчивым закономерностям, которые могут быть открыты исследователем.

Основания эпидемиологии как науки относят к XVII в., к тому же периоду, когда сформировались базовые принципы современной науки. *Санитарное движение* с его теорией *миазмы*, имеющей еще более раннее происхождение, определило особенности эпидемиологии начала XIX в. (табл. 1). Бурное развитие микробиологии привело к надеждам на скорое разрешение медицинских проблем человечества,



упрощенному пониманию причин инфекционных болезней и концентрации медицины на лабораторных исследованиях. Это период расцвета эпидемиологии инфекционных болезней с ведущей теорией инфекционного агента. Лишь к середине XX в. эпидемиология занимает современное место в структуре медицинских наук. Эпидемиология уже не концентрируется на распространении инфекционных болезней, а постепенно становится основной наукой медицины и общественного здоровья. Ключевая особенность эпидемиологии — изучение связи феноменов без обязательного знания подлежащих

Таблица 1. Три эры в эволюции современной эпидемиологии (по Susser M., Susser E. Am. J. Pub. Health. 1996. № 86. P. 668)

Эра	Парадигма	Аналитический подход	Профилактический подход
Санитарная статистика (первая половина XIX в.)	Миазма: отравление вредными эманациями почвы, воздуха и воды	Демонстрация кластеризации (совместного проявления) случаев заболеваний и смертей	Дренаж, канализация, очистка
Эпидемиология инфекционных болезней (конец XIX в. — первая половина XX в.)	Инфекционная: агенту соответствует специфическая болезнь	Лабораторная изоляция культур от больных и из мест заболеваний, экспериментальное воспроизведение заражения	Разрыв передачи (вакцины, изоляция зараженных в карантине и инфекционных больницах, антибиотики)
Эпидемиология всех болезней (вторая половина XX в.)	«Черный ящик»: воздействие на организм соотносится с исходом без обязательного знания этиологии и патогенеза	Относительный риск возникновения болезни для изучаемого воздействия на уровне индивидуума в популяции	Контроль над факторами риска через изменение образа жизни (питание, физическая активность и др.) или условий среды (загрязнение, пассивное курение и др.)



биологических процессов, реформулируется в терминах, заимствованных у кибернетики<sup>1</sup>.

Эпидемиологические исследования базируются в основном на наблюдении за развитием и распространением болезней человека, но их целью является установление причин болезней и способов влияния на развитие болезней. Логическим завершением исследований должен быть эксперимент, подтверждающий или отвергающий гипотезу исследователя. Это приложимо ко многим наукам, но не к изучению самого человека. По этическим соображениям экспериментирование на человеке ограничено даже в части, касающейся применения лечебных вмешательств. Экспериментирование с потенциально вредными воздействиями неприемлемо: человек не может быть средством, а может быть лишь целью любой деятельности. Поэтому основой эпидемиологического анализа является методически строгий анализ данных наблюдений.

Структура медицинских наук непостоянна. С течением времени образуются новые направления исследований, истощаются старые и формируются новые большие области со своим предметом исследований и своими методами, которые принято называть «науками». Основным принципом выделения медицинских наук является субстратный — отдельная наука изучает группу близких болезней. Таковы гастроэнтерология, дерматология, эндокринология. Этот принцип структурирования медицинского знания тесно связан с органометрическим подходом к классификации болезней. Одновременно в выделении медицинских наук действует функциональный принцип. В соответствии с ним хирургия в Средние века была отделена от медицины, и ею в Европе занимались цирюльники. В настоящее время методы диагностики, например лучевой или лабораторной, представляют собой отдельные специальности врачей и отдельные области исследований.

---

<sup>1</sup> В кибернетике (науке об общих принципах управления объектами, независимо от их природы) исследование системы подразумевает невозможность изучения ее внутреннего устройства по причине его сложности или закрытости. В медицине многие объекты не изучены или не могут быть изучены по этическим причинам. Такой объект с неизвестной структурой и функцией обозначается как «черный ящик», и изучаются «вход», т.е. поступающая в объект информация (материал, признак), и «выход» (проявления объекта, его действия, признаки). В результате можно установить закономерности функционирования объекта, не разрушая его и не изучая внутренней структуры и внутренних функций. В эпидемиологии, например, внешнее воздействие выступает как «вход», а исход (заболевание) как «выход».



Столь же непостоянны содержание и положение эпидемиологии. Медицина как наука и как общественная практика занимается проблемами отдельно взятого человека. Эпидемиология на популяционном<sup>1</sup> уровне изучает медицинские проблемы. Вместе с тем, поскольку проявления болезней у людей различны, медицинская наука и практика всегда опирались на обобщенные представления о том, как развивается болезнь у людей вообще, насколько действенно данное лекарство у *подобных больных*. Поэтому эпидемиология с ее методами является универсальной наукой медицины и общественного здоровья (здравоохранения): эпидемиология располагает методами, которые неспецифичны для отдельного пациента, конкретной болезни или патогенного агента. Методы вирусологии, биохимии или физики могут использоваться для изучения отдельной болезни, но они не являются медицинскими методами. Медицинское исследование состоит в описании течения болезни у совокупности схожих пациентов, сравнении групп пациентов или работников разных производств, сопоставлении исходов болезней у пациентов, получающих разное лечение. Методы подобных исследований являются предметом эпидемиологии. Иными словами, методы медицинских исследований являются эпидемиологическими методами.

Эпидемиология как наука возникла из изучения распространения инфекционных болезней. Это объясняется тем, что инфекционным болезням присуще быстрое внезапное распространение, привлекающее внимание и очевидным образом ставящее проблемы популяции впереди проблемы отдельного больного. Катастрофические эпидемии чумы, проказы, холеры оказали глубокое влияние на все аспекты человеческой культуры, а не только на медицину. Успехи вакцинации, антимикробной терапии и улучшение условий жизни уже в первой половине XX в. привели к снижению смертности от основных инфекционных болезней в экономически благополучных странах. Неинфекционные болезни (сердечно-сосудистые болезни, рак, язвенная

---

<sup>1</sup> Популяцией в строгом (биологическом) смысле является стабильная группа животных одного вида, проживающая на одной территории и воспроизводящая себя в ряду поколений. В эпидемиологии популяцией обычно называют совокупность людей, проживающих на территории, или совокупность людей, отличающихся по некоторому признаку, например наземные работники авиации или больные сахарным диабетом 2-го типа. В этом смысле «популяция» является синонимом «генеральной совокупности» и обычно на нее распространяются результаты выборочного исследования.



болезнь) после Второй мировой войны прочно заняли первое место в структуре смертности. Распространение неинфекционных болезней и травм обоснованно сравнивали с эпидемиями. Так, эпидемиология включила в круг изучаемых вопросов неинфекционную и неэпидемическую заболеваемость. Эпидемиология определяется как наука, изучающая причины, условия и механизм формирования заболеваемости населения путем анализа ее распределения по территории, среди различных групп и коллективов, а также во времени и у субъектов с различающимися характеристиками. Эпидемиологи последовательно разделяют заболевание как явление, наблюдаемое у индивидуального организма, и заболеваемость. Под заболеваемостью понимают совокупность заболеваний в популяции. Заболеваемость проявляется в статистических показателях, используемых эпидемиологами — показателях количества, частоты, отражающих различные стороны распространения болезней в популяции. Как болезнь отражает состояние здоровья человека, так показатели заболеваемости отражают состояние здоровья популяции.

Эпидемиология интенсивно использует статистические методы<sup>1</sup>. Это объясняется тем, что все явления в популяции людей или животных невоспроизводимы в деталях, а каждый случай не идентичен другому. Соответственно, использование статистических методов позволяет открыть в наблюдаемом разнообразии закономерности и применить эти закономерности в новых условиях. Методы статистики были привнесены в медицину и эпидемиологию из сельскохозяйственного эксперимента и технического контроля качества. Это наложило отпечаток на применение статистических методов. Ныне врачу доступен не только более широкий круг статистических методов, но и мощные средства вычислений — компьютеры. Появилась возможность сделать акцент на правильном применении более широкого круга статистических методов, отказаться от большей части вычислительных упражнений при изучении эпидемиологии. У изучающих медицину и эпидемиологию, специалистов общественного здоровья и организаторов медицинской помощи появились новые информационные инструменты — огромные массивы информации в интернете. Это не только электронные версии документов, ранее доступных только

<sup>1</sup> Статистика — наука и практика сбора, обобщения и анализа данных, подверженных случайным изменениям. Иногда термин используют для обозначения самих данных и обобщения данных.



в виде книг, но и базы данных<sup>1</sup> по заболеваемости, данные о геноме человека и других живых существ, статистические программы и многое другое. В этом учебнике приводятся ссылки на наиболее надежные и важные источники информации, программы и алгоритмы. Упоминание в учебнике компьютерных программ и иных продуктов не означает, что именно они рекомендуются для использования, если об этом не сказано специально.

Врач применяет разные методы анализа ситуации и разные способы действий. Традиционным для изучения медицины и практики является *дедуктивный* метод. Применяющий дедукцию врач основывает свои конкретные действия на теоретических знаниях о механизмах болезней и механизмах действия лекарства или операции. Так, в начале XX в. хирурги лечили ишемическую болезнь сердца шейной симпатэктомией, объясняя страдание сужением сосудов под действием симпатических нервов. *Эмпирический* метод основан на использовании опыта — личного или чужого. Обычно врач действует с пониманием механизмов заболевания и действия лекарств, но обосновывает свои действия результатами клинических испытаний, подтвердивших действенность лекарства. Именно на основании исследований эффективности вмешательств из практики лечения язвенной болезни исчезла ваготомия, а антибиотики прочно утвердились как наиболее эффективное средство. Дедуктивный и эмпирический методы анализа составляют *естественно-научный компонент* медицинской практики. Не менее важным является *гуманистический компонент* медицинской практики, состоящий в понимании пациента как дружественного человеческого существа. Этот компонент тесно связан с *этическим компонентом* медицинской деятельности, состоящим в применении принципов и норм этики. Эпидемиологические методы позволяют анализировать не только естественно-научные, но и многие гуманистические и этические аспекты медицинской деятельности, так как действия врача, их субъективные основания и последствия, страдания пациента и исходы болезней могут быть зарегистрированы и изучены.

---

<sup>1</sup> Базой данных (БД) называют сведения об объектах, систематизированные и записанные в стандартном виде, например, в таблицах. Объекты могут быть любыми, например работники производства, связанного с ионизирующими излучениями или структуры белков. Обычно имеют в виду компьютерные БД, в которых можно проводить поиск необходимых данных и содержание которых можно анализировать в исследовательских целях.



## Глава 1

# РАСПРОСТРАНЕННОСТЬ БОЛЕЗНЕЙ

### 1.1. ПРИЧИННОСТЬ И МЕТОДЫ ЕЕ ИЗУЧЕНИЯ

Знание причины болезни дает наиболее эффективный способ предотвращать и лечить ее. Поэтому главные усилия исследований в медицине направлены на установление причин и механизмов развития болезней. В некоторых случаях практикуемые медицинские вмешательства эффективны, несмотря на отсутствие знаний о механизме их действия или о причине болезни. Например, салицилаты успешно применялись задолго до выяснения механизмов лихорадки, а современное понимание механизма действия самих салицилатов сформировалось только в 1970-е годы. Обычно в истории изучения болезней можно обнаружить смену представлений о причине и смену вмешательств. На каждом этапе изучения болезни неизвестная истинная причина скрывается за представлениями о причине, которые используются исходя из имеющихся знаний. Например, язвенная болезнь, возникшая как массовое явление в Европе в конце XIX в., первоначально поражала молодых женщин, и ее объясняли «нервными причинами». В середине XX в. «пептическая язва» стала болезнью мужчин молодого и среднего возраста и рассматривалась как преимущественно «психосоматическая». Ее лечили, воздействуя на секрецию желудка лекарствами, и столь же успешно оперативно, удаляя желудок полностью или частично. Во второй половине XX в. открыли инфекционное происхождение язвенной болезни, и ведущим лечением становится антибактериальная терапия. Язвенная болезнь встречается все реже. Таким образом, знание причин выступает в качестве важного условия разработки эффективных вмешательств, но и без знания причины или при неверном понимании причины медицина помогает больным доступными средствами, пока накапливаются новые сведения, позволяющие уточнить представления о причине.



### 1.1.1. Общие сведения о причинности

Каждый человек уже в раннем детстве устанавливает связи между значимыми для него явлениями и предшествующими событиями (действиями). Так, легко устанавливается связь между прикосновением к горячей плите и болью, между поворотом крана и получением воды. Знание этих простых связей многим людям неплохо служит всю жизнь. Лишь при ближайшем рассмотрении оказывается, что реализация даже таких причинно-следственных связей предполагает соблюдение некоторых условий. Обжечься о плиту можно, если она включена, разогрета, а для того, чтобы она разогрелась, в электрической сети должно быть напряжение (гореть дрова). Основные причины ожогов (условия безопасности) люди изучают и запоминают быстро, детская боязнь прикосновения к плите сменяется обычным для многих взрослых мастерством.

На житейском уровне усваивается важное условие: причина существует до следствия и может сохраняться в то время, когда наблюдается следствие. Однако «следствие» не может возникнуть ранее причины. Если такая последовательность наблюдается, то у наблюдаемого события («следствия») есть другая причина. Трудность для большинства людей заключается в усвоении правила: «после того не значит вследствие того»<sup>1</sup>. Поэтому люди нередко увязывают следующие одно за другим события в причинно-следственные цепи. Это заблуждение сопровождает человечество на протяжении всей его истории. В медицине непонимание этого правила является корнем заблуждений относительно оценки действительности лекарств. Последовательность событий «помазал — и насморк прошел» оказывает огромное влияние на людей, даже если они понимают, что насморк проходит и без лечения. Лишь правильно организованные эксперименты позволяют выяснить, действительно ли вмешательство (мазь) изменяет течение болезни (ускоряет исчезновение симптомов ринита), т.е. найти истинные причинно-следственные отношения.

Несмотря на широкое употребление представлений о том, что наблюдаемые события имеют причину (концепция причинности), определить причину сложно. Д. Хьюм (D. Hume) в XVIII в. определил «причину как событие, вслед за которым наблюдается другое и когда после всех событий, подобных первому, наблюдаются события, подобные второму». В этом определении причине всегда сопутствует эффект, но в жизни

<sup>1</sup> Классическая латинская фраза, иллюстрирующая логическое заблуждение: *post hoc ergo propter hoc* (после этого значит по причине этого).



и в эпидемиологии эффект наблюдается не обязательно сразу после причины, и наблюдается после причины не всегда. Представление о жесткой связи причины и следствия называют *детерминизмом*. Такая жесткая связь не наблюдается в медико-биологических исследованиях. Поэтому современная эпидемиология использует модель причинности, называемую *компонентной*. Наиболее развитая ее версия принадлежит К. Ротману (К. Rothman).

Понимание причины как толчка и обязательного условия, без которого невозможно следствие, более всего соответствует простому случаю причины как *необходимого и достаточного условия*. *Необходимым* называется такое условие, которое не всегда ведет к изучаемому следствию, но без условия А следствия В не бывает (рис. 1.1, 1). Примером обязательного условия является присутствие в организме микобактерии туберкулеза для развития туберкулеза. *Достаточным* называется условие, при наличии которого следствие развивается всегда, хотя может наблюдаться и при других условиях (рис. 1.1, 2). Присутствия микобактерии туберкулеза в организме недостаточно для развития туберкулеза.

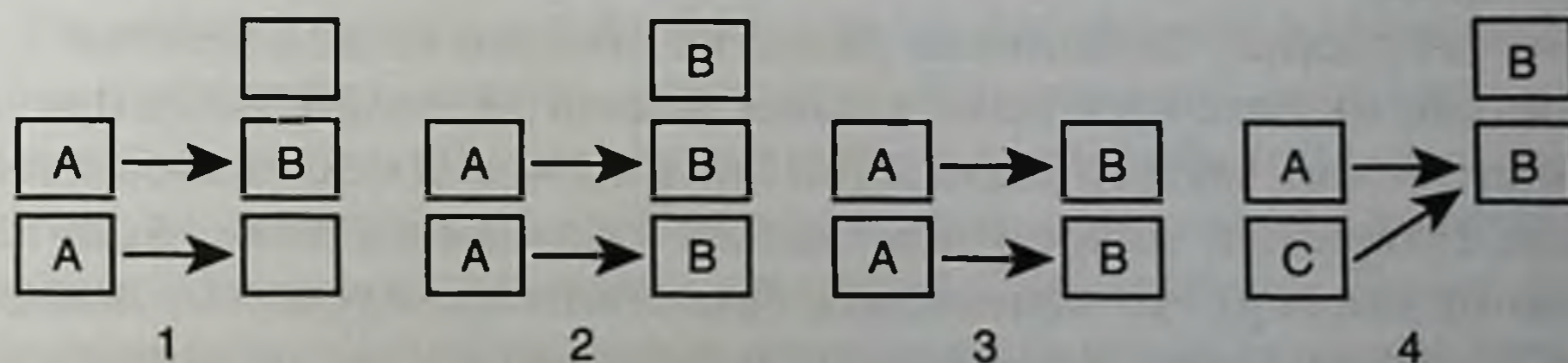


Рис. 1.1. Четыре типа причинных отношений. 1: А — необходимая причина для В. 2: А — достаточная причина для В. 3: А — необходимая и достаточная причина для В. 4: А и С — компонентные причины для В

Понимание причины как *необходимого и достаточного условия* соответствует ситуации, в которой при наличии условия А следствие В возникает всегда, и при отсутствии А не наблюдается В (рис. 1.1, 3). В медицине такие отношения наблюдаются почти исключительно в случаях сильных внешних воздействий — физических, химических. Достаточно уменьшить силу механического воздействия, и перелом кости не возникает. Однако и тут важна оговорка: *почти никогда* не возникает. В редких случаях переломы возникают при обычных для движений нагрузках на скелет. Такое возможно у людей со сниженной прочностью костной ткани вследствие снижения минерализации костей в старо-



сти и при некоторых болезнях. Иными словами, даже применительно к механической травме представление об особой механической нагрузке как необходимой и достаточной причине перелома оказывается упрощением.

Чаще в биологии и медицине имеют место *компонентные причины*. Каждая из них недостаточна для развития изучаемого следствия, но в совокупности они приводят к изучаемому следствию (рис. 1.1, 4). Если болезнь В вызывается факторами<sup>1</sup> А и С и наличия этих двух факторов достаточно, то для фактора С *дополнительной* (комплементарной) причиной будет А. Без нее С не вызывает болезни. Это определение важно, так как оно позволяет понять, как действие одного из факторов может изменять частоту<sup>2</sup> наблюдаемой болезни. Например, частота инфекционной болезни может зависеть не только от наличия источника инфекции, но и от состояния питания людей, от других факторов. Иными словами, частота возникновения болезни под действием причины С (контакт с вирусом) может зависеть от того, как часто наблюдается причина А (нарушение питания), необходимая для реализации патогенного действия С.

Биологические механизмы развития болезни не изменяются в зависимости от того, насколько высока частота дополнительных причин. Меняется лишь частота реализации биологических механизмов в популяции. Например, со временем частота заболеваний может изменяться, но не потому, что изменились биологические процессы, лежащие в основе заболевания, а потому, что изменилась частота компонентных причин, необходимых для реализации этих процессов. Так, детская смертность в беднейших странах, где она связана в основном с инфекционными болезнями, может быть снижена предоставлением всем детям витамина А.

Главной мерой силы воздействия на человека является относительная частота возникновения болезни у подвергшихся воздействию лиц

---

<sup>1</sup> Факторами называют признаки, которые изучают как возможные причины заболевания. Такими признаками могут быть особенности субъекта, его поведения, особенности окружающей среды.

<sup>2</sup> В эпидемиологии и статистике *частотой* называют число (количество) случаев, например число случаев заболевания. Поскольку число выявленных случаев заболевания зависит от числа обследованных людей, то для более наглядного представления результатов изучения заболеваемости число выявленных заболеваний делят на число обследованных людей. Этот показатель называют *относительной частотой*.



(статистическая связь воздействия и болезни — при воздействии болезнь развивается чаще). Эта мера в эпидемиологическом исследовании зависит не только от изучаемого воздействия, но обычно и от многих других. Поэтому оценка силы воздействия по относительной частоте заболевания может быть искажена вследствие действия одной из многих дополнительных причин. Условия, которые искажают изучаемую зависимость, называют вмешивающимися факторами (конфаундерами, *confounding factors*, англ., от лат. *confundere* — перемешиваться). В таком взаимодействии компонентных причин заключена самая большая сложность медицинских исследований.

Рак легких вызывается курением, но с меньшей частотой рак легких возникает и у некурящих. Никакого другого фактора, который был бы сильнее связан с развитием рака легких, чем курение, неизвестно. Еще предстоит изучить, как действует курение и какие причины приводят к раку у некурящих. В эпидемиологическом исследовании, сопоставив зарегистрированные факторы с возникновением изучаемого следствия, можно установить относительную значимость отдельных компонентных причин.

Изучаемый фактор может быть прямо действующей причиной болезни (состояния), или он может лишь инициировать цепь других промежуточных событий, или включаться в цепь событий, запущенных другим фактором (рис. 1.2). Например, нагревание кожи при контакте с горячей поверхностью прямо приводит к денатурации белков и разрушению других компонентов кожи. Однако при меньшей температуре (60–80 °С) запускаются сложные механизмы, которые являются непосредственной причиной болезни. Курение может быть не «первой» причиной, а существенно изменять течение процессов «естественного» канцерогенеза. Накопление знаний о взаимодействии причин и процессов патогенеза дает больше возможностей вмешаться в реализацию событий, ведущих к следствию (болезни).

Фактор, который ускоряет развитие болезни, например фактор А на рис. 1.2 (3), является не второстепенным фактором, а равноправной компонентной причиной. Без действия такого *иницирующего фактора* (*промотора*<sup>1</sup>) болезнь может не развиваться вообще в рамках продолжительности жизни человека. Поэтому профилактические вмешательства,

---

<sup>1</sup> Также в химии промотором называется фактор (вещество), которое само не участвует в превращениях и не является катализатором, но влияет на активность катализатора, который, в свою очередь, регулирует течение процесса.



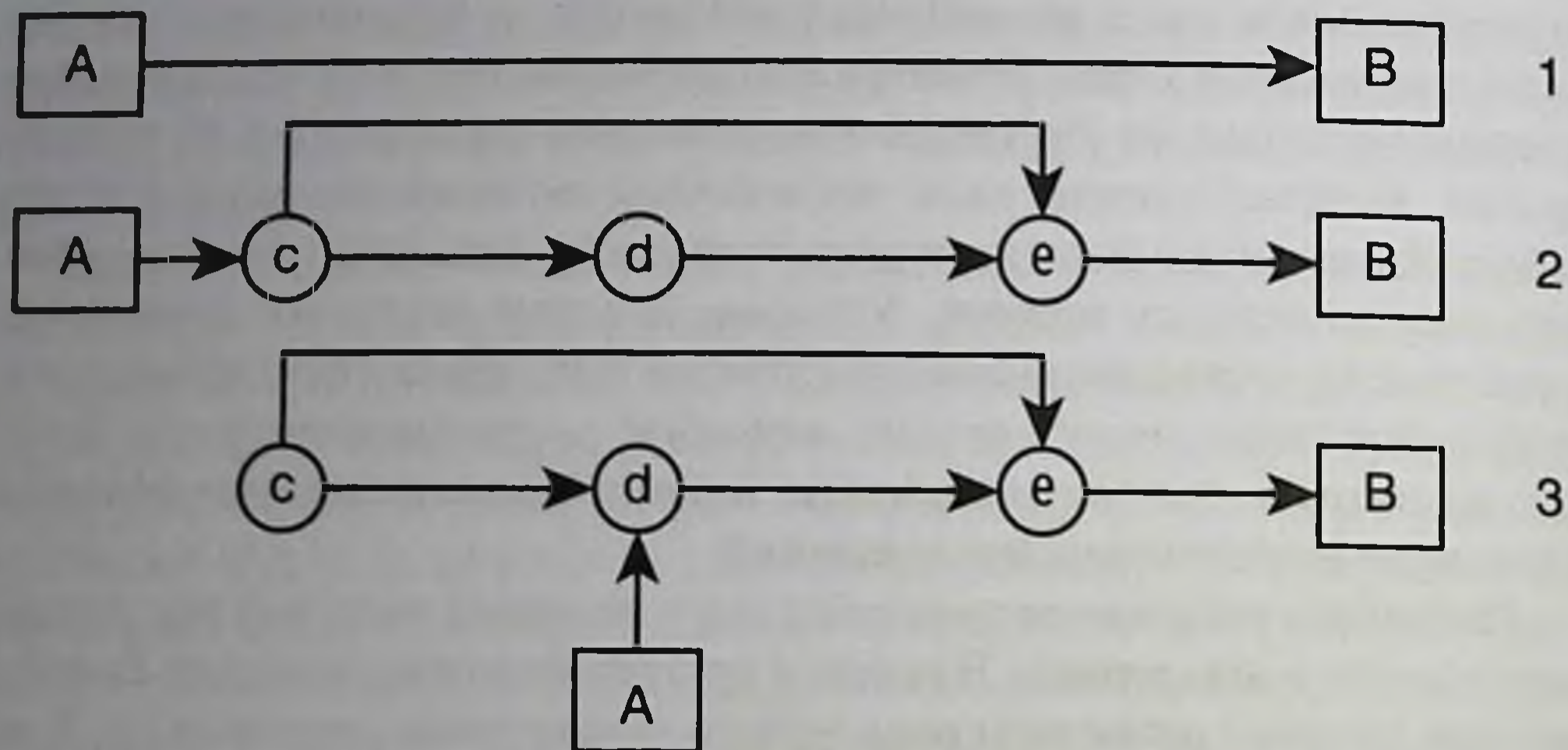


Рис. 1.2. Прямые и не прямые причинные отношения. 1 — фактор А может быть прямо действующей причиной болезни В. 2 — фактор А может действовать через запускаемую им цепь событий с, d, e. 3 — фактор А может модифицировать «естественные» отношения факторов с, d, e, существующих независимо от него

замедляющие развитие болезни, являются столь же полноценными вмешательствами, как и профилактика возникновения болезни. Более радикальных вмешательств не бывает: всегда, когда говорят, что «смерть предотвращена», в действительности имеют в виду лишь ее отсрочку.

### 1.1.2. Причинность в медицине

В медицине представление о компонентных причинах выражается в виде понимания этиологии как процесса взаимодействия организма с комплексом внешних факторов. В свою очередь, особенности организма в момент внешнего воздействия есть результат наследственных особенностей организма и процесса развития в конкретных условиях (рис. 1.3). Благоприятные внешние условия могут позволить организму развиваться нормально, несмотря на наличие врожденной особенности, например неспособности к усвоению глютена, но вынужденная смена диеты (внешнее воздействие) может нарушить благополучное существование организма и привести к болезненным проявлениям.

Иногда применительно к причине болезни используется выражение «этиологический фактор», чем подчеркивается существование един-



ственного фактора, который собственно и является настоящей «причиной», этиологией (от греч. *aitia* — причина). Такому пониманию причинности, характерному для XIX — первой половины XX в., свойственно наряду с причиной признавать возможность второстепенных условий, которые не равны причине, а лишь способствуют ее реализации. Обычно вместе с упрощенным пониманием причины поддерживается иллюзия существования простой пары «одна причина — одна болезнь». Традиционно в медицине наряду с этиологией выделяют патогенез болезни, т.е. механизм ее развития. В разное время этиология и патогенез рассматривались более или менее отдельно, но в настоящее время их разграничение весьма условно [2].

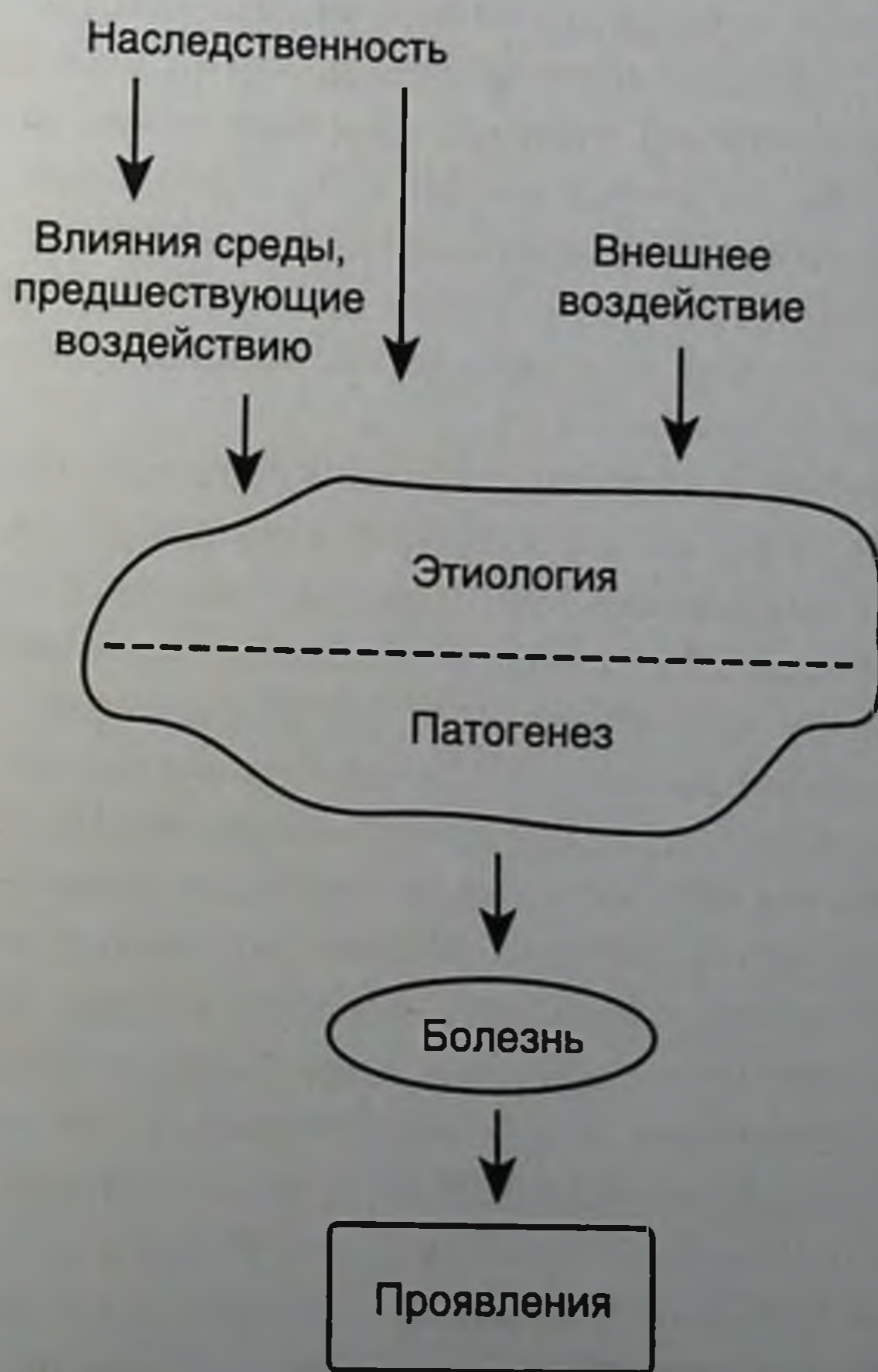


Рис. 1.3. Генез болезни и ее проявлений



История медицины изобилует примерами того, как понимание причины болезни с годами радикально изменялось, иногда неоднократно. Это не означает, что медицина не способна устанавливать истинные причины или что она не может получать убедительных доказательств в эксперименте. Медицина, так же как любая наука, имеет ограниченные возможности в установлении истины. Не только медицинские и эпидемиологические исследования, но и любые экспериментальные и неэкспериментальные исследования имеют ограниченные возможности. Получаемые результаты дают основания устанавливать связи, но найденные связи и их интерпретации могут быть приблизительны или неверны, даже если эксперимент проведен без больших ошибок. Самый яркий пример — расхождения в результатах клинических испытаний одного вмешательства. В одном испытании обнаруживается снижение летальности (доли случаев болезни, закончившихся смертью) при использовании вмешательства на 5%, в другом — на 8%, а в третьем — на 17%. Приблизительность оценок обычно не препятствует практическому применению знаний, но требует от исследователей и практиков критического отношения к результатам исследований других авторов, собственным исследованиям и к собственной практике.

Повторное наблюдение цепи событий действие—результат для обыкновенного ума кажется доказательством причинной связи. В действительности результат (выздоровление) может не зависеть от действий врача (инъекции лекарства). В научных исследованиях применяется традиционный прием проверки гипотез. Он состоит в формулировании предположения, противоположного основной гипотезе исследователя. Это противоположное предположение называют *нулевой гипотезой*. Нулевую гипотезу проверяют средствами опыта и статистического анализа данных. Если нулевую гипотезу удастся отклонить, то считается, что есть основания принять основную гипотезу. Однако не всегда гипотеза, подтвержденная таким образом, единственная. Обычно с большим или меньшим основанием можно предположить, что выявленная связь объясняется другими причинами. Например, обнаруженная ранее связь применения эстрогенов женщинами в менопаузе с низкой частотой ишемической болезни сердца была сильной. Имелись удовлетворительные объяснения механизмов этой связи. В последние 20 лет XX в. гормонозаместительная профилактика в менопаузе была широко распространена. Однако клинические испытания, завершённые в 2002—



2003 г., выявили *повышение* смертности от сердечно-сосудистых болезней у женщин, принимающих эстроген-прогестиновые препараты по сравнению с принимавшими плацебо (недействующая имитация препарата). Наиболее правдоподобным является следующее объяснение ранее обнаруженной связи: не гормонозаместительная терапия предотвращает сердечно-сосудистые болезни, а женщины с низкой вероятностью этих болезней чаще применяют гормональные препараты. Низкая вероятность сердечно-сосудистых болезней может объясняться тем, что женщины, применяющие гормональные препараты, обычно имеют выше доход и образование, лучше питаются, физически более активны.

### 1.1.3. Признаки причинного характера связи

Классические представления о причинности в медицине сформировались в эпоху бактериологических открытий. В конце XIX в. у больных инфекционными болезнями выделяли множество микроорганизмов, и каждый из них рассматривался как возможная причина болезни. Были необходимы критерии, которые позволяли бы отделять патогенный агент (причину) от множества других микроорганизмов, всегда присутствующих у людей и животных. Критерии, которые отделяли бы случайные находки микроорганизмов, сопутствующих болезни, от находки микроорганизма, ее вызывающего, т.е. причинно связанного с болезнью. Яснее всего критерии представлены в триаде постулатов, предложенных Я. Генле (F.G. Jacob Henle) и позднее измененных Р. Кохом (R. Koch). Связь микроорганизма и болезни можно считать причинной, и микроорганизм можно признать причиной болезни, если:

- ▶ микроорганизм *всегда* обнаруживается у лиц с изучаемой болезнью;
- ▶ микроорганизм *не* обнаруживается у других лиц;
- ▶ микроорганизм, выделенный из материалов, полученных от больных, и культивировавшийся в лабораторных условиях, вызывает болезнь в эксперименте.

В этой триаде первый элемент является признаком *необходимой причины* (присутствие микроорганизма обязательно для развития болезни), а последний — признаком *достаточной причины* (инфицирование вызывает болезнь). Применение этих критериев было успешным во многих случаях, но при ряде болезней встречались труднопреодолимые



препятствия. Например, выделение чистой культуры возможно только в том случае, когда имеется подходящая для микроорганизма питательная среда. Потребовались долгие поиски подходящей среды для ряда возбудителей, и ныне при изучении новых возбудителей создание способа культивирования — большая проблема. Избежать этапа выделения культуры нельзя, так как возбудитель только тогда существует для исследователя, когда его можно сохранять и изучать.

Следующая трудность в применении третьего критерия состоит в том, что многие болезни человека не воспроизводятся у животных, и применение третьего критерия в эксперименте затруднено. Более того, у людей и животных патогенные микроорганизмы не всегда вызывают болезнь (иммунитет<sup>1</sup>). Поскольку явление иммунитета было не вполне ясным в конце XIX в., сомневающийся врач мог выпить культуру предполагаемого возбудителя болезни. Если он не заболел, то микроорганизм не признавался возбудителем данной болезни. Нередко встречающийся естественный иммунитет людей к холере тогда не был известен, и самозаражение культурой вибриона без видимых последствий долго мешало признанию его возбудителем холеры. Поправка Коха к триаде Генле объясняла ситуацию, в которой болезнь не вызывается (не может быть вызвана) у человека и животных. В этом случае строгое соответствие критериям 1 и 2 признавалось достаточным для доказательства причинного характера связи. До середины XX в. экспериментальное подтверждение патогенного действия рассматривалось как важнейшее и окончательное доказательство причинно-следственных отношений. Соответственно эксперименты по заражению людей предполагаемыми возбудителями проводились довольно широко. Необходимость ограничивать эксперименты на человеке была широко признана только после Второй мировой войны.

Поскольку в XX в. проблемы неинфекционной патологии стали занимать большее место, возникла необходимость устанавливать причины неинфекционных болезней. Важным толчком послужила дискуссия о связи курения и рака легких. Для выработки политики в отношении курения главный врач США назначил комиссию, подготовившую отчет «Курение и здоровье» в 1964 г. Отчет содержал список признаков

---

<sup>1</sup> Иммунитетом называется невосприимчивость организма к развитию инфекционного заболевания при контакте с микроорганизмом. Иммунитет может иметь в своей основе различные механизмы.



причинного характера связи, использованных комиссией для оценки связи рака легких и курения, основанный на идеях А. Хилла (A.V. Hill). А. Хилл в 1965 г. опубликовал расширенный список признаков, которые позднее стали известны как «критерии причинности». Эти признаки не являются критериями, т.е. по ним нельзя принимать решение в соответствии с установленным алгоритмом и с известной надежностью. Они являются схемой анализа, осмысления имеющихся данных. В 1990-х годах больше внимания стали уделять тому, каким способом были получены данные о причине болезни или эффективности лечебного вмешательства (с использованием какой структуры исследования, подробнее см. разд. 1.2.2).

### **Фенфлурамин и поражение клапанов сердца**

Событие не является причиной само по себе, независимо от следствия, и не существует для сознания человека как причина, пока не известны следствия, т.е. пока не образуется гипотеза.

Например, препарат фен-фен, применявшийся для лечения ожирения, вызывал поражения клапанов сердца (см. с. 54). Производитель препарата компания Wyeth сообщила, что пострадали более 45 000 человек. После установления этих побочных эффектов появилось объяснение: эндокардиальную фиброплазию клапанов вызывает серотонин, и ответственным за этот эффект является фенфлурамин (компонент фен-фен), который, в отличие от других препаратов, подавляющих аппетит, имеет выраженные серотонинергические свойства. Это объяснение было возможно до применения препарата, так как избыток серотонина при карциноидном синдроме (опухоль желудка, продуцирующая серотонин) сопровождается поражениями клапанов, и то же бывает при интоксикации эрготамином, имеющим серотонинергические свойства. Однако это объяснение не было востребовано, пока явление не было обнаружено. В ходе клинических испытаний поражения клапанов не выявили, что позволяло признать препарат безопасным и разрешить его к применению.

В других случаях знания о типовых последствиях воздействий, обладающих сходными свойствами, могут направить поиск возможных следствий с самого начала. Например, легочная гипертензия — другое побочное действие фен-фена — была обнаружена в ходе клинических испытаний. Механизму ее появления сразу было найдено приемлемое объяснение, и по этой причине применение препарата было рекомендовано ограничить «болезненным» (тяжелым) ожирением и не назначать препарат дольше 6 мес подряд.



### Основные признаки причинного характера связи

**А. Временные соотношения экспозиции и заболевания.** Экспозиция<sup>1</sup> может быть причиной болезни (состояния) только в том случае, если предшествует ей. Положение о том, что причина предшествует следствию, является интуитивно ясным и только отражает общее определение причины. Это положение очень часто не оправдывается. Если причина не является необходимой и достаточной, наблюдаются случаи<sup>2</sup>, когда экспозиция отсутствует или следует за развитием болезни, потому что болезнь вызвана другими причинами. Знание временных отношений может помочь только в *исключении* причинной связи. Так, экспозиция к тератогену на поздних сроках беременности не может быть причиной рождения ребенка с дефектом развития нервной трубки.

**Б. Биологическая объяснимость связи экспозиции и заболевания.** Знание биологического механизма, объясняющего связь экспозиции и заболевания, укрепляет мнение о причинно-следственном характере связи. Однако связь может быть причинной несмотря на отсутствие объяснений. В медицине принято начинать исследование состояния, например редкой врожденной болезни, с наблюдений и эпидемиологических данных, а заканчивать изучением механизмов ее наследования и сущности «поломки» метаболизма. Кроме того, имеющиеся объяснения не всегда правильны. Так, одним из первых объяснений происхождения синдрома приобретенного иммунодефицита (СПИД) было объяснение его употреблением гомосексуалами амилнитрита. А. Хилл выделял также в качестве признака причинности согласованность объяснений предполагаемой причинной связи с другими данными о развитии заболевания (англ. *coherence*). Фактически этот признак — вариант

---

<sup>1</sup> Экспозицией в эпидемиологии называют контакт человека с фактором, который предположительно вызывает заболевание, например контакт с химическим веществом на производстве или курение. Термин «экспозиция» относится также к количественной характеристике внешнего воздействия. В отличие от экспозиции, термином «доза» обозначают количество вещества или энергии, поглощенное организмом (поглощенная доза излучения, поступившая в организм доза вещества). Экспозиция может быть не только вредной, но и полезной, например экспозиция к вакцине. На этом примере видно, что экспозиция может одновременно вызывать полезные эффекты (иммунитет) и вредные (абсцесс).

<sup>2</sup> Случаями в эпидемиологии называют отдельных пациентов с изучаемым заболеванием, по аналогии с клиническим описанием случаев заболевания. При эпидемиологическом сравнительном анализе группа случаев может сравниваться с группой контрольных пациентов (не имеющих заболевания).



признака «объяснимость связи». Отсутствие других данных, согласующихся с гипотезой, не является свидетельством против нее. Наличие противоречащих гипотезе данных снижает ее правдоподобие, но всегда есть вероятность, что эти данные ошибочны или неправильно интерпретируются. Придавая большое значение объяснимости наблюдаемой связи, А. Хилл выделял также наличие *аналогий* в качестве признака истинной причинно-следственной связи. Можно считать, например, радиационный генез одного типа рака подтверждением связи другого вида рака с облучением. К сожалению, аналогии возможны для всего, и их наличие зависит от воображения исследователя.

**В. Подтверждаемость обнаруженной связи.** Ни одно исследование не может дать единственного и окончательного решения проблемы. Накопление согласующихся результатов исследований в различных популяциях и в разных условиях дает больше уверенности в причинно-следственном характере связи, чем расходящиеся результаты. Совпадающие результаты подтверждают друг друга. Несовпадающие результаты должны быть объяснимы, и объяснимые расхождения подтверждают предполагаемый механизм. К сожалению, возникновение эффекта только в одних условиях (при наличии других компонентных причин) и невозникновение в других условиях встречается часто, и не всегда эти определяющие условия понятны. Различные исследования подтверждают результаты друг друга только в том случае, когда в экспериментах удастся учесть все детали механизма причинности, что достигается редко.

**Г. Альтернативные объяснения связи экспозиции и заболевания.** Убедительность гипотезы о причинной связи экспозиции и болезни подрывается наличием иных объяснений наблюдаемого эффекта, т.е. возможностью объяснить эффект вмешивающимися факторами. Например, если большее долголетие женщин, принимающих эстрогены в климактерическом периоде, можно объяснить особенностями их образа жизни, образования и поведения, то объяснение благоприятным действием эстрогенов на метаболизм становится сомнительным. Сомнение можно разрешить лишь в контролируемом испытании. Это было сделано после долгих лет широкого применения эстрогенов без уверенности в их пользе. Испытания показали, что эстрогены в лучшем случае бесполезны. Прямолинейное объяснение различий в заболеваемости и долголетию женщин, принимающих эстрогены и не принимающих их, действием эстрогенов на метаболизм было неверным.



### Вспомогательные признаки причинного характера связи

**А. Правильная форма зависимости экспозиция—эффект.** Если изучаемая экспозиция является причиной болезни, то с увеличением экспозиции (интенсивности и продолжительности воздействия) *обычно* увеличивается частота эффекта (развития болезни). Вместе с увеличением частоты *обычно* ускоряется развитие эффекта (болезни) и увеличивается тяжесть поражения. Зависимость экспозиция—эффект бывает линейной или S-образной не для всех исходов. Для многих экспозиций наблюдается J-образная зависимость. Например, смертность минимальна у употребляющих алкоголь в умеренных количествах, но повышена у непьющих и у пьющих много. Обычно происхождение таких зависимостей неизвестно. Возможно, исходы (смерть в этом примере) при низкой экспозиции и исходы при высокой экспозиции имеют разные механизмы развития. Наличие правильного биологического градиента не доказывает, что связь экспозиции и эффекта носит причинный характер, так как не исключает других причин. Например, некоторые неблагоприятные исходы закономерно учащаются у людей, которые выкуривают больше сигарет за день, но причиной этих исходов является не курение, а другие особенности курильщиков, выраженные тем больше, чем больше они курят (злоупотребление алкоголем, нездоровое питание, низкое образование). Отсутствие правильной зависимости доза—эффект отвергает только гипотезы, предполагающие такую зависимость.

**Б. Сила связи экспозиции и эффекта.** Сила связи экспозиции с заболеванием измеряется размером различия частоты заболевания в группе, подвергавшейся экспозиции, и частоты заболевания в контрольной группе, не подвергавшейся экспозиции. Если изучается предполагаемая причина болезни, то вероятность заболевания при экспозиции повышается, а если это терапевтическое (профилактическое) вмешательство, снижается. Слабая связь не исключает причинно-следственных отношений, а сильная связь может быть следствием действия привходящих факторов. В любом случае наблюдаемая сила связи между экспозицией и заболеванием есть не биологическое свойство, а лишь изучаемое проявление свойств популяции. Связь может быть сильной в одной популяции и слабой в другой — в зависимости от распространенности других компонентных причин. А. Хилл выделял также в качестве признака причинности специфичность связи: истинная причинная связь — чаще связь экспозиции с отдельной болезнью, а не со многими болезнями. В конце XX в. защитники производителей табака использо-



вали этот признак причинности для того, чтобы поставить под сомнение вредность курения табака, поскольку курение приводит к многообразным вредным последствиям, а не к специфической болезни. Позднее этот признак причинности был отвергнут, так как вредная экспозиция обычно приводит к множеству неблагоприятных последствий.

**В. Исчезновение эффекта при прекращении экспозиции.** Если экспозиция является причиной болезни, то ее прекращение должно приводить к уменьшению частоты новых случаев болезни. Нужно иметь в виду, что после воздействия изменение заболеваемости может проявляться через долгие годы. Так, через многие годы развивался рак влагалища у девочек, подвергшихся *in utero* воздействию диэтилстильбэстрола, применявшегося (без достаточных оснований) для лечения угрожающего аборта. Эксперимент на людях с использованием потенциально вредной экспозиции сейчас считается недопустимым. Применение профилактических воздействий в виде прекращения потенциально вредной экспозиции приемлемо. Экспериментальная проверка гипотезы может дать убедительные результаты в виде снижения частоты заболевания при прекращении экспозиции, но они не являются исчерпывающим доказательством причинно-следственного характера связи болезни с экспозицией.

Ни один из приведенных признаков причинного характера связи не является необходимым или достаточным, за исключением признака «правильные временные отношения», который является частью определения причины. А. Хилл не употреблял выражения «критерии причинности», и в настоящее время нельзя представить себе набор критериев, который позволял бы устанавливать, является ли экспозиция причиной заболевания. Приведенные признаки помогают осмыслить существующие данные о болезни и возможных ее причинах и планировать новые исследования для испытания наиболее правдоподобных гипотез. Такое последовательное обобщение имеющихся данных и испытание новых гипотез в исследованиях продвигают медицину в познании влияний среды на здоровье человека и в оценке эффективности медицинских вмешательств.

#### 1.1.4. Факторы риска

Экспозиция организма причинному фактору<sup>1</sup> вызывает развитие болезни. Возникновение болезни протяженно во времени. Причинные

---

<sup>1</sup> Фактором в эпидемиологии называют событие или признак, рассматриваемый в связи с развитием болезни (состояния). Чаще используют выражение



факторы действуют в период *индукции*. В этот период болезни еще нет, например, нет опухолевых клеток, обладающих характерными свойствами (рис. 1.4). В эпидемиологическом исследовании можно получить данные о периоде индукции и последующих событиях только относительно конкретной причинно-следственной пары, например употребления загрязненной афлатоксинами пищи и раком печени. Рак печени сам по себе, безотносительно некоторой причины, не имеет специфического периода индукции. С возникновением болезни, выражающейся появлением диагностических признаков, период индукции заканчивается. Например, по прошествии некоторого времени после экспозиции в ткани обнаруживаются клетки со специфическими изменениями или иные характерные для болезни проявления.

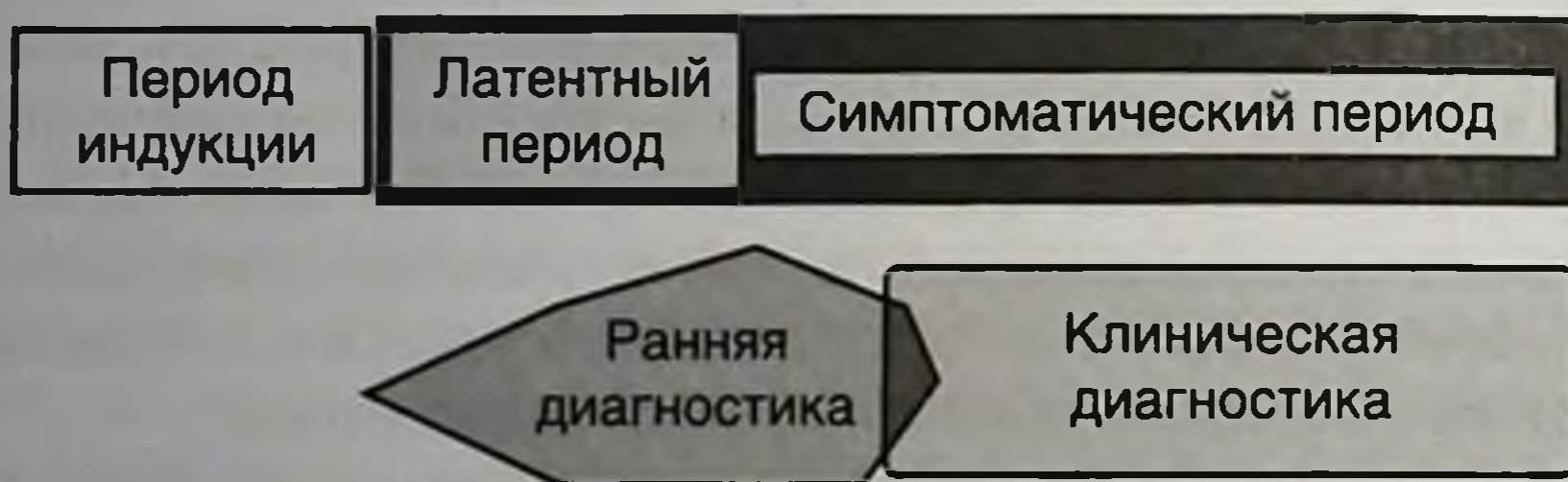


Рис. 1.4. Периодизация болезни и ее проявлений

Болезнь не сразу становится явной. *Латентный период* болезни (инаппарантный или бессимптомный) может быть более или менее длительным<sup>1</sup>. За латентным периодом следует *симптоматический период*, т.е. период, в котором болезнь проявляется внешне заметными признаками. Клиническая диагностика обычно осуществляется позднее, чем появляются симптомы болезни. Выявить болезнь ранее, чем

«фактор риска», подчеркивая, что этот признак связан с повышенной вероятностью заболевания. Использование термина «фактор» подчеркивает неопределенность роли рассматриваемого фактора в развитии болезни. В статистическом анализе «фактор» является синонимом независимого признака, отдельным значениям которого может присваиваться цифровой код. Например, различным значениям признака «суточное употребление алкоголя» может присваиваться значение от 1 до 4, и в дальнейшем сопоставляются группы с разным значением этого фактора. Сопоставление групп проводится по другим признакам, например по заболеваемости, и такие признаки в отличие от независимого признака группировки (независимого фактора) называются зависимыми.

<sup>1</sup> Периодизация инфекционного заболевания несколько сложнее (см. разд. 4.1).



пациент сам обратится за помощью, можно с помощью специальных методов (подробнее см. главу 2.2).

Изучая совокупность факторов, предположительно влияющих на развитие болезни, исследователь рассматривает их влияние в рамках своих представлений о возможных причинах заболевания. «Кандидаты» на роль участвующих причин называются *факторами риска*<sup>1</sup> заболевания. Такое название используется, поскольку эти факторы (условия, экспозиции, признаки) связаны с вероятностью развития (риском) заболевания и отвечают некоторым признакам причинности, но их роль окончательно не ясна. Для многих болезней, например атеросклероза, злокачественных новообразований, алкоголизма, описаны сотни факторов риска. Они более или менее интенсивно изучаются. Значение некоторых из них подтверждается в независимых исследованиях. Наконец, наиболее изученные факторы риска становятся объектом экспериментального исследования — профилактического изменения того или иного фактора. Снижение частоты болезни при устранении фактора риска оказывается наиболее доказательным этапом в выяснении роли фактора риска в генезе заболевания. Так, повышенный риск атеросклероза и связанной с ним смертности от сердечно-сосудистых заболеваний и общей смертности у лиц с гиперхолестеринемией был надежно показан еще в середине XX в. и многократно подтвержден (рис. 1.5), но лишь в конце XX в. разработка действенных средств против гиперлипидемии позволила доказать, что модификация этого фактора риска действительно способна снизить вероятность развития атеросклероза<sup>2</sup>.

Концепция факторов риска позволяет рассматривать любые признаки (например, оволосение ушной раковины) и любые экспозиции (например, работа на цинковом производстве, обучение в музыкальной школе, использование красителей для волос). Это особенно важно для эпидемиологических исследований, в которых развитие заболевания рассматривается в совокупности всех условий — особенностей экспо-

---

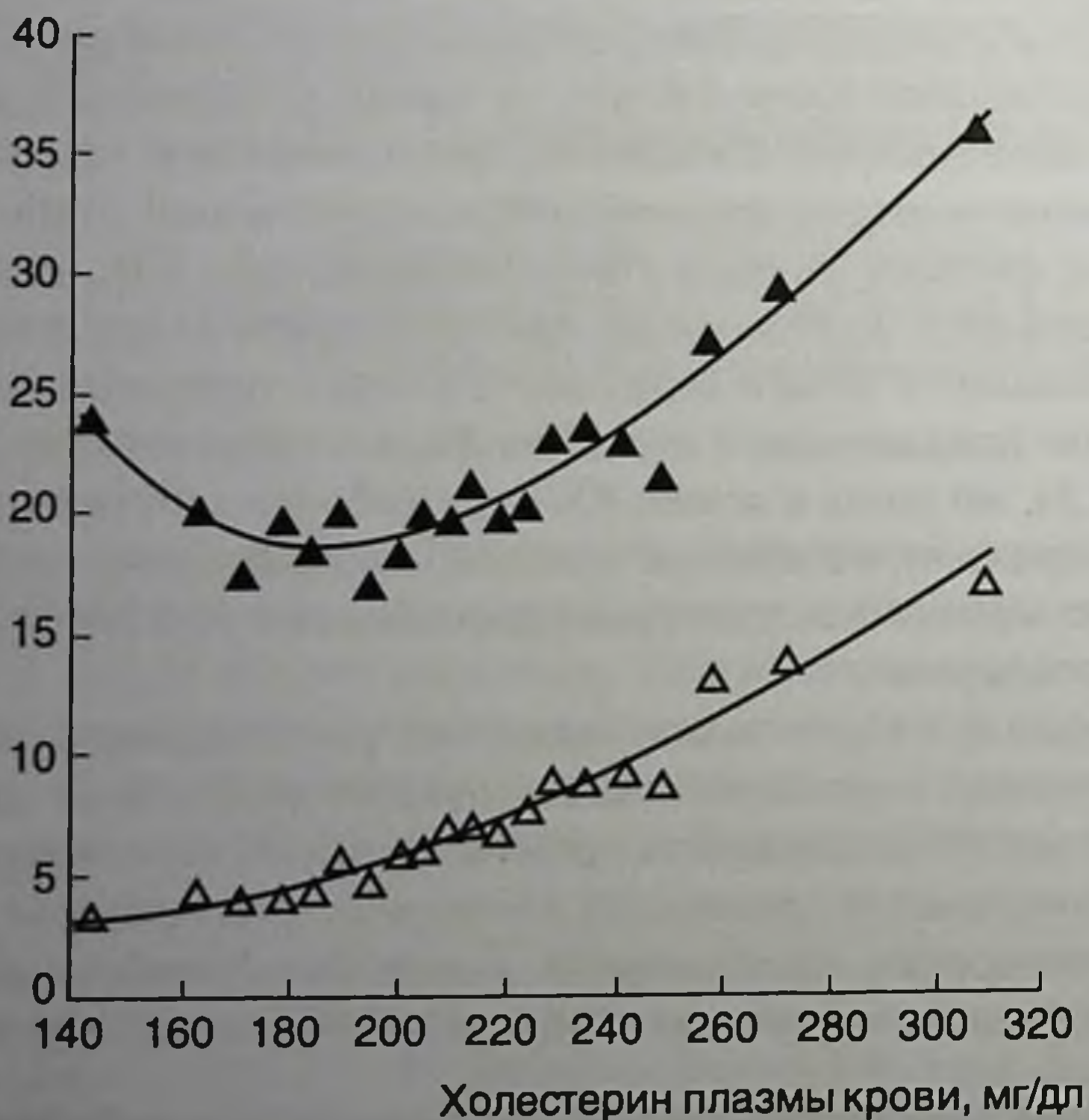
<sup>1</sup> В эпидемиологии риском называют вероятность развития события, обычно неблагоприятного, в течение некоторого периода времени.

<sup>2</sup> Остаются некоторые сомнения в том, что гиполипидемические препараты снижают частоту осложнений атеросклероза, именно изменяя липиды крови. Возможно, что механизм их действия иной, однако их действенность надежно доказана. Предполагалось, что концентрация холестерина в плазме крови может быть снижена, если не употреблять в пищу продукты, богатые холестерином. Полезность такого изменения диеты никогда не была доказана в контролируемом испытании.



зиции (природных и антропогенных), социальных факторов, индивидуальных особенностей людей. Рассмотрение всевозможных факторов риска важно потому, что исследователь может ничего не знать об этиологии болезни или иметь лишь приблизительные гипотезы.

Широкое использование исследований, «просеивающих» множество факторов риска, вызвало у некоторых специалистов ощущение, что эпидемиология середины XX в. превратилась в мельницу для перемалывания статистических связей «признак—болезнь». Исследования такого типа необходимы, так как выявление факторов риска позволяет целенаправленно генерировать гипотезы. Гипотезы затем можно проверить в более сложных исследованиях, например, в эксперименте изучая происхождение болезней, их профилактику и терапию.



**Рис. 1.5.** Смертность от ишемической болезни сердца (светлые треугольники) и общая смертность (темные треугольники) в зависимости от концентрации холестерина в плазме крови в начале исследования у мужчин в исследовании MRFIT. Смертность за 6 лет стандартизована по возрасту и приведена в расчете на 1000 человек (ордината). Данные сгруппированы по 1/20 от исходной когорты, составлявшей 361 662 мужчины (по Martin M.J. et al. Lancet. 1986. № 2. P. 933–6)



В некоторых случаях возникают трудности в определении характера связи признака с последующим заболеванием. Для внешних условий, например проживания в помещении, построенном с использованием асбестосодержащих материалов, обнаружение связи этой экспозиции с развитием мезотелиомы<sup>1</sup> дает веские основания полагать, что экспозиция асбесту является причиной болезни. Для «внутренних» особенностей человека, биологических и поведенческих, выяснение характера связи фактора риска и болезни сложнее. Так, малоподвижный образ жизни сопровождается повышенной вероятностью развития клинически явной ишемической болезни сердца и обычно рассматривается как фактор риска ишемической болезни сердца. Вместе с тем трудно исключить, что стремление избежать физических нагрузок является ранним симптомом ишемической болезни сердца.

Факторы риска, как любые условия, не обязательно специфичны для болезней, с которыми они связаны. Например, артериальная гипертензия обычно рассматривается как фактор риска сердечно-сосудистых болезней и их осложнений, но одновременно она является фактором риска рака (рис. 1.6).

Факторы риска — это обычно признаки, связанные с повышенной заболеваемостью. Однако могут быть и экспозиции (признаки), снижающие заболеваемость. Так, регулярный прием ацетилсалициловой кислоты (Аспирин) сопровождается повышенной вероятностью желудочно-кишечных кровотечений, но одновременно вероятность ишемического инсульта и рака толстой кишки снижается. Такие признаки, связанные со снижением вероятности болезни, иногда называют «факторами антириска».

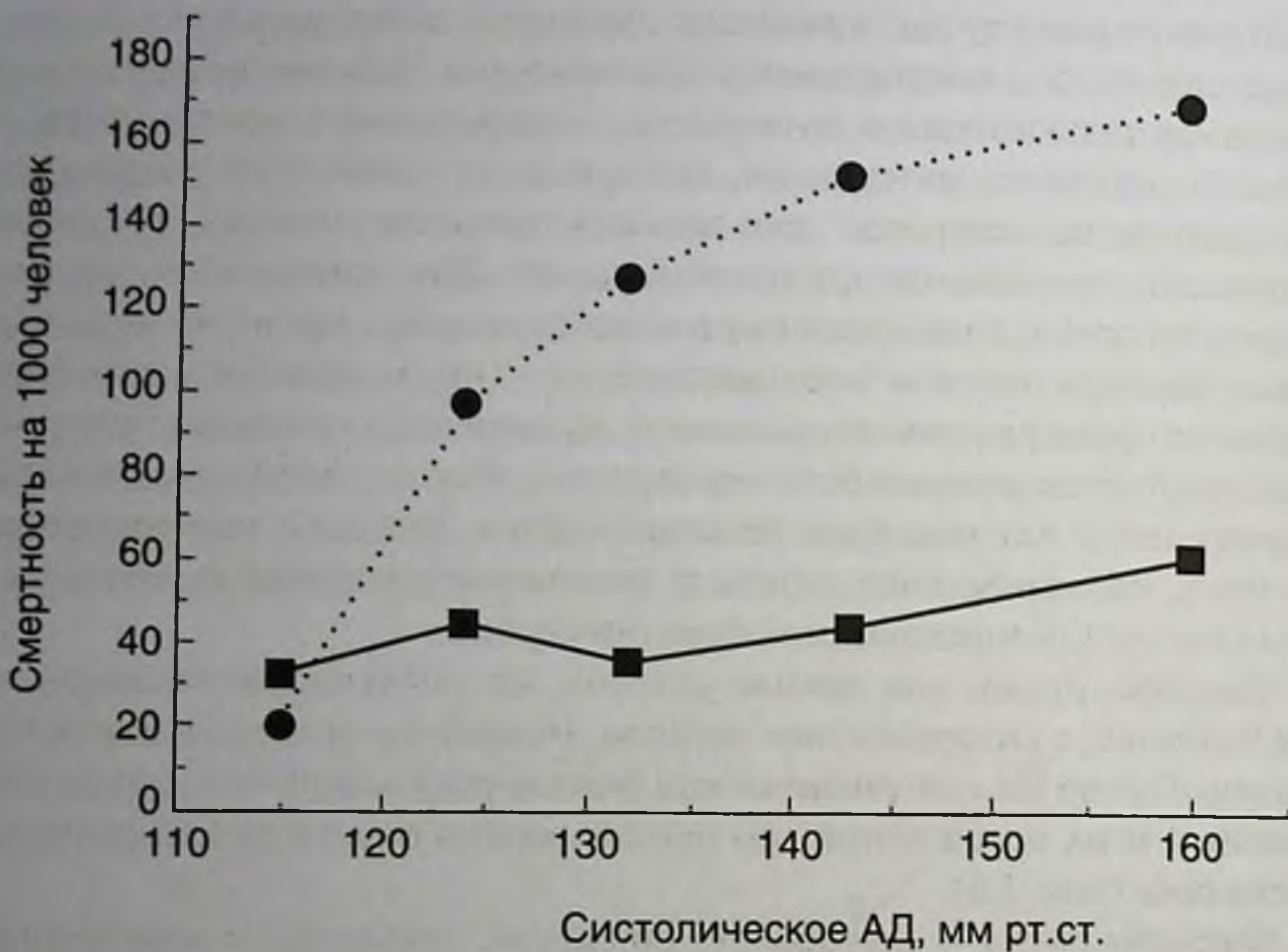
Факторы риска не привязаны к линейной модели развития болезни, представленной на рис. 1.4. Поскольку в процессе развития болезни могут наблюдаться не только обострения, но и ремиссии, например не только злокачественное изменение клеток эпителия, но и спонтанное их исчезновение, постольку могут быть исследованы факторы, влияющие на каждый из этих процессов.

Выражение «фактор риска» вышло за пределы эпидемиологии и стало частью обыденной речи, как стал частью общественной жизни учет выявленных в эпидемиологических исследованиях факторов риска.

---

<sup>1</sup> Мезотелиома — редкая опухоль (10–15 случаев на 100 000 населения в год в Европе), поражающая плевру и исходящая из клеток мезотелия. Возникновение обычно связано с экспозицией асбесту, что было установлено в 1960 г.





**Рис. 1.6.** Смертность от рака в зависимости от артериального давления (АД) в проспективном исследовании большой группы мужчин старше 40 лет (линия) и старше 59 лет (пунктир) (по данным Goldbourt U. et al. J. Nat. Cancer Inst. 1986. № 77 (1). P. 63–70)

Миллионы людей бросают курить и стараются предотвратить курение у детей, занимаются физическими упражнениями. Снижение смертности от сердечно-сосудистых болезней в экономически развитых странах частично объясняется именно тем, что люди под влиянием эпидемиологических данных изменяют образ жизни. К сожалению, освоение этой информации обществом проходит небезболезненно. Так, под влиянием результатов нескольких исследований особые профилактические свойства приписывались шпинату. В 1980-е годы на основе результатов исследований типа сравнения с контролем (ИСК)<sup>1</sup> считалось, что упо-

<sup>1</sup> Исследованиями типа сравнения с контролем (ИСК) называют исследования, в которых случаи (больные изучаемым заболеванием) сравниваются с контролем (лицами без такого заболевания). Если у больных обнаруживается отличие от здоровых, например они в пищу употребляют мало овощей, появляется основание предполагать связь заболевания с дефицитом овощей или каких-то их компонентов. Подробнее см. разд. 2.8.



требление каротиноидов<sup>1</sup> снижает вероятность развития рака, и миллионы людей наращивали долю краснокрашенных овощей в пище, принимали препараты каротина. Лишь последующие контролируемые испытания выявили отсутствие профилактических свойств у каротина. Для того чтобы влияние эпидемиологических данных на общественную практику было более плодотворным, эпидемиологические исследования должны иметь более совершенную структуру, а их осмысление должно быть более критическим.

## 1.2. МЕТОДОЛОГИЯ НАУЧНОГО ИССЛЕДОВАНИЯ

Классическое научное исследование, не только медицинское, строится таким образом, чтобы устранить влияние всех случайных (посторонних) факторов, уравнивать все возможные особенности обследуемых лиц, кроме одного изучаемого фактора. Тогда можно предположить, что разница между обследованными людьми вызвана именно этим фактором. Например, если две одинаковые группы пациентов получают разное лечение и по-разному выздоравливают, можно предположить, что различие исходов связано с лечением. Эта модель научных исследований появилась в XVII в.

Примерно в то же время Ф. Бэкон заложил основы *индуктивизма* — направления философии, объясняющего процесс познания наблюдением за природой, формулированием гипотез и последующей их проверкой. Эта проверка подтверждает истинность знания или отвергает гипотезу. Такими можно представить действия Э. Дженнера, знавшего об успешных опытах прививания коровьей оспы для получения иммунитета к оспе и успешно поставившего эксперимент, сделавший оспопрививание известным. Эксперимент заключался в прививке коровьей оспы ребенку, а по прошествии недель — в прививке ему натуральной оспы. Последняя прививка не вызвала оспы, чем подтвердилась гипотеза о защитном действии прививки коровьей оспы против натуральной оспы.

У такого подхода к получению знания есть недостаток: повторение наблюдавшегося ранее хода событий используется как критерий истин-

---

<sup>1</sup> Каротин — предшественник витамина А (провитамин), получаемый человеком из пищи, в частности из желтых и красных овощей и фруктов. Поскольку витамин необходим для нормального обновления эпителия, гипотеза о том, что дефицит каротина увеличивает вероятность развития рака, была правдоподобной.



ности связи. Многие ученые, изучавшие процесс познания человеком мира — *эпистемологию*, были не удовлетворены тем, что научное знание подтверждается повторением. Повторение связи не означает причинного ее характера, как неизменный восход солнца не связан с предшествующим криком петуха.

Карл Поппер, труды которого оказали наибольшее влияние на философию науки XX в., внес в эту ситуацию относительную ясность: научные утверждения (теории) не могут быть обоснованы и признаны правильными только с использованием инструментов логики. Теории могут лишь согласоваться или не согласоваться с наблюдением. Поскольку результаты правильного наблюдения могут совпадать с несколькими гипотезами, постольку согласие гипотезы с наблюдением не подтверждает именно ее. Наоборот, несовпадение наблюдения с гипотезой отвергает гипотезу. По К. Попперу, наука прогрессирует, отвергая неверные гипотезы. Подтверждение гипотез лишь служит условием для их сохранения в научном обиходе. Ключевым признаком научности (приемлемости для науки) любого объяснения, любой причинной связи является их «фальсифицируемость»: только то объяснение научно, которое может быть подвергнуто опровергающему эксперименту. Это очень важное положение для медицины и парамедицинских «практик» и теорий: те из них, которые не проверяются и объявляются непроверяемыми, не могут включаться в научный обиход (гомеопатия, каналы тела человека в чжень-цзю терапии и т.д.).

Для последователей К. Поппера повторные и подтверждающие друг друга наблюдения не утверждают природные законы, а лишь создают психологический комфорт исследователю, помогая верить в то, что закон найден. Более того, ученые свыкаются с теориями и защищают их от критической проверки, что замедляет развитие в любой сфере науки.

Позднее, в середине XX в., Т. Кун обратил внимание на то, что сами наблюдения (факты) зависят от теории [3]. В истории науки много примеров того, как ученые не отказывались от установленной теории, несмотря на наличие наблюдений, противоречащих ей. Такие наблюдения рассматривали как исключения. Например, такое положение занимали факты более редкого, чем у некурящих, развития рака легких у курильщиков трубки, и более частого, чем у глубоко вдыхающих дым, развития рака у вдыхающих дым неглубоко. Обычно подобные исключения позднее находят объяснение в рамках установленной теории. Лишь при накоплении критической массы необъяснимых «исключений» ученое



сообщество отказывается от старой совокупности взглядов и объяснений (*парадигмы*), отвергая ее целиком или оставляя ей узкую область применения. Так произошло с ньютоновской физикой при принятии эйнштейновской теории относительности и квантовой механики. Неизвестно, справедливы ли эти наблюдения смены парадигм для будущего науки, или это лишь история науки, но несомненно, что радикально новые идеи утверждаются не постепенно, через изменение предшествующих взглядов, а путем коренного пересмотра большой совокупности взглядов. Большинство ученых, занимающихся повседневным анализом экспериментальных данных, действительно работают в рамках относительно узкой системы взглядов, занимаясь «нормальной наукой», а не совершая перевороты в понимании мира.

В XX в. классическая модель исследования «изучением единственной разницы» перестала быть единственной. Появились новые методы исследований и *статистического анализа данных*. Главным методом стал дисперсионный анализ, созданный Р. Фишером. В настоящее время в большинстве биомедицинских исследований рассматривается вопрос: не может ли связь явления В с особенностью А объясняться влиянием постороннего фактора С? Раньше для ответа на такой вопрос нужно было бы организовать специальное исследование, в котором пациенты сравниваемых групп были бы абсолютно одинаковы по признаку С. Теперь при планировании исследования заранее учитывают возможные посторонние факторы и затем при анализе данных «вычитают» их влияние. Это позволяет ответить на вопрос о том, самостоятельна ли связь В с А, или она может полностью объясняться действием фактора С. Например, пациенты с избыточной массой тела могут иметь большую летальность после инфаркта миокарда, но это различие может полностью объясняться тем, что у них выше концентрация холестерина в крови или они больше курят.

### 1.2.1. Данные, факты и знания

В изучении болезней эпидемиология последовательно применяет логические и математические приемы (методы) к анализу *данных*. Данные — это сведения о событиях (явлениях), полученные в ходе исследования. Они могут быть представлены в цифровой (результаты измерений, подсчетов), словесной (описательной) или графической форме (фотографии, рисунки или рентгеновские снимки). Обычно



термином «данные» определяют исходные, «сырые», необработанные сведения. Данные не следует путать с информацией и знаниями. *Информация* — это систематизированные и структурированные данные. В процессе предварительной обработки их могут преобразовывать, исправлять ошибки, вносить поправки на условия наблюдения и т.д.

*Знание* — это информация, находящаяся в контексте представлений, теорий, принципов. Знание включает не только результаты измерений (данные), обычно находящиеся в систематизированном виде (информация), но и интерпретацию этой информации. Например, достаточно точные измерения высоты Луны и Солнца над горизонтом были возможны тысячи лет назад, они систематизированы и дошли до нашего времени. Эти данные можно включить в систему знаний о Земле как центре вращения небесных тел, но можно включить и в систему знаний о Земле, вращающейся вокруг Солнца, и о Луне, вращающейся вокруг Земли. Основоположник научной эпидемиологии Вильям Фарр (William Farr) писал, что только кристаллизованные интеллектом факты составляют вечную правду науки.

Термином «факт» обычно обозначают достоверное знание о происшедшем событии или установленной характеристике события (например, «поезд прибыл в полдень»). *Факт* — нечто, точно известное. Факт не равен «данным» и «информации», поскольку обязательно включает в себя способ проверки и элементы интерпретации, в том числе на основании собственного опыта исследователя, т.е. является частью знания. В примере с поездом подразумевается, что мы знаем критерий «прибытия» (пересечение границы станции или остановка у перрона) и что известно, по какой шкале учитывается время (русские железные дороги работают по московскому времени, а самолеты приземляются по местному). Именно поэтому И.П. Павлов, предлагавший ученым «снимать шляпу» перед фактом, напоминал о необходимости ее потом надеть.

### 1.2.2. Структуры исследований

В целом исследования делят на *поисковые* (англ. *exploratory*) и *проверяющие гипотезу* (аналитические, англ. *analytical, confirmatory studies*). Пример поискового исследования — изучение сотен генетических маркеров в поиске тех, которые связаны с изучаемым синдромом. В таком исследовании *нет проверяемой гипотезы*, и, соответственно, его результаты не подвергают сложному статистическому анализу, а про-



сто рассматривают каждый признак: нет ли достаточно сильной связи. Результаты такого исследования не доказывают различия больных и здоровых по этому признаку, а лишь указывают на возможные связи болезни с геном, нуждающиеся в проверке.

Исследования, проверяющие гипотезу, используют методы статистических выводов. Для этих исследований *заранее формулируют гипотезу*. Например, предполагают различия между больными и здоровыми по некоторому маркеру и обследуют соответствующих больных и здоровых людей. Обнаружив различие, делают вывод о связи развития болезни с этим маркером. Результаты таких исследований являются доказательными.

Способы организации исследования называют *структурами исследования* (или дизайном исследования) (англ. *research structures, study design*). Структуры медико-биологических исследований являются важнейшим объектом изучения в эпидемиологии.

### Основные варианты структуры исследования

#### 1. Поперечные исследования

А. Описания спектра проявлений болезни.

Б. Описания диагноза болезни и стадии болезни.

В. Описания случаев.

#### 2. Продольные исследования

А. Проспективные:

- естественного развития заболевания, прогноза;
- причинных факторов заболевания и инциденса;
- неконтролируемого вмешательства;
- с преднамеренным вмешательством:

- параллельные;

- последовательные;

Б. Исследования типа сравнения с контролем.

В идеале в результате научного исследования некоторое воздействие можно признать причиной болезни, если болезнь удастся вызвать, подвергнув людей этому воздействию. В таком идеальном исследовании — *контролируемом эксперименте* — здоровые люди распределяются на две группы, члены одной из них подвергаются воздействию, а члены другой — нет. Если болезнь возникает только у подвергавшихся воздействию, его можно признать причиной болезни. По этическим причинам подобного рода эксперименты последние 50 лет почти не проводятся, за исключением сомнительных экспериментов разработчиков оружия.



Приемлемы лишь эксперименты, в которых к человеку применяются лечебные и профилактические воздействия. Этот тип эксперимента можно применить для проверки гипотезы о факторе, вызывающем болезнь: если в эксперименте устранение фактора приводит к выздоровлению (ремиссии) или устранение фактора прекращает возникновение новых случаев в группе людей, то гипотеза о причине может считаться подтвержденной.

Важно, что все *структуры исследований* базируются на представлениях о развитии заболеваний во времени и последовательности экспозиции и исхода (заболевания). Поэтому все структуры исследований в той или иной степени соответствуют модели контролируемого эксперимента. Чем полнее соответствие этой модели, тем меньше вероятность ошибочного результата.

В зависимости от того, какой вопрос ставится перед исследованием, оно может быть описательным (дескриптивным, от англ. *descriptive*; например, только описание спектра проявлений болезни) или аналитическим. Аналитическими называют исследования, которые проверяют гипотезы, например о связи заболевания человека с его профессией или с инфекцией. Аналитические исследования могут быть поперечными (например, симптом сопоставляют с наличием антител к вирусу) или продольными (типа сравнения с контролем, проспективными).

### **Поперечные исследования**

*Поперечными* (англ. *cross-sectional*) или одномоментными называют исследования, в которых каждого пациента обследуют один раз. В результате можно описать картину болезни у одного пациента или у совокупности больных, уточнить симптоматику, связать отдельные симптомы с диагнозом и тяжестью болезни, сопутствующими болезнями и состояниями, т.е. ответить на множество важных вопросов. Эти исследования широко применяются в клинике и в эпидемиологии. В простейшем случае поперечное исследование — это описание болезни, но не в ее развитии, а в совокупности признаков, вариантов, тяжести течения.

В прошлые века наблюдение за развитием болезней было основным методом их изучения. Для скоротечных инфекционных болезней этот метод позволял получить верные представления. В XX в. основным объектом медицинского исследования становятся хронические болезни. Обнаруживаются такие инфекционные болезни, которые манифестируют через годы после заражения. Длительность подобных заболеваний



сравнима с продолжительностью профессиональной жизни врача. Поэтому, просто наблюдая совокупность больных в течение относительно короткого времени, врач легко может совершить ошибку, например, приняв варианты течения болезни за ее стадии. Главным средством против таких ошибок является организация продольных исследований.

### Продольные исследования

*Продольными* (от англ. *longitudinal* — лонгитудинальные) называют исследования, в которых выделяется группа людей, за ними в течение некоторого времени наблюдают и повторно оценивают их состояние. Длительность наблюдения не обязательно должна быть большой. Для наблюдения естественного развития острого респираторного заболевания достаточно 10 дней. Однако если цель исследования — выявление возможных поздних осложнений, например осложнений беременности в связи с инфекцией, то увеличивается и срок наблюдения.

Если группа больных (или людей, проживающих в одной местности или обладающих общими признаками) специально формируется и целенаправленно отслеживается, то такое исследование называют *проспективным*. Например, большую группу студентов можно наблюдать в течение учебного года для изучения возможной связи пользования общественным транспортом с заболеваниями респираторными инфекциями. Наиболее сложны *популяционные проспективные (когортные)* (рис. 1.7) исследования. Для такого исследования выбирают и затем отслеживают большую выборку из популяции. В ходе наблюдения фиксируют изменения, обычно — возникновение новых болезней и их осложнений. Эти заболевания сопоставляют с исходными особенностями изучаемых людей. Такие исследования называют также *исследованиями причинных факторов*. Самыми известными являются большие популяционные исследования зависимости заболеваемости ишемической болезнью сердца от курения, концентрации холестерина и липопротеидов в крови и артериального давления.

Проспективные исследования обладают большой доказательной силой относительно причинно-следственного характера связей, обнаруживаемых между экспозицией и последующим заболеванием. Они не исключают всех возможных ошибок, но ближе всего подводят исследователя к пониманию причинно-следственных отношений. К сожалению, проспективные исследования очень дороги и поэтому предпринимаются редко.



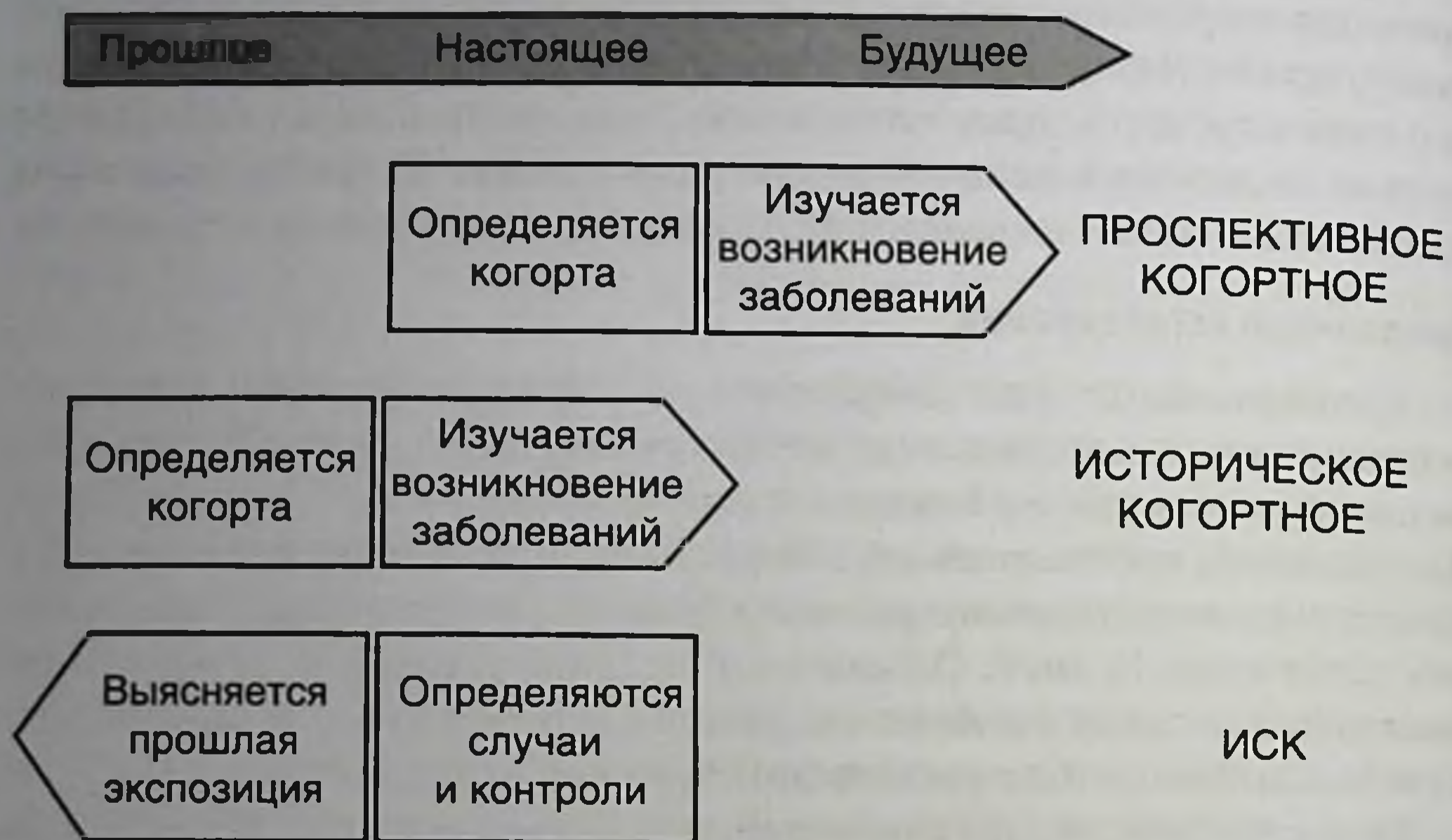


Рис. 1.7. Сходство и различия между когортными исследованиями и исследованиями типа сравнения с контролем при изучении происхождения заболевания. ИСК — исследования типа сравнения с контролем

Продольное исследование можно выполнить путем анализа данных, имеющихся в историях болезни, иных документах, накапливаемых последние десятилетия базах данных. Это называется *ретроспективным сбором данных*. Например, по архивным данным можно определить массу тела при рождении и оценить ее связь с последующими заболеваниями (базы данных обращений за медицинской помощью), приемом лекарств (базы данных оплаты лекарств) и другими исходами. Если удастся найти подавляющее большинство членов исходной группы и оценить исходы, то результаты такого сравнительно простого исследования (продольного по ретроспективно собранным данным) получатся высокого качества, т.е. они будут схожи с результатами проспективного исследования, если бы оно было проведено. Недостатком ретроспективного сбора данных является невозможность изменить методику первичного обследования — она целиком определяется практикой, существовавшей в момент заполнения используемых документов. Например, во всем мире пациентов при госпитализации не взвешивают, чаще спрашивают пациента, какая у него масса тела. Достоинствами являются оперативность такого исследования и дешевизна. Например, в таком



исследовании показано, что люди, родившиеся с меньшей массой тела, по крайней мере в детском и подростковом возрасте имеют худшие показатели выполнения математических тестов (рис. 1.8).

Наиболее важны для эпидемиологической практики *исследования типа сравнения с контролем (ИСК)* (см. разд. 2.8). В ходе таких исследований получено больше всего сведений о потенциально вредных экспозициях, о внешних и о внутренних факторах риска, в том числе о генетической предрасположенности. Группа больных (случаи) в таких исследованиях выделяется в конечный момент — при выявлении болезни. Одновременно подбирается контрольная группа. В прошлом можно выделить у больных какой-то общий фактор, например курение, использование краски для волос или ультрафиолетовых ламп для загара. Если в контрольной группе в прошлом этот фактор встречается реже, то можно предположить связь заболевания с этим фактором. В клинике ИСК являются современной формой научного осмысления медицинской практики.

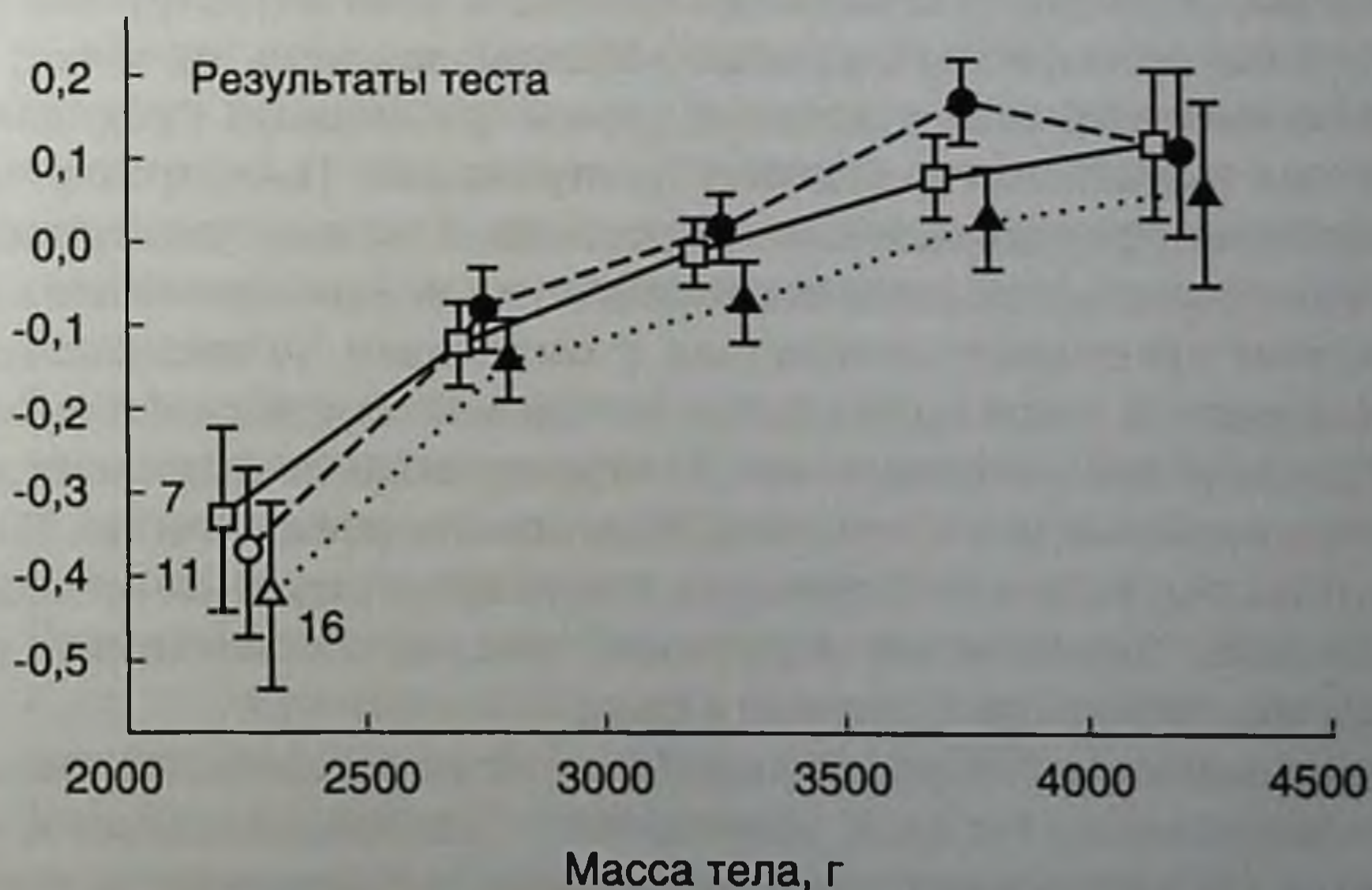


Рис. 1.8. Связь показателей интеллектуального развития с массой тела при рождении (абсцисса). Ордината — средние Z-баллы для результатов математических тестов в возрасте 7, 11 и 16 лет в группах девочек с разной массой тела при рождении. Z-баллы — результаты теста, выраженные в долях от стандартного отклонения (см. разд. 1.6.3) в соответствующей возрастной группе. Вертикальными линиями обозначены 95% доверительные интервалы (см. разд. 1.4.4) (по данным Jefferis B.J.M.H. et al. *BMJ*. 2002. Vol. 325. P. 305–8)



Никогда нет уверенности в том, что наблюдаемая в ИСК группа больных имеет общие для всех таких больных характеристики, а сведения о прошлом никогда не бывают надежными. Поэтому результаты отдельных ИСК обычно не рассматриваются как исчерпывающе доказательные. Тем не менее поскольку такая организация исследований относительно дешева и проста, а для редких болезней является единственно возможной (перспективно невозможно отследить столь большую группу лиц, чтобы в ней развилось достаточное для исследования число заболеваний), постольку ИСК являются важнейшим методом исследования хронических болезней.

Перспективные исследования могут сочетаться с *вмешательством* в естественный ход событий<sup>1</sup>. Таково большинство исследований по испытанию средств лечения и профилактики (*исследования с вмешательством или экспериментальные исследования*). По определению их особенностью является не наблюдение за естественным развитием событий, а внесение изменений в него с последующей оценкой результатов этих изменений. Экспериментальные исследования делят на две группы: контролируемые и неконтролируемые. «Контролируемые» означают, что результат вмешательства в основной группе сравнивается с результатом отсутствия вмешательства в другой (контрольной). Неконтролируемые исследования представляют собой простейший вариант эксперимента, в котором исследователь оценивает последствия вмешательства или в лучшем случае сравнивает последствия с состоянием до вмешательства. Из исследований такой структуры нельзя сделать определенных выводов об эффективности вмешательства. Контролируемые исследования дают наилучшие основания для оценки эффективности вмешательства, но для того, чтобы результаты эксперимента давали врачу надежные основания для действий, современный эксперимент должен соответствовать ряду требований, которые рассмотрены в главе 3.

Исследования, в которых вмешательств не осуществляется, называются *обсервационными* (от англ. *observational* — наблюдательные). К ним относятся как поперечные, так и продольные исследования, за исключением экспериментальных.

---

<sup>1</sup> В эпидемиологии вмешательством называют преднамеренное (вмешательство исследователя) или непреднамеренное (вмешательство законодателя или другого лица, например, вводящего обязательное тестирование водителей на наличие паров алкоголя в выдыхаемом воздухе) изменение в естественном течении событий.



## Контрольные вопросы и задачи к разделам 1.1, 1.2

Выберите правильный ответ. Правильных ответов может быть больше одного.

1. В исследовании типа сравнения с контролем, выполненном на больных с инфарктом миокарда, обнаружено, что 40% из них пили кофе 2 раза и более в течение суток перед инфарктом, в то время как в контрольной группе пациентов с другими болезнями таких было только 15%. В контрольную группу включали пациентов с острыми заболеваниями и подбирали каждому больному инфарктом парного контрольного по полу и возрасту. Вывод, сделанный исследователями: употребление кофе является одной из причин инфаркта миокарда. Этот вывод:
  - а) правильный;
  - б) может быть неправильным из-за неправильных временных отношений между предполагаемой причиной и следствием;
  - в) может быть неправильным из-за того, что госпитализированные пациенты не представляют всех больных инфарктом;
  - г) может быть неправильным из-за того, что диагноз «инфаркт» в анамнезе у контрольных пациентов не исключен.
2. Признаками истинной причинной связи между воздействием и событием являются все признаки, кроме:
  - а) сила связи;
  - б) зависимость доза—эффект;
  - в) согласованность результатов разных исследований;
  - г) правильная организация и структура исследования;
  - д) выполнение исследования в университете.
3. При исследовании рабочих, получивших травму на производстве, исследователь обнаружил, что получившие тяжелую травму женаты (замужем) в 42% случаев против 28% случаев у получивших легкую травму. Это означает:
  - а) ничего, так как различие статистически незначимо;
  - б) ничего, так как нет контрольной группы;
  - в) может быть ничего, так как неизвестно, чем еще различаются сравниваемые группы;
  - г) б и в;
  - д) ничего, так как для выяснения такой связи необходимо экспериментальное исследование.



4. Эпидемиология — это исследование:
  - а) этиологии и патогенеза болезней человека;
  - б) структуры причин смерти человека;
  - в) распространения болезней человека и их причин;
  - г) распространения разных способов оказания медицинской помощи;
  - д) эффективности медицинских вмешательств.
5. Чтобы изучить влияние курения беременной на массу тела новорожденного, акушер опрашивал беременных при первом визите и затем сопоставлял массу тела новорожденных, родившихся от курящих и некурящих матерей. Это исследование:
  - а) типа сравнения с контролем;
  - б) проспективное;
  - в) с вмешательством;
  - г) недоказательно, так как заявление о курении не проверено объективными средствами;
  - д) недоказательно, так как случаи аборт не включены.
6. Экспозиции действию факторов В и С по отдельности вызывают болезнь, но только если им предшествует экспозиция фактору D. Фактор D отдельно не вызывает болезни. Фактор D является:
  - а) необходимой причиной;
  - б) достаточной причиной;
  - в) компонентной причиной;
  - г) необходимой, но недостаточной причиной;
  - д) ничем из перечисленного.

### 1.3. ОПИСАТЕЛЬНЫЕ ИССЛЕДОВАНИЯ

Особенность описательной эпидемиологии (от англ. *descriptive* — дескриптивная) — изучение данных о распространенности болезней и особенностях болеющих людей без вмешательства в происходящие события и без проверки гипотез о причинах заболеваний. Данные описательных исследований особенностей людей, под которыми понимаются как внутренние особенности, так и характеристики их жизни, включая, например, экспозицию условиям труда, могут быть использованы в аналитическом исследовании для проверки гипотез о причинах болезней (поперечном, когортном или в ИСК).



Помимо собственно болезней эпидемиология изучает их последствия (смертность, инвалидность, бесплодие и т.д.) и исходы воздействия, которые не являются болезнями, но важны по разным причинам (например, прекращение курения, трудности социальной адаптации, употребление наркотических средств). Вкратце, описательная эпидемиология изучает вопросы о том, «кто», «когда» и «где». (Кто заболел? Когда заболел? Где, при каких условиях подвергся экспозиции, вызвавшей болезнь?)

### **1.3.1. Общие свойства описательных исследований**

Обычно в ходе описательного исследования предпринимаются меры к тому, чтобы само проведение исследования не оказывало влияния на естественное развитие событий. Например, целенаправленное собеседование о применяемых способах контрацепции может повлиять на методы, используемые участниками исследования, и результат последующего наблюдения будет отражать не обычную для населения практику, а практику, сложившуюся под влиянием непреднамеренного исследовательского вмешательства.

Описание совокупности больных изучаемой болезнью или описание болезней, встречающихся в популяции, является простейшим видом описательного исследования. Такого рода исследования важны для медицинской практики и общественного здоровья. Систематическое наблюдение за состоянием здоровья населения, тщательная регистрация заболеваний позволяют не только численно оценить происходящие изменения, но и своевременно выявлять повышение заболеваемости и даже выявить новые болезни. Именно так в ходе обычного анализа сведений о заболеваемости в США был выявлен СПИД. Так были выявлены крайне немногочисленные случаи «нового варианта» болезни Крейтцфельда—Якоба (быстро прогрессирующего смертельного заболевания мозга, новый вариант которого наблюдается в молодом возрасте). Эпидемиологическое исследование этих случаев было исчерпывающим, что позволило быстро принять меры против эпидемии в международном масштабе.

### **1.3.2. Наблюдение за заболеваемостью населения**

Только на основании исследований распространенности болезни в популяции можно получить представление о потребности населения в медицинской помощи. Например, только обследовав население,



можно установить число людей, нуждающихся в протезировании тазобедренного сустава.

Для того чтобы выявлять случаи опасных инфекционных болезней, осложнения медицинских вмешательств и новые болезни, необходимо поддержание системы регистрации заболеваний, анализа динамики заболеваемости и оповещения врачей и регулирующих органов о выявленных изменениях. Такую систему называют эпидемиологическим надзором. На основе наблюдения за заболеваемостью можно выявить новые угрозы общественному здоровью. Списки болезней, относительно которых применяется система оперативного уведомления, различаются в разных странах. В их основе лежат опасные эпидемические болезни (геморрагические лихорадки, чума и др.), другие болезни в списке могут присутствовать или нет.

Система наблюдения может реализовываться в трех вариантах. В первом случае информацию собирают врачи или другие специалисты попутно с выполнением ими своих основных обязанностей. Для этого медицинские руководители обеспечивают подчиненных руководствами по учету, формами отчетности и контролируют регулярность и правильность предоставления отчетной информации. Например, врачи обобщают результаты своей работы обычно не чаще 1 раза в год, но по важнейшим инфекционным болезням предоставляют информацию ежемесячно, а по отдельным болезням — немедленно при выявлении. Также поступают и другие участники наблюдения: сотрудники вирусологических лабораторий, судебно-медицинские эксперты и др. На основании такой отчетной информации специалисты-эпидемиологи, профессионально занимающиеся анализом заболеваемости, проводят анализ и дают рекомендации.

Недостатком подобной системы является сбор информации людьми, не имеющими специальной подготовки, обычно загруженными работой по специальности и не относящимися к этой своей обязанности как к важной. В результате такая информация недостаточно надежна и многократно перепроверяется, прежде чем станет основой для важных решений. Тем не менее этот способ сбора информации для наблюдения является основным, и он действует даже в периоды экономических и социальных кризисов (рис. 1.9). На этом основании предпринимаются соответствующие меры: в данном примере вакцинация, и в случае их эффективности система наблюдения дает сигнал обратной связи.



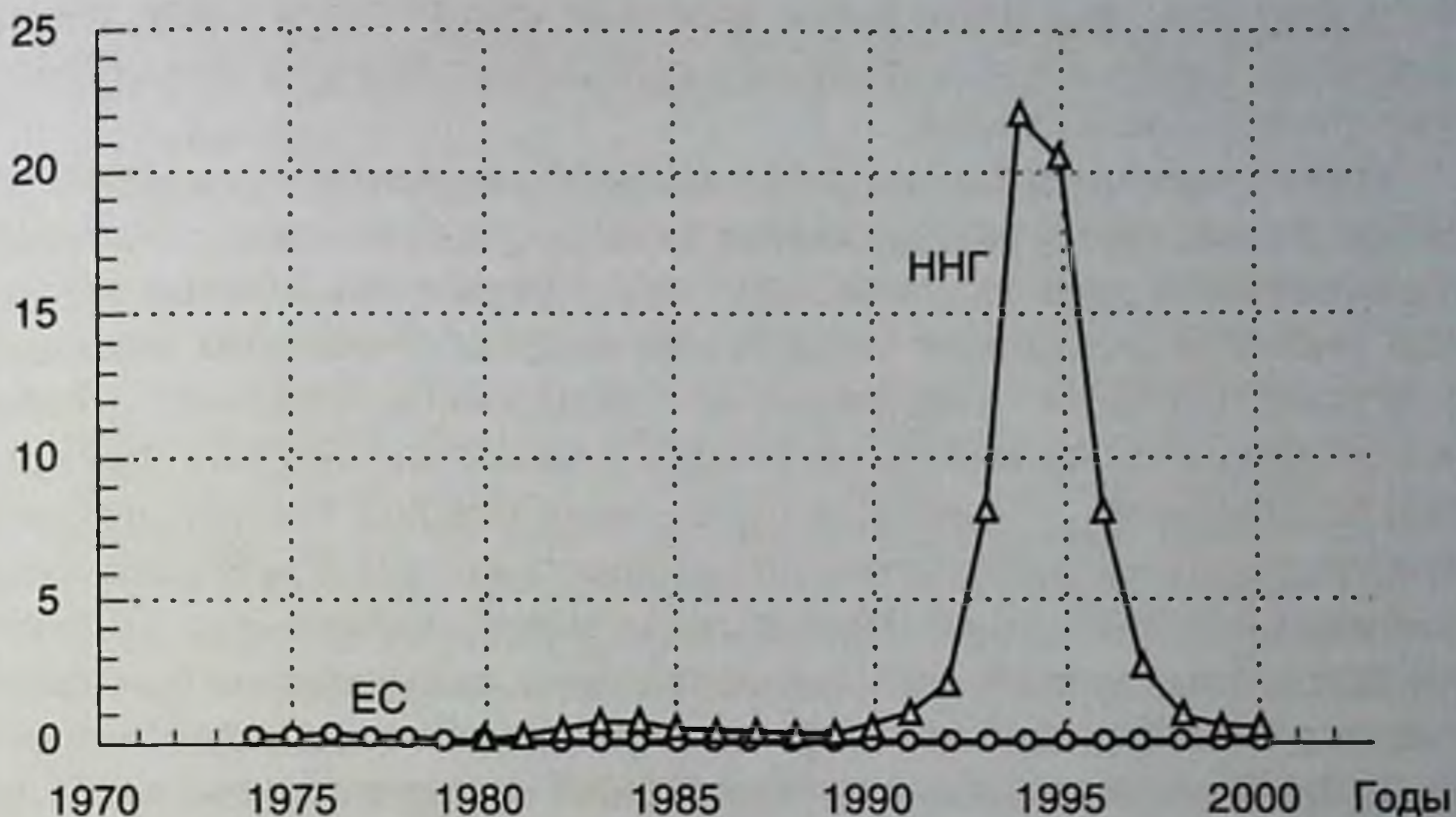


Рис. 1.9. Инцидент дифтерии в республиках бывшего СССР (ННГ, стрелки) и в странах Европейского союза (ЕС, кружки), выраженный числом новых случаев на 100 000 населения (по The European Health Report 2002. WHO Regional Publications, European Series, No. 97. WHO Regional Office for Europe, Copenhagen. Данные доступны на <http://www.euro.who.int/HFADB>)

Второй вариант — сбор информации в ходе специальных периодических обследований населения. Примером могут служить медицинские обследования малых групп населения в ходе переписей<sup>1</sup>. Если эти группы правильно выбраны, по ним можно судить о состоянии здоровья населения в целом (см. разд. 1.4.1). В США проводятся периодические Национальные обследования состояния питания и здоровья населения (NHANES) на выборках из населения страны. Проведение таких описательных исследований дорого, но они позволяют скорректировать оценки, получаемые на основании непрерывно действующей системы наблюдения за состоянием здоровья населения. Для целей наблюдения за заболеваемостью может применяться детальное наблюдение за заболеваниями, которые регистрируют отдельные избранные врачи первич-

<sup>1</sup> Перепись — сплошное, т.е. включающее все население, обследование, проводимое государством. Целью служит уточнение учета населения для сбора налогов, призыва в армию и др. Перепись является вариантом сплошного поперечного исследования. Поскольку перепись обычно проводится путем собеседования подготовленного специалиста (интервьюера) с гражданами, по методу проведения она является опросом (англ. *poll*).



ного контакта<sup>1</sup>, так называемые дозорные специалисты (англ. *sentinel specialists*). При постоянной работе с отдельными врачами можно получить более точные данные.

Третий вариант эпидемиологического обследования — создание регистра отдельной группы граждан, например работников предприятий, где используют радиоактивные вещества, или регистра больных отдельной болезнью, например, перенесших инсульт. Регистром называют базу данных о лицах, выделенных по определенному признаку, в которой собраны уточненные сведения о заболевании и сопутствующих признаках (например, о факторах риска, медицинских вмешательствах). Анализ данных регистра позволяет уточнить данные о заболеваемости и исходах болезней, получаемые в ходе обычного наблюдения. Полного совпадения между статистическими данными, полученными при оказании медицинской помощи и при эпидемиологических обследованиях, не может быть, поскольку диагностические критерии и объемы обследования, применяемые в клинике, отличаются от критериев, применяемых при обследованиях населения (см. разделы 1.4.3 и 2.2.9).

С развитием информационных систем появилась возможность анализировать данные, не предназначенные для регистрации заболеваний, для того, чтобы получить оперативные косвенные сведения о заболеваемости. Такими данными могут быть запросы в интернете, объемы покупки лекарств, других товаров. Эти сведения полезны только в том случае, если объем их велик и при наличии специальных методов анализа. Такие данные называют *большими* (англ. *big data*). Они обычно не могут заменить профессионального эпидемиологического надзора, но могут дополнять его и быть весьма оперативными.

### **Пример раскрытия причинно-следственных отношений в описательных и аналитических исследованиях**

Потребность в средствах для снижения массы тела в мире быстро возрастала в конце XX в., и фармацевтические компании, которые 23 года не предлагали новых лекарств, за 2 года (1996–1998) предложили в США 3 препарата для снижения массы тела. В 1992 г. было выписано 60 тыс. рецептов на фенфлурамин,

<sup>1</sup> Обычно таковыми являются врачи общей практики (семейные врачи) или другие врачи, к которым пациент обращается непосредственно: участковые терапевты, педиатры и гинекологи.



а в 1997 г. — более 10 млн на фен-фен (комбинированный препарат фенфлурамина и фентермина). Первый год широкого использования препарата стал его последним годом.

Первоначально врачи отметили клинически выраженное поражение клапанов сердца у небольшого числа пациентов. Это вызвало необходимость целенаправленных исследований частоты поражений клапанов. Для изучения последствий применения фен-фена разработали специальные диагностические критерии «слабой аортальной регургитации» и «слабой и умеренной митральной регургитации». Обследования пациентов, получавших фен-фен, выявили частоту аортальной регургитации в разных группах от 6,6 (т.е. примерно как в популяции) до 31%. Для более точной оценки риска поражения клапанов провели ИСК (см. разд. 2.8). В 5 исследованиях частота аортальной регургитации оказалась выше, чем в контроле (8,8 против 3,5%). Главное влияние на преваленс (частоту случаев в группе) оказывала длительность приема фен-фена (рис. 1.10). Преваленс митральной регургитации не различался у получавших фен-фен и контрольных пациентов. Преваленс клинически выраженного поражения клапанов составил только 0,09% по результатам объединенного анализа более 10 тыс. случаев применения фен-фена.

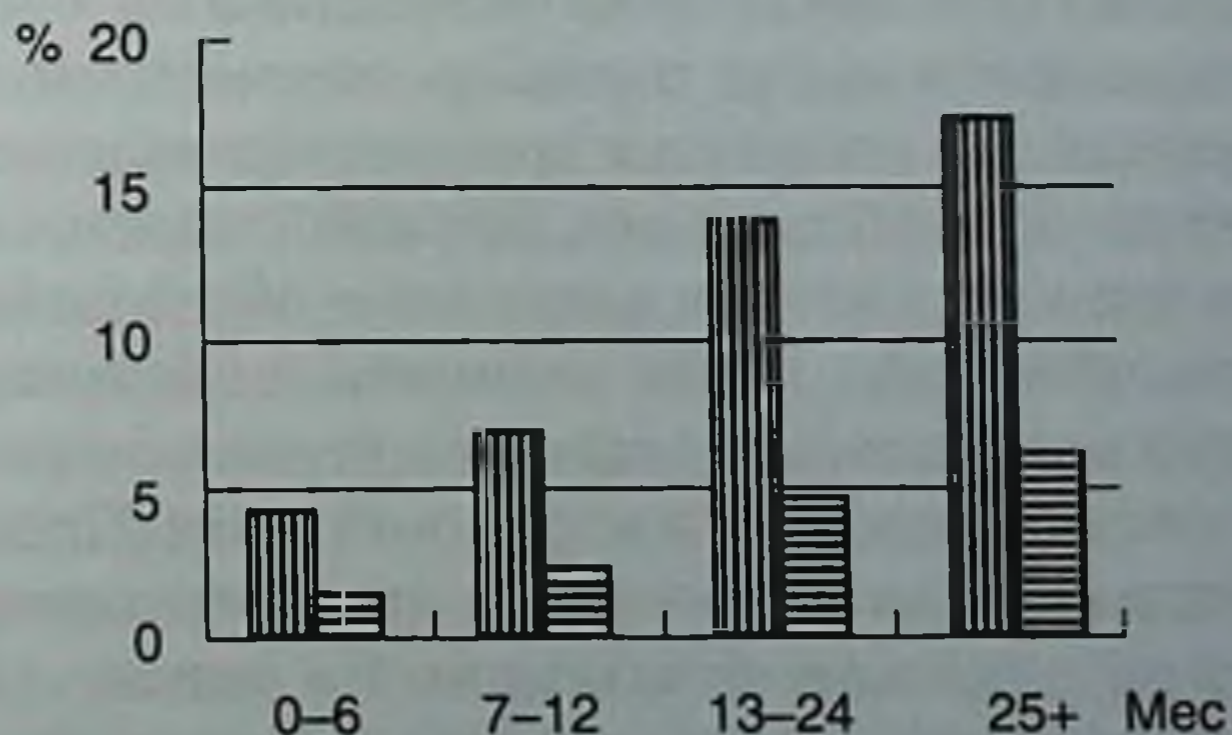


Рис. 1.10. Преваленс аортальной регургитации в зависимости от длительности приема фен-фена (вертикальная штриховка) и ее относительный риск (отношение преваленса у принимавших препарат к преваленсу в контрольной группе)

Исследование поражения сердечных клапанов было бы неполным, если бы не был оценен инцидент в процессе приема лекарственного средства (ЛС), так как какая-то неизвестная часть пациентов имела до лечения дефекты клапанов (инцидентом называется частота новых случаев в группе). С этой целью проанализировали данные о пациентах, у которых эхокардиографическое исследование сердца



было выполнено до и после лечения. У 46 пациентов инцидент составил 2% при продолжительности лечения в среднем 5,3 мес (т.е. 1 пациент) и 15% — при продолжительности лечения в среднем 13,8 мес.

Препарат был быстро изъят производителем с рынка, и новых случаев поражения клапанов описывается крайне мало. Наблюдение за пациентами с поражениями клапанов продолжается, и есть основания считать течение этих поражений доброкачественным — поражение уменьшается в размерах. Таким образом, в короткий срок были реализованы эпидемиологические исследования разной структуры, позволившие исчерпывающе оценить риск, выявить причину и дать рекомендации по профилактике.

### 1.3.3. Вероятность

Как и в других науках, в эпидемиологии в различные периоды ее истории и в одинаковых целях использовалось множество более или менее различающихся показателей, и их названия менялись. Некоторые из показателей продолжают использовать и теперь, сохранив за собой их традиционные уже имена, но появились новые названия и новые показатели. Из статистики врачи почерпнули методы и представления с их особыми терминами, которые стали важными для понимания того, что имеет место в медицине и общественном здоровье. Более того, специфические задачи медицины и общественного здоровья подтолкнули математиков к созданию специальных методов, которые ныне широко используются и в других областях. Процесс освоения медициной чуждого ей математического инструментария и математических концепций протекает болезненно. Не только студенты-медики испытывают трудности с освоением точных наук. Врач, приступая к научной работе, также осознает трудность и необходимость применения статистических методов для планирования и описания результатов своих исследований. Хорошо известно, как велико желание врачей овладеть началами статистического анализа и как велики трудности. Статистики, со своей стороны, понимают проблемы медицинских специалистов, и мы имеем прекрасные учебники для врачей [4–6]. Одна из наиболее важных статистических концепций для врача — концепция вероятности.

В клинике и в обычной жизни мы видим, что у одних людей имеются одни признаки, у других — другие. Одни признаки часты (прямые



волосы, курение табака), другие — редки (зеленые глаза, способность достать I пальцем до предплечья). Понятие «частота» соответствует тому, какая часть от всех наблюдаемых людей имеет этот признак. Эту часть можно выразить по-разному. Можно таких людей сосчитать. Это будет количество или *частота* (иначе называемая частотой, *абсолютная частота*). В других науках, например в физике, под частотой понимают совсем иное (число событий за единицу времени). Даже для начинающего наблюдателя ясно, что абсолютная частота (число), например количество зеленоглазых, будет зависеть от того, скольким людям заглянуть в глаза. Одно дело, если 20, и другое — если 200. Поэтому естественно вычислять *отношения* количества зеленоглазых к общему количеству осмотренных людей. Эта *относительная частота* называется *вероятностью* (англ. *rate*).

Различие между абсолютной частотой и относительной частотой велико, и использование абсолютной частоты может скрывать то, что становится очевидным при использовании вероятности. В табл. 1.1 приведены данные о смертности от несчастных случаев на производстве работников различных отраслей хозяйства. Абсолютная частота наиболее велика в строительстве и промышленности, но в расчете на человеко-годы на производстве смертность (относительная частота) в 25 раз выше средней (2/100 000) в торговом флоте и в 50 раз — у рыбаков. Такой *анализ в подгруппах* — важный вариант эпидемиологического исследования, так как позволяет выявить наиболее уязвимые группы людей, в отношении которых профилактические вмешательства могут быть наиболее эффективны. Обратим внимание, что в представленных данных отсутствуют некоторые другие профессиональные группы с высокой смертностью от несчастных случаев на производстве, например шахтеры (около 30 на 100 000 человеко-лет) и лесорубы (около 90 на 100 000).

Относительную частоту можно выразить по-разному. Чаще всего число субъектов с изучаемым признаком делят на общее число наблюдений. В виде натуральной дроби это отношение оставляют редко, чаще результат выражают в долях от единицы или в процентах. Так, если из 80 школьников оказалось 3 рыжих, то относительная частота будет 0,04, или  $0,04 \times 100\% = 4\%$ . Точнее, при делении получилось 0,0375, но очевидно, что из наблюдения 80 человек можно получить только приблизительное впечатление об относительной частоте; о точности в долях процента не может быть и речи.



Таблица 1.1. Число погибших в производственных происшествиях работников по отраслям хозяйства Великобритании в 1976–1995 гг. и смертность (по Roberts S. Lancet. 2002. № 360. P. 543–4)

Отрасль	Число погибших работников	Частота на 100 000 человеко-лет (95% доверительный интервал)
Рыбаки	454	103 (87–119)
Торговый флот	507	52 (39–65)
Энергетика и водоснабжение	700	10 (7–14)
Сельское, лесное хозяйство, рыболовство	737	9 (8–10)
Строительство	2404	8 (7–10)
Промышленное производство	2264	2,2 (1,9–2,5)
Обслуживание	1619	0,7 (0,6–0,7)

В медицинской практике и в практике общественного здоровья приходится иметь дело с редкими событиями, например со случаями смерти от редкого вида рака. Если имеет место 2 случая на 100 000 населения, то это относительная частота, выражаемая как 0,00002, или 0,002%. Для того чтобы выразить относительную частоту редких событий удобным для произношения числом, вводят подобные процентам выражения: 1/1000 (промилле), 1/10 000, 1/100 000, 1/1 000 000. В результате получается удобопроизносимое: «два на 100 тыс. населения».

Наиболее привычное отношение, применяемое для выражения относительной частоты, называется «пропорция (доля)». Пропорция — это отношение двух величин, из которых находящаяся в числителе входит в состав знаменателя. Существует также иной способ выражения относительной частоты, который используется в эпидемиологии и статистике. Это *шансы*. Так, относительная частота 3/80 может быть выражена как шансы 3:77. Именно так выражают вероятность ожидаемого события, когда делают ставки на бегах (англ. *betting chances*). Этот же способ выражения вероятности оказывается удобным для ряда вычислений в эпидемиологии. Для нас важно, что если событие редкое, как в нашем примере, то разница между 3/80 и 3:77 невелика (0,0375 и 0,0390). Если же речь идет о более частых событиях, то эти показатели



существенно различаются. Так, 3/7 и 3:4 различаются уже заметно (0,43 и 0,75). Поэтому шансы и вероятность можно оценивать одинаково, только если речь идет о редких событиях. В табл. 1.2 и на рис. 1.11 соотношение между вероятностью и шансами представлено в диапазоне вероятности от 0,01 до 0,99.

Таблица 1.2. Вероятность и шансы при разной абсолютной частоте события в группе из 100 человек

Абсолютная частота	Относительная частота (вероятность)	Шансы
1	0,01	0,01
2	0,02	0,02
3	0,03	0,03
4	0,04	0,04
5	0,05	0,05
6	0,06	0,06
7	0,07	0,08
8	0,08	0,09
9	0,09	0,10
10	0,10	0,11
20	0,20	0,25
30	0,30	0,43
40	0,40	0,67
50	0,50	1,00
60	0,60	1,50
70	0,70	2,33
80	0,80	4,00
90	0,90	9,00
93	0,93	13,29
95	0,95	19,00
98	0,98	49,00
99	0,99	99,00



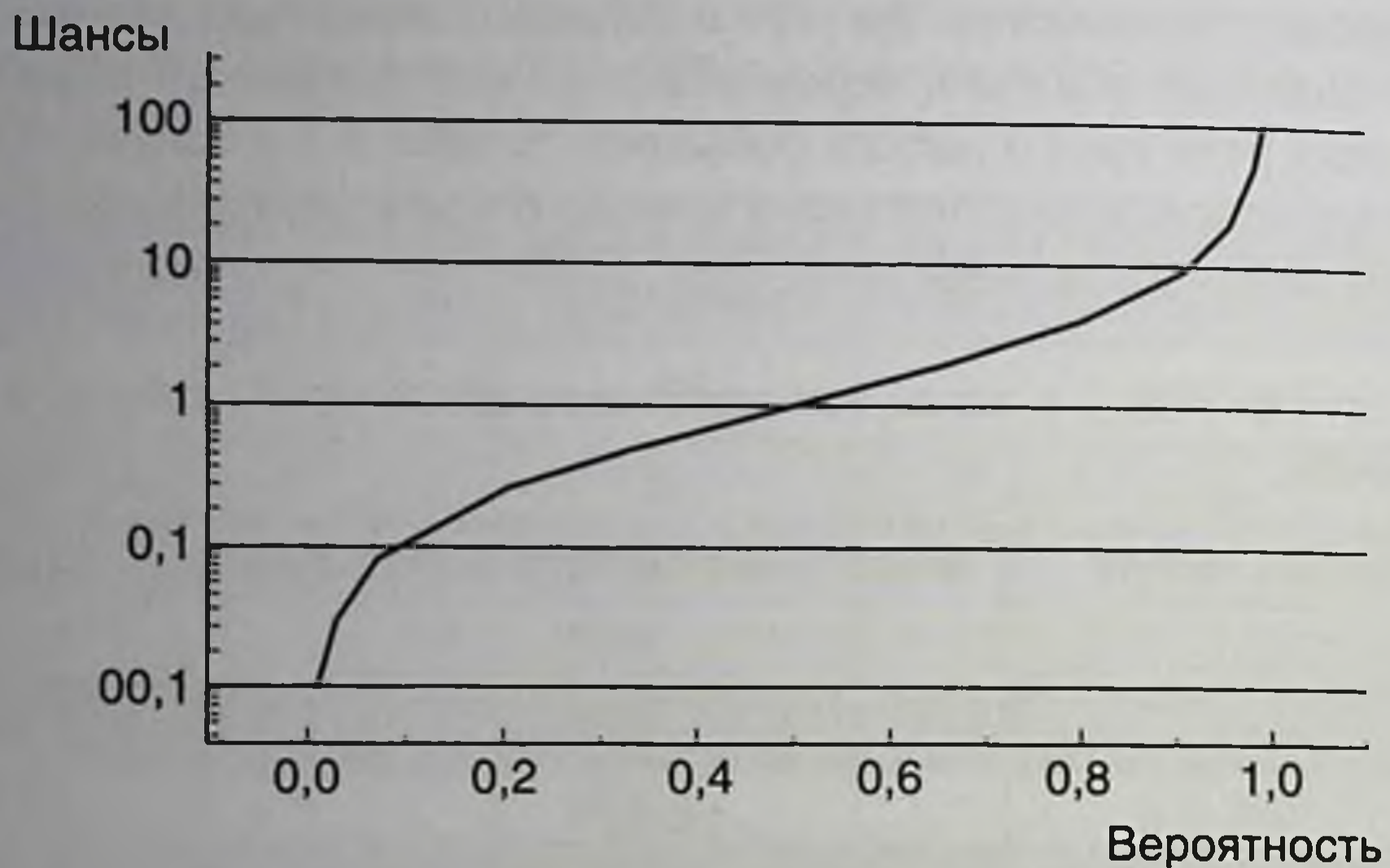


Рис. 1.11. Соотношение между шансами и вероятностью. Данные соответствуют числам из табл. 1.2. Обратите внимание на логарифмическую шкалу шансов

События, имеющие место в популяции, могут быть взаимоисключающими. Это означает, что люди в популяции относятся только к одному из *классов*. Классами называют группы людей, выделяемые по какому-нибудь признаку (от названия процесса — «классификация»). Если женщина обозначена в заполняемой форме обследования как имеющая двух детей, то она не может быть отнесена одновременно к бездетным.

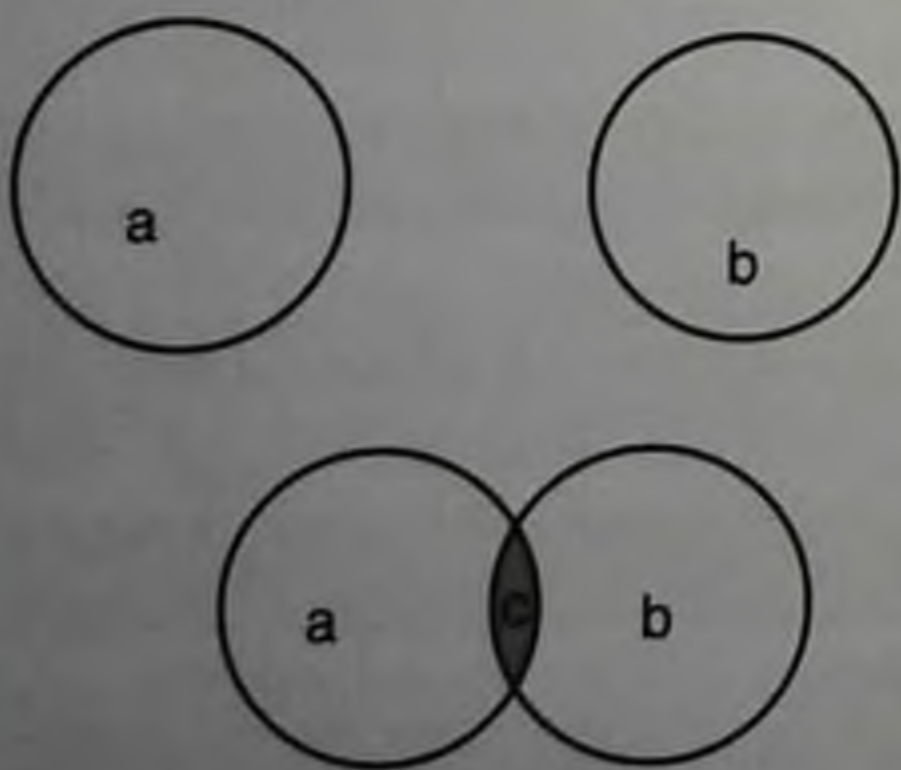


Рис. 1.12. Два взаимоисключающих события (вверху) и частичное совпадение независимых событий (внизу)

Наоборот, получение пенсии (а) не исключает продолжения трудовой деятельности (б) (рис. 1.12). Это не взаимоисключающие признаки, что должно учитываться при планировании исследования и при анализе его результатов. Например, общее число работающих и получающих пенсию не может быть получено сложением пенсионеров и работающих: при этом работающие пенсионеры (с) будут посчитаны дважды.

Классы могут быть исчерпывающими. Например, в медицинской



статистике людей классифицируют по полу как мужчин и женщин. Незначительное число новорожденных, пол которых не удается определить, после проведения дополнительных исследований все же относят к одной из категорий. Уточнение пола важно не только для статистического учета, но и по многим другим причинам, например, это важно для формирования самоидентификации человека в процессе взросления. Если классы исчерпывающи, то сумма всех классов дает численность популяции. Классы могут быть неисчерпывающими. Например, сумма численности школьников, студентов и работающих меньше численности популяции.

События могут быть независимыми. С точки зрения здравого смысла и по статистическому определению независимыми являются события, совместная вероятность которых равна произведению их вероятностей. Например, если в популяции половина — женщины и половина людей — кареглазые, то и среди женщин половина (50%) кареглазых. Относительная частота кареглазых женщин в популяции:  $0,5$  (женщин в популяции)  $\times$   $0,5$  (кареглазых в популяции)  $= 0,25$ . Другими словами,  $1/4$  от популяции — кареглазые женщины.

В случае если вероятность совпадения признаков отличается от произведения вероятностей, говорят об их связи. На этом простом принципе — сравнении реальной частоты с той, что ожидается от случайного совпадения, построено много статистических тестов. На рис. 1.12 число людей, страдающих болезнью *a* (ишемическая болезнь сердца) и болезнью *b* (пародонтоз) одновременно (*c*), может оказаться больше, чем ожидается на основе случайного совпадения. Это может быть основанием для размышлений о связи двух болезней, но мы можем предположить, что дело всего лишь в том, что пожилые люди, входящие в выборку, чаще страдают обеими болезнями. Эпидемиологические методы позволяют разделить такого рода сложные взаимодействия двух и более признаков.

Вероятностные модели позволяют представлять связи не в виде жестких правил типа «если *A*, то *B*», а в виде «если *A*, то *B* с вероятностью  $0,8$ », что лучше соответствует реальности.

### 1.3.4. Преваленс

Для изучения явлений в популяции ключевыми являются понятия «преваленс» и «инциденс». *Преваленсом*, или *распространенностью*, называют относительную частоту в популяции (группе людей) некоторого



признака (от англ. *prevalence*). Относительная частота (вероятность) выражается в долях от единицы, в процентах, и в отношении редких событий — в числе случаев на 1000 или 100 тыс. человек.

Преваленс:

$$P_r = D/n,$$

где  $D$  — число больных,  $n$  — численность популяции.

Так, если в популяции курят 4 из 10, то преваленс можно выразить как 0,4 или как 40%. Для оценки<sup>1</sup> преваленса надо установить число больных (людей, имеющих некоторый признак) в популяции. Применительно к болезням для этого существуют два основных метода. Преваленс можно оценить на основании *обращаемости*, т.е. установить по медицинским или страховым документам или иным путем, сколько человек в популяции страдают изучаемой болезнью. Это главный способ, используемый в медицинской статистике. Поскольку не все люди обращаются за медицинской помощью, таким методом исследователь получает заниженную величину преваленса. Известно, например, что некоторые люди обращаются за медицинской помощью, когда размер внешне видимой опухоли превышает голову человека, или когда эта опухоль уже мешает ходить или носить одежду. Немало людей умирает от хронических болезней, обратившись за помощью только за несколько дней до наступления смерти.

Поэтому более точные представления о преваленсе болезней (состояний) можно получить при обследовании (англ. *survey*) населения (табл. 1.3). Такие исследования называют *исследованиями распространенности* (англ. *prevalence study, disease frequency survey*). В таком случае, выявляя стандартные признаки болезни, можно обнаружить больных, которые еще не заметили у себя болезни, еще не считают себя достаточно больными, чтобы обратиться к врачу, или просто не имеют доступа к медицинской помощи. Преваленс болезни, измеренный таким образом, называется *преваленсом по выявляемости*.

<sup>1</sup> Термин «оценка» используется в двух близких смыслах. Во-первых, оценкой называется получение приблизительного, предварительного результата. Именно в таком смысле говорят об оценке средней, подчеркивая, что значение получено на выборке, но оно служит для получения представлений о всей популяции. Во-вторых, оценкой называют вынесение суждений на основе заранее сформулированных критериев. В этом смысле оценивают данные эпидемиологического наблюдения, чтобы принять решение о вмешательстве, или знания студента, чтобы помочь ему в освоении учебного материала.



Таблица 1.3. Преваленс ожирения у детей и подростков США (в процентах). Данные периодических исследований состояния здоровья и питания населения США (NHANES), приводятся по: National Center for Health Statistics, CDC

Возраст, лет	Годы						
	1963-1965; 1966-1970	1971-1974	1976-1980	1988-1994	1999	2005	2015
6-11	4	4	7	11	13	15	18
12-19	5	6	5	11	14	18	21

Примечание. Беременные исключены начиная с 1971 г. За 1963–1965 гг. данные для детей 6–11 лет, а за 1966–1970 гг. — для подростков 12–17 лет.

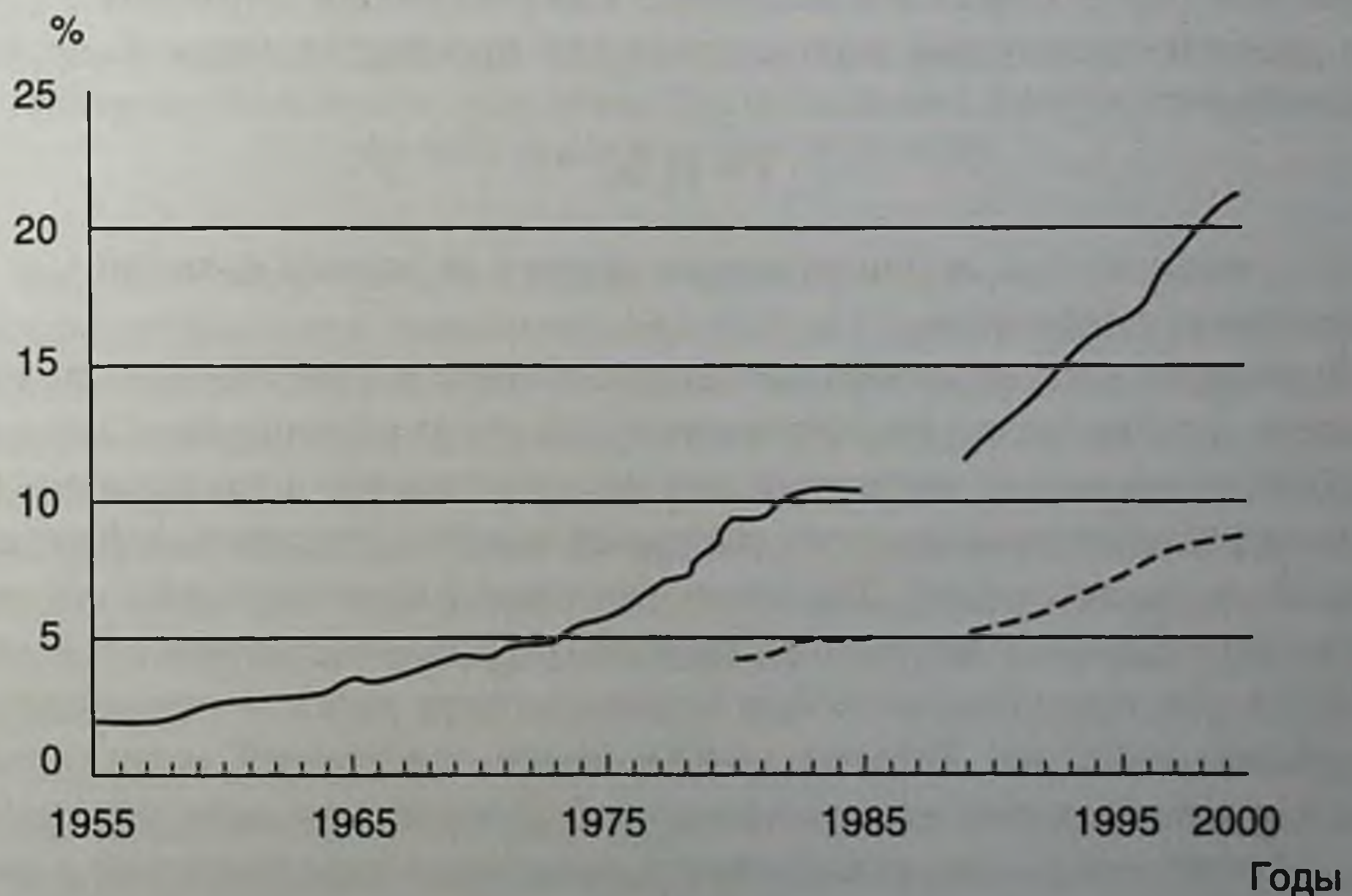


Рис. 1.13. Доля родов путем кесарева сечения в Англии (1955–2000 гг.). Линия — всего, пунктир — плановое сечение, в процентах от числа родов. Данные UK Department of Health

С преваленсом не следует путать *показатели структуры* заболеваемости и структуры исходов болезней (состояний). Например, доля родов, завершившихся кесаревым сечением (рис. 1.13), вычисляется как пропорция  $K/n$ , где  $K$  — число родов, завершившихся сечением,  $n$  — общее число родов.



### 1.3.5. Инциденс

Число новых событий в популяции в фиксированный период времени называют *инциденсом* (англ. *incidence, incident number*). Применительно к новым случаям болезней эквивалентом инцидента является *заболеваемость*<sup>1</sup>. Поскольку изучаемые новые события не обязательно являются болезнями (например, первое использование инъекционного наркотика или начало вождения автомобиля), использование термина «инциденс» неизбежно.

Если исследование обобщает данные за некоторый период, то представляют данные обо всех случаях за весь период исследования — это называют *кумулятивным инциденсом* (т.е. накопленным за время наблюдения (англ. *cumulative incidence*). Кумулятивный инциденс с учетом размера популяции вычисляется как пропорция (англ. *incidence proportion*):

$$I = D_i/n,$$

где  $I$  — инциденс,  $D_i$  — число новых случаев за период времени  $i$ ,  $n$  — численность популяции.

Выражают кумулятивный инциденс обычно в виде относительной частоты — в процентах от популяции или в долях от единицы. С целью обобщения инциденс выражают как частоту случаев в традиционный для измерения период времени, например за год (относительный инциденс, англ. *incidence rate*). Для этого инциденс  $I$  делят на число лет наблюдения  $t$ . Соответственно инциденс за год вычисляется как  $I = D_i/nt$ .

Инциденс представляет собой основную меру *риска* — вероятности изучаемого события. Термин «риск» имеет отчетливый негативный смысл, и его обычно *не* используют для обозначения частоты желательных событий, например прекращения выявления микобактерий в мокроте у больных легочным туберкулезом.

Преваленс и инциденс связаны между собой. В популяции стабильного размера:

$$P_r = I \times L,$$

где  $P_r$  — преваленс,  $I$  — инциденс,  $L$  — продолжительность болезни.

<sup>1</sup> В российской государственной статистике инциденс заболеваний в виде числа новых случаев заболевания на 100 тыс. населения в год отражается как «заболеваемость», а преваленс заболеваний отражается как «общая заболеваемость».



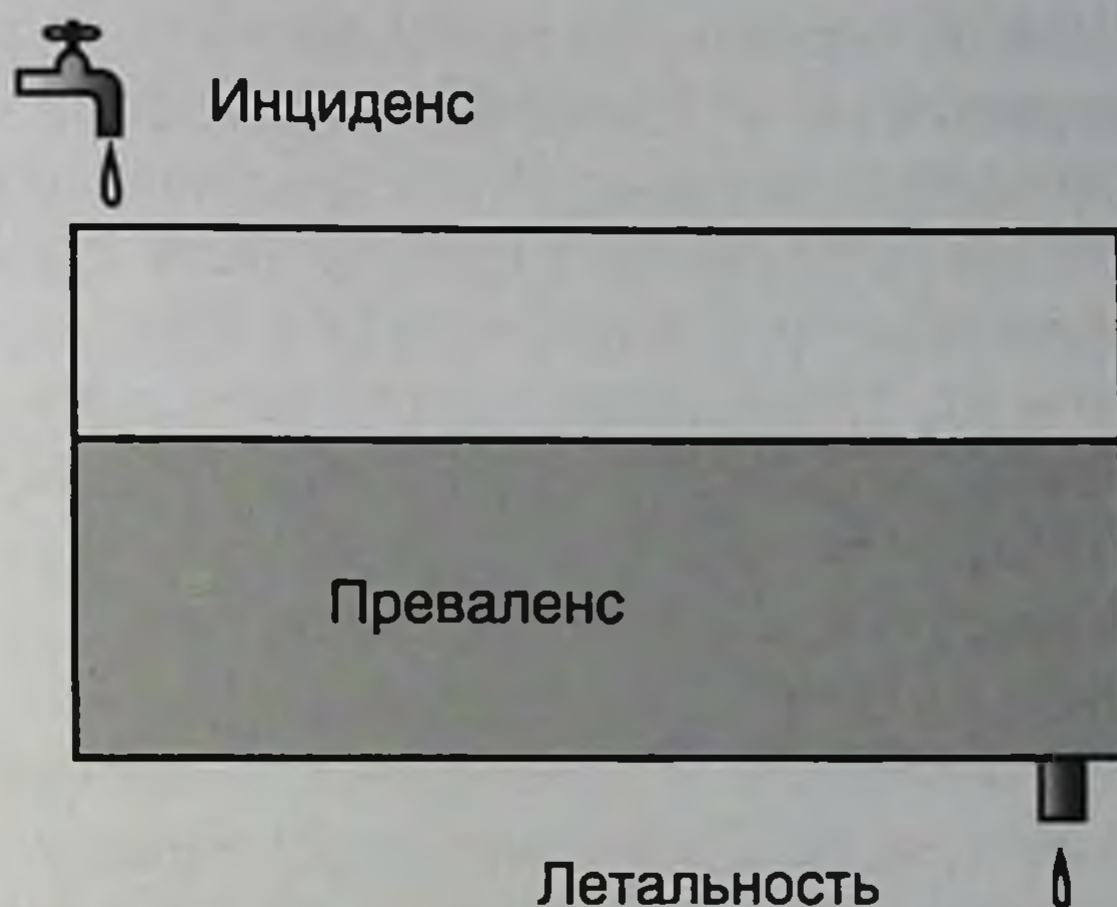


Рис. 1.14. Модель зависимости между инцидентом и превалентом. В популяции (обозначенной как емкость) доля больных (заполненная часть емкости) тем больше, чем выше инцидент и ниже летальность

Чем выше инцидент, тем выше и превалент болезни. В случае хронического заболевания нарастанию преваленса противостоит смертность. Если больные с изучаемой болезнью быстро умирают, то превалент болезни невелик (рис. 1.14). Наоборот, при болезнях, не увеличивающих смертность, превалент может быть высоким (например, нейро-сенсорная тугоухость). Поэтому *высокий превалент болезни не означает ее опасности*. В случае острой (излечимой) болезни превалент снижается не только за счет смерти заболевших, но и за счет выздоровления **выживших**.

Вариантом событий, описываемых инцидентом, является смертность. *Смертность* — это частота смертельных исходов в популяции, и она выражается в показателях относительной частоты, например 2 на 100 000 населения в год. Помимо общей смертности (от всех причин, англ. *total mortality, all-causes mortality*) вычисляют показатели смертности от отдельных причин.

*Летальность* (англ. *case-fatality*) — это относительная частота смертельных исходов в группе больных данной болезнью. В стабильной популяции смертность от острого заболевания однозначно определяется инцидентом и летальностью:

$$\text{Смертность} = \text{инцидент} \times \text{летальность.}$$



Обычно показатель летальности вычисляют для острых состояний, поскольку они ограничены во времени, или для групп больных, находящихся на стационарном лечении. Например, летальность вычисляют за первые 28 дней после развития инсульта. Такое ограничение неизбежно при изучения исходов хронических болезней, но приводит к недооценке летальности. Летальность следует вычислять по отношению к числу заболевших, но иногда летальность вычисляют для госпитализированных больных (больничная летальность). Например, летальность больных инфарктом миокарда, поступивших в больницу, ниже, чем общая летальность при этом заболевании, так как часть больных умирает до госпитализации, и не все больные госпитализируются.

Врачи не всегда доверяют естественному процессу выявления новых случаев, так как в популяции регистрация состояния может быть неполной, и диагностика в рамках обычной практики может быть ненадежной. Например, начало курения подростками отражается в медицинских картах, но только при целенаправленном опросе и при контроле правдивости ответов объективными методами (анализ выдыхаемого воздуха) можно установить истинный инцидент. Поэтому инцидент стремятся оценить не по формальным данным о популяции, а наблюдая за выборкой из популяции, по необходимости применяя периодические обследования. Поскольку численность выборки может быть небольшой, а период наблюдения не равен и не кратен целому году, то вычисляют суммарное количество лет наблюдения за пациентами. Предполагается, что, например, 120 человеко-лет равны году наблюдения за 120 пациентами или 12 годам наблюдения за 10 пациентами<sup>1</sup>. Инцидент, определенный таким способом, называют «плотность инцидента» (англ. *incidence density*) (пример применения см. табл. 1.1). Число случаев болезни, развивающихся в единицу времени в популяции (например, на человеко-год), называют также силой заболеваемости (англ. *force of morbidity*, *hazard rate*, *instantaneous incidence rate*, *person-time incidence rate*). Для обозначения смертности, определенной наблюдением за выборкой и вычисленной как отношение частоты к наблюдаемым годам жизни, применяются аналогичные термины «плотность смертности» и «сила смертности».

<sup>1</sup> Такое предположение, конечно, справедливо не всегда. Например, после вакцинации иммунитет возникает не сразу и ограничен во времени, поэтому короткое наблюдение может дать иной результат, нежели длительное. Тем не менее во многих случаях такое предположение приемлемо.



Применение более совершенных диагностических методов обычно увеличивает инцидент. Так, в сравнении только с физикальным исследованием при использовании ультразвуковых приборов растет число выявляемых аномалий клапанов сердца и изменяется структура аномалий — среди них уменьшается доля тяжелых, клинически явных. Выявление болезней зависит от интенсивности и обеспеченности усилий общества. Например, предоставление анонимного бесплатного обследования для выявления СПИДа повышает выявляемость. Может изменяться определение состояния, например изменяется критерий живорождения, и соответственно изменяется статистика рождаемости и перинатальной смертности. Если не учитывать этих изменений технологии, материального обеспечения и отношения общества к выявлению болезни, то изменения чисел, описывающих инцидент и превалент, предстанут как изменения показателей здоровья во времени.

### 1.3.6. Показатели здоровья населения

Оценка индивидуального здоровья заключается прежде всего в диагностике болезней, что рассматривается в разд. 2.2. При традиционном *медицинском* подходе отсутствие болезней означает здоровье. Всемирная организация здравоохранения (ВОЗ) при формулировании своей политики в уставе дала расширительное определение здоровью как полному благополучию, а не только отсутствию болезней. Поскольку критериев полного благополучия нет и, видимо, не может быть дано, это определение затрудняет оценку здоровья человека. Применительно к группам населения, к популяции оценка состояния здоровья основывается на смертности и инциденте болезней. В свете таких данных популяция здоровее относительно других популяций, если продолжительность жизни и сохранение трудоспособности населения в ней выше, а заболеваемость ниже. Дополнительная информация, включаемая в групповые оценки здоровья, складывается из распространенности признаков, которые связаны с вероятностью развития болезней, — курение, избыточная масса тела (факторы риска) (см. разд. 1.1.4). Эти признаки имеют только дополнительное значение при сравнении популяций, близких по смертности и заболеваемости.

Значимость болезни для общественного здоровья измеряют и представляют разными способами. Наиболее значимыми являются болез-



ни распространенные и опасные — часто вызывающие смерть и другие важные исходы (инвалидность, слепоту, бесплодие). Опасность не имеет ясных критериев и выражается в субъективных оценках. В отношении болезней оценка опасности выражается в выделении, например, карантинных инфекций, т.е. группы болезней, для предотвращения распространения которых признается оправданным применение карантина.

Преваленс болезней и их вклад в формирование смертности и инвалидности могут быть представлены разными способами. Вклад болезни в общее число смертельных исходов выражается *смертностью* от этой болезни<sup>1</sup> (англ. *specific mortality*). Смертность может быть выражена *интенсивным показателем*, т.е. как частота смерти от данной болезни на 100 000 населения в год. Общая смертность складывается из смертностей от отдельных болезней. Это называют *показателем структуры смертности* населения. Показатели последнего типа называют *экстенсивными показателями*. Интенсивные и экстенсивные показатели используются для разных целей, и их применение к одним и тем же данным приводит к разным результатам. Например, в интенсивных показателях *смертность* от травм на дорогах самая большая у лиц старше 76 лет (рис. 1.15). В показателях экстенсивных первое место по доле травм на дорогах занимают молодые люди (до 16 лет) (рис. 1.16). Это отражает, с одной стороны, высокую частоту травм у молодых людей, а с другой — высокую *летальность* при травмах равной тяжести у старых людей.

Показатель *смертности* от отдельной болезни относится к популяции и отражает значение отдельных болезней для общественного здоровья. Для клинической практики, для больного и врача важен показатель *летальности* (см. разд. 1.3.5). Связь между летальностью и смертностью

---

<sup>1</sup> Смертность от болезни является косвенным показателем (заменителем) (англ. *proxy* или *surrogate*) заболеваемости этой болезнью. Смертность от болезни — важный показатель распространенности болезней, точная прижизненная диагностика которых затруднена. В популяции, где большая часть населения умирает в пожилом и старческом возрасте, для которого характерно наличие множества хронических болезней, выявление основной причины смерти затруднено. Поэтому показатель смертности от болезни в таких популяциях не дает правильной оценки заболеваемости. Косвенные показатели используют также для оценки экспозиции, когда напрямую она невозможна. Например, поскольку невозможно измерить степень воздействия дыма курящих людей на некурящего человека, постольку в качестве косвенного показателя используют совместное проживание с курящим человеком.



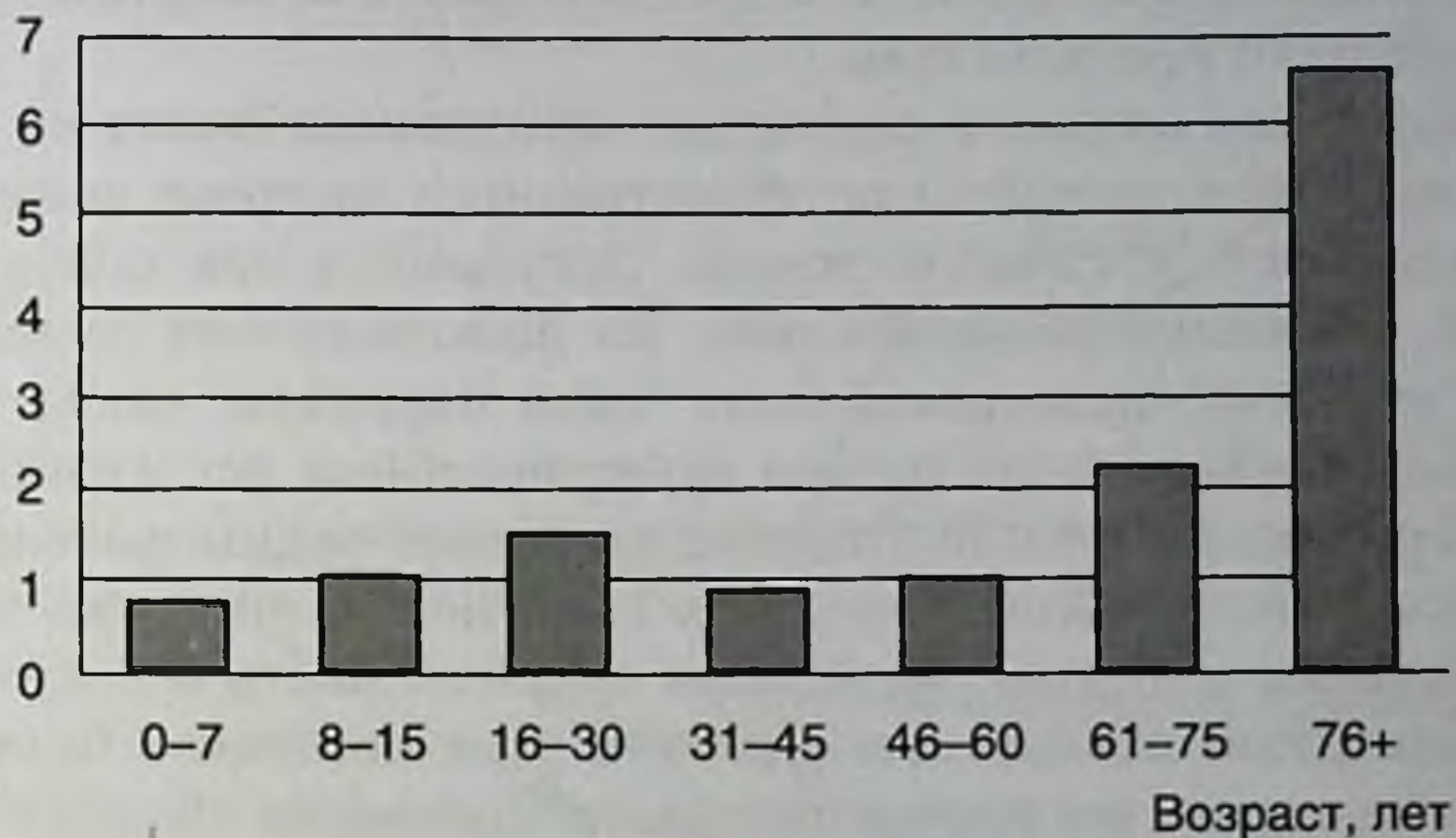


Рис. 1.15. Смертность от травм пешеходов разного возраста (число случаев на 100 000 населения) в Великобритании. По Department of the Environment, Transport and the Regions: *Transport Statistics: Road Accidents Statistics*. <http://www.transtat.dft.gov.uk/facts/accident/pedest/pedest98.htm>



Рис. 1.16. Распределение пострадавших на дорогах Великобритании пешеходов по возрасту (источник — см. рис. 1.15)

опосредуется через инцидент. Например, летальность при внебольничной пневмонии в развитых странах примерно 14% (5–20%). Поскольку это частая болезнь, постольку вклад пневмонии в смертность значителен: в США выявляется более 1 млн случаев в год, половина требует госпитализации, не менее 50 000 умирает. Наоборот, бешенство харак-



теризуется высокой летальностью, но из-за редкости смертность в развитых странах от него невелика.

Медицинский подход к статистике заболеваний ставит на первое место смертность, однако для общественного здоровья важны также инвалидность и качество жизни. Летальность при психических болезнях повышена незначительно, но поскольку они распространены и при них трудоспособность часто нарушена, вклад психических болезней в общие потери трудоспособных лет жизни очень велик. В оценке состояния здоровья населения важны не только текущие показатели заболеваемости и смертности, но также тенденции изменения и угрозы нарастания заболеваемости и смертности. На интегральной оценке совокупности этих признаков базируется оценка санитарно-эпидемиологического состояния (благополучия) населения.

### 1.3.7. Показатели наглядности

При изучении временных изменений наиболее демонстративным является представление динамики в сравнении с исходной величиной, принятой за единицу или 100%. Такие показатели называют *показателями наглядности*. В особенности они демонстративны, когда сравниваются несколько временных рядов, имеющих разный фактический уровень. Приведение всех сравниваемых рядов к показателям наглядности позволяет абстрагироваться от различий и ясно представить динамику (рис. 1.17). На этом рисунке показатели наглядности использованы для того, чтобы продемонстрировать многолетнюю динамику (*секулярный тренд*) смертности от важнейших болезней, приводящих к смерти в промышленно развитых странах.

Такое упрощенное и наглядное представление таит в себе некоторые опасности. Возникает иллюзия, что снижение величины одного показателя соответствует увеличению другого. Например, в показателях наглядности прирост на 1% смертности от болезни, дающей смертность 0,01, будет выглядеть как скомпенсированный равным снижением смертности (на 1%) от болезни, дающей смертность 0,0001. В показателях смертности в действительности будет прирост смертности на 0,01 ( $1,01 \times 0,01 - 1,01 \times 0,0001 = 0,0101 - 0,000101 = 0,009999$ ). Другая ошибка возможна, если при описании изменений во временном ряду вычислять прирост не по отношению к одному моменту, например





Рис. 1.17. Изменения смертности от четырех причин в конце XX в. в США. Смертность (ордината) скорректирована на возрастную структуру популяции (стандартизована) и выражена в процентах к уровню 1980 г. Данные Diabetes Research Working Group Report

исходному периоду анализа, а вычислять прирост или снижение в каждом периоде по сравнению с предыдущим периодом. Такое вычисление относительных показателей наглядности дает искаженное представление об изменениях во времени.

Самый распространенный вариант показателей наглядности — вычисление относительного риска. Для этого принимают за единицу риск исхода (например, смертность) в группе самого низкого риска или в иной группе сравнения. Этот прием применяется для представления результатов исследований разного типа — когортных, типа сравнения с контролем и контролируемых испытаний.

### 1.3.8. Стандартизованные показатели

*Стандартизация* выполняется для того, чтобы при сравнении статистических показателей устранить различия во вмешивающихся фак-



торах. Например, сравнивать смертность в двух странах можно, только устранив разницу в возрастном составе населения.

### Непрямой метод стандартизации

Сравнивать летальность в двух больницах можно, только устранив разницу в структуре коечного фонда. В противном случае больница, в которой хорошо лечат больных острыми нарушениями мозгового кровообращения, будет иметь очень высокие показатели летальности просто потому, что там много коек специализируется на лечении нарушений мозгового кровообращения. Для стандартизации используют референтные (стандартные) вероятности летального исхода. Ими могут быть средние показатели по городу в отделениях соответствующего типа.

Например, стандартная летальность (0,1), умноженная на число поступивших в отделение больных (800), дает ожидаемое число умерших в отделении ( $800 \times 0,1 = 80$ ). Такие ожидаемые числа умерших вычисляются для отделения каждого профиля. После этого достаточно суммировать ожидаемые числа умерших в отделениях больницы и сумму можно сопоставить с фактическим числом умерших в больнице. Сравнение может быть сделано в абсолютных величинах (число фактически умерших в сравнении с ожидаемым) или в показателях частоты, для чего ожидаемое и фактическое числа делятся на число поступивших в больницу.

Стандартизация исключает ошибку в оценке работы больницы, связанную с особенностями профиля и емкости отделений больницы, но не исключает других ошибок. Например, даже с учетом специализации коек летальность может быть более высокой в стационаре, куда поступают больные из района, населенного бедными людьми, работающими на производствах с вредными условиями труда, сравнительно с больницей, обслуживающей благополучную часть города. Связывать высокую летальность в таком случае с работой врачей будет ошибкой.

Чаще всего непрямую стандартизацию используют для получения *стандартизованного отношения смертности (standardized mortality ratio — SMR, так же вычисляется стандартизованная заболеваемость)* для сравнения смертности в нескольких группах населения или в нескольких странах. Стандартизованное отношение смертности равно отношению фактической смертности (числу смертей) к ожидаемой (ожидаемому



числу). Например, оно равно отношению смертности в изучаемой группе рабочих электродного производства к смертности в популяции у мужчин с соответствующим распределением по возрасту. При сравнении нескольких стран используются условные популяции, например условная «стандартная европейская популяция». Без стандартизации по возрасту затруднительно было бы сопоставление смертности от рака в Европе и республиках бывшего СССР, где продолжительность жизни существенно меньше (рис. 1.18).

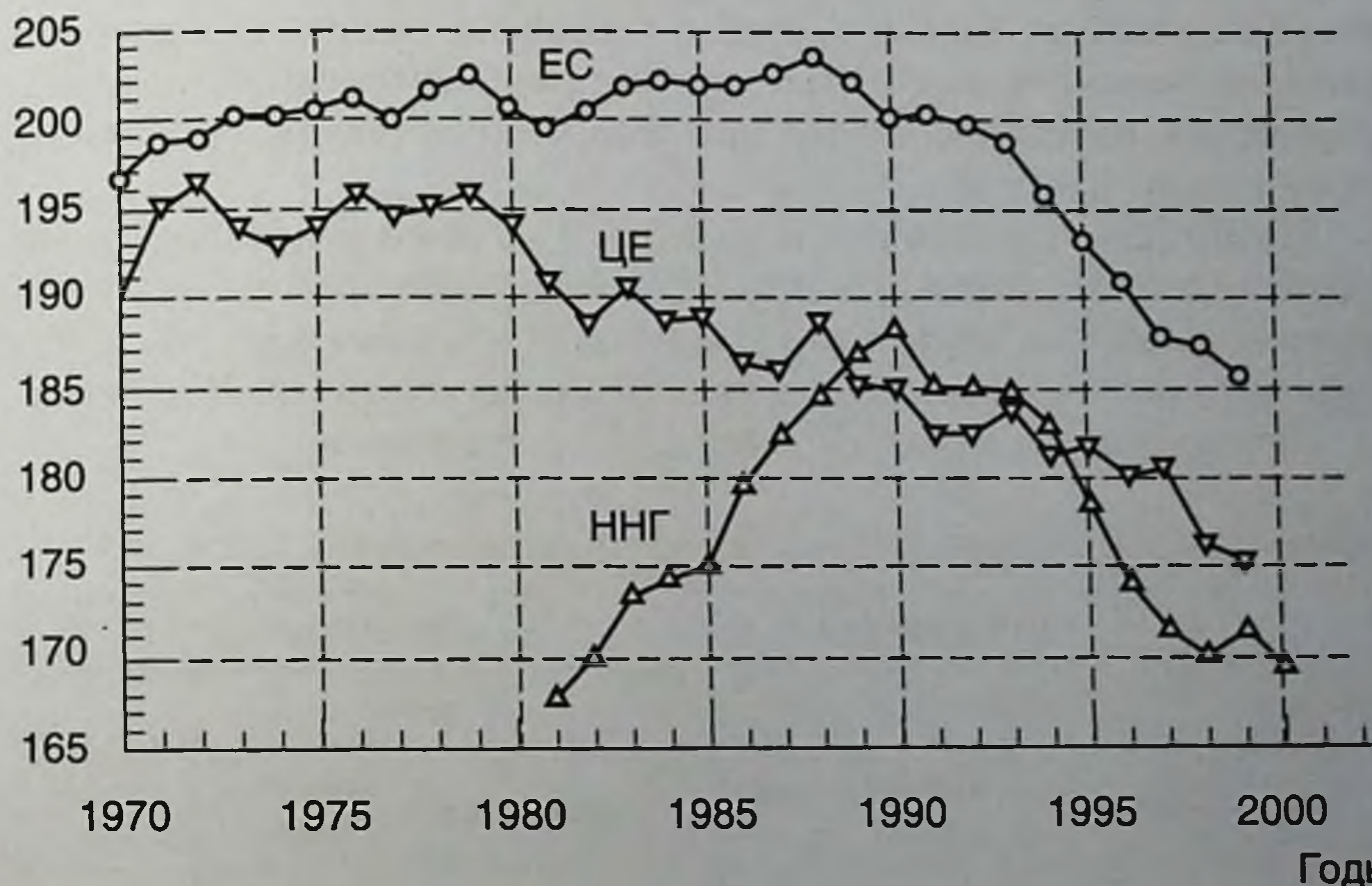


Рис. 1.18. Смертность от рака в республиках бывшего СССР (ННГ), в странах Европейского союза (ЕС) и в Центральной Европе (ЦЕ) на 100 000 населения всех возрастов. По данным The European Health Report 2002. WHO Regional Publications, European Series, no. 97. WHO Regional Office for Europe, Copenhagen. Отчет и исходные данные доступны: <https://www.euro.who.int/ru/data-and-evidence>

Стандартизация показателей смертности по возрасту не исключает других ошибок. Например, у рабочих химического производства смертность может быть найдена такой же, как в популяции соответствующего возраста. Однако подобный результат *дает основания предполагать наличие вредных условий на производстве*, так как работающие люди должны иметь и всегда имеют смертность ниже, чем смертность в целой



популяции. В целой популяции из-за высокой смертности неработающих смертность повышена сравнительно со смертностью работающих.

### Прямой метод стандартизации

В случае когда надо сопоставить два или более районов города или два производства по смертности (общей или специфической) либо когда надо оценить изменения в смертности во времени, применяют прямую стандартизацию. Примером сравнения районов может быть задача сравнения смертности в «экологически чистом» районе со смертностью в районе, где расположено хранилище боевых отравляющих веществ. Вариантом может быть изучение не общей смертности, а специфической, например от рака мочевого пузыря, или не смертности, а инциденса болезней.

Рассмотрим гипотетический пример, в котором сравниваются два производства, на которых занято одинаковое число работников и выявлено одинаковое число случаев изучаемой болезни (табл. 1.4). Более того, число случаев одинаково в каждой возрастной группе. Различается лишь возрастная структура занятых на производствах.

Таблица 1.4. Сравнение двух производств по частоте рака носоглотки

Возраст, лет	Производство А			Производство В		
	число		случаев на 1000	число		случаев на 1000
	работников	случаев рака		работников	случаев рака	
20–30	2500	0	0	1500	0	0
31–40	4000	6	1,5	3000	6	2
41–50	1000	2	2	2000	2	1
51–60	500	2	4	1500	2	1,33
<b>Всего</b>	<b>8000</b>	<b>10</b>	<b>1,25</b>	<b>8000</b>	<b>10</b>	<b>1,25</b>

Примечание. Общее число случаев рака на 1000 работников вычисляется не суммированием значений в колонке, а делением значений в строке «всего» — (число случаев/число работников) × 1000.

Учитывая различие работников по возрасту можно сравнить два производства по частоте рака. Примем за стандарт возрастную струк-



туру работников производства В. Для каждой возрастной группы частоту случаев на производстве А умножаем на число работников этого возраста на производстве В. В результате (табл. 1.5, последняя колонка) получаем число случаев, которое наблюдалось бы на производстве А, если бы там была такая же возрастная структура работников, как на производстве В. Благодаря тому что работники моложе, производство А имеет такое же суммарное число случаев, как производство В. С поправкой на особенности возрастной структуры работников частота рака в 1,5 раза выше на производстве А.

Таблица 1.5. Сравнение двух производств по частоте рака носоглотки

Возраст, лет	Производство А	Производство В		Число случаев на производстве А, если бы имела место возрастная структура В
		число		
	случаев на 1000	работников	случаев рака	
20–30	0	1500	0	0
31–40	1,5	3000	6	4,5
41–50	2	2000	2	4
51–60	4	1500	2	6
<b>Всего</b>	<b>1,25</b>	<b>8000</b>	<b>10</b>	<b>14,5</b>

Конечно, в качестве основы для сравнения в прямой стандартизации можно взять любую из сравниваемых групп. В нашем примере можно было бы взять за основу возрастную структуру работников производства А, и тогда мы получили бы число случаев для производства В при возрастной структуре производства А. Соответственно, это число случаев для производства В при возрастной структуре, свойственной производству А, было бы ниже 10. Точность вычисленного показателя стандартизованного отношения смертности отражается его доверительным интервалом (ДИ), в котором с заданной вероятностью, обычно 95%, находится популяционная величина (см. разд. 1.4.4).

Классические способы стандартизации по возрастной структуре — прямым и непрямым методами — не является единственным. Суще-



ствуют другие приемы стандартизации, которые позволяют в других ситуациях более эффективно сравнивать группы людей, внося поправки на их различия.

### Контрольные вопросы и задачи к разделу 1.3

1. В стране с населением 200 млн человек за год зарегистрировано 50 случаев чумы, в том числе 3 случая закончились смертью заболевших. Рассчитайте:
  - а) заболеваемость чумой;
  - б) летальность от чумы;
  - в) смертность от чумы.
2. Скорректированные на возраст показатели смертности используются для:
  - а) устранения различий в распределении по возрасту в сравниваемых популяциях;
  - б) коррекции ошибок в свидетельствах о смерти;
  - в) устранения различий в числе людей, о которых не собрана информация;
  - г) установления истинного числа умерших в данной возрастной группе.
3. При равном преваленсе болезни ее инцидент в 8 раз выше у детей, чем у взрослых. Наиболее вероятное объяснение:
  - а) у детей смертность ниже;
  - б) у детей специфическая смертность выше;
  - в) факторы риска болезни более распространены у взрослых;
  - г) длительность болезни у детей больше.
4. В выборке 1000 человек 50 страдают колитом, 100 — простатитом, 120 — той или другой болезнью. Такое распределение означает:
  - а) эти болезни возникают независимо;
  - б) больные колитом больше подвержены простатиту;
  - в) простатит снижает вероятность развития колита;
  - г) имеет место связь между этими болезнями.
5. Предположим, что при переливании крови вероятность заражения вирусной инфекцией составляет 5%. Если субъект в течение года получает 3 инфузии, то вероятность заражения в течение 2 лет:
  - а) 30%;      б) 5%;      в) 14%;
  - г) 26%;      д) 96,3%.



## 1.4. ВЫБОРОЧНЫЕ ИССЛЕДОВАНИЯ

Обычно исследователь не может изучать всю популяцию или всю профессиональную группу или всех больных данной болезнью. Поэтому он не может узнать точно, как часто встречается у больных тот или иной признак в популяции, или как много больных в популяции. Единственным доступным методом исследования является обследование приемлемых по размерам групп людей. Действительно, достаточно обследовать одну городскую школу, чтобы приблизительно оценить преваленс кариеса у школьников. Однако обследование одной случайно взятой школы может дать неверное впечатление о состоянии зубов у подростков страны. Это может произойти потому, например, что в данной городской школе учатся дети обеспеченных родителей. Жители бедных районов и сельской местности имеют худший доступ к стоматологической помощи, и у них чаще при обследовании обнаруживается отсутствие зубов как результат несвоевременного и упрощенного лечения.

### 1.4.1. Репрезентативность оценки

Выборка, которая правильно отражает изучаемую популяцию, называется *репрезентативной*. Если исследователь знает некоторые характеристики популяции, то в целях исследования он может подобрать репрезентативную школу по известным признакам (число учащихся, их успеваемость, заболеваемость, выполнение календаря прививок и т.д.). Проблема в том, что при этом нельзя исключить, что избранная школа отличается от других школ по неизвестному признаку. Например, ученики данной школы могли переболеть некоторой инфекцией, или подвергаться действию некоторого экологического фактора, или в силу особенностей педагогического коллектива иметь особые установки в отношении питания. Поэтому правильным решением является не использовать одну единицу выборки (одного человека, пациентов одного врача), а исследовать возможно большую выборку.

Простейший способ получить репрезентативную выборку — случайно выбирать из популяции субъектов для исследования. Случайная выборка (англ. *random*, отсюда — рандомизация) извлекает из популяции как типичные, так и нетипичные случаи. Важно, что в совокупности они дают самое правильное представление об изучаемой генеральной сово-



купности — о популяции, или о детях школьного возраста, или о рабочих типографий. Еще важнее, что при случайном выборе людей они отличаются от других в генеральной совокупности по всем, в том числе неизвестным, нам признакам, случайным образом. Иными словами, мы не знаем, как часто школьники города применяют не прописанные врачами лекарства, но если школы для обследования выбраны случайным образом, то обследованные школьники схожи со всеми школьниками города и по этому признаку. При случайной выборке из генеральной совокупности с увеличением числа людей в выборке ее свойства приближаются к свойствам генеральной совокупности. Это важно для применения статистических методов анализа данных.

Принципиальная простота простой случайной выборки из популяции сопровождается трудностями ее реализации. Обычно невозможно получить список, например, всех жителей города, чтобы сделать из него случайную выборку. Поэтому существует множество способов, которые для конкретного исследования позволяют получить репрезентативную выборку, соответствующую его задачам.

Ключевым является использование правильных элементов выборки. При изучении работы врачей единицами выборки являются врачи, а при изучении исходов заболеваний — пациенты. При исследовании болезней глаз иногда анализируют течение заболевания не отдельных людей, а отдельных глаз. Это увеличивает размер выборки, так как у некоторых больных может быть более одного больного глаза. В исследовании течения артрита подсчитывают число вовлеченных суставов, что увеличивает объем исследования многократно. Такая практика является ошибочной, поскольку суставы одного человека не являются независимыми друг от друга. Два глаза заболевают и подвергаются лечению в одном организме. Подобные события не являются независимыми и, как правило, не должны быть элементами выборки. Аналогичная ошибка возникает, когда в исследование повторно включают одних и тех же людей, когда у них наступает очередное обострение заболевания.

### 1.4.2. Случайная выборка

Случайный выбор не является «произвольным». В то время как человек может думать, что его решения случайны, они управляются скрытыми закономерностями. Например, при произвольном выборе



пациентов врач может подсознательно выбирать более тяжелых для обследования. В итоге может оказаться, что выбраны для обследования нерепрезентативные пациенты. Поэтому выборка обязательно должна выполняться с использованием надежных способов. Упрощения — *псевдорандомизация* — например, выборка в зависимости от дня недели, номера истории болезни, имени больного и т.д. — неприемлемы.

Проведение правильной рандомизации незатруднительно, и для этого достаточно средств. Во многих руководствах по статистике приводятся таблицы случайных чисел. Случайные числа могут быть получены многими компьютерными средствами, например в русской версии MS Excel<sup>1</sup> есть операторы СЛУЧМЕЖДУ() И СЛЧИС(). В интернете существуют бесплатные сайты с генераторами случайных чисел, например [www.gandom.org](http://www.gandom.org). Компьютерные программы позволяют генерировать случайные числа с разной закономерностью, например, имеющие нормальное распределение или равномерно распределенные в указанных границах.

Случайное число можно генерировать заново для каждого очередного пациента или можно подготовить заранее ряд чисел в виде, например, таблицы или карточек. Для того чтобы заранее заготовленные числа не были известны, их на карточках помещают в последовательно пронумерованные закрытые конверты. Конверт открывают для очередного пациента, когда принято решение о включении его в исследование. Особенно важной оказывается рандомизация при проведении контролируемых испытаний. В испытаниях рандомизация используется для того, чтобы сформировать одинаковые группы, в которых будут применяться сравниваемые вмешательства (см. с. 325).

### 1.4.3. Типичный и стандартный случаи

Описание «типичных случаев» в современной эпидемиологии используется редко, хотя в клинике традиционно для демонстрации

---

<sup>1</sup> Excel — это электронная таблица, входящая в состав популярного коммерческого пакета программ Microsoft Office. Не меньшие возможности предоставляет бесплатная программа 602Tab, доступная на <http://www.software602.com>, и многие другие программы. В учебнике Excel используется как инструмент для упражнений, но не рекомендуется для статистического анализа данных. Автором известного руководства по эпидемиологии К. Ротманом в Excel создана программа для изучающих эпидемиологию и статистику EpiSheet, доступная на <http://www.krothman.org/episheet.xls>.



студентам выбирают такие случаи заболеваний. Причина нерепрезентативности «типичных» случаев состоит в том, что «типичный» случай должен объединять в себе ряд «типичных» (часто встречающихся при данной болезни) признаков. В результате такой больной действительно легко узнаваем студентами, ему нетрудно поставить диагноз. Беда в том, что больных с таким сочетанием признаков очень мало. Большую часть больных на приеме у врача составляют больные с «нетипичными» проявлениями, когда тот или иной «характерный» симптом отсутствует. Только совокупность больных, избранных случайным образом из популяции, дает репрезентативную картину проявлений болезни. Поэтому врачи после обучения в кафедральных клиниках на примере «типичных» больных (традиционно на кафедрах «факультетской» терапии и хирургии) обязательно проходят последипломное обучение в амбулаториях и стационарах.

Статистические приемы, описанные ниже, пришли в клиническую практику и физиологический эксперимент только в начале XX в. До этого врачи в экспериментах коллекционировали наблюдения и неформально обобщали их, выделяя наиболее типичные, по их мнению, случаи течения болезни или развития реакции на экспериментальное воздействие. Так, в лабораториях И.П. Павлова из массы наблюдений отбирались те, которые хорошо отражали изучаемый процесс, и этими наблюдениями иллюстрировали утверждения. Возможно, именно из-за неформального синтеза экспериментальных данных исследования некоторых последователей И.П. Павлова не были позднее подтверждены. Так, например, произошло с исследованиями А.Д. Сперанского, вызывавшего полную картину инфекционных и других болезней путем раздражения мозговых структур.

В клинической практике диагноз обосновывается соответствием проявлений болезни у пациента набору диагностических критериев (см. разд. 2.2). Этот подход широко используется не только для диагностики, но и для принятия других решений — о необходимости операции, переводе на искусственное дыхание и т.д. В эпидемиологии используется *стандартное определение случая* — набор критериев, позволяющих отнести лицо к группе заболевших или к группе подвергшихся некоторому воздействию. Такие критерии обычно включают клинические признаки, данные лабораторных исследований, указание на время и место заболевания (см. разд. 2.2.9).



### 1.4.4. Точность оценки

При исследовании всех случаев генеральной совокупности эпидемиолог с исчерпывающей полнотой представляет ее характеристики. Так, при обследовании всей популяции можно точно установить преваленс болезни. При исследовании выборки точность представления о генеральной совокупности тем ниже, чем меньше выборка. Например, при обследовании 5 больных ( $n$  — число обследованных) искомое осложнение может быть выявлено у одного ( $S = 1$ ). Оценка частоты осложнений как  $0,2$  ( $S/n$ ) приближительна. Выявление всего лишь еще одного больного с осложнением изменит оценку в 2 раза ( $2/5$  или  $0,4$ ). Для того чтобы описать точность оценки относительной частоты  $p$ , вычисляют показатель «стандартная ошибка».

$$m = \sqrt{(p) \times (1-p) / n},$$

где  $n$  — число обследованных,  $p$  — относительная частота признака у этих обследованных,  $m$  — стандартная ошибка.

В случае когда  $S = n$  или  $S = 0$ , числитель должен стать равным единице или нулю. Ошибка оценки вероятности, конечно, в таком случае не исчезает, просто точных способов вычисления ошибки не существует. В таких ситуациях вероятность можно приблизительно оценить, сделав поправку на размер выборки. Вероятность с поправкой Ван дер Ваардена  $p = (S + 1) / (n + 2)$ . Таким образом, в случае когда симптом обнаружен у 5 больных из 5, наиболее вероятная оценка частоты в популяции составит  $6/7$ , т.е.  $0,9$ . При обследовании большего числа пациентов поправка будет иметь меньший размер. Например, при частоте в выборке  $20/20$  относительная частота с поправкой будет  $0,95$ .

Стандартная ошибка уменьшается при увеличении числа обследованных  $n$ , но лишь пропорционально корню квадратному из  $n$ . Поэтому при увеличении числа больных на 15 человек (от исходного результата  $2/5$  до  $8/20$ ) стандартная ошибка уменьшится в 2 раза (от  $0,22$  до  $0,11$ ), а при обследовании следующих 15 больных (всего  $14/35$ ) уменьшится только в 1,4 раза (до  $0,08$ ).

Для того чтобы обозначить пределы, в которых можно доверять полученной выборочной оценке, вычисляют «доверительный интервал» (ДИ, в англоязычной литературе — CI, *confidence interval*). Вычисление ДИ называют также интервальной оценкой параметра. Обычно



используют ДИ, в котором значение для генеральной совокупности находится с вероятностью 95%, реже — 90 и 99%. Для получения 95% ДИ надо  $m$  умножить на  $(t_{1-\alpha/2})$ , где  $\alpha$  — принятый уровень *альфа-риска*. Альфа-риск соответствует вероятности ошибки, которую мы готовы принять: считается обычно приемлемой 5% вероятность того, что популяционная частота находится за пределами данного интервала (см. также разд. 1.8). Нижняя граница ДИ будет равна  $p - m \times (t_{1-\alpha/2})$ , а верхняя —  $p + m \times (t_{1-\alpha/2})$ . При расчете 95% ДИ величина в скобках близка к 2 при численности пациентов в группе более 20, но быстро растет с численностью выборки менее 20. Величину  $t_{1-\alpha/2}$  можно получить в статистических таблицах (см. приложение 1), вычислить с помощью статистических программ. Программы статистического анализа позволяют вычислять ДИ при вычислении частоты, средней величины, коэффициентов регрессии и др.

Выше рассмотрено вычисление стандартной ошибки и ДИ для качественных признаков (для пропорции (доли)). Для количественных признаков, таких как возраст или концентрация, ДИ вычисляется аналогичным образом. Отличается лишь вычисление стандартной ошибки, что рассмотрено в разд. 1.6. В случае когда рассматриваются данные обо всей популяции, ДИ не вычисляется (см., например, рис. 1.15, 1.16). Это не означает, что подсчет числа погибших на дорогах страны абсолютно точен. Дело в том, что оценка получена не на основе анализа выборки, и потому ее неопределенность зависит в основном от систематических ошибок, влияющих на регистрацию. Чтобы учесть их, нужно применять другие методы.

#### 1.4.5. Статистическая мощность

Если болезнь наблюдается у 1 человека из 1000, то очевидно, что, обследовав 10 и даже 1000 человек, мы вряд ли выявим даже один случай. Для того чтобы обнаружить редкое событие, надо обследовать по крайней мере в 3 раза большее число людей, чем то, которому соответствует один случай. В нашем примере надо обследовать более 3000 человек. Способность обнаружить явление называют *мощностью* или *статистической чувствительностью*. Чаще всего термин «мощность» используется применительно к планированию исследования. Исходя из запланированного уровня альфа-риска (см. разд. 1.4.4), заранее вычисляют, сколько пациентов включить в исследование, чтобы явление можно было



обнаружить. Исследование на меньшем числе пациентов явление не обнаружит, но и не докажет, что его нет. Обследовать большее, чем нужно, число пациентов нерационально и принесет излишние неудобства пациентам. Компьютерная программа EpiInfo (см. сноску на с. 103) позволяет быстро оценить, что если в популяции размером 250 000 предполагается преваленс 2% и приемлемой верхней границей ошибки доверительного интервала является 4%, то с вероятностью 95% оценить преваленс можно, обследовав случайную выборку в 188 человек.

Статистическая чувствительность зависит не только от числа включенных в исследование пациентов, но и от того, какую структуру исследования используют и каких пациентов включают в исследование. Например, в клинических исследованиях можно выявить более тонкие изменения, если подбирать однородных пациентов. В зависимости от того, какая структура исследования планируется, используются различные способы расчета необходимой численности участников эксперимента (пациентов) (см. с. 327). Калькуляторы для таких расчетов широко доступны в интернете.

## Контрольные вопросы и задачи к разделу 1.4

1. Репрезентативность выборки может быть достигнута:
  - а) увеличением ее численности;
  - б) более тщательным обследованием каждого субъекта, включенного в выборку;
  - в) случайным выбором из популяции;
  - г) планомерным выбором представителей популяции из числа женщин, детей, разных профессий и т.д.
2. Для уменьшения размера случайной ошибки, связанной с аналитическими процедурами, можно:
  - а) брать пробы биологических жидкостей в контролируемых условиях;
  - б) увеличить численность исследуемой группы;
  - в) проводить параллельные многократные измерения одной пробы и усреднять результат;
  - г) начать выполнять программу обеспечения качества в лаборатории.
3. Доверительный интервал:
  - а) может быть сужен, если применять более точные измерения;



- б) может быть сужен, если исследовать выборку большей численности;
  - в) вычисляется обычно для интервала вероятности 95%.
4. Для вычисления инциденса в знаменателе пропорции используется:
- а) число потерянных для наблюдения людей;
  - б) человеко-годы наблюдения;
  - в) число бессимптомных случаев;
  - г) число новых случаев;
  - д) общее число случаев.

## 1.5. РЕГИСТРАЦИЯ И ПОДГОТОВКА ДАННЫХ В ИССЛЕДОВАНИИ

### 1.5.1. Типы переменных в исследовании

При обследовании пациентов исследователь получает разнообразные *данные*. При анализе их называют *признаками* или переменными (англ. *variables*). Типом переменной и структурой исследования определяется, какие манипуляции можно с этими переменными проводить, и какие статистические инструменты использовать.

#### Качественные признаки

Простейший вариант — альтернативные признаки, называемые также дихотомическими, экзистенциальными, бинарными. Относятся к состояниям и признакам, которые отсутствуют или присутствуют. Результаты исследования качественного признака надо описывать в виде относительной частоты, например «инцидент 0,1» или «летальность 3,5%»<sup>1</sup>. Иногда только альтернативные признаки называют «качественными», отделяя их от номинативных и ординальных.

---

<sup>1</sup> Слово «качество» и прилагательное «качественный» имеют много значений. Для эпидемиологии помимо значения «дихотомический» важно значение «качества» как способности предмета (медицинской услуги) удовлетворять потребности (см. разд. 5.2). Существует также особый класс медицинских исследований, называемых качественными, так как они изучают трудноизмеримые признаки и процессы, такие как переживания больного, его опыт и ощущения. Эти исследования в учебнике не рассматриваются. Подробнее см. [7].



**Номинативные** признаки называются также *полихотомическими* (рис. 1.19). Главное их отличие — признак имеет варианты, различающиеся между собой, но эти варианты нельзя ранжировать, соотнести по величине: диагноз, тип клетки. Качественные признаки являются простейшим случаем номинативных. Результаты исследования номинативного признака представляются только в виде доли от выборки, имеющей такой результат, например «из 260 препаратов 2 непригодны, остальные классифицированы: тип 1 — 32%, тип 2 — 26%, остальные — тип 3».

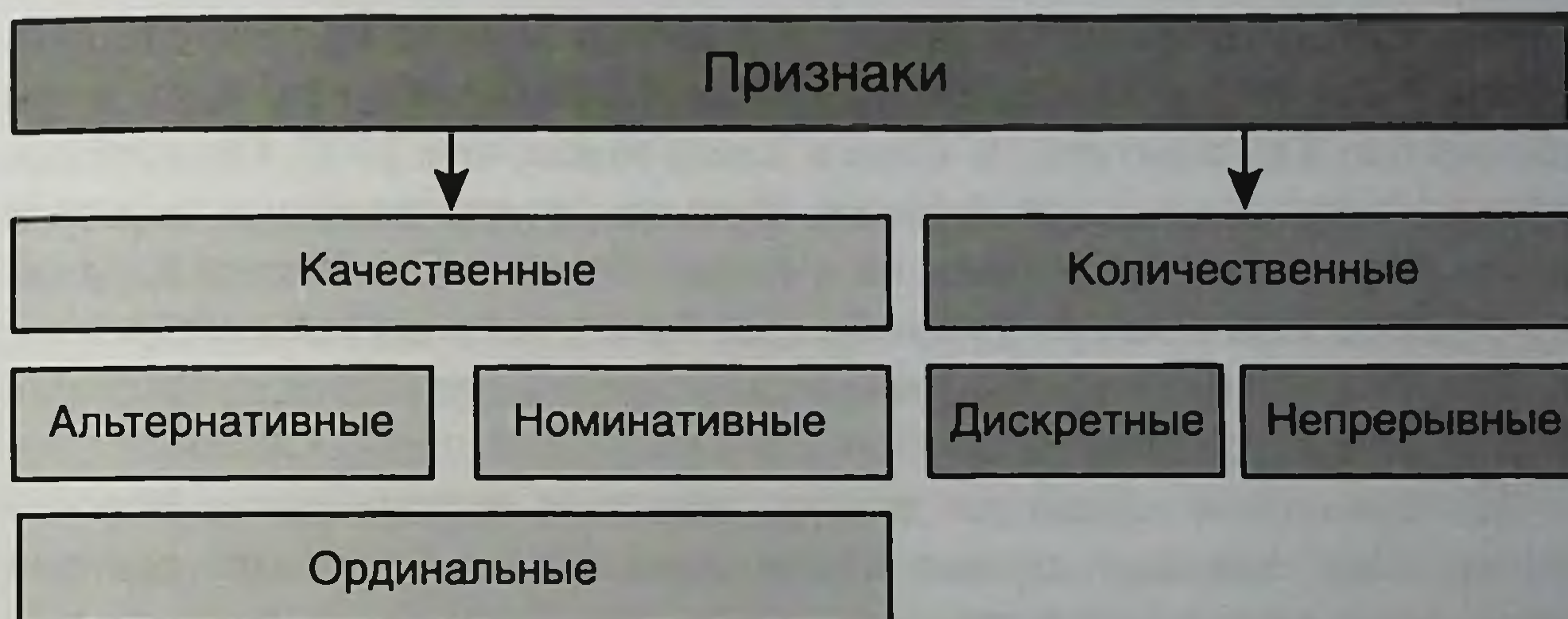


Рис. 1.19. Типы переменных

**Ординальные** признаки называются также *порядковыми* (от лат. *ordinalis* — порядковый). По значению таких признаков случаи (наблюдаемых больных) можно ранжировать, т.е. расположить в естественном порядке величины признака. Примером являются оценки силы землетрясения в баллах, тяжести состояния по условной шкале или стадия болезни. Важно, что «очень тяжелое состояние» не находится в определенном отношении с «состоянием средней тяжести» — не в 2 или 22 раза тяжелее. Мы знаем, что тяжелее, но не знаем насколько или во сколько раз. Неправильно вычислять среднюю величину для ординальных признаков. Результаты исследования ординального признака должны представляться в таком же виде, как и номинативного, — в виде частоты или относительной частоты каждой оценки.

### Количественные признаки

Количественные признаки называются также *интервальными*. Могут иметь *дискретную шкалу* (число приступов, число детей) или *непрерыв-*



ную (концентрация, масса, длина). Представляют собой некий высший тип переменных, с которыми возможны наиболее эффективные статистические манипуляции. Отличием интервальных признаков является то, что не только существует возможность ранжировать отдельные случаи, но и шаг между отдельными ступеньками шкалы одинаков: числам соответствует определенная единица измерения. Поэтому над интервальными признаками можно выполнять разнообразные арифметические операции: складывать, делить, умножать.

Обычно интервальная шкала обладает также свойством *шкалы отношений*. Это означает, например, что объект массой 30 г тяжелее объекта массой 15 г в 2 раза. На шкалах с произвольной нулевой точкой это не так. Температура 30 °С не в 2 раза выше, чем 15 °С. Отчасти поэтому люди, привыкшие к шкале Цельсия, затрудняются в переводе температуры из шкалы Цельсия в шкалу Фаренгейта. Шкала Кельвина (абсолютной температуры) обладает свойствами шкалы отношений, и перевод в градусы шкалы Кельвина из производной шкалы (Цельсия или Фаренгейта) нетруден.

Интервальные признаки иногда называют *параметрами*. Это нежелательно, так как термин «параметр» обычно используется для обозначения параметров распределения (средняя, мода, среднее квадратическое отклонение). Результаты исследования количественного признака могут быть представлены в таком же виде, как для номинативных признаков, но полнее они представляются в виде гистограммы (см. разд. 1.7.1).

### **Редукция переменных**

В медицинской практике многие признаки, которые могут быть измерены с большой точностью, традиционно измеряются качественно или полуколичественно. Например, концентрация белка в моче оценивается качественными методами, хотя более дорогие количественные методы тоже существуют. Они могут стать более широко используемыми, если их полезность для ведения пациентов, например, для выявления нефропатии у больных сахарным диабетом будет показана.

При анализе данных нередко создаются новые переменные на основе полученных первично. Исследователь может, например, редуцировать количественные данные до качественных: лиц, у которых концентрация холестерина более 220 мг/дл, оценивать как лиц с гипер-



холестеринемией, а прочих — как с нормохолестеринемией. Такая редукция упрощает анализ, но приводит к потере информации.

## 1.5.2. Производные переменные

### Сравнение со стандартом

Одной из основных причин отказа от первично измеренной переменной является стремление сократить вариабельность<sup>1</sup>. Например, физическую работоспособность человека непосредственно *измеряют* в единицах мощности — ваттах. Эта величина широко варьирует в зависимости от пола, возраста, массы тела человека. Поэтому *выражение* физической работоспособности в ваттах может заменяться на проценты от должной для лиц данного пола, роста, массы тела и возраста. Устранение вариабельности, связанной с этими признаками, не представляющими самостоятельного интереса, позволяет лучше выявить отклонения, связанные с заболеванием. Точно так же поступали в первой половине XX в., оценивая массу тела (тогда использовался термин «вес»): поскольку она существенно различается у мужчин и женщин и прямо связана с ростом, устанавливали популяционные нормативы. Обычно ими были средние величины массы тела для лиц данного пола, роста и возраста или «идеальные» (наиболее благоприятные для долголетия). Массу отдельного человека *выражали* в процентах к такому стандарту. Выраженную в виде процентов к стандарту массу легче было *оценивать*, применяя простые критерии (например, >120% — ожирение).

### Индексы

Традиционным подходом к устранению влияния привходящих признаков является вычисление индексов, т.е. отношений между признаками, позволяющих правильнее оценить один из признаков. Например, с 1980-х годов массу тела принято выражать «индексом массы тела» (индексом Кетле II) (ИМТ), вычисляемым как:

$$\text{ИМТ} = \text{вес (кг)} / \text{рост (м)}^2.$$

---

<sup>1</sup> Вариабельностью (реже вариативностью, изменчивостью) называют свойство всех переменных иметь разные значения у разных людей и у одного человека при повторных измерениях. В узком смысле вариабельность означает неравенство полученных значений признака, широкий разброс значений. В эпидемиологии вариабельность является одним из основных объектов исследования.



Этот индекс позволяет выразить массу тела, устранив влияние на нее роста — главного признака, влияющего на массу тела. Оценка проста:  $ИМТ \geq 30$  — ожирение. Эта простота привлекательна для медицинской практики и большинства эпидемиологических исследований.

Такого рода индексы широко применяются в медицине: например, ударный объем сердца относят к поверхности тела (сердечный индекс) для того, чтобы устранить влияние различий размера тела на оценку сердечного выброса. Выбор поверхности тела в качестве знаменателя дроби не имеет под собой теоретического основания, просто так влияние размера тела устраняется лучше всего. Индексы имеют эмпирическое происхождение и обычно представляют собой компромисс между точностью (степенью коррекции влияния привходящих переменных), наглядностью и простотой вычислений.

Чтобы устранить влияние на результат лечения значительных различий между пациентами, иногда вычисляют изменения в ходе лечения в процентах к исходной величине. Вычисление процентов к исходной величине приводит к искажениям и обычно недопустимо [8].

### 1.5.3. Другие типы данных

#### Шкалы

Некоторые признаки, исследуемые в медицине, эпидемиологии, социальных науках, не существуют объективно (материально) и не могут быть надежно измерены, как измеряется масса или концентрация, но могут быть описаны, выражены и оценены через субъективные оценки самого человека, внешнего наблюдателя или через результаты выполнения человеком заданий. Такими *конструктами* являются, например, депрессия, шизоидность, интеллект. Для выражения таких признаков используют *шкалы*.

Некоторые из шкал очень просты. Так, например, зрелость новорожденного измеряется по шкале Апгар. Итоговое значение «зрелости» по ней образуется путем простого сложения баллов, начисленных за сердцебиение, дыхание, мышечный тонус и т.д. Аналогичным образом измеряется «качество жизни» (SF-36 и другие методы), степень нарушения гомеостаза (шкала АРАСНЕ) и множество других сложных качеств, оцениваемых в медицине. Интегральный показатель затем оценивается с помощью простого критерия. Например, при определенной сумме баллов прогноз выживания называется *неблагоприят-*



ным, или субъект оценивается как обладающий циклоидными чертами личности. Строго говоря, такие шкалы являются не количественными, а ординальными. Однако число возможных уровней на такой шкале обычно довольно велико, и с оценками по шкалам исследователи обращаются нередко как с количественными признаками.

### Аналоговые шкалы

Для выражения своих ощущений и предпочтений человека можно попросить сдавливать пружину тем сильнее, чем сильнее боль или зуд, или отразить на нарисованной шкале штрихом свою оценку испытываемой боли (рис. 1.20). Пациент отмечает на шкале свое представление, и через час, день или месяц можно сопоставить результаты новой оценки с предыдущей. Иногда аналоговую шкалу переводят в линейную, заменяя крайние значения, например, на 0 и 10 и равномерно разделяя шкалу на 10 единиц. Такая трансформация некорректна и грозит ошибками. Нет оснований считать, что 10 мм в начале шкалы соответствуют 10 мм в конце шкалы. Кроме того, субъективные шкалы у людей

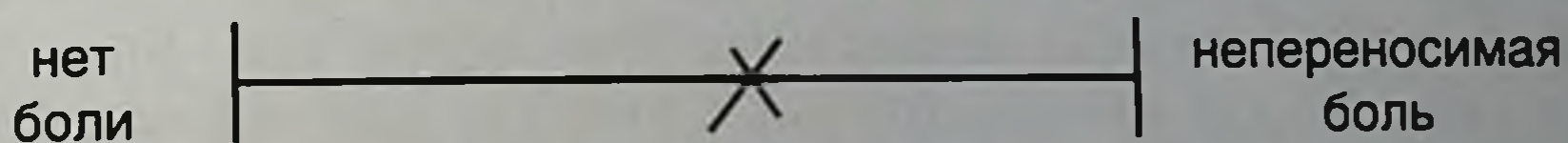


Рис. 1.20. Визуальная аналоговая шкала для оценки боли

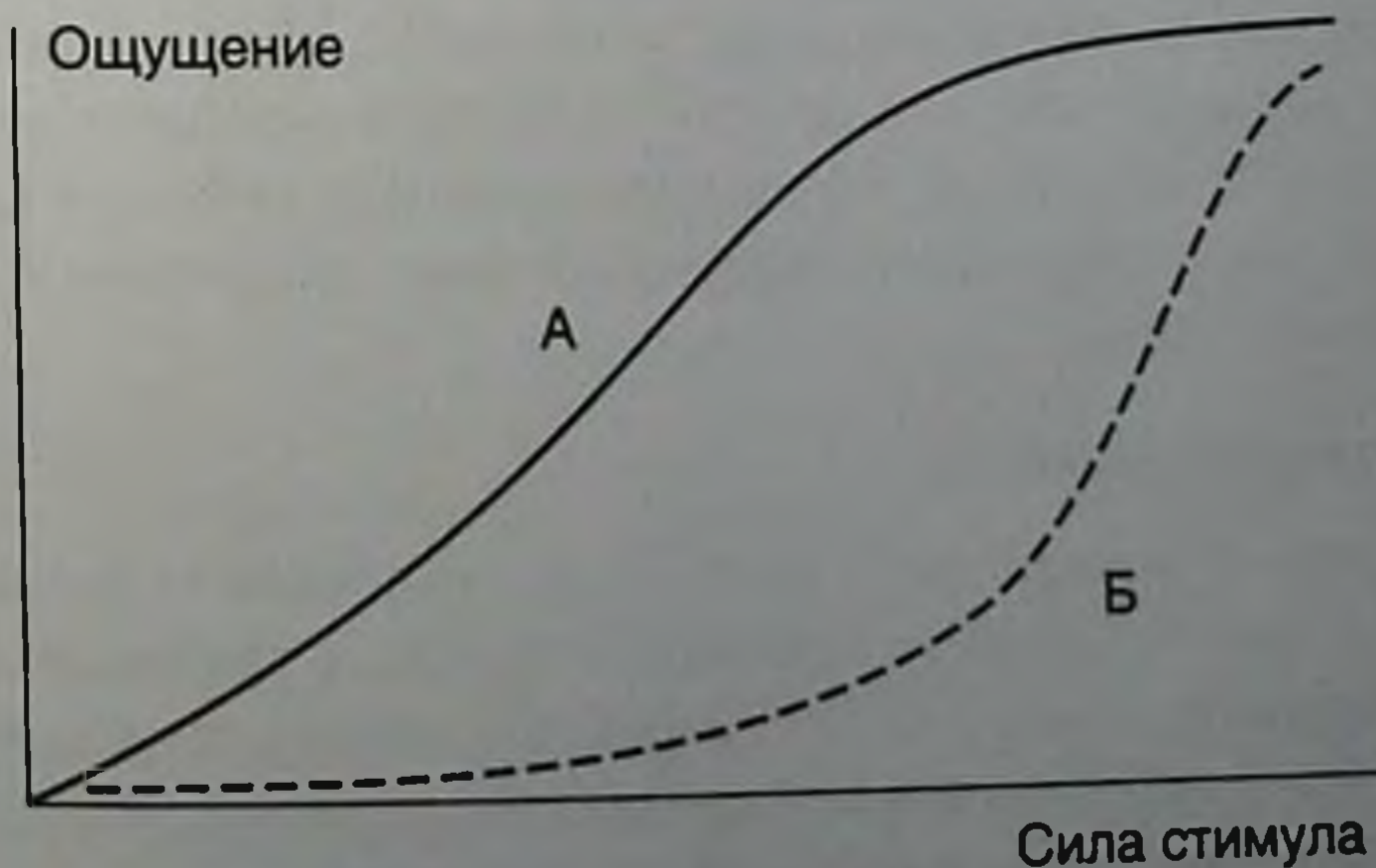


Рис. 1.21. Индивидуальные варианты различения стимула: одни люди хорошо различают слабые стимулы (А), другие — сильные (Б)



разные. Они подобны неравномерно растянутому портновскому «метру». У одних людей большую часть шкалы занимают и хорошо дифференцируются слабые ощущения (боль, зуд, запах), у других — сильные ощущения (рис. 1.21). Более того, со временем страдающий от болей человек переживает адаптацию к ним, сопровождающуюся изменением субъективной шкалы. Поэтому сопоставлять оценки разных людей надо с большой осторожностью, и равным образом надо избегать обобщения таких оценок методами, разработанными для количественных признаков.

#### 1.5.4. Измерения

Источником данных для анализа являются результаты исследований и измерений и документы, в которых зарегистрированы такие результаты. В процессе работы исследователь регистрирует наблюдения в рабочих записях (журналах, формах) или в электронной форме на компьютере. Эти записи могут быть словесными (описательными), графическими (изображения, полученные техническими средствами, — от фотографии до томограммы или зарисовки), цифровыми. Иногда уже по ходу наблюдения исследователь классифицирует свои наблюдения и вводит коды для наблюдаемых вариантов, например для вариантов облысения — буквенные. За кодом стоит соответствующая методика его получения. Эта методика должна быть разработана и детально описана так, чтобы другой специалист с аналогичной подготовкой мог повторить наблюдение (измерение) с теми же результатами. Например, мог точно так же оценить облысение у пациента АА, как это сделал коллега 12 мес назад. Невоспроизводимые события не являются объектом науки, и невозпроизводимые измерения не могут использоваться в научных целях.

#### Компоненты вариабельности

Наблюдения и измерения у отдельных субъектов (*в разных случаях*) и у одного субъекта в разных условиях и в разное время дают различающиеся результаты. Эти результаты являются объектом статистического анализа. В совокупном вариационном ряду<sup>1</sup> результатов измерений мы наблюдаем сумму действия множества факторов.

<sup>1</sup> Совокупность результатов измерения, ранжированных по величине, называют «вариационный ряд».



Основным компонентом общей вариабельности является *аналитическая вариабельность*, являющаяся следствием неточностей работы приборов, дозировки реактивов, ограниченной точности работы лаборантов и др. Обычно исследователь-врач не заинтересован в анализе происхождения этих отклонений от истинного значения измеряемой величины и удовлетворяется мерами по снижению аналитической вариабельности, проводимыми специалистами лабораторного отделения и техниками, ответственными за поверку измерительных приборов.

*Внутрииндивидуальная вариабельность* — это колебания измеряемой величины у отдельных людей в разное время и в разных условиях. Эта вариабельность может мешать исследованию. Например, суточная вариабельность массы тела и концентрации глюкозы в плазме крови мешает измерить их «истинное» значение. Поэтому приходится определять условия, в которых измерение считается «правильным» (истинным), например, утром натощак. Внутрииндивидуальная вариабельность нередко бывает самостоятельным объектом медицинских исследований. Исследования изменений некоторого признака в процессе лечения, например показателя выраженности депрессии, являются вариантом измерения внутрииндивидуальной вариабельности.

*Межиндивидуальная вариабельность* — это различия измеряемой величины между отдельными людьми. Эта вариабельность часто служит объектом медицинских исследований, например, при описании проявлений болезни в сравнении с «нормой» или другими болезнями (рис. 1.22).

При анализе совокупности данных исследователь имеет дело с результатами измерений, в которых проявились все компоненты вариабельности. Для того чтобы эффективнее решить задачу, исследователь стремится к тому, чтобы изучаемый компонент вариабельности был максимален, а другие компоненты вариабельности — минимальны. Тогда на фоне прочих вариаций можно увидеть, например, различие между здоровыми и больными.

В медицинских исследованиях всегда стремятся уменьшить аналитическую вариабельность. Высокая аналитическая вариабельность, зависящая от непонятных причин или от различий в работе лаборантов, может приводить к тому, что разница результатов у двух пациентов будет объясняться не различиями их состояния здоровья, а тем, что пробы попали на анализ к разным лаборантам.



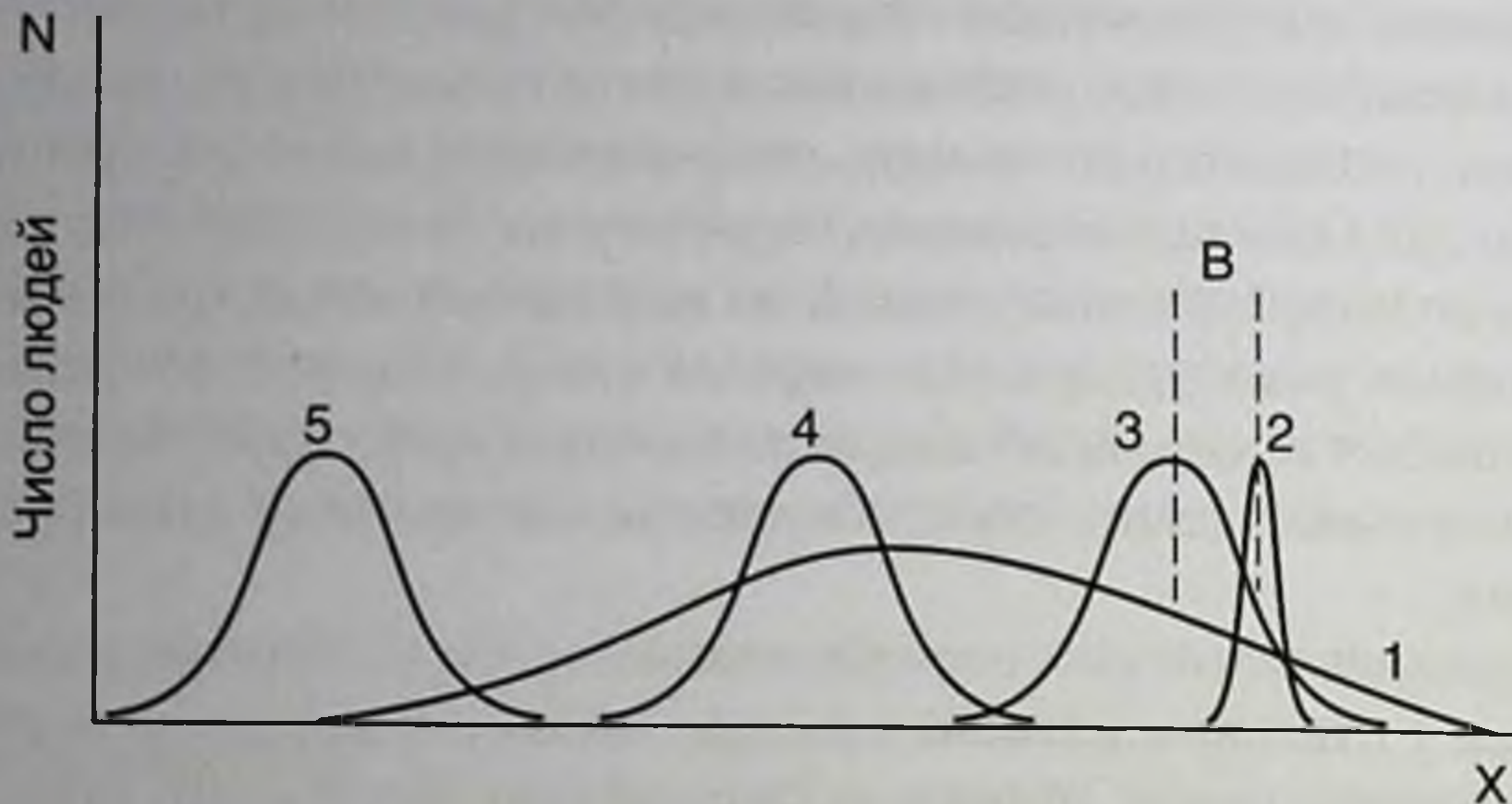


Рис. 1.22. Соотношение внутрииндивидуальной, межиндивидуальной и аналитической вариабельности с изменениями при заболевании. Истинные значения признака  $X$  у здорового человека колеблются в узких пределах (2). Лабораторное исследование вследствие аналитической ошибки дает увеличенный разброс результатов (3) и систематическую ошибку (смещение  $B$ ). В процессе развития заболевания признак изменяется (4), но на ранних стадиях болезни обычно находится в пределах межиндивидуальной вариабельности (1). Лишь при дальнейшем развитии болезни величина признака выходит за пределы типичных для популяции величин (5)

При изучении внутрииндивидуальных изменений стараются помимо аналитической подавить межиндивидуальную вариабельность. Для этого из изучаемой группы пациентов исключают тех, кто резко отличается от прочих (например, имеет тяжелое сопутствующее заболевание), проводят измерения в стандартных условиях, применяют коррекцию известных межиндивидуальных различий (например, применяют для измерения артериального давления манжетки различной ширины в зависимости от диаметра плеча). Чтобы внутрииндивидуальные изменения были заметнее и лучше выявлялись в исследовании, выбирают пациентов с благоприятными для исследования характеристиками. Например, при изучении эффективности медицинских вмешательств подбирают:

- ▶ пациентов, у которых в предварительном испытании выявляется хорошая реакция на препараты изучаемой группы;
- ▶ пациентов, которые правильно выполняют назначения;
- ▶ пациентов с большей тяжестью болезни, так как у таких пациентов величина изменений под влиянием лечения обычно бывает больше;



- ▶ а также силу воздействия (дозу) выбирают достаточно большой, чтобы вызвать заметные изменения.

*При изучении межиндивидуальных различий* помимо уменьшения аналитической вариабельности стремятся уменьшить внутрииндивидуальную. Для этого проводят измерения в стандартизованных условиях, дают пациентам отдых перед измерением и учитывают при анализе данных сопутствующие факторы, например интенсивность курения или занятий физкультурой обследуемыми.

Правильная организация исследования состоит в том, чтобы сделать изучаемый компонент вариабельности максимальным, а другие слагаемые вариабельности свести к минимуму. В этом же смысле суть статистического анализа заключается в соотнесении величины изучаемого компонента вариабельности с величиной прочих компонентов вариабельности. Если это отношение велико, то статистические критерии<sup>1</sup> подтвердят гипотезу исследователя<sup>2</sup>. Если исследование спланировано так, что изучаемый компонент вариабельности мал сравнительно с межиндивидуальными различиями обследованных субъектов и аналитическими вариациями, то исследователь не сможет выделить этот «полезный сигнал» из «шума» мешающих вариаций. Например, годовой ритм артериального давления может быть не виден на фоне больших различий между людьми.

*Регрессия к средней величине* является важным феноменом, который проявляется во всех продольных исследованиях. Вследствие действия аналитических вариаций и из-за кратковременных изменений в состоянии людей получаемые при измерении величины могут быть завышены, занижены или соответствовать долговременным индивидуальным особенностям. Если величина, измеренная в первый раз, не была существенно смещена вследствие случайных и кратковременных вариаций,

---

<sup>1</sup> Статистическими критериями, или критериями для проверки гипотез, называют математические показатели, вычисляемые на основе экспериментальных данных для того, чтобы установить, не являются ли полученные результаты, подтверждающие гипотезу исследователя или опровергающие ее, только лишь случайными. Подробнее см. разд. 1.8.2.

<sup>2</sup> Здесь имеются в виду гипотезы о различии, например, между двумя группами пациентов или между состоянием до и после вмешательства. В эпидемиологии и медицине проверяются гипотезы о неразличимости (нулевая гипотеза) вместо гипотезы о различии, например между двумя группами пациентов или между состоянием до и после вмешательства. Если нулевая гипотеза отвергается, принимается основная гипотеза — о различии.



то при последующем измерении она изменится непредсказуемо, *в среднем* для таких субъектов — не изменится. Если же величина была завышена, то она *в среднем* приблизится к средней. Чем более она была завышена (например, вследствие большой аналитической ошибки), тем больше она сдвинется к средней, типичной для популяции величине. В случаях, где наблюдалась заниженная вследствие случайных и кратковременных вариаций величина, *в среднем* будет сдвиг к средней — повышение. Если бы имели место только аналитические вариации, то регрессия к средней полностью реализовывалась бы при втором измерении. Поскольку внутрииндивидуальные вариации могут быть длительными, возврат к средней, например, после болезни или после изменения образа питания, может занимать месяцы и годы. На рис. 1.23 видно, что повышение артериального давления с годами наблюдается

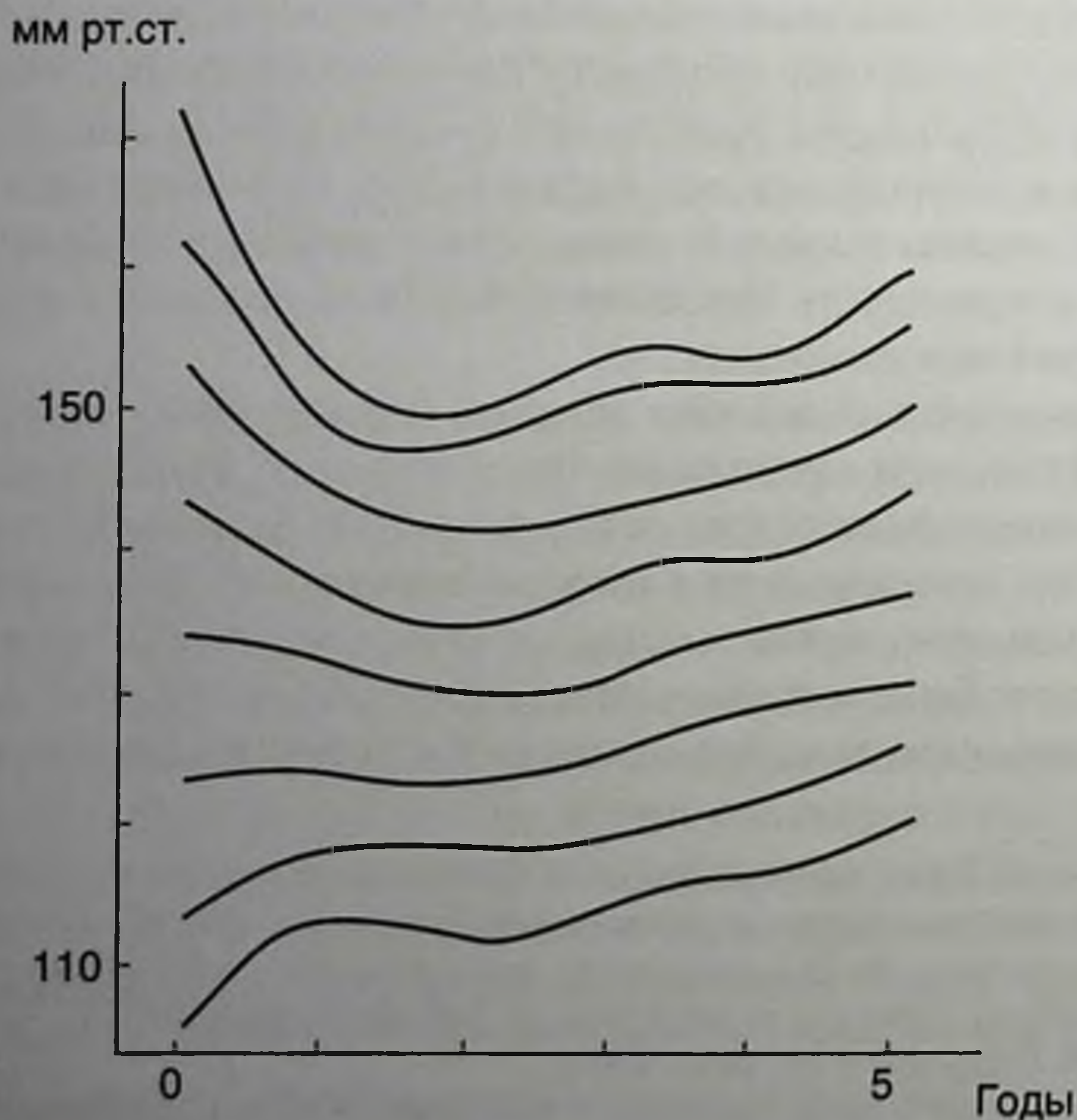


Рис. 1.23. Изменения систолического артериального давления в течение многолетнего проспективного исследования у людей с разным исходным давлением. По данным R.I. Havlic et al. *Clinical Approaches to High Blood Pressure in the Young*. Wright-PGS Publ., 1982



примерно одинаково у лиц с разным исходным давлением, но в течение первого года наблюдения у лиц с высоким давлением наблюдается снижение давления как регрессия к средней. Наоборот, у лиц с самым низким давлением в первый год давление быстро увеличивается, также как регрессия к средней.

### Метод измерения

*Измерения*, проводимые у отдельных лиц, обладают важными свойствами, которые необходимо учитывать при организации измерения и при последующем анализе данных. Прежде всего, измерения должны проводиться правильными средствами и по установленным правилам. Даже для простого измерения роста (длины тела) существуют правила. До возраста 36 мес длина тела измеряется лежа, а после 2 лет (рост) — стоя. В периоде, перекрывающемся двумя методами, можно применять оба способа. Неисполнение правил измерения, в том числе правил взятия пробы биологической жидкости, ее транспортировки и хранения, может приводить к получению неправильных, невоспроизводимых результатов, не сравнимых с результатами, полученными другими исследователями, т.е. непригодных для вынесения диагностических и терапевтических решений.

Аналитическая характеристика метода измерения является важной частью исследования и должна быть получена до того, как начинается исследование. Часто исследование не проводится самым точным из существующих методом, позволяющим получить наиболее близкое к истинному значению измеряемой величины, поскольку такие *референтные методы* (дающие эталонное по точности измерение) обычно трудоемки или дороги.

Отклонения результатов измерения от истинного значения являются ошибками (погрешностями) измерения, независимо от причины отклонения. *Случайной ошибкой* называют отклонения от истинной величины, которые в среднем равны нулю, т.е. в среднем не изменяют измеряемой величины. *Систематической ошибкой* (от англ. *bias* — смещение) называют смещение среднего результата измерения по отношению к истинной величине. Например, в результате наличия других веществ в крови результат измерения концентрации глюкозы может быть завышен, т.е. не соответствовать истинной концентрации.

Представления о случайной и систематической ошибках относятся не только к инструментальным измерениям переменных у пациентов,



но и к измерениям факторов физической и социальной среды, клинических исходов, эффективности вмешательств, т.е. ко всем ситуациям измерения. Так, в испытании лечебного вмешательства величина лечебного эффекта (снижение летальности) измеряется различием в летальности между контрольной группой и группой вмешательства. Оценка различия может быть смещена (занижена), если часть пациентов контрольной группы будет самостоятельно вместо плацебо принимать лекарство, получаемое пациентами группы вмешательства.

Хороший тест, например способ измерения массы жира в теле, должен обладать свойством *воспроизводимости* (англ. *reproducibility*), т.е. давать близкие результаты в разных условиях у одного человека. В одинаковых условиях при повторном выполнении он должен давать очень близкие результаты. Это свойство называют *сходимостью* (англ. *precision, repeatability*). Тест может давать неточные результаты, но в *среднем* соответствующие истине. Это свойство называют *правильностью*. Даже при невысокой сходимости, т.е. наличии случайных колебаний измеряемой величины, тест может давать правильные результаты. Идеальный тест должен быть *точен* (англ. *accuracy*), т.е. каждый результат его применения близок к истине (результату референтного измерения). У такого теста, естественно, малы систематические и случайные погрешности, т.е. высоки сходимость и правильность.

Результаты измерения могут быть оценены при повторных исследованиях стандартной пробы биологической жидкости. Например, при концентрации вещества X, равной 20 у.е., лаборатория А дает результаты, близкие к истине (рис. 1.24), лаборатория В — неточные, но правильные. Эти результаты могут быть использованы, если вместо одного измерения проводить, например, пять и вычислять среднюю величину. Это доступно в исследовательской практике, но неприемлемо в обычной клинической работе. Лаборатория С дает неточные и неправильные результаты, но хорошо воспроизводимые. Обычно такие результаты нельзя использовать, но при отсутствии точного метода результаты, получаемые в лаборатории С, можно использовать, вводя поправку на смещение результатов относительно истинной концентрации. Конечно, это возможно, если только это смещение постоянно, например возникает вследствие использования в лаборатории особого прибора или отличающихся реактивов. Результаты лаборатории D неточны, неправильны и невоспроизводимы. Они непригодны для использования.



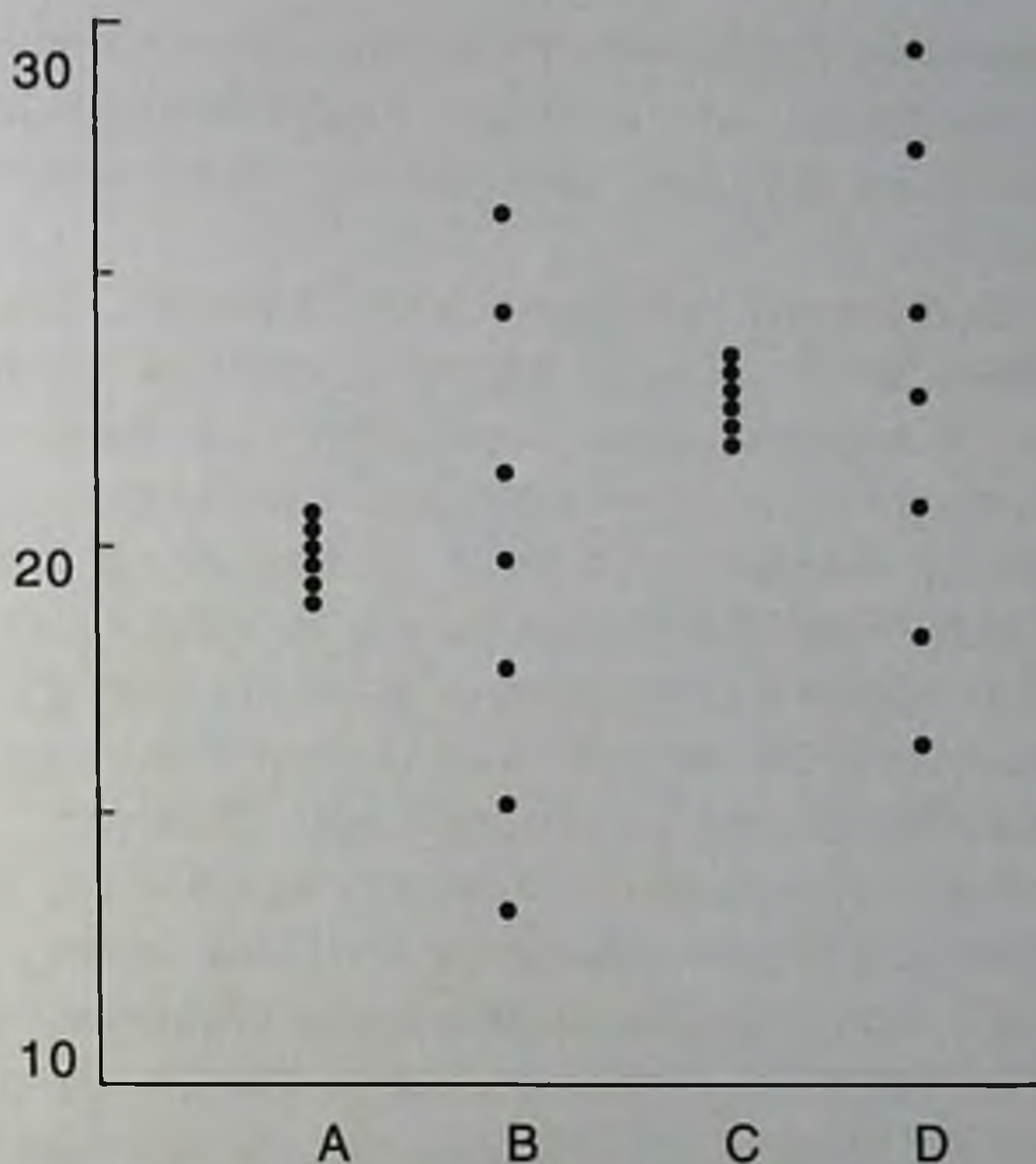


Рис. 1.24. Результаты повторных измерений стандартной пробы в четырех лабораториях. Лаборатория А дает точные результаты. В — результаты неточны, но правильны. С — результаты неточны и неправильны, но воспроизводимы. D — результаты неточны, неправильны и невоспроизводимы. Пояснения в тексте.

Характеристики метода измерения зависят от многих причин: от прибора, реактивов, особенностей исследуемой группы пациентов. Многие применяемые в медицине измерения имеют небольшую точность и воспроизводимость. Например, традиционное для нашей страны измерение температуры тела в подмышечной впадине ртутным максимальным термометром имеет невысокую воспроизводимость и относительно длительно. Более точное измерение в прямой кишке культурно неприемлемо для многих стран, и даже там, где практиковалось, от него отходят, например, в пользу измерения в полости рта. Электрические термометры быстрее дают устойчивое показание. Как физический измерительный прибор — электротермометр точен. К сожалению, оральная температура слишком вариабельна: результат измерения у человека плохо воспроизводится и недостаточно хорошо соответствует температуре ядра тела. В некоторых странах переходят к измерению температуры в наружном слуховом проходе, но ни удоб-



ство, ни воспроизводимость этого способа также неудовлетворительны. Врач не может избежать этих проблем. Единственным способом плодотворно измерять на практике является учет особенностей метода измерения.

*Оценки специалистами изображений*, например, гистологических препаратов, рентгенограмм или эндоскопической картины, обычно имеют невысокую воспроизводимость. При повторной оценке одних и тех же препаратов или рентгенограмм с интервалом в несколько месяцев специалисты обычно дают иное заключение в 10–30% случаев. Специалисты, оценивающие изображения, не всегда дают одинаковые заключения. Это зависит не только от квалификации, поэтому мнение «младшего» не может быть просто отвергнуто — не соглашаются между собою и специалисты высокого класса. В связи с этим в исследовательских целях необходимо проверять надежность таких оценок. Обычной процедурой в исследовании является проведение повторных независимых оценок несколькими специалистами и согласование их мнений в случае расхождения, например, с привлечением «внешнего» эксперта, дискуссии или процедуры голосования. Согласие экспертов является мерой надежности метода оценки и может быть измерено (см. разд. 2.2.2).

Исследователь должен документировать принцип метода измерения, реактивы. Аналитическая характеристика метода измерения должна быть документирована в цифровом виде. Обычно достаточно привести сведения о том, какими были результаты исследования стандартных проб (биологических жидкостей, рентгеновской плотности и т.д.). Эти результаты приводят в виде параметров распределения (медианы, средней величины и стандартного отклонения, вычисление которых подробнее рассматривается в разд. 1.6).

Если исследование предполагает повторное измерение, то необходимо учесть сам эффект повтора. При повторном исследовании внешнего дыхания или выполнении тестов на операторскую деятельность обследуемые чувствуют себя увереннее, приобретают навыки. Обычно такие изменения имитируют выздоровление (эффект лечения). Если не принимать их во внимание, то можно не обнаружить терапевтического эффекта в группе вмешательства в сравнении с контрольной группой. Чтобы уменьшить влияние эффекта тренировки, всех участников тренируют до начала исследования. Такая короткая тренировка не исключает дальнейшего приспособления обследуемых к повторному



применению физической нагрузки или психологического опросника<sup>1</sup>, но уменьшает размеры этих изменений. Дело в том, что процесс приспособления (тренировки) развивается быстро в начальных стадиях, при первом предъявлении задания. Затем совершенствование возможностей замедляется, и эти менее выраженные эффекты приспособления могут быть учтены при статистическом анализе данных.

### 1.5.5. Документальные источники данных

Почти в каждом исследовании используются не только объективно измеряемые (регистрируемые) исследователем признаки, но и сведения, сообщаемые обследуемыми людьми и получаемые из документов. Сведения из этих источников также подвержены случайным и систематическим ошибкам.

Документальные источники принято делить на первичные и вторичные. *Первичными источниками* называют протоколы лабораторных исследований, истории болезни (карты пациентов). Вторичными — отчетные данные или сводные таблицы, составленные по первичным документам. В первичных источниках могут быть ошибки, связанные с ошибками считывания показаний приборов, ошибки записи, неясно сделанные записи (не поддающиеся дешифровке или неправильно прочитываемые), а также систематические ошибки, связанные с неполным сообщением обследуемыми об осуждаемом поведении (курение, употребление алкоголя, наркотиков, число половых партнеров), искаженные сообщения о доходе и пр. Искажения могут намеренно вноситься медицинскими работниками в случае, если регистрируемый признак используется для оценки работы врачей или для суждения о благополучии в районе (стране). Например, хронология поступления больного с тяжелой травмой и оказания ему помощи до наступления смерти может искажаться в истории болезни, а причина смерти «отравление алкоголем» может заменяться на более социально приемлемую.

---

<sup>1</sup> Опросником называют набор специально разработанных вопросов, предназначенный для сбора информации. Опросники могут быть краткими, всего в несколько вопросов, или содержать сотни разнообразных вопросов и заданий, например для изучения психологических особенностей человека. Опросники обычно заполняются человеком самостоятельно, в отличие от подобных форм, предназначенных для интервью (собеседования).



С 1990-х годов в развитых странах больницы все чаще ведут электронные истории болезней. Помимо всей информации, доступной обычно в первичном документе, они предоставляют быстрый доступ к дополнительной информации — электронным копиям рентгеновских снимков, например. При наличии таких электронных документов становится возможным оперативное проведение исследований, которые ранее были бы затруднительными (см. пример на рис. 1.8).

*Вторичные источники информации* привлекательны, так как в них информация уже собрана по первичным источникам, и ее легче включить в анализ. В современных информационных системах вторичные источники могут быть доступны в электронной форме, что делает их использование еще более привлекательным. Недостатками вторич-



**Рис. 1.25.** «Айсберг» медицинской информации. Исследователь, который пользуется только данными официальной регистрации заболеваний, наблюдает только вершину айсберга. Уведомления не представляются на некоторые случаи заболеваний, не всегда лабораторный тест дает положительный результат у заболевшего, не у всех больных удается взять материал для анализа, не все больные обращаются за помощью и не у всех развивается болезнь при экспозиции (контакте с больным или с вредным фактором среды)



ных источников являются их неполнота (в отчетные формы включаются не все сведения, доступные в первичных документах), накопление случайных ошибок в процессе переноса сведений из первичных документов, внесение составителями умышленных искажений. Последнее никогда не может быть полностью исключено, поскольку результаты отчетных форм являются основанием для оценки работы врачей, деятельности медицинских учреждений и благополучия в регионах. Обычно исследователь может предполагать направление возможных искажений, но он не может оценить размер искажений, пока не сопоставит отчетные сведения с первичной документацией. Поэтому многие эпидемиологи, использующие вторичные источники, применяют разнообразные приемы коррекции известных им ошибок. Результаты исследования непосредственно зависят от того, какой вид информации использует исследователь и о каком из аспектов изучаемого явления (рис. 1.25).

Извлечение данных из документации должно поручаться подготовленным исполнителям. Качество их работы необходимо проверить при допуске к работе и затем периодически контролировать. Исполнитель может знать план исследования, но надо организовать работу так, чтобы данные извлекались из всех документов одинаковым образом. В противном случае заинтересованный исполнитель может более детально изучать записи в документации у одних больных и менее тщательно — у других, в результате может возникнуть ложное различие между группами больных.

### 1.5.6. Формирование базы данных исследования

В результате измерений и сбора данных по первичным и вторичным источникам у исследователя накапливается архив, состоящий из материальных объектов (например, гистологических препаратов), первичных документов (например, заполненных бланков), копий вторичных документов. Обычно такие материалы не пригодны для анализа. В докомпьютерную эру составляли итоговые формы (карты) на каждый случай (обследуемого, исследуемую больницу или территорию). Затем на основании этих карт, содержащих всю полученную информацию, составляли *аналитические таблицы*. По таблицам проводили статистический анализ, готовили графики. Теперь использование компьютеров стало повсеместным, благодаря эффективности компьютерного анализа, его полноты и значительным уменьшением числа ошибок.



Обязанностью исследователя является содержание материалов исследования в такой форме и в таком месте, чтобы к ним имели доступ только уполномоченные лица, и чтобы деликатная информация об участниках исследования не могла быть разглашена. Это касается не только информации об участниках-пациентах, но также и о других участниках, например врачах, предоставивших исследователю информацию. Сохранение первичных материалов и итоговых документов является обязанностью исследователя. Даже в том случае, когда исследование проведено на средства и по инициативе самого исследователя, информация должна сохраняться в интересах общества.

### Создание файла

Данные исследования вводятся в компьютерную базу данных, характеристики которой зависят от объема, особенностей данных и планируемого анализа. В случае ограниченных по объему данных (небольшое число случаев и небольшое число признаков) можно воспользоваться *электронными таблицами*, например программой Excel. Если известно, с помощью какого пакета статистических программ будет проводиться анализ, то рационально вводить данные с помощью электронной таблицы, которая обычно есть в составе такого пакета. Каждая программа записывает данные на компьютере в один или несколько файлов (записей в компьютерных кодах). Данные расположены в файле в виде (*формате*), приспособленном для средств этой программы. В такой записи помимо собственно вводимых данных присутствует разнообразная служебная информация программы. Поэтому обычно файл, созданный одной программой, нельзя использовать в другой программе. Некоторые программы имеют встроенные функции, позволяющие прочитать «чужой» файл и записать его в собственном формате. Эта процедура называется «импорт». Поскольку большинство компьютерных программ с годами совершенствуется, в них изменяется и формат файлов. Поэтому, например, даже если используемая программа имеет функцию импорта файлов Excel, она в версии 2012 г. может быть неспособной импортировать файлы формата более позднего.

Если не известно точно, какими средствами будут анализироваться данные, то целесообразно создать файл, который импортируется во все программы. Таким форматом является файл в кодировке ASCII (текстовый, произносится как «аскии»). Он не содержит служебной информации, которую всегда записывают статистические программы



или электронные таблицы. В нем содержатся только данные, разделенные пробелами или другими разделительными знаками (например, запятой). Такой формат файла использует программа EpiInfo, разработанная в Центре по контролю заболеваний США и ВОЗ для эпидемиологов<sup>1</sup>. Программа позволяет спланировать исследование, разработать учетные формы, ввести и проверить данные, провести статистический анализ.

Данные вводятся в электронную таблицу обычно в виде последовательных записей, каждая строка которых соответствует одному пациенту (случаю) (рис. 1.26). Правильной практикой является присвоение каждому участнику цифрового кода. Это позволяет сделать записи анонимными и легко идентифицировать случаи. Последнее особенно

Microsoft Excel - таблица NEW.xls									
File Edit View Insert Format Tools Data Window Help									
C56									
	A	B	C	D	E	F	G	H	I
1			Вопросы I шкалы вопросника Dermatology specific q						
2			1	2	3	4	5	6	7
3	Кодовый номер	Дата опроса	сухость	чувствительность	болезненность	зуд	жжение	другое	дискомф
52	1270108/1.2	07.02.02	3	3	1	2	2	0	2
53	1270108/1.2	14.02.02	1	1	0	0	0	0	2
54	1270108/1.2	28.02.02	2	2	1	3	2	0	3
55	1270108/1.2	28.03.02	0	0	0	0	0	0	0
56	2170312/2.1	07.02.02				3			2
57	2170312/2.1	14.02.02				3			2
58	2052401/2.3	24.01.02	0	0	0	0	2	1	1
59	2052401/2.3	14.01.02	1	0	0	0	1	0	0
60	5050116/1.2		3	2	1	3	3	2	1
61	5050116/1.2	18.04.02	0	0	2	2	2	0	2

Рис. 1.26. Пример таблицы с данными исследования больных атопическим дерматитом. В колонках C56, D56... — отсутствие данных. Данные А. Мартынова, с разрешения автора

<sup>1</sup> Программу можно бесплатно получить на сайте Центра контроля над заболеваемостью (CDC) <http://www.cdc.gov/epiinfo/index.htm>. Современная версия EpiInfo существует не только для Windows, но и для мобильных устройств. Учебник по анализу данных с помощью EpiInfo см. <http://www2.sjsu.edu/faculty/gerstman/EpiInfo/>



важно, если создается не один файл, а несколько. Их позднее можно объединить, используя в качестве метки случая его номер.

В колонках записываются значения признаков. Последовательность признаков не имеет значения, но для удобства заполнения и последующего пролистывания таблицы признаки следует располагать в последовательности их появления на первичных формах и в последовательности форм.

В случае если в проекте участвует множество исследователей, таблицу данных естественно сделать доступной в сети, скоординировав доступ к ней и ее безопасность. Ввод данных может быть с мобильных устройств, у постели больного или при обследовании на производственной площадке.

*Количественные признаки* записываются с той точностью, с которой проводилось измерение. Важно вводить результаты измерения в одинаковых единицах, например рост у всех пациентов в метрах. В противном случае возникнет необходимость позднее исправлять данные, и появятся новые ошибки. *Ординальные признаки* вводят в том виде, в каком они получены (цифрами). *Номинативные* признаки обычно кодируют цифрами. Например, для качественных «да» заменяют на 1, «нет» — на 0. Некоторые программы позволяют анализировать номинативные признаки, записанные в виде буквенных или словесных выражений. Диагнозам, симптомам и подобным признакам обычно присваивают индивидуальные коды. Группировать диагнозы или симптомы и использовать укрупненные категории, например «кашель есть» и «кашля нет», в то время как у пациентов регистрировали варианты кашля и отделение мокроты, не следует. При необходимости объединение симптомов и группировка могут быть выполнены позднее. Обозначение номинативных признаков цифрами создает опасность случайной манипуляции с ними позднее как с количественными признаками. В некоторых статистических программах существует возможность указать в таблице тип признака (номинативный, количественный), и при статистическом анализе программа будет сопротивляться попытке выполнить неверную манипуляцию.

У одного обследуемого может быть более одного симптома или более одного источника дохода. В таком случае для каждого варианта признака создается отдельная колонка, в которой кодами 1 и 0 записывается наличие или отсутствие данного симптома (или размер дохода из данного источника). В случае если планируется лишь упрощенный



анализ и точно известно, что полного анализа не предстоит, можно ограничиться в соответствии с планом исследования записью только «основного источника дохода» или «первого симптома».

Всегда имеются случаи, в которых по одному или более признакам нет данных. Это может происходить в результате утраты документов, отказа опрашиваемых отвечать на отдельные вопросы, неисправности оборудования при обследовании и по другим причинам. При отсутствии данных в ячейку таблицы записывается код пропущенного значения, например –999. Важно, чтобы вводить код было легко и чтобы его значение никогда не могло быть спутано с реальными данными. Пропуски в данных, образно называемые «дырками» (так как они выглядят в электронной таблице пустыми по сравнению с окружающими заполненными ячейками), крайне нежелательны. Они снижают эффективность анализа данных и могут приводить к искажениям результатов. Многие методы статистического анализа позволяют анализировать только «полные случаи», и это приводит к уменьшению размера выборки. «Паллиативным средством» является заполнение пропусков количественных данных средними значениями, случайными значениями или значениями, вычисленными по величине других переменных. При этом количественно выборка увеличивается, но анализируемая переменность сокращается. Ни один из способов замещения пропусков данных не предотвращает снижения эффективности анализа, поэтому при проведении исследования должно быть сделано все возможное для получения максимально полных данных.

### **Проверка данных и предотвращение ошибок**

При вводе данных в электронную таблицу возможно возникновение новых ошибок. Эффективным способом выявления таких ошибок является создание одинаковых компьютерных файлов двумя операторами и последующее их сравнение. Поскольку операторы вряд ли сделают одинаковые ошибки, обнаружение различий в файлах позволит исправить практически все ошибки ввода данных. Этот способ дорог, но применяется там, где объем данных велик, а последствия ошибок существенны. В последние десятилетия все более распространяется регистрация данных в электронной форме с самого начала, с процедуры аппаратного измерения и опроса пациента. При этом уменьшается вероятность ошибок и сокращаются расходы на перевод данных в электронную форму.



В любом случае ввод данных завершается их проверкой. Таблица небольшого объема может быть вычитана, а в случае большой таблицы данных проводится анализ каждой переменной и совокупностей переменных. Для каждой переменной строится гистограмма, и на ней выявляются случаи со значениями слишком малыми и слишком большими, для того чтобы быть правильными (рис. 1.27, а). Каждая такая «выскакивающая» величина сверяется с первичной документацией и при обнаружении ошибки исправляется. Затем для тесно связанных между собой переменных строятся графики рассеяния (рис. 1.27, б). На них выявляются возможные ошибки, которые не были видны на гистограммах отдельных переменных. Кроме того, применяются логические условия для выявления случаев, где они не соблюдаются, например дата установки диагноза ранее даты рождения, отсутствие диагноза, по которому все пациенты включались в исследование, и т.д.

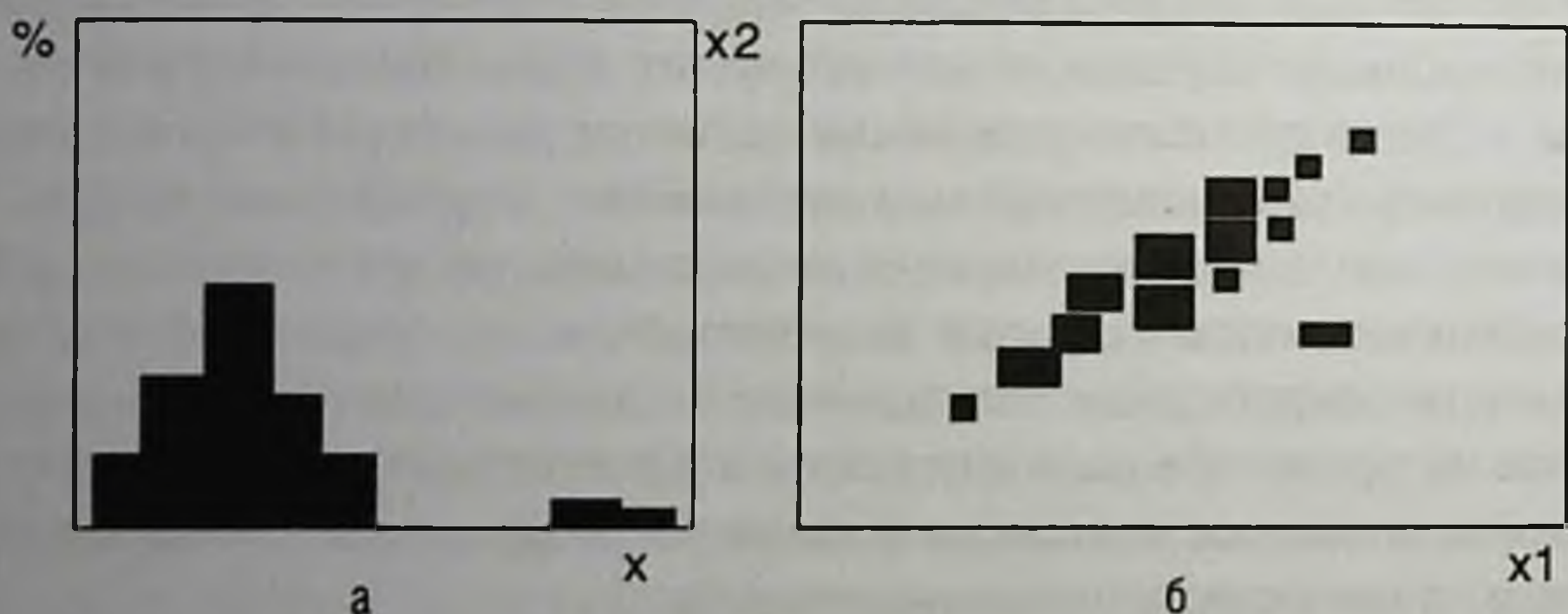


Рис. 1.27. «Выскакивающие» величины на гистограмме (а) и на скаттерграмме (б)

Уже в ходе проверки данных обнаруживается, что некоторые «выскакивающие» величины (слишком большие или маленькие по сравнению с основным числом случаев) правильны, в том смысле, что соответствуют первичным документам. Строгих правил, определяющих, что такое «выскакивающая» величина, т.е. нетипичная в смысле соответствия прочим случаям, не существует. Иногда используется критерий, согласно которому к «выскакивающим» относятся величины, отклоняющиеся от средней более чем на три средних квадратических отклонения (СКО) (см. разд. 1.6.2). «Выскакивающие» *правильные* величины не должны исключаться. Они могут исключаться только при одновременном исключении всего случая (пациента) и только тогда,



когда при проверке данных выяснится, что выскакивающее значение отражает наличие у пациента состояния (признака), дающего основание его исключить по критериям исключения, использовавшимся при формировании выборки. Например, в случае, если у него найдено тяжелое сопутствующее заболевание, не установленное в период включения пациентов в исследование.

Случаи с выскакивающими значениями оказывают большое влияние на результаты некоторых видов статистического анализа, значительно большее, чем случаи со значениями, близкими другим типичным значениям. Для того чтобы установить, насколько велико влияние выскакивающих значений на результат анализа, его проводят с включением этих значений и без включения, а затем сопоставляют результаты. Совпадение результатов будет свидетельствовать о том, что подтверждение гипотезы исследователя (или неподтверждение ее) не связано только лишь с наличием выскакивающих значений.

## 1.6. ОЦЕНКА РАСПРЕДЕЛЕНИЯ ВЕЛИЧИН

### 1.6.1. Оценка центральной тенденции

Целям описательных исследований служат методы описательной (дескриптивной) статистики. При регистрации любых количественных или ординальных признаков в эпидемиологическом, биологическом, социологическом, психологическом или медицинском исследовании обычно получается совокупность более или менее различающихся результатов. Эту совокупность результатов, ранжированных по величине, называют «вариационный ряд». Особый интерес представляет выяснение той величины (ранга), которая наиболее типична для изучаемого признака. Эта наиболее типичная величина называется *центральной тенденцией*. Применительно к ординальным признакам центральная тенденция соответствует тому рангу, который встречается чаще всего.

Применительно к количественным признакам в зависимости от того, как мы определяем «типичное», центральная тенденция может быть определена по-разному. Если мы полагаем, что типичное соответствует той величине, которая чаще всего регистрировалась, то тогда вычисляется *мода* ( $M_o$ ). Если предпочтительнее считать типичной середину ряда — *медиана* ( $M_e$ ). Если мы думаем, что правильнее типичной



считать величину, уравновешенную влиянием всех больших и малых величин, — *среднюю арифметическую* ( $\bar{X}$ ). Решение о выборе способа определения центральной тенденции зависит от формы распределения результатов. Средняя правильно отражает центральную тенденцию только в случае *симметричного распределения*, когда она совпадает с модой и медианой. В случае несимметричного распределения вычислять среднюю не следует (рис. 1.28).

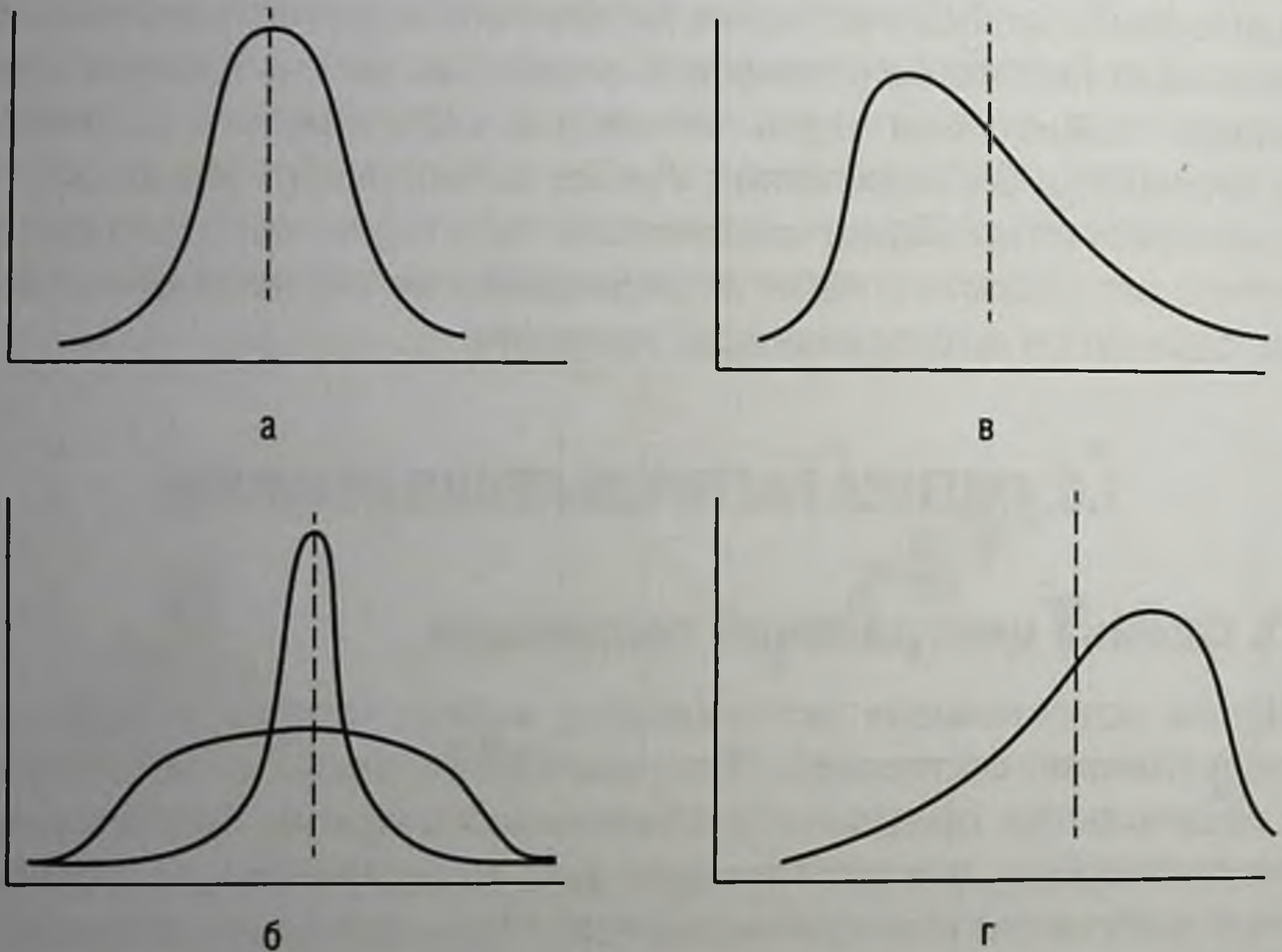


Рис. 1.28. Примеры форм распределения. Симметричные (а, б), островершинное и плоское (б), асимметричные (в, г), с правосторонней (в) и левосторонней (г) асимметрией. Вертикальная пунктирная линия обозначает положение средней арифметической

Средняя так часто используется для описания центральной тенденции, что в обыденном языке стала синонимом типичного значения. Средняя арифметическая вычисляется суммированием всех результатов и делением суммы на число результатов, что записывается в виде формулы как:

$$\bar{X} = \sum_{i=1}^n x_i / n,$$



где  $\bar{X}$  — средняя (икс с верхним подчеркиванием). Символ  $\sum$  — математический знак суммирования измеренных величин  $x_i$ . Выражения над и под знаком суммирования означают, что суммируется весь ряд величин  $x$ , обозначенных индексом  $i$ , от 1 до  $n$ ;  $n$  — число измеренных значений переменной  $x$ . Эта простая процедура реализована не только в специализированных пакетах программ для статистического анализа (например, Statistica, SPSS, SAS, BMDP), но и в пакетах программ широкого назначения, например в электронных таблицах (MS Excel и др.).

Представим, что при измерении артериального давления в случайной выборке из школьников-спортсменов получены результаты: 100, 105, 105, 110, 110, 110, 110, 110, 115, 115, 115, 120, 120, 120, 120, 125, 125, 125, 125, 130, 130, 135, 140. Для вычисления средней их можно суммировать и разделить на 23 (число результатов,  $n$ ). Будет нагляднее, если мы применим прием докомпьютерной эпохи и рассортируем результаты в таблицу (табл. 1.6).

Таблица 1.6. Сгруппированные результаты измерения артериального давления в выборке школьников

Число случаев	Артериальное давление, мм рт.ст.								
	100	105	110	115	120	125	130	135	140
1	*	*	*	*	*	*	*	*	*
2		*	*	*	*	*	*		
3			*	*	*	*			
4			*		*	*			
5			*						
Всего	1	2	5	3	4	4	2	1	1

Обратим внимание на повторяющиеся результаты, что облегчит суммирование:  $100+2\times 105+5\times 110+3\times 115+4\times 120+4\times 125+2\times 130+135+145$ . Для большого числа измерений во времена, когда компьютеры были недоступны, такой способ ускорял дело. Это называли *вычислением по сгруппированным данным*.

Теперь, когда нам доступен компьютер, введем результаты в электронную таблицу Excel в первую колонку (а) сверху вниз, с первой



ячейки по 23-ю (табл. 1.7). Ниже, в ячейке a24 достаточно ввести команду «=sum(a1:a23)», чтобы получить сумму (в русской версии Excel «=сумм(a1:a23)»; внимание: в русской версии команды пишутся русскими буквами, а обозначения ячеек — латинскими!). Чтобы получить среднюю, можно в ячейке a25 ввести формулу (=a24/23). Однако правильнее использовать встроенные в программу средства статистических вычислений. Введем в ячейку a26 специальную команду «=average(a1:a23)» («=срзнач(a1:a23)» в русской версии). Ее выполнение даст нам среднюю арифметическую без промежуточных вычислений. В разнообразных электронных таблицах и статистических программах написание команд различается, но вычисления приводят к одному результату. Лишь в сложных расчетах, где применимы способы, различающиеся в тонких деталях, применение разных статистических программ может давать несколько различающиеся результаты.

Таблица 1.7. Пример результатов измерения количественного признака: артериальное давление в малой группе подростков

N	Артериальное давление, мм рт.ст.		N	Артериальное давление, мм рт.ст.	
	систолическое	диастолическое		систолическое	диастолическое
1	100	75	13	120	70
2	105	70	14	120	75
3	105	75	15	120	80
4	110	75	16	125	80
5	110	70	17	125	85
6	110	70	18	125	80
7	110	75	19	125	70
8	110	80	20	130	75
9	115	75	21	130	80
10	115	80	22	135	80
11	115	70	23	140	85
12	120	80			



Асимметричное распределение может иметь правостороннюю (см. рис. 1.28, в) или левостороннюю *скошенность* (см. рис. 1.28, г). Первая встречается значительно чаще в результатах биологических и медицинских измерений. Для описания скошенных (несимметричных) распределений следует вычислять моду или медиану.

### Нормализация данных

При испытании устойчивости к укачиванию кандидатов на летное обучение с помощью метода прерывистой кумуляции ускорения Кориолиса (испытуемого вращают в кресле Барани, и он ритмично наклоняет туловище к коленям и выпрямляется, этот тест применяют отоларингологи) получены следующие данные (рис. 1.29). Распределение имеет ярко выраженную правостороннюю асимметрию. В этой ситуации, для того чтобы получить симметричное распределение и иметь возможность использовать закономерности нормального распределения и многие другие статистические методы, обычным решением является переход к логарифму измеряемой величины (времени).

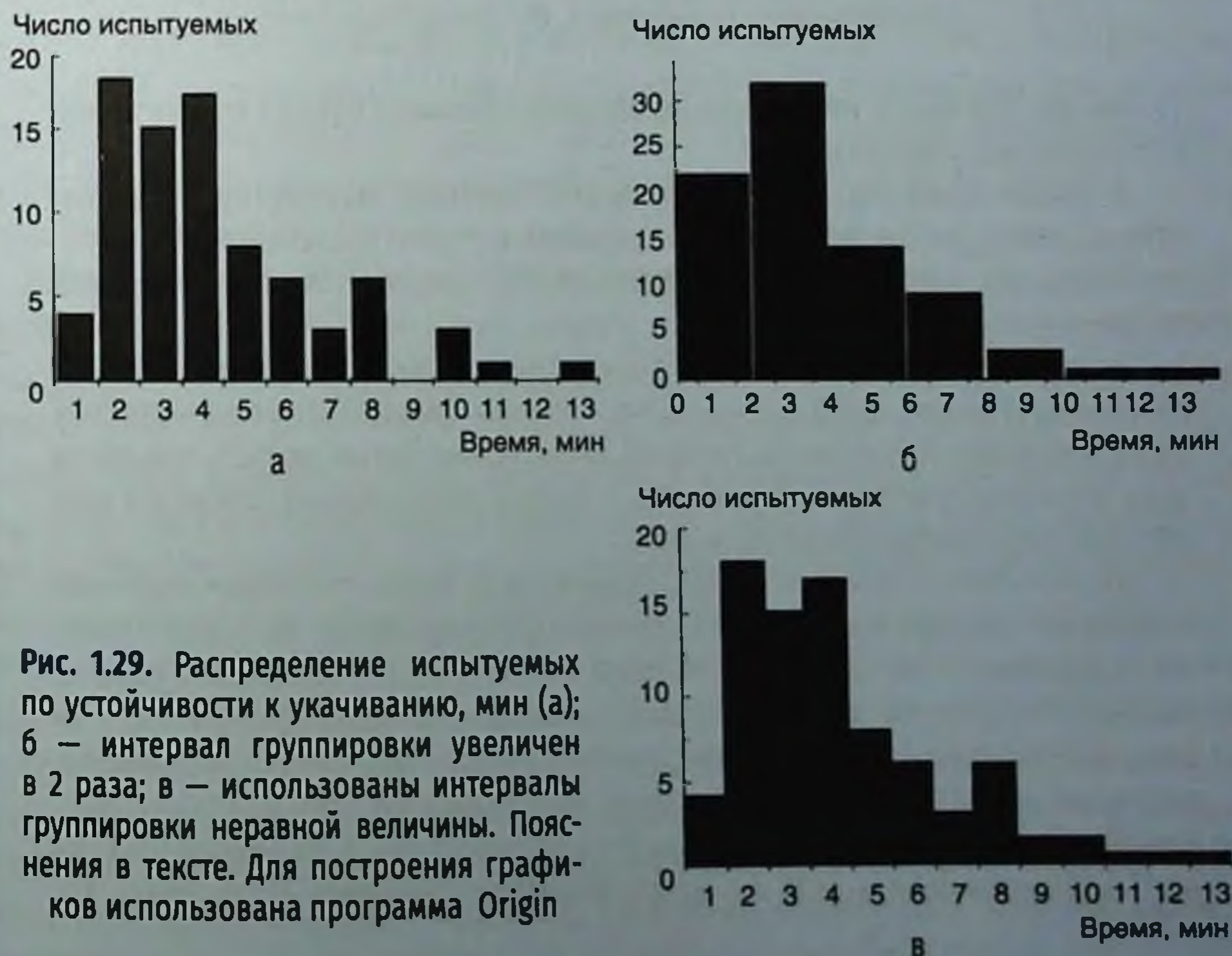


Рис. 1.29. Распределение испытуемых по устойчивости к укачиванию, мин (а); б — интервал группировки увеличен в 2 раза; в — использованы интервалы группировки неравной величины. Пояснения в тексте. Для построения графиков использована программа Origin



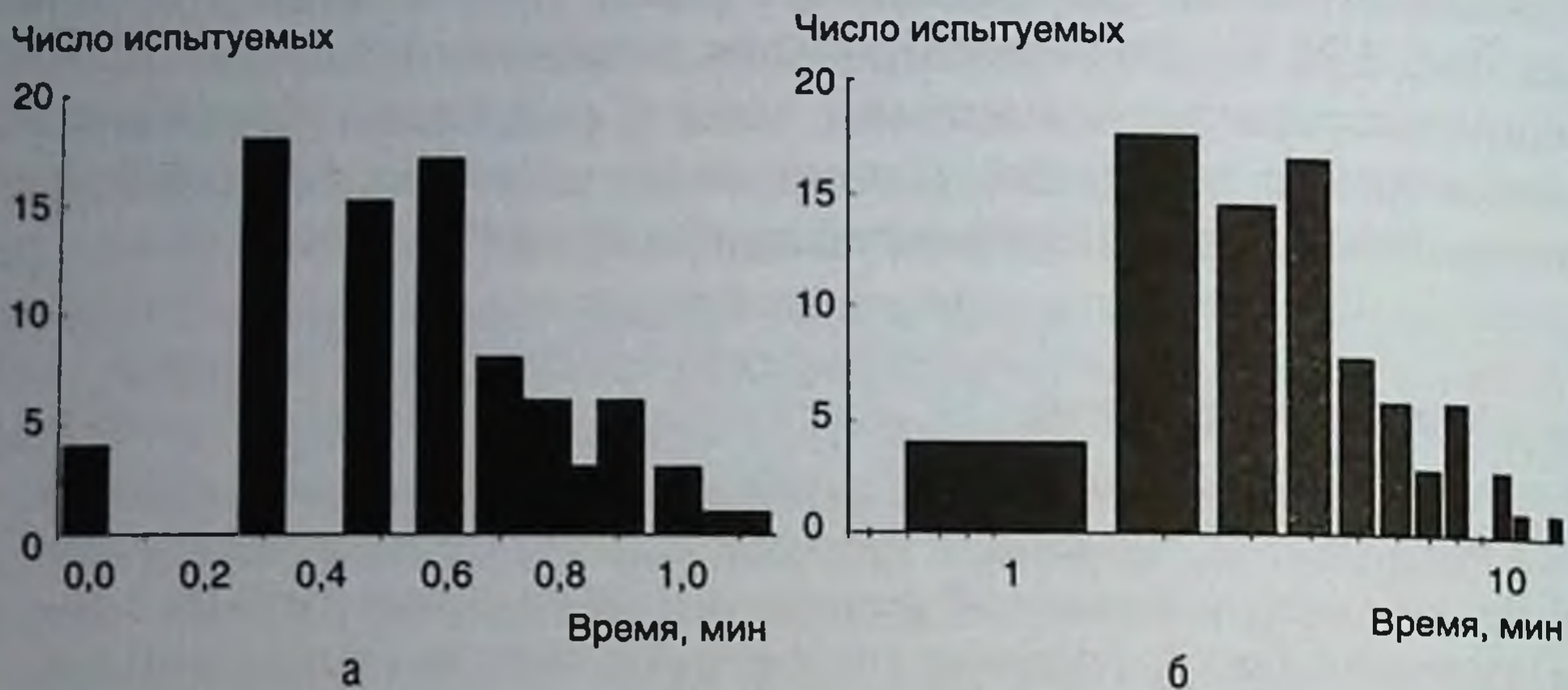


Рис. 1.30. Распределение испытуемых по устойчивости к укачиванию, по оси абсцисс — десятичный логарифм времени качания, мин: а — график, построенный с равной шириной столбиков; б — со столбиками, заполняющими диапазон величин, к которым относятся соответствующие данные. Так, слева (а) 4 испытуемых, укачавшихся в течение 1 мин, отнесены к «0» ( $\lg 1 = 0$ ), а справа (б) образуют столбик с основанием около «1». Оба варианта представления правильны, но второй создает иллюзию большого числа испытуемых, укачавшихся за 1 мин

На рис. 1.30 видно, что с помощью логарифмирования удалось получить симметричное распределение.

В случае когда исследователь получает подобное нашему примеру «неравномерное» распределение, на гистограмме которого существуют «провалы» и в некоторых диапазонах нет ни одного случая, возникает желание укрупнить диапазоны группировки.

В нашем примере удвоение диапазонов группировки приводит к результату, отраженному на рис. 1.29, б. Очевидно, что при этом происходит потеря некоторой части информации. Например, по такой гистограмме нельзя сказать, как много испытуемых укачивается на 1-й минуте, — трудно предположить, что их в 4 раза меньше, чем укачавшихся на 2-й минуте.

Для того чтобы избежать такой потери информации, используют неравные диапазоны группировки: узкие — в той части распределения, где случаев больше, и широкие — там, где случаев мало. В нашем примере желательно использовать более широкие интервалы при большой длительности качания. Было бы ошибкой «расширять» имеющийся столбик вправо или влево на диапазон, в котором нет частот. Если на 9-й минуте не укачался ни один из испытуемых, а на 10-й — три, то высота объединенного столбика должна быть равной 1,5. Если на 11-й укачался один, а на 12-й — никто, то высота объединенного столбика 0,5.



Такой подход оставляет на наше усмотрение, как поступать с 13-й минуты. Мы можем создать новый интервал (13 и 14 мин), высота столбика которого составит 0,5. Но можем и создать более широкий интервал для 11–13 мин, как это сделано на рис. 1.29, в.

Конечно, эти решения дают разные результаты, но эта неопределенность — типичное следствие малого числа наблюдений. В этом диапазоне величин мы действительно не можем точно сказать, какова частота события с большей точностью. Статистические и графические программы обычно не предоставляют инструмента для построения гистограмм с неравными диапазонами группировки. Это можно сделать, однако, создавая равные столбики для каждого интервала и вычисляя среднее число испытуемых для объединяемых интервалов.

*Мода* соответствует середине интервала, в котором находится больше всего результатов измерения. Лучше всего видно моду на гистограмме. Гистограммой называют столбиковый график, в котором каждой величине количественной переменной соответствует столбик, по высоте равный числу наблюдавшихся значений. Собственно, в табл. 1.6 мы уже видим прообраз такого графика — чем больше наблюдалось значений каждой величины, тем длиннее серый столбик. Гистограмму правильнее и удобнее строить компьютерными средствами (рис. 1.31).

При малом числе измерений гистограмма обычно имеет «нерегулярный» характер. Поэтому мода может оказаться различной в зависимости от того, какую ширину имеют интервалы группировки результатов измерений. На рис. 1.31 при показанной группировке по исходным результатам измерений мода равняется 110. Если расширить интервал группировки в 2 раза, то из 8 интервалов образуется 4 новых, и мода будет соответствовать середине между вторым и третьим (117,5). К аналогичному результату (нерегулярности, многовершинности гистограммы) может приводить особый вид систематической ошибки, имеющий место при измерениях — «смещение к круглым величинам». Например, при измерении артериального давления врачи чаще записывают результат в виде цифры, оканчивающейся на 5 и 0, и цифры целых десятков получают чаще, чем оканчивающиеся на 5. Таким образом, мода является параметром распределения, который легко получить, но оценка моды подвержена ошибкам.



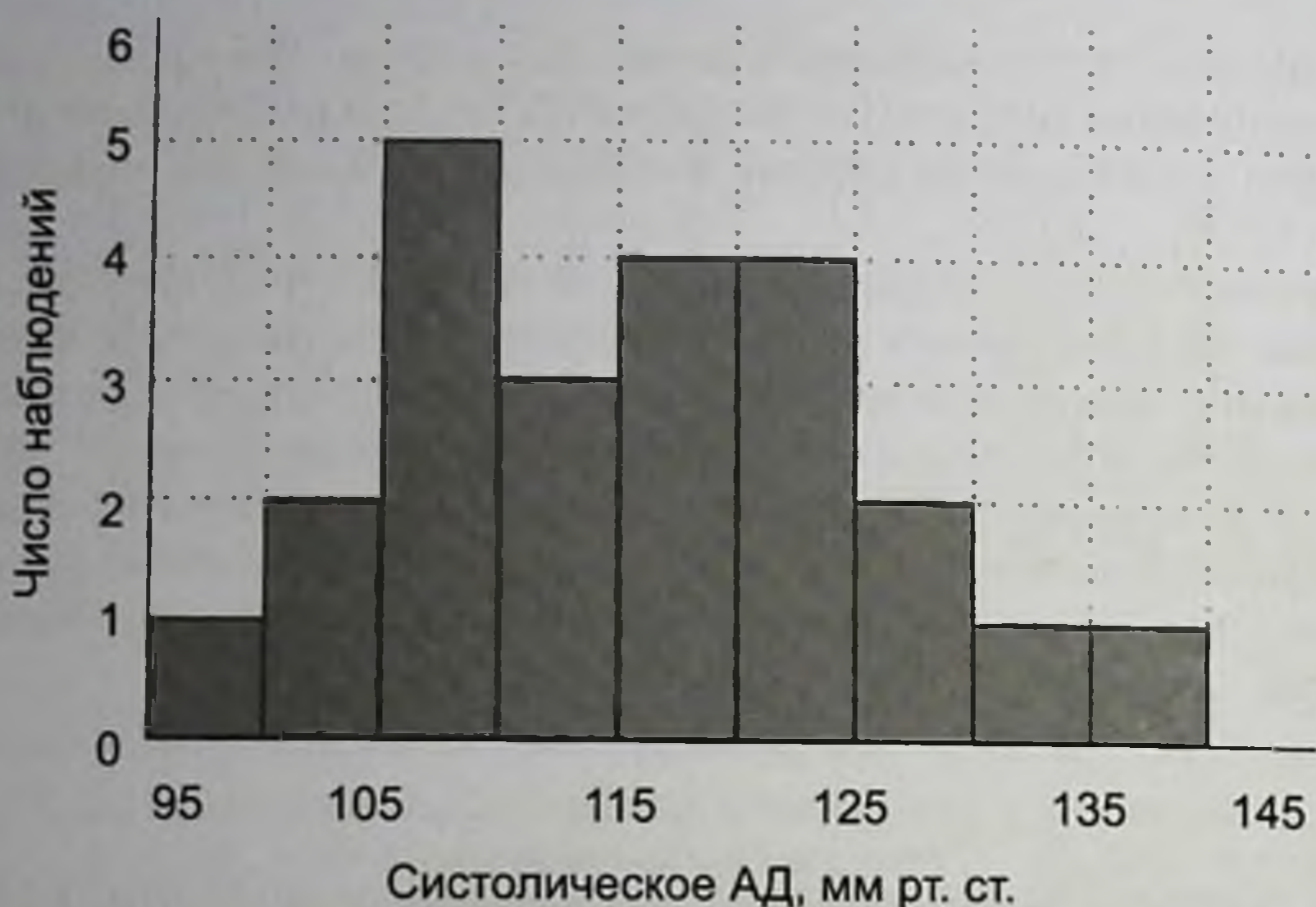


Рис. 1.31. Пример гистограммы с двумя вершинами: результаты измерения артериального давления в малой группе (см. табл. 1.6). График создан в пакете программ Statistica

*Медиана* вычисляется как значение результата измерения, находящегося в середине вариационного ряда, а в случае четного числа измерений — как середина между двумя результатами середины ряда. Сортировка результатов измерения вручную в нашем примере не сложна, но в большой выборке только использование компьютера позволяет выполнить задачу. Например, в Excel медиану получим для нашего примера, введя команду «=median(a1:a23)» (в русской версии «=медиана(a1:a23)»). Медиана нагляднее, если рассматривать не дифференциальную гистограмму, как на рис. 1.31, а кумулятивную, на которой высота столбика соответствует числу случаев с меньшей и данной величиной признака (рис. 1.32).

Как мода, так и медиана оцениваются точнее при большом числе измерений и при высокой точности измерения (большом числе градаций признака). В случае, подобном рассмотренному нами (см. рис. 1.32), когда имеется только 9 градаций признака, точность оценки низка. Для того чтобы избежать сокращения числа градаций, при измерении применяются специальные приемы. Например, при измерении артериального давления используют «слепой» метод: врач выслушивает тоны, не видя шкалы манометра, и делает повторные замеры. Результатом



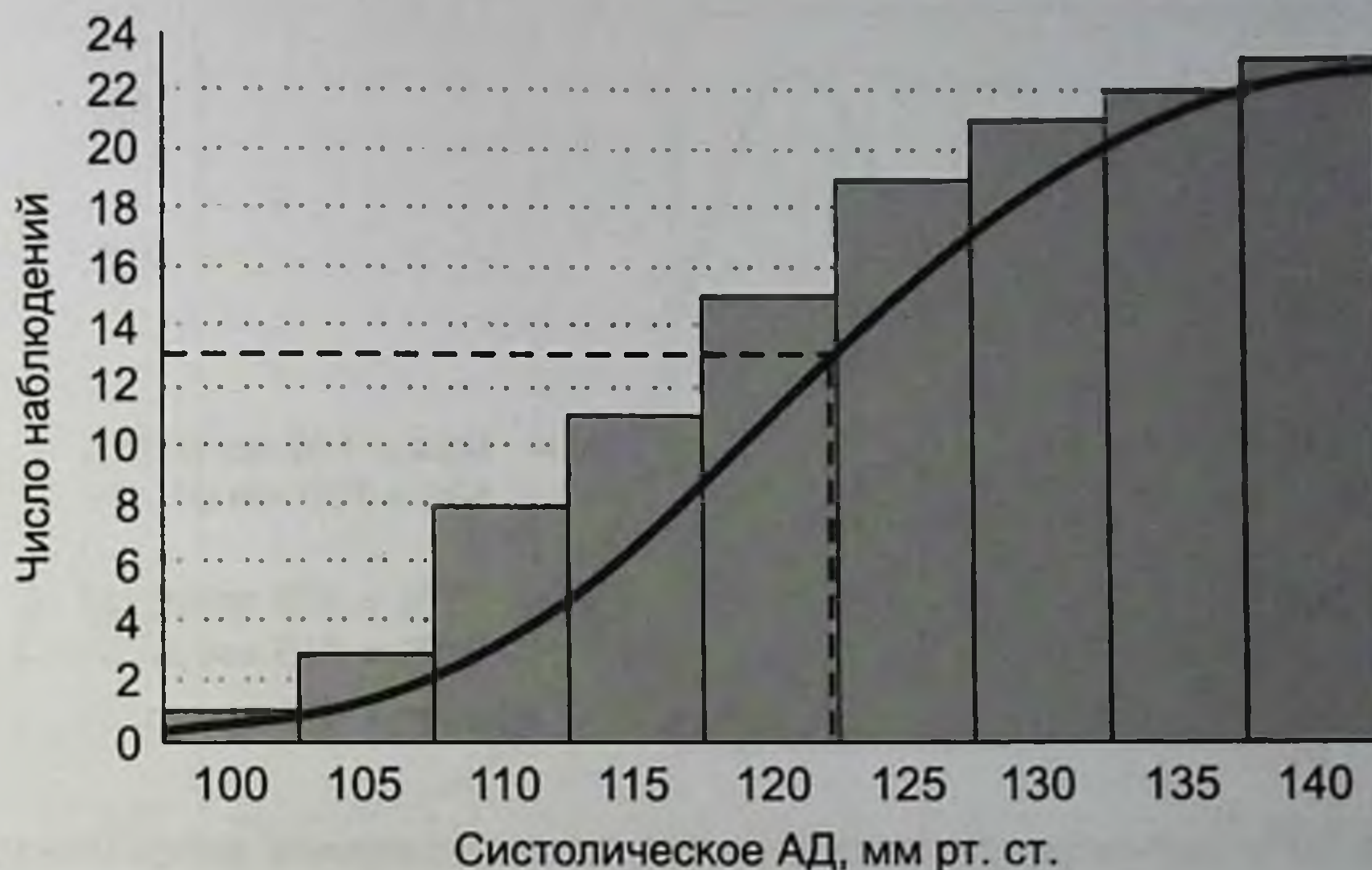


Рис. 1.32. Кумулятивное распределение результатов измерения систолического артериального давления в группе пациентов. Кривая — аппроксимирующая<sup>1</sup> экспериментальное распределение кривая нормального распределения. График создан в пакете Statistica

измерения считается средняя из выполненных замеров с точностью до 1 мм. Исчезает при этом и «смещение к круглым величинам».

### 1.6.2. Оценка вариабельности

Вариабельность представляет главный интерес, поскольку в различиях между пациентами врач, например, ищет объяснения особенностям симптоматики, а в различиях территорий эпидемиолог, например, ищет объяснения различиям в преваленсе кариеса. Применительно к любым распределениям для описания вариабельности можно воспользоваться таким же методом, какой применялся для вычисления медианы, т.е. основанным на ранжировании результатов измерения. За показатель вариабельности принимается интервал, в котором находятся половина

<sup>1</sup> Аппроксимацией называют приблизительное, упрощенное описание. В данном случае плавная кривая не отражает всех особенностей фактического распределения, а лишь прилагает закономерности нормального распределения к данной выборке в соответствии с ее средней арифметической и стандартным отклонением.



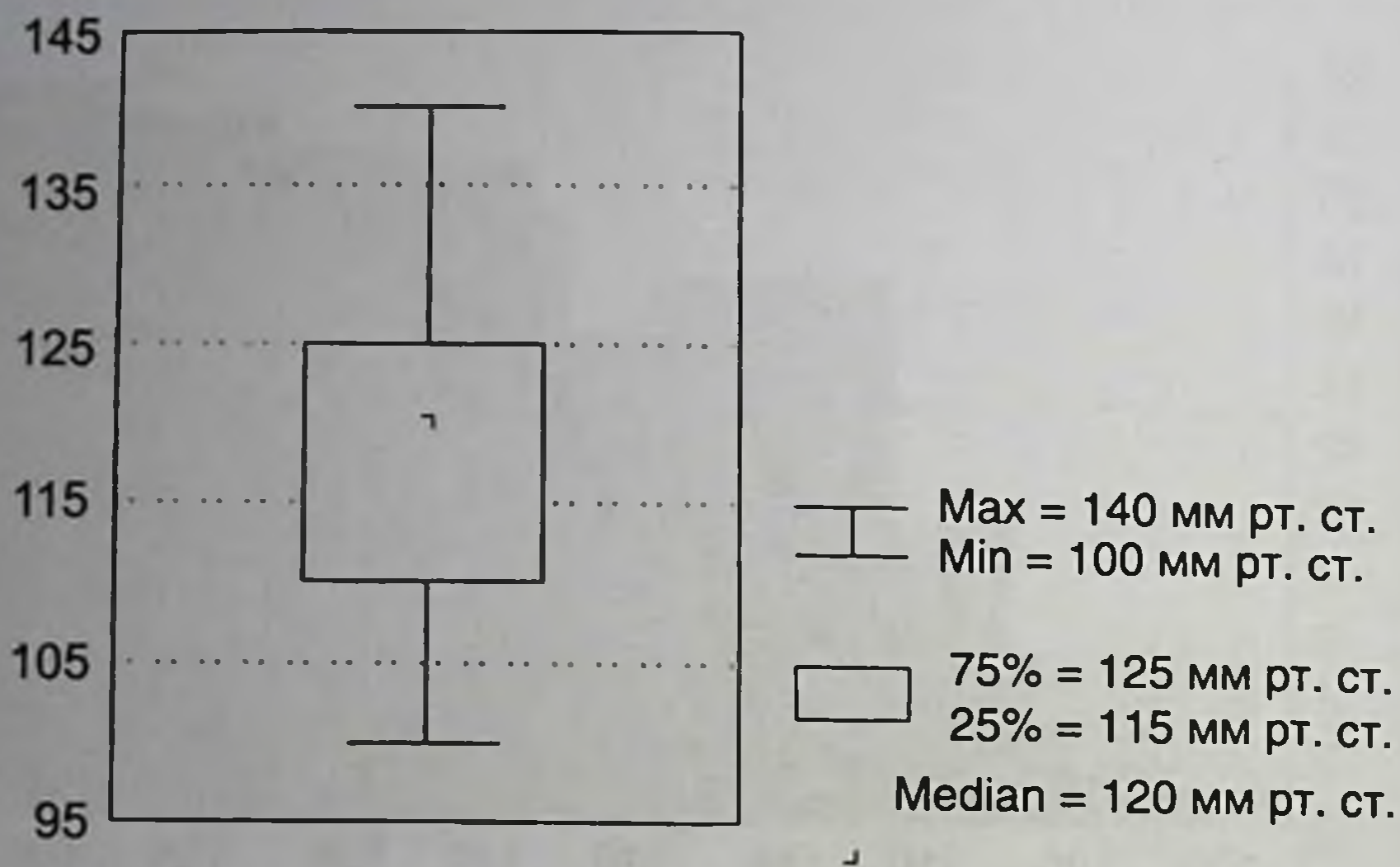


Рис. 1.33. Параметры распределения результатов измерения артериального давления в малой группе пациентов (ящичная диаграмма / box plot)

средних результатов. Его вычисляют как интервал между 25-м и 75-м из 100 результатов (см. рис. 1.32, 1.33). Конечно, число результатов в исследовании обычно не равно 100. Поэтому 25-й процентиль вычисляется как  $n \times 25 / 100$ . Например, для 82 результатов 25-й процентиль будет соответствовать 20-му результату в ранжированном ряду ( $82 \times 25 / 100$ ). Части распределения размером в 25% являются четвертями. Поэтому такие границы называют квартильными (квартелями). Графически представляя результаты описания вариационного ряда с помощью медианы и квартилей, обычно указывают также минимальные и максимальные значения. Следует помнить, что эти значения *могут быть* «выскакивающими» по неизвестным (случайным) причинам, т.е. не отражать закономерностей распределения изучаемой величины. В любом случае чем больше пациентов обследовано, тем больше разность между максимальным и минимальным значениями. Поэтому «размах» вариаций нельзя использовать для описания вариабельности результатов.

### 1.6.3. Нормальное распределение

Симметричное распределение чаще всего, но не всегда, близко к *нормальному* (распределению Гаусса) (рис. 1.34). Двумя параметрами нормального распределения являются средняя величина и среднее



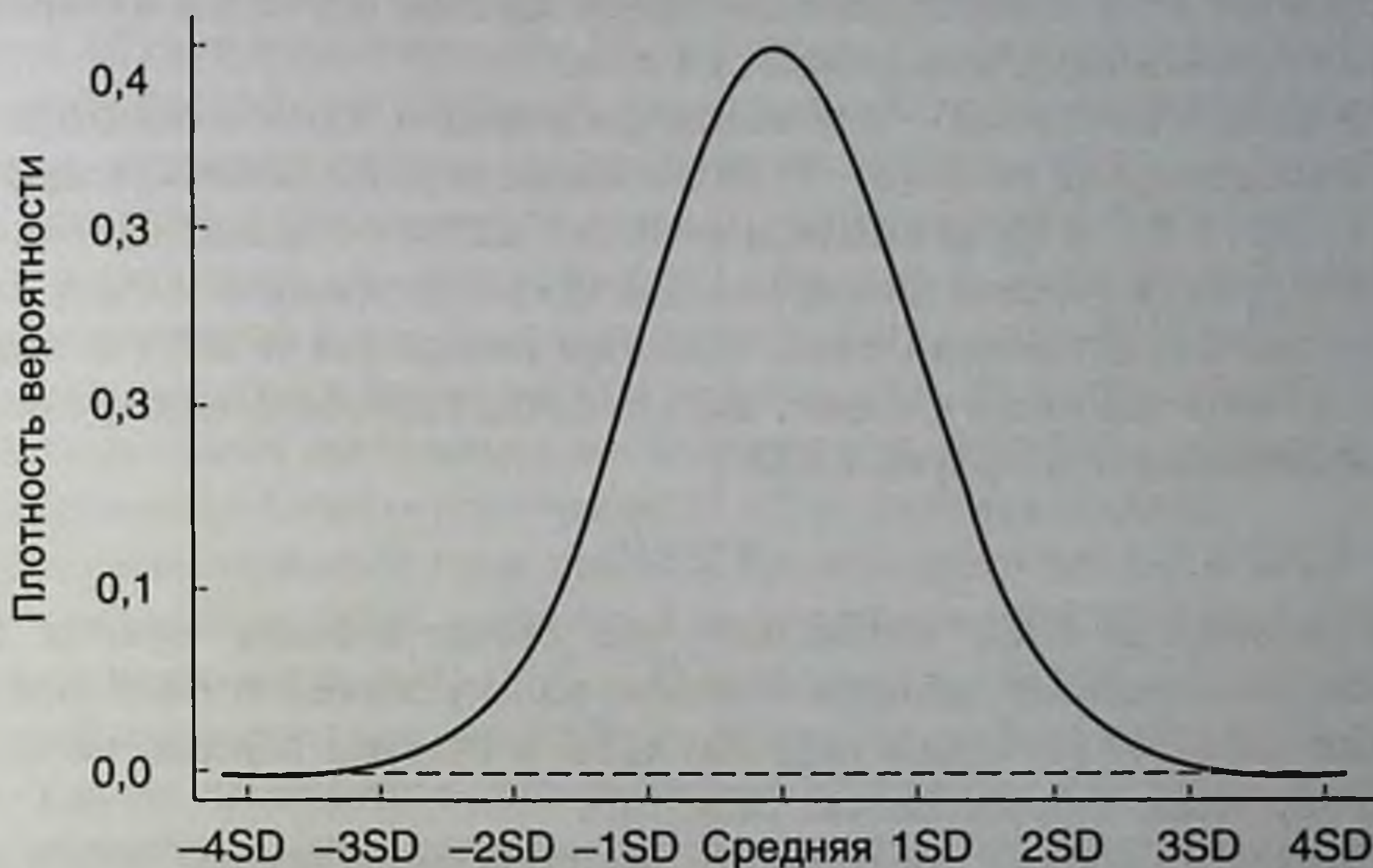


Рис. 1.34. Кривая нормального распределения и положение границ в стандартных отклонениях, используемых при ограничении референтных интервалов

квадратическое отклонение (СКО, называемое также «стандартное отклонение», SD; математики обозначают его популяционную величину  $\sigma$  — сигма греческая). СКО является мерой вариабельности результатов. Название «квадратическое» происходит из методики его вычисления.

Для получения оценки СКО сначала вычисляется величина отклонения каждого результата от средней и возводится в квадрат. Считается, что отклонения от центральной тенденции лучше измерять не в натуральных единицах, а возводить их в квадрат. Во-первых, потому что таким образом более удаленные от средней величины всегда более редкие, дают больший вклад в получаемую оценку вариабельности. Во-вторых, потому что если считать просто среднюю арифметическую отклонений от средней, то она при симметричном распределении будет равна нулю.

Затем вычисляется средняя этих квадратов. Она называется *дисперсией*:

$$D = \sum_{i=1}^n (x - \bar{X})^2 / (n-1).$$

В числителе в традиционной для математиков форме записана сумма ( $\Sigma$  — знак суммирования) отклонений каждой измеренной величины  $x$ ,



от средней  $\bar{X}$ . В такой записи для признака  $x$  мы изучаем  $n$  измерений, пронумерованных с номерами  $i$  от 1 до  $n$ .

Обратите внимание — в знаменателе вместо  $n$ , как это было при вычислении средней, стоит  $(n - 1)$ . Это — *число степеней свободы* (подробнее см. с. 195, 236). В Excel команда «=var(a1:a23)» («=дисп(a1:a23)» в русской версии) вычислит дисперсию для нашего примера (см. табл. 1.6).

Поскольку дисперсия имеет единицы измерения те же, что и средняя, но возведенные в квадрат, для описания вариабельности результатов измерения используют СКО:

$$\sigma = \sqrt{D}.$$

Среднюю и СКО вычисляют все статистические пакеты программ, электронные таблицы и многие калькуляторы. В Excel команда «=stdev(a1:a23)» («=стандотклон(a1:a23)» в русской версии) вычислит СКО для нашего примера (см. табл. 1.6).

Свойства (форма) нормального распределения таковы, что в пределах одного СКО от средней находится 68% результатов, в пределах от  $-2$ СКО до  $+2$ СКО — 95%, в пределах  $\bar{X} \pm 3$ СКО — 99% результатов. Эти границы постепенно стали почти общепринятыми в ряде статистических задач, например, при оценке результатов лабораторных исследований (см. разд. 2.2.7). Важно, что кривая распределения (см. рис. 1.34) отличается от гистограммы. На гистограмме мы можем указать, в каком числе случаев признак имеет искомое значение, или какая доля случаев имеет данное значение. В нашем примере измерения артериального давления у одного обследуемого было систолическое артериальное давление (АД) 100 мм рт.ст. При использовании распределения мы можем только сказать, какая доля случаев располагается *в заданном интервале*. Этот интервал необязательно должен быть симметричным относительно средней.

Теоретически кривая нормального распределения (см. рис. 1.34) не имеет пределов справа и слева. Практически при увеличении числа обследованных выявляются редкие величины, сильно отклоняющиеся от средней. Соответственно увеличивается диапазон между минимальным и максимальным значениями. Вместе с тем поскольку большая часть значений находится вблизи средней, эти редкие величины мало влияют на среднюю и СКО, если, конечно, не отклоняются от средней слишком далеко, т.е. если укладываются в закономерности нормального распределения.



Если на графике горизонтальная шкала расширяется (например, вместо 10 см график строится 20 см шириной), то кривая распределения становится более широкой, но ее пропорции не изменяются. Плоские и островершинные распределения имеют другие пропорции (см. рис. 1.28). Отклонение по степени островершинности от нормального распределения (англ. *kurtosis*) (см. рис. 1.28, б) и отклонение от симметричной формы — скошенность (англ. *skewness*) (см. рис. 1.28, в, г) могут быть описаны математически. Эти функции реализованы в статистических программах и в электронных таблицах, например для скошенности в Excel — это «=скос(a1:a23)» (русская версия).

Для рассмотренного нами примера средняя величина 118 и СКО = 10. Результаты вычислений параметров нормального распределения следует представлять в виде: «118; 10( $\bar{X}$ ; СКО) мм рт.ст.». На рис. 1.35 над гистограммой приведены средняя и СКО с точностью до 6-го знака после запятой. Так их представила статистическая программа, где такая точность представления данных заложена «по умолчанию», т.е. на все случаи, когда пользователь сам не изменит настройки программы. Использовать в отчете (статье) параметры распределения с такой точностью неправильно. Вычисления средней допустимы с одним дополнительным знаком

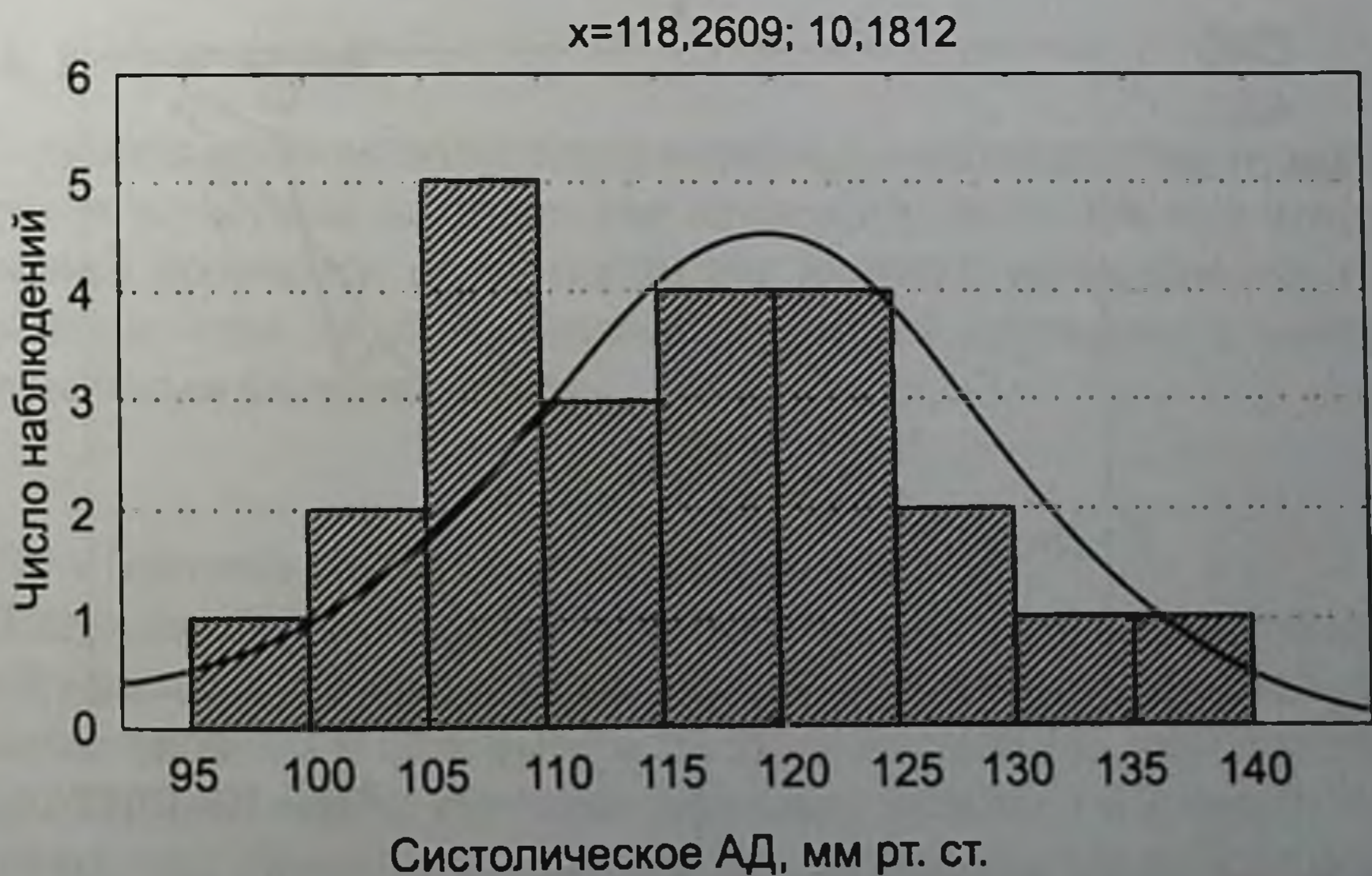
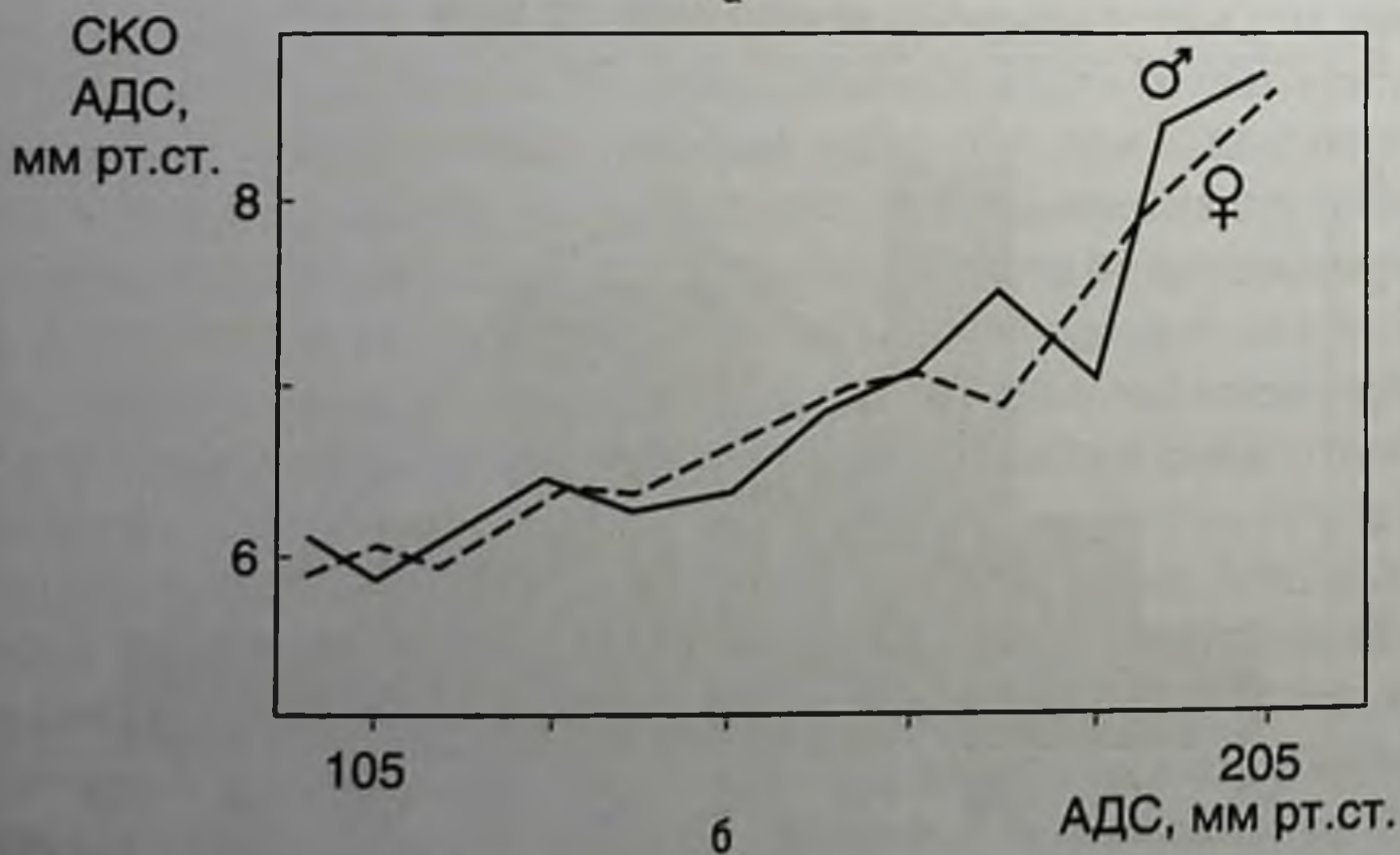
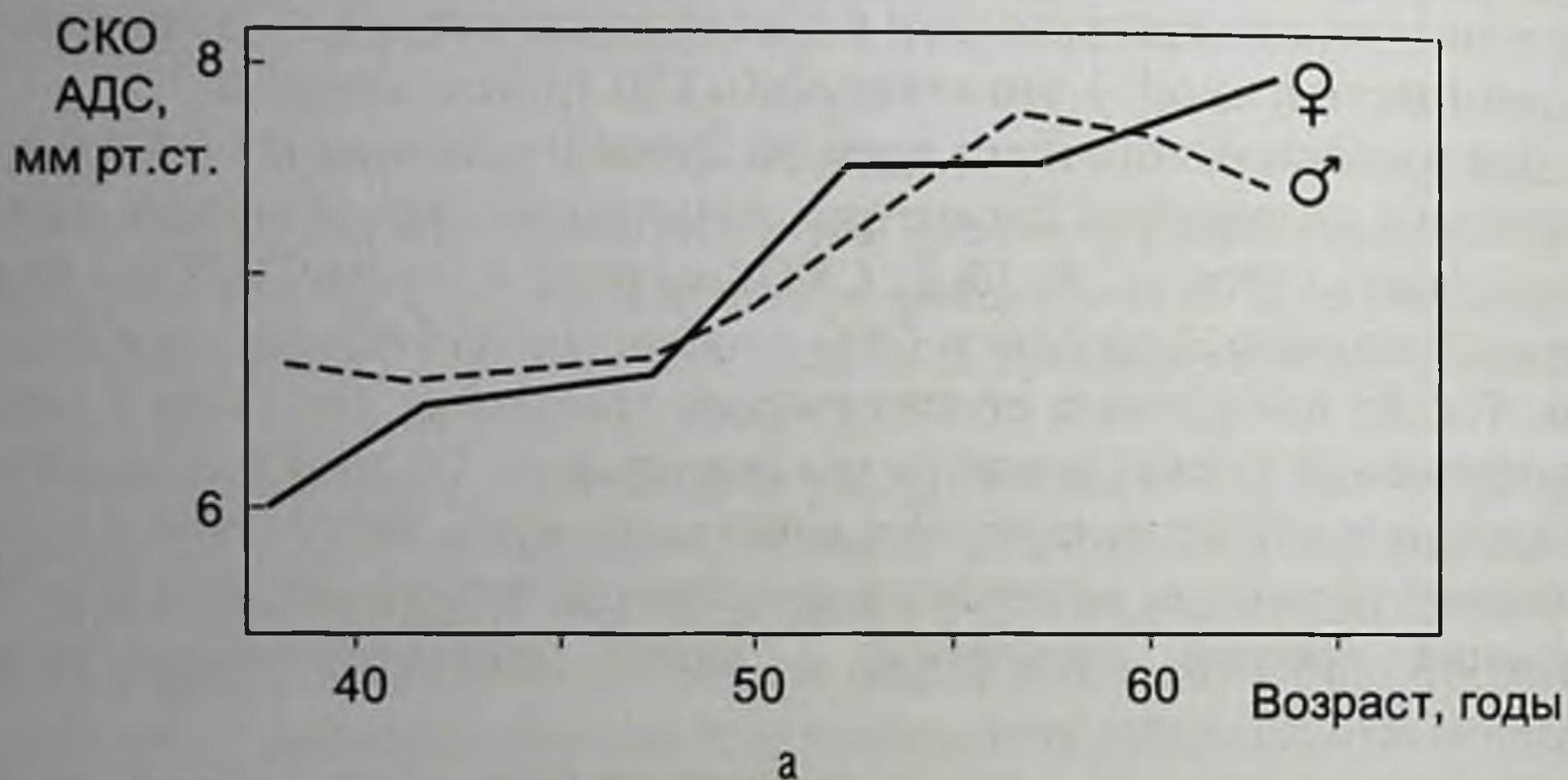


Рис. 1.35. Результаты измерения артериального давления. Гистограмма и аппроксимирующая ее кривая нормального распределения. Над графиком — средняя ( $\bar{x}$ ) и среднее квадратическое отклонение. Получено в программе Statistica



по сравнению с точностью исходных измерений. Поскольку наши измерения были сделаны с точностью до 5 мм рт.ст., полученные нами параметры бессмысленно описывать точнее, чем до 1 мм рт.ст., и их надо округлить.

В некоторых случаях может возникнуть необходимость сравнить вариабельность результатов измерения двух признаков, например вариабельность концентрации лейкоцитов в крови и вариабельность концентрации натрия в крови. Поскольку средние и СКО для этих



**Рис. 1.36.** Зависимость вариабельности систолического артериального давления (СКО АДС) от возраста людей (а) и величины систолического артериального давления (АДС) (б) по результатам повторных измерений во Фремингемском проспективном исследовании (по данным Kannel W.B., Gordon T. In Epidemiology of aging. US Dep. HHS, 1980. P. 65–90)



признаков различаются очень сильно, непосредственное их сравнение невозможно. Для такого сравнения вычисляют коэффициент вариации:

$$V = \text{СКО} / \bar{X}.$$

Коэффициент вариации — безразмерная величина, меньшая единицы, и умножается на 100, если желательно манипулировать размером вариабельности в процентах. Для иных целей использовать коэффициент вариации, методически нестрогий, не следует.

Многие статистические методы базируются на предположении о том, что стандартное отклонение не зависит от средней величины. Это упрощает сравнение групп пациентов с различными средними величинами. В действительности правилом является прямая зависимость между средней и СКО (рис. 1.36). Обычно зависимость между средней и СКО не приводит к большим ошибкам при применении статистических методов, если выборки велики, а распределения симметричны. Избавиться от потенциально опасной асимметрии и зависимости между средней и СКО может помочь *нормализация* (см. с. 111). В тех случаях, когда предположение о нормальном распределении данных не оправдывается, должны применяться *непараметрические методы* анализа данных.

#### 1.6.4. F-критерий

Вариабельность величин в двух выборках можно сравнить с помощью F-критерия — теста на равенство дисперсий. Этот простой критерий позволяет проверить, являются ли две выборки выборками из одной популяции, судя по их дисперсии. Критерий применим к выборкам с нормальным распределением.

$$F = D_1 / D_2,$$

где  $D_1$  — большая дисперсия из двух сравниваемых.

Полученный результат мы сравниваем с таблицей критических значений критерия (см. приложение 2), где  $N_1$  и  $N_2$  соответственно — число степеней свободы в двух выборках, вычисляемое как число значений минус единица. Если величина критерия больше табличной, то это означает, что гипотеза о том, что две сравниваемые выборки взяты из одной популяции, может быть отклонена. Значит, дисперсии различаются, и мы не можем использовать статистические критерии, требующие равенства дисперсии, например *t*-критерий.



### 1.6.5. Точность оценки средней

Получаемые в ходе вычислений оценки параметров (мода, медиана, средняя, СКО) относятся к выборке — в нашем примере случайной выборке из школьников-спортсменов. Исследователь изучает выборку, чтобы получить знание о генеральной совокупности. Соответственно, получаемые величины параметров являются *оценками*. Чем больше размер репрезентативной выборки, тем точнее оценка популяционной средней. Точность оценки средней определяется ошибкой средней (стандартной ошибкой), обозначаемой  $m$ :

$$M = \sigma / \sqrt{n}.$$

Стандартная ошибка является не мерой выборочной вариабельности (для этого служит СКО), а мерой точности оценки популяционной средней. Поэтому приводить величину оценки средней точнее, чем это позволяет размер ошибки, не следует: если средняя 5,3456, а стандартная ошибка 0,2, то среднюю надо округлять до 5,3.

Для получения доверительного интервала (ДИ, интервала, в котором с заданной вероятностью находится популяционная средняя) стандартная ошибка откладывается от выборочной средней вправо и влево. Обычно вычисляется ДИ, в котором с вероятностью 95% находится популяционная средняя. Для 95% ДИ приблизительно можно откладывать  $2m$  вправо и влево при числе измерений в выборке более 20. При численности выборки менее 20 коэффициент быстро растет выше 2. Для точного вычисления ДИ надо  $m$  умножить на  $(t_{1-\alpha/2})$ , где  $\alpha$  — принятый уровень альфа-риска. Альфа-риск и статистический  $t$ -критерий обсуждаются соответственно в разд. 1.8 и 2.5.2.

Границы ДИ нетрудно вычислить, используя табличные значения  $t$  для данного уровня альфа-риска и численности выборки. Статистические программы позволяют быстро вычислить ДИ. Для нашего примера (см. табл. 1.6) мы можем это сделать в Excel, выполнив команду: «=confidence(0,05;10;23)» («=доверит(0,05;10;23)» в русской версии). В результате получим 4,1. Эту величину надо вычесть из средней для получения нижней границы ДИ и прибавить к средней для получения верхней границы ДИ. В генеральной совокупности средняя лежит в интервале от 112 до 122 с вероятностью 95%. На рис. 1.37 представлена ширина ДИ для средней величины систолического АД в зависимости от численности исследованной выборки и принятого уровня



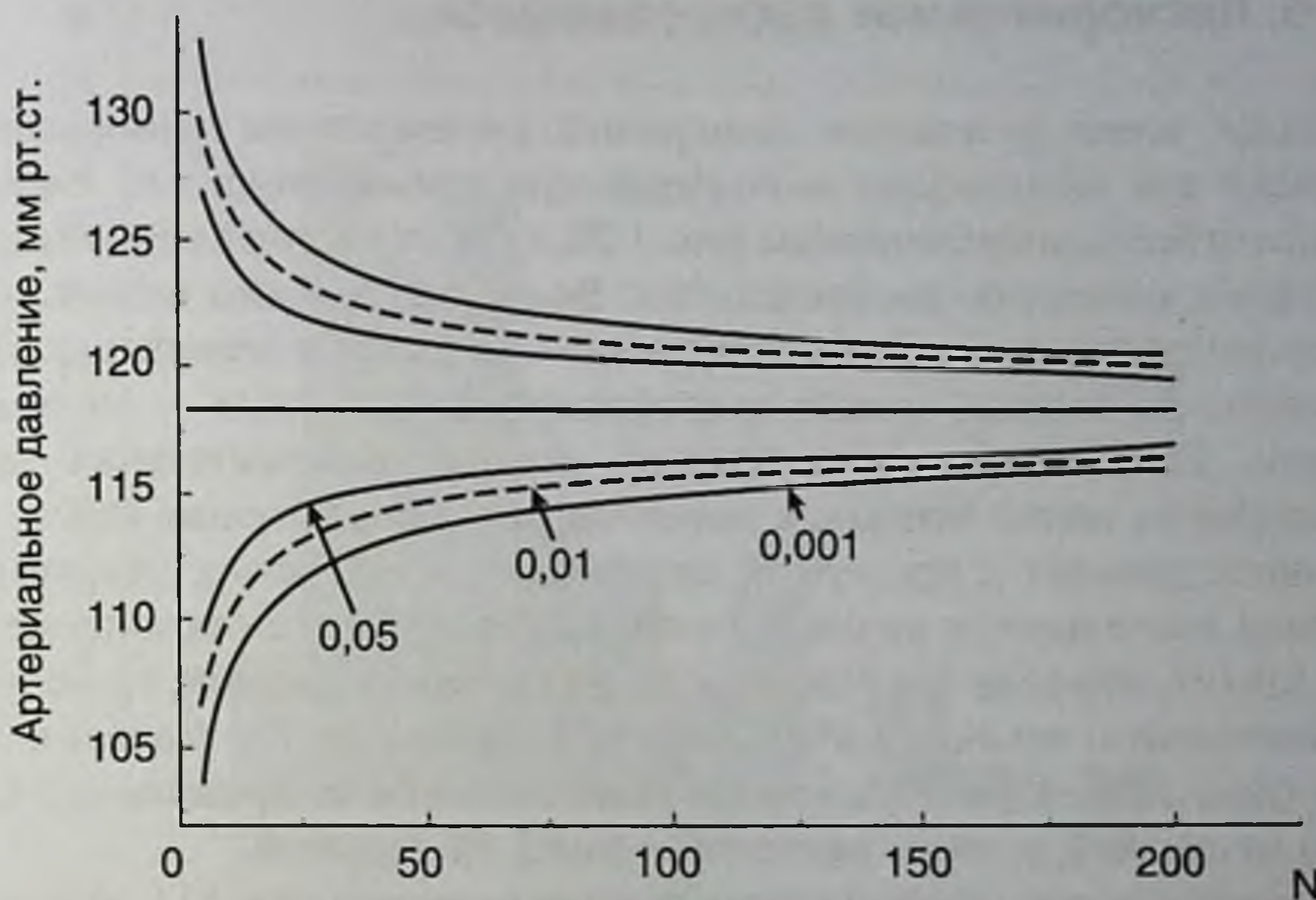


Рис. 1.37. Ширина доверительного интервала средней величины систолического артериального давления в зависимости от величины выборки (N). Три уровня альфа-риска приведены у кривых

статистической значимости. Легко видеть, что при увеличении выборки точность оценки средней увеличивается при увеличении выборки до 50–100 человек. Далее сужение ДИ происходит медленнее. Стандартное отклонение влияет на ширину ДИ независимо от численности выборки. Поэтому если есть возможность для исследования отобрать однородную группу пациентов (с меньшей межиндивидуальной вариабельностью), то оценка средней в ней будет точнее, будет легче обнаружить отличия от других выборок и легче обнаружить изменения в состоянии пациентов данной группы.

Помимо нормального распределения в статистическом анализе используется ряд других *теоретических распределений*, параметры которых позволяют аппроксимировать эмпирические распределения, т.е. распределения результатов, получаемых в исследованиях. Поскольку свойства теоретических распределений хорошо известны, в случаях, когда эмпирическое распределение соответствует какому-то теоретическому, появляется возможность использовать эффективные средства анализа, разработанные для этого теоретического распределения.



### 1.6.6. Логнормальное распределение

Чаще всего результаты измерений распределены ненормально, а имеют так называемую положительную асимметрию, т.е. «хвосты» в области больших величин (см. рис. 1.28, в). В этом случае возможны два подхода к описанию распределения. Во-первых, можно использовать непараметрические способы описания — медиану и межквартильную разность. Во-вторых, можно трансформировать данные — «нормализовать». Если скошенность «сырых» величин положительная, часто логарифм их имеет близкое к нормальному распределение (см. с. 48). В нашем примере с проверкой людей на устойчивость к укачиванию средняя длительность качания равна 4,3, стандартное отклонение 2,5. Это значит, что если мы отложим 2 СКО влево от средней, то получим отрицательную величину длительности качания, чего не бывает в природе (рис. 1.38). При нормальном распределении за пределами 2 СКО слева от средней должны находиться еще 2,5% величин.

Как мы убедились, анализируя пример во врезке на с. 111, логарифм исходных величин имеет симметричное распределение. Применяя лог-

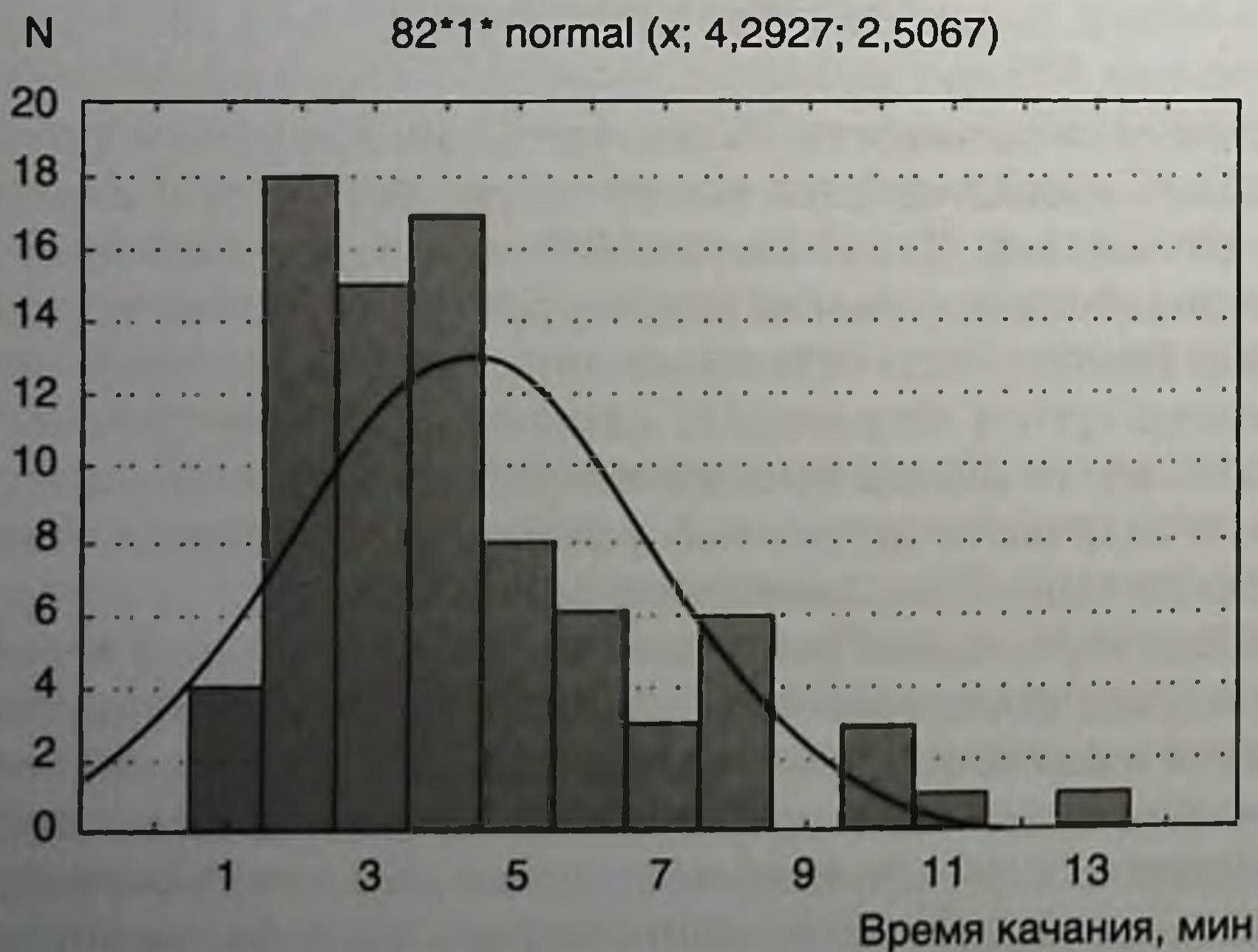


Рис. 1.38. Распределение числа испытуемых N по длительности качания и кривая нормального распределения. Пример из врезки на с. 111



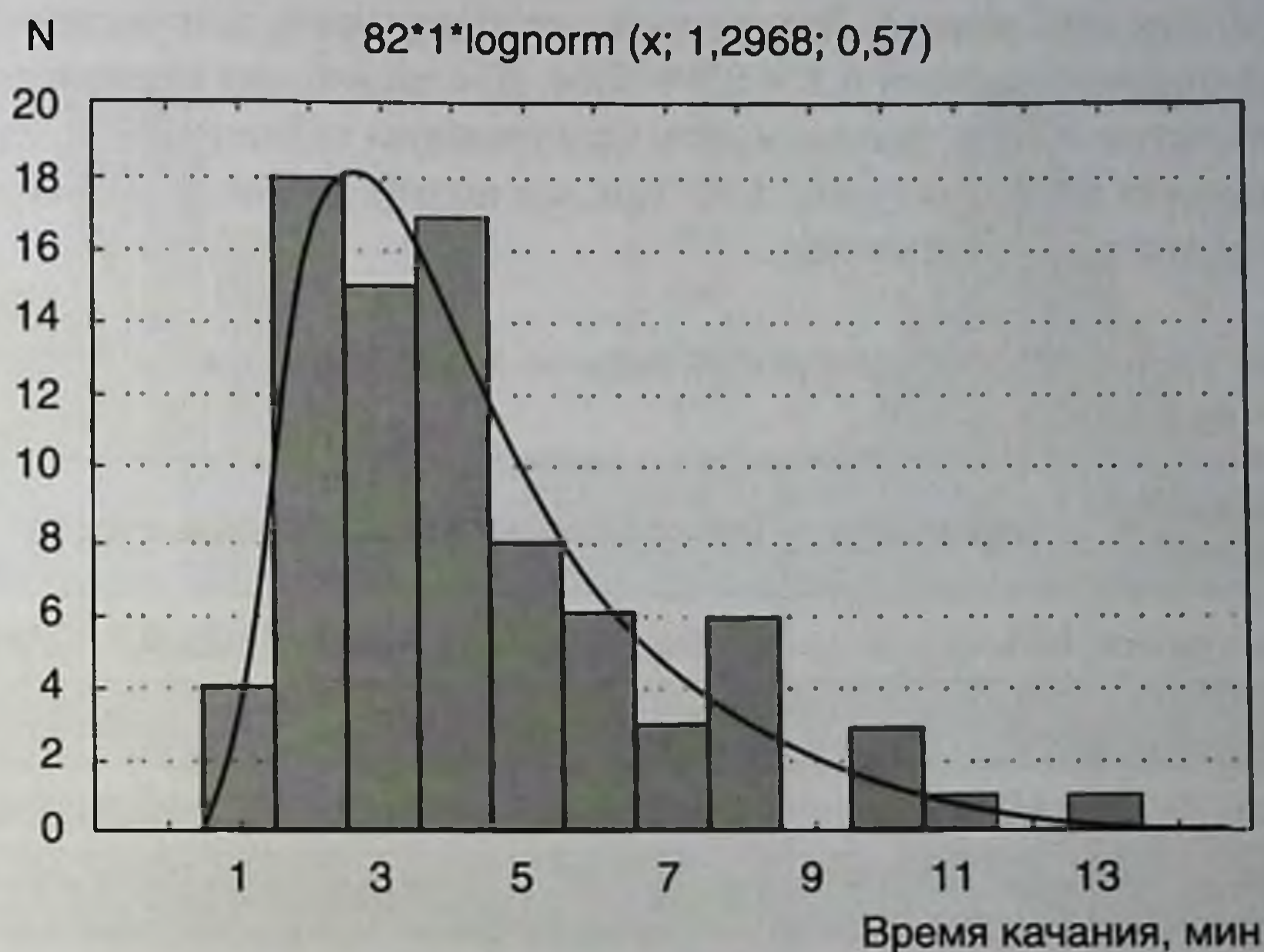


Рис. 1.39. Распределение числа испытуемых N по длительности качания и кривая логнормального распределения. Пример из врезки на с. 111

нормальное распределение, мы получаем хорошую аппроксимацию распределения исходных величин (рис. 1.39).

Если нормализация удалась, с логарифмами можно проводить необходимые вычисления, а потом при необходимости перейти от них к исходным величинам. Если недостаточно медианы и межквартильного расстояния для описания распределения исходных данных, то вычисляют среднюю и ее ДИ для логарифмов исходных значений, и затем границы ДИ и среднюю переводят в исходные величины.

### 1.6.7. Биномиальное распределение

Для дискретных данных простейшим вариантом распределения является распределение качественных данных. Например, если относительная частота зеленоглазых в популяции 0,2, то у случайно взятой пациентки вероятность обнаружить зеленые глаза тоже 0,2. Вероятность найти зеленые глаза у двух пациенток одновременно равняется  $0,2 \times 0,2 = 0,04$ . Вероятность встретить зеленые глаза у одной из двух пациенток вычисляется несколько сложнее:  $0,2 \times 0,8 = 0,16$  для каждой,



т.е. 0,32 для двух вместе. Вероятность не обнаружить зеленоглазой среди двух пациенток будет  $0,8 \times 0,8 = 0,64$ . Для получения перечисленных вероятностей можно использовать электронную таблицу Excel, где есть специальная функция (табл. 1.8). Как мы видим на рис. 1.40, распределение резко правостороннее.

Таблица 1.8. Вероятность обнаружения события с популяционным прева­ленсом 0,2 у пациенток А и Б

Комбинация событий	Вероятность	Способ расчета	Команда в Excel
А и Б	0,04	$0,2 \times 0,2$	=БИНОМРАСП(2;2;0,2;ЛОЖЬ)
не А Б	0,16	$0,2 \times 0,8$	=(БИНОМРАСП(1;2;0,2;ЛОЖЬ))/2
А не Б	0,16	$0,2 \times 0,8$	=(БИНОМРАСП(1;2;0,2;ЛОЖЬ))/2
не А не Б	0,64	$0,8 \times 0,8$	=БИНОМРАСП(0;2;0,2;ЛОЖЬ)
Всего	1,00	—	—

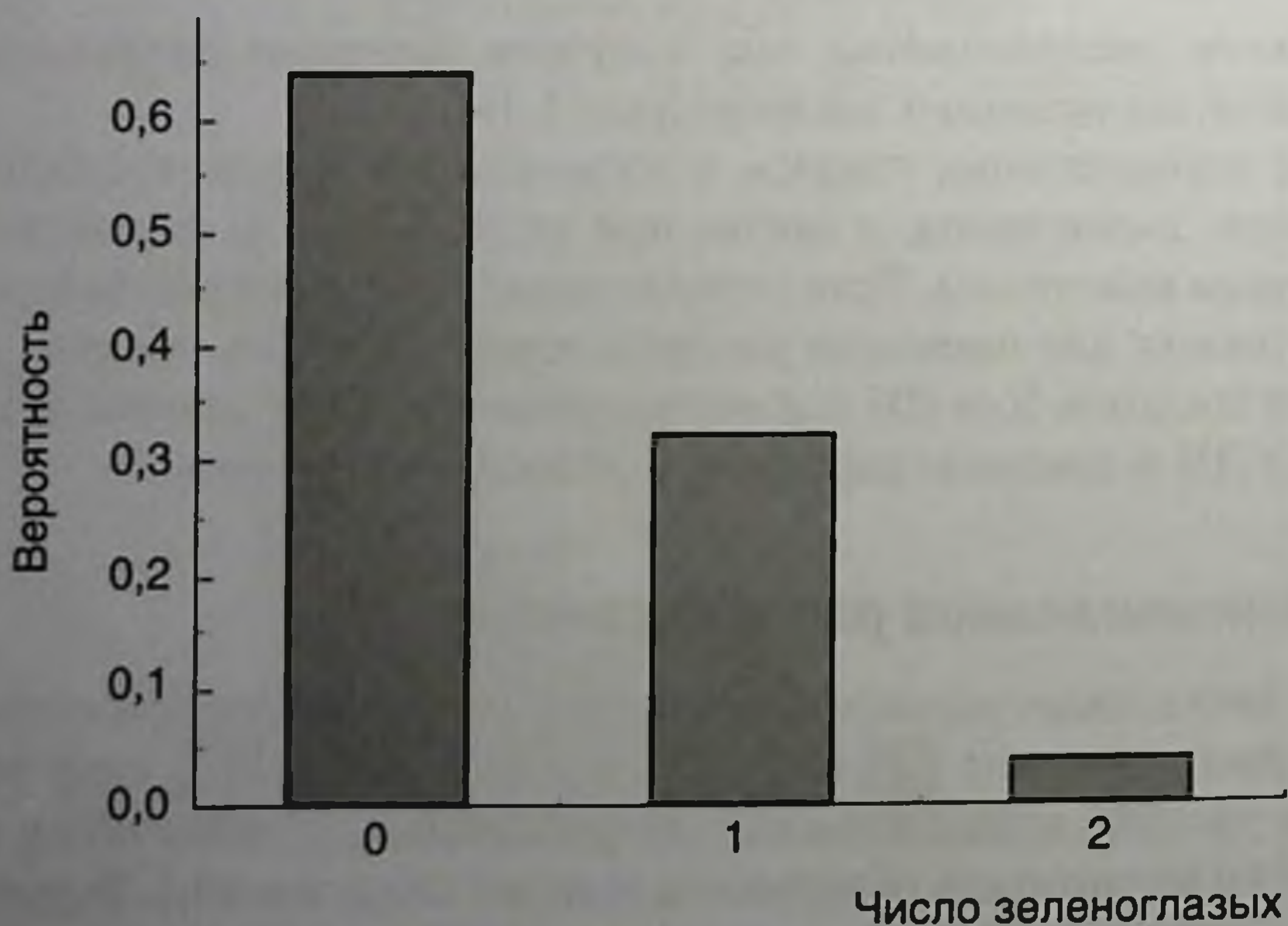


Рис. 1.40. Биномиальное распределение вероятностей обнаружения зеленоглазой из двух пациенток при популяционной вероятности 0,2



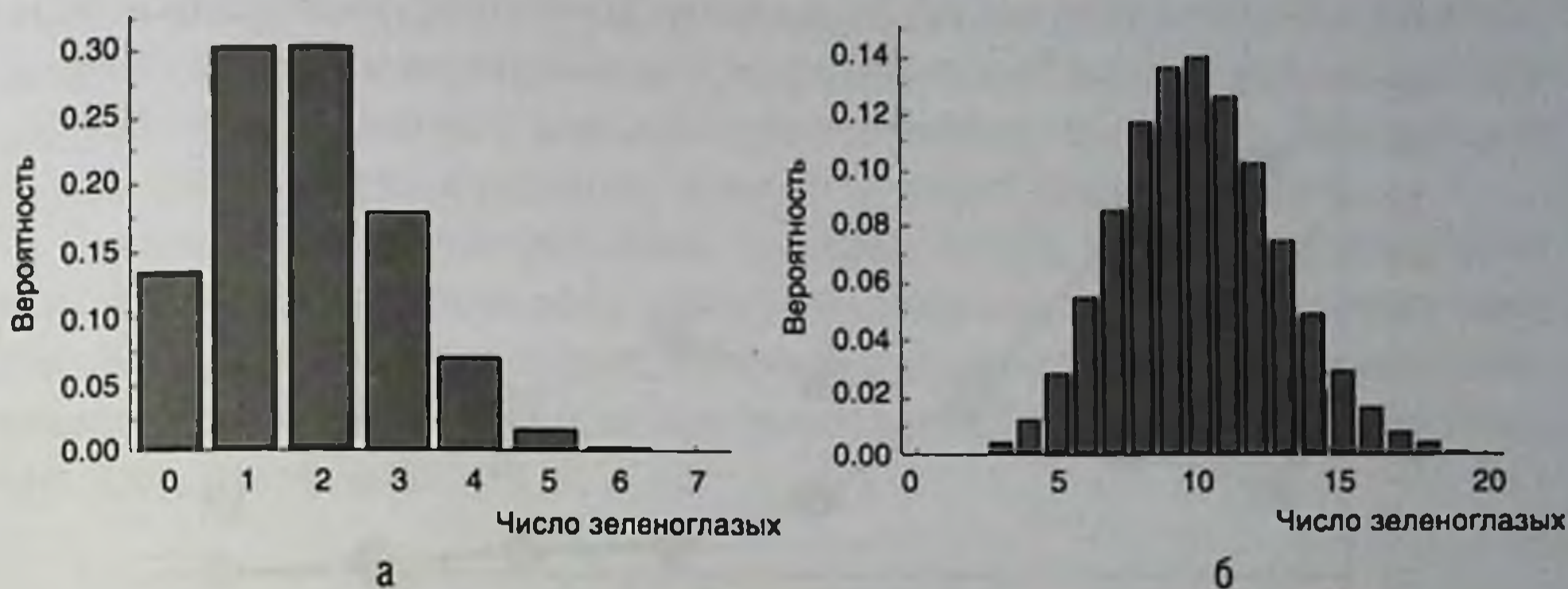


Рис. 1.41. Биномиальное распределение вероятностей обнаружения зеленоглазых из 9 (а) и из 50 пациенток (б)

Если число обследуемых субъектов увеличивается, то резко увеличивается число вариантов. Так, для 5 пациентов будет  $2 \times 2 \times 2 \times 2 \times 2$ , т.е. в общем виде  $2^n = 32$  комбинаций. К счастью, существуют обобщающие формулы и компьютерные вычисления. При увеличении размера выборки биномиальное распределение меняет форму и по форме приближается к нормальному. На рис. 1.41 представлены распределения вероятностей для выборок размером 9 и 50 пациентов при той же популяционной вероятности 0,2.

### 1.6.8. Распределение Пуассона

Другой вариант использования дискретных данных в практике — подсчет числа событий для разных людей или событий в единицу времени. Например, число пневмоний, выявленных за день в поликлинике, распределяется в соответствии с распределением Пуассона. Так же люди распределяются по числу заболеваний за год, числу дней нетрудоспособности и по числу диагнозов хронических болезней (рис. 1.42). Распределение Пуассона крайне асимметрично, когда среднее число событий мало. Например, если в среднем за год в поликлинике выявляется 2 случая пневмонии в сутки, то с помощью встроенной функции программы Excel мы легко получим вероятность любого числа событий ( $n$ ) для 1 дня поликлиники ( $=\text{ПУАССОН}(n;2;\text{ЛОЖЬ})$ ), где логическое условие «ложь» указывает на вычисление вероятности события именно  $n$ . Это распределение приведено на рис. 1.43. Когда средняя (среднее число диагнозов, среднее



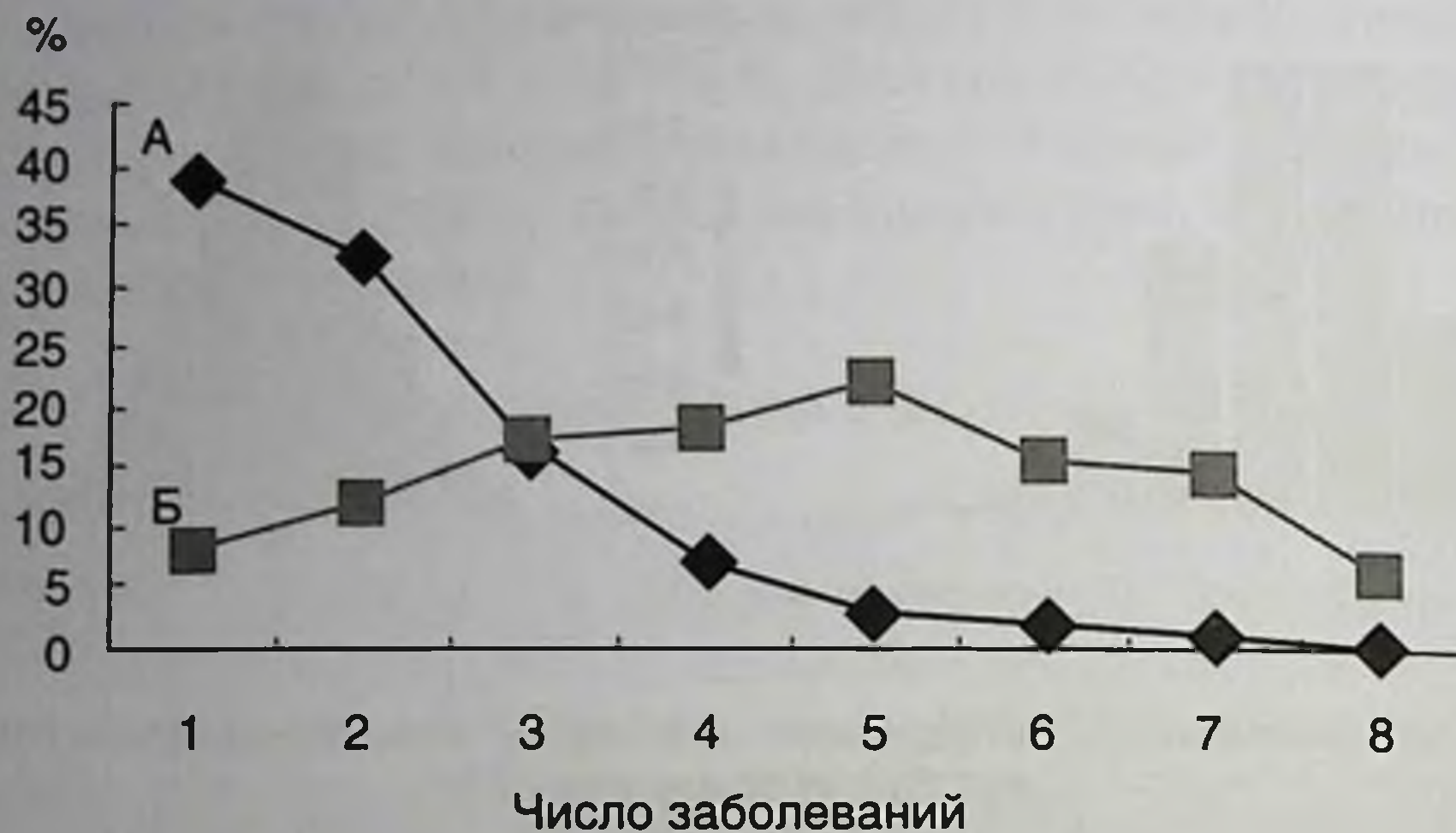


Рис. 1.42. Распределение летчиков и штурманов военной авиации по числу диагнозов хронических болезней: А — годные к летной работе; Б — признанные негодными по состоянию здоровья

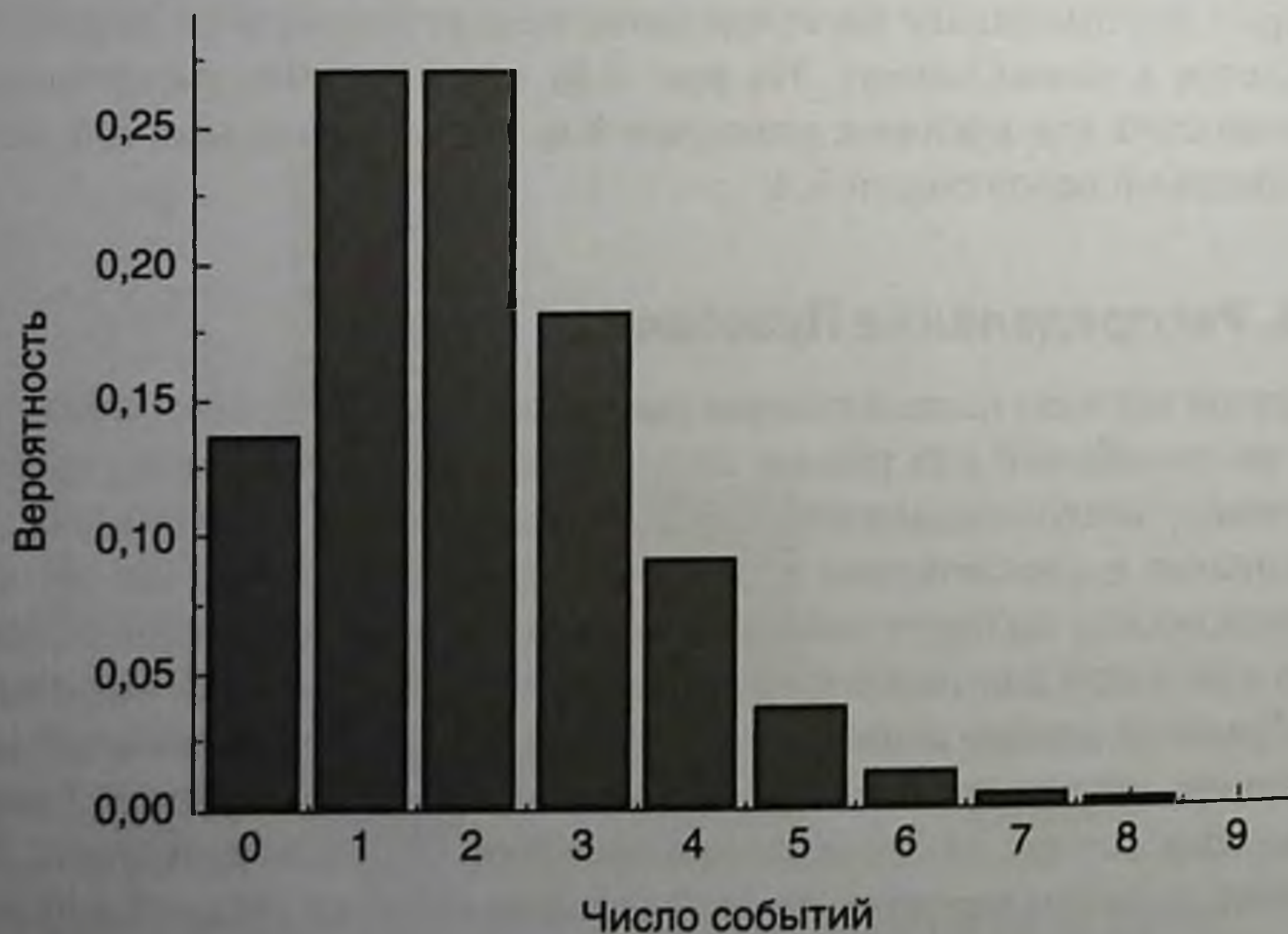


Рис. 1.43. Вероятность выявления разного числа случаев пневмонии за день при среднегодовой вероятности для поликлиники — 2 случая



число выявляемых случаев) увеличивается, форма распределения Пуассона приближается к нормальному распределению. Поэтому у годных к летной работе летчиков распределение имеет форму, сходную с приведенной на рис. 1.43, а у признанных негодными — почти симметричную.

Средняя для пуассоновского распределения является единственным параметром. Дисперсия для пуассоновского распределения равна средней. Поэтому два распределения будут идентичны, если у них одинаковые средние. Пуассоновское распределение правильно использовать для редких событий.

## Контрольные вопросы и задачи к разделам 1.5, 1.6

1. Свойствами нормального распределения НЕ являются:
  - а) равенство моды, медианы и средней;
  - б) 50% наблюдений превышают среднюю;
  - в) 68% наблюдений лежат вокруг средней в интервале шириной в два стандартных отклонения;
  - г) 68% наблюдений лежат вокруг средней в интервале от  $-2$  СКО до  $+2$  СКО.
2. У больных получены следующие числа приступов в неделю: 2, 1, 5, 3, 8, 3, 7. В этой группе мода, медиана и средняя равняются:
  - а) 2; 3; 4,14;
  - б) 3; 3; 4,14;
  - в) 3; 5; 4,83.

## 1.7. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ДАННЫХ

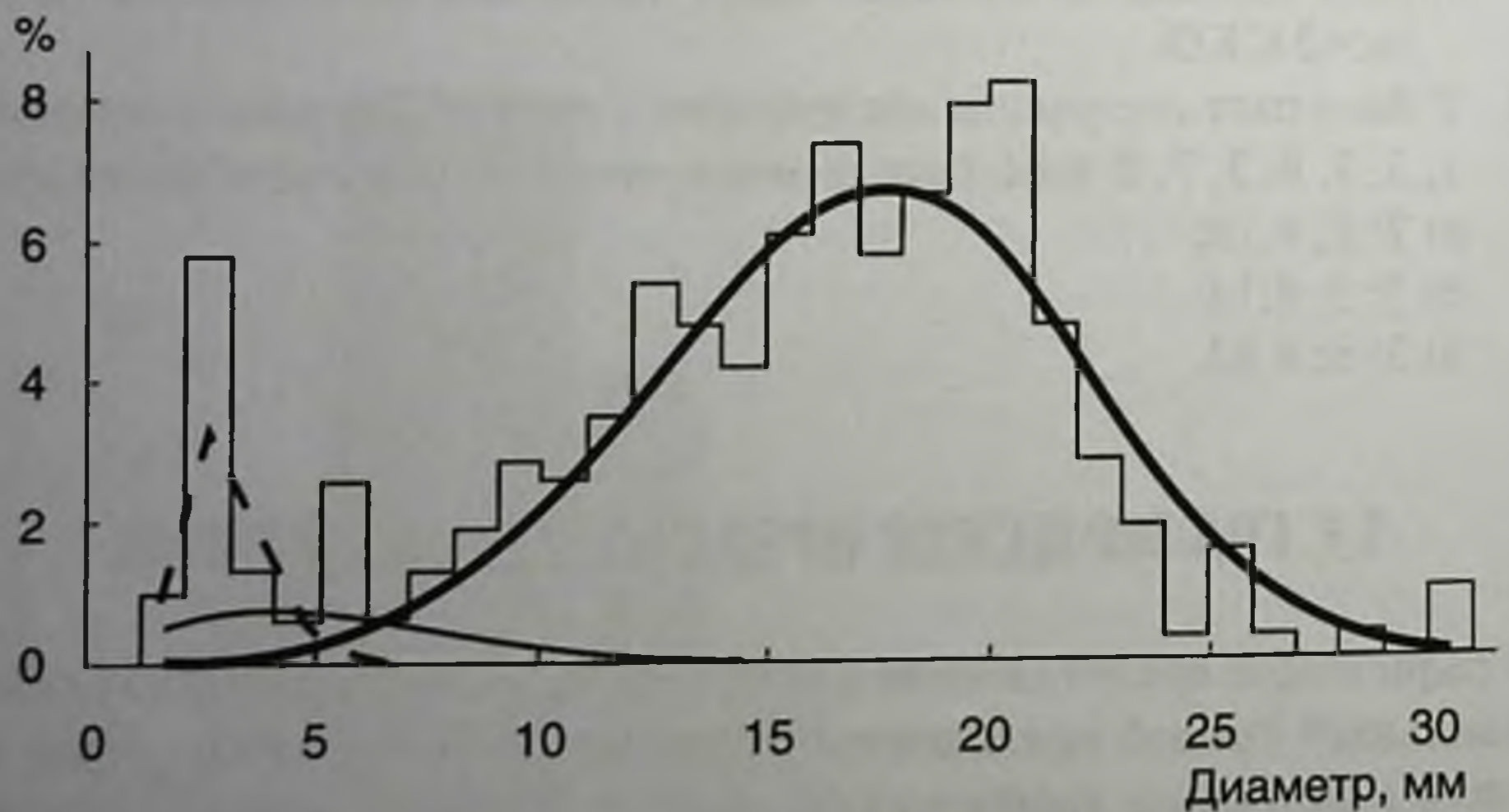
Графическое представление данных является частью анализа данных. Правильный способ представления данных позволяет ряды цифр превратить в понятные графические образы. Графики должны использоваться в соответствии с простыми правилами и быть легко понятными.

### 1.7.1. Виды графиков

*Гистограмма* используется для представления распределения случаев по величине отдельного признака, обычно количественного, реже — ординального. На гистограмме хорошо видно, какие величины



являются типичными (*мода*), насколько симметрично распределение, имеются ли «выскакивающие» величины (рис. 1.44). Визуальный анализ является важным, предваряющим применение статистических методов и влияет на выбор способа описания распределения (см. с. 45). На гистограмме по оси абсцисс откладывается величина признака. По ординате откладывается число случаев с данным значением или доля случаев с таким значением в процентах (долях от единицы). Последнее удобнее, если сравниваются гистограммы для выборок различного размера. Ширина диапазона группировки выбирается произвольно, обычно таким образом, чтобы число интервалов с отсутствующими случаями было минимальным. Иногда величина признака выражается в натуральных единицах, а иногда — в трансформированном виде: в виде логарифма или корня квадратного из исходного значения (см. врезку на с. 111). Если на ординате и абсциссе использованы натуральные единицы, то площадь гистограммы соответствует числу случаев, а площадь каждого столбика — числу случаев в данном интервале измеренной величины. Гистограмму не следует путать со столбиковой диаграммой (см. ниже).



**Рис. 1.44.** Распределение корейцев 25–29 лет, привитых вакциной БЦЖ, по диаметру индурации при туберкулиновом тесте. Абсцисса — диаметр в миллиметрах; ордината — процент лиц с такого размера индурацией. Гистограмма аппроксимирована тремя распределениями, слагающими ее: распределение реагирующих вследствие инфекции микобактерией туберкулеза (толстая линия), реагирующих вследствие БЦЖ вакцинации (тонкая линия) и вследствие инфицирования другими микобактериями (пунктир) (по Neuenschwander B.E. et al. Amer. J. Epidemiol. 2002. № 155 (7). P. 654–63, с разрешения автора)



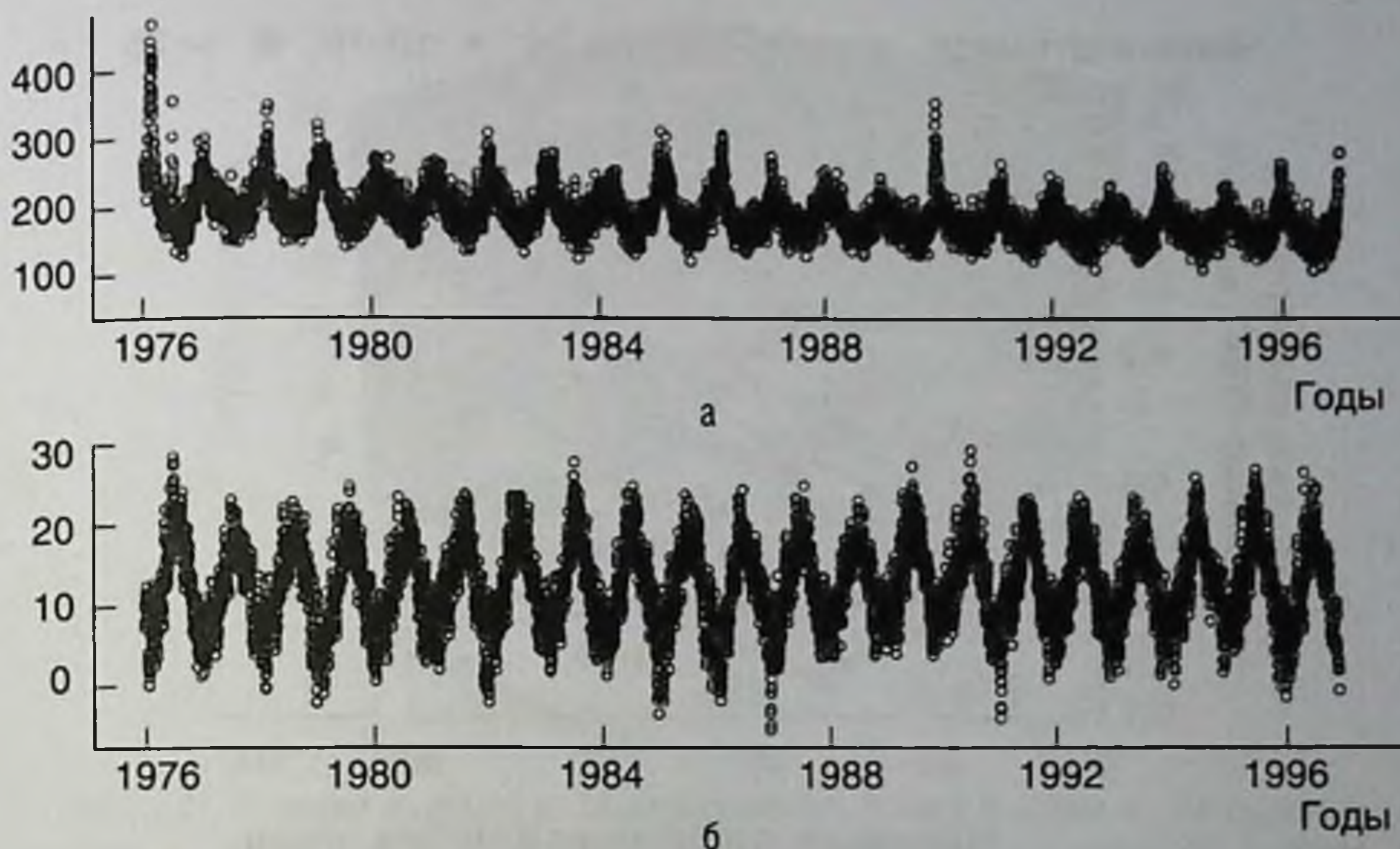


Рис. 1.45. Общая смертность (а — всего смертей) и температура воздуха (б — °С, усреднена температура на 6 и 15 ч) в Лондоне, 1976–1996 гг. Засечки на абсциссе соответствуют 1 января (по Hajat S. et al. J. Epidemiol. Community Health. 2002. № 56. P. 367, с разрешения автора)

**Скаттерграмма** (диаграмма рассеивания) используется для представления случаев по двум признакам одновременно. На скаттерграмме видно, как связаны между собой значения двух признаков. Иногда она дает почти исчерпывающее представление о довольно сложной связи (рис. 1.45). В случае наложения нескольких экспериментальных точек друг на друга обозначают буквами разное число наблюдений, так как число точек на единице площади важно для оценки скаттерграммы. При обобщении исследований с разной численностью пациентов используют разные размеры точек (рис. 1.46).

**Столбиковая диаграмма** используется для представления сведений о значении количественного признака в сравниваемых группах, выделенных по номинативному признаку (например, распространенность болезни по территориям), или частоты номинативного признака в сравниваемых группах (частота симптома при разных вариантах болезни) (пример см. рис. 1.15). На столбиковой диаграмме должны быть изображены величина стандартной ошибки или ДИ для того, чтобы было видно, как различия между столбиками соотносятся по величине с точностью оценки.



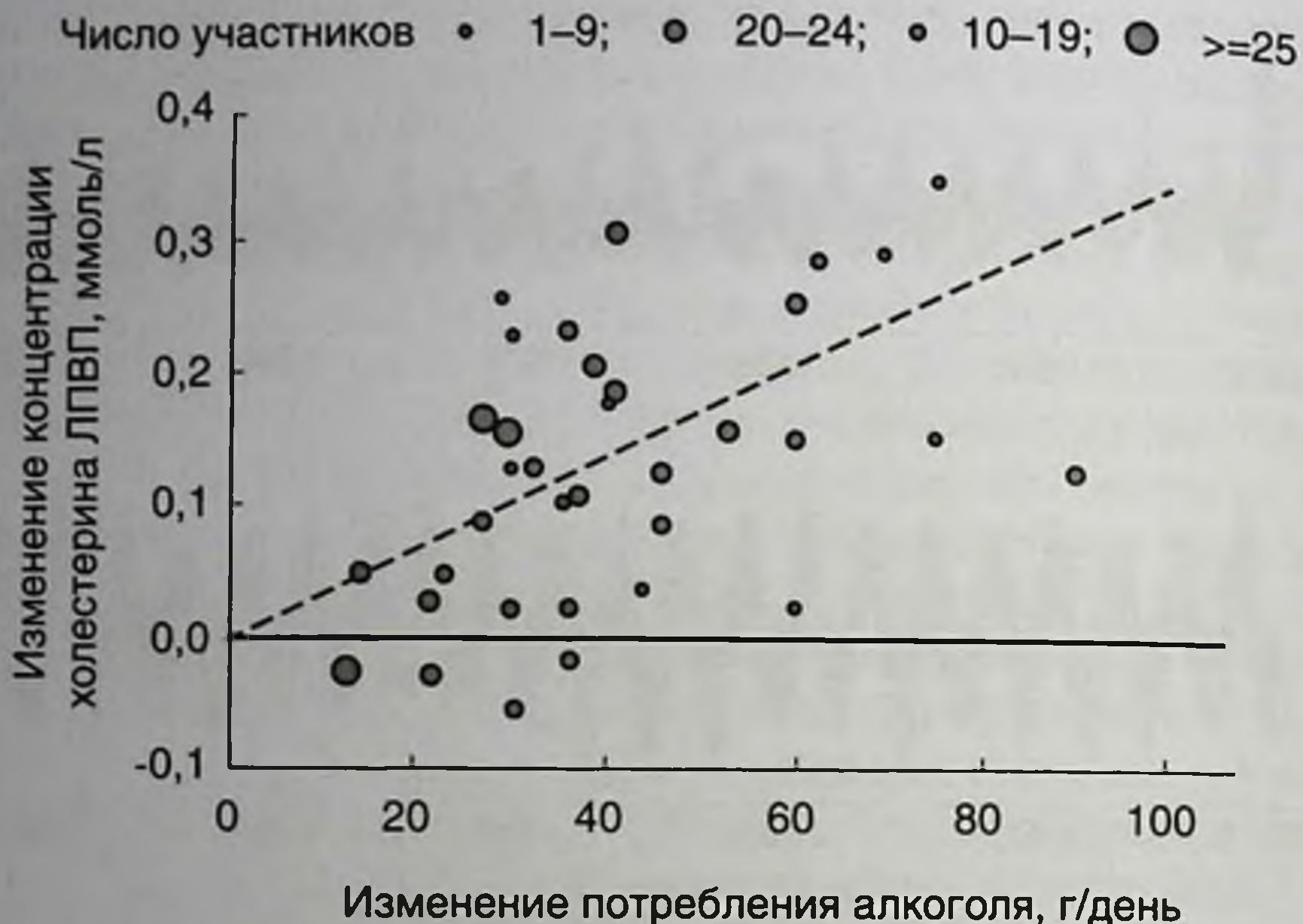


Рис. 1.46. Скаттерграмма, демонстрирующая корреляцию изменений в концентрации холестерина липопротеидов высокой плотности (ЛПВП, ордината) с изменениями в потреблении алкоголя (абсцисса). Использованы данные 36 экспериментальных исследований. Линия регрессии (пунктир) получена с учетом веса отдельных исследований, отражающего число участников исследования (по Rimm E.B. et al. BMJ. 1999. № 319. P. 1523–1528, с разрешения издателя)

Сложная столбиковая диаграмма используется, если нужно показать не только суммарную частоту, но и из чего она складывается (например, в каждой группе не только общее число работающих, но и число работающих полный и неполный рабочий день). Вариантом столбиковой диаграммы является ленточная — на ней группирующий признак располагается по ординате, а величина сравниваемого признака — по абсциссе.

Для всех видов диаграмм с распространением средств компьютерной графики стали чаще применять «объемные» (псевдотрехмерные) графические элементы — как на рис. 1.47. Этот прием может быть украшением в цветной презентации на экране, но обычно затрудняет понимание рисунка. Трехмерная столбиковая диаграмма оправдана, когда демонстрируется одновременное влияние двух факторов на изучаемый признак. На примере рис. 1.47 видно, какие это дает преимущества: влияние концентрации альбумина в плазме крови на летальность значительно больше, чем влияние возраста. Правилom представления трехмерных диа-



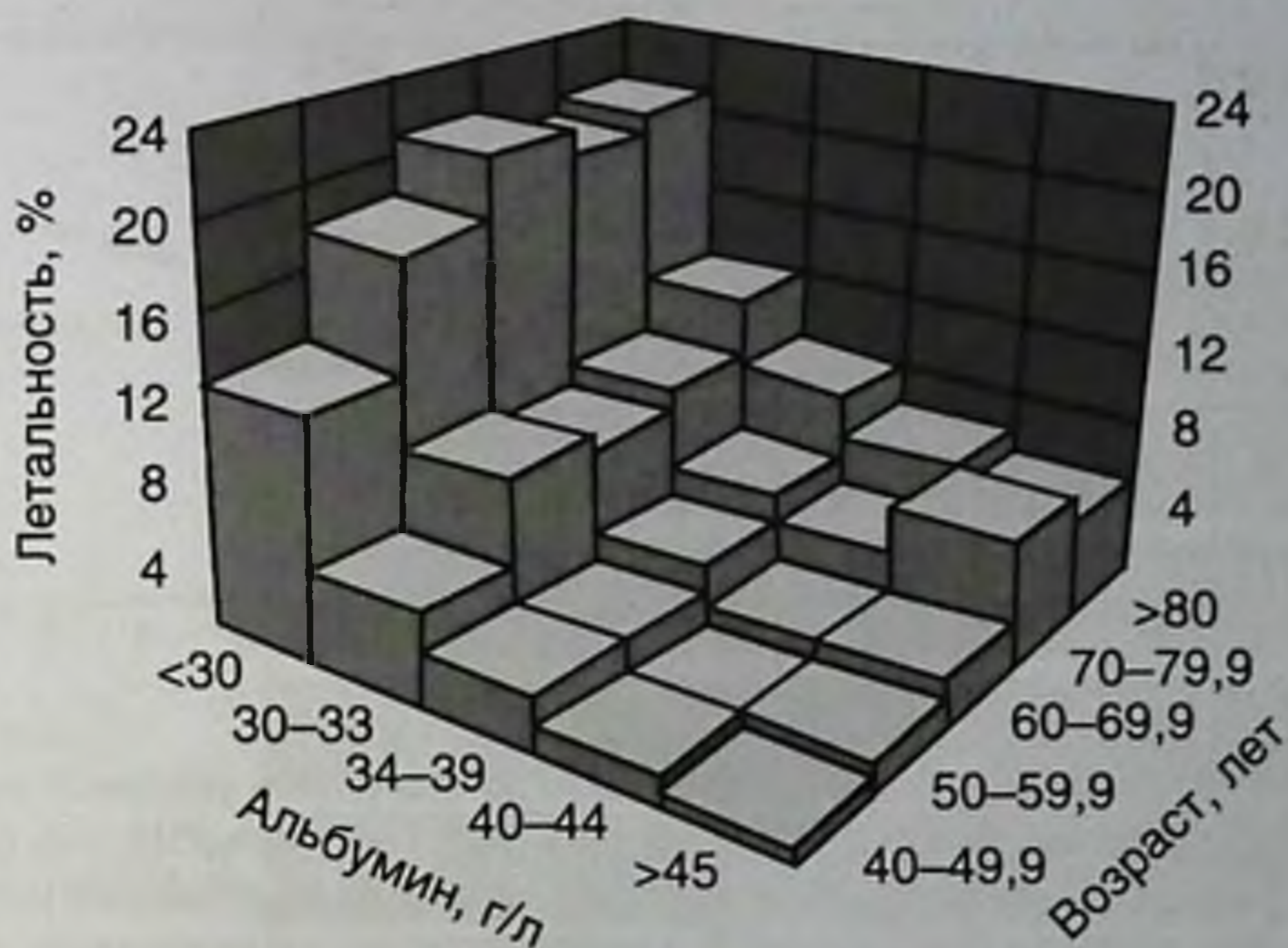


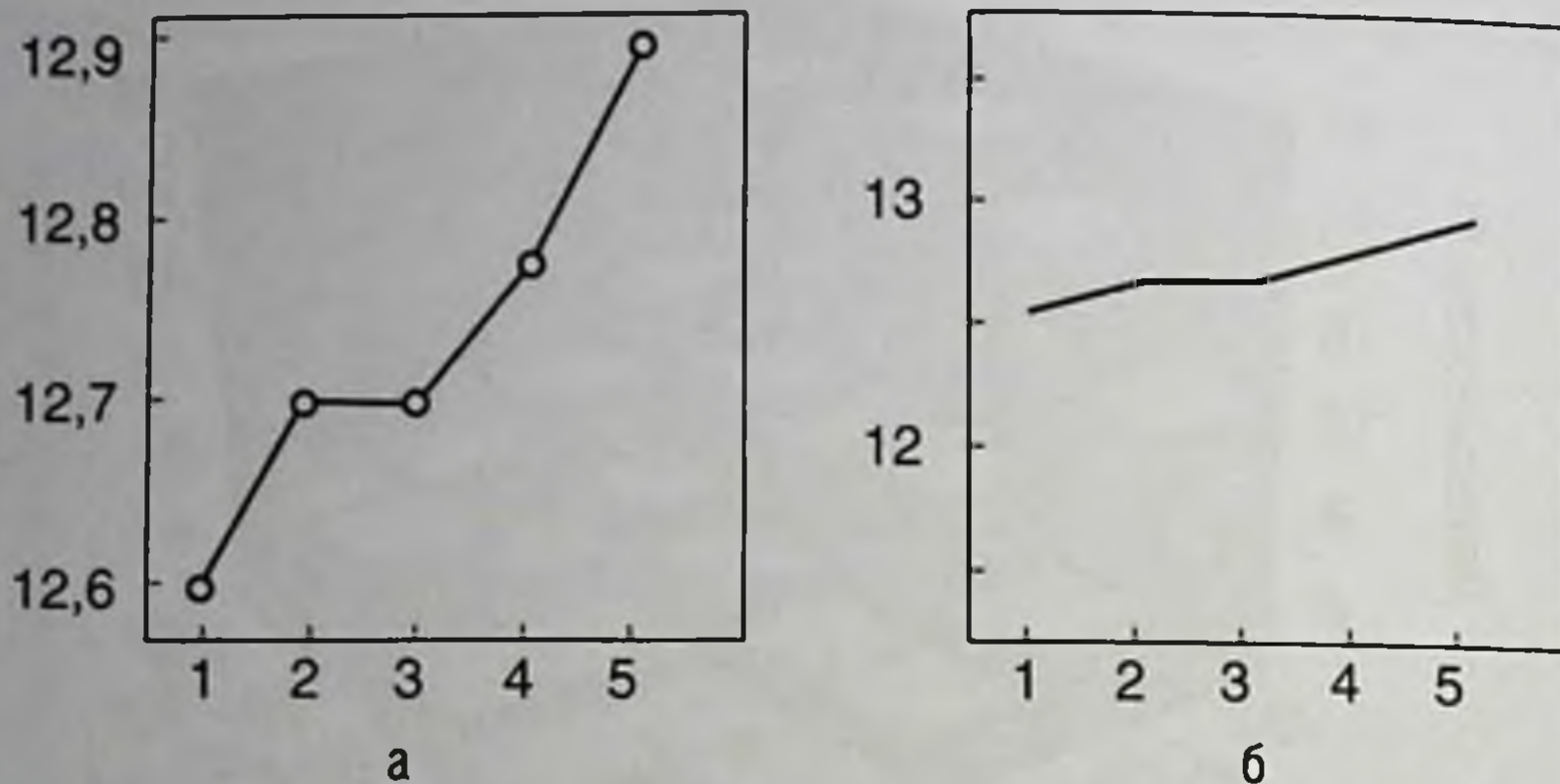
Рис. 1.47. Влияние возраста и концентрации альбумина в крови на летальность больных в стационаре. Исследование крови проведено в течение 48 ч после поступления у 50% впервые поступивших лиц в возрасте старше 40 лет; смерть наступила у 879 из 15 511 обследованных (по Hermann F.R. et al. Arch. Intern. Med. 1992. № 152. P. 125-30, с разрешения издателя)

грамм является указание на число лиц в каждой группе, так как обычно при группировке по двум признакам некоторые группы получаются малого размера, и соответственно точность оценки изучаемого признака в них мала. Обычно численность групп приводится в сопутствующей таблице.

**Ящичная диаграмма** (англ. *box-whisker plot*, иногда называется *коробочной диаграммой*) используется для представления сведений о распределении значений количественного признака (см. рис. 1.33). Эта простая диаграмма лучше всего позволяет представить соотношение параметров распределения.

**Линейная диаграмма** обычно используется для представления изменений количественного признака во времени. При этом время традиционно откладывается по абсциссе слева направо. На диаграмме должны обозначаться не только средние значения количественного признака, но и стандартные отклонения или доверительные интервалы, чтобы было видно соотношение средних с точностью их оценки. Рис. 1.48 иллюстрирует возможные варианты представления одинаковых данных без указания на точность измерения. Сложность понимания для читателя увеличивается тем, что шкала ординат обрезана снизу.





**Рис. 1.48.** Линейная диаграмма, представляющая изменения условной величины при пяти последовательных измерениях. Данные могут быть представлены различно (а, б). Если нет указаний на точность измерения в виде доверительного интервала, мы не знаем, велик ли прирост на 0,3, и вероятно ли его существование вообще

**Секторальная диаграмма** (англ. *pie chart*) используется для представления сведений о распределении случаев выборки по номинативным признакам (например, основные расходы изученных домохозяйств, распределение пациентов по диагнозам, возрасту) (см. рис. 1.16).

Иногда секторальную диаграмму создают в виде ленточной или столбиковой (рис. 1.49). В этом случае ее столбики (ленточки) имеют одинаковую длину, а различается длина составляющих столбик. Поскольку на секторальной диаграмме нельзя представить доверительные интервалы для каждого сектора (компонента ленточки), желательно на диаграмме или рядом представлять данные в цифровой форме или приводить полные табличные данные, а секторальную диаграмму использовать только для наглядности.

**Представление циклических процессов.** Циклические процессы могут быть развернуты по шкале времени, как это сделано на скаттерграмме на рис. 1.45, но могут быть представлены в виде круга, логично замыкающего суточный или годовой цикл. Впервые такую диаграмму применила Флоренс Найтингейл — основоположница сестринского дела и один из основоположников менеджмента в медицине. На круговой диаграмме, которую она называла «кокскомб» (*Coxcomb*), была отражена смертность английских солдат в Крымской войне от ран и от прочих причин (рис. 1.50). Так же как секторальная диаграмма, кокскомб отражает количественный показатель площадью соответствующего



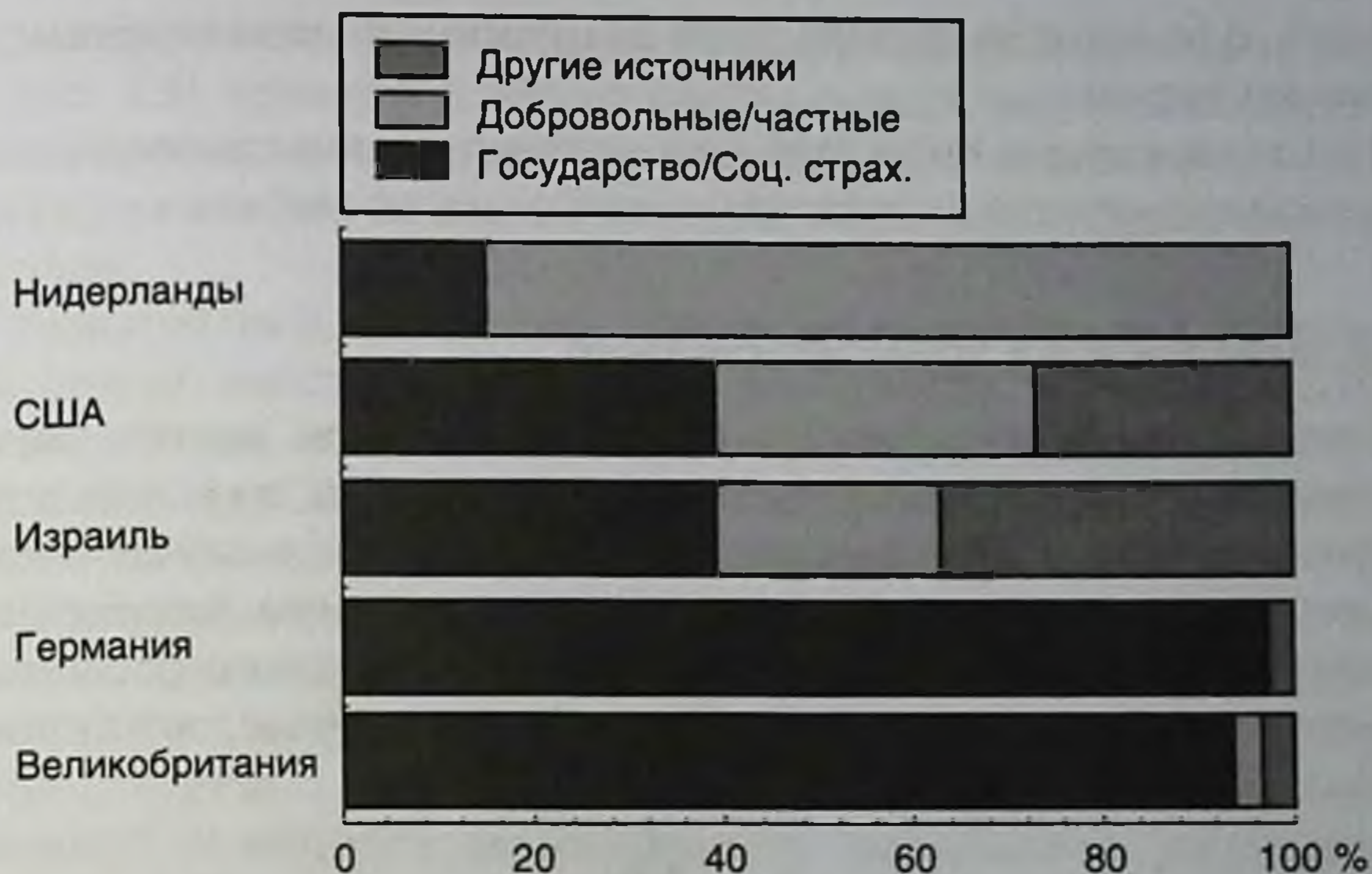


Рис. 1.49. Структура источников средств здравоохранения в некоторых экономически развитых странах в начале 1990-х годов (по: Медицина в условиях дефицита ресурсов. М.: Триумф, 2000. С. 32)

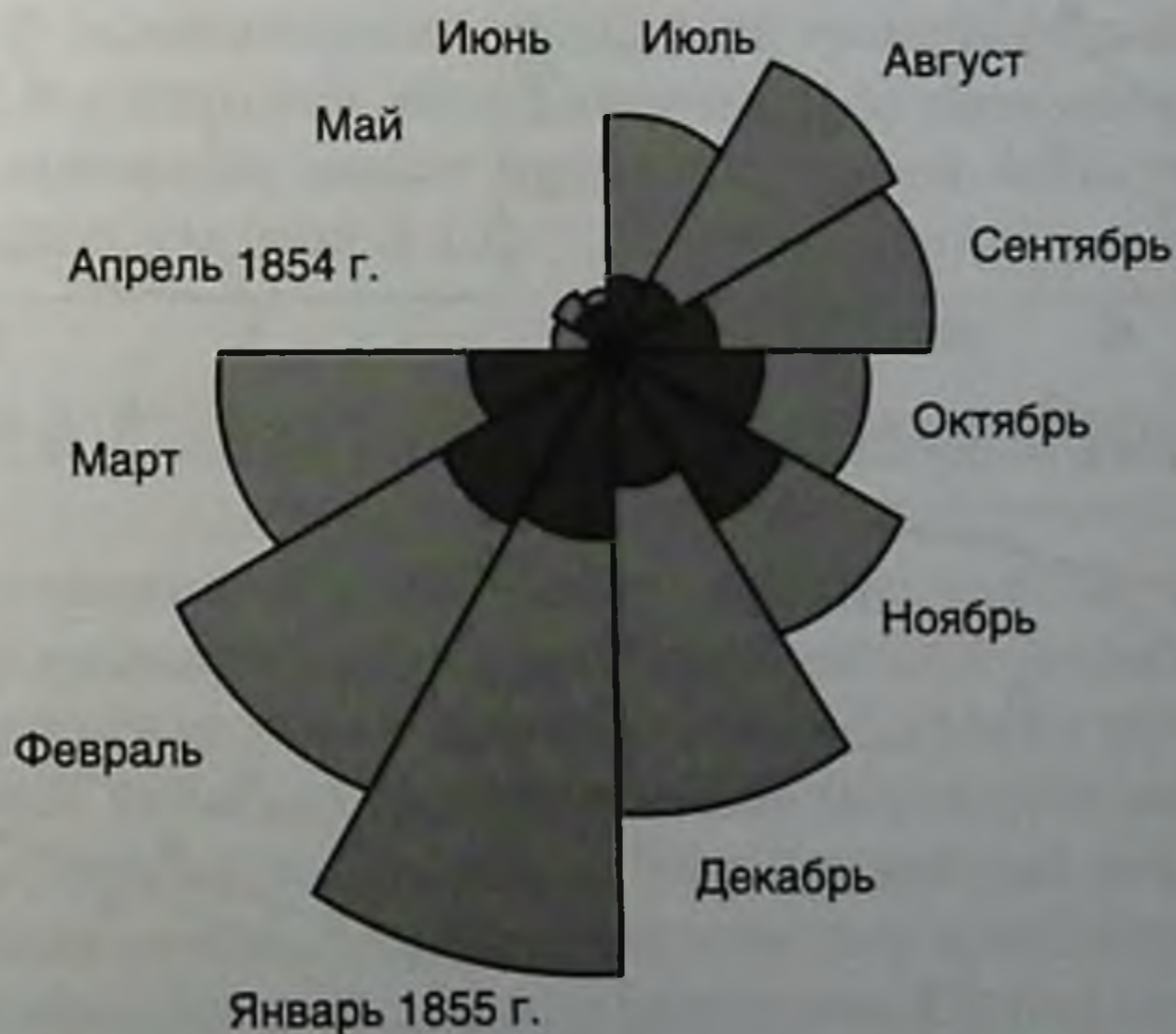


Рис. 1.50. Смертность солдат английской армии от ран (темно-серый цвет в центре) и от небоевых причин (болезней) по месяцам с апреля 1854 до марта 1855 г. По Notes on Matters Affecting the Health, Efficiency and Hospital Administration of the British Army (1858)



сектора. В отличие от секторальной диаграммы, углы постоянны (они отражают время).

Некоторые другие виды графиков, используемые в специальных целях, рассматриваются в соответствующих разделах учебника.

### 1.7.2. Шкалы и их трансформация

При построении графиков и при статистическом анализе данных мы сталкиваемся с разными формами зависимостей, из которых простейшей является прямолинейная зависимость. Как в случае анализа распределения величин (см. разд. 1.6), использование теоретических распределений для аппроксимации экспериментального распределения помогает более эффективно анализировать данные, точно так же

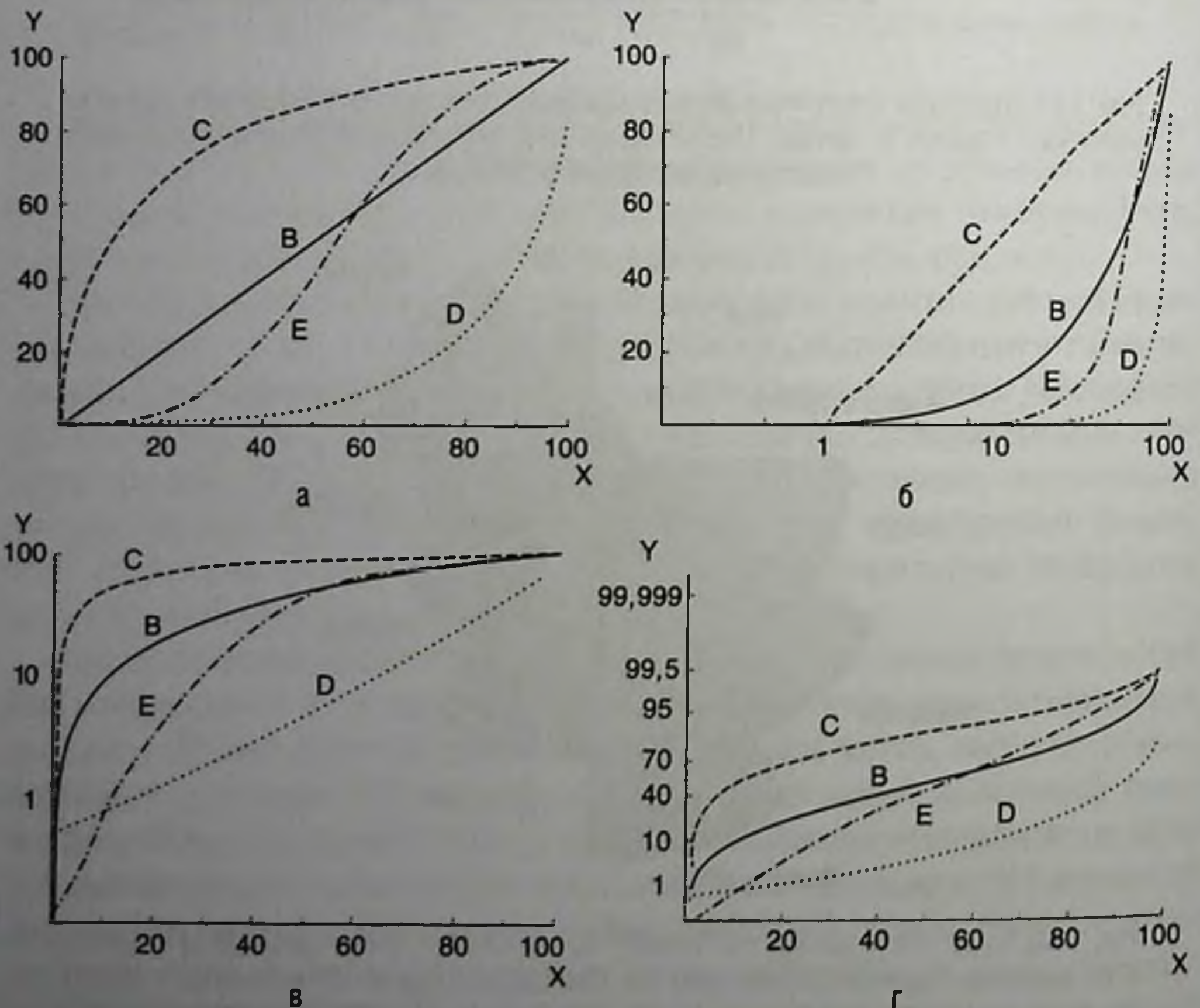


Рис. 1.51. Примеры преобразования формы зависимостей. Пояснения в тексте



знание важнейших форм зависимостей упрощает анализ зависимостей. На рис. 1.51 приведены четыре наиболее часто встречающиеся в медицине формы зависимостей для двух условных величин  $X$  и  $Y$ . В данном случае для нас не важно, носит ли связь переменных причинный характер.

Зависимость В на линейных шкалах (а) не нуждается в упрощении. Три другие зависимости сложнее. Зависимость С трансформируется в линейную, если вместо «сырого» значения  $X$  использовать логарифм (б). Зависимость D зеркальна к зависимости С относительно прямой В. Она приближается к прямой, если вместо «сырого» значения  $Y$  использовать его логарифм (в).

Зависимость Е представляет собой более сложный вариант. Она соответствует по форме кумулятивной кривой нормального распределения (см. разд. 1.6.3 и рис. 1.32). Для приведения этой зависимости к прямой используют закономерность нормального распределения (см. рис. 1.51, г). На ординате видно, что нормальная шкала сжата в центре и растягивается к флангам. Анализ, использующий трансформацию на основе свойств нормального распределения, называется *пробит-анализом*. Он используется в биологии и фармакологии, но редко в медицине и эпидемиологии. Используя трансформацию изученных переменных и превращая экспериментальную зависимость в прямую линию, исследователь может применить к ней более эффективные средства анализа (см. разд. 2.8.2).

## 1.8. СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ

Результат анализа данных зависит прежде всего от правильности организации исследования, а не от правильности вычислений. Статистический анализ не может компенсировать дефекты построения эксперимента или наблюдения. Необходимость статистического анализа в биологии и медицине вытекает из того, что каждый человек и каждое животное неповторимы. Нормальные процессы и болезни в организмах отдельных животных протекают неодинаково. Соответственно наблюдая выздоровление у одного человека, нельзя быть уверенным в том, что то же будет наблюдаться у второго, третьего. Возникает необходимость описать, какой же эффект получается *обычно*, наиболее часто, для чего применяются методы описательной статистики.



Чтобы показать, как эффект одного лечения отличается от эффекта другого, надо сопоставить лечебные эффекты применения этих средств. Если различие сравниваемых средств полностью скрывается за различием лечебного эффекта у разных людей, то это означает отсутствие заметных различий. Исследование движется от гипотезы к построению выборочного исследования для ее проверки и к получению выводов, которые можно отнести ко всей популяции с определенной вероятностью. Такие исследования называют *аналитическими*, а соответствующие статистические методы — *аналитической статистикой* или *статистическими выводами* (англ. *inferential statistics*).

### 1.8.1. Формирование гипотезы

Представим себе человека, который с удивлением обнаружил, что высшие руководители профсоюзов СССР с фамилией на «Ш» сменяли друг друга трижды. Взяв в руки телефонный справочник, человек убедился в том, что фамилии, начинающиеся на «Ш», не часты — примерно 4% в его родном городе, вполне типичном для России. Значит, вероятность случайного совпадения трех подряд таких фамилий составляет:  $0,04 \times 0,04 \times 0,04 = 0,000064$ . Вероятность настолько мала, что возникает сомнение в том, что это сочетание случайно. Беда такого «исследования» в том, что нарушено основное правило научного исследования: сначала гипотеза, потом ее проверка.

Если нарушать это правило, то возникает такая же ситуация, что в известной шутке о техасском снайпере. Он никогда не промахивался в стрельбе по мишени потому, что сначала стрелял, а потом рисовал мишень вокруг дырки от пули. Оценивать вероятность события, которое заведомо произошло, — бессмысленно. Гипотеза, которая подвергается исследованию, должна быть ясно сформулирована до начала исследования и *записана*. История науки знает немало примеров того, как исследователи изменяли гипотезу по ходу исследования, в согласии с теми данными, которые они получали. Результат такого исследования неизменно положителен: гипотеза изменяется в соответствии с данными, и данные, естественно, подтверждают гипотезу. Но результаты исследования с таким **тяжким** дефектом никому не нужны.

Аналогичная ситуация складывается, когда инструмент создается и проверяется на одной и той же группе людей. Представим, что исследователь обнаружил, что в старших классах средней школы № 114



его города школьники, которые имеют разницу в длине ног более 1 см, чаще умеют играть на каком-либо музыкальном инструменте, и построил формулу, позволяющую прогнозировать освоение музыкальных инструментов по разнице в длине ног. Сила связи превышает ожидаемую случайно, и, полагая, что находка представляет интерес, исследователь представил материал в печать. Вскоре он получил отрицательный отзыв рецензента журнала, который указывал на то, что построенная модель прогнозирования не была проверена на независимой выборке. Действительно, проверка этой прогностической формулы на учащихся другой школы покажет, что прогнозирование не удастся. Иначе говоря, обнаруженная в школе № 114 связь была случайной.

Нередко подобные случайные находки возникают, когда исследователи «просеивают данные». Например, они могут обследовать школьников по множеству признаков социологического, демографического, антропометрического, психологического, биохимического характера и затем среди этих признаков и их комбинаций искать те, которые связаны с успехами в музыке. «Если долго мучить имеющиеся данные, они обязательно скажут, что тебе надо». Среди десятков признаков найдется один-два, которые будут коррелировать с успехами в музыке. Однако утверждать, что найден способ прогнозирования, рано. Такого рода исследования дают только основания для выдвижения гипотезы. После этого гипотеза о связи асимметрии длины ног с успехами в музыке должна быть проверена на независимой выборке.

### 1.8.2. Проверка гипотез

Наиболее распространенный тип анализа — проверка предположения (гипотезы) о том, что найденные различия между двумя выборками неслучайны и, следовательно, отражают действительные различия между двумя генеральными совокупностями. В типичном случае такими совокупностями являются: 1) все больные данной болезнью, получающие традиционное лечение, и 2) все больные, получающие новое лечение. Неслучайность различия между двумя репрезентативными выборками из одной популяции, получавшими разное лечение, означает, что новое лечение в подобных условиях на самом деле лучше традиционного, т.е. для всех таких больных.

Различие групп означает их неравенство по некоторому показателю (инциденсу осложнений, сроку выздоровления, степени ослабления



боли). Чтобы обосновать этот вывод, применяется формальный логический прием. Доказать правильность гипотезы проведением подтверждающего ее эксперимента нельзя. Но можно высказать противоположную гипотезу, и отклонить ее опытом (см. разд. 1.2). Поэтому сначала предполагается, что группы неразличимы (*нулевая гипотеза*). Затем проверяется, может ли эта гипотеза быть отклонена. Отклонение нулевой гипотезы и означает неравенство групп, т.е. принятие *альтернативной гипотезы*. Альтернативная гипотеза состоит в том, что различия между группами не случайны, а отражают истинные различия между генеральными совокупностями, из которых получены выборки. Это могут быть различия в центральной тенденции, что обычно в медицине, в дисперсии, и в других параметрах распределения. Для испытания нулевой гипотезы используются разнообразные *статистические критерии* (реже называемые «статистические тесты»).

### Статистические критерии

Статистическими критериями или критериями для проверки гипотез называют математические показатели, вычисляемые на основании экспериментальных данных, для того, чтобы установить, не являются ли полученные результаты, подтверждающие гипотезу исследователя или опровергающие ее, только лишь случайными. Обычно исследователя интересует, не является ли *эффект*, т.е. различие между двумя группами пациентов или изменение в результате вмешательства, случайным. Для этого вычисляется величина статистического критерия. В принципе она вычисляется так:

$$\text{Статистика критерия} = \frac{\text{Наблюдаемая величина} - \text{Ожидаемая величина}}{\text{Стандартная ошибка наблюдаемой величины}}$$

В некоторых ситуациях случайно ожидаемая величина равна нулю, как, например, в ситуации случайного различия между группами, выбранными из одной популяции. Тогда статистика критерия (величина критерия) равняется отношению величины эффекта к точности его оценки. Для того чтобы оценить статистику критерия, ее сравнивают со стандартным распределением этой статистики. Это может быть нормальное распределение (см. разд. 1.6.3) или специальное распределение для этой статистики. Если полученная в результате расчета статистика критерия очень мала или очень велика, то можно откло-



нить нулевую гипотезу. Очень малые и очень большие величины располагаются в правом и левом «хвостах» распределения (см. рис. 1.34). Отсюда английское название для двухсторонних статистических критериев — двуххвостные (*two-tailed*). Случайные отклонения могут быть как отрицательными (левый хвост распределения), так и положительными (правый хвост). Для статистического критерия, имеющего нормальное распределение, пороговая величина допустимой вероятности ошибочного отклонения нулевой гипотезы 5% достигается при значении критерия 1,96 (в единицах среднего квадратического отклонения). Если же использовать односторонний тест, то оказывается достаточным значение критерия 1,64. Поэтому односторонние критерии должны использоваться как исключение в ситуации, когда изучаемая разница может наблюдаться только в одном направлении. Обычно уверенность в направлении различия отсутствует. Выбор одностороннего критерия должен быть сформулирован в плане (протоколе) исследования.

Статистику критерия ранее оценивали по специальным таблицам, но в настоящее время вычисляют ее и соответствующие вероятности ошибок с помощью компьютерных программ. В учебнике оценка по таблицам используется для пояснения принципа.

### Два типа ошибок

Никакой статистический критерий не дает абсолютной уверенности в различии групп или в их идентичности. Напротив, все статистические критерии позволяют утверждать что-либо лишь с некоторой вероятностью ошибки вывода. Чаще всего исследователя интересует различие, и, соответственно, речь идет об отклонении нулевой гипотезы. Поэтому оценивают именно вероятность (риск) ошибочного решения о том, что различие существует (ошибка типа I, произносится как «первого типа» — табл. 1.9). Этот риск называют *альфа-риском* ( $\alpha$ -риском). Для него устанавливается критическая величина вероятности ошибки  $P$  на уровне, традиционно равном 0,05 или 0,01. Традиция использовать критическую величину для оценки  $\alpha$ -риска основана на использовании вычислений на бумаге: когда точные вычисления были затруднены, по таблицам оценивали, не превышен ли критический уровень  $\alpha$ -риска.

Правильно вероятность ошибки  $P$  приводить не в отношении порогового значения ( $P < 0,05$ ), а в виде точной величины (например,  $P = 0,0017$ ). Все современные статистические программы рассчитывают точную величину  $P$ . При  $P < 0,0001$  точная величина уже не имеет



значения, поскольку для медицинских решений это слишком низкие вероятности. Если вероятность ошибки типа I близка к пороговой, то при обсуждении различий не может быть разного подхода к случаям, когда  $P = 0,04$  и  $P = 0,06$ . Подход к оценке таких результатов должен быть близким, несмотря на то, что формально один достигает условного уровня статистической значимости, а другой — нет.

Таблица 1.9. Примеры и терминология применительно к ошибкам I и II типов

Ситуация	Ошибка	
	Тип I: нулевая гипотеза отклонена, когда она истинна (т.е. различия отсутствуют)	Тип II: нулевая гипотеза не отклонена, когда она неверна (т.е. различия имеются)
Обозначение вероятности ошибки	$\alpha$	$\beta$
Традиционно допустимый уровень	0,05, высокозначимо при 0,01	0,2
Отношение к чувствительности статистического критерия	Непрямое	1-чувствительность критерия
Применение в диагностической практике	Ложноположительное решение	Ложноотрицательное решение
Отношение к диагностической чувствительности	Непрямое	1-чувствительность
Отношение к диагностической специфичности	1-специфичность	Непрямое
Испытание способа лечения	Неэффективное названо эффективным	Эффективное названо неэффективным
Контроль качества в промышленности	Риск продавца	Риск покупателя
Решение, лечить ли пациента	Не лечить, когда могла бы быть польза	Лечить, когда пользы нет

Достижение установленной, достаточно малой величины  $P$  свидетельствует лишь о том, что различие между группами существует, точ-



нее, существует с высокой вероятностью. Статистическая значимость различия ничего не говорит о происхождении различий. Различия могут быть обусловлены тем фактором, который изучает исследователь, но могут быть и следствием иных различий между группами (следствием действия вмешивающихся факторов) (см. разд. 2.7.4). Поэтому при чтении научного сообщения надо внимательно анализировать организацию исследования. Именно она определяет, какие результаты получает исследователь. Если исследователь комплектовал сравниваемые группы больных так, что они могли различаться по возрасту, уровню дохода, стадии болезни или ее форме, то статистически значимое различие в эффекте лечения может быть обусловлено именно этими факторами, а не различиями сравниваемых лекарств.

Если исследователь желает более уверенно утверждать, что группы различаются, то он волен устанавливать очень высокий порог статистической значимости, например  $P < 0,001$  (не отклонять нулевую гипотезу, если  $P$  больше или равно  $0,001$ ). Но в таком случае увеличивается вероятность того, что он установленного уровня значимости не достигнет и не обнаружит различие, которое в действительности — между генеральными выборками — существует (возрастает  $\beta$ -риск). Такая ошибка — неотклонение нулевой гипотезы, когда она неверна, называется *ошибкой типа II*.

Такая же ситуация (ошибка типа II) возникает, если число пациентов в сравниваемых группах невелико, и поэтому реально существующее различие между вмешательствами не обнаружено. Обычно приемлемой считается вероятность ошибки типа II (уровень  $\beta$ -риска) не выше  $0,2$ .  $\beta$ -Риск, равный  $0,5$ , означает, что с вероятностью  $0,5$  (50%) исследователь ошибается, считая, что различий между выборками нет.

### 1.8.3. Статистическая мощность

Величину  $(1-\beta)$  называют *мощностью статистического критерия* или *мощностью эксперимента, исследования* (по аналогии с англ. *power*). Понятие «статистическая мощность» отражает возможность с помощью эксперимента выявить различие. Поэтому на русском ее иногда обозначают как «чувствительность» по аналогии с чувствительностью диагностического теста. Применяя более мощные статистические критерии, можно (в некоторых пределах) выявить значимое различие там, где менее мощные его не выявят. Обычно достигается такое преодоление



формального порога значимости, например  $\alpha = 0,05$ , ценой принятия некоторых статистических гипотез, которые могут быть несправедливы для этого частного случая. Например, различие между двумя группами может быть не установлено по критерию знаков (см. разд. 2.4.5), но оказаться статистически значимым при использовании  $t$ -критерия, применение которого возможно только при соблюдении ряда условий (см. в разд. 2.5.2). Увеличивая мощность (снижая  $\beta$ -риск), исследователь почти неизбежно увеличивает вероятность ошибки типа I. Единственный способ увеличить мощность эксперимента без увеличения риска ложноположительного результата исследования — увеличить численность сравниваемых групп.

Во многих областях человеческой деятельности имеет место схожее отношение к ошибкам типа I и II. Как в медицине допускается более высокая вероятность ошибки типа II, так и в просвещенной судебной практике традиционно считается, что «лучше, если сто преступников будут оправданы, чем будет осужден невинный». Наиболее остра проблема ошибок типа II в испытаниях методов лечения на малых группах пациентов. В случае малого числа пациентов трудно отклонить нулевую гипотезу, и она не отклоняется, так как не доказано обратное. Между тем в действительности преимущество у одного из вмешательств может существовать, но лишь не быть выявленным в эксперименте. Теоретически любое, самое малое различие между сравниваемыми группами может быть показано с любым, самым высоким уровнем статистической значимости (при самой малой величине  $P$ ), если обследовать достаточно большую группу пациентов. Возможности исследователя в увеличении численности группы ограничены.

Наиболее частая ошибка в медицинских исследованиях — отсутствие заблаговременного, до начала исследования, расчета необходимой численности исследуемых пациентов. Если заранее, исходя из ожидаемой величины эффекта (например, ожидаемого преимущества нового вида лечения в виде частоты выздоровления или срока нормализации признака), рассчитать, сколько потребуется пациентов в таком исследовании, то, возможно, окажется, что для получения статистически значимого результата достаточного числа больных никогда не удастся собрать. И наоборот, может оказаться, что достаточно обследовать значительно меньшее число пациентов, чем сначала предполагалось. Например, используя статистический калькулятор, входящий в программу EpiInfo (см. сноску на с. 103), мы без труда рассчитаем,



что для обнаружения 5-кратного различия в смертности между двумя группами больных (1 и 5%) при  $\alpha$ -риске 5% и  $\beta$ -риске 20% понадобится не менее 332 пациентов в каждой группе. Проводить такие расчеты не только рационально, но и принципиально важно, так как определять, достаточно ли пациентов, «по ходу исследования» ориентируясь на достижение статистической значимости получаемого результата, неправильно. В таком случае получается, что решение о завершении эксперимента принимается в зависимости от того, какой результат достигается (см. также разд. 3.4.3).

Обычно исследователь ожидает получить различия в изучаемых группах в подтверждение своей гипотезы, например, о преимуществах нового способа лечения. Если критический уровень статистической значимости не достигнут, то это не означает отсутствия различия. Правильный вывод: *предполагаемое различие не обнаружено*. Отсутствие статистически значимого различия ( $P > 0,05$ ) не означает, что эффекта нет (вмешательства действуют одинаково), а означает лишь, что полученные данные не позволяют отклонить нулевую гипотезу (об отсутствии эффекта). Чтобы доказать, что эффекта нет, надо рассчитать  $\beta$ -уровень, и если он окажется достаточно низок, например ниже 0,1, то можно утверждать, что эффекта нет с данным уровнем вероятности ошибки типа II. Обоснование выводов принятием нулевой гипотезы (тем, что ее не удалось отклонить) является частой ошибкой в медицинских исследованиях. Отсутствие статистически значимых результатов исследования не означает, что исследование выполнено плохо. Это предубеждение. Оно приводит к тому, что многие исследования, не принесшие статистически значимых результатов, не публикуются. Поэтому из опубликованных данных можно получить преувеличенные представления о действенности почти любого вмешательства, в том числе лекарства по сравнению с плацебо. Это один из вариантов так называемого *публикационного смещения*.

Известно, что при планировании клинических испытаний всегда существует этическая дилемма: если испытываемое лечение неэффективно, то подвергать ему большие группы пациентов неэтично. С другой стороны, на малых группах статистически значимо выявить преимущества невозможно. Единственным приемлемым решением этой проблемы является добросовестное планирование исследования, опирающееся на лучшие имеющиеся данные об ожидаемом эффекте и частоте изучаемых исходов.



### 1.8.4. Выбор статистического критерия

Выбор критерия диктуется характером данных, полученных в ходе исследования, структурой исследования и вопросом, поставленным для исследования. В выборе критерия можно руководствоваться приложениями 3–5. Существует много критериев, большая часть которых применяется редко, и много вариантов их применения. В учебнике рассмотрены только наиболее часто используемые критерии. Для подробных консультаций имеются специальные руководства и компьютерные программы<sup>1</sup>. Такие программы задают вопросы о том, каковы тип данных и организация исследования, какие вопросы ставит перед собой исследователь, и рекомендуют оптимальные статистические методы. Некоторые пакеты статистических программ после того, как исследователь ввел в таблицу свои данные и присвоил переменным их тип (номинативная, ординальная), при попытке провести анализ сразу предлагают соответствующие варианты. Так, например, действуют программы JMP (SAS Institute, Inc.) и SPSS (IBM). Выбор статистического критерия — важнейший элемент плана исследования, и в планировании желательно руководствоваться рекомендациями специалиста по статистике.

Обсуждая структуры исследований, мы упоминали, что контрольная группа может быть разной. Если помещение пациента в нее не обусловлено помещением другого пациента в опытную группу, то это случай *независимых выборок*. Если в опыте и контроле исследуются одни и те же люди, но в разное время (например, до и после приема лекарств), то это *связанные выборки*. Если пациенту опытной группы подбирается пара равного возраста и пола в контрольной группе — это тоже вариант связанных выборок (подобранные пары, англ. *matched pairs*). Получать связанные выборки труднее, но выгодно, поскольку статистический анализ связанных выборок более эффективен. Важно, что *статистические критерии, предназначенные для связанных выборок, не должны использоваться для независимых выборок, и наоборот*.

Важным аспектом статистической проверки гипотез является *выбор одно- или двухстороннего статистического критерия* (англ. *one- or two-tailed test*). Действительно, если исследователь изучает действие инги-

<sup>1</sup> См., например, <http://www.statstutor.ac.uk/> и <https://stats.idre.ucla.edu/other/mult-pkg/whatstat/>



битора холинэстеразы на ее активность, то он может ограничиваться ожиданием только одного эффекта — более низкой активности фермента у подвергнутых экспозиции людей по сравнению со здоровыми. Если же сравнивается действие двух инсектицидных препаратов, то возможно, что первый окажется активнее второго, или наоборот. Соответственно, нужно изучить обе возможности. Практически важно, что критическая величина  $P$  для критерия двухсторонних различий выше и поэтому достигается труднее. Из-за этого некоторые исследователи используют односторонний критерий там, где следовало бы использовать двусторонний.

Иногда, в особенности в исследованиях методов диагностики, анализируют совместно данные различного типа на предмет соответствия значений одних признаков другим. Для этих целей выбор статистического критерия нуждается в особом уточнении (см. приложение 4). Для непараметрических критериев, т.е. критериев, которые могут применяться к различного типа данным, а в случае интервальных данных используются независимо от формы распределения, приведена более детальная таблица (см. приложение 5).

## Контрольные вопросы и задачи к разделам 1.7, 1.8

1. Выберите правильные утверждения относительно статистического анализа:
  - а) если  $P = 0,05$ , то имеется 95% вероятность, что результат не является случайным;
  - б) нулевая гипотеза утверждает, что между группами нет различий;
  - в) если  $P$  достаточно велико, то нулевую гипотезу нельзя отклонить;
  - г) знание метода формирования выборки не имеет значения для оценки статистической значимости.
2. В отношении статистической проверки гипотез правильны утверждения:
  - а) если величина  $P$  очень мала, то различия между сравниваемыми группами должны быть очень большими;
  - б) если исследовать достаточно большие группы, то можно достичь статистической значимости  $P = 0,05$  даже при отсутствии различий;



- в) стандартная ошибка описывает, насколько хорошо выборочная средняя приближается к популяционной;
  - г) при прочих равных условиях использование двустороннего критерия реже приводит к отклонению нулевой гипотезы, чем использование одностороннего.
3. В отчете сказано, что между сравниваемыми группами есть различие в инциденте болезни, статистически значимое. Это различие может быть результатом (выберите все правильные ответы):
- а) случайной ошибки;
  - б) наличия причинной связи;
  - в) наличия непрямой связи;
  - г) наличия привнесенной связи.
4. В исследовании терапии артрита найдено, чем сильнее снижается концентрация вещества  $X$  в плазме, тем меньше отек, лучше подвижность, меньше боль. Каким способом лучше всего продемонстрировать графически связь каждого такого исхода с концентрацией?
- а) секторальной диаграммой, где секторы соответствуют концентрации;
  - б) столбиковой диаграммой, где столбики соответствуют выраженности симптома, а их высота — концентрации;
  - в) линейной диаграммой, где по абсциссе — выраженность симптома, а по ординате — концентрация;
  - г) диаграммой рассеяния, где по абсциссе — выраженность симптома, а по ординате — концентрация.



## Глава 2

# ВЫЯВЛЕНИЕ БОЛЕЗНЕЙ И ИХ ПРИЧИН

Аспект эпидемиологии, относящийся к выявлению причин и закономерностей медицинских проблем, называют *аналитической эпидемиологией* [9]. Ключевые структуры исследований в аналитической эпидемиологии — когортные и поперечные исследования и исследования типа сравнения с контролем (рассмотрены в разд. 2.7 и 2.8). В таких исследованиях изучаются связи медицинских проблем с экспозициями — возможными причинами этих проблем. На первом месте в любом исследовании стоит проблема оценки состояния людей и выявления исходов — проблема диагностики.

## 2.1. КЛАССИФИКАЦИЯ БОЛЕЗНЕЙ И ДИАГНОЗ

### 2.1.1. Определение болезни

Страдания людей многообразны, но в просторечии страдание может быть синонимом болезни. *Медицина как наука, профессия и социальная практика* помогает людям в страданиях, связанных с болезнями. Медицина опирается на представления о том, что болезни имеют материальный субстрат, пусть во многих случаях пока неизвестный. Поэтому иногда болезнь может предстать как суборганизмальный процесс (например, изменения функции и структуры печени или наличие в крови вирусных частиц). Соответственно нередко упрощенно болезнь приравнивается к наличию объективно определяемых признаков, отличающих больного человека от здоровых (*объективистская концепция болезни*). С другой точки зрения, болезнь — субъективный феномен, и болен прежде всего тот, кто страдает (*субъективистская концепция*



болезни). Этот подход также ограничен, так как и больной без сознания, и больной начальной бессимптомной формой рака являются больными.

Границы болезней не всегда надежны, и число болезней, которые известны, непрерывно увеличивается от дюжины во времена Гиппократов до многих сотен сегодня. Еще 200 лет назад врачи полагали, что может быть создана надежная и исчерпывающая классификация болезней. Т. Сиденгам (Т. Sydenham, 1624–1689) полагал, что такая классификация может быть похожей на классификацию растений или насекомых. Подобные классификации болезней создавались неоднократно в XVIII в., в том числе Карлом Линнеем (Carolus Linnaeus, 1707–1778). В таком понимании болезни существуют в природе независимо от человека и должны лишь быть открыты (*эссенциалистическая концепция болезни*). Соответственно, здоровье существует само по себе, безотносительно болезней, и болезни отдельны друг от друга, их разнообразие может быть описано как номинативный признак. Тогда такие болезни — внешние по отношению к здоровому человеку — могут быть устранены.

Нередко процесс взаимодействия человеческого организма с болезнетворным агентом («внешней» болезнью) очеловечивается (*антропоморфизм*), и для его обозначения используется выражение «борьба организма с инфекцией (болезнью)», а изменения в организме классифицируются на «вредные» и «полезные» для исхода болезни (например, повышенная температура тела рассматривается как полезная для выздоровления при гриппе). Такие подходы характерны для примитивного, бытового понимания процессов заболевания и выздоровления, но иногда встречаются и в медицине.

В противоположном понимании болезни есть проявления адаптации (приспособления) человека к среде — природной или социальной — и могут быть устранены лишь постольку, поскольку возможно изменить среду и человека. В этом понимании то, что врачи наблюдают как болезнь, есть лишь проявление приспособления организма человека к среде (температуре, химическому веществу, недостатку питательного вещества) или взаимного приспособления организма и элемента среды — микроорганизма [2, 8]. Для медицины болезнь — это общее свойство сходных пациентов. Этот принцип отражает несхожесть проявления болезни у разных людей, и необходимость классифицировать болезни по доступным в практике признакам (*номиналистическая концепция болезни*). В этом отношении медицинский подход интегрирован с эпидемиологическим. Эпидемиология изучает распространение



болезней (заболеваемость) и признаков, связанных с болезнями, в популяции, и методы эпидемиологии лежат в основе медицинского понимания болезни.

Представления о болезнях зависят от культуры. Что является болезнью в одной цивилизации, может быть хромосомной аномалией, преступлением, признаком святости или грехом в других цивилизациях. Социальное сопротивление диссидентов рассматривалось в тоталитарных обществах как признак нарушенной социальной адаптации, склонность рабов к бегству — как болезнь драпетомания, болезнью считалась гомосексуальность. Теперь последняя не считается болезнью, но гиперактивность и нарушения внимания отдельных детей стали рассматриваться как болезнь. То, что признается болезнью, попадает в сферу медицины. Врачи получают право такие состояния лечить, изменять рамки социального функционирования человека (признавая его инвалидом) или даже лишать его свободы, помещая в закрытое психиатрическое учреждение. В узком смысле болезнь есть страдание, при котором ищут медицинской помощи. Одновременно в обществе болезнь обладает некоторой привлекательностью, позволяя извинить нетрудоспособность, отсутствие на работе, получить пенсию.

В целом «болезнь вообще» нельзя определить, и ни одно из определений болезни не является операционным (позволяющим выполнить действие — выявить болезнь). Определение болезни как состояния, отличного от здоровья, также невозможно, поскольку общее определение здоровья еще труднее сформулировать. ВОЗ определила здоровье как состояние физического, психического и социального благополучия, включающее (но не ограниченное только этим) отсутствие болезни или физического недостатка. Это определение рисует идеал здоровья, не наблюдаемый в реальности, и такое состояние человека невозможно выявить существующими инструментами. Поскольку медицина уполномочивается обществом действовать в отношении того, что является болезнью, принципиально важно болезни определить и классифицировать.

### 2.1.2. Международная классификация болезней

Вершиной усилий по классификации болезней является создание и периодическое обновление Международной классификации болезней (МКБ). МКБ развивается на основе классификации Ж. Бертильона



(Berthillon J., 1893). В МКБ главенствуют исторически оправдавшие себя два основных принципа в выделении номенклатуры болезней: клинический (субстратный) и ассоциативный (этиологический). Использование субстратного принципа группирует болезни по отношению к пораженному органу (системе органов). Этиологический группирует болезни по признаку общего происхождения. Наряду с травмами и отравлениями происхождение инфекционных болезней обычно понятно, и по отношению к ним чаще применяется этиологический принцип. Группировка первого класса (некоторые инфекционные и паразитарные болезни) МКБ определяется преимущественно классификацией возбудителей: бактериальные (4 группы), вирусные (6 групп), микозы, протозоозы, гельминтозы, инфестации. Часть инфекционных болезней не включена в этот класс. Например, грипп по традиции включен в класс болезней органов дыхания.

Этиология болезни известна не всегда, и последовательное применение этиологического принципа классификации затруднительно. Преобладающая субстратная группировка болезней практически оправдана, поскольку позволяет успешно интегрировать клинические сведения с результатами фундаментальных исследований субстрата (изучаемых органов) и с эпидемиологией. В результате рубрики действующей МКБ-10 (10-й пересмотр, 1993) содержат болезни, отнесенные туда на основании согласованных критериев, а не систематического применения одного принципа. Каждый пересмотр МКБ сопровождается заметными изменениями в ней, но вместе с тем существует и преемственность структуры МКБ. Иногда МКБ пополняется новыми рубриками оперативно, как это произошло в 2020 г. с целью унифицировать учет заболеваемости и смертности от новой коронавирусной инфекции (COVID-19).

МКБ не является естественно-научной классификацией, а представляет собой рабочий документ, который стандартизует классификацию болезней и позволяет сопоставлять заболеваемость в районах страны, результаты деятельности отдельных стационаров и медицинских служб в разных странах. Это делает сведения о заболеваемости и медицинской деятельности пригодными для анализа в разных масштабах — от двух районных больниц до международных сопоставлений. Соответственно, МКБ является важным инструментом любого медицинского исследования. Применительно к отдельным классам болезней, отдельным группам болезней разрабатываются более детальные классификации,



включающие свои диагностические критерии. Обычно такие классификации разрабатываются международными или национальными медицинскими обществами в интересах клиники, а не статистического учета. Они обновляются оперативнее, чем МКБ, и являются источником для последующих ее изменений.

### 2.1.3. Виды диагнозов

Среди подходов к формированию диагностической категории самым древним является *симптоматический*. Примеры — хронический понос, нейросенсорная тугоухость. *Анатомически определенными* являются болезни, ключевые особенности которых, используемые в диагностике, выражаются в анатомическом названии: инфаркт миокарда, перелом шейки бедра, рак легкого. Анатомические данные могут быть доступны при диагностике (биопсия образования в прямой кишке) или недоступны (больной инфарктом миокарда выжил). *Физиологически (метаболически) определенные* болезни включают в себя гипертензию, гипертиреоз, диабет и многие подобные болезни. Диагноз их опирается на совокупность данных о специфических нарушениях в метаболизме, но почти всегда включает в большей или меньшей степени другие клинические данные, которые интегрируются в пакет диагностических критериев. *Этиологически определенные* болезни включают в себя инфекционные болезни, болезни, вызванные внешними воздействиями химической или физической природы (отравление свинцом, асбестоз, лучевая болезнь), и врожденные болезни.

Более сложным является подход к диагнозу, называемый *синдромным*. Так, базедова болезнь была описана как триада симптомов (рис. 2.1), наиболее постоянно встречающихся у больных базедовой болезнью. Позднее этот диагноз заменился на гипертиреоз, но это произошло уже на основании знаний о метаболизме гормонов.

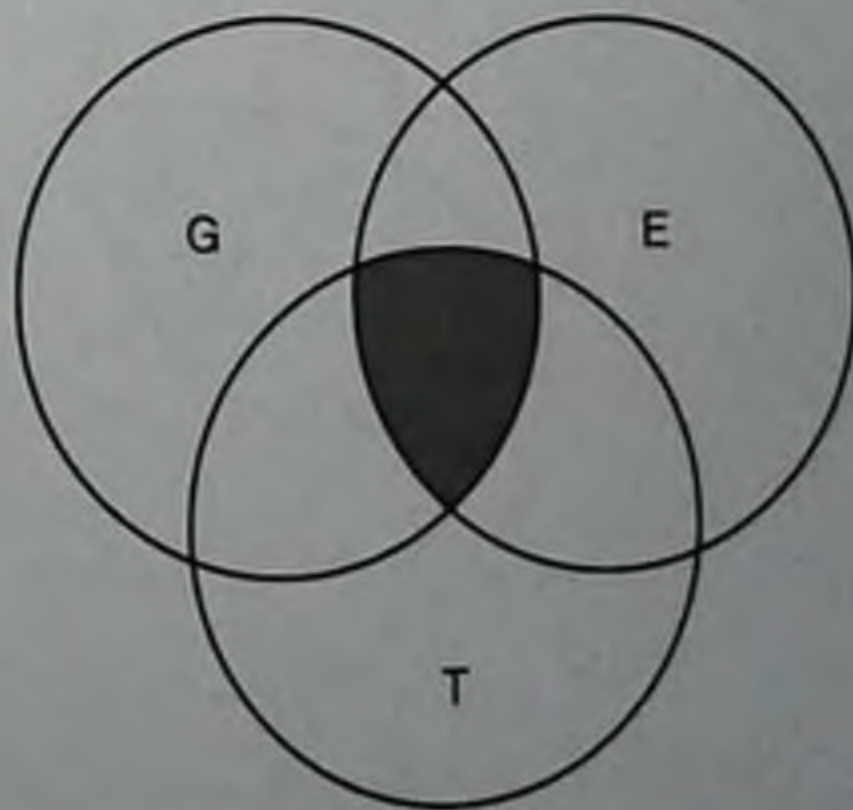


Рис. 2.1. Мерсебургская триада: среди пациентов с зобом (G), экзофтальмом (E) и тахикардией (T) есть группа имеющих все три симптома



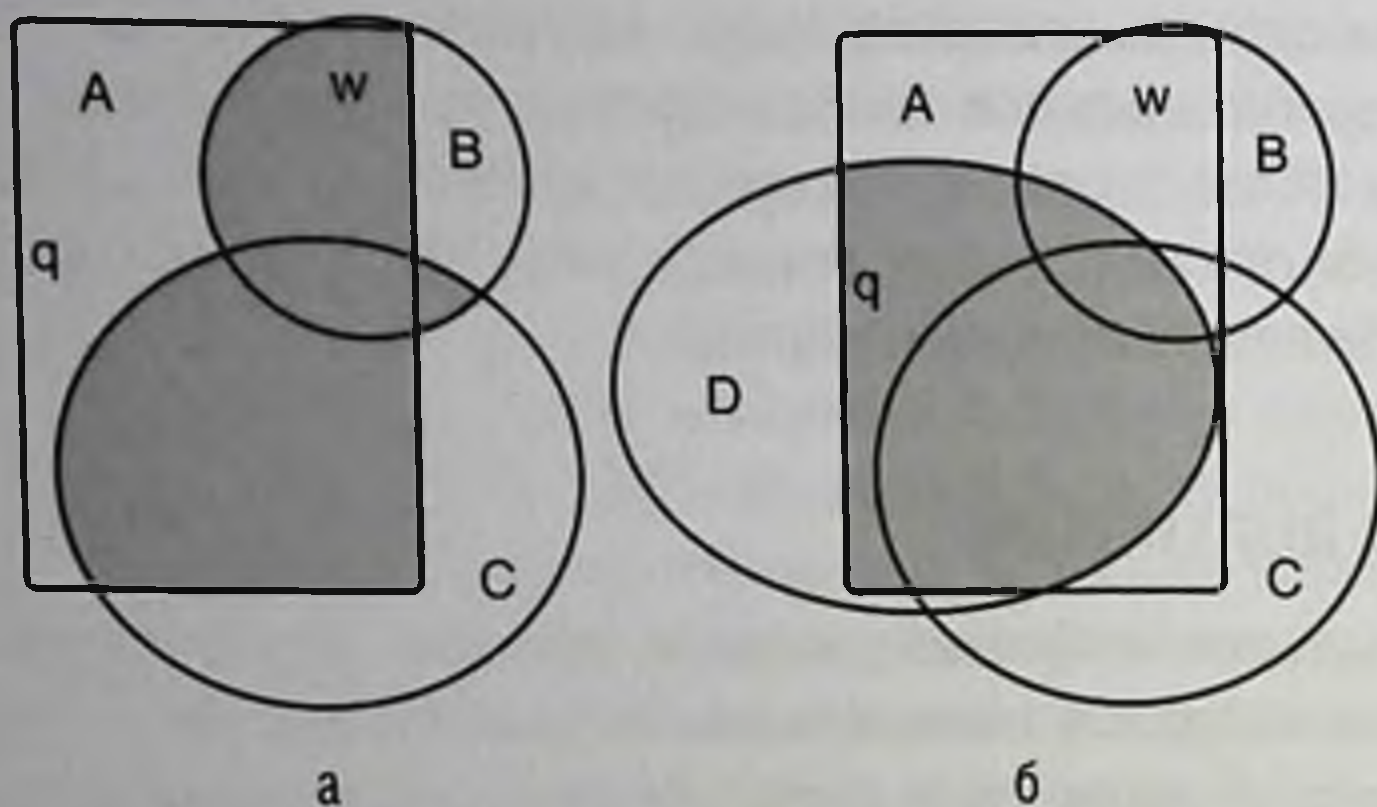


Рис. 2.2. Изменение спектра пациентов с введением нового диагностического теста. Использование старого пакета тестов ABC приводило к установлению диагноза пациентам с сочетанием признаков, заштрихованных на рисунке слева (а). Введение теста D изменило ситуацию: часть больных, ранее соответствовавших критериям диагностики, теперь им не соответствует (w), другие, ранее не получавшие диагноза, теперь его получают (q) (б)



Рис. 2.3. Расхождения в диагностике шизофрении у 72 первичных пациентов (не лечившихся и не обследовавшихся ранее). Из выборки исключены лица, принудительно госпитализированные, с тяжелыми психотическими расстройствами, явными аффективными и органическими расстройствами и классифицированные по всем критериям как здоровые (по данным Jansson L. et al. World Psychiatry. 2002. № 1 (2). P. 109–14)



Для многих болезней в настоящее время существуют более сложные критерии диагностики, требующие, например, для подтверждения диагноза наличие двух «больших» признаков или одного «большого» и двух «малых». Такие диагностические критерии проходят многостороннюю проверку, одобряются международными профессиональными организациями и становятся фактическим стандартом.

Обновление диагностических критериев приводит к изменению спектра симптомов, которые признаются характерными для больных данной болезнью (рис. 2.2). На рис. 2.3 приведены результаты применения четырех современных диагностических критериев шизофрении к однородной группе пациентов. Только 15 из 72 человек классифицированы как больные шизофренией по всем 4 критериям (см. также пример изменения критериев диагностики сахарного диабета во врезке на с. 173).

Синдромные диагнозы являются менее совершенными, чем анатомически определенные, физиологически (метаболически) и этиологически определенные болезни. Точность диагностики болезней, выделяемые на основании синдрома, обычно ниже, и границы синдрома исторически более подвижны.

## 2.2. КЛИНИЧЕСКИЙ И ЭПИДЕМИОЛОГИЧЕСКИЙ ДИАГНОЗ

Клиническая диагностика — это наука и практика выявления и различения болезней и неболезненных состояний (например, беременности). Диагностическая деятельность направлена на индивидуума. Эпидемиологическая диагностика — это анализ материалов текущей статистики в целях изучения заболеваемости. Она нацелена на выявление изменений в преваленсе и инциденсе болезней, состояний и признаков (факторов риска). Выявление изменений инциденса и преваленса может быть основанием как для популяционных вмешательств (например, хлорирование воды, программы против курения), так и для изменений во врачебной практике (срочная вакцинация или антибиотикопрофилактика, диагностика с учетом эпидемиологических данных).

В основе выявления болезней лежит обнаружение людьми у себя необычных признаков. Эти признаки называют *симптомами*. Боль и головокружение, голоса невидимых людей и зуд — все может привести человека к врачу. Симптомы делят на *субъективные* (головокружение, зуд) и *объективные* (кровь в кале, сыпь), выявляемые больным, их род-



ственниками и врачами. Обычно пациенты и их родственники не используют профессиональных медицинских приемов изучения симптомов, которые применяют врачи. Сведения, получаемые врачом при исследовании пациента с использованием традиционных средств (плессиметра, фонендоскопа, отоскопа), называют *физикальными*, или *клиническими данными*, а полученные с помощью более сложных приборов — *параклиническими* (инструментальные, в том числе лабораторные). Симптом, обусловивший обращение человека к врачу, называют *ятротропным*. Он может быть не самым важным для диагностики,

но он приводит человека к решению о необходимости обратиться к врачу и превращает его в *пациента*<sup>1</sup>. Не всегда пришедший к врачу за помощью человек страдает от болезни, т.е. он не всегда *больной*.

Многие болезни обладают свойством самопроизвольно излечиваться, нередко больные знают, как лечить свою болезнь, и лечат себя сами более или менее успешно. Наконец, больные могут обращаться за помощью к альтернативным целителям (хилерам). Поэтому такие больные не становятся пациентами врача никогда или становятся ими позднее. Множество факторов способно повлиять на решение обратиться к врачу. Одни ускоряют обращение, например доступность врача и неприятные условия на производстве, другие — отсрочивают, например социально осуждаемый характер болезни и страх услышать опасный диагноз. В результате процесс развития болезни в том виде, в котором врачи его наблюдают, в значительной мере зависит от социальных и культурных факторов (рис. 2.4).

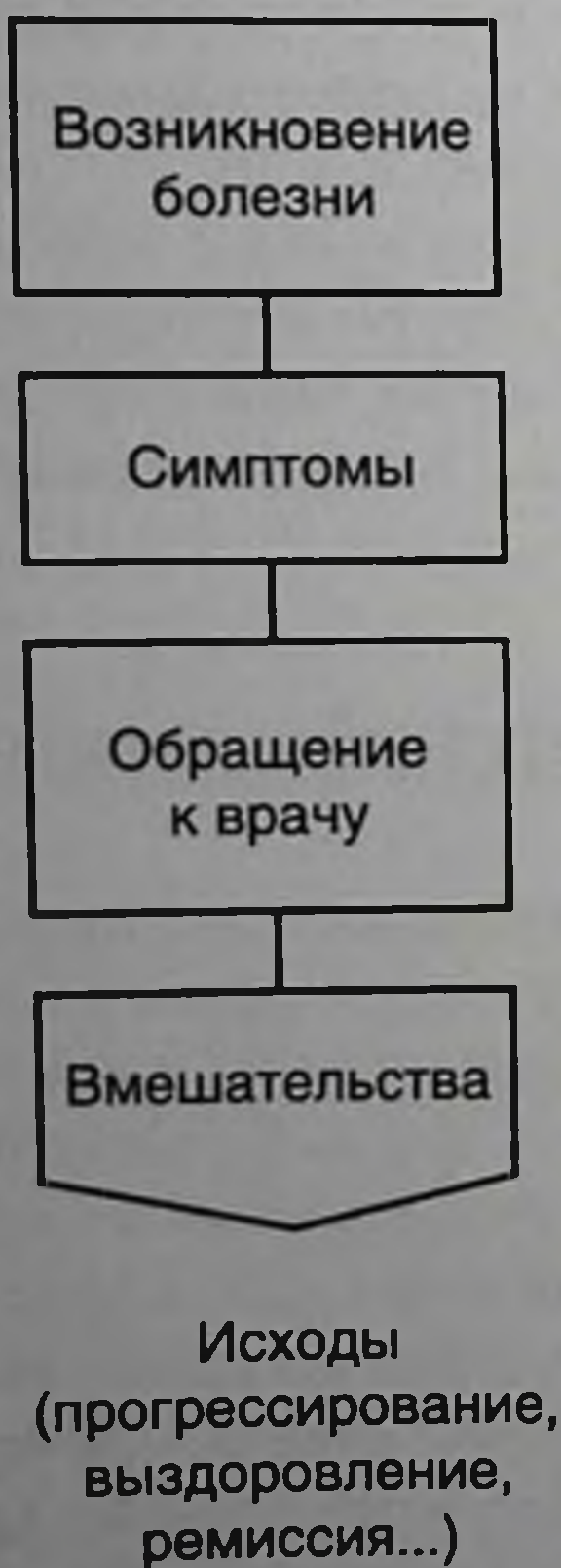


Рис. 2.4. Общая схема развития болезни во времени

<sup>1</sup> Поскольку традиционный термин «пациент» имеет оттенок смысла «подчиненный, исполняющий, терпящий» (от лат. *patiens* — страдающий), постольку существует тенденция заменять его на «пользователь», «потребитель» медицинской помощи.



Именно поэтому среди впервые пришедших на прием больных оказываются больные с терминальной стадией рака, погибающие через неделю. Равным образом среди пациентов всегда присутствуют те, кто, по мнению врача, мог бы не обращаться за помощью ввиду объективно «незначительных» расстройств.

При обращении пациента врач выслушивает «жалобы», т.е. описание симптомов и их развитие, и обнаруживает симптомы, которые пациент не заметил или счел неважными (например, сыпь на ладонях). Врач применяет специальные методы, более или менее сложные, для выявления хрипов при дыхании, аномальной пульсации вен на шее и многих других признаков, наличие которых сочтет важным проверить. Практика диагностики, т.е. набор и порядок использования диагностических тестов, клинических и параклинических, изменяется со временем. Диагностика нейроциркуляторной дистонии, которой обучали студентов середины XX в., утратила смысл, поскольку исчезла эта нозологическая единица. Клиническая дифференциация гриппа и других респираторных вирусных инфекций, которую в этот период требовали от врачей первичного звена, ушла в прошлое. Одновременно появились небывалые возможности точного бактериологического и вирусологического диагноза, портативные ультразвуковые сканеры.

Существует объективная тенденция к замещению клинического диагноза параклиническим (морфологическим, микробиологическим). Выявление чумной палочки означает диагноз чумы, а выявление злокачественно трансформированных клеток в биоптате образования в бронхе — рак. Это не означает безошибочности таких заключений. Например, клетки с признаками малигнизации могут быть обнаружены в новообразовании воспалительного происхождения. Движение к преимущественно параклиническому диагнозу отражает лишь его большую точность и развитие специализации в медицине.

Клинический диагноз, диагноз специалиста и морфологический диагноз не одинаковы и не могут быть одинаковы ни по клинической ситуации, ни по способу практики. Так, для врача общей практики «гастрит» — это диспепсия с дискомфортом в эпигастрии, но без определенной эндоскопической картины. Для гастроэнтеролога гастрит соответствует определенным эндоскопическим проявлениям. Для морфолога гастрит — совокупность макро- и микропроявлений, которые он может охарактеризовать как воспалительные, атрофические, ги-



пертрофические. Так же неоднозначна ситуация с определением хронического характера процесса. В клинической практике хронический процесс — это процесс, длящийся дольше, например, 6 мес. Для патологоанатома хронический — это сопровождающийся специфической клеточной инфильтрацией, и иногда хронический процесс на микропрепарате проявляется признаками, свойственными острому (например, язвенный колит).

### 2.2.1. Референтный диагноз

Каким бы ни был по существу диагностический признак, его ценность определяется на основе простых принципов. Главным принципом является сравнение любых диагностических признаков с *референтным диагнозом*. Референтным называют диагноз, установленный с помощью наиболее надежного метода. Например, несмотря на развитие неинвазивных методов исследования, основанных на различных принципах (ультразвуковые, радиосцинтиграфические, магнитно-резонансные), наилучшим образом степень сужения коронарных артерий оценивается с помощью рентгеноконтрастной коронарографии. Любой из перечисленных методов или новый метод имеют для врача ценность постольку, поскольку результаты их применения соответствуют золотому стандарту коронарографии. Референтный тест — не идеальный, но обычно лучший из имеющихся.

Исторически роль референтного диагноза играли патологоанатомические данные или данные прижизненного морфологического исследования. Они сохраняют свое значение, но только в комплексе с такими новыми методами, как гистохимические методы и полимеразная цепная реакция, позволяющая с почти абсолютной точностью идентифицировать микроорганизмы. Вместе с тем морфологические исследования (макро- и микроскопические) имеют не очень высокую воспроизводимость. Например, при оценке микроскопических препаратов морфологи расходятся во мнениях относительно 30% препаратов. При повторной оценке тех же препаратов до 10% получают иное заключение у того же специалиста. Возможность ошибки в заключении параклинического специалиста не может исключаться, так же как возможность ошибки врача.

В ситуации, когда надежного референтного теста не существует, применяются «рабочие» диагностические критерии болезни (состоя-



ния). Например, для большинства психических болезней созданы такие стандартные диагностические критерии, которые используются для клинической диагностики, эпидемиологических исследований и для разработки новых методов диагностики.

Для многих болезней не существует референтного теста, позволяющего *раннюю диагностику*. В таком случае на первое место выходит наблюдение за больными до того момента, когда существующий референтный метод позволит установить или отвергнуть диагноз. Результаты нового диагностического теста могут быть сопоставлены с результатами, полученными позднее, после достаточно длительного наблюдения. Так до сего времени обстоит дело с миеломной болезнью, рассеянным склерозом, другими медленно развивающимися болезнями с неясной этиологией. Метод выжидания момента формирования ясной клинической картины, позволяющей осуществить диагностику, является традиционным решением в медицине.

Существуют состояния, для которых отсутствует референтный тест, например фибромиалгия, синдром хронической усталости. Такие состояния определяют через совокупность симптомов (синдром), наиболее часто встречающихся при этих состояниях. Разрабатываются критерии *синдромной диагностики* (см. разд. 2.1.3), которые используются на основании их воспроизводимости (см. с. 41). Этот способ диагностики не является надежным, а сами такие диагнозы обычно не рассматриваются как устойчивые самостоятельные нозологические единицы. Описания таких синдромов со временем претерпевают изменения, и нередко синдромы существенно изменяются и попадают в международную классификацию под другими именами, с другими референтными признаками и в другие рубрики. Примерами являются нейроциркуляторные дистонии (теперь паническая атака), синдром концентрационного лагеря (теперь посттравматическое стрессовое расстройство).

### 2.2.2. Оценка согласия

Известно, что врачи при оценке рентгенограммы, аускультативной картины или электрокардиограммы могут давать несовпадающие заключения. Например, при оценке электрокардиограммы (ЭКГ) несколько опытных кардиологов могут найти признаки ишемической болезни сердца у 20% пациентов (один или более врачей), но *все* врачи



найдут такие признаки только у небольшой части пациентов, например у 5%.

Простейшим способом оценки согласия является вычисление *общего согласия*. Например, если из 100 оценок 2 врача согласились в «патологии» в 10 случаях и в «норме» в 70 случаях, то общее согласие составляет  $10 + 70 = 80$  (табл. 2.1, А). Общее согласие — плохой индекс согласия. Он не учитывает, что заключения специалистов могут совпадать случайным образом.

Таблица 2.1. Согласие двух наблюдателей: А — общие данные о согласии; Б — вероятности случайного согласия

А		Врач 1		
		Патология	Норма	Всего
Врач 2	Патология	10	11	21
	Норма	9	70	79
	Всего	19	81	100
Б		Врач 1		
		Патология	Норма	Всего
Врач 2	Патология	4	17	21
	Норма	15	64	79
	Всего	19	81	100

Представим теперь, что второй врач не стал утруждать себя оценкой ЭКГ, а зная, что нормальные ЭКГ преобладают, просто все ЭКГ классифицировал как нормальные. Очевидно, что его согласие с первым специалистом будет высоким — 81%. Представим более сложный эксперимент. Два врача согласились случайным образом дать заключения, так, чтобы частота «патологии» составила 20%. Каждый использовал таблицу случайных чисел, и число «патологических» случаев было не точно 20 из 100, а у одного — 19, а у другого — 21 (табл. 2.1, Б). Случайное совпадение заключений — случайное общее согласие, оказалось равным 68%. Получается, что только случайным образом заключения



согласуются на 68%. После этого общее согласие 80% не выглядит впечатляющим. Из этого примера следует, что согласие врачей между собой не обязательно означает, что они используют эффективные методы диагностики. Согласие может быть высоким даже случайно.

Цифры в ячейках таблицы — вероятности случайного совпадения — можно получить исходя из определения независимых событий (см. разд. 1.3.3): вероятность совпадения независимых событий равна произведению вероятностей этих событий. Иными словами, если врач 1 дал 21/100 заключений «патология», а врач 2 — 19/100, то вероятность случайного совпадения составляет  $21/100 \times 19/100 = 0,04$ . В нашем случае, когда оценивались 100 случаев, 0,04 соответствует 4 пациентам. Точно так же с использованием *краевых сумм* мы можем вычислить вероятность случайного совпадения для трех других ячеек табл. 2.1, Б.

Оценка согласия должна основываться на том, насколько наблюдаемое согласие лучше, чем случайное. Если случайное согласие в отношении наличия «патологии» ожидается в 4 случаях, а фактическое составило 10 случаев, то прирост — 6 случаев. Таково же приращение согласия в отношении «нормы»:  $70 - 64 = 6$ . По сравнению с полным несогласием случайное согласие составляет почти 70%, наблюдаемое — 80%, при полном согласии — 100% (рис. 2.5). Превышение наблюдаемого согласия над случайным описывается индексом  $k$  (каппа — греческая буква), вычисляемым как:

$$k = (p_o - p_e) / (1 - p_e).$$

В нашем примере  $k = (80 - 68) / (100 - 68) = 12 / 32 = 0,375$ .

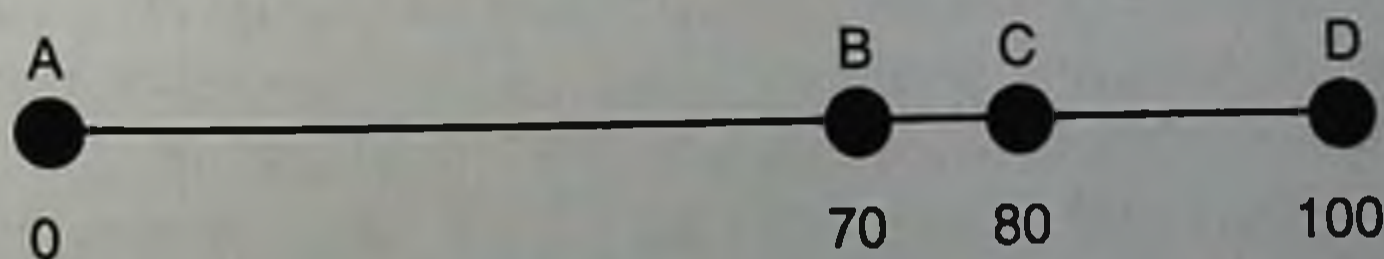


Рис. 2.5. Индекс согласия  $k$  данным табл. 2.1: А — отсутствие согласия, В — случайное согласие, С — наблюдаемое согласие, D — полное согласие

Каппа изменяется от 1 (полное согласие) до 0 (только случайное согласие). Отрицательные величины возможны, но лишены смысла, так как отражают систематическое разногласие между экспертами (мето-



дами диагностики), когда случаи, оцениваемые одним положительно, другим оцениваются отрицательно. Каппа не является единственным или самым лучшим индексом согласия [10], но полезен для оценки воспроизводимости оценок разного рода. Во множестве исследований найдено, что многие методы, представляющиеся врачам убедительными и надежными, имеют невысокое согласие между врачами. Так, для диагноза «язва желудка» при эндоскопии  $k = 0,59$ .

### 2.2.3. Сопоставление с референтным диагнозом

По сравнению с референтным тестом другие используемые или разрабатываемые методы обычно несут возможность прижизненной, более ранней, более приемлемой, более дешевой диагностики. В диагностике изучаемый тест и референтный тест не равны, поэтому изучается не согласие между ними, а совпадение результатов изучаемого теста с результатами референтного. Более того, совпадение положительных результатов двух тестов может иметь принципиально иное значение, чем совпадение отрицательных результатов. Поэтому общее согласие результатов не используется для оценки диагностических тестов. Традиционным методом оценки диагностического теста в сравнении с референтным диагнозом является построение латинского квадрата (четырёхпольной таблицы) (табл. 2.2).

По приведенным в табл. 2.2 результатам испытания можно вычислить так называемые *операционные характеристики теста* чувствительность и специфичность, а также производные от них вероятности (прогностичности результатов) (табл. 2.3).

*Чувствительность* (англ. *sensitivity*, *Se*) — это доля больных, которые выявляются диагностическим методом как больные. Чувствительный тест при исследовании больных дает много истинно положительных (TP) и мало ложноотрицательных (FN) результатов.

*Специфичность* (англ. *specificity*, *Sp*) — это доля здоровых, которые определяются диагностическим методом как здоровые. Специфичный симптом редко выявляется у лиц без данной болезни. Такие признаки (симптомы) называют *патогномоничными*. Например, кальцификация диафрагмы, выявляемая рентгенологически, патогномонична для поражения человека асбестовой пылью.

Чувствительность и специфичность называют *операционными характеристиками теста*. Операционные характеристики теста оценива-



ются в диагностическом эксперименте. В него включают лиц, диагноз которых считается необходимым уточнить с помощью изучаемого теста. Всех включенных в исследование лиц подвергают как изучаемому тесту, так и референтному тесту. Затем результаты сопоставляют с построением латинского квадрата. В процессе проведения эксперимента специалисты, проводящие тесты, не должны знать результатов применения других тестов (оценка данных вслепую). Иначе возникнет смещение в сторону повышенного согласия исследуемого и референтного тестов («смещение к предполагаемому диагнозу»).

Таблица 2.2. Данные проверки правильности диагностики болезней почек по радужке глаза опытными иридодиагностами и офтальмологами

Изучаемый метод иридодиагностика	Референтный диагноз		
	Здоровые	Болезни почек	Всего
Здоровы [Negative]	36 (TN) A	19 B (FN)	A + B = 55
Болезнь почек [Positive]	(FP) C 59	D (TP) 29	C + D = 88
Всего	A + C = 95	B + D = 48	143

Изучаемый метод — оценка офтальмолога	Референтный диагноз		
	Здоровые	Болезни почек	Всего
Здоровы [Negative]	59	27	86
Болезнь почек [Positive]	36	21	57
Всего	95	48	143

Примечание. TN — истинно отрицательные; FN — ложно отрицательные; FP — ложно положительные; TP — истинно положительные. Диагноз «болезнь почек» клинически подтвержден, и тяжесть болезни подтверждается нарушением выделительной функции (концентрация креатинина в крови выше 1,5 мг/100 мл) — надежный референтный метод. По данным Simon A. et al. J.A.M.A. 1979. № 242 (13). P. 1385–1389.



Таблица 2.3. Операционные характеристики теста на определение болезни почек по радужке глаза

Характеристика	Иридодиагносты, %	Офтальмологи, %
Чувствительность (Se) = $D/(B + D)$	61	44
Специфичность (Sp) = $A/(A + C)$	38	62
Прогностичность положительного результата (PVP) = $D/(C + D)$	33	37
Прогностичность отрицательного результата (PVN) = $A/(A + B)$	65	69

Примечание. Пояснения в тексте.

Специфичность теста можно также оценить при обследовании заведомо здоровых людей, а чувствительность — при обследовании людей, диагноз у которых установлен референтным методом. Этот упрощенный способ дает менее надежную оценку операционных характеристик теста (обычно завышенную), поскольку обе группы формируются разными способами и в процессе формирования могут приобретать непредсказуемые особенности.

Для операционных характеристик по результатам эксперимента вычисляется ДИ, как это описано в разд. 1.4.4 для относительных величин. Если в диагностическом эксперименте обследовано мало людей, то оценка операционных характеристик неточна (неопределенна).

Для диагностических тестов не существует минимально необходимой величины чувствительности или специфичности. Тест, дающий положительный результат у больных чаще, чем у здоровых, может быть полезен. Принятие решения об использовании теста (признака) зависит от многих факторов, лежащих за пределами оценки собственно диагностической эффективности теста. Например, при ожидании завоза опасной инфекционной болезни, например, «тяжелого острого респираторного синдрома» (SARS), распространявшегося по миру зимой 2002 г., даже неспецифические проявления инфекционных болезней, диагностически обычно неэффективные, могут использоваться для принятия мер ограничительного характера.

В медицинской практике, *если у теста высокая чувствительность*, то по его отрицательному результату можно надежно исключить подо-



зреваемую болезнь. При использовании такого теста мала вероятность пропустить больных. Ошибочно считать, что чувствительный тест хорош для включения болезни в дальнейшую диагностическую проработку, для выявления больных.

*Если у теста высокая специфичность*, то его положительный результат (его почти никогда не бывает у здоровых!) дает основания включить подозреваемую болезнь в дальнейшую дифференциальную диагностику.

Не следует переоценивать эффективность текущей клинической диагностики. Например, по оценке, выполненной в университетской клинике Швейцарии за 1972, 1982 и 1992 гг. точность диагностики по сравнению с комплексным патологоанатомическим исследованием (референтный критерий) была невысокой. При патологоанатомическом исследовании 90% умерших доля существенных расхождений с клиническим диагнозом составила в эти годы соответственно 30, 18, и 14%. Существенным называют расхождение, которое могло изменить исход болезни или существенно изменить применяемые вмешательства. Улучшение диагностики в основном связано с повышением специфичности кардиологических диагнозов (85, 82, 97%) при меньшем повышении их чувствительности (69, 82, 86%) и повышением чувствительности в диагностике инфекционных болезней (25, 67, 86%). В условиях более редкого проведения аутопсий и при меньшей обеспеченности инструментальными средствами диагностики ее результаты хуже приведенных.

#### **2.2.4. Операционные характеристики и преваленс**

Диагностический тест (прием, метод) может быть использован в разных целях.

А. Если исследования проводят у людей, считающих себя здоровыми, с целью выявить тех, кто, не подозревая того, является больным, то такое исследование называется *скринингом* (от англ. *screen* — просеивать). Пример: поголовные флюорографии органов грудной клетки.

Б. Если исследование проводят у всех людей, оказавшихся в больнице или поликлинике в связи с любой болезнью, то это называют поиском сопутствующих заболеваний (англ. *case finding*). Примеры:



общий анализ мочи, флюорография, назначаемые всем госпитализируемым. Поиск сопутствующих болезней близок по многим признакам к скринингу.

В. Если исследование проводят у людей, обратившихся с жалобами для выяснения причины их страданий, уточнения диагноза, то это собственно диагностическое исследование — самый важный случай. Примеры: электрокардиография — при жалобах на сердцебиения, копроскопия, колоноскопия — при жалобах на следы крови на каловых массах.

Г. Если исследование проводят у людей, получающих лечение, для оценки его успешности, то это тест на эффективность лечения (мониторинг). Пример: термометрия у больного пневмонией, получающего антибиотики.

Эта классификация исходит из ситуаций клинической (индивидуальной) диагностики, но она отражает и ситуации эпидемиологической диагностики, нацеленной на собственно диагностику, выявление сопутствующих болезней (факторов риска в популяции) и наблюдение (мониторинг).

Скрининг проводят обычно в условиях очень низкой вероятности обнаружения болезни. В случае диагностического исследования и при мониторинге лечения вероятность диагностической находки (изменения наблюдаемого признака) очень высока. Различия преваленса оказывают решающее влияние на оценку результата теста.

Чувствительность и специфичность называют стабильными характеристиками диагностического теста. Это означает, что они не зависят (или мало зависят) от преваленса. Если чувствительность равна 0,75 (75%), то независимо от того, сколько среди пациентов больных этой болезнью,  $3/4$  из этих больных будут выявлены с помощью теста. Если специфичность 0,9, то среди лиц, не имеющих данной болезни, положительный результат (ложный!) будет получен у 10%.

Чувствительность и специфичность теста не отражают вероятность наличия болезни после выполнения диагностического теста. Врача же в основном интересует, какова *вероятность болезни* у лиц с положительным результатом теста и у лиц с отрицательным результатом. Это называется *прогностичностью* положительных и отрицательных результатов (иногда называют *прогностической ценностью* или *предсказательной ценностью*, от англ. *predictive value*) (см. табл. 2.2). Прогностичность положительного результата обозначается PVP, а от-



рицательного — PVN. Например, если прогностичность положительного результата теста 0,8, это значит, что у больного, результат теста у которого был положительным, вероятность болезни составляет 0,8, или 80%. К получению показателя такого типа направлена вся диагностическая деятельность.

Почему нельзя ограничиваться этими показателями, а надо оценивать диагностический тест чувствительностью и специфичностью? Потому что прогностичность зависит от преваленса. Например, при чувствительности и специфичности 90% (это высокие показатели) тестирование даст: в условиях низкого преваленса (больных 10%) PVP = 50% и PVN = 99%; в условиях высокого преваленса (больных 80%) PVP = 97% и PVN = 69%. Поэтому высокая надежность окончательного диагноза (высокая прогностичность конечного результата) — не доказательство эффективности теста. Возможно, что высокая прогностичность является простым следствием того, что в изученной группе очень высок преваленс, например, вследствие предварительного отбора пациентов. Для оценки диагностического теста надо оценивать чувствительность и специфичность.

Тест, полезный в условиях клинической диагностики, может быть негоден для скрининга (при низком преваленсе) или в условиях очень высокого преваленса. Яркий пример — электрокардиография. Очень полезная в обследовании больных с возможными болезнями сердца, она оказывается малополезной в профилактических обследованиях населения. Нагрузочное ЭКГ-тестирование — эффективный метод для обследования мужчин, предъявляющих жалобы на боли в груди. Однако этот метод дает малопривлекательные результаты, когда применяется при низком преваленсе хронической ишемической болезни сердца — при профилактических обследованиях мужчин и женщин моложе 35 лет. Напротив, у пожилого мужчины, предъявляющего жалобы, соответствующие типичной стенокардии, нагрузочное ЭКГ-тестирование также почти бесполезно — оно не может ни заметно увеличить вероятность хронической ишемической болезни сердца при положительном результате, ни снизить существенно вероятность болезни при отрицательном. Это означает, что нагрузочное ЭКГ-тестирование (впрочем, как и любой другой диагностический тест) максимально полезно в условиях неопределенности диагноза. Низкий или высокий преваленс равнозначен определенности диагноза уже до тестирования.



### Байесовский подход

В XVIII в. монах Томас Байес (Thomas Bayes) разработал принципы анализа процесса познания и результатов опыта, названные его именем как *байезианизм*, или *байесовский подход*. Основой байесовского подхода является признание факта: в реальном мире рассуждения формальной логики не действуют в полной мере, так как наблюдатель не может быть уверен в безошибочности своего наблюдения. Поэтому не только *индукция*, но и *дедукция* не могут быть безошибочны. Байес сформулировал не способ преодоления проблемы неопределенности, но способ действий в условиях неопределенности. Для этого рациональный наблюдатель может оценить свои исходные аргументы (вероятность). Затем можно использовать вероятности, полученные в опыте, для уточнения своих исходных оценок. Иными словами, маловероятное событие, даже подтверждаемое наблюдением, в свою очередь всегда подверженное ошибке, не становится высоковероятным. Для того чтобы маловероятное событие стало высоковероятным, необходимы особо надежные наблюдения (эксперименты).

Байесовский подход подвергался критике, поскольку конечные результаты его применения зависят от того, какие исходные вероятности присваиваются изучаемым событиям. При прямолинейном представлении познания как уточнения знаний о механизмах функционирования физического мира байесовский подход выглядит как подмена знания субъективной уверенностью. Между тем показано, что все люди и все исследователи действуют с трудом в условиях неопределенности и интерпретируют результаты наблюдений и экспериментов, в зависимости от своих исходных представлений. Например, из множества гипотез выбирают для испытания самую правдоподобную.

В этом смысле всякое научное исследование, и эпидемиологическое исследование в том числе, заключается не в установлении точного знания, а в получении интервалов, в пределах которых находится наиболее правдоподобная оценка, с учетом использованного метода эксперимента и методов измерения. В медицинской диагностике байесовский подход используется в форме последовательного уточнения вероятности болезни на основании включения результатов диагностических тестов. Существуют компьютерные диагностические системы, основанные на этом методе [10]. В испытании эффективности терапевтических вмешательств байесовский подход используется в форме последовательного, по ходу испытания, уточнения оценок, сформированных до испытания. В свою очередь, эти *априорные оценки* (до данного опыта, до получения новых знаний) опираются на знания экспертов о патогенезе болезни, о механизмах действия лекарства и о результатах предыдущих испытаний.

Обычный (вероятностный, не-байесовский) статистический анализ экспериментальных данных основывается на испытании гипотезы.



При этом сведения, существовавшие до эксперимента, а также представления о патогенезе, правдоподобность испытываемой гипотезы не принимаются во внимание. Они могут быть лишь позднее включены в осмысление результатов эксперимента.

Байесовский анализ изначально рассматривает результат эксперимента в свете априорных (до опыта) сведений о вероятностях исхода эксперимента. В последние десятилетия байесовские методы все шире используются в анализе биомедицинских данных [11].

Т. Байес разработал принцип статистического анализа результатов эксперимента, учитывающий исходные представления о вероятности события (см. врезку на с. 167). Использование байесовского подхода наиболее распространено при оценке результатов диагностических вмешательств. При этом в качестве исходной вероятности выступает преваленс у пациентов, подобных данному пациенту. Если пациент не обладает очевидными признаками, влияющими на исходную оценку преваленса, то можно использовать сведения о преваленсе в популяции. Например, применительно к новому пациенту с жалобами на боли в груди можно учитывать его пол и возраст.

**Отношение правдоподобия** (LR, от англ. *likelihood ratio*) – отношение вероятности события при некотором условии к вероятности без этого условия, например отношение частоты симптома при наличии болезни к частоте симптома при отсутствии болезни. При качественной оценке теста чувствительность, специфичность и отношение правдоподобия тесно связаны:

$LR_p$  (для положительного результата) =  $Se/(1 - Sp)$ .

$LR_n$  (для отрицательного результата) =  $Sp/(1 - Se)$ .

Обычно  $LR > 10$  или  $LR < 0,1$  дает для диагностики почти окончательное решение.

$LR$  от 5 до 10 и от 0,1 до 0,2 дает умеренные основания для диагностического решения.

$LR$  от 2 до 5 и от 0,5 до 0,2 мало дает для изменения оценки вероятности болезни у пациента.

$LR$  от 0,5 до 2 почти не изменяет вероятности (шансы) болезни у пациента.

Если известна исходная вероятность болезни (преваленс), то можно количественно оценить результаты диагностического теста.



**Первый способ:** воспользоваться показателями чувствительности и специфичности и получить прогностичность положительного (или отрицательного) результата. Например, преваленс  $Pr = 0,3$ , чувствительность теста  $Se = 0,7$ , специфичность  $Sp = 0,9$ . Воспользуемся обозначениями, приведенными в табл. 2.2. Положительный результат приносит следующие вероятности:

$$TP = Pr \times Se = 0,3 \times 0,7 = 0,21;$$

$$FP = (1 - Pr) \times (1 - Sp) = 0,7 \times 0,1 = 0,07.$$

Соответственно, прогностичность положительного результата (апостериорная вероятность)

$$PVP = TP / (TP + FP) = 0,21 / (0,21 + 0,07) = 0,75.$$

**Второй способ:** основан на использовании шансов и отношения правдоподобия (см. врезку на с. 169). Шансы ( $\Omega$ , греч. буква омега) — это отношение вероятности того, что событие произойдет, к вероятности того, что событие не произойдет. Шансы и вероятности содержат одну и ту же информацию, но по разному выражают ее (см. табл. 1.2). Шансы легко переводятся в вероятности, и наоборот, и используются часто в эпидемиологии, так как имеют некоторые преимущества. В диагностике шансы до исследования называют априорными ( $\Omega_a$ ), а после применения теста — апостериорными ( $\Omega_p$ ).

Для использования данных о диагностическом тесте в виде шансов нужно:

- ▶ перевести исходную вероятность в априорные шансы:  $\Omega_a = [Pr / (1 - Pr)]$ . Например, если априорная вероятность болезни у пациента 0,8, т.е. 8 из 10, то априорные шансы 8:2 или 4,0.
- ▶ получить отношение правдоподобия для диагностического теста (из публикации — статьи, руководства). Например, если из 200 больных с кровотечением у 120 тест положителен (чувствительность =  $120/200 = 0,6$ ), а из 400 больных без кровотечения тест положителен у 80 (1-специфичность) =  $80/400 = 0,2$ , то отношение правдоподобия для этого теста диагностики кровотечения  $LR = 0,6/0,2 = 3$ .
- ▶ вычислить апостериорные шансы для пациента, у которого результат теста положительный, умножив исходные шансы на LR:  $\Omega_p = 4 \times 3 = 12$ . Шансы 12:1 — хорошее основание для установления диагноза.

Апостериорные шансы нетрудно перевести в апостериорную вероятность: вероятность =  $\Omega_p / (\Omega_p + 1) = 12/13 = 0,92$ , или 92%.



Использование вычислений, основанных на шансах и отношении правдоподобия, предпочтительнее расчетов на основе чувствительности и специфичности. При использовании инструментов чувствительности и специфичности есть только два «результата» — «положительный» и «отрицательный». Среди «положительных» результатов теряются «резко положительные». Лейкоцитоз, выражающийся концентрацией лейкоцитов  $12 \times 10^9 \text{ л}^{-1}$ , не равен «лейкоцитозу вообще» (концентрации выше  $8 \times 10^9 \text{ л}^{-1}$ ) и для врача несет другую информацию при диагностике «болей в животе». Отношение правдоподобия может вычисляться для каждого «уровня» результата. Поэтому такая оценка теста более продуктивна.

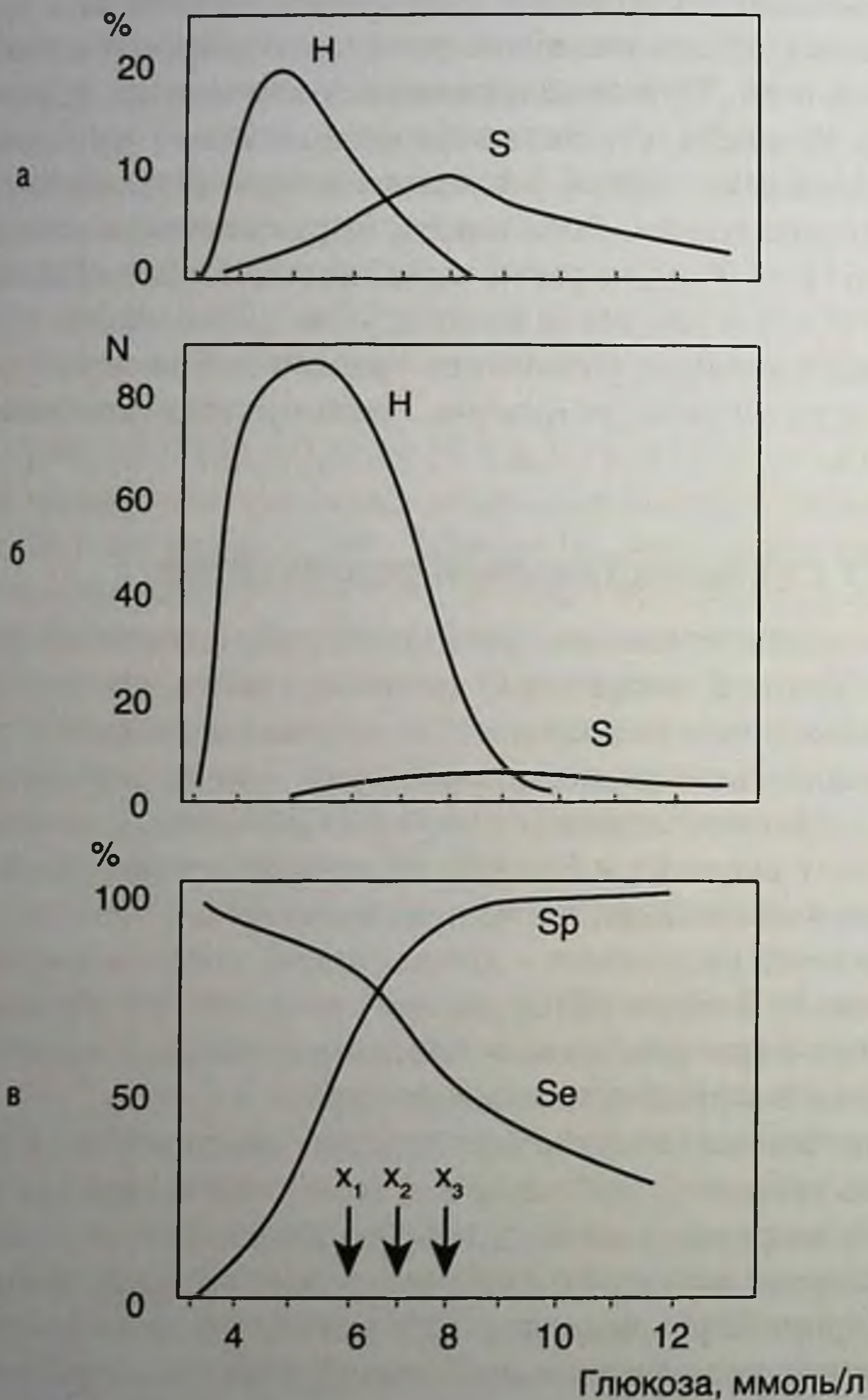
### 2.2.5. Тест с количественным результатом

Многие диагностические тесты дают количественные результаты, например: диаметр индурации, толщина стенки желудочка сердца, концентрация вещества в крови. Для оценки количественного результата теста необходимо установить критерий — точку разделения. *Точкой разделения* называют величину признака организма, с помощью которой разделяют здоровых и больных (годных и негодных, нуждающихся в дальнейшем обследовании и не нуждающихся).

Пример точки разделения — концентрация глюкозы в венозной крови 5,83 ммоль/л. В случае обнаружения натошак более высоких концентраций ранее считалось целесообразным проводить дополнительные исследования (нагрузочную сахарную пробу).

В действительности при концентрации глюкозы более 5,83 ммоль/л вероятность сахарного диабета не возрастает скачкообразно. На рис. 2.6 видно, что у здоровых людей и у лиц, больных диабетом, концентрация глюкозы широко варьирует. Описывая такие распределения, говорят, что они широко перекрываются. При указанной точке разделения ( $x_2$ ) можно оценить чувствительность, специфичность и прогностичность положительного и отрицательного результатов. Точку разделения можно сместить в сторону больших концентраций ( $x_3$ ). Тогда чувствительность снизится, но зато возрастет специфичность и прогностичность положительного результата. Если сдвинуть точку разделения влево ( $x_1$ ), чувствительность повысится, повысится прогностичность отрицательного результата, но снизится специфичность и прогностичность положительного результата.





**Рис. 2.6.** Распределение концентрации глюкозы у здоровых людей (H) и больных сахарным диабетом (S) при исследовании венозной крови натощак: а — распределение обследованных в процентах; б — распределение в абсолютных цифрах в исследованной выборке; в — изменения чувствительности (Se) и специфичности (Sp) при сдвиге точки разделения ( $x_1, x_2, x_3$ ) (по данным Remein Q.R., Wilkerson H.L. J. Chron. Dis. 1961. № 13. P. 6–21)



## Существует ли золотой стандарт диагностики сахарного диабета?

Диагноз «сахарный диабет» основывался на результатах исследования концентрации глюкозы в крови натощак и после глюкозной нагрузки, а клиническая картина в большей части случаев недостаточно специфична.

Как установить точку разделения, если нет золотого стандарта диагноза?

Основным способом является наблюдение за больными (см. разд. 2.2.1). Применительно к диабету в качестве основного признака тяжести и основного исхода, указывающего на прогрессирование болезни, использовали связанные с диабетом сосудистые осложнения — нефропатию и ретинопатию. Оказывается, результаты нагрузочной пробы лишь незначительно лучше, чем концентрация глюкозы натощак или концентрация гликозилированного гемоглобина разделяют людей, у которых уже имеются сосудистые осложнения, и людей, не имеющих этих осложнений. Развитие осложнений в течение 5 лет прогнозируется всеми тремя методами также с равным успехом. Поэтому с 1998 г. диабет диагностируют без использования нагрузочного теста, а лишь на основании концентрации глюкозы натощак более 7 ммоль/л. Концентрацию натощак выше 7 ммоль/л имеет меньше людей, чем концентрацию выше 5,83 ммоль/л, но зато все с концентрацией выше 7 ммоль/л признаются больными диабетом. Часть людей, отвечавших критериям диагноза, основанным на нагрузочном тесте, новым критериям не соответствует. Напротив, часть людей соответствует новым критериям болезни, но не соответствовала старым. Важно, что новый критерий по крайней мере не хуже старого, но экономнее позволяет выявить людей, которым терапия может помочь избежать осложнений.

Картина будет неполной, если не упомянуть о зависимости риска развития нефропатии и ретинопатии от концентрации глюкозы в крови людей, не имеющих диабета, во всем диапазоне концентраций (см. рис. 2.7). Соответственно, выделение из популяции только «больных диабетом» — лиц с очень высокой вероятностью осложнений, подлежащих лечению, может показаться условным решением. Это не так. В действительности эффективность гипогликемической терапии у лиц, соответствующих диагностическим критериям диабета доказана, в том числе по критерию предотвращения осложнений.

Возможность добиться положительных результатов, например снижения вероятности сосудистых поражений путем снижения концентрации глюкозы в крови у здоровых людей (всех людей) не доказана, а потому не может быть рекомендована врачами в качестве индивидуального стиля поведения (питания) или в качестве «оздоровительного» вмешательства в популяции. Иными словами, само определение болезни является в известной степени производным от возможности медицины помочь при данном состоянии. Это вполне типичная ситуация [12].



Зависимость установленных диагностических критериев от возможностей не только референтного теста, но и от возможностей медицинских вмешательств, является фундаментальной закономерностью современной диагностики.

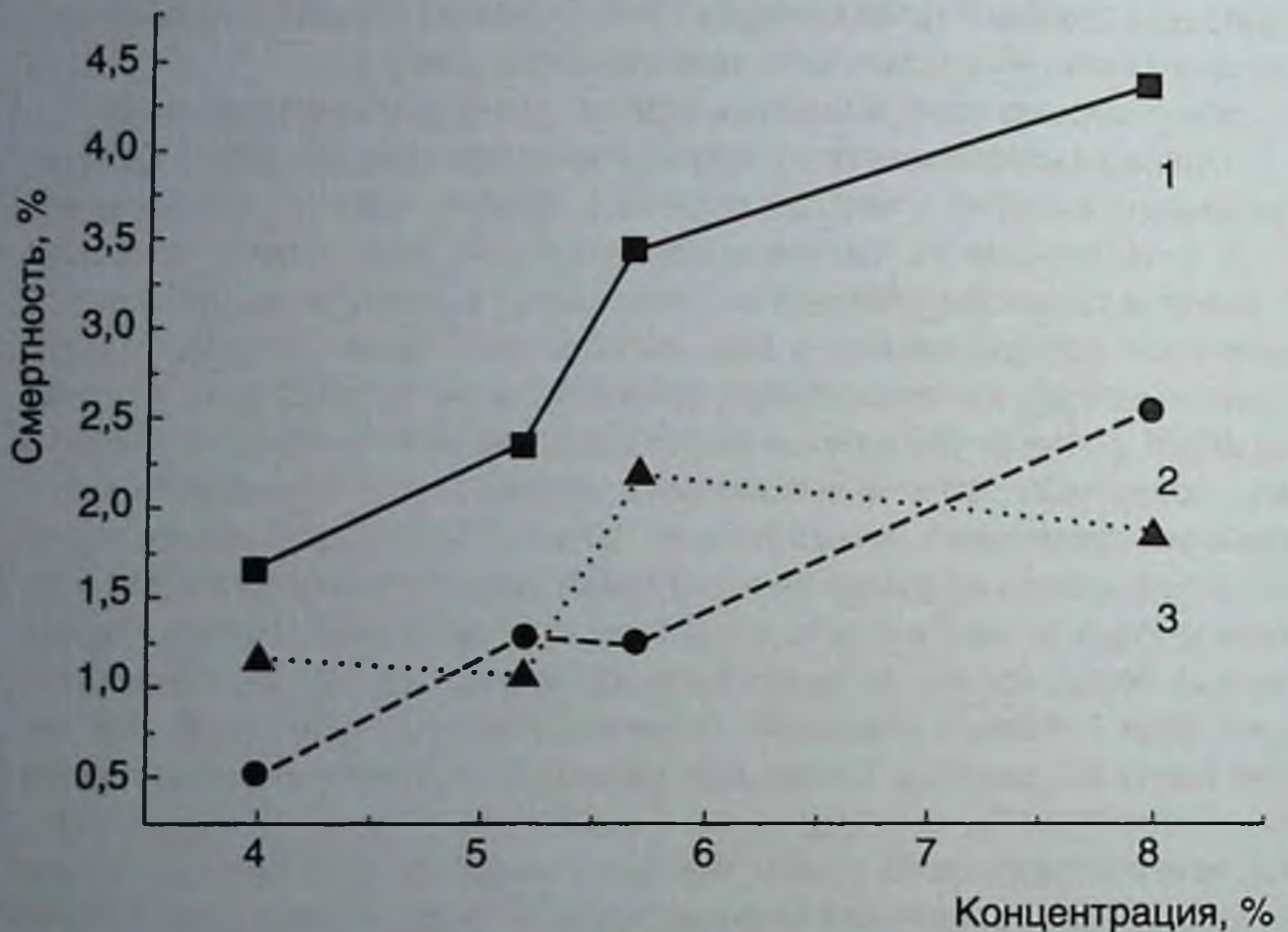


Рис. 2.7. Зависимость смертности (ордината — проценты, с поправкой на возраст, от всех причин (1), от сердечно-сосудистых болезней (2) и от прочих причин (3) у мужчин с разной концентрацией гликозилированного гемоглобина в крови (абсцисса, %). Продольное исследование, Англия, 1995–1999 гг. (по данным Khaw K.T. et al. BMJ. 2001. № 322. P. 1–6)

### 2.2.6. Тест с качественным результатом

Часть тестов, традиционно называемых «качественными», тоже может рассматриваться как полуколичественные (ординальные). Например, при оценке рентгенограммы или гистологического препарата специалист может быть более или менее уверен в своем заключении и без труда сформулирует его в терминах, например «безусловно, патология Р», «вероятно, патология Р», «сомнительно патология Р», «не исключена патология Р», «исключена патология Р». При сопо-



ставлении таких заключений с референтным тестом оказывается, что они соответствуют *вероятности* патологии, определяемой качественно, или *выраженности* патологии, если она определяется количественно (например, степени резорбции костной ткани).

Впервые изменения чувствительности и специфичности теста при изменении точки разделения были изучены применительно к рентгенофлюорографии (рис. 2.8). Было показано, что чем строже подходит врач к оценке рентгенограммы, т.е. чем сильнее он желает не пропустить ни одного случая туберкулеза, тем больше снимков он находит «подозрительными», в то время как в действительности туберкулеза легких нет. Иными словами, чем больше стремление к высокой чувствительности, тем ниже специфичность.

Для того чтобы описать, как изменяются характеристики теста в зависимости от устанавливаемой точки разделения, используют так называемый ROC-анализ (англ. Receiver Operating Characteristic, ROC-Curve). ROC-кривая, или характеристическая кривая, — это кривая зависимости чувствительности от вероятности ложноположительных результатов, т.е. величины  $(1 - \text{специфичность})$ . Например, чем ниже опускается точка разделения по шкале концентрации глюкозы, тем полнее выявляются больные. Одновременно все большее число здоровых людей оценивается как больные (рис. 2.9). Выбор точки разделения не меняет положения кривой. Происходит лишь перемещение точки по кривой.

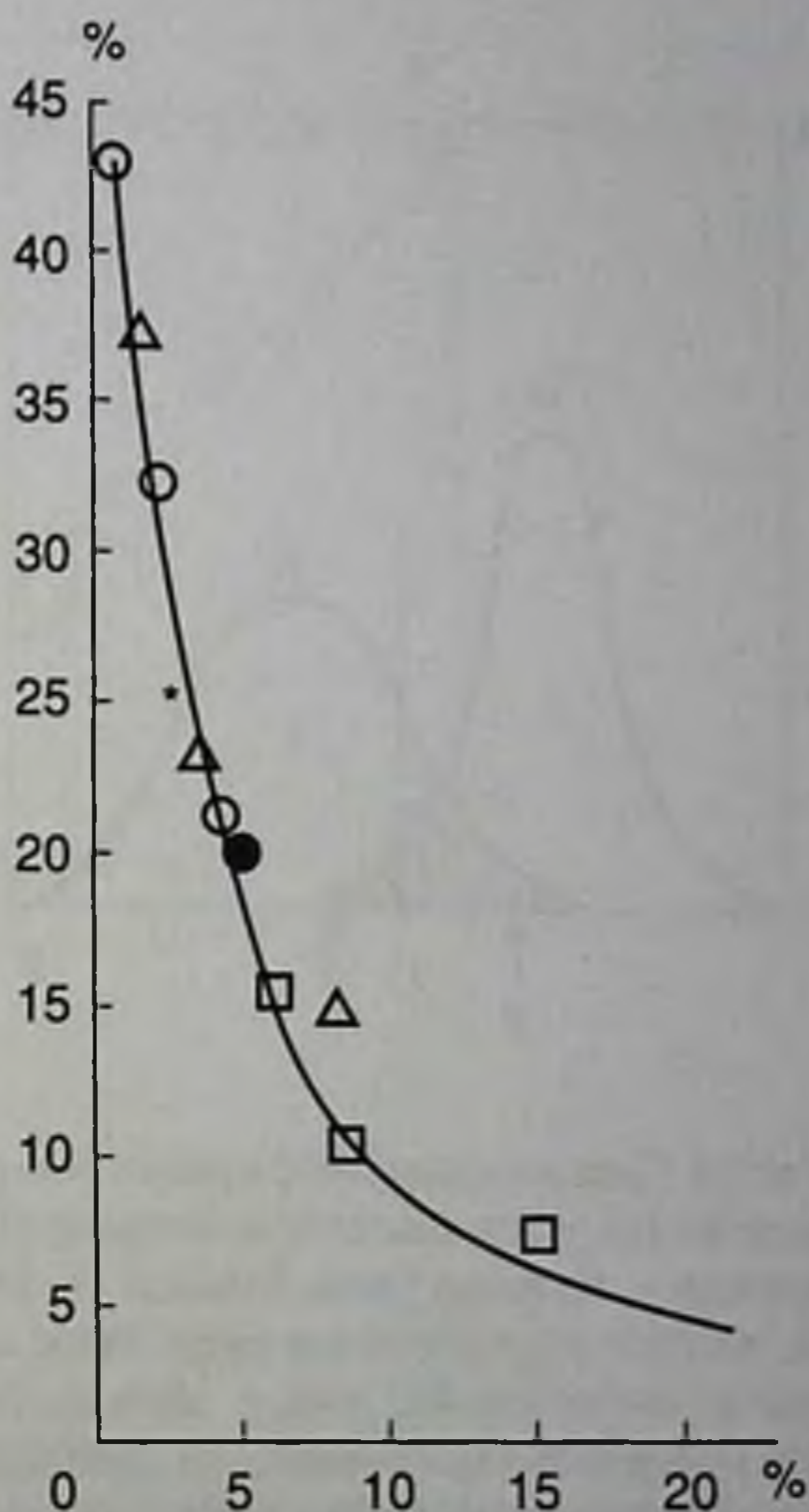


Рис. 2.8. Зависимость вероятности ложноположительных оценок рентгенофлюорограмм при выявлении туберкулеза легких (абсцисса) от вероятности ложноотрицательных (ордината) (по данным 5 исследований, обобщенных Л. Ластед. Введение в проблему принятия решений в медицине. М.: Мир, 1971)



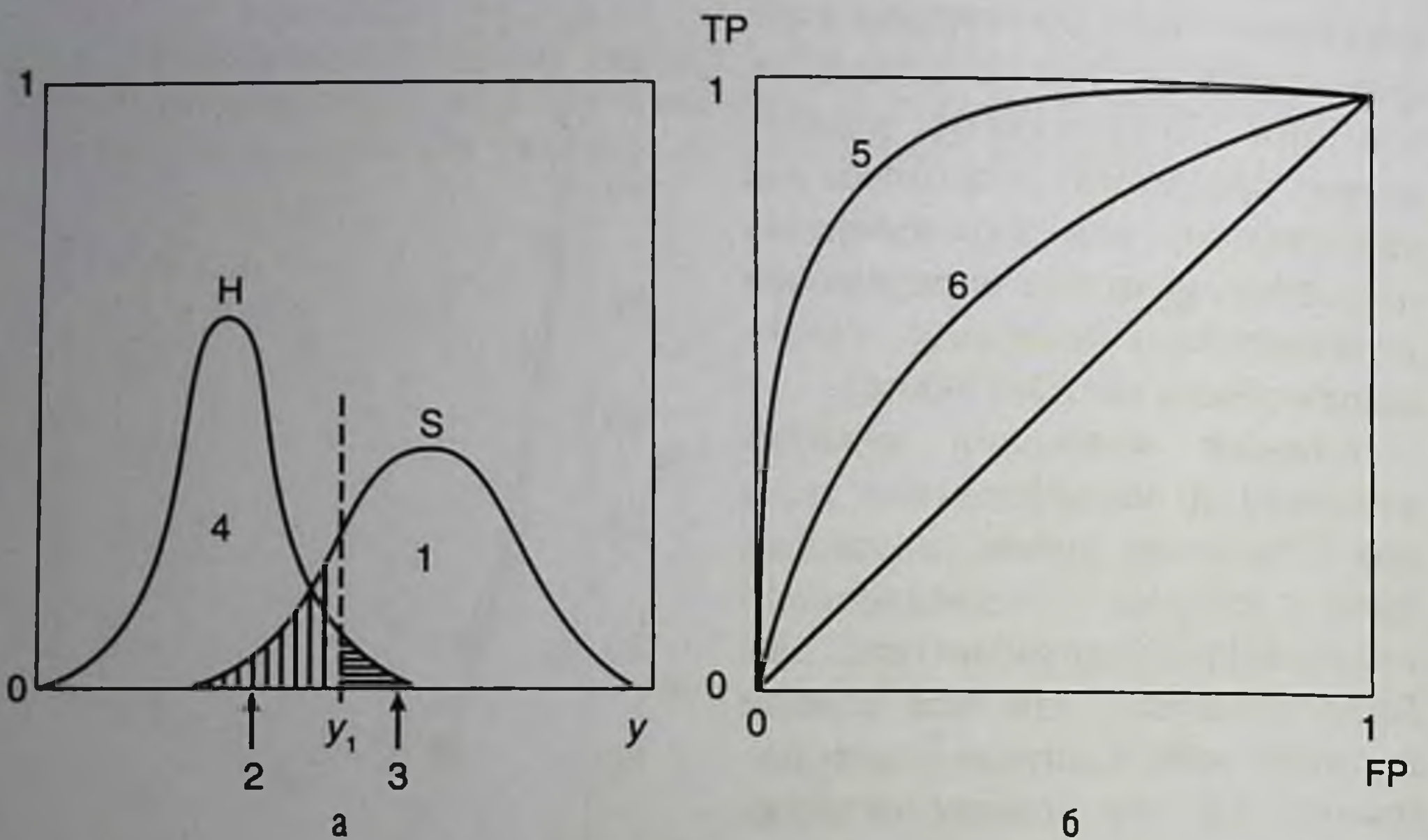


Рис. 2.9. Происхождение ROC-кривой: а — распределение результатов измерения признака  $y$ , полученных у здоровых (Н) и больных (S), перекрывается. При установлении точки разделения  $y_1$  большая часть больных (1, истинно положительные результаты) и здоровых (4, истинно отрицательные результаты) классифицируются правильно. При этом имеют место ошибки первого рода — часть здоровых оказываются за точкой разделения и классифицируются как больные (3, ложноположительные результаты) и второго рода — часть больных классифицируются как здоровые (2, ложноотрицательные результаты); б — при сдвиге точки разделения  $y_1$  изменения соотношений вероятности истинно положительных результатов (ТР, чувствительность) и ложноположительных результатов (FP, т.е.  $1 - \text{специфичность}$ ) для эффективного диагностического теста описываются кривой (5). Для менее эффективного теста, хуже разделяющего больных и здоровых, ROC-кривая лежит ближе к диагонали (6)

Информативность диагностического теста оценивается настолько высоко, насколько высоко лежит его ROC-кривая. Простейшей оценкой информативности является оценка части площади квадрата, заключенной под кривой. Чем ближе площадь под кривой к 100%, тем эффективнее диагностический тест. Чем ближе кривая к диагонали, тем ниже эффективность диагностического метода. Диагональ (50% площади квадрата) соответствует полной неразличимости здоровых и больных изучаемым методом. Точки, лежащие ниже диагонали, соответствуют «обратной» диагностике, т.е. ситуации, когда здоровые чаще называются больными, чем истинные больные. На рис. 2.10 приведены



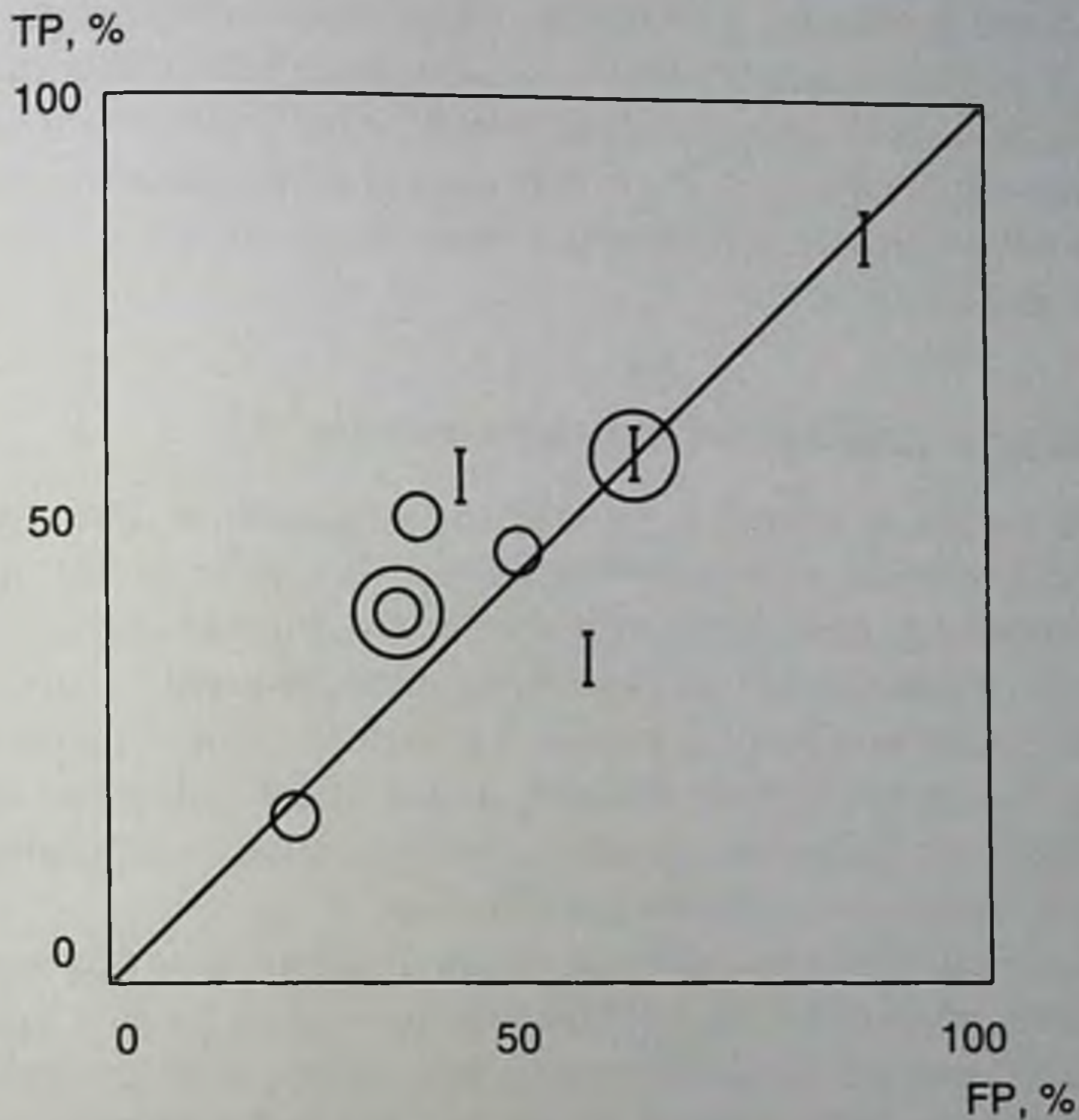


Рис. 2.10. Данные испытания точности диагностики болезней почек по радужке глаза опытными иридодиагностами (I) и офтальмологами (O). Данные табл. 2.2. Кружком обведены средние величины для иридодиагностов и офтальмологов

данные о диагностике болезней почек иридодиагностами (хилерами, использующими альтернативный метод диагностики, претендующий на выявление болезней по радужке глаза). Хорошо видно, что все экспериментальные точки лежат около диагонали. Это означает, что метод неэффективен.

Построение характеристической кривой является наиболее полным методом описания эффективности диагностического теста. Когда имеется характеристическая кривая, можно обсуждать вопрос о том, какую точку разделения выбрать. Простые способы не дают хороших результатов. Например, ошибкой является стремление к минимуму числа ложных результатов. При этом подходе уравниваются ложноположительные и ложноотрицательные результаты. Цена их последствий никогда не бывает одинаковой. Ложноположительный ре-



зультат может угрожать, например, лишь дополнительным визитом к врачу. А ложноотрицательный — невыявлением опасной, но излечимой болезни. Поэтому выбор точки разделения осуществляется путем взвешивания последствий. Важность последствий определяется не только объективными особенностями исходов, но и отношением людей к этим исходам.

### 2.2.7. Норма и референтные величины

Представления о норме и ее вариациях являются важной частью медицинской теории и социальной жизни. В случае описания нормы по номинативным признакам вариабельность описывается как преваленс отдельных вариантов у здоровых людей (например, доля отдельных вариантов строения артерий руки). Концепция нормы используется очень широко. В инструментальной, в том числе лабораторной, диагностике свойства здоровых людей используются в виде «референтных величин». Их получают следующим образом.

Группа людей, отвечающих некоторым стандартным признакам здоровья, называется *референтной*. Применение теста к людям, составляющим такую группу, в стандартных условиях и при стандартной методике дает совокупность референтных величин. Эти референтные величины в статистическом смысле распределены некоторым образом, и это распределение может быть описано с помощью известных математических способов. Например, иногда фактические распределения могут быть удовлетворительно описаны с помощью средней величины и среднего квадратического отклонения, т.е. соответствуют закономерностям «нормального», или гауссовского, распределения. В этом случае устанавливают границы референтного интервала от  $\bar{X} - 2\text{СКО}$  до  $\bar{X} + 2\text{СКО}$ , в которых находится 95% нормально распределенных результатов измерения (см. разд. 1.6.3).

В более частом случае, когда фактическое распределение не соответствует какому-либо теоретическому, или если исследователь принципиально не желает искать способ описания распределения, он может найти границы, включающие 95% средних величин (или 90, 99%) методом процентилей. Метод заключается в ранжировании всех результатов измерений и нахождении величин, соответствующих 2,5% нижних и 2,5% верхних результатов. За этими пределами остается 5% экстремальных результатов. Внутри референтного интервала — 95%.



Референтный интервал — это не норма «вообще», а характеристика определенной референтной группы и условий исследования. Он не претендует на то, чтобы выходящие за его пределы случаи считались «ненормальными». Они лишь *нехарактерны* для референтной группы, ведь 5% членов референтной группы имеют характеристики, выходящие за пределы интервала. Поскольку для измерения признака, например, концентрации с-реактивного белка в крови могут использоваться разные методы, для каждого из методов могут быть свои, особенные референтные интервалы.

Каким бы способом ни устанавливались границы референтного интервала, практическое их применение приводит к тому, что у большинства больных и у многих здоровых людей обнаруживается при обследовании большее или меньшее число «отклонений». Чем больше используется разнообразных тестов, тем больше отклонений. Тратятся большие средства на дополнительное обследование людей, у которых выявлены «отклонения». В особенности остро встает эта проблема при скрининге, т.е. в условиях низкого преваленса (см. разд. 5.2.3).

### **2.2.8. Место диагностических тестов в ведении больного**

Когда устанавливают границы референтных интервалов, исходя из распределения величин у здоровых людей (см. разд. 2.2.7), то игнорируют различия болезней между собой. В действительности при одних болезнях наблюдаются величины более близкие к «норме», а при других — более далекие от «нормы». В группе больных разными болезнями типичная величина признака будет примерно такая же, как у здоровых, лишь вариабельность будет больше (рис. 2.11). Диагностические пороги (точки разделения) устанавливаются разные для разных болезней. Такой порог может быть установлен и внутри референтного интервала для здоровых людей, и за его пределами. Референтные интервалы сохраняют свое значение для количественной оценки результатов измерения количественных признаков безотносительно известной или предполагаемой патологии у того или иного больного. При рассмотрении конкретной диагностической гипотезы врач использует специальные диагностические критерии.

Диагностические критерии создаются не для выделения величин признаков, отличных от референтных, а для выделения величин призна-



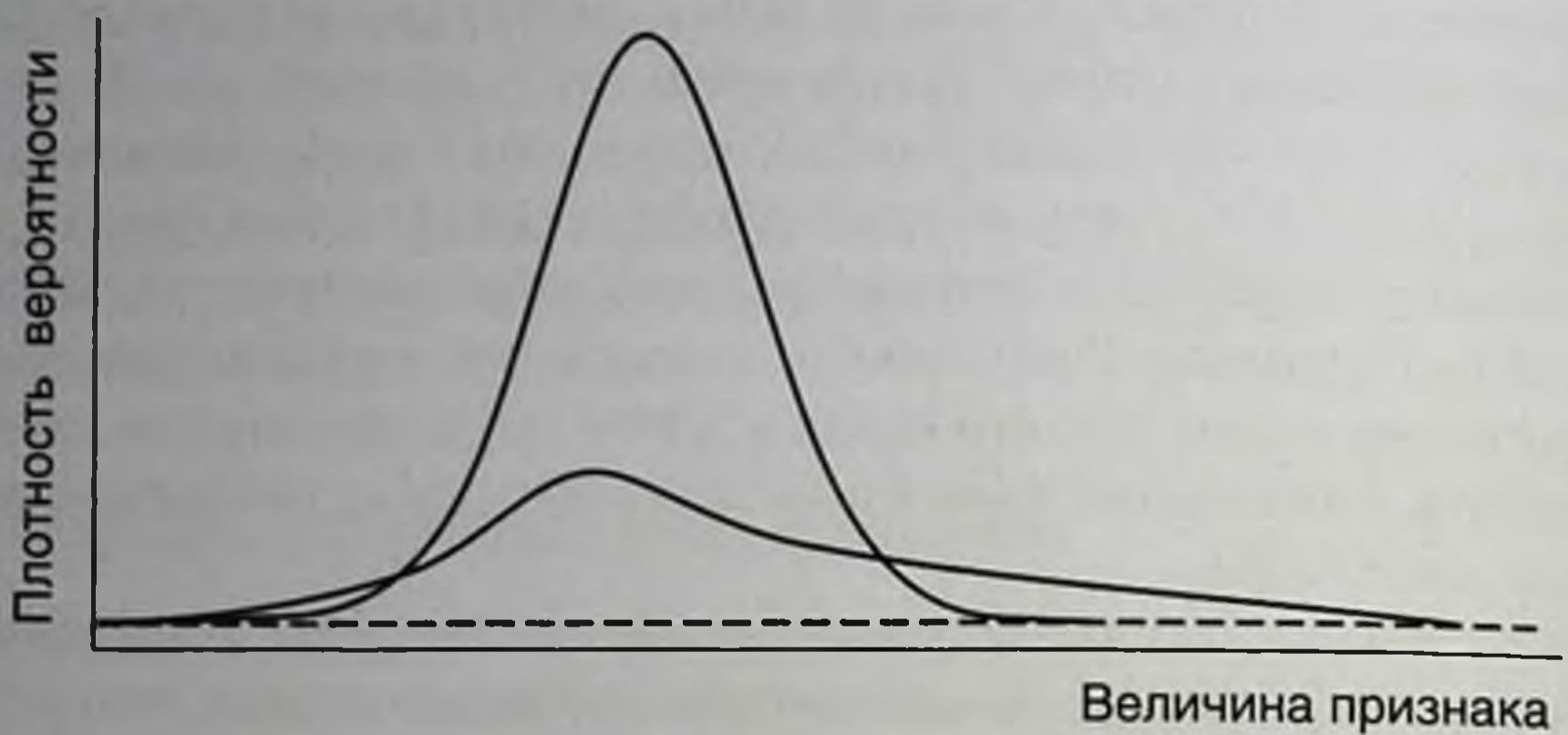


Рис. 2.11. Типичное соотношение кривой распределения в норме (высокая) и кривой распределения в случаях различной патологии — например, у всех пациентов многопрофильного стационара (низкая)

ков, при которых лечение оправдано. Например, величина артериального давления 140/90 мм рт.ст. является распространенным критерием диагностики гипертензии. Однако ее не рассматривают как критерий принятия решения о медикаментозном лечении: сначала выполняют повторные измерения, дополнительные исследования, применяют немедикаментозные методы лечения. Диагностический критерий как инструмент для разделения «здоровых» и «больных» в целях вмешательства зависит от возможностей лечения, от знаний о последствиях лечения. Например, в конце XX в. представления о том, при каком артериальном давлении надо начинать лекарственную терапию, изменились в пользу вмешательства более агрессивного и при меньших величинах давления. Представления о «пограничных» состояниях и «нейроциркуляторной дистонии гипертонического типа», при которой якобы гипотензивная терапия не нужна, остались в прошлом. То же имеет место в ведении больных сахарным диабетом.

Диагностический тест обычно применяется в совокупности с другими тестами. Тем не менее полезность изучаемого теста должна быть показана при его самостоятельном применении. При использовании теста в совокупности с другими должно быть определено, в каком случае применяется этот тест (применительно к каким больным, при каких результатах других тестов). Должна быть показана его эффективность в выделенной другими средствами группе больных — с измерением чувствительности и специфичности. Обязательно должно быть



оценено улучшение диагностики в целом вследствие применения этого теста.

Если применение теста в сочетании с другими тестами специально не оговаривается стандартными диагностическими критериями, то врач может последовательно использовать один тест за другим, переходя от одной вероятности болезни к другой, пока не достигнет удовлетворительной вероятности диагноза. Исключением являются тесты, связанные между собой по принципу действия. Например, если гипокоагуляция подтверждена данными тромбоэластографии, то другие тесты на гипокоагуляцию уже не дадут существенной дополнительной (независимой) информации, хотя и могут уточнить ее происхождение.

Критерием эффективности диагностики является возможность выяснения истинного диагноза. Критерием полезности применения теста является возможность улучшения исхода лечения. Поскольку полезность диагностики опосредована другими вмешательствами, оценить итоговую эффективность (полезность) диагностического теста сложно. Общее стремление к точности диагноза, существующее в медицине, не является оправданным само по себе. Длительное уточнение диагноза может быть бесполезным, ибо откладывает лечение и увеличивает страдания и расходы больного и общества. В других случаях может отсутствовать специфическое лечение и проводится лечение неспецифическое, что делает уточнение диагноза бесполезным. Сложившаяся практика обследований не является доказательством полезности. Иногда такие традиции относительно безвредны, а иногда — травматичны. Например, многие диагностические тесты, проводимые детям при ночном недержании мочи, болезненны и эмоционально травмируют их (цистоскопия), в то время как показано, что результаты этих тестов не влияют на исход лечения.

### 2.2.9. Стандартное определение случая

Диагностические критерии используют для вынесения диагностических и терапевтических решений. В свою очередь, результаты диагностики используют врачи, заполняя формы статистического учета. Как мы видели (разд. 2.2.2), результаты клинической диагностики всегда лишь относительно правильны и различаются в разных больницах и в разных странах. Для того чтобы результаты анализа заболе-



ваемости были более сопоставимыми, используются стандартные определения случая заболевания. *Стандартное* (или *рабочее*) *определение случая* (англ. *case definition*) — набор критериев, позволяющих отнести лицо к группе заболевших или к группе подвергшихся некоторому воздействию. Такие критерии обычно включают клинические признаки, данные лабораторных исследований, указание на время и место заболевания. Например, в первые месяцы распространения новой коронавирусной инфекции (COVID-19) использовали определение подозрительного случая как человека с высокой температурой тела в сочетании с кашлем, если такой человек находился перед этим в Китае.

Эпидемиологи обычно используют несколько упрощенные по сравнению с клиническими критерии, но более стандартизованные. В описание случая могут входить инструментальное (лабораторное) подтверждение, наличие клинических признаков, эпидемиологических данных. В некоторых случаях учитываются анатомические особенности поражения. Например, дифтерия с поражением дыхательных путей является случаем, обязательным для уведомления учреждений, осуществляющих эпидемиологический надзор, а кожная — нет. Стандартные критерии случая используются для наблюдения и учета (см. разд. 1.3.2), обычно ретроспективно. Они не заменяют диагностических критериев и не служат единственным основанием для медицинского вмешательства или вмешательства в защиту общественного здоровья.

В клинике диагностические критерии совершенствуются и постепенно становятся международными. Точно так же стандартные определения случая в эпидемиологическом наблюдении унифицируются, но пока не являются международными стандартами. Например, в США только в 1990 г. была создана национальная система стандартных определений случаев заболеваний для системы наблюдения<sup>1</sup>.

### 2.2.10. Прогноз

Если исследованный признак связан с развитием событий в последующий период времени, то это соответствие может использоваться для прогнозирования будущих событий. В медицине наибольшей предсказательной силой обладает диагноз. Именно характер болезни лучше

<sup>1</sup> <http://www.cdc.gov/mmwr/preview/mmwrhtml/00047449.htm>.



всего позволяет предсказать дальнейшие события, прежде всего — вероятность осложнений и смерти. Основной прогнозируемый показатель — летальность (см. разд. 1.3.5). Он подходит только для острых болезней и травм. Применительно к хроническим болезням традиционно оперируют 5-летней выживаемостью, но у этого показателя нет преимуществ перед другими сроками, например 1 год или 10 лет. Использование анализа выживания позволяет описать летальность во всем доступном интервале времени (см. разд. 2.6.2) и построить полную кривую выживания.

Прогноз течения болезни может быть уточнен распознаванием вариантов течения болезни. Например, от того, какими клетками представлена ткань злокачественной опухоли, зависят темп ее роста и быстрота метастазирования, реакция на применение цитотоксических препаратов и облучение.

Прогноз может основываться на выявлении стадий болезни. Этот подход менее надежен, поскольку обычно большее значение имеет не стадия болезни в настоящий момент, а скорость перехода к следующим стадиям. Процесс перехода между стадиями заболевания лучше всего описывается с помощью статистического метода *марковских цепей*<sup>1</sup>, т.е. сменяющих друг друга состояний, из которых обычно одно является «поглощающим» (смерть). Этот метод анализа разнообразных процессов (не только медицинских) применяется в анализе принятия решений (см. разд. 5.1.4).

В терапии острых состояний одним из самых надежных методов прогнозирования исхода болезни или травмы является прогнозирование на основании выраженности изменений в состоянии организма. При тяжелых изменениях метаболизма и состояния нервной системы вероятность смерти резко увеличивается. Наиболее известным является индекс тяжести состояния больного АРАСНЕ.

Прогноз для пациентов, получающих лечение, является одновременно показателем эффективности этого лечения. Если установлено, что прогноз для больных раком желудка, оперированных в 2000 г., лучше, чем для оперированных в 1970 г., то можно предполагать, что причиной является улучшение лечения. При этом, однако, надо иметь в виду возможность ошибки. Во-первых, болезни с течением времени могут трансформироваться. Эта трансформация может

---

<sup>1</sup> А.А. Марков (1856–1922) — выдающийся русский математик, автор трудов по теории вероятности и математическому анализу.



включать в себя истинное изменение течения болезни. Например, скарлатина конца XX в. доброкачественна и отличается от скарлатины, наблюдавшейся в середине века. Во-вторых, трансформация проявлений болезни может быть следствием изменения диагностических технологий. Например, в эпоху до разработки ультразвуковых диагностических аппаратов средством выявления пороков сердца была аускультация и надежным критерием формирования приобретенного порока была кардиомегалия. С внедрением ультразвуковых методов результаты диагностики стали принципиально иными. Появилась возможность ранней и уверенной диагностики пороков, и со многими пороками теперь не связаны представления о неблагоприятном прогнозе.

Использование изменившихся средств диагностики может создавать иллюзию улучшения прогноза, что называется *смещением вследствие миграции между стадиями* (*stage migration bias*, описана А. Файнстайном<sup>1</sup>). Представим, что мы наблюдаем исходы лечения у пациентов со злокачественной опухолью, у которой принято клинически выделять два варианта: метастазировавшая и неметастазировавшая (табл. 2.4). Если в прошлом микрометастазы оставались нераспознанными, то это приводило у пациентов, в совокупности классифицировавшихся как имеющие неметастазировавшую опухоль, к повышению летальности до 15%. Применение более совершенного метода исследования позволило выявлять микрометастазы и относить таких пациентов в группу с метастазировавшей опухолью. В результате в 2000 г. летальность снижается у пациентов, отнесенных к обеим клиническим стадиям, хотя эффективность лечения не изменилась.

Большое внимание привлекают возможности прогнозирования состояния здорового человека. Эти возможности связаны, прежде всего, с использованием концепции факторов риска развития заболеваний. Например, в многолетнем проспективном исследовании (MRFIT — испытание возможностей многофакторной профилактики сердечно-сосудистых болезней) были получены результаты применительно к некурящим мужчинам, не имеющим артериальной гипертензии (табл. 2.5). В соответствии с результатами более ранних исследований показано, что риск развития ишемической болезни сердца увеличен при высокой концентрации холестерина в плазме крови. Вы-

<sup>1</sup> A.R. Feinstein (1926–2001) — выдающийся американский математик и врач, создатель клинической эпидемиологии.



деляя относительно небольшую группу лиц с самой высокой концентрацией, можно выявить значительную часть лиц, у которых увеличен риск. В данном случае, выделяя 20% обследованных, можно выделить 35% всех случаев, кому угрожает коронарная смерть.

Таблица 2.4. Гипотетический пример влияния изменений в диагностике на улучшение прогноза при раке

Период, годы	Клиническая стадия					
	Неметастазировавшая			Метастазировавшая		
1970	Нет метастазов	Микро-метастазы	—	—	Макро-метастазы	—
	100 пациентов, летальность 10%	100 пациентов, летальность 20%	Общая летальность 15%		100 пациентов, летальность 70%	Общая летальность 70%
2000	Нет метастазов	—	—	Микро-метастазы	Макро-метастазы	—
	100 пациентов, летальность 10%	—	Общая летальность 10%	100 пациентов, летальность 20%	100 пациентов, летальность 70%	Общая летальность 45%

Таблица 2.5. Смертность от коронарной болезни сердца в течение 6 лет в исследовании MRFIT (по данным Stamler J. et al. JAMA. 1986. № 256 (20). P. 2823–8)

Концентрация холестерина в плазме крови	Умерли от коронарной болезни сердца за 6 лет, %		Всего
	Да	Нет	
В верхних 20% величин	188 (0,7)	26 808	26 996
В нижних 80% величин	342 (0,3)	132 322	132 664
Итого	530	159 130	159 660



Отличительной особенностью прогнозирования состояния здоровья здоровых людей по факторам риска является низкая эффективность прогноза по сравнению с прогнозом аналогичных исходов на основании диагноза. Поэтому общей стратегией применения прогностических методов является внимание к небольшой группе людей с наиболее повышенным риском неблагоприятного исхода — в нашем примере имеющей концентрацию холестерина выше 80-го перцентиля. В низкой эффективности прогноза по факторам риска нет ничего удивительного, так как при диагностике выявляются болезненные состояния, отличающиеся от нормы, а при прогнозировании — состояния, входящие в спектр нормальных. Относительный риск коронарной смерти у лиц с гиперхолестеринемией составляет только  $0,7/0,3 = 2,3$ .

Большая часть случаев коронарной смерти развивается у людей, не относящихся к группе повышенного риска. Поэтому имеет право на существование противоположная стратегия профилактики — популяционная. Если фактор риска установлен, а выявление группы повышенного риска позволяет выделить лишь небольшую часть от общего числа исходов, которые желательно предотвратить, то предлагается не заниматься выявлением группы повышенного риска, а устранять или ослаблять действие фактора риска у всех членов популяции. В реальных условиях берет верх комбинированный подход: вся популяция становится объектом относительно дешевых и безвредных мер просвещения, а лица самого высокого риска по мере выявления становятся объектом более эффективных, дорогих и не столь безопасных мер воздействия, например профилактической фармакотерапии.

В некоторых случаях термин «прогноз» используют не только по отношению к предсказанию будущих событий, но и по отношению к предсказанию признаков, которые могут быть исследованы в настоящее время, но не известны. Например, *прогнозированием* называют предсказание по результатам одного теста результатов другого теста (референтного). С этим связано само появление термина «прогностичность». Как мы видим, методы оценки эффективности диагностического и прогностического тестов одинаковы.

## Контрольные вопросы и задачи к разделам 2.1 и 2.2

1. Для диагностики язвенной болезни желудка в общей врачебной практике предложен признак — специфические изменения



- поверхности языка, который, будучи применен у 100 больных с подтвержденной болезнью, оказался положительным у 36. Это означает, что:
- а) у этого признака специфичность 36%;
  - б) у 64% больных имеет место особый вариант болезни;
  - в) прогностичность признака 36%;
  - г) прогностичность признака 64%;
  - д) чувствительность признака 36%.
2. Диагностический тест, разработанный в популяции, где прева-ленс язвенной болезни высок, будучи примененным в популяции, где язвенная болезнь встречается редко, будет иметь:
- а) повышенную чувствительность;
  - б) его положительные результаты будут обладать меньшей про-гностичностью;
  - в) его отрицательные результаты будут обладать меньшей прогно-стичностью;
  - г) будет давать меньше ложноположительных результатов.
3. Для большинства диагностических тестов существует баланс меж-ду высокой чувствительностью и высокой специфичностью. Ка-кие из приведенных положений правильны?
- а) необходима высокая чувствительность метода диагности-ки рассеянного склероза, так как возможности терапии его ограничены;
  - б) поскольку гипотиреозидизм в детстве приводит к тяжелым по-следствиям для развития ребенка, тест для выявления гипоти-реоидизма должен обладать высокой специфичностью;
  - в) поскольку иридодиагностика не требует больших затрат, она может быть полезна, даже если не разделяет здоровых и больных;
  - г) поскольку менингококковый менингит опасен и имеется эф-фективное лечение, важно применять чувствительное диагно-стическое правило.
4. В лаборатории больницы А используют старый тест для иденти-фикации стрептококков с чувствительностью 92% и специфичностью 98%, а в лаборатории больницы Б — новый, с чувствительностью 98% и специфичностью 98%. Если в эти лаборатории поступят данные от 100 пациентов, то:
- а) лаборатория А правильно выявит больше пациентов с ин-фекцией;



- б) лаборатория А правильно выявит меньше пациентов с инфекцией;
- в) лаборатория А правильно выявит меньше пациентов без инфекции;
- г) необходимо знать преваленс, чтобы сказать, какая лаборатория выявит инфекцию лучше.

### 2.3. ПОПЕРЕЧНЫЕ ИССЛЕДОВАНИЯ

Поперечными или одномоментными (англ. *cross-sectional study*) называют исследования, в которых субъектов (пациентов, школьников, работников) обследуют однократно, как бы создавая одномоментный поперечный срез характеристик популяции. Создается возможность сопоставить наличие болезни с другими признаками людей, например концентрацией вещества в крови или профессией. Чисто поперечные исследования встречаются редко, поскольку врачи, эпидемиологи, социологи стремятся в пределах своих возможностей выстроить временные ряды явлений, стадий заболевания, реконструировать течение описываемого случая болезни. Пример такого исследования приведен на рис. 2.12. Несмотря на то что обследованы *разные* мужчины, данные представлены над абсциссой, на которой отложен возраст, и соединены линиями так, что возникает впечатление *снижения концентрации с возрастом*. Таков замысел авторов, и концентрация тестостерона в плазме действительно снижается с возрастом, если повторно обследовать одних и тех же людей.

К сожалению, такие построения не всегда корректны. Например, можно предположить, что мужчины, продолжающие работать в исследованной компании в возрасте старше 50 лет, отличаются от более молодых сотрудников — к этому возрасту часть умирает, часть уходит на пенсию по возрасту или инвалидности, часть меняет место работы. В результате люди, работающие в компании в старшем возрасте, отличаются от более молодых сотрудников по многим признакам. Возможно, например, что некоторое увеличение концентрации тестостерона, видимое на графике, отражает как раз особенности этих «профессиональных долгожителей». Трудно предположить, что истинная концентрация тестостерона снижается, а затем начинает повышаться после 50 лет.



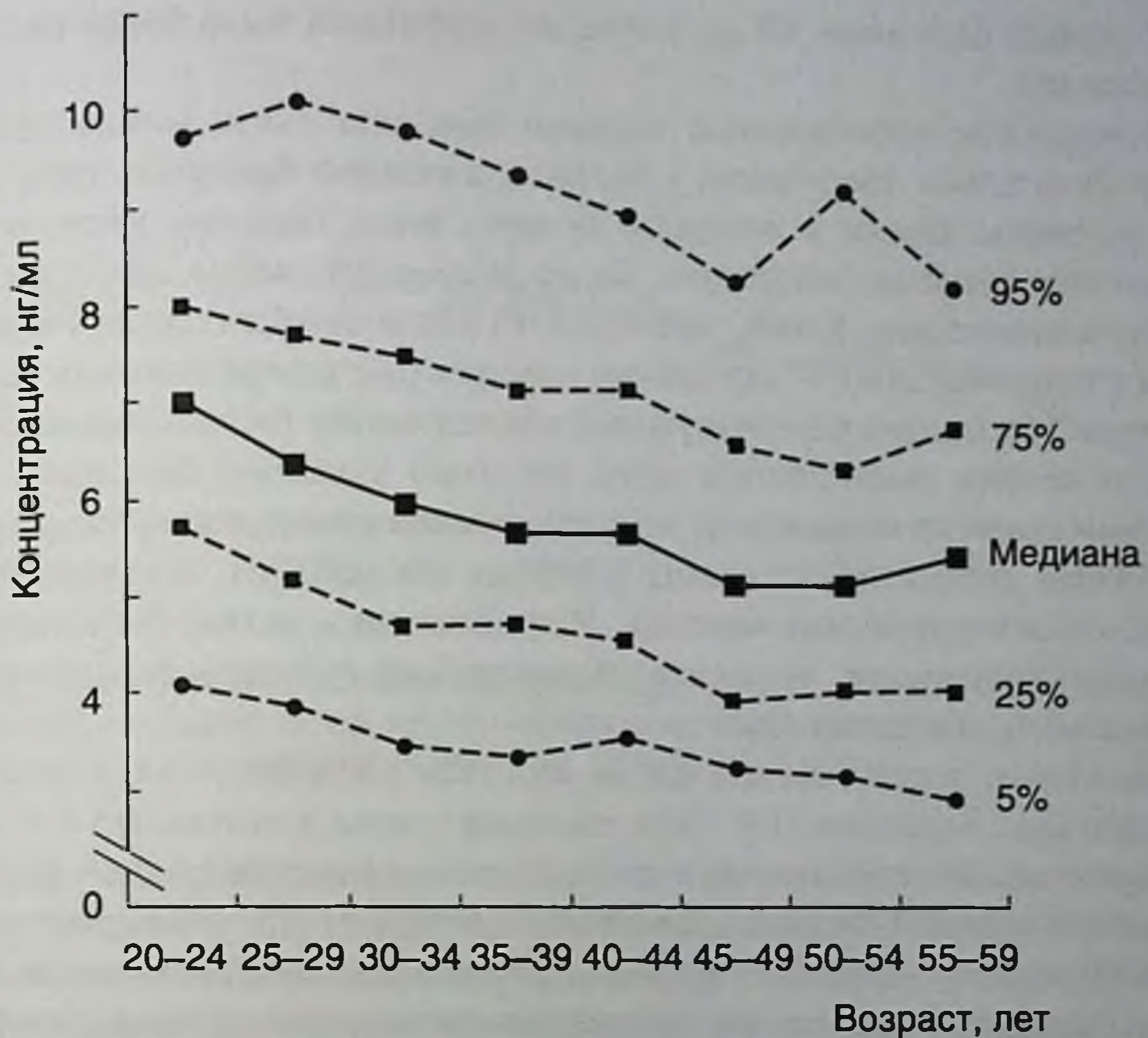


Рис. 2.12. Концентрация тестостерона в плазме крови 1408 мужчин (Telecom Study, Paris, 1985–1987). Кривые — медиана и два процентильных интервала — 90 и 50% (по Simon D. et al. Am. J. Epidemiol. 1992. № 135. P. 783–91, с разрешения издателя)

Обычно поперечное исследование не может дать неискаженных сведений об изменениях во времени. При построении по данным поперечного исследования представления о развитии болезни могут в действительности отражать не стадии, а варианты течения болезни или вероятность наличия болезни. Исследование выборки из европейской популяции, выполненное в конце XX в., показывает, что с возрастом преваленс хронической инфекции *Helicobacter pylori* увеличивается. Прямолинейная интерпретация привела бы к выводу о приобретении этого признака с возрастом. В действительности инфицирование геликобактером происходит преимущественно в молодости и у большинства остается бессимптомным. Поэтому нарастание инфицированности с возрастом просто отражает больший преваленс инфекции в старших возрастных группах, инфицированных в середи-



не и первой половине XX в., когда эта инфекция была более распространенной.

Поперечное исследование нередко выявляет связи между отдельными болезнями. Например, у больных язвенной болезнью чаще диагностированы камни в желчном пузыре, чем у больных пневмонией и многими другими болезнями. Более детальный анализ выявляет, что эта связь не истинна. В действительности большая часть случаев холелитиаза у таких больных — случайные находки при ультразвуковом исследовании. Поскольку ультразвуковое исследование гастродуоденальной зоны и печени выполняется чаще больным язвенной болезнью, чем больным другими болезнями, постольку чаще обнаруживается холелитиаз. У тех субъектов, кто чаще и глубже обследуется, всегда имеется больше диагностических находок. Истинная связь может быть выявлена только в ситуации, когда все сравниваемые субъекты подвергаются одинаковому обследованию.

Вариантом такой ложной связи является систематическая ошибка, носящая имя Берксона [13]. Она возникает, если в стационар и в поле зрения исследователя попадают избирательно пациенты с одной формой болезни. Например, больные язвенной болезнью госпитализируются нечасто. Известно, с другой стороны, что лица с проблемами психологической и социальной адаптации госпитализируются чаще. Соответственно, госпитализируются чаще больные язвенной болезнью, сопровождающейся психологическими проблемами. Если обследовать только госпитализированных больных, то обнаружится, что язвенная болезнь ассоциируется с проблемами психологической и социальной адаптации.

## 2.4. ИЗМЕНЕНИЯ ЗАБОЛЕВАЕМОСТИ ВО ВРЕМЕНИ

Эпидемии или вспышки заболеваемости были главными событиями не только медицинской, но и всей мировой истории. Из множества объяснений постепенно выделились те, которые позволяли дать простое, понятное и эффективное средство если не к предотвращению эпидемии, то к новым исследованиям, проливающим свет на возбудителя болезни или механизм ее распространения. В течение XVII—XIX вв. во многих странах Европы были выполнены исследования смертности с использованием статистических методов, положившие начало широкому применению этих методов в демографии и эпидемиологии. Медицинская



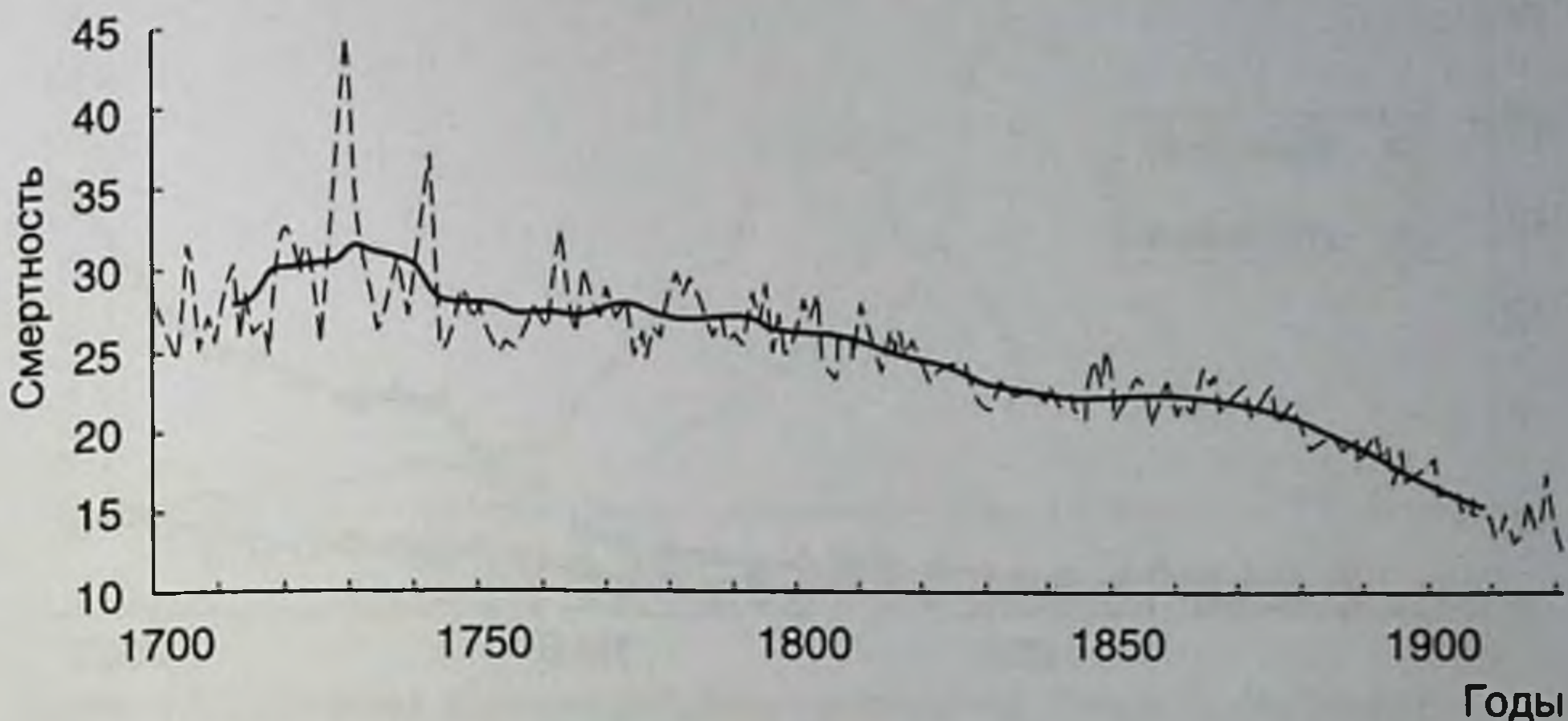


Рис. 2.13. Смертность (на тысячу населения в год) в Англии и Уэльсе в 1700–1920 гг. Пунктир — средняя за год, линия — сглаженная средняя за 25 лет (по Harris B. *Internat. J. Epidemiol.* 2001. № 30. P. 688)

и социальная проблематика оказалась одним из самых важных приложений статистики, и многие статистические методы были разработаны впервые именно для решения медицинских и демографических проблем.

Вторым условием развития эпидемиологических методов в Европе было создание еще в средние века в отдельных городах и странах системы регистрации рождений и смертей и даже регистрации причины смерти. В передовых странах Европы стали систематически проводиться переписи населения. Эти документы легли в основу проводимых до наших дней исследований, например наследования болезней и биологических признаков во многих поколениях. Благодаря развитию статистического учета в то время, мы можем представить, как изменялась смертность за последние 300 лет: смертность не только снизилась, но ее колебания разительно уменьшились (рис. 2.13).

### 2.4.1. Вспышка заболеваний

Резкое увеличение заболеваемости — вспышка — является одним из самых важных объектов изучения в эпидемиологии. Во-первых, потому что при этом угроза жизни и здоровью людей наиболее велика, во-вторых, потому что нередко для прекращения массовых заболеваний есть средства, и, в-третьих, потому что изучение вспышки заболеваний



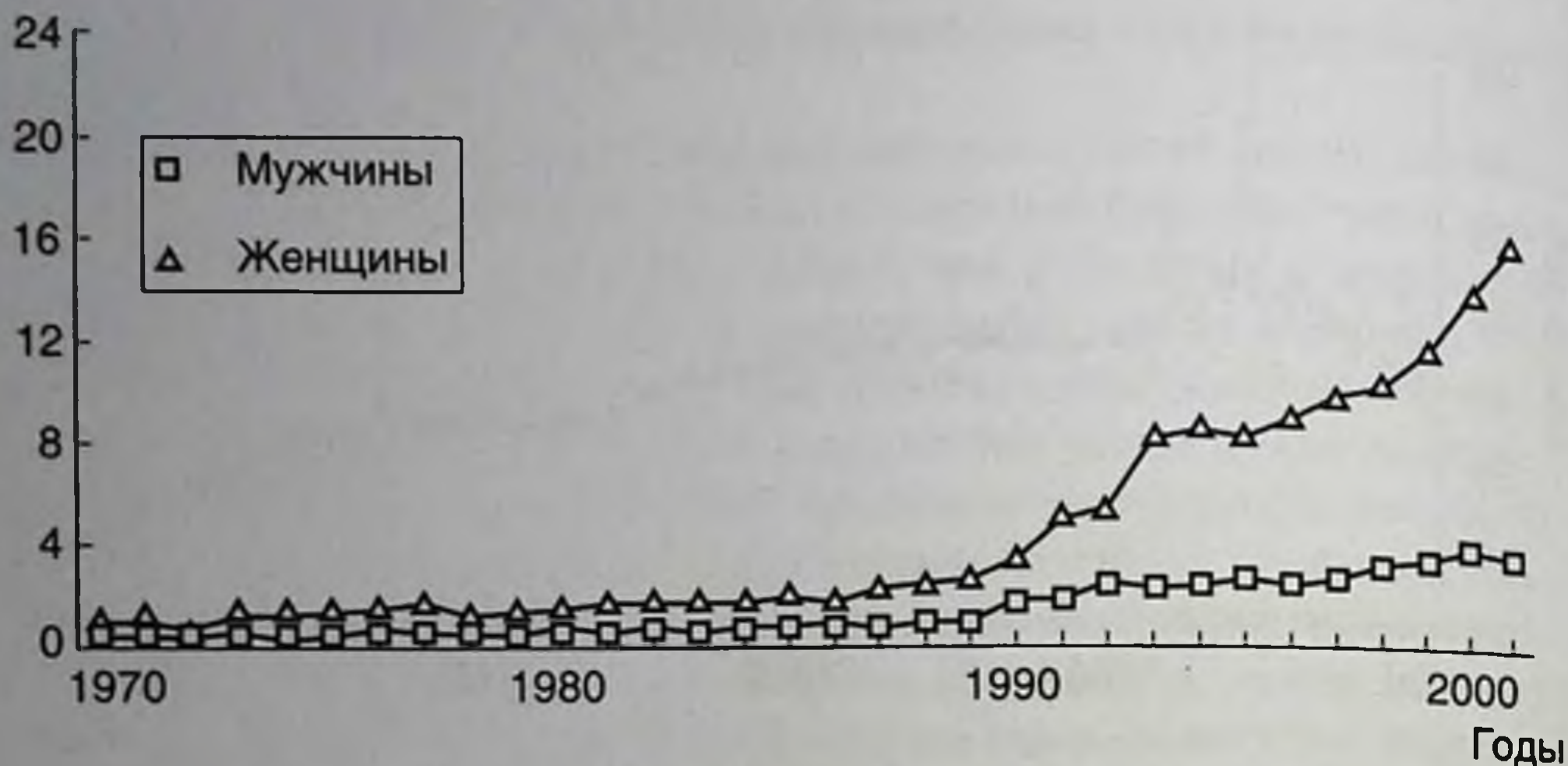


Рис. 2.14. Инцидент рака щитовидной железы после Чернобыльской катастрофы на 1 000 000 населения (ордината) в Беларуси. Абсцисса — годы диагностики (по данным Mahoney M.C. et al. *Int. J. Epidemiol.* 2004. № 33 (5). P. 1025–33)

позволяет иногда устанавливать причину лучше, чем при изучении отдельных случаев. Традиционным объектом внимания эпидемиологии были вспышки инфекционных болезней, однако массовые страдания в результате последствий экономической деятельности человека или техногенных катастроф также являются яркими примерами современного эпидемиологического анализа (рис. 2.14) (см. врезку на с. 290).

### Вспышка легионеллеза

Английские эпидемиологи любезно разместили в интернете детальные данные о вспышке легионеллеза летом 2002 г. в Warrington-Furnes. Эта инфекционная болезнь, вызываемая легионеллой, проявляется как пневмония с относительно высокой летальностью — 10–15% у здоровых до этого людей. Инфицирование происходит при вдыхании водного аэрозоля, содержащего возбудитель. Такие аэрозоли образуются техническими средствами, например кондиционерами некоторых типов, охладительными башнями тепловых станций. В промышленно развитых странах легионеллез составляет 2–15% всех случаев внебольничной пневмонии.

На гистограмме распределения случаев по срокам заболевания хорошо видно, как в начальной стадии рост числа случаев происходит по экспоненте, а затем число новых случаев быстро спадает (рис. 2.15).



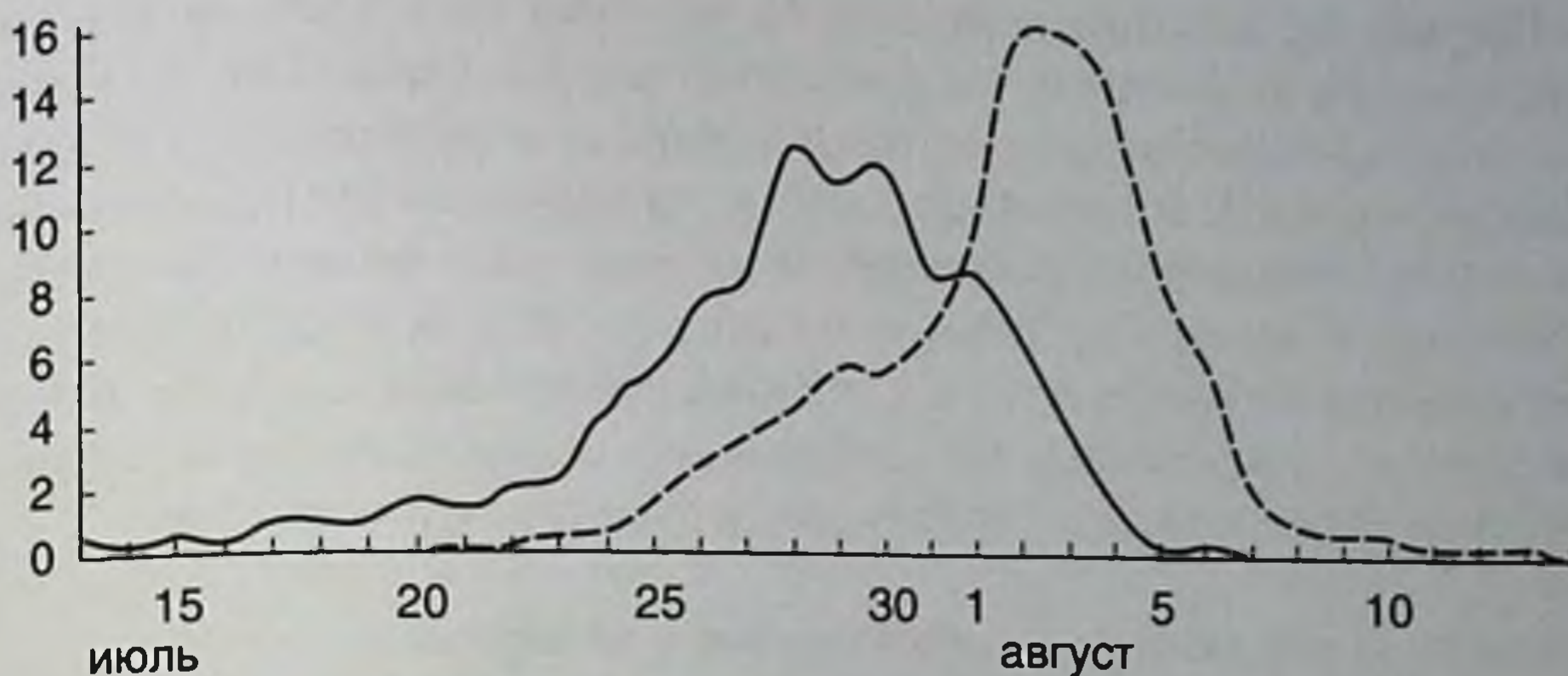


Рис. 2.15. Динамика развития вспышки легионеллеза. Линия — распределение по дате заболевания, пунктир — по дате обращения за помощью. Ордината — число случаев (источник: <http://www.healthprotection.org.uk/Barrow/main.html>)

Такая динамика характерна для вспышек, возникающих в результате заражения в короткое время большого числа людей. Распределение по срокам заболевания в таком случае предстает как распределение по длительности инкубационного периода.

### 2.4.2. Применение критерия хи-квадрат ( $\chi^2$ ) и точного критерия Фишера

$\chi^2$  — один из наиболее употребляемых в медицине статистических критериев. Он подходит для анализа качественных данных. Отличие критерия — независимость от формы распределения (непараметричность). Критерий  $\chi^2$ , как и ряд других критериев, основывается на сопоставлении частот наблюдаемых событий с теми, что могли бы иметь место, если бы события происходили случайно. Например, если подбрасывание монеты дает выпадение «орлом» 20 раз против 2 «решеткой», то у нас есть основания полагать, что монета «неправильная», поскольку результат сильно отличается от ожидаемого случайного отношения 1:1. В случае сопоставления двух и более событий, например съеденного блюда и заболевания, ожидаемое число наблюдений вычисляется, исходя из определения независимых событий. *Независимыми событиями* в статистическом смысле называют события, совместная вероятность которых равна произведению вероятности каждого из событий (см. разд. 1.3.3).



**Пример.** На пикнике заболели 24 человека из 52. Мы можем сопоставить их заболевание со съеденной пищей (табл. 2.6). В нашем простейшем случае можно составить четырехпольную таблицу ( $2 \times 2$ , латинский квадрат). В действительности, на первом этапе исследования составляют довольно большие таблицы, в которых фигурируют десятки факторов риска — от пищи в течение последних дней до содержания домашних животных и рыб, условий проживания и работы. В ходе предварительного анализа выделяются несколько наиболее вероятных факторов риска, а затем обычно удается выделить один или два.

Таблица 2.6. Пример результатов анализа группового заболевания

Питание	Заболели	Не заболели	Всего
Бутерброд с сыром	5( <i>a</i> )	22( <i>b</i> )	27( <i>a</i> + <i>b</i> )
Бутерброд с паштетом	19( <i>c</i> )	6( <i>d</i> )	25( <i>c</i> + <i>d</i> )
Итого	24( <i>a</i> + <i>c</i> )	28( <i>b</i> + <i>d</i> )	52( <i>n</i> )

Нулевая гипотеза в данном исследовании состоит в том, что съеденная пища и заболевание не связаны между собой. Мы обоснованно предполагаем, что если пища не является причиной заболевания, то заболевание будет лишь случайным образом совпадать с характером пищи. Применение критерия основывается на сопоставлении наблюдаемого сочетания пищи и заболевания со случайно ожидаемым. Так, в табл. 2.6 в левой верхней ячейке ожидаемая вероятность составляет произведение суммарной вероятности в строке ( $(a + b)/n = 0,52$ ) на суммарную вероятность в колонке ( $(a + c)/n = 0,46$ ), т.е.  $0,52 \times 0,46 = 0,24$ . Отсюда ожидаемое число случаев  $0,24 \times 52 = 12,46$ . Вычисления могут быть упрощены: ожидаемое число в ячейке *a* можно вычислять как  $(a + b)(a + c)/n^2$ . В табл. 2.7 приведен расчет критерия  $\chi^2$  для нашего примера.

Формула вычисления критерия в общем виде:

$$\chi^2 = \sum [(O - E)^2 / E],$$

где  $\Sigma$  обозначает суммирование. Иными словами, критерий равен сумме отношений квадратов разностей фактической и ожидаемой частоты к ожидаемой.

Для таблицы  $2 \times 2$  формула может быть упрощена:  $\chi^2 = n(ad - bc)^2 / [(a + b)(a + c)(b + d)(c + d)] = 52 \times (30 - 418)^2 / [27 \times 24 \times 28 \times 25] = 17,25$ .



Величина  $\chi^2$  служит для оценки статистической значимости отличия наблюдаемых частот от случайных. Для этого надо принять во внимание число наблюдаемых случаев, точнее, число степеней свободы. Под *числом степеней свободы* обычно понимают число изученных случаев минус число параметров, ограничивающих вариации. В случае *таблиц сопряжения*, подобной рассмотренной выше табл. 2.6, число степеней свободы составляет  $(r - 1)(c - 1)$ , где  $r$  — число строк,  $c$  — число колонок в таблице. По одной колонке и строке вычитается, так как значения «всего» в колонках и «всего» в строках (при вычислении ожидаемой частоты) зависят от остальных колонок (строк), помимо той, для которой проводятся вычисления. В нашем случае число степеней свободы 1. Сравнение  $\chi^2$  со значениями, приведенными в приложении 6, показывает, что гипотеза о случайном совпадении вида бутербродов с заболеванием может быть отклонена с вероятностью ошибки типа I  $P < 0,001$ .

Таблица 2.7. Ожидаемые частоты для примера табл. 2.6

Наблюдаемое число (O)	Ожидаемое число (E)	Разность наблюдаемого и ожидаемого (O-E)	$(O-E)^2/E$
$a = 5$	$E(a) = 12,46$	-7,46	4,47
$b = 22$	$E(b) = 14,54$	7,46	3,83
$c = 19$	$E(c) = 11,54$	7,46	4,82
$d = 6$	$E(d) = 13,46$	-7,46	4,13
Всего 52	52	0	$\chi^2 = 17,25$

Как видно,  $\chi^2$  нетрудно применить в простом случае, используя вычисления на калькуляторе. В реальных условиях эпидемиологического анализа следует применить компьютерную программу, например EpiInfo (см. примечание на с. 103), в которой предусмотрены создание форм для обследуемых случаев, перевод данных в сводную таблицу и расчет статистических критериев.

Если таблица сопряжения имеет размер  $2 \times 2$ , то  $\chi^2$  желательно вычислять с поправкой Йетса (F. Yates — статистик, предложивший эту поправку), чтобы не давать слишком оптимистических оценок статистической значимости. Размер вносимой поправки уменьшается



по мере увеличения числа наблюдений. Поправка состоит в вычитания из разности ожидаемого и фактического числа наблюдений  $1/2$  от общего числа наблюдений ( $n/2$ ), т.е. заменой  $O-E$  на  $|O-E|-n/2$ . Для нашего примера формула примет вид:

$$\chi^2 = n(|ad - bc| - n/2)^2 / [(a + b)(a + c)(b + d)(c + d)],$$

где  $|ad - bc|$  обозначает абсолютную величину разности.

Вычисления с поправкой дадут результат:

$$52 \times (|30 - 418| - 26)^2 / [27 \times 24 \times 28 \times 25] = 15,02.$$

$\chi^2$  — подходящий статистический критерий для независимых совокупностей. Поэтому если исследуются связанные совокупности, например диагнозы группы людей в начале и в конце наблюдения, то его использование нежелательно. Столь же нежелательно использовать  $\chi^2$  для анализа ординальных данных, так как это приводит к потере чувствительности эксперимента — достоинства ординальных данных перед качественными не будут использованы.

Важным достоинством критерия  $\chi^2$  является его приложимость к практически любому числу категорий (в  $n \times m$ -таблице  $n$  и  $m$  могут быть больше двух). Однако таблицу с большим числом ячеек в эксперименте трудно бывает «хорошо» заполнить, ведь для вычисления критерия надо, чтобы в каждой ячейке был хотя бы один пациент и желательно, чтобы не более 20% ячеек имели ожидаемое число событий менее 5. В ситуации, когда много ячеек таблицы содержат менее 5 наблюдений, прибегают к объединению близких групп, например объединяют пациентов с разными проявлениями респираторной патологии (кашель, насморк, боль в груди). При этом численность наблюдений в объединенных ячейках увеличивается.

Критерий  $\chi^2$  не следует применять, если общее число обследованных менее 20. Уверенно можно применять, если оно более 40. Если от 20 до 40, то только в том случае, когда ожидаемое число наблюдений в каждой ячейке 5 и более (в случае четырехпольной — 4 и более). Если критерий  $\chi^2$  неприменим по указанным причинам, для качественных данных лучше воспользоваться точным критерием Фишера<sup>1</sup> (ТКФ).

<sup>1</sup> Р. Э. Фишер (R.A. Fisher) (1890—1962) — английский астроном, математик и генетик, творец современной статистики; ему мы обязаны концепцией правдоподобия, методами анализа малых выборок, принципом рандомизации и выборочного исследования, дисперсионным анализом.



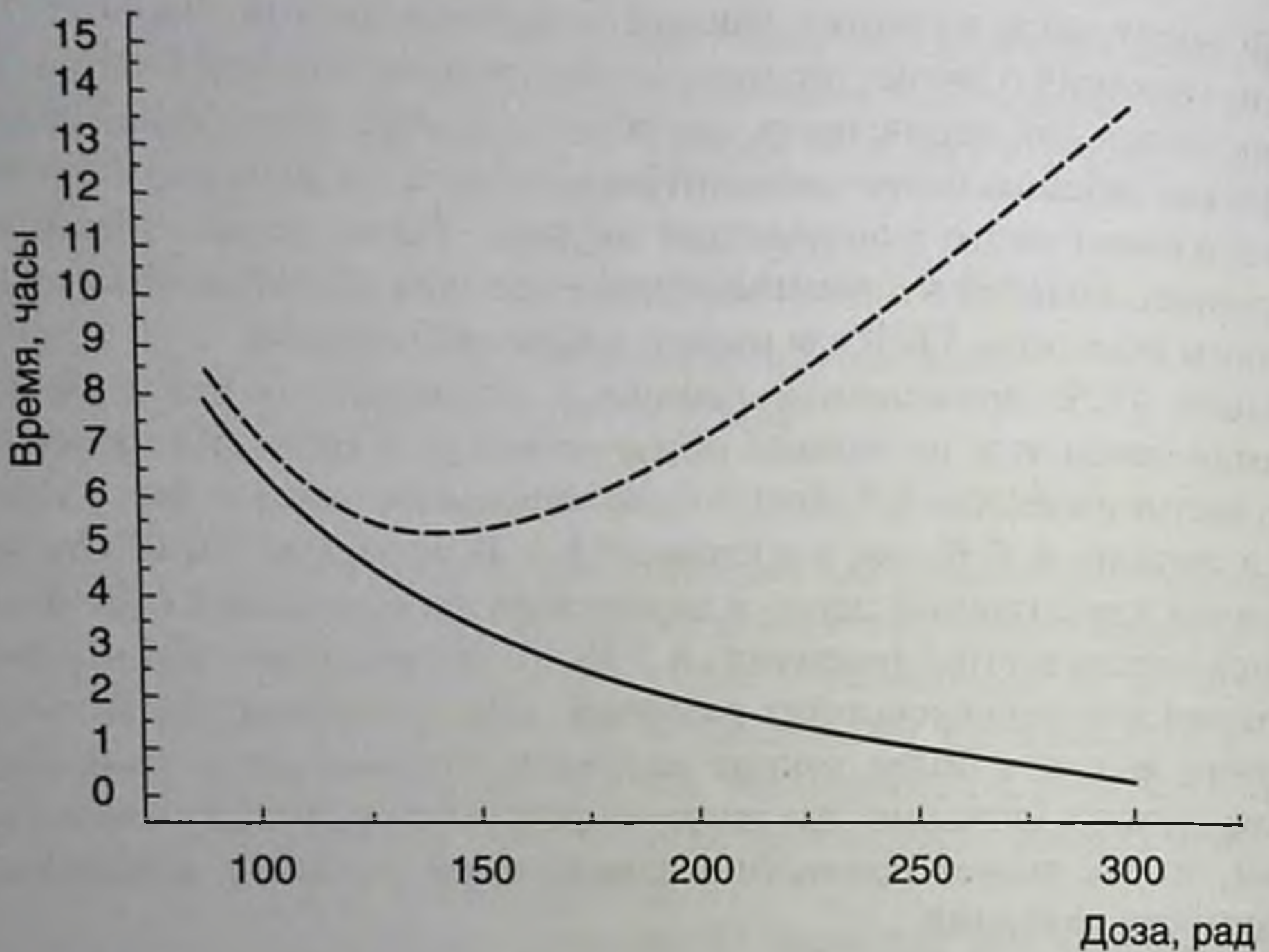
ТКФ носит такое название, так как опирается на минимальное число предположений о форме распределения и независимости данных. ТКФ вычисляется как вероятность случайного распределения наблюдений в ячейках таблицы более сконцентрированного по диагонали согласия, чем это имеет место в полученной таблице. Такой метод сложен в вычислениях, поэтому в докомпьютерные времена доступны были только таблицы величины ТКФ для малого числа наблюдений.

Ныне ТКФ применяется широко в медицинских исследованиях. Он отличается от  $\chi^2$  по области применения.  $\chi^2$  — критерий на неравенство частот в ячейках таблицы, т.е. *двусторонний*, он выявляет различия как в сторону  $A < B$ , так и в сторону  $A > B$ . Не таков ТКФ. Он имеет различия для ситуаций одно- и двустороннего оценивания. Если ожидаются определенные различия ( $A > B$ ), то должна применяться форма критерия для односторонних различий. Она позволяет статистически значимо выявить более тонкие различия. Однако если такая форма применяется в ситуации, где должен применяться двухсторонний критерий, то это может привести к завышенной оценке статистической значимости различий.

### 2.4.3. Последствия внешнего воздействия

Последствия внешнего воздействия (экспозиции) зависят от его специфических особенностей и от силы. Под силой воздействия обычно понимают дозу, т.е. количество попавшего в организм вещества, поглощенную энергию. Дозиметрия представляет собой сложную проблему, которая по-разному разрешается в различных областях биологии и медицины. В фармакологии и токсикологии изучается не только зависимость эффекта введения вещества в организм от дозы, но и от того, насколько полно в конкретных обстоятельствах это вещество поступает в организм (всасывается из желудочно-кишечного тракта) и насколько оказывается способным к взаимодействию с рецепторами (биодоступность). При дозиметрии механических воздействий (например, вибрации) и электромагнитных излучений изучается не только «приложенная» доза, но и характер ее распределения в организме, поглощения в поверхностных и глубоких тканях. Именно поэтому на рис. 2.16 данные о развитии первичной реакции на рентгеновское облучение приведены в зависимости от дозы, отнесенной к поверхности тела.





**Рис. 216.** Развитие во времени первичной реакции на рентгеновское облучение. Абсцисса — доза в рад, отнесенная к поверхности тела. Линия — начало первичной реакции, пунктир — окончание (по данным Коурт-Браун В. и Аббатт Дж. Вопросы радиобиологии. М. Изд-во иностр. лит., 1956. С. 367–70)

Такие данные нельзя было бы получить при изучении случаев облучения людей при авариях. Оценки полученной при авариях дозы настолько приблизительны, что, наоборот, нередко полученную дозу определяют по изменениям, которые наблюдаются в организме облученных (биологическая дозиметрия). Приведенные данные получены, конечно же, не на добровольцах — ввиду очевидного вреда облучения такого рода эксперименты невозможны. Эти данные получили, подвергая облучению позвоночник у больных анкилозирующим спондилитом. С момента открытия рентгеновского излучения его применяли для разных целей и при разных болезнях. В лечении анкилозирующего спондилита облучение применяли широко, но эта практика не имела оснований и была позднее прекращена. Благодаря ей имеются настоящие данные, хотя дозиметрически и несовершенные. Они иллюстрируют фундаментальную закономерность: чем больше сила (доза) воздействия, тем раньше развивается реакция на него и тем дольше реакция длится.



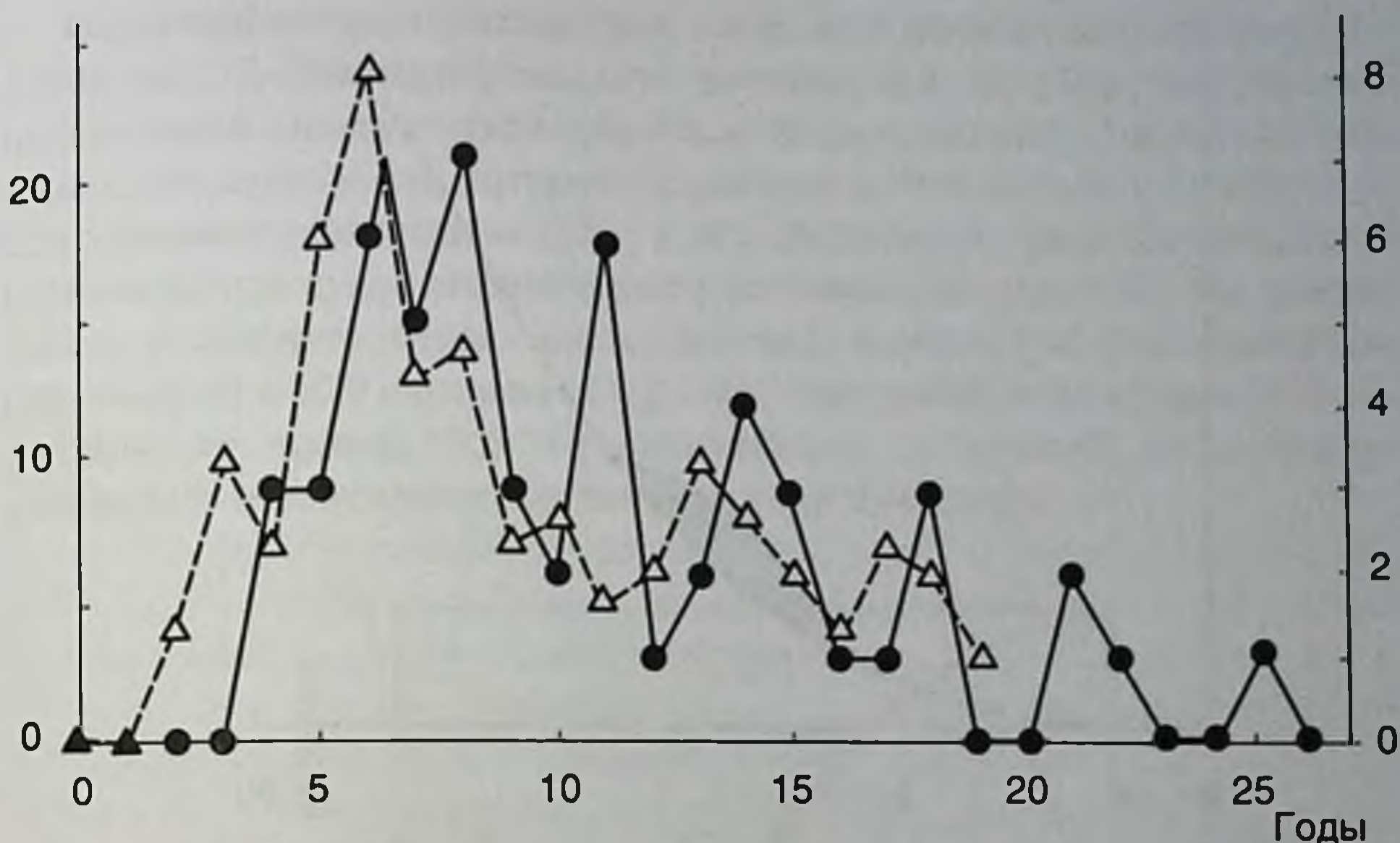


Рис. 2.17. Распределение во времени числа случаев лейкоза у переживших атомную бомбардировку (треугольники, шкала слева) и остеосарком у подвергшихся действию короткоживущего изотопа радия (кружки, шкала справа) (по данным Mays C.W., Spiess H. in *Radiation cancerogenesis: Epidemiology and biological significance*, New York, Raven Press, 1984. P. 241–52)

В последствиях внешнего воздействия сочетаются специфические для него и неспецифические последствия. Если специфический исход воздействия — опухоль, то ее возникновение требует длительного времени, обычно месяцы и годы. Поэтому развитие опухоли относительно постоянно во времени проявления после воздействия. Тем не менее и применительно к таким медленно развивающимся исходам, как развитие опухоли, дозовая зависимость существует: чем выше доза, тем раньше и с большей частотой развиваются опухоли. После однократного воздействия можно наблюдать, как частота болезней увеличивается и затем снижается (рис. 2.17). Эти экзогенные случаи заболевания накладываются на естественную возрастную кривую частоты заболеваний (рис. 2.18).

При длительном воздействии, как это имеет место у работников вредного производства, инцидент остается высоким в течение всего времени контакта, например с ионизирующим излучением, и годы после него. Некоторые воздействия даже после однократной экспозиции вызывают повышение вероятности заболевания на долгие годы (рис. 2.19).



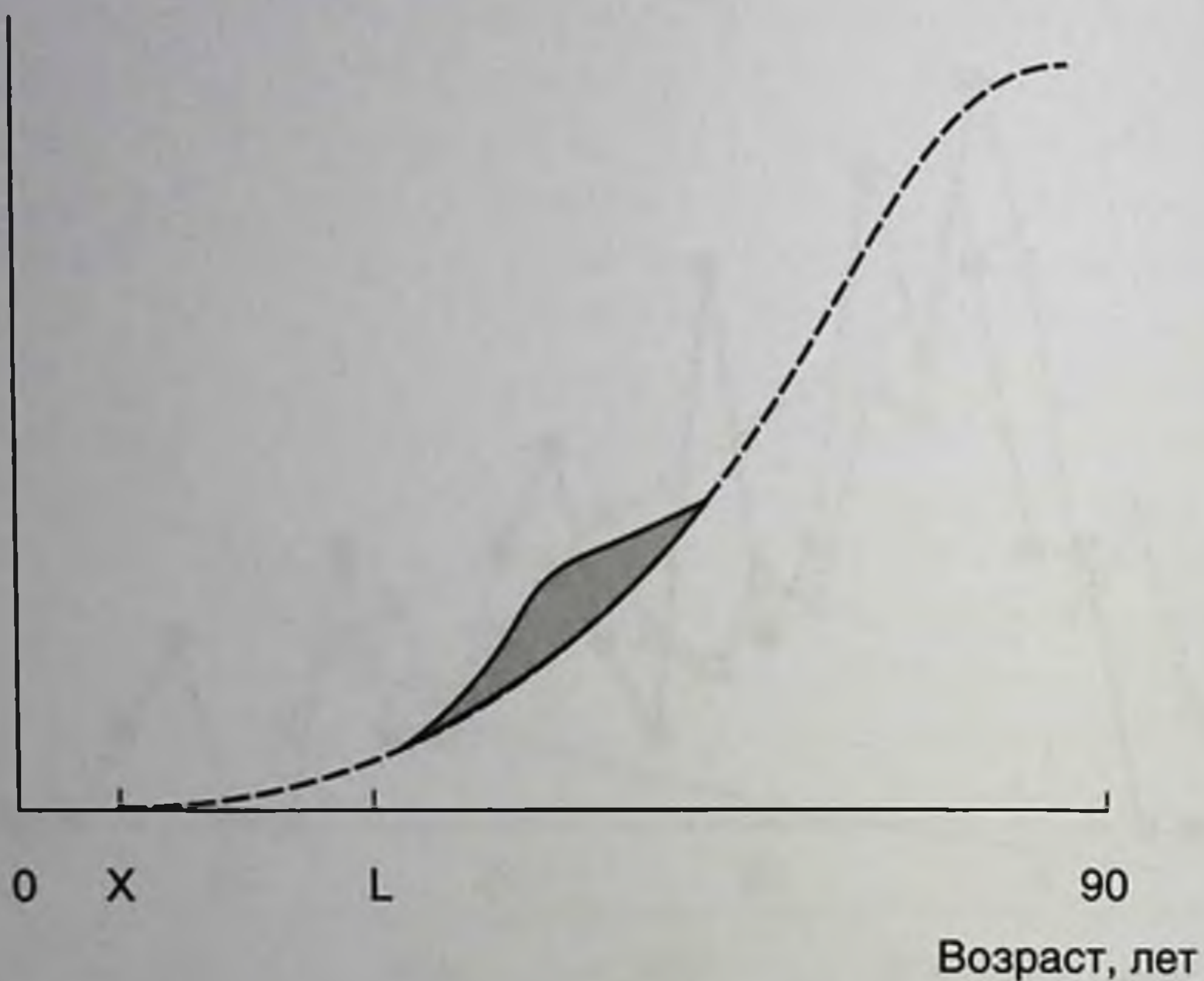


Рис. 2.18. Кратковременное повышение риска заболевания после воздействия (X). Пунктир — естественный инцидент болезни, линия — инцидент после воздействия (облучения). L — латентный период

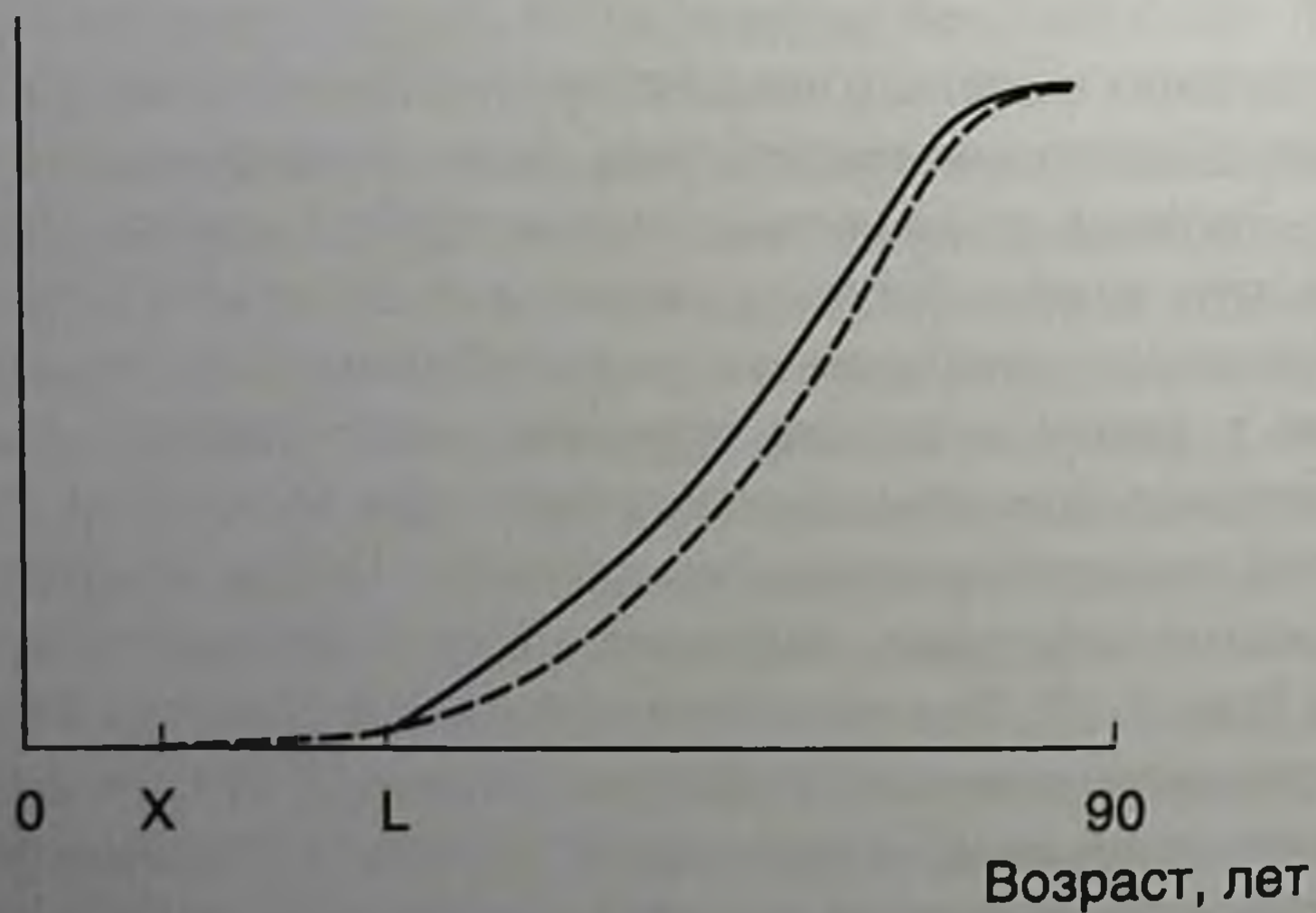


Рис. 2.19. Пожизненное повышение риска заболевания после воздействия (X). Пунктир — естественный инцидент болезни, линия — инцидент после воздействия (облучения). L — латентный период



Неспецифические последствия наступают раньше при сильных воздействиях. Зависимость частоты последствий от силы воздействия на линейной шкале имеет S-образную форму, которая хорошо изучена в токсикологии и для которой разработаны специальные статистические приемы ее описания (рис. 2.20). Эта кривая представляет собою разновидность кумулятивного распределения (см. рис. 1.32). На кривой можно установить, какая сила воздействия вызывает эффект (заболевание, смерть) в 50% случаев ( $LD_{50}$ ). Для этого достаточно провести горизонталь от уровня 0,5 до пересечения с кривой доза–эффект и от пересечения опустить перпендикуляр к абсциссе.

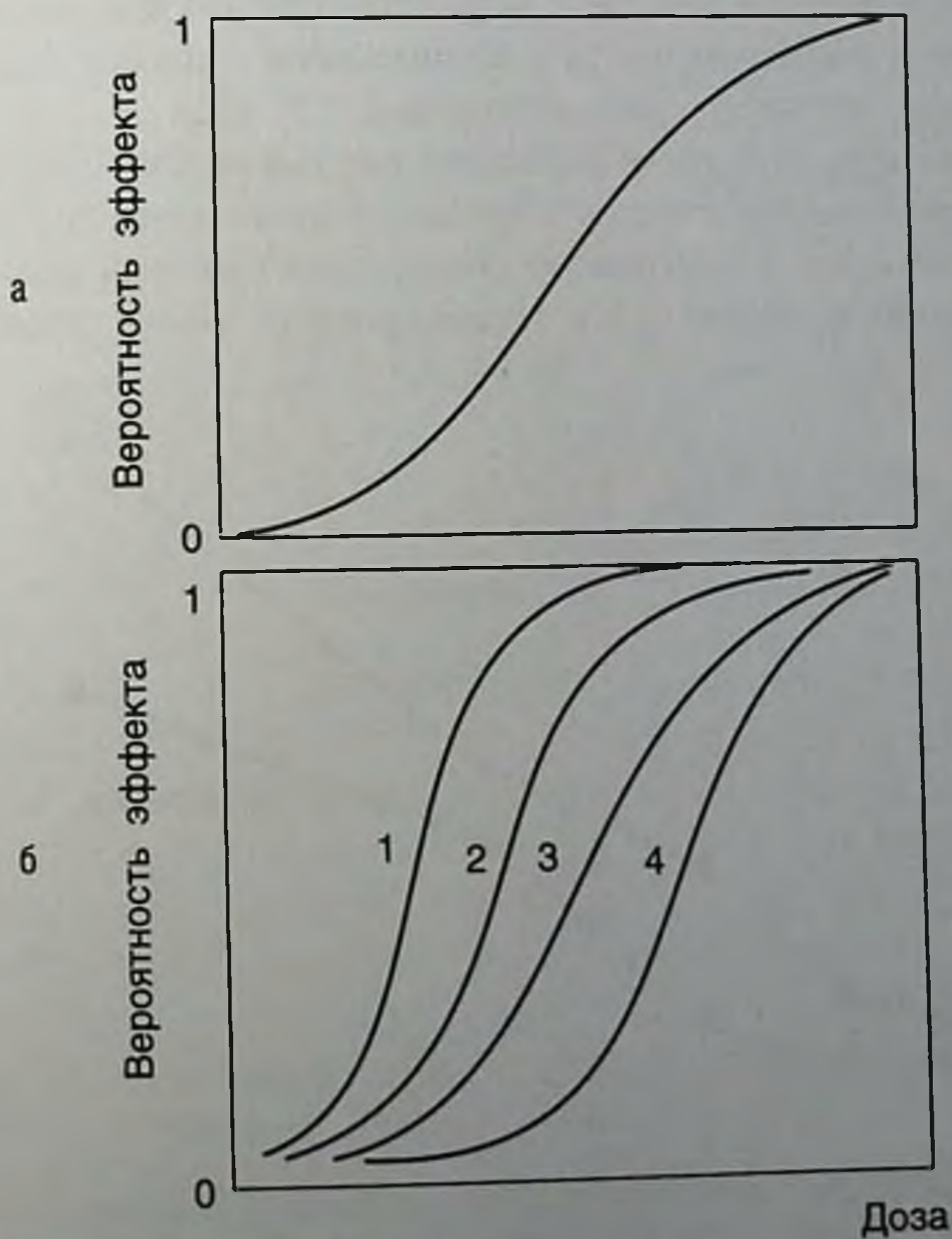


Рис. 2.20. Форма зависимости частоты эффекта от силы воздействия (а) и соотношения таких зависимостей для разных эффектов воздействия (б)



Обычно воздействие вызывает не один, а несколько эффектов, например кожные поражения, диарею, головокружение. Для разных эффектов воздействия характерно свое время развития, и каждый из эффектов проявляется во времени с той же закономерностью (рис. 2.20, б, кривые 1–4). Точно так же происходят изменения эффекта воздействия при изменении состояния организма. У лиц с более высокой резистентностью последствия возникают при более высоких дозах: S-образная форма зависимости сохраняется, но кривая смещается вправо, к более высоким дозам. S-образную форму можно объяснить как проявление того, что в популяции мало лиц с очень низкой и очень высокой устойчивостью. Такие кривые могут быть превращены в прямые, если значения по ординате (функция или «ответ» на дозу, силу воздействия) трансформировать в соответствии с плотностью вероятности нормального распределения (см. рис. 1.51 (г)).

Поскольку при слабом воздействии патологические последствия, например, опухоли наблюдаются редко, чрезвычайно трудно составить представление о том, какую форму имеет кривая доза–эффект в области малых доз (рис. 2.21). Использование уникальных данных,

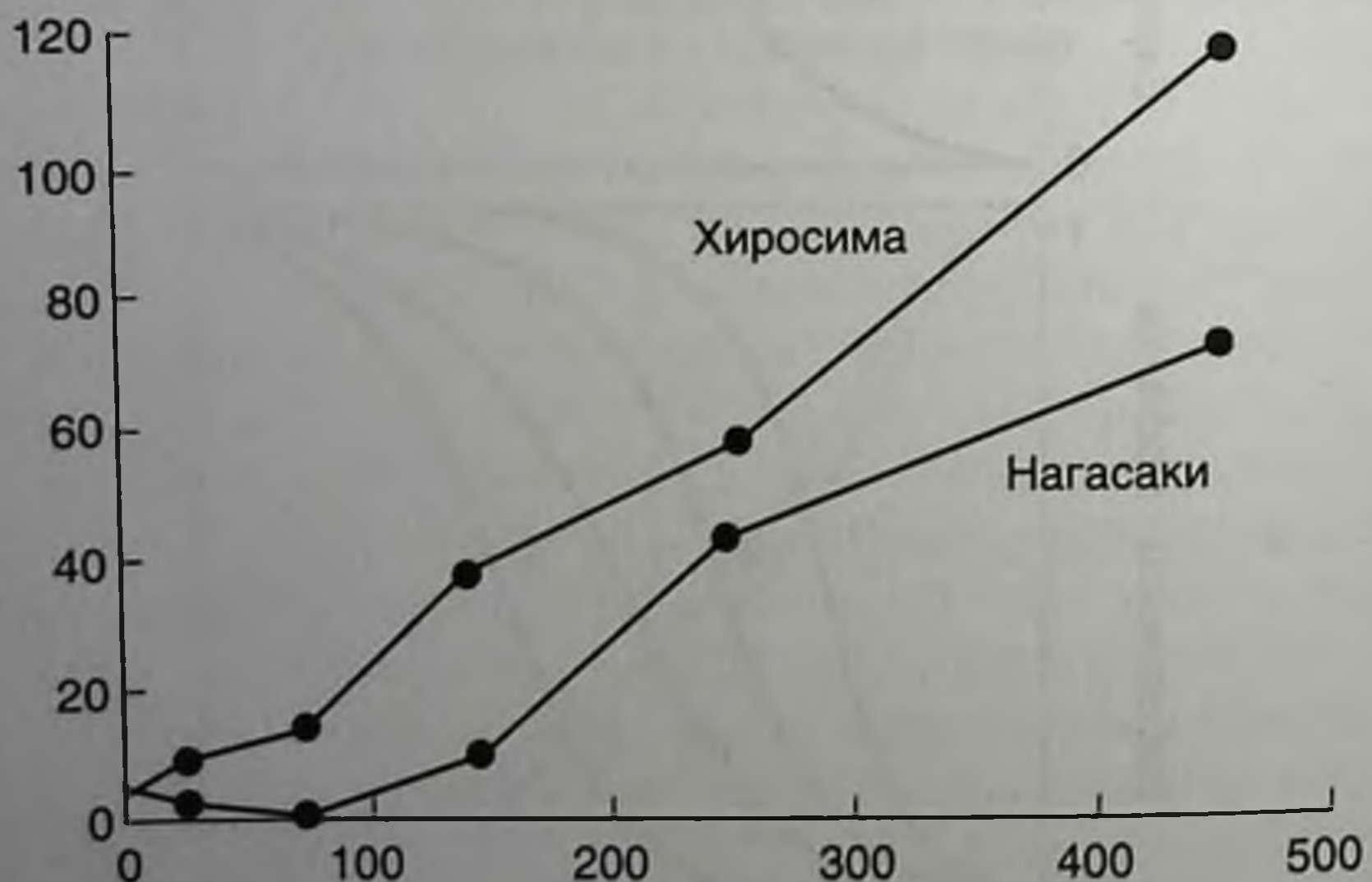


Рис. 2.21. Зависимость частоты лейкозов от дозы облучения у переживших атомную бомбардировку японцев. Абсцисса — доза, рад; ордината — инцидент лейкозов на 10 000 человек в год (по данным Beebe G.W. et al. Mortality experience of atomic bomb survivors 1950–1974, Life span Study Report 8 (Techn. Rep. REFR TR 1-77). Hiroshima, Radiation Effects Research Foundation, 1977)



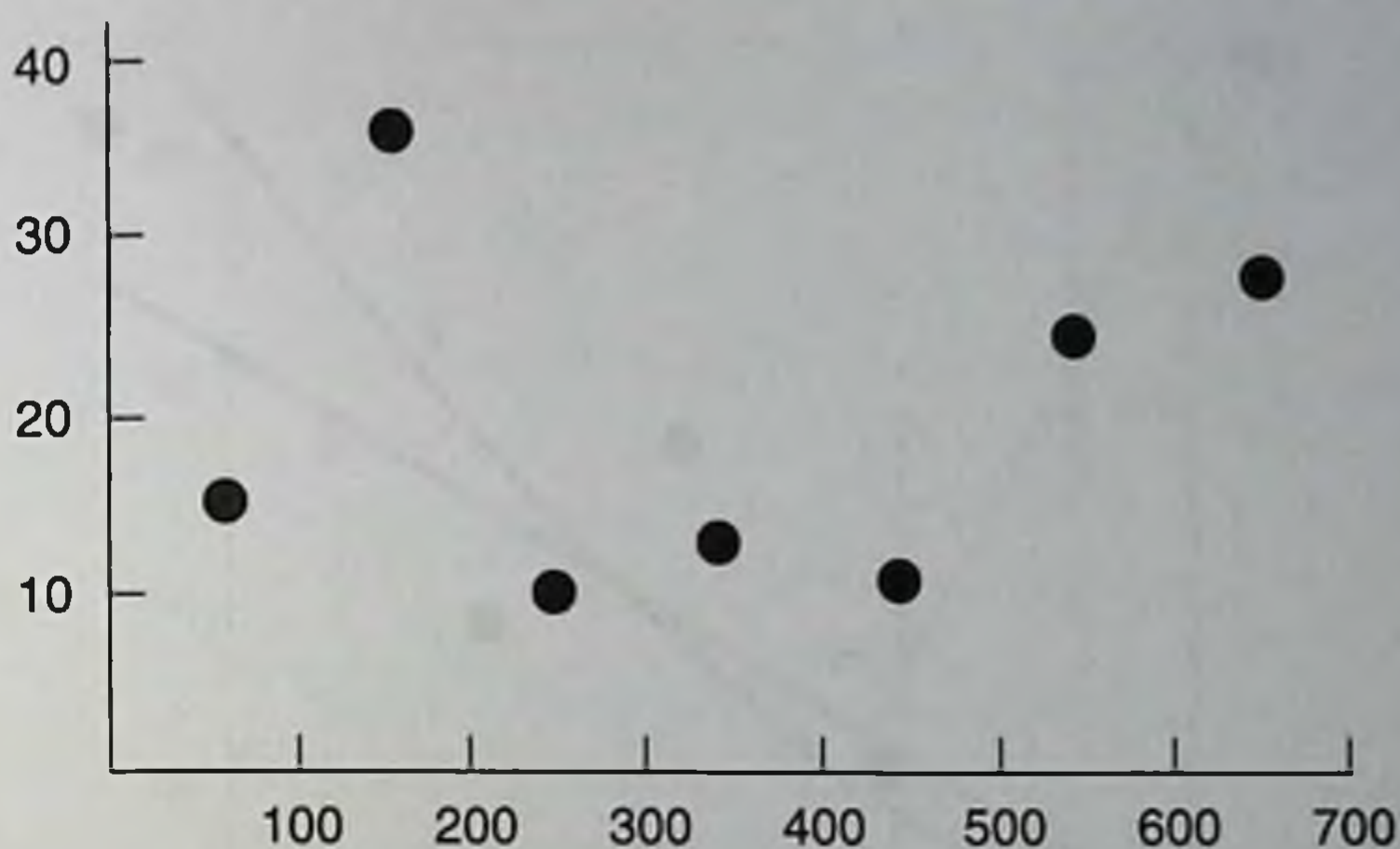


Рис. 2.22. Зависимость инциденса лейкозов от дозы облучения. Абсцисса — костномозговая доза, рад; ордината — избыточная смертность от лейкозов, на 10 000 человеко-лет (по данным Smith P.G., Doll R.. BMJ. 1982. № 284. P. 449–60)

полученных при наблюдении за пациентами, подвергшимися рентгено-терапии, показывает, что при облучении в относительно малых дозах форма зависимости не очевидна (рис. 2.22).

Низкая точность оценок величины эффекта (частоты патологических последствий) при воздействиях малой интенсивности является неизбежным следствием редкости патологических последствий слабых воздействий. В условиях такой неопределенности приходится выбирать, какую модель (форму зависимости) использовать для описания имеющихся данных, и этот выбор оказывается в известной степени произвольным (рис. 2.23). Для нормирования (установления предельно допустимых доз облучения) применительно к низким дозам ионизирующей радиации используется линейная функция, поскольку она предсказывает более высокую частоту эффектов. Выбор этой математической модели предохраняет от излишнего «оптимизма». Некоторые из исследований достаточно согласованно свидетельствуют в пользу квадратического характера зависимости (рис. 2.24), но в нормировании соображения надежности нормативов защиты оказываются преобладающими, и выбирается линейная зависимость.

В отношении вредных последствий курения существуют надежные данные (рис. 2.25), и квадратический характер зависимости риска рака легких от дозы выкуриваемого табака не вызывает особых сомнений.



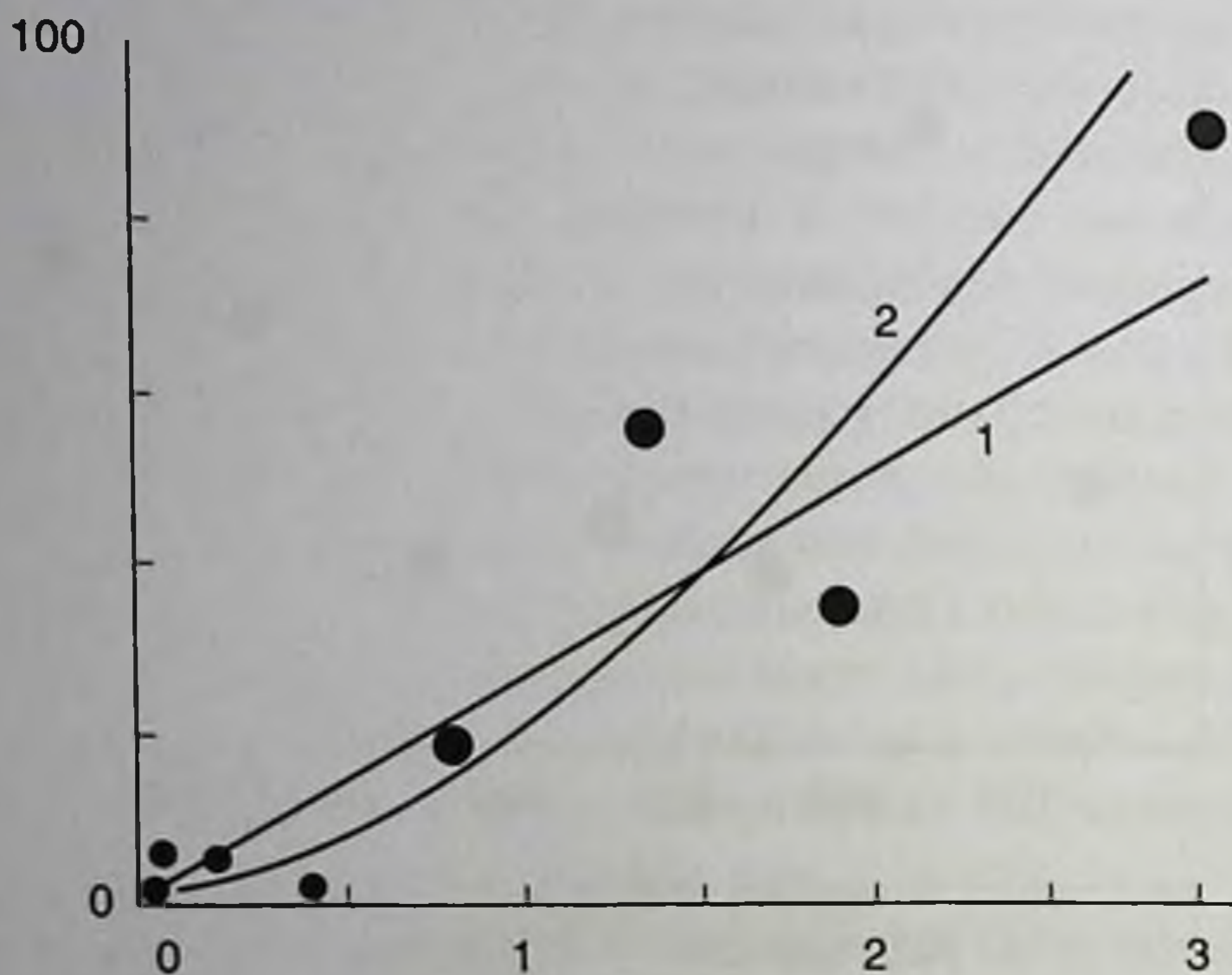


Рис. 2.23. Зависимость инцидента лейкозов на 100 000 человеко-лет от годовой костномозговой дозы поглощенной радиации (рад, абсцисса). Экспериментальные точки могут быть с равным успехом аппроксимированы прямой (1) и квадратичной (2) функциями (см. разд. 2.8.2) (по данным Land C.E. Science. 1980. № 209. P. 1197–203)

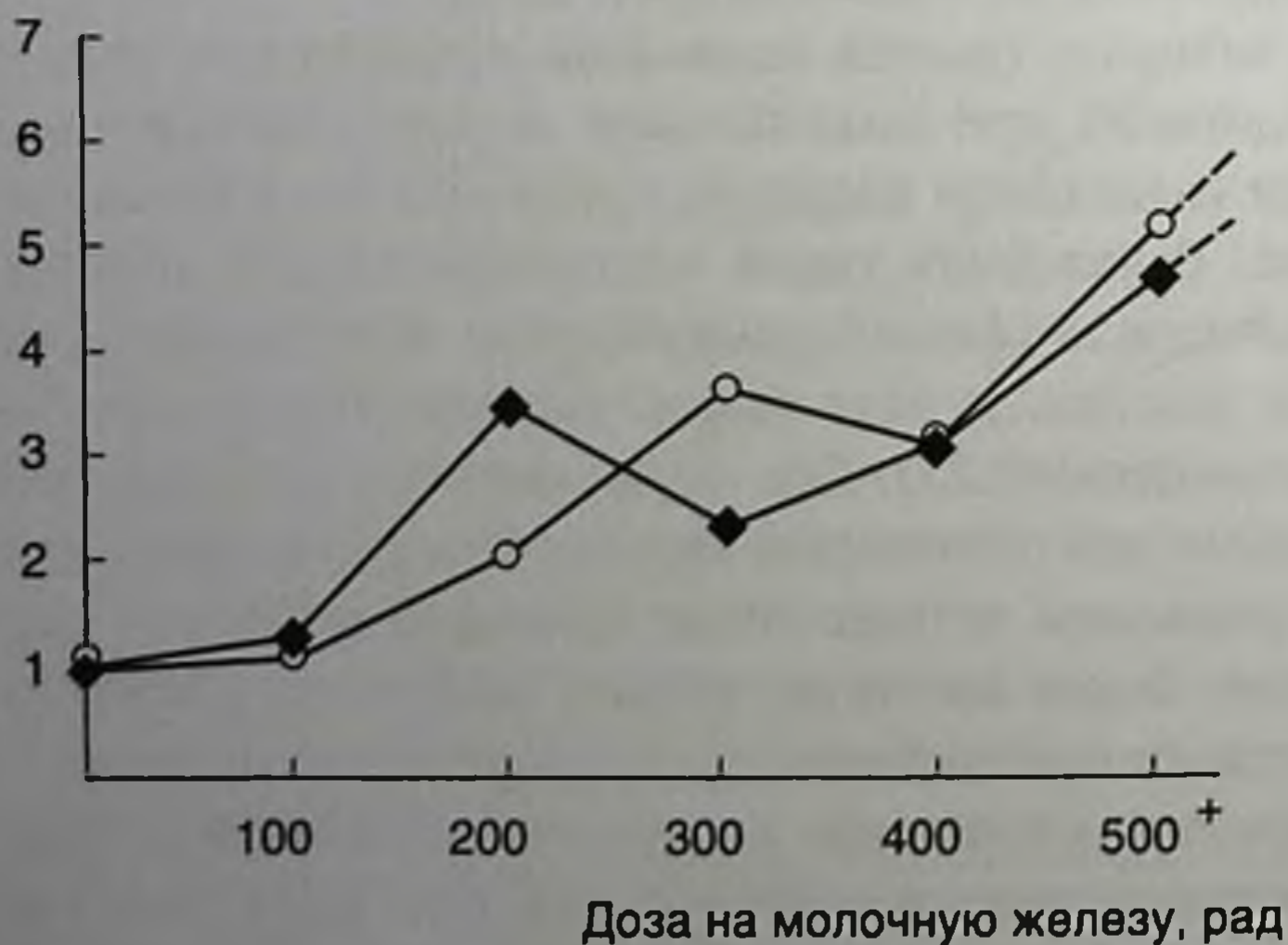


Рис. 2.24. Зависимость частоты рака молочной железы (ордината, относительный риск) от дозы, поглощенной молочной железой, у переживших атомную бомбардировку (ромбы) и по данным Канадского исследования флюорографии (кружки) (по Breslow N.E., Day N.E. Statistical Methods in Cancer Research. Vol. 2. 1987. Lyon, WHO, IARC. IARC Sci. Publ. N 82; с разрешения издателя)



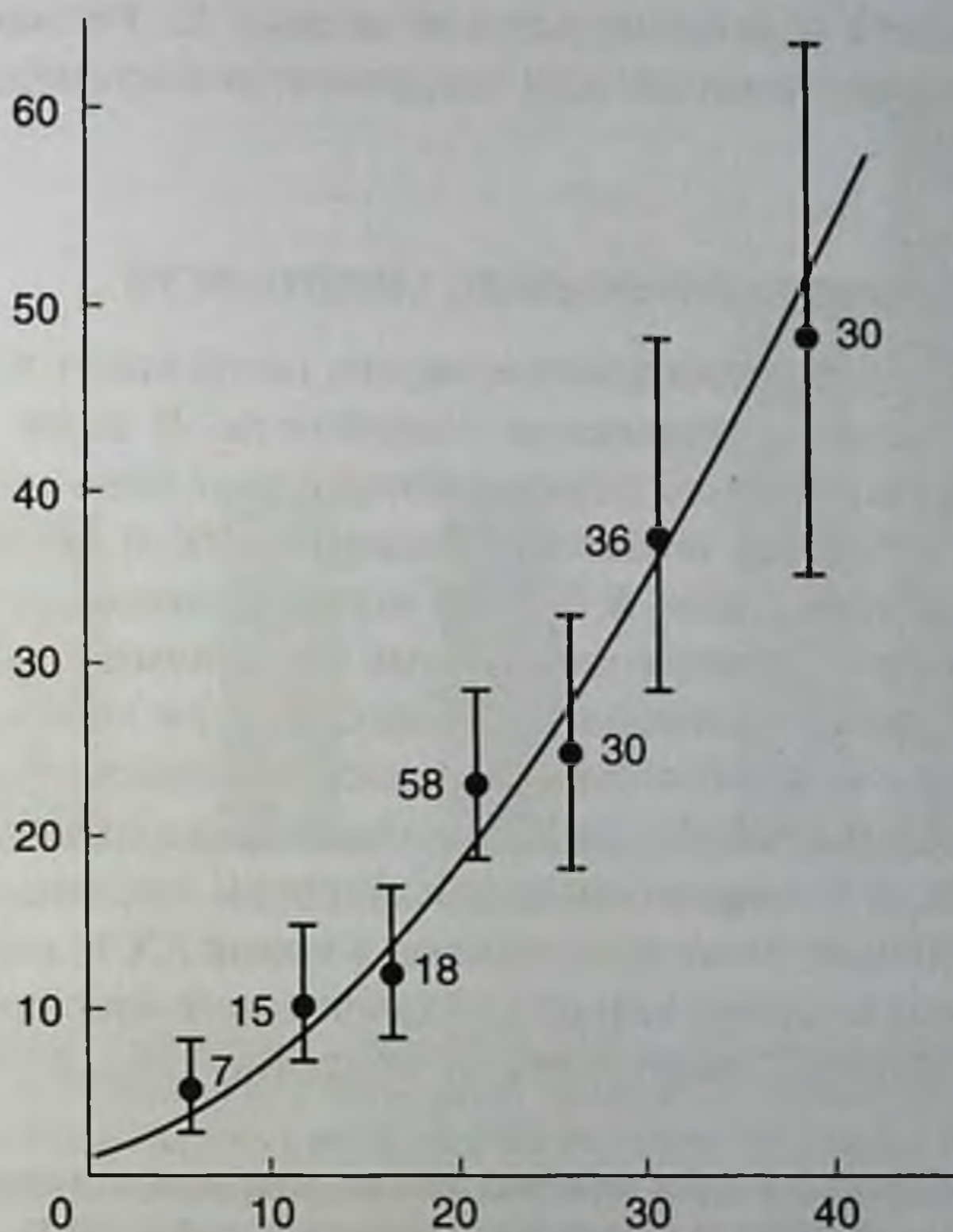


Рис. 2.25. Относительный риск рака легких (ордината) в зависимости от числа выкуриваемых в день сигарет (абсцисса). Число случаев рака приведено около кривой (по данным Doll R., Peto R. J. *Epidemiol. Commun. Health*. 1978. № 32. P. 303–13)

Для интерпретации зависимости доза–эффект возможны два подхода. Первый — вероятностный, объясняет возникновение болезни у одних организмов и невозникновение у других игрой случая. Второй — детерминистский. Он предполагает, что при слабом воздействии болезнь не возникает потому, что при действии такого причинного фактора необходимы дополнительные причинные факторы, которые встречаются редко или редко имеют достаточно большую силу. При этом подходе случайность есть лишь обозначение для причинных факторов, которые нам неизвестны. Все же для случайности в объяснениях непредсказуемости событий остается особое место. Это хаотические изменения в микроструктурах, которые могут приводить к неопределенным последствиям. Такой случайный механизм может объяснять



непредсказуемость в детерминистской модели К. Ротмана, базирующейся на представлениях об игре компонентов достаточной причины (см. разд. 1.1.1).

#### 2.4.4. Многолетние изменения смертности

До настоящего времени в человеческих популяциях по всему миру происходят огромные изменения смертности. В мире завершается *эпидемиологический переход* (англ. *epidemiological transition*) — переход от высокой смертности и высокой рождаемости к низкой смертности и низкой рождаемости. В бедных странах ранняя смертность обусловлена в основном инфекционными болезнями, высока детская смертность. С ростом социального благополучия начинает преобладать смертность от хронических болезней преимущественно в старшем возрасте (демографический переход), снижается фертильность, заболеваемость. В бедных странах субсахарской Африки (к югу от Сахары) и в Восточной Азии этот переход в конце XX в. сопровождался весьма умеренным снижением общей смертности: при высокой общей смертности нарастала смертность от сердечно-сосудистых болезней и СПИДа.

В развитых странах при продолжающемся снижении смертности происходит трансформация причин смерти (см. рис. 1.17). Для этих изменений предложено множество объяснений, но уверенности в их правильности нет, и нет эффективных средств противостоять негативным тенденциям.

Основным подходом к интерпретации изменений во времени является поиск совпадающих по времени с заболеваемостью событий, которые могут объяснить заболеваемость. Естественно, поиск происходит в том направлении, которое подсказывают знания об изучаемой болезни. Например, в середине XX в. исторический параллелизм распространенности курения и инцидента рака легких стал важнейшим основанием для связи рака легких и курения (см. рис. 2.25). Применительно к смертности от сахарного диабета убедительных объяснений нет: нарастание смертности от диабета не может быть полностью объяснено тем, насколько снизилась смертность от сердечно-сосудистых болезней и инсульта (см. рис. 1.17). Высказывается патогенетически правдоподобная гипотеза, что нарастающая смертность от диабета связана с нерациональным питанием, снижением физических нагрузок



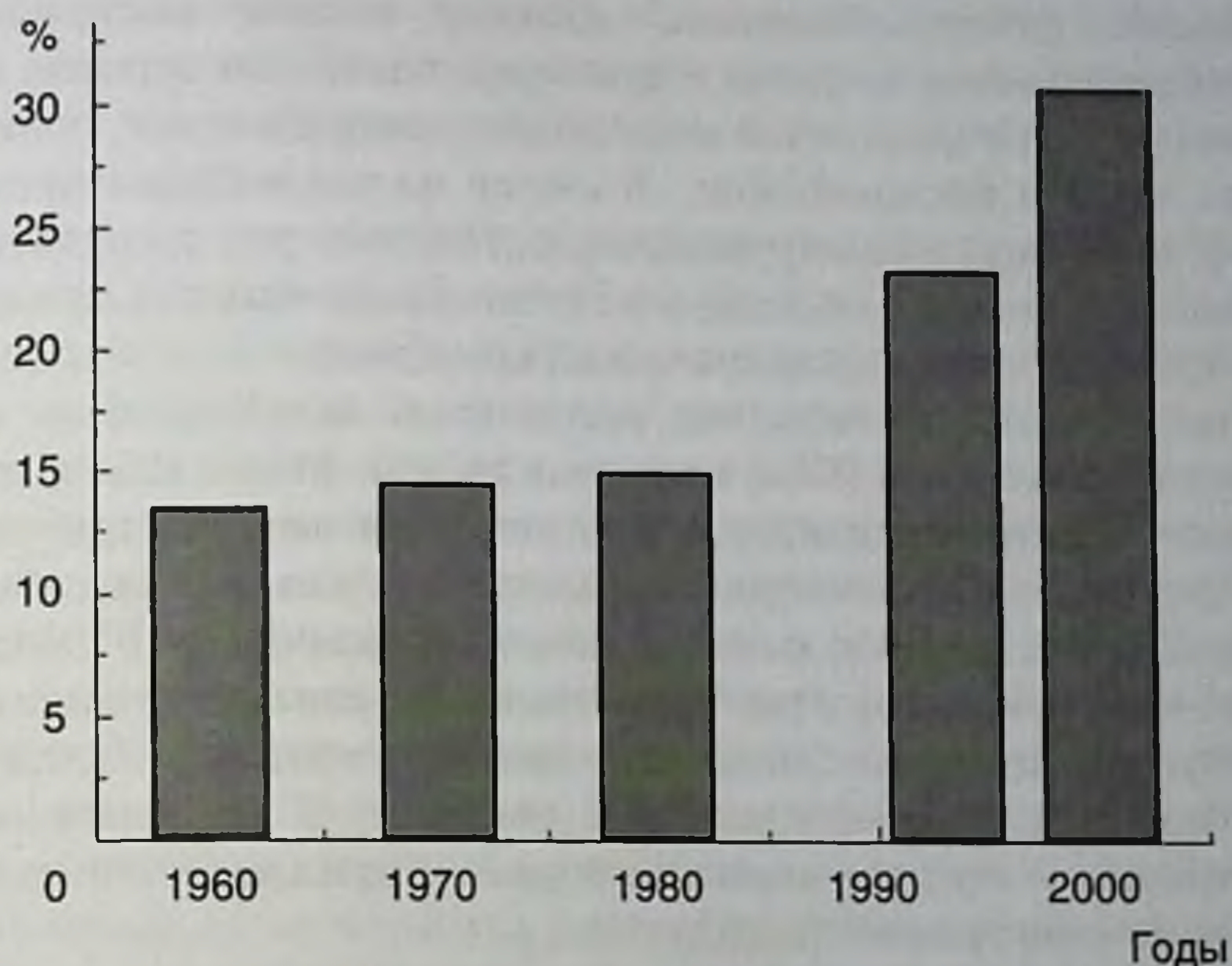


Рис. 2.26. Преvalенс ожирения в популяции США у мужчин и женщин 20–74 лет (в процентах), скорректировано на различия по возрасту. Данные пяти NHANES — повторяющихся национальных исследований состояния здоровья и питания. Диагностический критерий ожирения — индекс массы тела более 30 кг/м<sup>2</sup> (по данным Flegal K.M. et al. JAMA. 2002. № 288. P. 1723–7)

и их хорошо известным результатом — ожирением. Ожирение становится все более распространенным в экономически развитых странах в конце XX в. (рис. 2.26). Эта гипотеза не является вполне убедительной: в эти же годы смертность от инсульта и сердечно-сосудистых болезней снижается, хотя и гипертензия и дислипотеидемии, ускоряющие развитие атеросклероза, учащаются при ожирении.

#### 2.4.5. Критерий знаков, критерий Вилкоксона (W) и критерий Манна–Уитни (U)

При выявлении различий между выборками с двух территорий, двумя группами больных или при сопоставлении результатов измерений, выполненных с некоторым интервалом времени, возникает проблема оценки статистической значимости обнаруженных различий. Для этих целей может быть использовано множество методов.



Наиболее распространенным методом оценки статистической значимости различия является  $t$ -критерий, однако он основан на использовании закономерностей нормального распределения, что в большинстве случаев несправедливо. В случае малых выборок проверить характер распределения затруднительно. Поэтому при распределении, отличающемся от нормального, и в случае малых выборок предпочтительно применение непараметрических критериев.

Соответственно, сначала мы рассмотрим использование критерия знаков Вилкоксона ( $W$ ) и критерия Манна–Уитни ( $U$ ), а далее —  $t$ -критерий. Критерии знаков,  $W$  и  $U$  являются непараметрическими. Они наилучшим образом приспособлены для анализа ординальных переменных, т.е. данных, которые можно ранжировать. В случае количественных признаков с распределением, не соответствующим нормальному, эти критерии применяют вместо  $t$ -критерия.  $U$ -критерий используют в случае независимых выборок, а критерий знаков и  $W$ -критерий — при связанных выборках и при сравнении выборки с популяционной средней (стандартом).

### Критерий знаков

Если выборка получена из популяции, то в выборке половина значений будет меньше медианы в популяции, а половина — больше. Этот простой признак может быть использован для решения вопроса о принадлежности группы случаев к популяции или к другой группе, медиана в которой известна. Для этого число результатов, равных популяционной медиане, вычитается из численности выборки, получается новая численность  $n'$ . Подсчитываем число значений больше медианы (+) и меньше медианы (—). Из двух полученных чисел меньшее  $r$  используем для расчета величины критерия. Если это меньшее число достаточно мало, то гипотезу о соответствии выборки популяции с данной медианой можно отклонить.

$$z = \frac{|r - n'/2| - 1/2}{\sqrt{n'/2}}$$

Если выборка больше 10 ( $n' > 10$ ), то критическое значение для  $z$  получаем в приложении 1. Если выборка меньше 30, то оцениваем меньшую сумму непосредственно по приложению 7.

Критерий можно применять для связанных совокупностей. Применение основано на простом правиле, понятном из обыденного на-



блюдения: если при многократных испытаниях средство А оказывается лучше, чем средство В в большинстве случаев, то, вероятно, оно действительно лучше. Нулевая гипотеза состоит в том, что средства А и В действуют одинаково.

Представим, что в испытании этих двух средств пациенты были подобраны в пары равной тяжести болезни, и затем с помощью таблицы случайных чисел лечение А или В предоставлялось одному и другому. Эта структура исследования является вариантом связанных выборок. По истечении установленного срока состояние пациентов повторно оценили. В случае если лечение А дало более выраженный эффект, результат обозначается (-), если лечение В дало более выраженный эффект, то (+).

Число (+) и (-) подсчитывается. В нашем примере 1(+) и 8(-). В одной из пар различия не выявилось. Осталось 9 пар. Меньшее из чисел сравнивается с критическим значением в приложении 7. В нашем примере впечатление о большей эффективности вмешательства А подтвердилось, статистически значимое различие выявлено *по критерию знаков*.

Описанная выше методика соответствует двустороннему варианту теста, т.е. вывод о преимуществе одного из испытываемых вмешательств делается на основании результатов эксперимента. Если рассматривается один определенный вариант исхода, то с табличными величинами сравнивается число знаков, противоположных этому варианту. Например, если оценивается, действительно ли лучше лечение А, то число (+) сравнивается с табличным. Если оценивается, лучше ли лечение В, то с табличным значением сравнивается число (-).

Точно так же можно с помощью критерия знаков оценивать статистическую значимость изменений некоторого признака во времени в одной группе (например, изменим представление данных табл. 2.8 таким образом, что А — состояние до лечения, В — после лечения) (табл. 2.9).

Таблица 2.8. Сравнение двух вмешательств в подобранных парах пациентов по функциональным возможностям пациентов

Лечение	Возможности пациентов, условные единицы									
	1	2	3	4	5	6	7	8	9	10
А	16,2	15,2	17,8	17,3	16,1	12,7	13,3	18,2	18,3	14,4
В	14,1	14,2	16,3	16,6	16,1	14,3	12,5	16,7	15,1	11,0
Оценка	-	-	-	-		+	-	-	-	-



Таблица 2.9. Функциональные возможности пациентов до и после лечения

Лечение	Возможности пациентов, условные единицы									
До	16,2	15,2	17,8	17,3	16,1	12,7	13,3	18,2	18,3	14,4
После	14,1	14,2	16,3	16,6	16,1	14,3	12,5	16,7	15,1	11,0
Оценка	-	-	-	-		+	-	-	-	-

Интерпретация теста проста: если из 9 пациентов у 8 после лечения снизилась концентрация метаболита ХХ, то вероятность случайного характера изменений можно отклонить при  $P < 0,05$ . Другой пример: если из 17 пациентов у 14 с началом лечения развилась диарея, то с вероятностью ошибки первого типа меньше 5% можно отклонить гипотезу о случайном характере этого события. Конечно, если с такой же относительной частотой диарея наблюдается в эти дни и у пациентов, подвергающихся другим вмешательствам, то надо исследовать другие гипотезы, помимо предположения о побочных эффектах фармако-терапии.

### Критерий Вилкоксона (W)

Критерий знаков, рассмотренный выше, игнорирует размер изменений. Между тем размер разности величины признака до и после воздействия имеет значение. Точно так же в подобранных парах имеет значение, насколько велики различия в парах пациентов. Различия в связанных выборках, например величину изменений в процессе воздействия, можно оценить с помощью критерия знаковых рангов Вилкоксона.

Таблица 2.10. Функциональные возможности пациентов до и после лечения

Лечение	Возможности пациентов, условные единицы									
До	16,2	15,2	17,8	17,3	16,1	12,7	13,3	18,2	18,3	14,4
После	14,1	14,2	16,3	16,6	16,1	14,3	12,5	16,7	15,1	11,0
Разность	-2,1	-1	-1,5	-0,7	0	1,6	-0,8	-1,5	-3,2	-3,4
Ранг абсолютной величины разности	3	7	4,5	9		6	8	4,5	2	1



Используем в качестве примера данные табл. 2.8. Вычислим для каждой пары значений изменение в процессе лечения (разность) (табл. 2.10). Абсолютные величины разностей ранжируем. Сумма рангов положительных и отрицательных разностей сравниваются. Меньшая из сумм является статистикой критерия Вилкоксона. Ее сравнивают с критическими значениями, приведенными в приложении 8. В нашем примере эта сумма складывается из одного числа и равняется 6. Для числа степеней свободы 9 (случай с нулевой разностью исключается) сумма 6 соответствует  $P < 0,1$ . В случае выборок большего размера статистика критерия вычисляется, и для этого лучше использовать компьютерные программы.

### Критерий Манна–Уитни U

Критерий Манна–Уитни состоит в сравнении двух независимых выборок по рангам входящих в них величин. Существует два варианта критерия, и более сложный вариант U-критерия дан во врезке.

#### U-критерий

Каждый результат одной группы сравнивается с каждым результатом другой группы. Если справедлива нулевая гипотеза, т.е. вмешательства идентичны, то число случаев, когда результаты из группы А будут больше, будет примерно равно числу случаев, когда результаты группы В будут больше. Если альтернативная гипотеза справедлива — лечение А дает более высокие результаты, то будет мало случаев, когда в группе А результаты будут меньше, чем в группе В.

Представим, что мы получили те же результаты, что приведены в табл. 2.8, но они относятся к двум несвязанным группам пациентов, получавшим вмешательства А и В (табл. 2.11). Всего нам надо выполнить  $n_a \times n_b$  сравнений. В этом нам поможет график «стебель и листья». Этот график похож на гистограмму, но рисуется по вертикали. В нашем случае сравниваются две группы и график «дает листья» в две стороны от «стебля». Стебель составляют общие для наблюдений части чисел (в нашем примере — целые части чисел). Отсюда происходит его название «*stem and leaf plot*» (англ.).

Таблица 2.11. Сравнение двух групп пациентов по функциональным возможностям

Лечение	Возможности пациентов, условные единицы									
	16,2	15,2	17,8	17,3	16,1	12,7	13,3	18,2	18,3	14,7
А										
В	14,1	14,2	16,3	16,6	16,1	14,4	12,5	16,7	15,1	11,0



Рассортировав результаты, мы можем сосчитать для каждого случая В число случаев А с меньшими значениями (в табл. 2.12 справа). Величина критерия U будет равна сумме этих рангов:  $2 + 2 + 2 + 3 + 4 + 6 + 6 + 6 = 27$ . Сравнение с критическими значениями (приложение 8, двусторонний тест) приводит нас к выводу об отсутствии статистической значимости различия между вмешательствами А и В. Если бы мы использовали односторонний тест, то статистическая значимость различия была бы достигнута.

Таблица 2.12. Расчет критерия U

А	Вмешательство	Б	В случаях вмешательства Б число случаев вмешательства А, имеющих меньшие ранги							
	11	0	0							
	7	12	0							
	3	13								
	7	14	1	2	4		2	2	2	
	2	15	1				3			
2	1	16	1	3	6	7	4	6	6	6
8	3	17								
3	2	18								

Вариант U-критерия иногда называется критерий суммы рангов Вилкоксона или Манна–Уитни–Вилкоксона. Его используют для сравнения *независимых выборок*. Представим, что мы получили те же результаты, что приведены в табл. 2.8, но они относятся к двум несвязанным группам работников, подвергавшихся условиям труда А и В. В каждой группе 10 человек, и результаты измерений уже ранжированы (табл. 2.13).

Сумма рангов для условия А  $RA = 123,5$  и для условия В  $RB = 86,5$ . Меньшая сумма  $RB$  является значением критерия. Его осталось сравнить с критическим значением. В качестве критического значения используются значения статистики U (приложение 8). Чтобы ими воспользоваться, надо значение критерия  $RB$  перевести в значения U:  $UB = RB - 1/2 \times nB \times (nB + 1)$ . В нашем примере  $UB = 86,5 - 0,5 \times 10 \times 11 = 31,5$ . Сравнение с критическим значением (23) показывает,



что отклонить нулевую гипотезу невозможно: статистически значимого различия между людьми, работающими в условиях А и В, не обнаружено по исследованному признаку.

Таблица 2.13. Расчет критерия суммы рангов Вилкоксона

Ранг	1	2	3	4	5	6	7	8	9	10	11,5	13	14	15	16	17	18	19	20
А			12,7	13,3				14,7		15,2	16,1	16,2				17,3	17,8	18,2	18,3
В	11	12,5			14,1	14,2	14,4		15,1		16,1		16,3	16,6	16,7				

При большом ( $>25$ ) числе пациентов в выборке вычисления на бумаге с калькулятором и оценка вычислений по таблице затруднительны. Обычно таблицы для непараметрических критериев ограничены числом 10–30 единиц в каждой группе.

Как видим, применять критерии такого типа просто. Они не требуют сложных вычислений. Отчасти поэтому они приобрели значительную популярность у врачей и биологов в середине XX в., т.е. до распространения компьютеров. Непараметрические критерии применяются сейчас не очень часто, возможно, в силу их меньшей мощности, ведь при распространении вычислительной техники простота вычислений отходит на второй план. Вероятно поэтому же исследователи иногда не используют такие критерии, хотя они требуются по характеру изучаемых данных (ненормальное распределение интервальных данных, ординальный характер данных).

Критерий знаков очевидным образом обедняет возможности анализа, редуцируя данные до + и –, и имеет низкую мощность. U-критерий является более мощным и в некоторых ситуациях может не уступать самому мощному *t*-критерию. Все же использование U-критерия и других непараметрических критериев применительно к интервальным данным вместо *t*-критерия приводит к потере статистической значимости различий. На нашем примере, использованном для демонстрации применения трех критериев, хорошо видно, что сравнение двух рядов чисел приводит к разным результатам в зависимости от того, рассматриваются ли ряды как связанные или как независимые. В случае когда два ряда предстают как связанные, различия между ними оказываются статистически значимыми даже при применении критерия знаков, имеющего низкую мощность.



### 2.4.6. Временные ряды и прогнозирование

Если исследователь знает, как изменялась некоторая величина, например заболеваемость, в течение предшествующих лет, у него существует возможность с большей или меньшей уверенностью прогнозировать дальнейшие изменения. Такой «инерционный» прогноз достаточно успешен, но лишь в той степени, в какой изучаемая система изменяется по инерции. Его также называют *экстраполяционным*, поскольку закономерности, действовавшие до текущего момента, распространяются (экстраполируются) на предстоящий период, о котором нет сведений.

В качестве примера инерционного прогноза и некоторых статистических приемов, используемых для этого, рассмотрим данные об инфицированности вирусом иммунодефицита человека (ВИЧ) в СССР и России. Отличием деятельности по контролю над инфицированностью ВИЧ в нашей стране является интенсивное использование диагностических обследований населения: добровольных анонимных, обязательных для отдельных категорий работников и для беременных и госпитализированных больных, для доноров крови и органов и для некоторых других граждан.

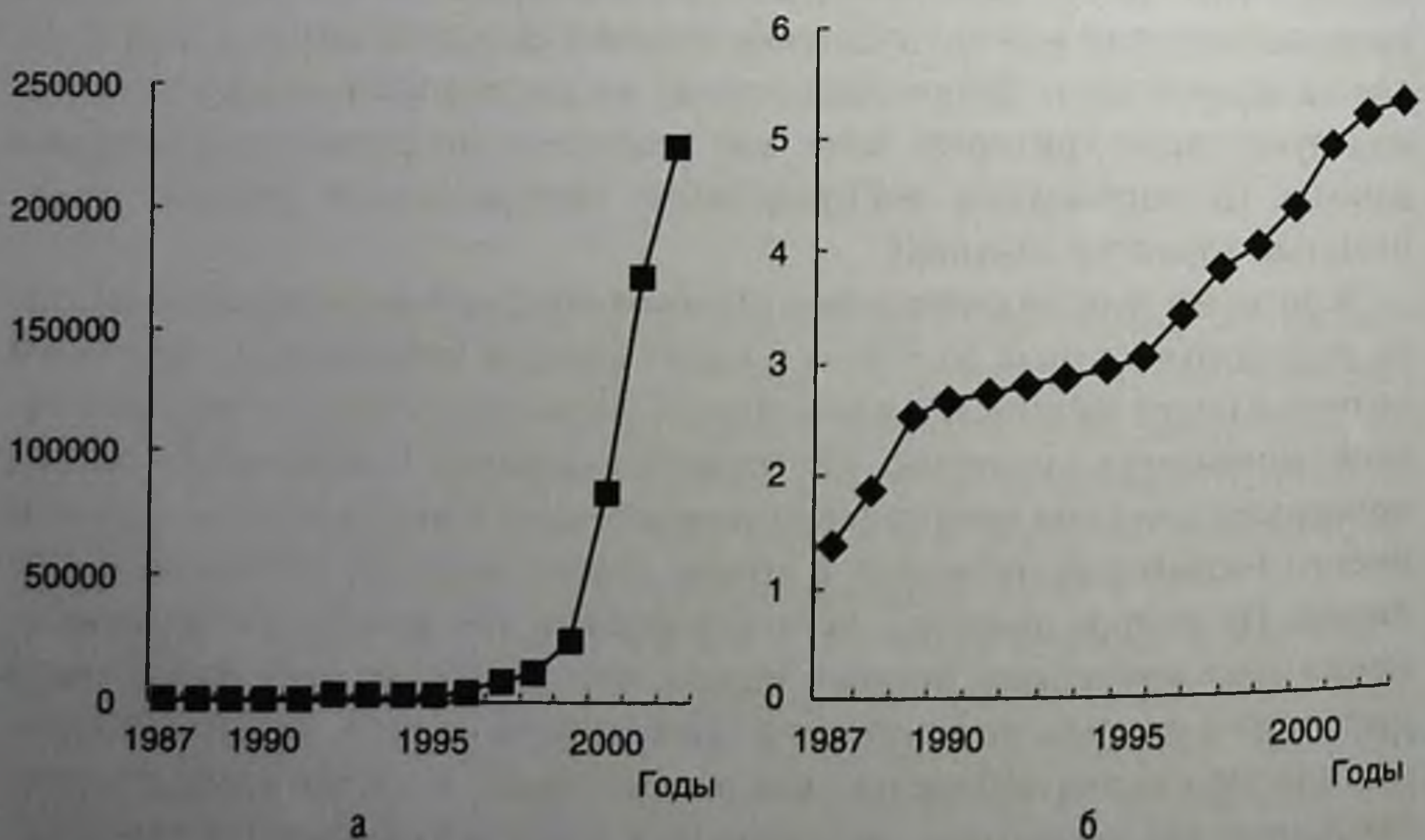


Рис. 2.27. Численность инфицированных ВИЧ в России: а — по ординате линейная шкала; б — логарифмическая шкала



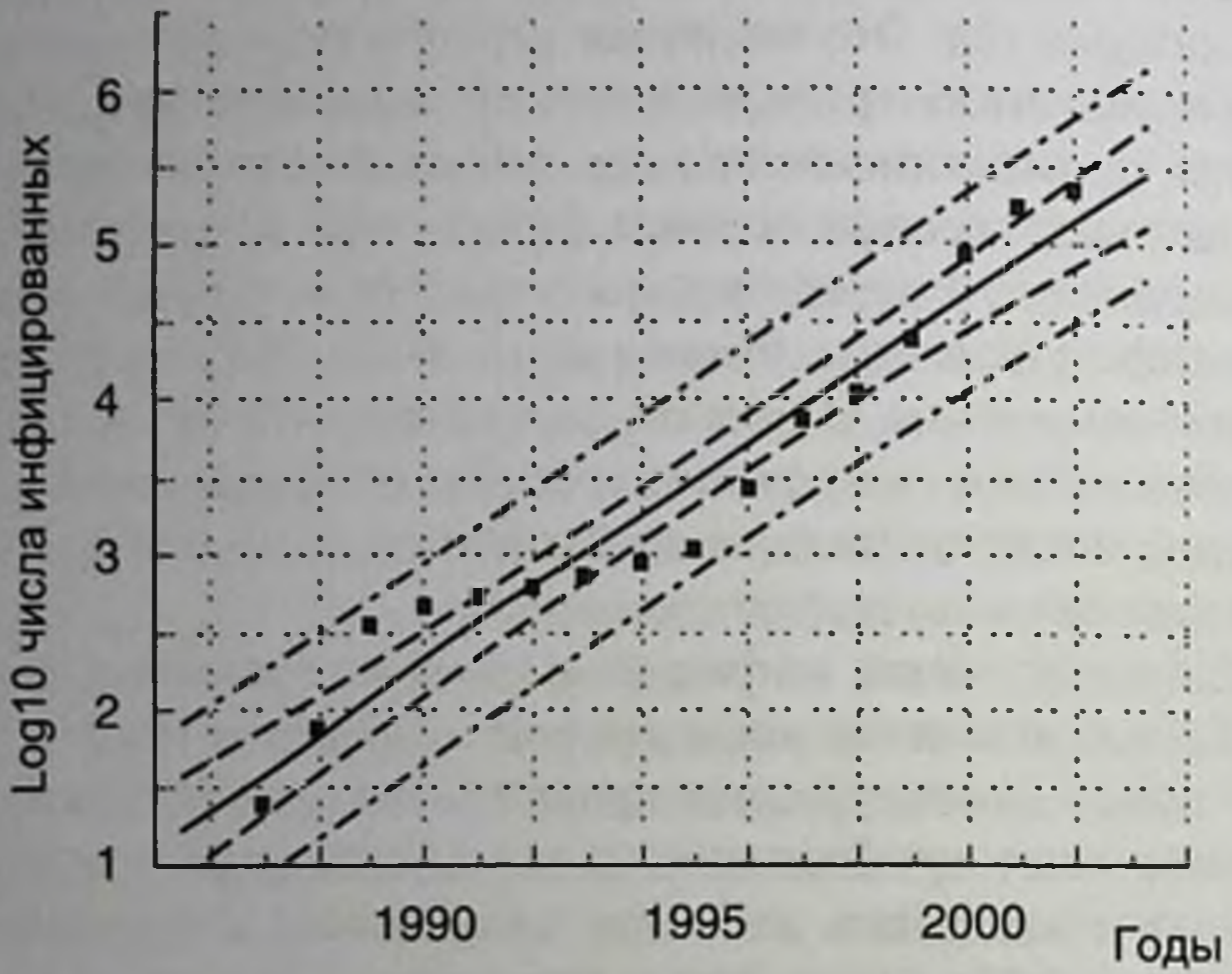
В последние годы XX в. число инфицированных стало быстро расти, удваиваясь каждый год. Это типичная картина количественных изменений на ранних этапах распространения эпидемии при отсутствии эффективных противоэпидемических вмешательств. Позднее, когда в популяции число восприимчивых людей уменьшается вследствие развития иммунитета у переболевших и смерти остальных, темп прироста числа инфицированных уменьшается. Поскольку инфекция ВИЧ не формирует иммунитета, наоборот, через относительно большой срок вызывает развитие иммунодефицита и смерть от осложнений, в наблюдаемый период число инфицированных растет, а смертность больных почти не влияет на число инфицированных.

Быстрый прирост числа инфицированных на линейной шкале выглядит как перелом, и левая часть кривой недоступна глазу в масштабе, который нужен для построения правой части кривой. Такие кривые обычно приобретают приблизительно линейный вид, если на вертикальной шкале откладывать не число заболевших, а логарифм числа заболевших (рис. 2.27). На рисунке видно, что мы получили линию, колеблющуюся около прямой. Если предполагать, что прирост числа инфицированных действительно является экспоненциальным, т.е. на логарифмической шкале описывается прямой, и все отклонения обусловлены привходящими факторами, то можно аппроксимировать фактическую кривую прямой (рис. 2.28, А) (см. разд. 2.8.2 о регрессионном анализе). На рисунке центральная линия — наиболее вероятная прямая роста числа инфицированных, узкие границы — 95% ДИ для нее. Внешние границы показывают, в каком диапазоне можно с 95% вероятностью прогнозировать число инфицированных.

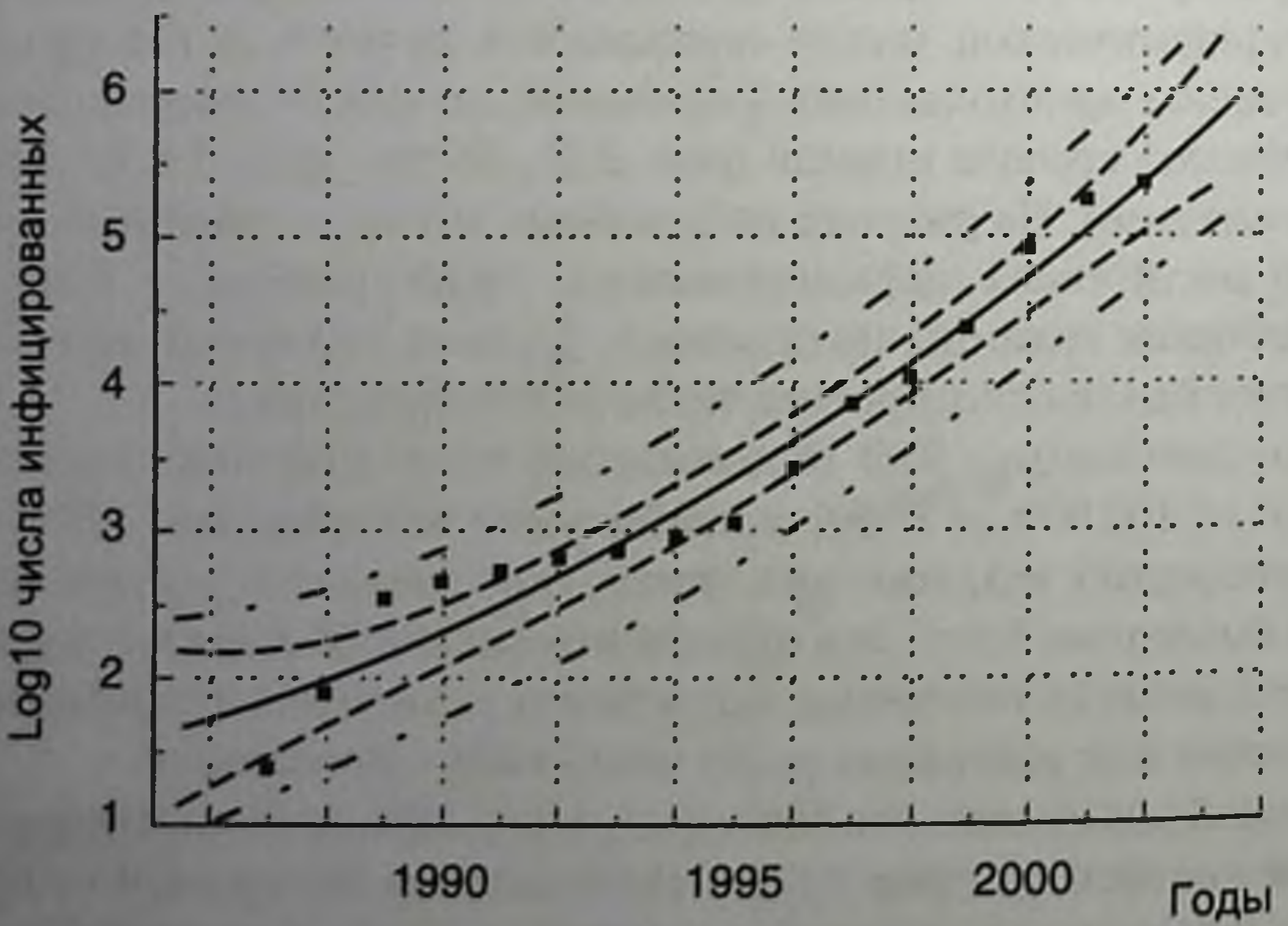
Получается, что с 95% вероятностью число инфицированных будет в 2004 г. от 100 000 до 2 млн, с наибольшей вероятностью 300 000. Прогноз игнорирует искривления линии, описывающей фактические данные за последние 5 лет. Эти отклонения расцениваются как случайные, и предполагается некоторая вероятность изменения тенденции в виде замедления или ускорения роста числа инфицированных.

Попробуем применить более рациональную модель. Предположим, что имеющиеся данные лучше описываются не прямой, а действительно имеет место некоторое ускорение (или замедление) прироста числа инфицированных по сравнению с экспонентой. Результат применения нелинейной модели приведен на рис. 2.28, Б. Этот прогноз дает на 2004 г. примерно 1 млн инфицированных, с вероятностью 95%





а



б

Рис. 2.28. Линейная (а) и квадратическая (б) аппроксимации изменений числа инфицированных ВИЧ



от 160 тыс. до 6,5 млн. Вряд ли нас может удовлетворить такой приблизительный прогноз, но таково свойство всех приближенных моделей.

Чем ограничены возможности экстраполяционного прогноза? Прежде всего тем, что мы не знаем, какие факторы будут определять распространение инфекции завтра, и основываем прогноз только на количественных данных, полученных в предшествующий период. Можно предполагать, что в соответствии с общими для инфекций закономерностями вслед за экспоненциальным ростом наступает замедление прироста числа инфицированных («насыщение») и применительно к болезням, которые несмертельны и вызывают иммунитет, наступает снижение инцидента. СПИД не таков. Замедление его распространения наступает лишь по исчерпанию некоторого «объема» подверженных ему людей. В России в конце XX — начале XXI в. СПИД распространялся преимущественно в среде лиц, употребляющих инъекционные наркотики. Таковых примерно 2% населения. Кажется, что по мере исчерпания этого «резервуара» распространение вируса должно замедлиться. Вероятно, замедление будет лишь ненадолго. По мере увеличения числа инфицированных, вирус выплескивается из «резервуара» групп высокого риска и распространяется через гетеросексуальные контакты. Этого можно было ожидать, исходя из известной динамики развития эпидемии в Африке (рис. 2.29).

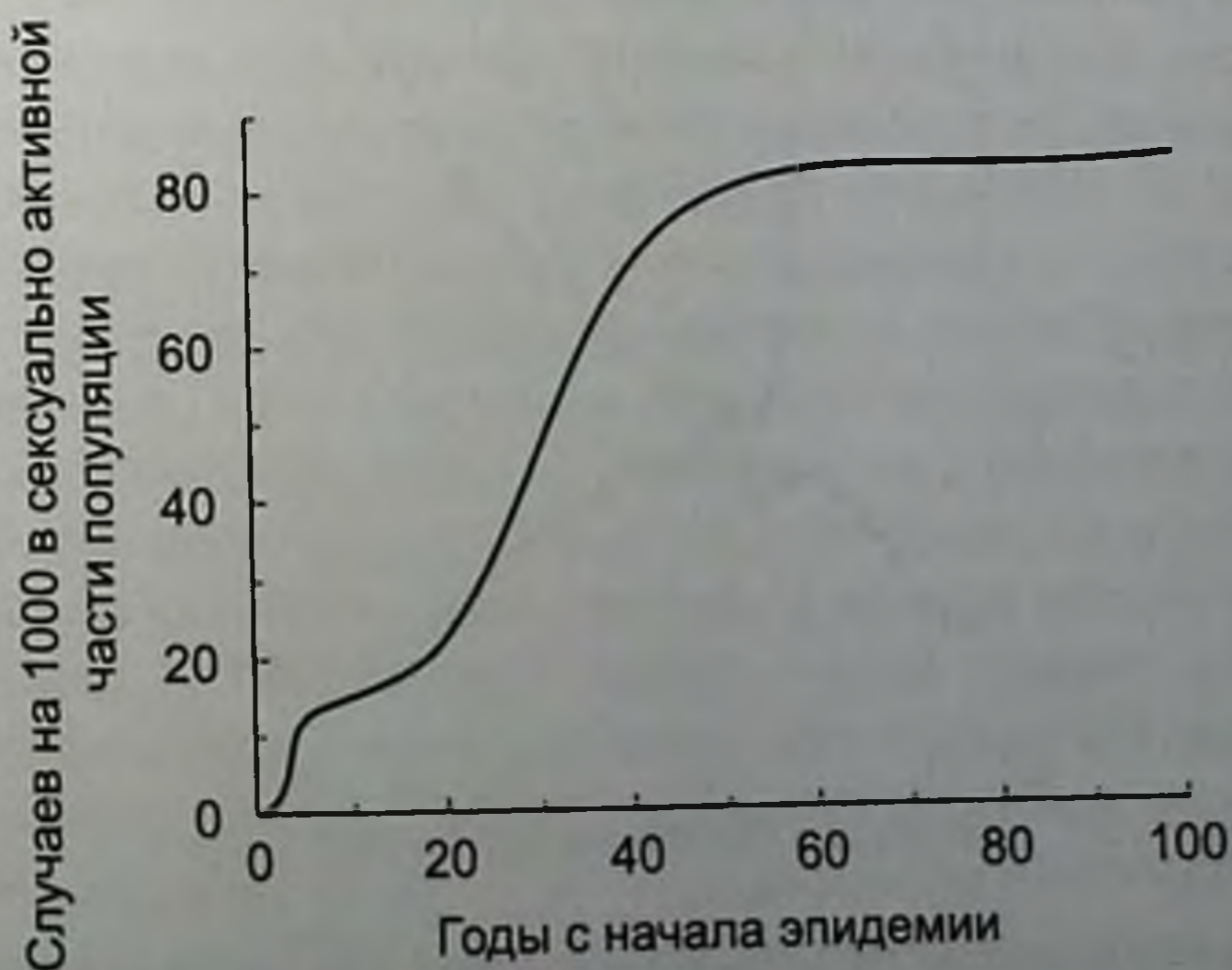


Рис. 2.29. Модель распространения СПИДа по данным о его распространении в субсахарской Африке (по данным Potts M. et al. Lancet. 1991. № 338. P. 608–13)



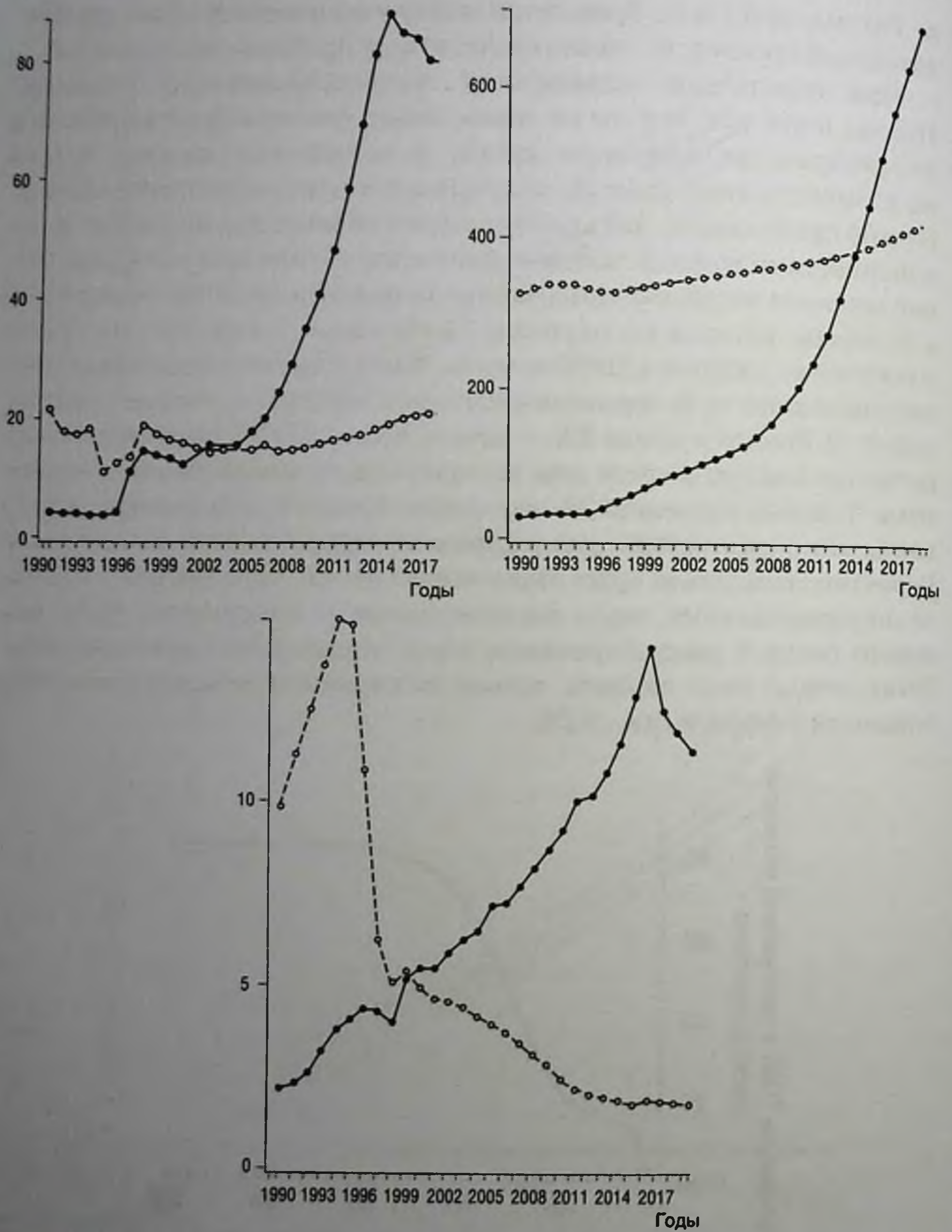


Рис. 2.30. Инциденс, преваленс и смертность от СПИДа (с поправкой на возраст, ордината, слева направо) в США и в Российской Федерации (темные кружки), все числа на 100 тыс. населения. Построено средствами <https://vizhub.healthdata.org/gbd-compare/>



Через 20 лет мы можем проверить, насколько оправдались прогнозы. Для такого анализа удобны инструменты проекта Глобального бремени болезней (Global Burden of Diseases, <https://vizhub.healthdata.org/gbd-compare/>). На рис. 2.30 мы видим, что после некоторого замедления распространения инфекции в 2002–2004 гг. наступил период нового экспоненциального роста. Его нельзя было предвидеть, экстраполируя данные за предшествующие годы. Но его можно было предсказать на основе знаний о механизмах распространения инфекции.

### 2.4.7. Циклические изменения

Самыми известными в эпидемиологии циклическими изменениями являются сезонные изменения заболеваемости. В примитивных обществах сезонная заболеваемость и смертность ярко выражены. По мере развития цивилизации и защищенности человека от суровой среды амплитуда годовых колебаний смертности уменьшается. В известной степени выраженность годового цикла смертности есть мера развитости общества, защищенности его членов от явлений погоды и сезонной неравномерности в доступности пищи.

В Северном полушарии смертность повышается зимой (см. рис. 1.45), но смертность повышается также и в жаркую погоду летом. В этом отражается общая закономерность повышения смертности при экстремальных значениях почти любых факторов окружающей среды. Одним из способов проанализировать влияние циклических изменений среды на заболеваемость — изучить зависимость инциденса от измеряемой переменной окружающей среды. На скаттерграмме рис. 1.45 по абсциссе откладывалось время, а по ординате смертность, и этот график сопоставлялся с графиком, где по ординате отложена температура. Совпадение ритма подтверждает гипотезу о связи смертности с температурой воздуха.

При непосредственном анализе связи смертности с температурой по абсциссе откладывается температура, а по ординате — смертность. Применительно к температуре воздуха такой график покажет, что смертность повышается в холод и в жару. Чем больше отклонение от оптимума температуры, тем выше смертность. Пример такого анализа приведен на рис. 2.31. В больших городах в жаркую погоду образуется множество токсичных газов под влиянием выхлопов автомобилей и выбросов промышленности. Одним из таких газов является озон. При высокой концентрации озона смертность повышается многократно,



прежде всего — смертность людей старше 65 лет. Как и для многих других условий внешней среды, существует некоторый оптимум концентрации озона, за которым в области низких концентраций вновь нарастает смертность.

Сезонные изменения не сводятся к прямому действию климата. Климатические изменения влияют на человека также опосредованно. На рис. 2.32 сопоставляется распространенность клещей в лесах с числом случаев мигрирующей эритемы. Видно, что число случаев хорошо совпадает с распространенностью нимф, а не взрослых клещей.

Вместе с похолоданием в высоких широтах люди начинают проводить больше времени в помещениях, чем создаются благоприятные условия для распространения инфекций воздушно-капельным путем. В большинстве стран, за исключением самых экономически благополучных, в холодное время люди меньше двигаются, употребляют в пищу другой набор продуктов, получают меньшую дозу солнечного света, дышат бо-

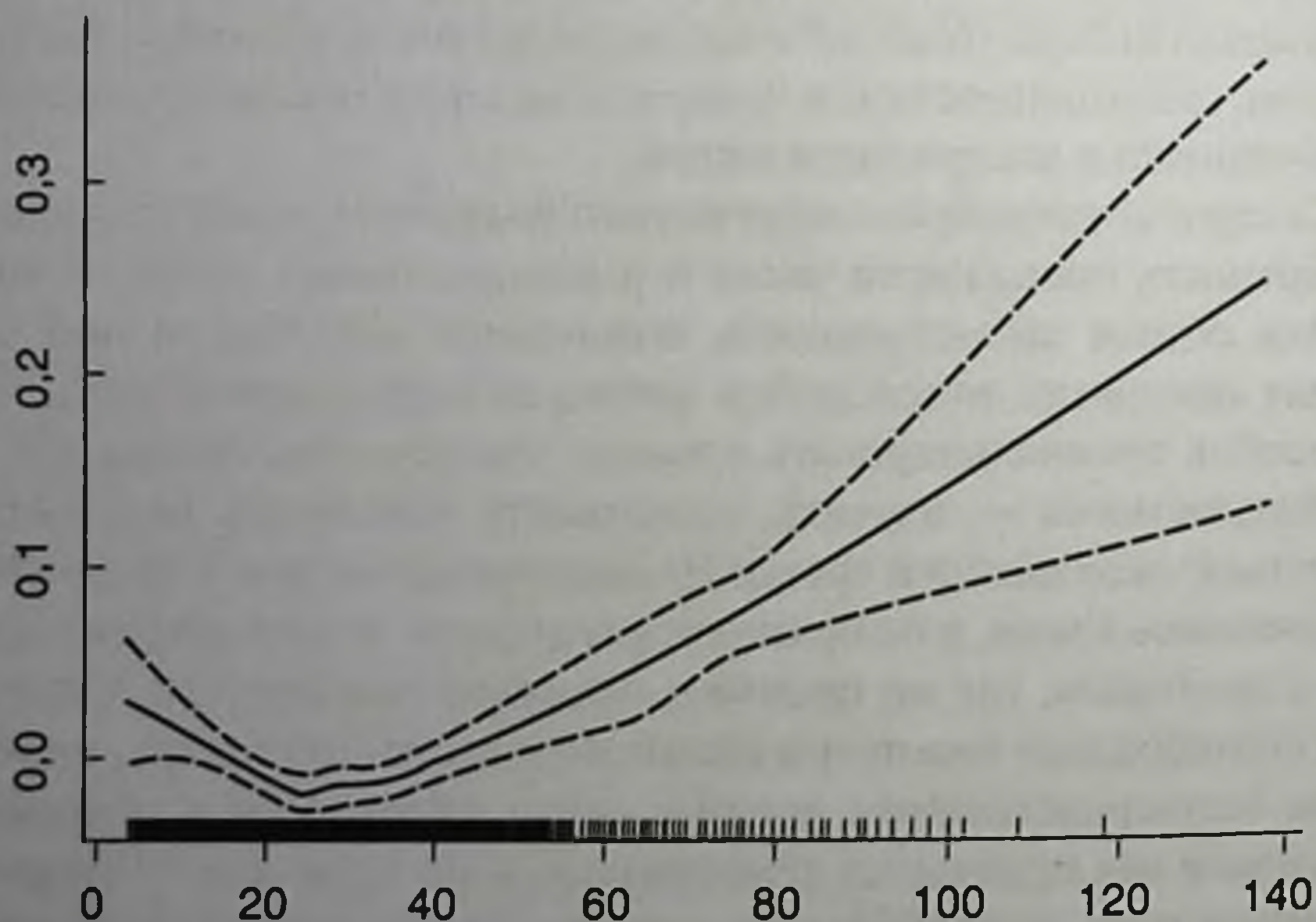


Рис. 2.31. Зависимость общей смертности от концентрации озона в воздухе. Ордината — логарифм изменения числа смертей в текущий день по сравнению с предыдущим, за исключением смертей от несчастных случаев. Абсцисса — концентрация озона, мкг/м<sup>3</sup>. Линия — сглаженная средняя, пунктир — границы  $\pm 2$  м. Засечки на абсциссе — экспериментальные точки, по которым построена нелинейная аппроксимация (Монреаль, 1984–93 гг., по Goldberg M.S. et al. Amer. J. Epidemiol. 2001. № 154. P. 817, с разрешения автора)



лее грязным воздухом в закрытых помещениях. Эти и многие другие факторы приводят к комплексным изменениям в состоянии организма, которые выражаются, в частности, в изменении инцидента инфекционных болезней, травм, обострения или ремиссии хронических болезней.

Несмотря на то что глобальные изменения климата очень инертны и хорошо изучены, лишь во второй половине XX в. были обнаружены многолетние изменения климата, связанные с эль-ниньо, — возникающими в рождественские дни восходящими потоками воды в океане (El Niño, исп. — от слов младенец-Христос). Они приводят рыбу к берегам Перу, оказывают значительное влияние на погодные явления на планете, на флору океана. Последние, в свою очередь, влияют на человека прямо или опосредованно, например, через доступность пищи и размножение насекомых-переносчиков инфекции. Имеются убедительные свидетельства того, что распространение инфекций на Земле коррелирует с активностью Солнца в его многолетнем цикле активности.

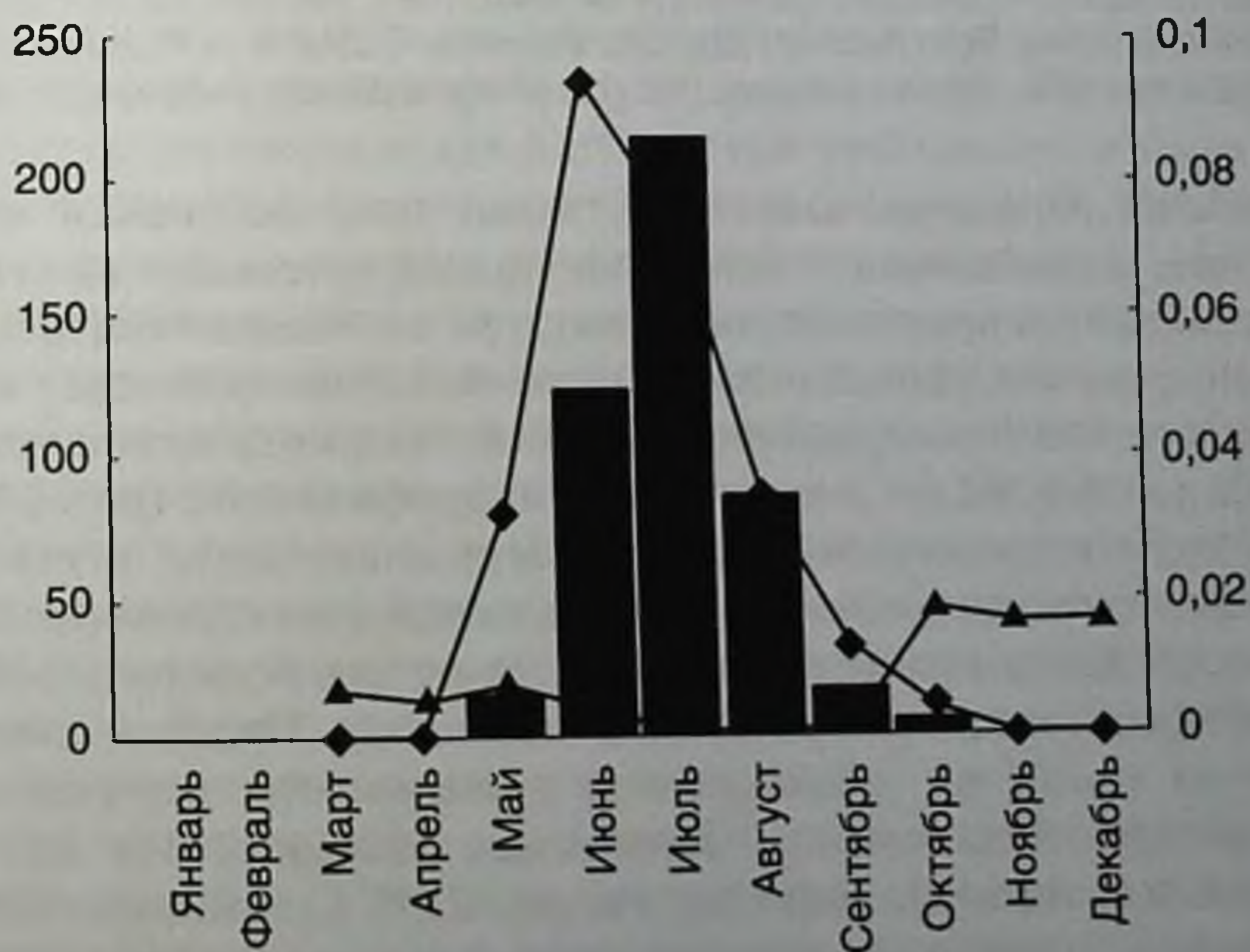


Рис. 2.32. Изменения в течение года частоты встречаемости нимф и взрослых клещей *I. Scapularis* и изменения числа случаев мигрирующей эритемы; по горизонтали — месяцы года. Столбики — число случаев эритемы (шкала слева), ромбы — число нимф и треугольники — взрослых клещей, на квадратный метр, шкала справа (Графство Вестчестер, США, 1991–96 гг., по Falco R.C. et al. Amer. J. Epidemiol. 1999. № 149. P. 771–6, с разрешения издателя)



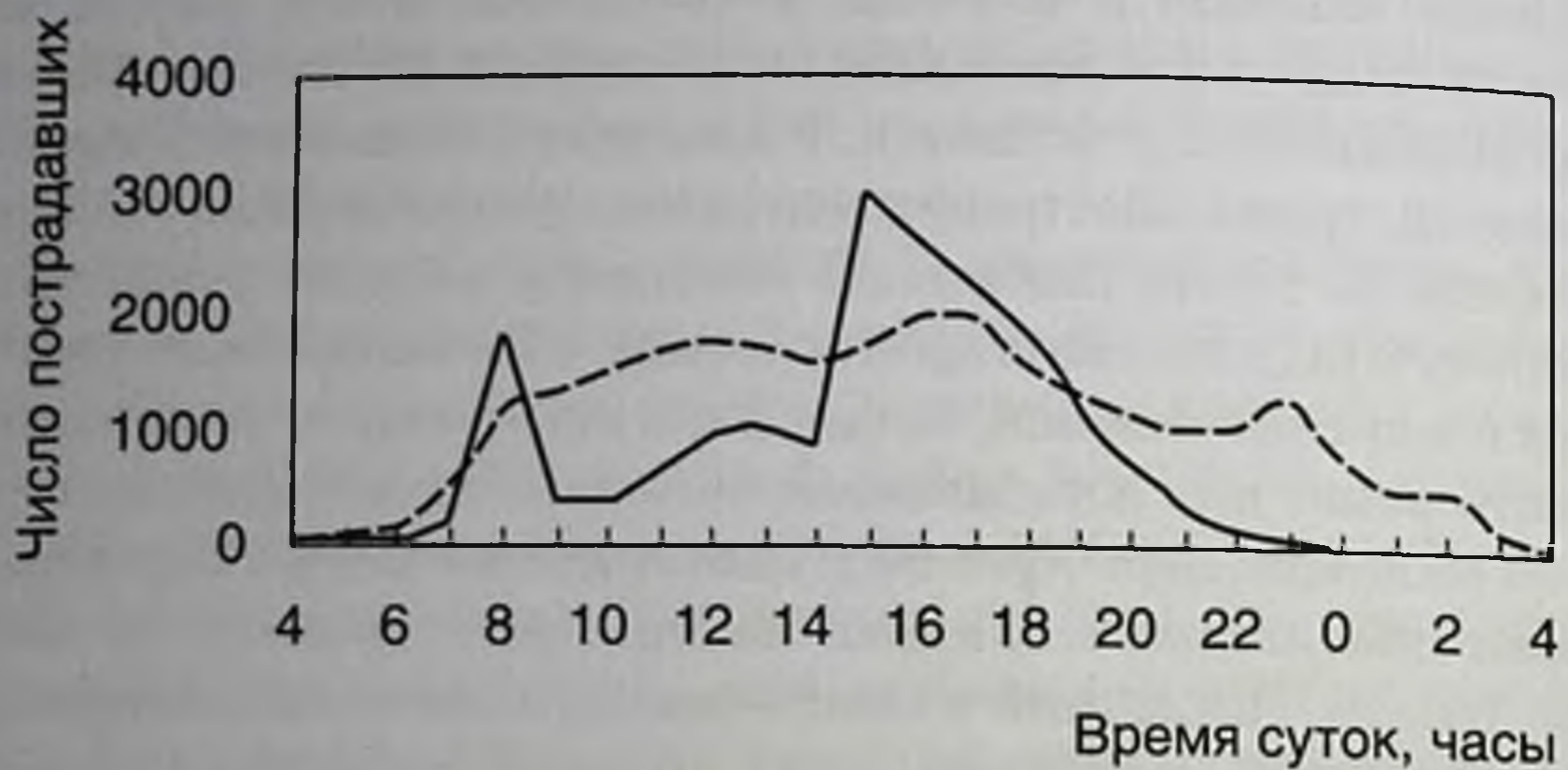


Рис. 2.33. Изменения числа пострадавших пешеходов на дорогах Великобритании в течение суток. Обратите внимание на два пика травматизма у детей (линия) во время, когда они идут в школу и из школы. У взрослых (пунктир) подъем травматизма при движении на работу в 7–8 ч, в 16–18 (движение с работы) и в 22–23 (время закрытия пабов). По данным Department of the Environment, Transport and the Regions: Transport Statistics: Road Accidents Statistics: Pedestrian Casualties in Road Accidents: Great Britain 1998. <http://www.transtat.detr.gov.uk/facts/accident/pedest/pedest98.htm>

Наиболее изученным является суточный цикл активности человека (точнее, *циркадианный* — близкий по продолжительности к суткам). В многообразных проявлениях суточного ритма переплетаются биологические ритмы и социальные факторы. Эпидемиологическое исследование суточного распределения событий, например возникновения травм на дорогах, может открыть пути к их профилактике (рис. 2.33).

Для оценки амплитуды и длительности циклических изменений, а также их статистической значимости используются специальные статистические методы: автокорреляционный анализ, спектральный анализ, преобразование Фурье, косайнор-анализ и др. Простым способом выявления колебаний заболеваемости и смертности, периодических и непериодических, является *фильтрация*. Элементарным приемом фильтрации является сглаживание (см. рис. 2.13). Сглаживанием называют вычисление средней величины для расположенных близко во времени величин. Предполагается, что различающиеся между собой близкие по времени величины различаются случайно или под влиянием кратковременно действующих факторов. Обычно применяют сглаживание методом скользящей средней. Для этого вычисляют среднюю 1-, 2- и 3-й величин в ряду  $(X_1, X_2, X_3)$ , получая первую точку, замещающую  $X_2$



в новом сглаженном ряду. Затем вычисляют среднюю для величин  $X_2$ ,  $X_3$ ,  $X_4$ , получая точку на место  $X_3$ . Новая кривая лишается кратковременных изменений от точки к точке, и на ней лучше видна долговременная тенденция. Для того чтобы сгладить еще более длительные колебания, усредняют большее число последовательных величин в ряду (4, 5... 10). Когда число точек становится сопоставимым с длительностью цикла изменений изучаемой переменной, цикл сглаживается (исчезает). Например, при изучении суточного цикла усреднение результатов измерений, выполненных в течение 1 ч позволяет лучше увидеть суточный цикл. Усреднение результатов, полученных в течение 4 ч, приведет уже к устранению важных деталей самого суточного цикла.

Конечно, методом скользящей средней сглаживаются не только периодические, но и непериодические колебания. На рис. 2.13 сглаживанием за 25 лет устранены колебания смертности от эпидемий, войн и неурожаев, и в результате становится видимой очищенная от этих кратковременных изменений тенденция.

Вторым простым приемом выявления периодических изменений является прием *синхронизации* множества измерений. Число событий в течение одного цикла может быть достаточно большим, и тогда простая скаттерграмма выявляет наличие цикла (см. рис. 1.45). Однако событий может быть меньше, а влияние случайных факторов — больше. Если исследователь знает длительность цикла или имеет гипотезу о его длительности, то он может синхронизировать все изученные им циклы, например годы цикла солнечной активности или месяцы года, часы суток и получить суммарную кривую (см. рис. 2.31–2.33). Число событий в каждом интервале времени многократно увеличивается. Случайные изменения, связанные с более короткими, чем изучаемый цикл, и с более длительными, чем этот цикл, событиями при вычислении средней устраняются. В результате удастся выявить циклические изменения, недоступные непосредственному наблюдению.

Третий простой прием — *автокорреляция*. Его применяют для анализа ряда величин, в котором предполагают наличие скрытой циклическости. Прием состоит в вычислении коэффициента корреляции (см. разд. 2.8.2) с разной величиной шага. Сначала вычисляют корреляцию между собой соседних величин (1-ю с 2-й, 2-ю с 3-й, 3-ю с 4-й...). Затем каждой 3-й, т.е. через одну (1-ю с 3-й, 2-ю с 4-й, 3-ю с 5-й...). Затем каждой 4-й, т.е. через две (1-ю с 4-й, 2-ю с 5-й...) и т.д. Обычно очень близкие по времени измерения дают близкие результаты



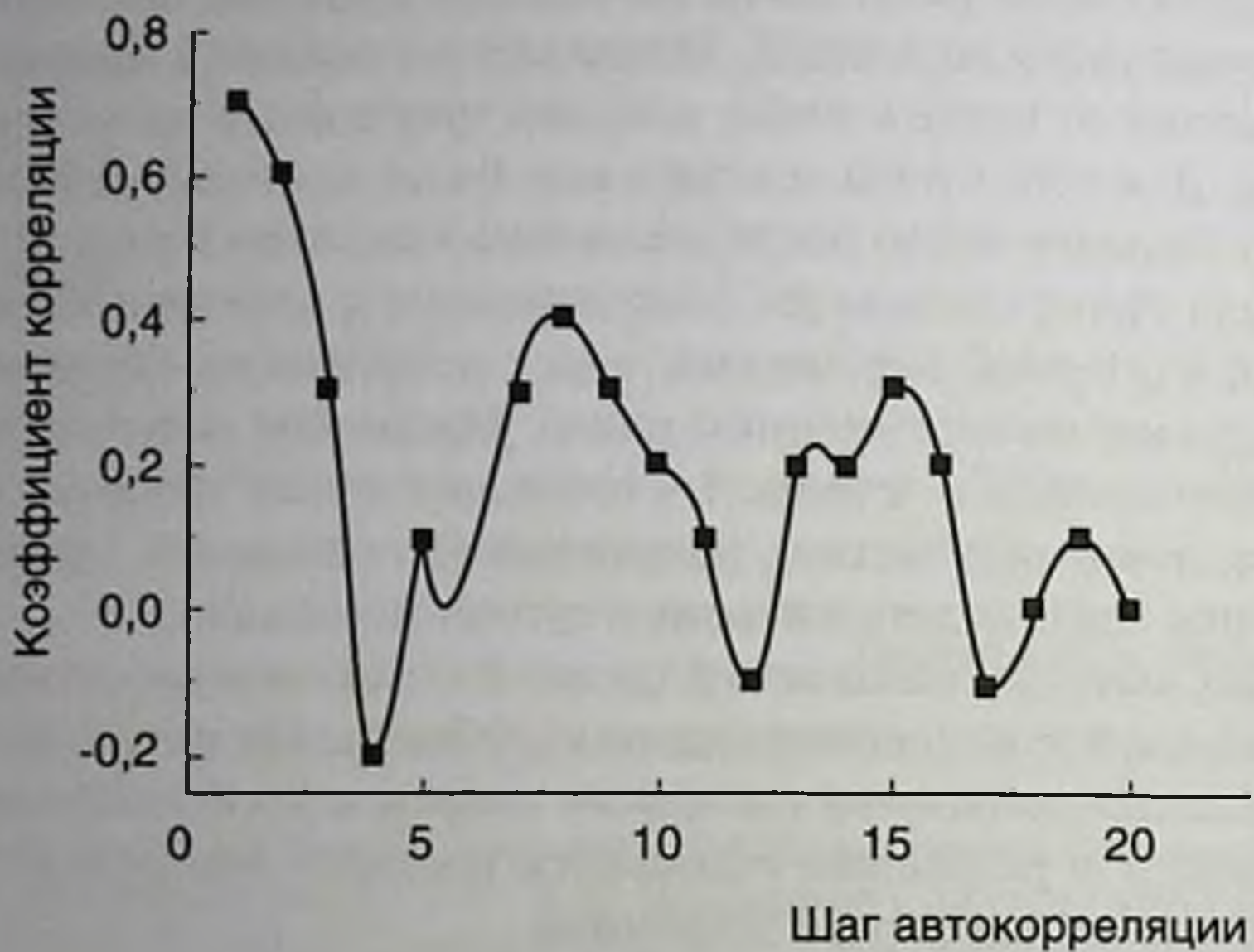


Рис. 2.34. Автокорреляционная функция для длительности интервалов между сердечными сокращениями. Видна выраженная дыхательная аритмия с периодом примерно 8 сердечных сокращений

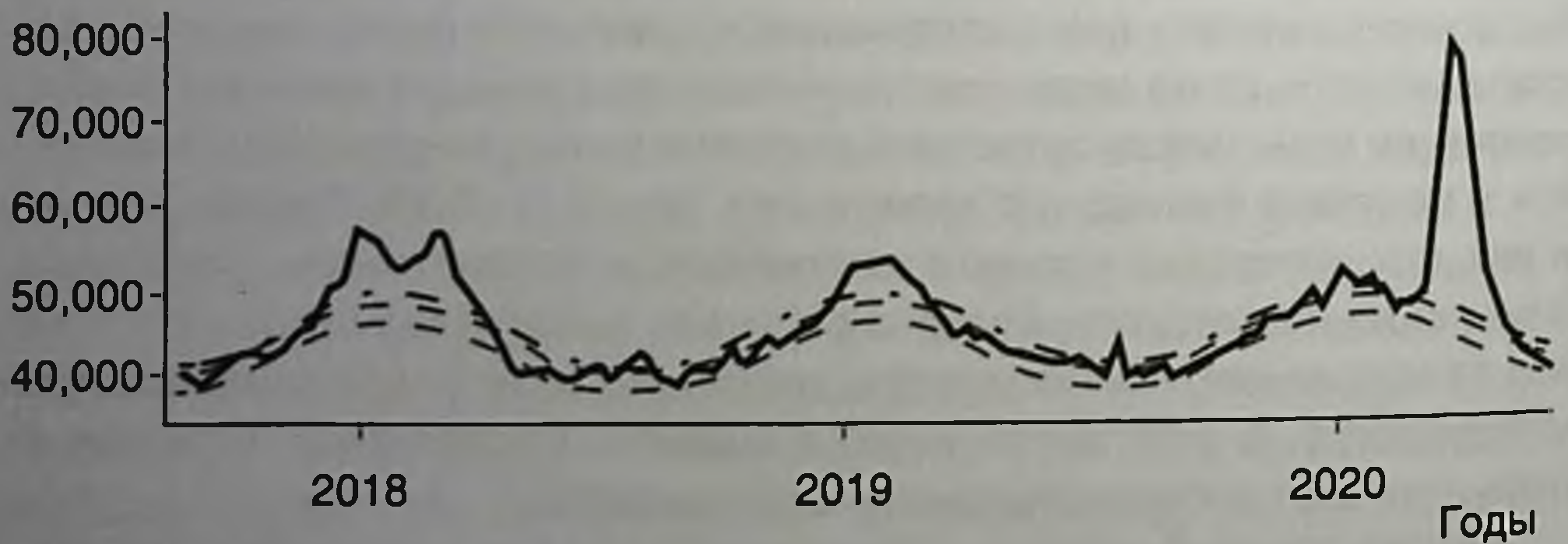


Рис. 2.35. Недельное количество смертей в странах европейского региона, участвующих в проекте <https://www.euromoto.eu/graphs-and-maps#>. Абсцисса — годы, ордината — количество смертей у лиц в возрасте 65+ лет. Пунктирные линии ограничивают зону в  $\pm 1$  СКО. Штрих-пунктир — 2 СКО над среднегодовой смертностью. При развитии пандемии новой коронавирусной инфекции COVID-19 общая смертность оказалась наилучшей мерой заболеваемости и смертности, так как истинная заболеваемость и специфическая смертность даже через полгода после начала эпидемии остаются приблизительно оцениваемыми



(положительно коррелируют между собой). По мере удаления измерений во времени корреляция между результатами измерений уменьшается. Например, если измерять массу тела ежедневно по утрам, то максимальной будет корреляция результатов измерений в последовательные дни (с шагом 1-м), но, вероятно, обнаружится пик корреляции с шагом 7-м, так как существует традиция обильной еды по выходным. Наиболее наглядно применение автокорреляции в анализе сердечного ритма (рис. 2.34). Интервалы между сердечными сокращениями становятся короче во время вдоха, и поэтому с шагом примерно от 4 до 8 выявляется выраженная положительная корреляция длительности кардиоинтервалов.

В эпидемиологии многих инфекций имеет место выраженная сезонность. Применительно к таким инфекциям для диагностики вспышки, выходящей за пределы обычного сезонного повышения заболеваемости, рассчитывают многолетнюю среднюю годовую кривую заболеваемости и ее предсказательный интервал (два стандартных отклонения вверх). При выходе кривой заболеваемости за пределы интервала диагностируют вспышку, т.е. более чем сезонное повышение заболеваемости (рис. 2.35).

## Контрольные вопросы и задачи к разделам 2.3 и 2.4

1. При расследовании вспышки пищевого отравления в доме отдыха было установлено:

	Блюдо	Доля заболевших, %	
		Ели данную пищу	Не ели данную пищу
а	Оладьи со сметаной	32	36
б	Салат «Оливье»	64	28
в	Гороховый суп	26	22
г	Печень жареная	42	36
д	Капустный салат	28	26

С каким блюдом вероятно связано отравление?



2. В населенном пункте с населением 10 000 в 2000 г. было на учете 200 больных сахарным диабетом. В течение 2000 г. диагностировано 50 новых случаев и 40 больных диабетом умерли. Инцидентс диабета в городе:
  - а) 20 на 1000;
  - б) 200 на 1000;
  - в) 4 на 1000;
  - г) 5 на 1000.
3. При тех же условиях, что в задаче 2, преваленс диабета на 1 января 2001 г.:
  - а) 25 на 1000;
  - б) 15 на 1000;
  - в) 21 на 1000;
  - г) 29 на 1000.
4. При тех же условиях, что в задаче 2, преваленс диабета на 1 января 2000 г.:
  - а) 15 на 1000;
  - б) 21 на 1000;
  - в) 20 на 1000;
  - г) 11 на 1000.

## **2.5. РАЗЛИЧИЯ ЗАБОЛЕВАЕМОСТИ В ПРОСТРАНСТВЕ**

Различия заболеваемости в отдельных местностях известны с древних времен, оставивших нам в наследство термин «эндемия» (заболеваемость, характерная для данной местности). В течение всей истории эпидемиологии устойчивая заболеваемость на отдельных территориях, отличная от других территорий или вновь выявленная, привлекала внимание исследователей. Благодаря этим работам было выявлено множество до того неизвестных болезней, вредных условий естественной среды, фактов загрязнения среды отходами, вызывающих массовые заболевания. Некоторые из этих болезней получили название по имени места, где они были впервые эпидемиологически выявлены и описаны (см. врезку на с. 226). Так, в процессе освоения географического распространения патологии эпидемиология обогащала экологию и медицинскую географию.



### Болезнь Минамата

Болезнь получила свое название по названию залива в Японии, где жители побережья в 1956 г. стали страдать непонятной болезнью, проявляющейся неврологической симптоматикой.

Пострадали более 2 млн человек, умерли 900. В течение 3 мес после первого официального сообщения врача о больных с подозрительной и непонятной симптоматикой была создана специальная исследовательская группа в университете и в течение 3 мес она установила, что заболевание связано с отравлением вод залива, откуда получали пищу жители региона, солями тяжелых металлов. В 1959 г. было твердо установлено, что болезнь является формой хронического отравления органическими соединениями ртути, поступающими с морепродуктами. Была показана связь с ртутным отравлением случаев врожденных церебральных параличей — врожденная форма болезни Минамата. Chisso Corp сливала ртутьсодержащие отходы в залив с 1932 г. по 1968 г., когда правительство Японии признало связь заболеваний с загрязнением.

Характерные черты случаев болезни

1. Все больные жили в одном районе у залива Минамата.
2. Большинство случаев наблюдалось в семьях рыбаков.
3. Поражались мужчины и женщины любого возраста, кроме младенцев.
4. Среди взрослых чаще поражались мужчины.
5. Не было регулярности в развитии заболевания в семьях: в одном случае члены семьи заболевали с интервалом в дни, в другом — в годы.
6. Летальность была высокой.
7. Сельскохозяйственные продукты и пищевая вода не были под подозрением с самого начала.
8. Все заболевшие ели рыбу и морепродукты из залива.
9. Кошки со сходными проявлениями болезни были найдены в пораженном районе.

Главными проявлениями болезни у взрослых были атаксия, концентрическое сужение полей зрения, тугоухость, экстрапирамидные симптомы (мышечная ригидность, произвольные движения), небольшой интеллектуальный дефект и эмоциональная неустойчивость. При вскрытии умерших обнаружили токсическую энцефалопатию без признаков воспаления. Эта же патология была найдена у кошек, пойманных в районе Минамата, и у кошек, которым в эксперименте скармливали рыбу.

Особенностью изучения эпидемии было то, что к этому времени не было известно отличие поражения органическими соединениями ртути от поражений парами металлической ртути. Подозрение падало на ряд тяжелых металлов, повышенная концентрация которых в заливе была найдена. В 1958 г. выяснили, что патологические изменения в мозгу сходны с впервые описанными в 1954 г. признаками отравления метилртутью.



Исследование содержания ртути в мозгу погибших подтвердило ее наличие в небольших концентрациях в случаях хронической болезни и в высоких — при остром заболевании. В начале 1960-х годов были уточнены представления о том, какое именно соединение ртути было причиной болезни Минамата. Тогда же была начата широкомасштабная программа восстановления региона — значительная часть залива была отрезана от моря и осушена, вылов рыбы ограничен, население получило экономическую и медицинскую помощь, компенсации. Эта программа продолжается и в XXI в.

### 2.5.1. Географический анализ

Применительно к инфекционным болезням принято называть территорию, на которой регистрируется данная болезнь, ее *нозоареалом* (в экологии ареалом называют зону обитания вида). Обычно нозоареал ограничен условиями, в которых может существовать возбудитель, его переносчик или промежуточный хозяин. Для неинфекционных болезней нозоареал может быть связан с особенностями почвы и воды, например, приводящими к недостатку йода в пище и появлению эндемического зоба. На ранних этапах развития современной эпидемиологии различия территорий и мест проживания по заболеваемости и смертности были основным способом исследований в *санитарии*. На основании таких исследований были предложены меры улучшения среды обитания и достигнуты большие успехи в сокращении смертности населения в городах Европы.

Во второй половине XX в. сравнение стран и территорий между собой по заболеваемости и смертности хроническими болезнями широко применяли в предположении, что, установив различия в заболеваемости и в предполагаемом факторе риска, можно будет воздействовать на эти факторы риска. На рис. 2.36 приведены результаты известного исследования, продемонстрировавшего, что в регионах с высокой концентрацией холестерина в плазме крови жителей смертность от ишемической болезни сердца повышена. Эта закономерность к этому времени была уже установлена на индивидуальном уровне (см. рис. 1.5). Холестериновая гипотеза получила множество подтверждений разного рода.

До 1980-х годов язвенная болезнь желудка и двенадцатиперстной кишки представлялась загадочной болезнью. Ее лечили терапевты,



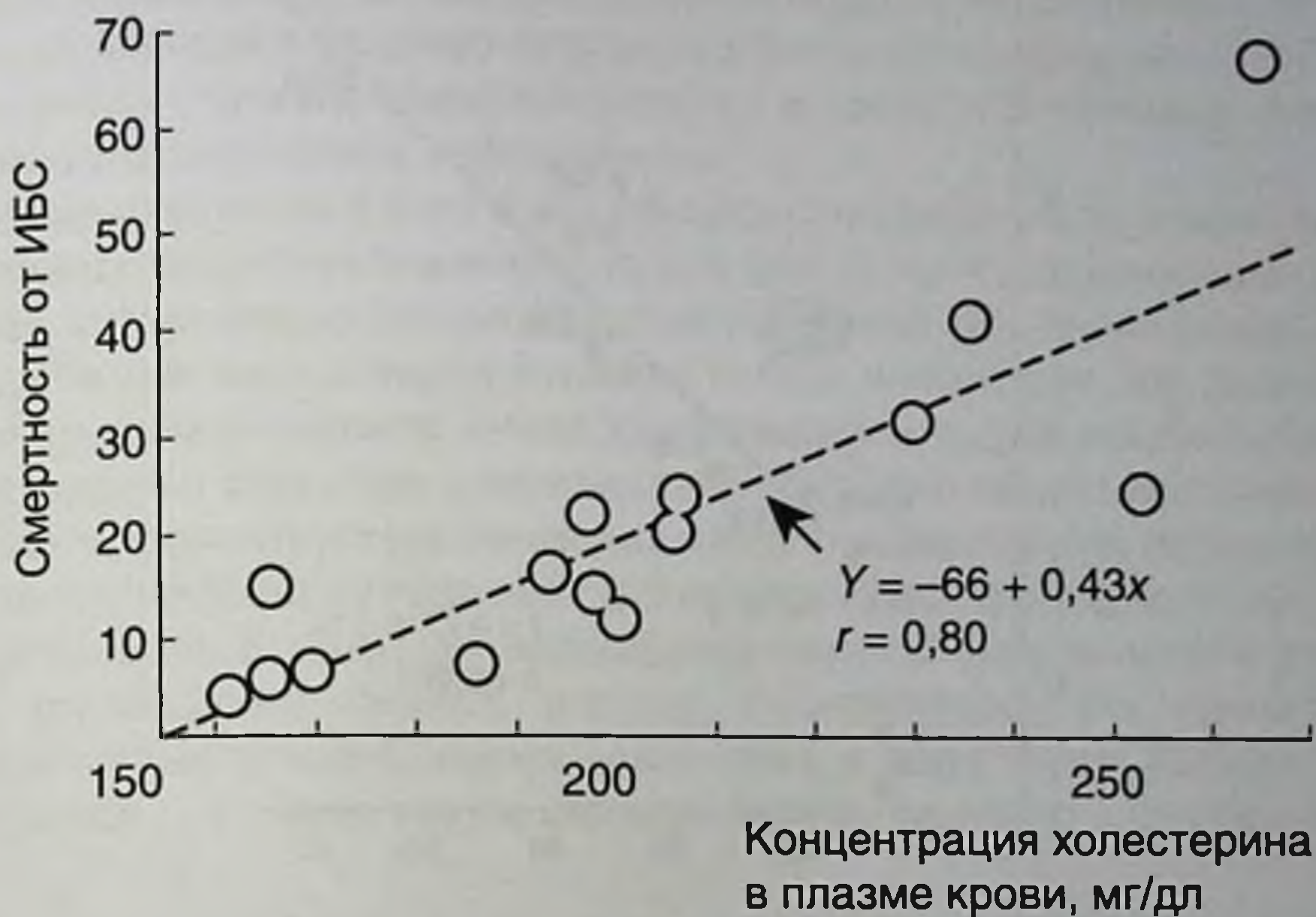


Рис. 2.36. Корреляция смертности от ишемической болезни сердца (на 100 000 населения, ордината) с медианой концентрации холестерина в плазме крови (мг/дл, абсцисса). Кружками обозначены когорты, изученные в 1970-е годы в 7 странах мира. Пунктиром обозначена линия регрессии, описывающая зависимость смертности от концентрации холестерина для этой выборки, на рисунке приведены также коэффициент корреляции ( $r$ ) и уравнение линейной регрессии (по Keys A. Seven Countries: A multivariate analysis of death and coronary heart disease. Cambridge: Harvard UP, 1980)

преимущественно лекарствами, подавляющими секрецию желудочного сока. Ее лечили и хирурги, ушивая язвы, резецируя их или часть желудка. Таким образом с помощью ножа устраняли пораженную язвенным процессом часть органа или часть органа, ответственную за секрецию. Были разработаны «физиологические» способы хирургического лечения в виде ваготомии и пилоропластики, направленные на уменьшение желудочной секреции. Происхождение болезни связывали с множеством факторов риска, из которых на первом месте были социальные и психологические. Язвенная болезнь была первым примером в числе так называемых психосоматических болезней. Как важный фактор риска язвенной болезни изучали употребление в пищу большого количества соли (рис. 2.37). В настоящее время эта корреляция представляет



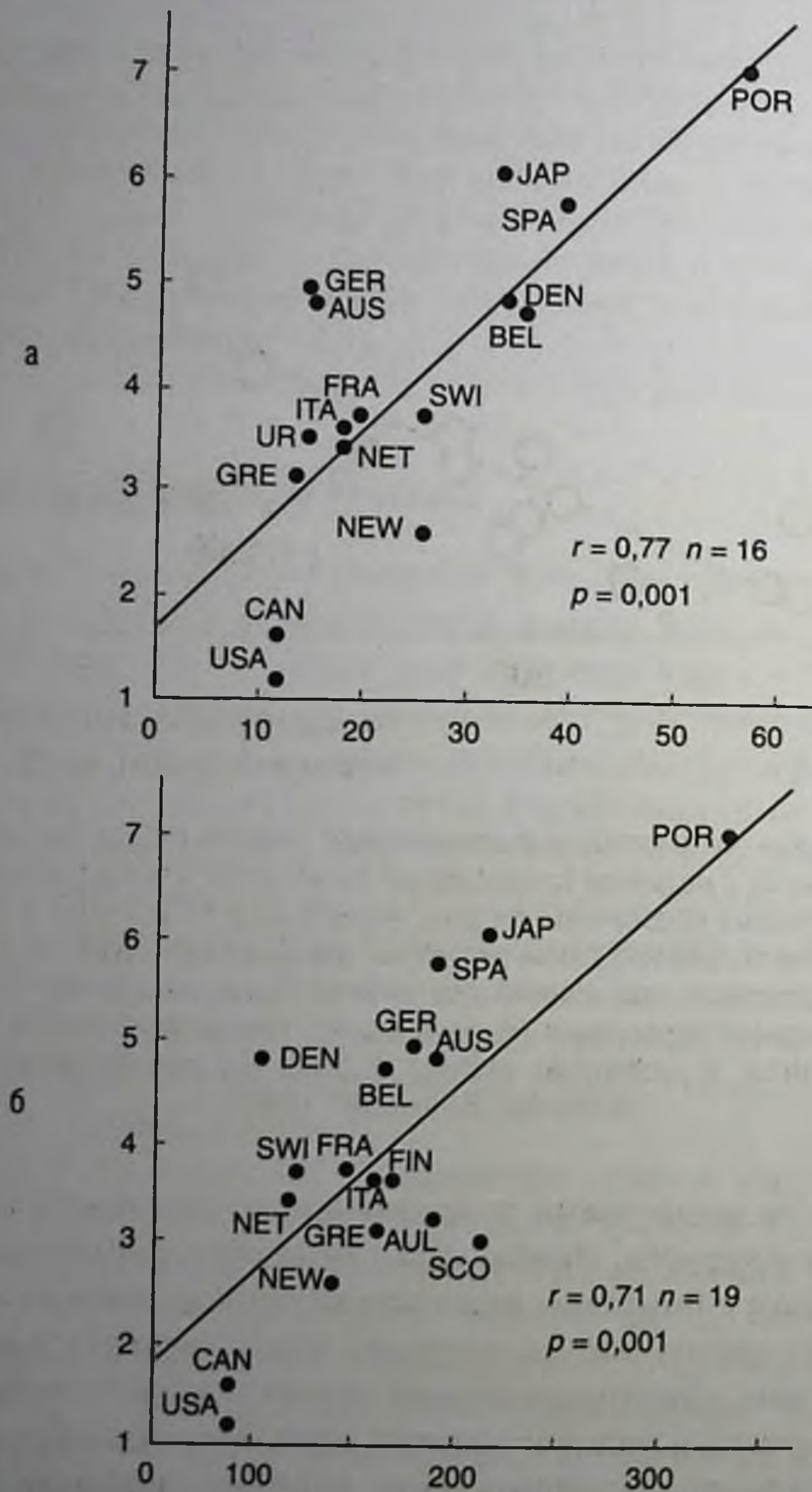


Рис. 2.37. Корреляция смертности от язвенной болезни желудка (ордината, на 100 000 населения) с потреблением поваренной соли (абсцисса, г/сут) (а) и смертностью от сердечно-сосудистых болезней (абсцисса, на 100 000) (б);  $r$  — коэффициент корреляции,  $p$  — его статистическая значимость,  $n$  — число стран. Потребление соли — общее в стране, а не только потребление с пищей (по Sonnenberg A. Digest. Dis. Science. 1987. № 32 (1). P. 1–7)



собой только исторический интерес, а вот корреляция смертности от язвенной болезни и сердечно-сосудистых болезней, видимо, не случайна. Она может отражать общий компонент в патогенезе болезней, который можно условно назвать *инфекционным*.

Применительно к раку были обнаружены удивительно тесные корреляции с потреблением мяса и жиров (рис. 2.38). На приведенном рисунке справа вверху находятся развитые страны. Не удивительно, что потребление мяса и жиров наиболее велико именно там, где наибольшая продолжительность жизни и, соответственно, рак занимает большее место в структуре смертности. Исследователи делали поправку на возрастную структуру населения в виде стандартизации по возрасту, но такая поправка не устраняет других возможных факторов. За исключением Японии, на фланге низкой смертности от рака молочной железы расположены бедные страны, отличающиеся по множеству признаков от употребляющих много мяса и жира стран. Каковы эти признаки — не известно до сих пор. Вторая проблема экологических

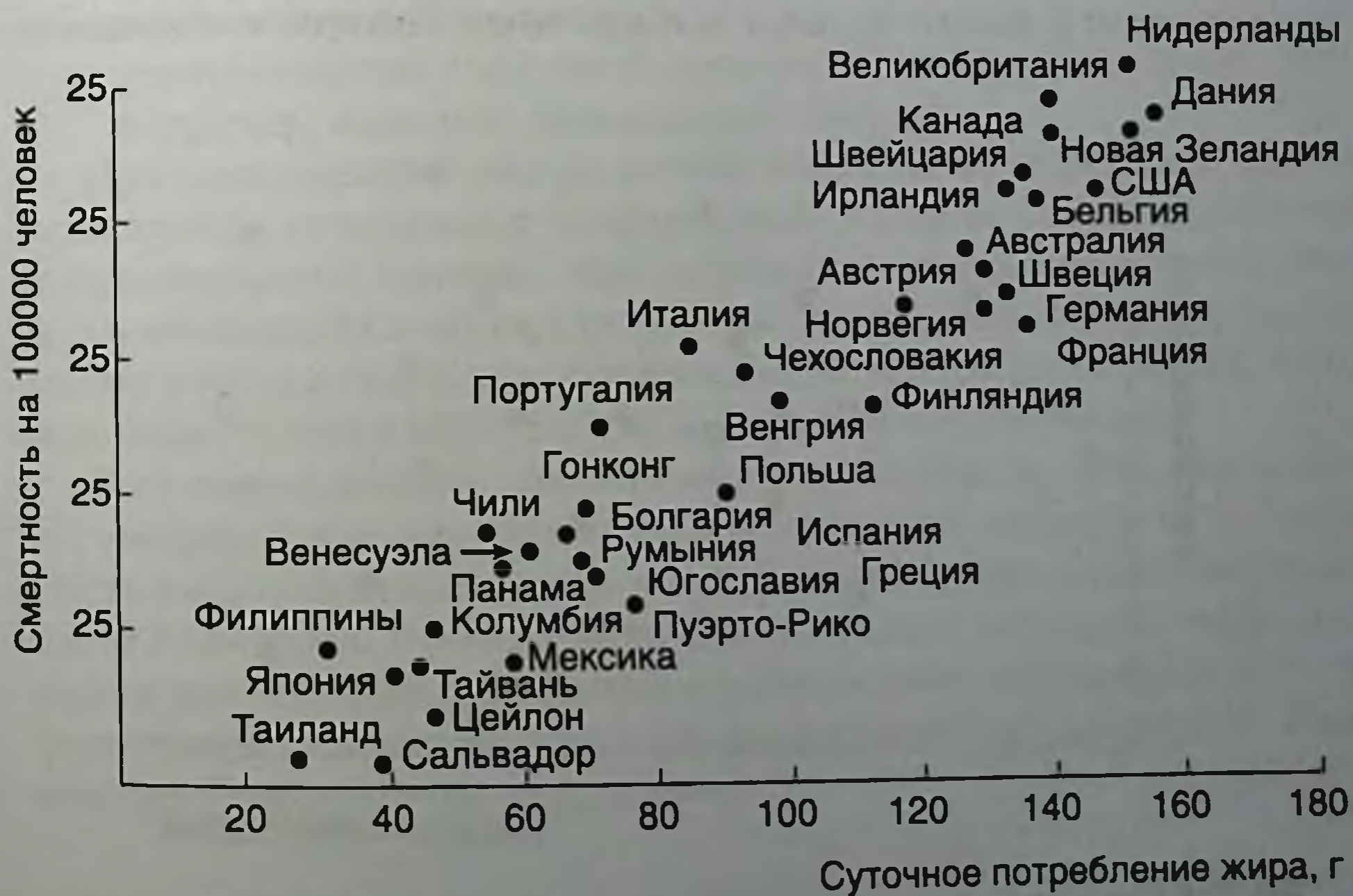


Рис. 2.38. Корреляция между общим потреблением жира и смертностью женщин от рака молочной железы в различных странах (по Carroll K.K. Cancer. Res. 1975. № 35 (11 part 2). P. 1374–83 с изм.)



исследований — неопределенность выборки. Почему в одном из анализов использованы данные о 16 странах, в другом — о 19, в третьем — о 39? Подобного рода неопределенная выборка сама по себе может «создавать» любые связи (рис. 2.39).

Для наиболее дальновидных эпидемиологов (Р. Долл) было очевидно, что даже тесная корреляция не может быть надежным доказательством причинно-следственной связи, но она заставляла подозревать, что существуют какие-то близкие диете причины, на которые можно воздействовать. Менее осторожные специалисты сделали из этих экологических исследований далеко идущие выводы, и в течение 20 лет идея ограничения количества мяса и жира в диете с целью профилактики рака была популярной.

Исследования конца XX в. различной структуры: когортные и типа сравнения с контролем, показали, что на индивидуальном уровне связь рака с потреблением жира практически отсутствует. Более того, у женщин, в детстве и юности регулярно употребляющих молоко, риск рака молочной железы в молодости даже снижен. Такого рода результаты привели к далеко идущим последствиям. Интерес к «экологиче-

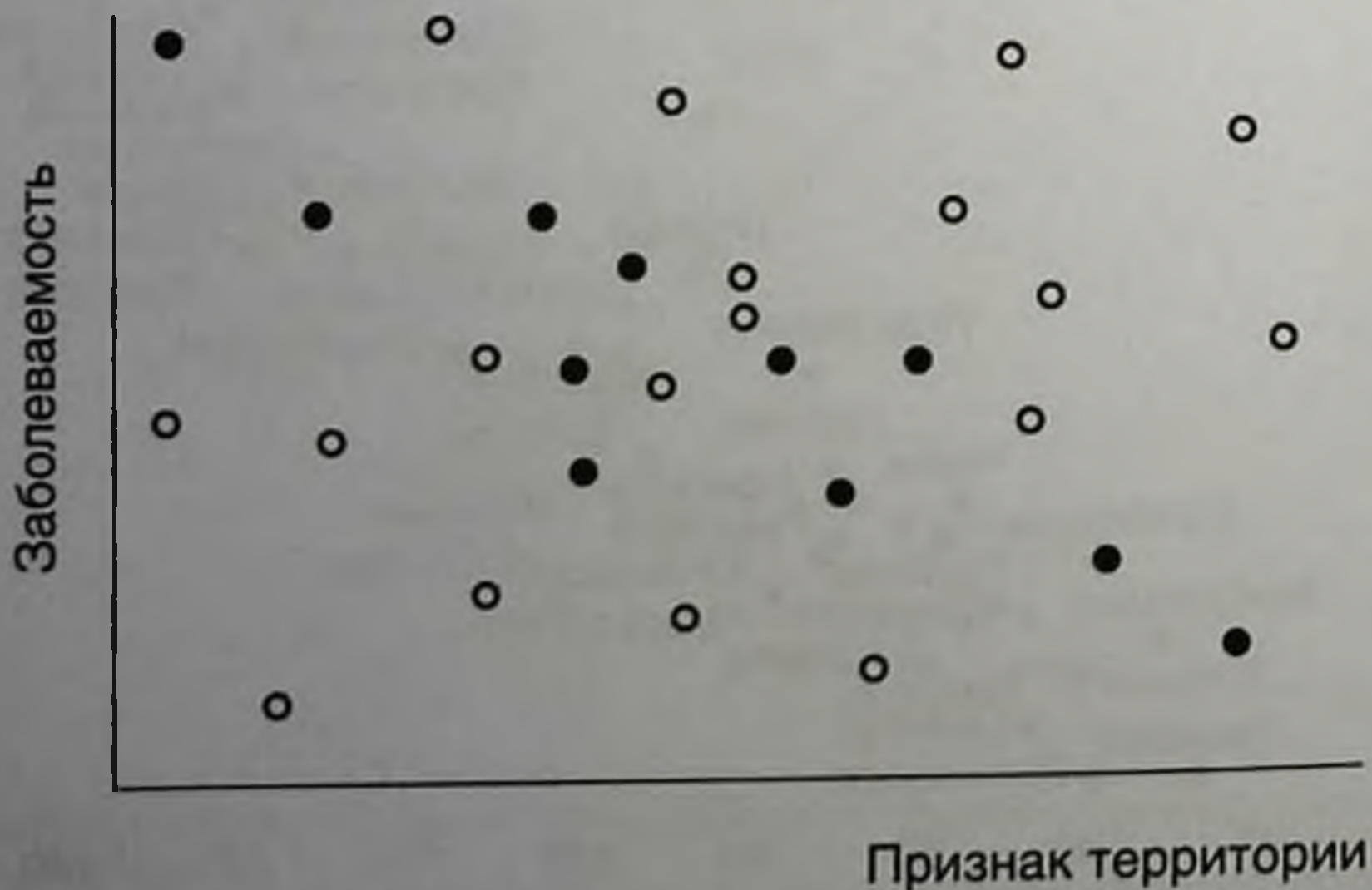


Рис. 2.39. Возможный механизм возникновения связи при экологических исследованиях. Если из большого числа территорий в анализ включены несколько, на рисунке обозначенных черными кружками, то при неслучайном их выборе возможно возникновение корреляции между заболеваемостью и изученным признаком



скому» анализу резко снизился. Был сформулирован вывод о том, что данные экологического анализа могут использоваться только для межпопуляционных сравнений. Их нельзя переносить на различия между индивидуумами, пока такие же различия индивидуумов не будут подтверждены в когортных исследованиях межиндивидуальных различий. Только после такого подтверждения можно формулировать рекомендации по вмешательству на уровне отдельных субъектов, касается ли это лечения или рекомендаций по питанию.

Экологические исследования являются приемлемым методом, если:

- ▶ региональные особенности проще определить из-за сложности или дороговизны определения этих особенностей у отдельных лиц;
- ▶ такой подход позволяет получить большее разнообразие изучаемой особенности, как, например, при международных сопоставлениях пищевых привычек (по сравнению с пищевыми привычками граждан одной страны);
- ▶ точность общего измерения выглядит более высокой, чем индивидуального (для субъекта), как это имеет место, например, с определением потребления алкоголя;
- ▶ популяционные измерения представляют основной интерес, как, например, инцидент прекращения курения.

При сопоставлении данных о заболеваемости на территориях удобно использовать специальные компьютерные программы «географические информационные системы». Эти программы позволяют расположить данные о заболеваемости на карте и наглядно представить их. Модуль EpiMap входит в бесплатный пакет программ EpiInfo (см. сноску на с. 103), такие возможности есть в SPSS, SAS, во многих других программах.

При сопоставлении показателей, полученных на двух территориях, например при сопоставлении концентрации холестерина в плазме крови в случайной выборке из населения, правильным будет использование *t*-критерия. У этого критерия есть ряд достоинств, которые определили его широкое применение в медицинских исследованиях, и ряд недостатков. У критерия есть несколько вариантов применения. Ниже они будут рассмотрены на примерах.

### 2.5.2. Применение *t*-критерия

*t*-Критерий Стьюдента предназначен для сопоставления средних величин двух выборок. Чем точнее определены средние (это зависит от ве-



личины СКО и численности выборки), тем меньшие различия между ними можно выявить. Критерий существует в двух вариантах — для связанных выборок и для независимых выборок. Эти варианты должны использоваться правильно, иначе возможно завышение или занижение статистической значимости различий.

Применительно к качественным событиям критерий дает правильную оценку  $\alpha$ -риска только в случае, если в обеих выборках численность выборок  $n$ , а также вероятности события  $p$  и отсутствия события  $(1 - p)$  достаточно велики, т.е.  $n \times p > 5$  и  $n \times (1 - p) > 5$ .

Применительно к количественным признакам использование критерия допустимо при соблюдении ряда условий:

1. Желательна непрерывная, а не дискретная шкала переменной.
2. Дисперсии равны.
3. Распределение нормальное.
4. Выборки достаточно велики.

Если эти условия не соблюдаются, то надо использовать U-критерий или другие непараметрические критерии (см. разд. 2.4.5).

Предположение о нормальности распределения следует проверять уже на этапе подготовки данных: построение гистограмм позволит выявить асимметричные распределения и при необходимости использовать нормализацию данных (см. разд. 1.6). Поскольку распределения, отличающиеся от нормального, встречаются очень часто, а равные дисперсии — редко, некоторые специалисты рекомендуют в медико-биологических исследованиях от  $t$ -критерия отказаться. Все же  $t$ -критерий продолжает широко использоваться, но не потому, что такие соображения несправедливы, а потому, что традиционно существует надежда на устойчивость критерия к отклонениям от условий обоснованного применения. Это привлекательное свойство статистического критерия называют *робастностью* (англ. *robustness*). Критерий действительно устойчив к нарушениям условий его применения, перечисленным выше, но только при изучении больших выборок. Правильным решением для исследователя должно быть использование соответствующих критериев и соответствующей трансформации данных, чтобы сделать применение  $t$ -критерия оправданным.

Использование  $t$ -критерия для сравнения двух выборок с двух территорий позволяет оценить статистическую значимость различия. Рассмотрим применение  $t$ -критерия при оценке различий в заболеваемости на двух территориях (вариант исходных данных — частота каче-



ственных признаков) и к оценке различий в количественном признаке в выборках из двух популяций (вариант исходных данных — количественный признак).

### 2.5.3. Сравнение частот

В общем виде

$$t = (p_a - p_b) / \sqrt{m_a^2 + m_b^2}.$$

В примере, приведенном в табл. 2.14, мы сравниваем преваленсы болезни на двух территориях. Используя приведенную выше формулу, мы получаем  $t = 0,02/0,0022 = 9,09$ . Приложение 1, которое мы использовали для расчета доверительных интервалов, содержит критические величины для оценки величины двустороннего критерия. Разница в преваленсе в данном примере статистически значима при вероятности ошибки типа I менее 0,01.

Таблица 2.14. Сравнение преваленса болезни на двух территориях

Рассчитываемые показатели	Территория А	Территория В
Размер выборки ( $n$ )	20 000	50 000
Больны ( $r$ )	2000	6000
Преваленс ( $P_r$ )	0,10	0,12
Здоровы ( $q = 1 - P_r$ )	0,90	0,88
Относительный риск ( $OR = P_B/P_A$ )	Референтная величина	1,2
Разность риска ( $D$ )	Референтная величина	0,02
Стандартная ошибка $P_r$ ( $m = \sqrt{\frac{P_r \times (1 - P_r)}{n}}$ )	0,002	0,001
Доверительный интервал для $P_r$ ( $\pm 1,96m$ )	0,096–0,104	0,118–0,122

При использовании таблиц  $t$ -критерия, подобной приложению 1, надо иметь в виду, что под числом степеней свободы обычно понимают число изученных случаев минус число оцениваемых параметров. В слу-



чае  $t$ -критерия оценивают различия между группами на основании среднего квадратического отклонения (СКО), и число степеней свободы составляет  $(n - 1)$ . Статистические программы не только позволяют вычислить величину критерия  $t$ , но и вычислить точно значение  $P$ . Нужды в использовании таблиц не возникает.

Предпочтительно не оценивать вероятности ошибки типа I для гипотезы о различии двух пропорций, а вычислять ДИ для разности двух пропорций. Процедура основывается на той же методологии, но результат получается более наглядный. Для разности пропорций в двух несвязанных выборках стандартная ошибка

$$m = \sqrt{\frac{p_a(1-p_a)}{n_a} + \frac{p_b(1-p_b)}{n_b}}.$$

В нашем примере  $m = 0,0026$ . Соответственно, ДИ простирается от  $D - z_{1-\alpha/2} \times m$  до  $D + z_{1-\alpha/2} \times m$ . Для вычисления ДИ  $z_{1-\alpha/2}$  получаем в приложении 1. Для ДИ 95% это значение 1,96 при числе степеней свободы более 500. Значит, нижняя и верхняя границы ДИ 0,015 и 0,025.

#### 2.5.4. Сравнение средних

В случае если две выборки сравниваются по количественному признаку и распределения результатов измерений близки к нормальному, то применение  $t$ -критерия является методом выбора, так как он обладает наибольшей статистической мощностью. Принцип, на котором построен  $t$ -критерий, — сравнение разности между сравниваемыми группами с точностью ее оценки.

##### Несвязанные выборки

Точность оценки есть стандартная ошибка разности  $m_D$ . Критерий вычисляется по формуле:

$$t = \frac{\bar{X}_a - \bar{X}_b}{m_D},$$

и затем его величина сопоставляется с  $t$ -распределением (приложение 1) в соответствии с числом степеней свободы. Последнее для сравнения двух групп вычисляется как  $(n_a + n_b - 2)$ , где  $n_a$  и  $n_b$  — численность групп А и В.



Относительно сложно вычисление  $m_D$ . Для этого сначала вычисляют суммарную дисперсию:

$$s^2 = \frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{n_a + n_b - 2},$$

где  $s_a$  и  $s_b$  — стандартные отклонения для групп А и В соответственно. Используя вычисленную дисперсию, можно получить  $m_D$ :

$$m_D = s \times \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}.$$

Применим эти формулы к данным, использованным ранее (см. табл. 2.8). Будем полагать, что группы не связаны между собой. Величина критерия будет 1,45, что меньше критического в приложении 1.

Вычисление  $t$ -критерия доступно не только в программах для статистического анализа, но даже в Excel. При использовании компьютерных программ пользователь получает величину  $t$  и точное значение  $P$ . Так, использование Excel для нашего примера дает  $P = 0,16$ . При «ручных» вычислениях и использовании таблицы критических значений  $t$  вероятность ошибки типа 1 оценивается в виде  $P > 0,05$ , что делает суждения по результатам анализа менее уверенными.

### Связанные выборки

Типичным примером связанных или парных данных являются данные, полученные у одних и тех же людей до и после лечения или до и после работы. Реже такого типа данные получают, сопоставляя результаты вмешательства у разных людей, если эти лица были подобраны в пары. Например, если результат лечения сильно зависит от тяжести болезни, то для сравнения двух препаратов логично подобрать пары пациентов с равной тяжестью болезни. Такой пример мы рассматривали в табл. 2.8. Применим к нему  $t$ -критерий (табл. 2.15).

Для этого вычислим разность в результатах лечения в парах пациентов. Эта разность становится объектом оценки как самостоятельная переменная. Мы можем вычислить ее среднюю величину, стандартное отклонение и стандартную ошибку так, как мы это делали в разд. 1.6. Для  $D$  эти параметры составят: 1,26; 1,47; 0,47. Теперь

$$t = (D - 0) / m = 1,26 / 0,47 = 2,7.$$



Таблица 2.15. Сравнение двух вмешательств в подобранных парах пациентов по функциональным возможностям пациентов

Лечение	Возможности пациентов, условные единицы									
	А	16,2	15,2	17,8	17,3	16,1	12,7	13,3	18,2	18,3
В	14,1	14,2	16,3	16,6	16,1	14,3	12,5	16,7	15,1	11,0
Разность (D)	2,1	1	1,5	0,7	0	-1,6	0,8	1,5	3,2	3,4

В числителе фигурирует разность  $D=0$ , поскольку формально мы сравниваем обнаруженную разность с гипотетическим отсутствием эффекта. Сравнение с критическими значениями в приложении 1 показывает, что в группе, получавшей препарат А, достигнут более высокий результат, чем в группе, получавшей В: нулевая гипотеза может быть отклонена при  $P < 0,05$ . Вычисление в Excel дает  $P = 0,02$ . На этом примере мы еще раз видим, что правильная организация исследования — испытание в парах пациентов — позволяет применить более эффективные статистические методы и выявить различия, которые при упрощенной структуре исследования остались бы не выявленными.

### 2.5.5. Дисперсионный анализ

Первоначально дисперсионный анализ (ANOVA) был создан для оценки эффектов двух видов лечения, применяемых по отдельности и вместе. Сейчас метод в основном применяется:

- ▶ для сравнения более чем двух групп, т.е. в ситуациях, когда одного парного сравнения с помощью, например,  $t$ -критерия недостаточно;
- ▶ в случае, когда в изучаемых группах пациентов проводится более двух последовательных измерений;
- ▶ в случаях измерения двух и более признаков, способных повлиять на зависимую величину (исход), с тем, чтобы устранить (учесть) такое влияние привходящих факторов.

Метод состоит в сопоставлении различий между средними величинами выборок (дисперсии средних) с дисперсией в изучаемых выборках. Так можно установить, существует ли различие между группами или их различия случайны. После того как различие групп установлено, можно применить другие методы для локализации различий, т.е. установить, какие именно группы различаются между собой [4].



В случае простого сравнения двух групп результат применения ANOVA совпадает с результатом сравнения по *t*-критерию. ANOVA — более общий случай исследования различий, когда возможно сравнение большего числа групп и большего числа признаков. Для наиболее частого случая влияния двух независимых переменных (предикторов) на результаты эксперимента оптимальный критерий можно избрать по табл. 2.16.

Таблица 2.16. Оптимальные варианты использования ANOVA при разнородных типах данных

Тип данных второго предиктора	Тип данных первого предиктора	
	Номинативный	Количественный
Качественный	Двухмерный ANOVA	ANCOVA
Номинативный	Двухмерный ANOVA	ANCOVA
Количественный	ANCOVA	Множественная регрессия

ANOVA — дисперсионный анализ; ANCOVA — ковариационный анализ.

Для ANOVA не очень важна нормальность распределения количественного признака, но необходимо соблюдение следующих условий.

А. Линейность зависимостей между переменными. Если различия между группами больных по эффективности лечения  $y$  зависят от какого-нибудь признака, например артериального давления  $x$ , то эта зависимость должна быть или линейной ( $y = ax$ , где  $x$  — давление) или квадратической ( $y = ax^2$ ), но не степенной ( $y = axz$ ), где  $z$  — некоторая другая переменная. Нормализация распределения исходных данных логарифмированием (см. врезку на с. 111) может существенно изменить результат анализа. Точно так же, если вместо сырых данных используются проценты, вычисленные к исходной величине, то нужна логарифмическая трансформация.

Б. Равенство дисперсий. Равенство дисперсий в изученных группах необходимо так же, как для применения *t*-критерия. Как правило, в биологии ситуация иная. Дисперсия в группе часто пропорциональна средней: чем больше средняя, тем больше и дисперсия (см. рис. 1.36). Дисперсия возрастает также после воздействия на группу организмов внешнего фактора. В случае если существует выраженная зависимость



дисперсии от средней, может быть применена логарифмическая трансформация. Логарифм изучаемого признака, вероятно, не будет иметь зависимости дисперсии от средней величины. Впрочем, делая вычисления корректными, трансформация вносит затруднения: после вычислений необходимо перевести полученные оценки в исходные величины, что может быть непросто.

**В. Отсутствие «хвостов» в распределениях.** Нередко при форме распределения, близкой к нормальной, наблюдается незначительное число результатов измерения, отличающихся очень большим значением («жирный хвост»). В этой ситуации особенно опасно использование  $t$ -критерия и ANOVA.

При использовании ANOVA для сравнения нескольких групп различия в численности групп не должны быть очень большими. Число групп тоже не должно быть большим. В противном случае при дроблении всей совокупности пациентов на множество групп численность каждой группы уменьшается, и все различия становятся статистически незначимыми. К такому же результату приводит применение многомерного анализа к анализу большого числа признаков. Например, если используются только качественные признаки, но число их велико (например, 5), то число комбинаций составляет  $2^5 = 32$ , соответственно, число наблюдений разбивается на 32 ячейки.

Частым способом организации исследования является повторное обследование одних и тех же лиц (в течение нескольких минут, месяцев, лет). При представлении таких данных авторы статей часто ограничиваются изображением средних величин в разные моменты времени и соединяют их линиями. В действительности такая «средняя динамика» может не отражать того, что наблюдается у отдельных лиц. Более того, при этом за счет «выпадения» (миграции) отдельных лиц может изменяться эта «средняя динамика». Например, в старшем возрасте средняя масса тела перестает расти и даже может снижаться, но совсем не из-за того, что худеют отдельно взятые субъекты, а потому, что из изучаемой группы «выпадают» (умирают) лица с большой массой тела. Вероятно, что подобной «миграцией» объясняется более высокая концентрация тестостерона в крови мужчин старшего возраста в примере поперечного исследования, приведенного на рис. 2.12. ANOVA является методом выбора для анализа результатов исследований с повторными измерениями. Могут для этих целей использоваться и другие критерии, но важно, чтобы они предусматривали именно анализ повторных измерений.



Поскольку ANOVA для ручных вычислений малопригоден, обычно применяют его варианты, реализованные в пакетах статистических программ. В отдельных пакетах реализовано до десятка вариантов анализа, не считая деталей их выполнения. Применение этих вариантов может приносить несколько различные результаты. Ввиду обилия вариантов ANOVA и потенциальных проблем с его применением исследования, в которых предполагается его использование, целесообразно планировать с участием специалиста по статистике.

### 2.5.6. Сравнение множества групп

Использование вероятности 0,05 в качестве порога статистической значимости означает признание допустимости ошибиться в одном случае из 20. Рассмотрим простой пример сравнения двух групп. Представим, что мы оцениваем статистическую значимость различий двух *одинаковых* групп (случайно выбранных из популяции) по 100 признакам. Общепринятый порог статистической значимости обеспечит нам обнаружение статистически значимых различий примерно по 5 признакам. Такая ошибка совершается нередко. Нельзя использовать *t*-критерий или другой критерий статистической значимости для попарного сравнения множества групп по одному признаку или двух групп по множеству признаков. Наиболее частой ошибкой является использование сравнений 3–4 групп по нескольким признакам. Например, нельзя сравнивать несколько территорий по заболеваемости десятками болезней с помощью попарного сравнения по инциденту каждой болезни. При таких *множественных сравнениях* возрастает вероятность ложного обнаружения «статистически значимого» различия.

Для множественных сравнений должен использоваться дисперсионный анализ. После того как найдено, что группы различаются, можно попарными сравнениями искать, какие из различий между группами достигают порога статистической значимости. Этот порог, конечно, не может быть таким же, как для одного сравнения двух групп. Точнее, мы можем исходить из той же вероятности ошибки первого типа, но этот уровень относится *ко всем группам*. Иными словами, если мы выполняем *k* сравнений между группами, то для сохранения уровня вероятности ошибки типа I надо снизить эту вероятность применительно к каждому парному сравнению в *k* раз. Например, при выполнении 4 парных



сравнений групп между собой надо вместо  $P = 0,05$  применить критический уровень ошибки  $P = 0,01$ . К сожалению, эта простая поправка Бонферрони слишком поднимает планку статистической значимости, если выполняется более 4–5 парных сравнений. Поэтому в статистических пакетах программ используются методы Дункана, Ньюмана–Кейлса и Шеффе, которые дают не столь консервативные оценки статистической значимости при множественных сравнениях.

## Контрольные вопросы и задачи к разделу 2.5

1. Представим, что смертность от рака легких у курящих более 20 сигарет в день водителей составляет 200 на 100 000, а у некурящих — 10 на 100 000. Каков относительный риск смерти от рака легких у курильщиков?
  - а) его нельзя высчитать из этих данных;
  - б) 20;
  - в) 190;
  - г) 10.
2. При сравнении двух групп величина  $P = 0,15$ . Это означает:
  - а) различие является случайным;
  - б) случайное происхождение различия не может быть отклонено;
  - в) нулевая гипотеза отвергается;
  - г) различие не существует.
3. Какие предположения должны быть справедливы в отношении анализируемых данных, чтобы можно было применять  $t$ -критерий Стьюдента (выберите все правильные ответы)?
  - а) независимые выборки;
  - б) стандартные отклонения приблизительно равны;
  - в) данные нормально распределены;
  - г) выборки равного размера.
4. Для сравнения преваленса HbsAg антигена у медицинских сестер в терапевтических отделениях и отделениях диализа правильнее применить критерий:
  - а)  $\chi^2$ ;
  - б)  $t$ -критерий Стьюдента;
  - в) дисперсионный анализ;
  - г) линейную регрессию;
  - д)  $t$ -критерий для связанных выборок.



## 2.6. ВЕРОЯТНОСТЬ ЗАБОЛЕВАНИЯ И СМЕРТИ В ТЕЧЕНИЕ ЖИЗНИ ЧЕЛОВЕКА

Смертность, т.е. вероятность наступления смерти, является популяционным показателем. Этот показатель отражает вероятность смерти у отдельного человека из популяции. С медицинской точки зрения самые важные изменения, происходящие в течение жизни человека, выражаются смертностью. Само старение определяется прежде всего как процесс нарастания смертности.

Поскольку факт наступления смерти регистрируется относительно просто и рождение и смерть имеют особую социальную важность, процессы рождаемости и смертности всегда находились в сфере интересов не только медицины, но и социальных наук, математики. Объектом науки демографии являются именно процессы рождаемости и смертности, а также более тонкие вопросы причин смерти, миграции, связи показателей рождаемости и смертности с разнообразными характеристиками населения. У демографии множество точек соприкосновения с эпидемиологией и социальными науками. Методы, разработанные для изучения смертности людей применительно к практике страховых компаний, стали ныне главными методами статистического анализа «событий типа времени жизни», т.е. длительности ремиссии, длительности выживания, длительности выздоровления.

### 2.6.1. Демографические данные о смертности

Продолжительность жизни и смертность связаны между собою простым отношением. Смертность  $M$  вычисляется, как число умерших в некоторый период (например, год,  $D$ ), отнесенное к числу людей, наблюдавшихся в этом году ( $N$ ):

$$M = D/N.$$

Чем большая часть населения умирает ежегодно, тем короче продолжительность жизни. Упрощенно, если в популяции стабильной численности в год умирает 2% населения, то это значит, что продолжительность жизни составляет 50 лет. Продолжительность жизни ( $L$ ) соотносится со смертностью, как:

$$L = 1/M.$$



Вычисленные таким образом смертность и продолжительность жизни правильно отражают только события, происходящие в гипотетической популяции, где численность постоянна, а смертность одинакова во всех возрастах. Действительность сильно отличается от упрощенной модели. В Колымских лагерях смертность превышала 400%, т.е. в течение года умирало четыре населения лагеря. Однако даже в таких условиях некоторые редкие люди выживали там годами. В любых условиях распределение людей по продолжительности жизни асимметрично: небольшое число людей умирает в детстве и немногие живут относительно продолжительную жизнь. На рис. 2.40 приведены данные о смертности и продолжительности жизни в последнем столетии в США. Три способа описания продолжительности жизни используются в разных целях. Кривые выживания (верхний рисунок) показывают, какая доля рожденных остается живой в разных возрастах. Распределение по возрасту смерти (средний рисунок) показывает, в каком возрасте сколько людей умирает. Например, хорошо видно, что в течение 100 лет резко уменьшилась доля людей, умирающих в детском и юношеском возрасте, и пик смертности в старшем возрасте сместился вправо, к старческому возрасту. Вероятность смерти в течение года для живых в начале данного возрастного периода (нижний рисунок) лучше всего отражает изменения смертности с возрастом.

Поскольку смертность имеет такое сложное распределение по возрасту, описать ее одним параметром невозможно. Когда в обыденной речи употребляют выражение «средняя продолжительность жизни», то обычно имеют в виду предстоящую продолжительность жизни для людей, рожденных в конкретном году. Этот показатель учитывает сложный характер распределения смертности по возрасту, но одновременно скрывает важные детали. Например, продолжительность предстоящей жизни у новорожденных меньше, чем у проживших год. Это объясняется тем, что новорожденному предстоит прожить первый год, когда смертность очень высока. Вычисление медианы возраста при наступлении смерти открывает удивительные факты. В экономически развитых странах (Европы, Северной Америки, Австралии) медиана выше 75 лет, т.е. отражает обыденные представления о наступлении смерти в старости (медиана соответствует правому «горбу» распределения на рис. 2.10 в центре). В большинстве стран Центральной Африки и в некоторых странах Южной и Юго-Восточной Азии медиана возраста при наступлении смерти менее 10 лет. Это означает, что в короткий период от рождения до 10 лет погибает половина родившихся.



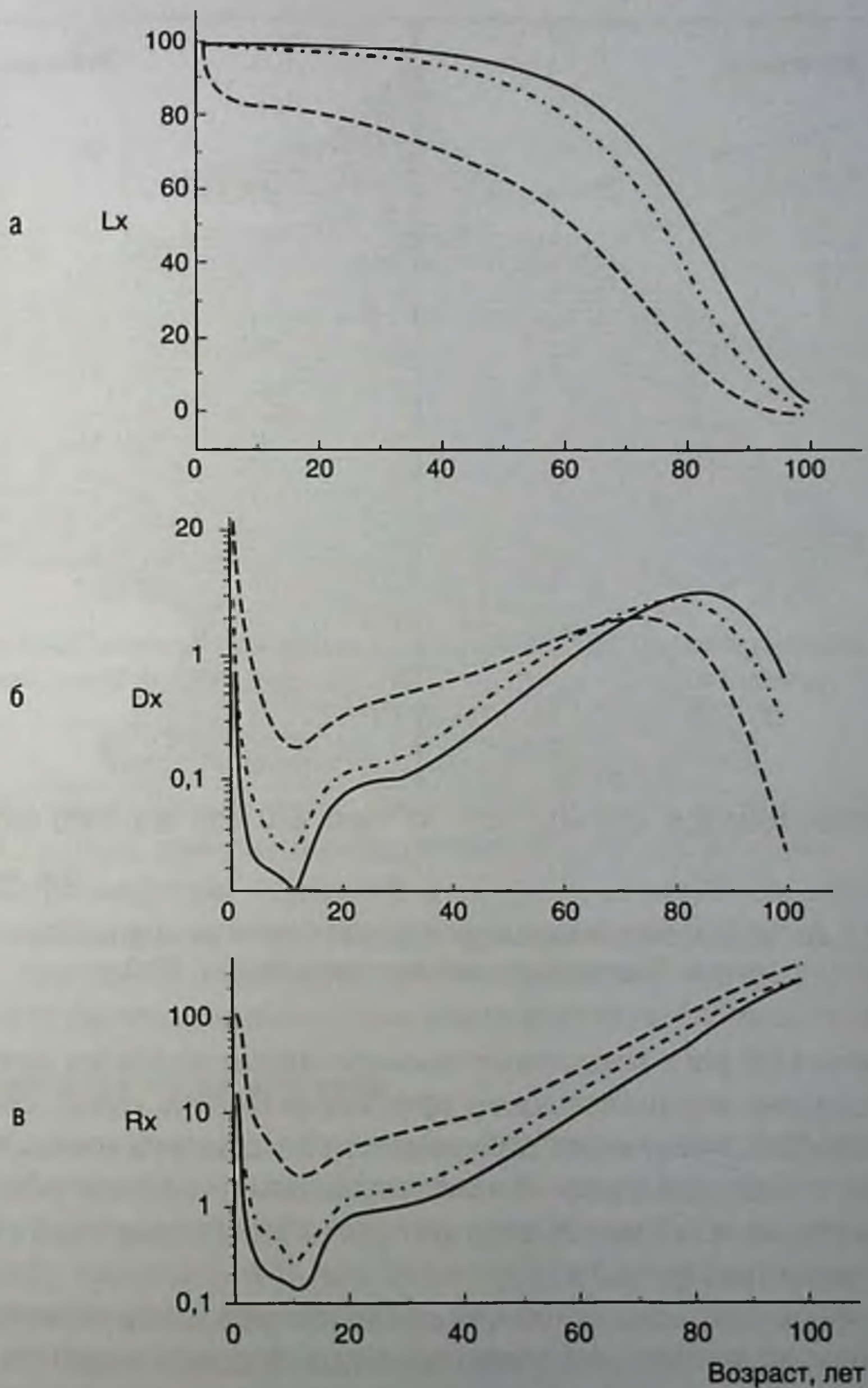


Рис. 2.40. Изменения смертности в XX в. США (1910 г. — пунктир, 1970 г. — штрих-пунктир, 1998 г. — линия): а — доля живых среди рожденных в изучаемом году ( $L_x$ ); б — доля умирающих в данном возрасте от рожденных в изучаемом году ( $D_x$ ); в — доля умирающих за год от живых в начале данного возрастного периода (по данным USA National Vital Statistics Report)



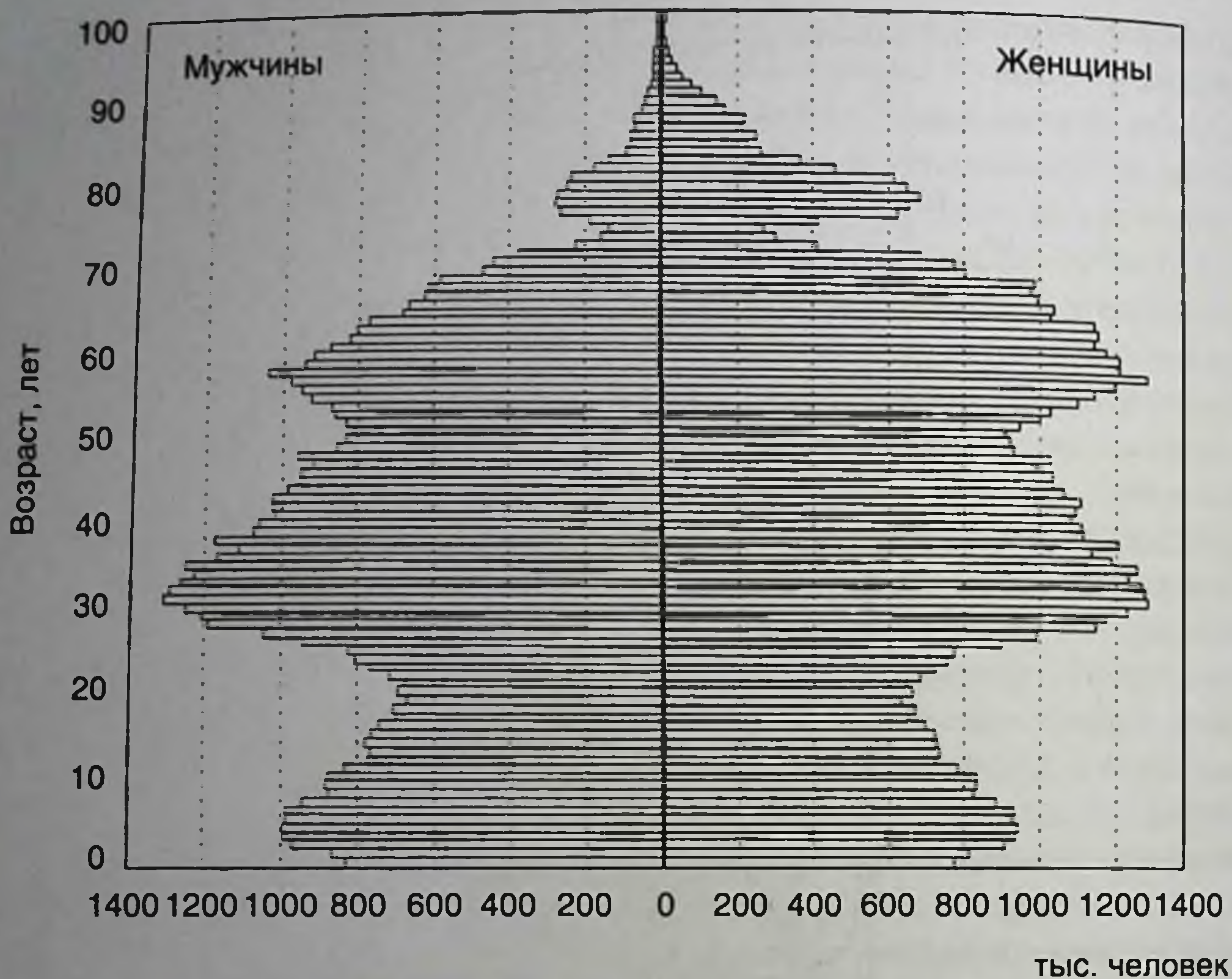


Рис. 2.41. Половозрастная пирамида населения России на начало 2019 г.  
(по данным: Демографический ежегодник России, 2019 г.)

Исторический рост продолжительности жизни человека нелинеен, он деформирован периодическими эффектами (война, голод, социальные катастрофы), в результате половозрастная структура населения несет на себе следы прошлого. В нем сокращенное войной поколение быстрее вымирает и оставляет мало детей, что эхом отражается в последующих поколениях (рис. 2.41).

Предстоящая продолжительность жизни при рождении является важнейшим показателем благополучия населения. В самом широком смысле неуклонное увеличение продолжительности жизни людей является важнейшим доказательством того, что прогресс существует. С продолжительностью жизни человека связано немало заблуждений, некоторые из которых происходят из неправильного использования эпидемиологических и демографических данных. Главным заблуждением является



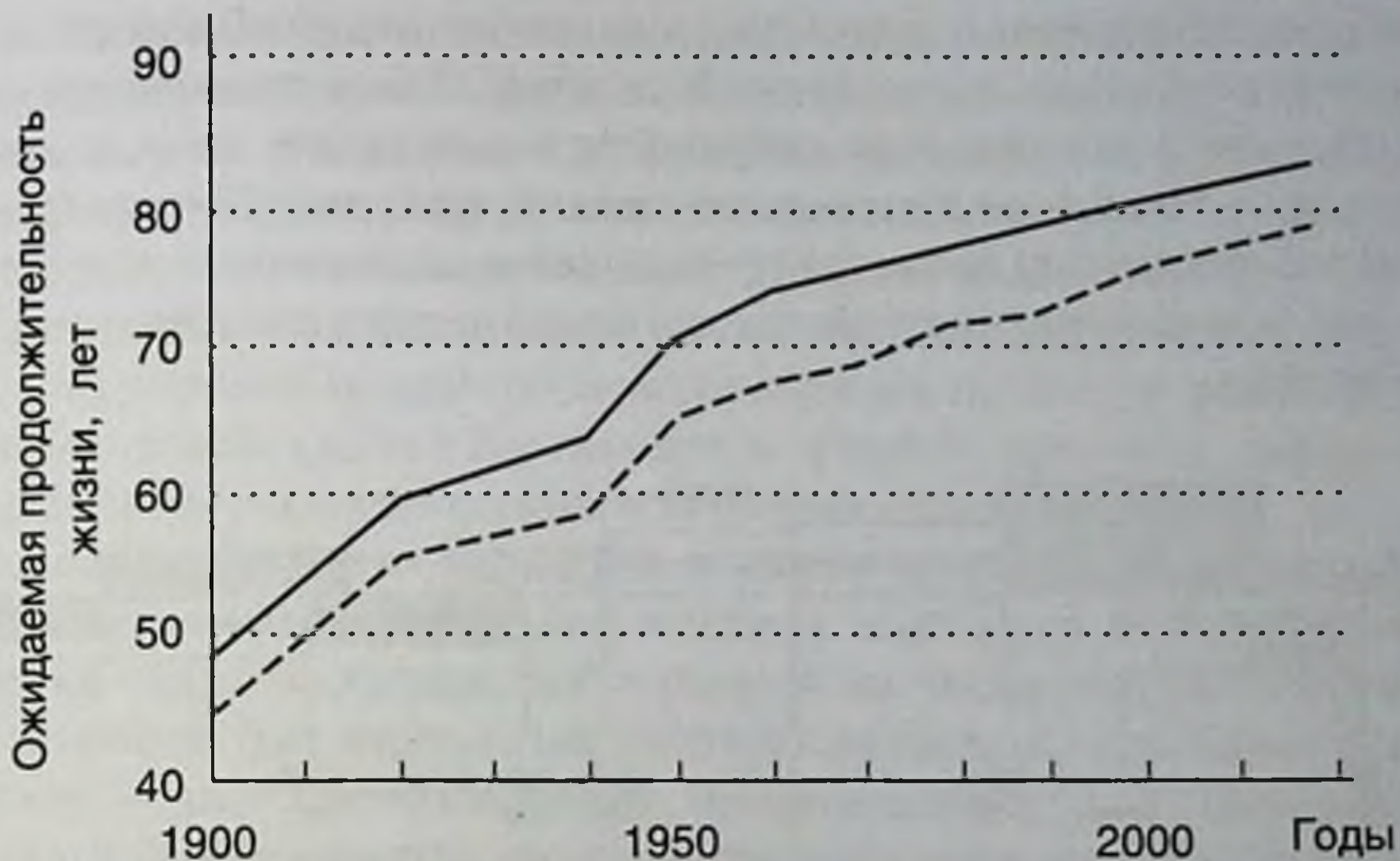


Рис. 2.42. Ожидаемая продолжительность жизни при рождении у женщин (верхняя линия) и мужчин (пунктир). После 2000 г. — прогнозируемые величины (по данным Oerpen J., Vaupel J. Science. 2002. № 296. P. 1029)

предсказание *максимальной* продолжительности жизни человека. Весь объем данных, которыми располагают демографы в странах, где веками ведется регистрация рождений и смертей, доказывает, что продолжительность жизни человека продолжает увеличиваться (рис. 2.42). В странах с тяжелыми условиями жизни продолжительность жизни ниже, но может быстро увеличиваться при нарастании благосостояния.

### Периодизация жизни человека

По совокупности признаков отдельных возрастов человека его жизнь более или менее условно разделяется на ряд периодов (рис. 2.43). Такие периоды, как беременность (внутриутробное развитие), перинатальный, неонатальный, младенческий, выделяются преимущественно по биологическим признакам — по процессам, происходящим в организме человека, по патологии, которая в это время развивается, по уровню риска патологии и смерти. Другие периоды выделяются преимущественно на основании характеристик социального функционирования человека: дошкольный, школьный, подростковый и т.д. Соответственно, периодизация младенческого периода является международно-признанной, а последующие периоды не вполне совпа-



дают в разных странах в зависимости от особенностей образовательной системы, типа семьи и других особенностей<sup>1</sup>. Такое деление применяется не только для описания смертности в популяции, но и во многих областях медицинской и исследовательской практики. Например, клинические испытания лекарств проводятся в аналогичных возрастных группах, и лекарства разрешаются для применения в подобных возрастных группах.

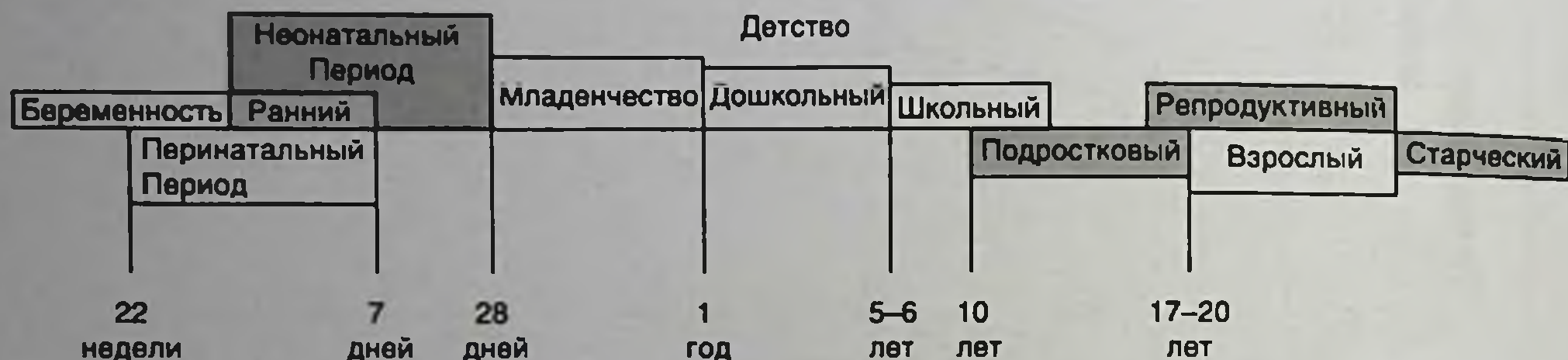


Рис. 2.43. Периодизация жизни человека

При оценке заболеваемости населения в целом инцидент и превалентенс рассчитывают к численности населения (например,  $n/100\ 000$ ). Если оценивается патология, которая специфична для какого-то возраста, или отдельный возраст или пол рассматриваются как наиболее поражаемый, то рассчитываются специальные показатели. Так, смертность от рака молочной железы может рассчитываться к численности населения, поскольку мужчины также страдают от этой болезни, но чаще расчет делается на число женщин (число случаев на 100 000). Это обусловлено тем, что соотношение мужчин и женщин в популяции непостоянно, и расчет на всю популяцию дает менее точные оценки для женщин. Правильный расчет смертности выполняется по пропорции, в которой в числителе — число заболевших (умерших) в данной группе, а в знаменателе — численность данной группы. Например, число женщин, умерших от рака влагалища, к численности женщин в США дает показатель специфической смертности 0,2 на 100 000.

Применительно к деторождению и связанной с ним патологии в качестве знаменателя, очевидно, нельзя использовать численность попу-

<sup>1</sup> Например, Конвенция ООН «О правах ребенка» устанавливает в целях защиты прав детей детский возраст до достижения 18 лет, если по закону ребенок не достигает совершеннолетия ранее (<http://www.memo.ru/prawo/child/891120.htm>).



ляции. Однако нельзя использовать и численность женщин, поскольку лишь часть женщин находится в детородном возрасте. В популяции, где много детей или стариков, показатели рождаемости и осложнений могут быть неправильными, несопоставимыми с другими странами или с этой же страной в другие годы. Поэтому рассчитывают отношения, в которых в знаменателе — число женщин фертильного возраста. Таким образом можно правильнее оценить истинную рождаемость и инцидент осложнений беременности и родов, получить цифры для сопоставления регионов страны или разных стран.

Поскольку в разных странах и в разное время различается частота беременностей у фертильных женщин, постольку приблизительны и оценки частоты осложнений в расчете на число женщин фертильного возраста. Для этого используются показатели, в которых в знаменателе — число беременностей, закончившихся рождением живого ребенка. Выбор в качестве «популяции» (знаменателя) числа живорождений является компромиссом. Использовать в качестве знаменателя все беременности или все роды нельзя, так как определение родов и разграничение родов и аборта недостаточно четко; в большинстве стран беременности, не закончившиеся рождением живого ребенка, регистрируются неполно. Поэтому показатели смертности младенцев и материнской смертности рассчитываются к числу рожденных живыми. Критерии живорождения могут различаться в разных странах. В России используются рекомендованные ВОЗ критерии живорождения, основанные на извлечении из тела матери продукта зачатия, соответствующего признакам срока беременности, массы тела и наличия признаков жизнедеятельности. Переход на обновленные критерии объясняет повышение показателя младенческой смертности в начале 1990-х и в 2012 г. (рис. 2.44).

Показатели младенческой и материнской смертности — наиболее чувствительные индикаторы социального благополучия населения и качества оказания медицинской помощи. В то время как показатели здоровья всего населения (предстоящая продолжительность жизни, инвалидность, смертность от хронических заболеваний, трудопотери в связи с заболеваниями) изменяются медленно, обычно в течение десятилетий, изменения материнской и младенческой смертности могут достигаться быстро, если население обеспечено доступом к доброкачественной пище, жилью и медицинской помощи (см. рис. 2.44). Показатель младенческой смертности может давать неверное представление



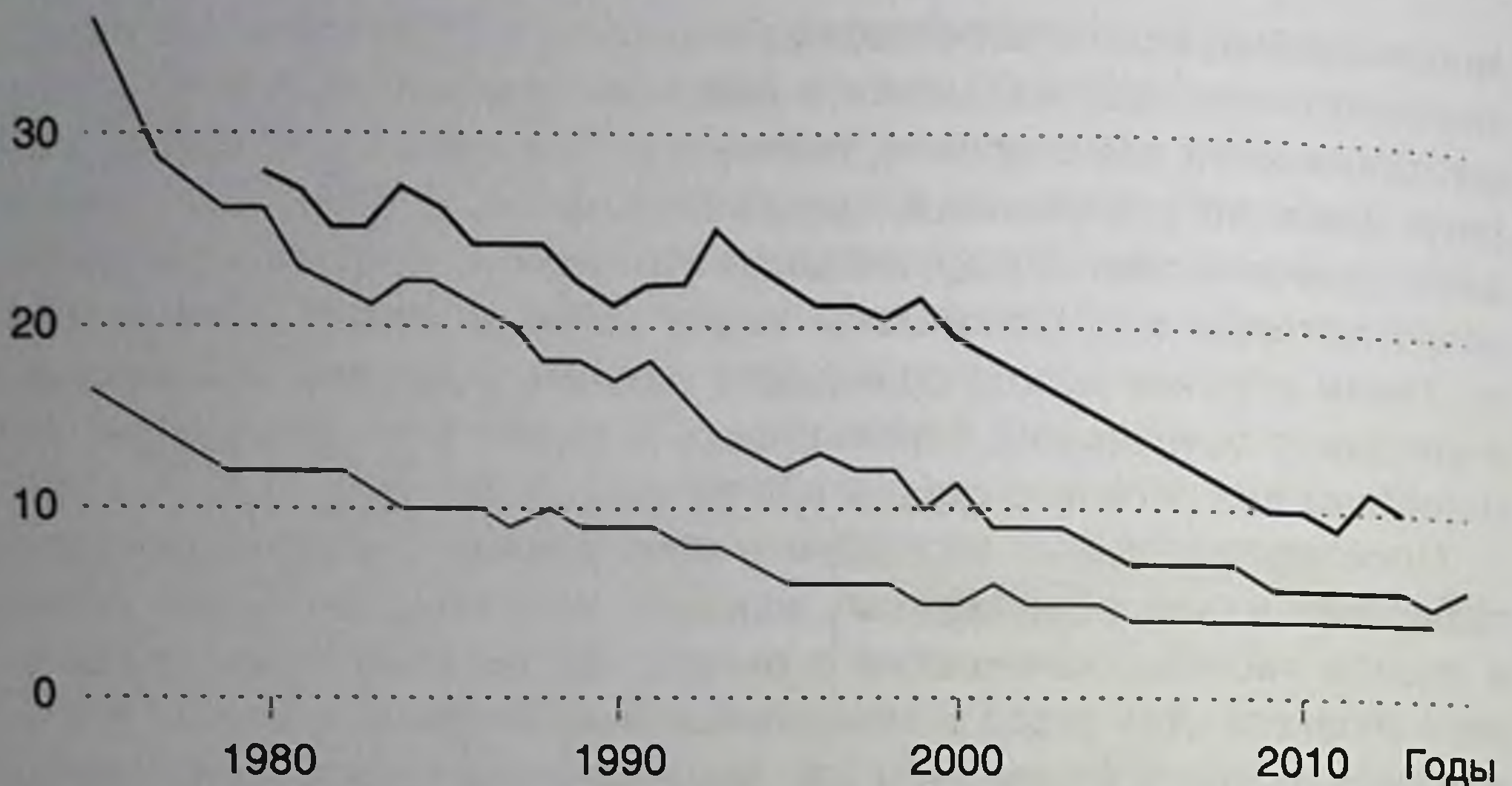


Рис. 2.44. Младенческая смертность в России, Венгрии и Франции (линии сверху вниз) на 1000 живорожденных. Построено средствами European Health Information Gateway (<https://gateway.euro.who.int/en/hfa-explorer/>)

о смертности в случае быстрых изменений, поскольку в числителе присутствуют смерти детей, родившихся в прошлом году, а часть детей из знаменателя умрут в следующем году. Показатель материнской смертности несколько завышается за счет того, что не учитываются случаи рождения мертвых детей, и занижается за счет того, что не учитываются многоплодные беременности.

### 2.6.2. Анализ выживания

Для описания продолжительности жизни можно применять те же способы, которые используются для описания других количественных признаков: среднюю, СКО, медиану, моду. Поскольку распределение по времени жизни всегда резко асимметрично, вычисление средней — ошибка, и следует использовать медиану.

Расчет медианы возможен, если исследователь знает исходы для всех пациентов. Обычно в медицинском исследовании остается неизвестной судьба части пациентов, переехавших в другой город или по другим причинам выпавшим из поля зрения исследователя. Таких пациентов называют потерянными при наблюдении, а в статистическом анализе такие случаи называют *цензурированными* (англ. *censored*, т.е. урезан-



ными). Поскольку они находились в исследовании в течение некоторого времени, желательно использовать информацию о них за это время, например за первый год наблюдения. С другой стороны, к моменту анализа данных некоторые пациенты могут быть еще живы. Это тоже не случайные пациенты. Многие из них — самые долгоживущие из включенных в исследование. Естественно, среднюю продолжительность жизни нельзя вычислять, если не все наблюдавшиеся больные умерли. Представим, что 100 пациентов наблюдались 5 лет и 60 из них умерли, а 40 пациентов еще живы и будут жить неизвестный нам срок. Если вычислить среднюю продолжительность жизни только умерших, то она будет, например, всего 2,5 года.

Для того чтобы полно использовать имеющиеся данные и получить более точные оценки вероятности выживания до установленных сроков, разработан специальный метод. Метод анализа выживания (англ. *survival analysis*, иногда называется «анализ дожития») применяется широко везде, где нужно описать процесс возникновения со временем некоторых событий, например зачатия ребенка, возникновения осложнений болезни или смерти, наступления ремиссии под влиянием лечения. Такие данные называют *данными типа времени жизни*. Метод заключается в том, что вероятность смерти в ближайший короткий период оценивается по данным обо всех пациентах, обследованных исходно, в следующем периоде такой же длительности — только по данным о тех, кто наблюдался в этом периоде (за исключением «потерянных»), так же в третьем периоде и т.д. В зависимости от организации исследования периоды вычислений могут быть разными — от минут до нескольких лет.

### Актuariальный метод

Самый простой и разработанный исторически первый метод такого анализа называется *актуариальным* (от англ. *actuary* — специалист страховой компании, рассчитывающий риски и выплаты; называется также методом Катлера—Эдерера (Cutler—Edeger)). Предположим сначала, что отслежена судьба всех пациентов (табл. 2.17). Даже к этим полным данным простые методы анализа не могут быть применены. Выживаемость за 5 лет можно узнать только для тех, кто оперирован в 1999 г.; при этом данные об оперированных в последующие годы пропадают. Можно использовать данные обо всех пациентах, но такие данные доступны только для выживаемости в течение 1 года. Даже для больных, отслеженных



в течение 5 лет, распределение по продолжительности жизни обычно резко асимметрично (рис. 2.45). Вычисление медианы не решает проблемы оценки продолжительности жизни. По данным о пациентах, умерших в течение 5 лет, медиана продолжительности жизни равна 1 году, что не отражает нашего впечатления об исходах лечения.

Таблица 2.17. Гипотетический пример лечения пациентов в течение 5 лет, отслежены все пациенты

Год операции	Число пациентов	Живы к концу периода				
		2000 г.	2001 г.	2002 г.	2003 г.	2004 г.
1999	42	20	9	5	3	2
2000	32	–	16	9	5	3
2001	26	–	–	14	8	5
2002	35	–	–	–	18	11
2003	40	–	–	–	–	21

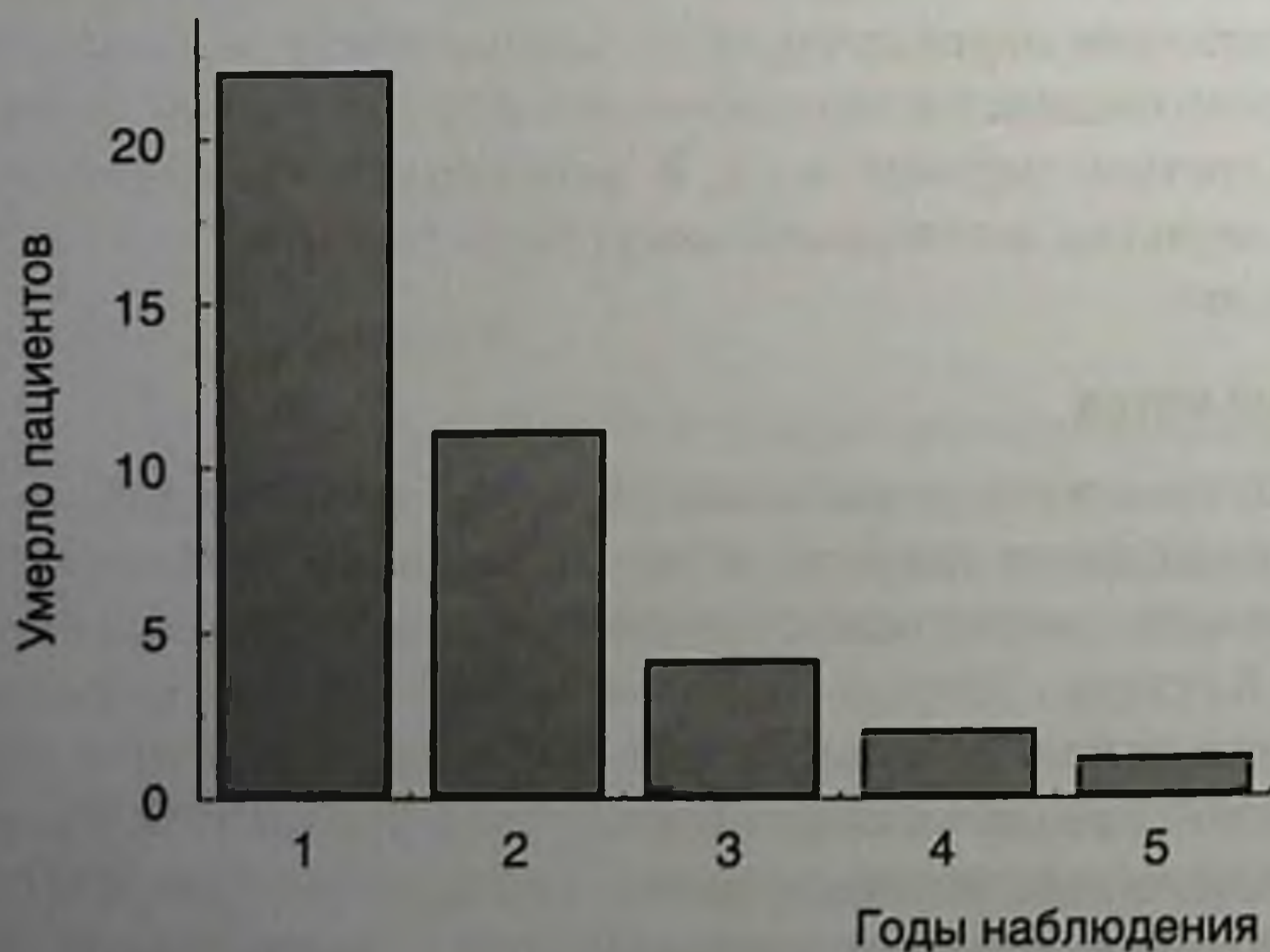


Рис. 2.45. Распределение по продолжительности жизни пациентов, отслеженных в течение 5 лет (пример табл. 2.17)



Для того чтобы использовать данные обо всех пациентах, надо данные реорганизовать, как показано в табл. 2.18. По упорядоченным данным легко рассчитать выживаемость пациентов в течение одного года:  $P_1 = 89/175 = 0,51$ . Для расчета выживания пациентов в течение второго года после операции невозможно использовать данные обо всех 89 пациентах, живых после одного года наблюдения, поскольку 21 пациент не был отслежен более года. Значит, число выживших 37 относится только к числу отслеженных (89–21), т.е. выживаемость составляет:  $P_2 = 37/68 = 0,54$ . Точно так же число переживших третий год (15) относится только к тем, кто был отслежен этот третий год (37–11), т.е.  $P_3 = 15/26 = 0,56$ .

Таблица 2.18. Гипотетический пример лечения пациентов в течение 5 лет, отслежены все пациенты. Данные табл. 2.17 приведены по годам с момента операции

Год операции	Число пациентов	Живы к концу периода (года)				
		1	2	3	4	5
1999	42	20	9	5	3	2
2000	32	16	9	5	3	–
2001	26	14	8	5	–	–
2002	35	18	11	–	–	–
2003	40	21	–	–	–	–
Всего	175	89	37	15	6	2
Вероятность выживания в периоде	–	0,51	0,54	0,56	0,60	0,66

Теперь, когда известно, с какой вероятностью пациенты выживают год за годом, можно рассчитать кумулятивные вероятности выживания (табл. 2.19) и построить кривую выживания (англ. *survival curve*) (рис. 2.46). Мы видим, что медиана выживания действительно несколько менее 1 года и вероятность выживания более 5 лет составляет 6%.

В реальных исследованиях некоторая часть пациентов по разным причинам выпадает из исследования. Соответственно вероятность выживания в отдельном интервале времени рассчитывается как отношение числа выживших до конца интервала к числу пациентов, отслеженных в этом интервале. Выпавшие в течение данного интервала могут быть отслеже-



ны более или менее долго в течение этого интервала. Обычно предполагается, что данные о таких пациентах перестают поступать с одинаковой частотой в течение всего интервала. Соответственно, можно полагать, что в среднем выпадающие в течение данного интервала пациенты наблюдаются  $1/2$  длительности интервала. Иными словами, если наблюдалось до конца интервала 70 и в течение интервала выпало 17 пациентов, то считаются наблюдавшимися в течение интервала  $70 + 17 \times 0,5 = 78,5$ . Расчеты, основанные на методе, близком к актуарному, составляют основу демографических таблиц выживания (см. разд. 2.6.1).

Таблица 2.19. Расчет кумулятивной вероятности выживания (данные табл. 2.17)

Период	Формула	Расчет значения	Сокращенный расчет
1 год	$P_1$	0,51	0,51
2 года	$P_1 \times P_2$	$0,51 \times 0,54 = 0,275$	$0,51 \times 0,54 = 0,275$
3 года	$P_1 \times P_2 \times P_3$	$0,51 \times 0,54 \times 0,56 = 0,154$	$0,275 \times 0,56 = 0,154$
4 года	$P_1 \times P_2 \times P_3 \times P_4$	$0,51 \times 0,54 \times 0,56 \times 0,6 = 0,092$	$0,154 \times 0,6 = 0,092$
5 лет	$P_1 \times P_2 \times P_3 \times P_4 \times P_5$	$0,51 \times 0,54 \times 0,56 \times 0,6 \times 0,66 = 0,061$	$0,092 \times 0,66 = 0,061$

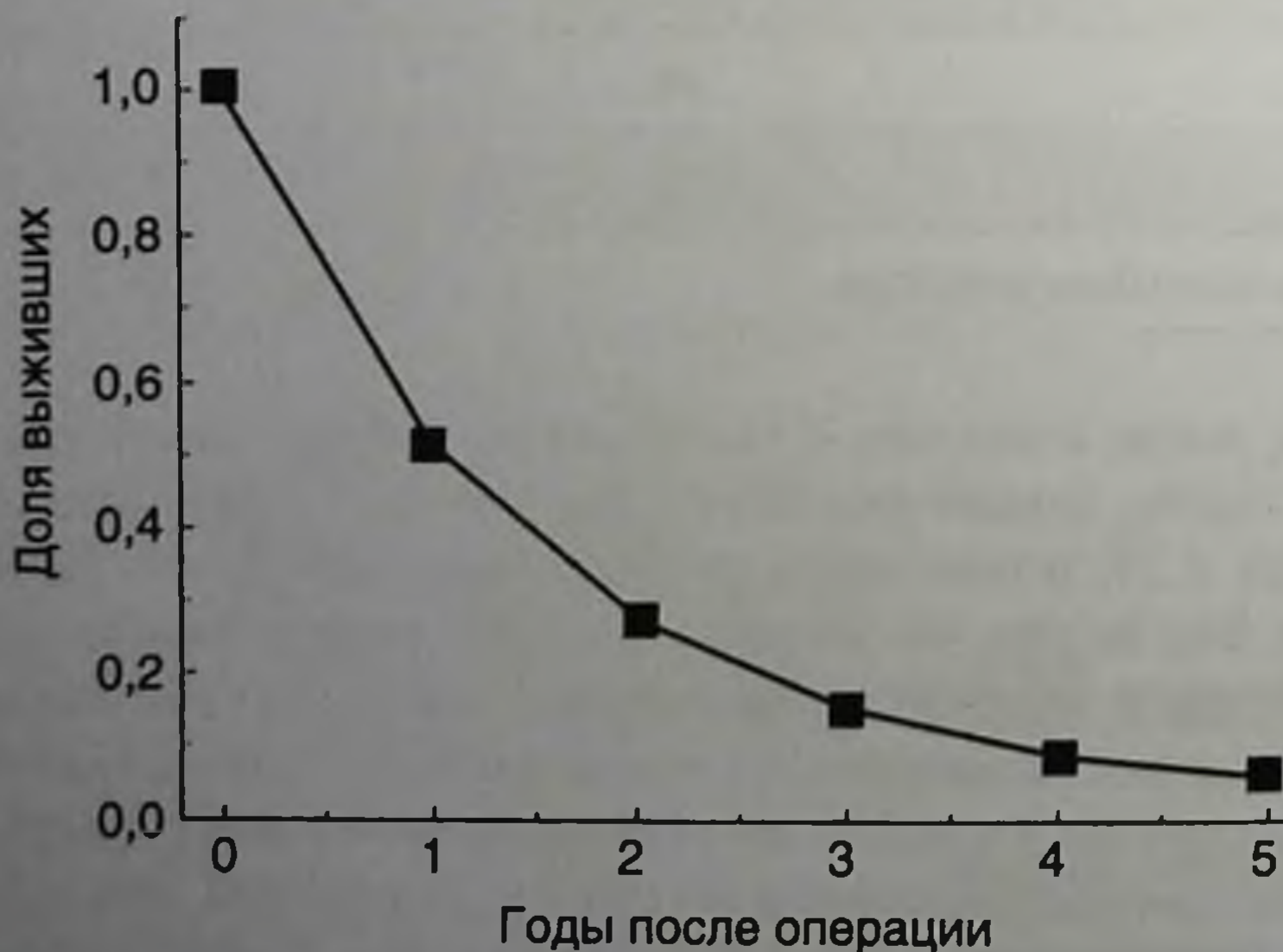


Рис. 2.46. Кривая выживания для примера табл. 2.17



Результаты анализа выживания должны быть представлены в таблице или графике с указанием численности пациентов в каждый период анализа. Естественно, чем больше срок, тем меньше пациентов оказываются отслеженными до этого срока. Соответственно, точность оценки доли доживших (не имевших осложнений) становится малой. Доверительный интервал для каждого периода наблюдения определяется числом пациентов, отслеженных в этом периоде, и вероятностью выживания, т.е. рассчитывается примерно так же, как для относительных величин (см. разд. 1.4.4).

### Метод Каплана-Майера

В актуарном анализе выживания расчеты делались применительно к интервалам наблюдения, равным, например, году или месяцу. Обычно это совпадает с планами долговременного наблюдения, но в клинике регистрируют реальные сроки смерти пациентов. Целям такого анализа, не привязанного к периодичности оценки пациентов, соответствует метод Каплана-Майера (Kaplan-Meier Method, называется также множительным методом — англ. *product-moment method*). Его принцип так же прост, как и изящен. Летальность рассчитывается на момент смерти каждого пациента как пропорция, в числителе которой число умерших на этот момент пациентов, а в знаменателе — число всех доживших до этого момента (в том числе и умерших в этот момент), за вычетом тех, кто до этого момента не отслежен. Под моментом понимается единица времени, используемая в исследовании. Этот подход рассмотрим на примере рис. 2.47, данные и расчет приведены в табл. 2.20 и 2.21.

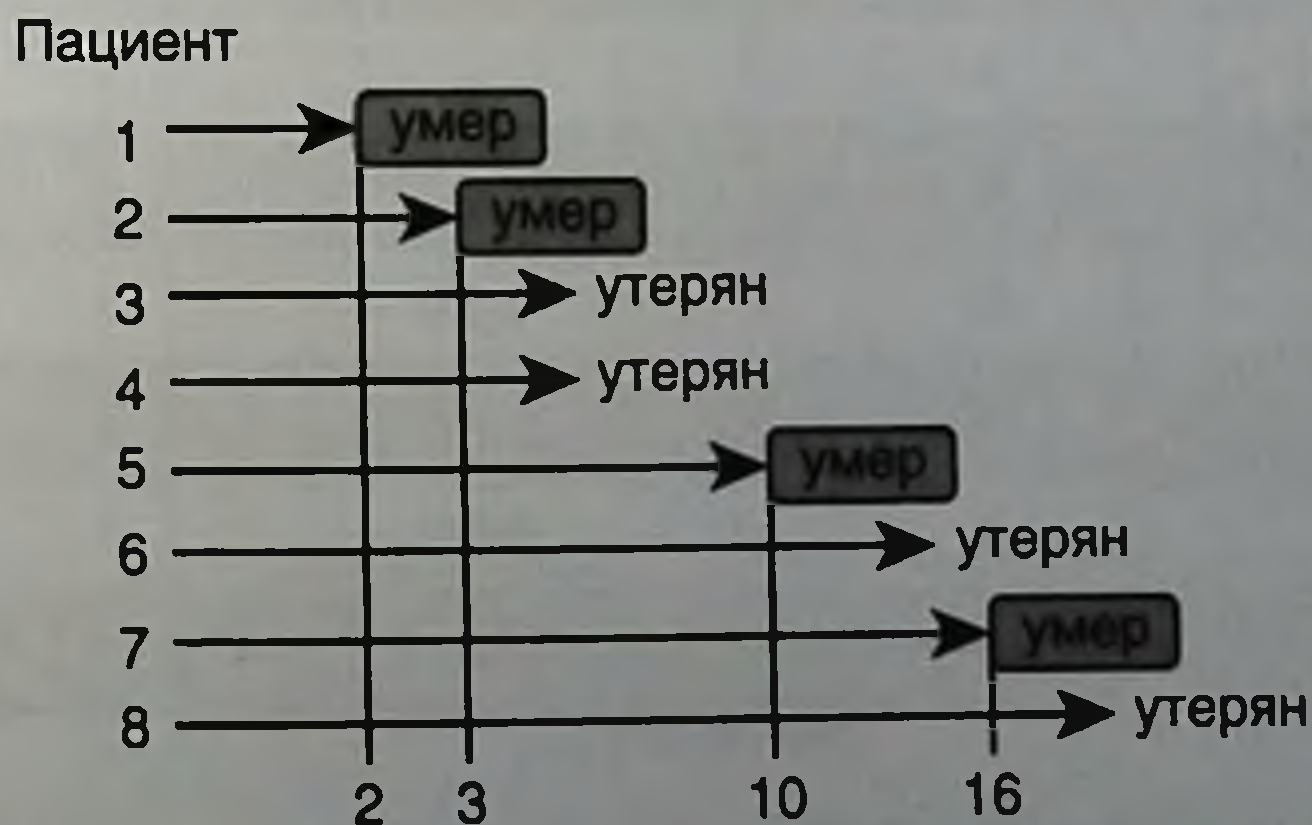


Рис. 2.47. Гипотетический пример наблюдения за 8 пациентами



Таблица 2.20. Расчет выживания методом Каплана–Майера по данным рис. 2.47

Время от начала наблюдения, мес	Число живых пациентов к моменту расчета	Число умерших в момент расчета	Пропорция умерших в этот момент	Пропорция выживших на этот момент	Кумулятивная пропорция доживших до этого момента
1	2	3	4	5	6
–	–	–	Кол. (3)/Кол. (2)	1 – Кол. (4)	–
2	8	1	0,12	0,88	0,88
3	7	1	0,14	0,86	0,76
10	4	1	0,25	0,75	0,56
16	2	1	0,50	0,50	0,28

Таблица 2.21. Расчет выживания методом Каплана–Майера по данным рис. 2.47. Расчет с помощью пакета прикладных программ Statistica. Plusом отмечены номера пациентов, выпавших из наблюдения

Пациенты	Время, мес	Кумулятивная вероятность выживания	Стандартная ошибка
1	2	0,875	0,117
2	3	0,750	0,153
3+	4	–	–
4+	4	–	–
5	10	0,562	0,199
6+	15	–	–
7	16	0,281	0,222
8+	18	–	–

В третьей строке табл. 2.20 приведены подсказки к вычислениям. Кумулятивная пропорция (колонка 6) получается умножением пропор-



ции в строке сверху на пропорцию из колонки 5, т.е. для второй строки (месяц 3-й) это  $0,88 \times 0,86 = 0,76$ .

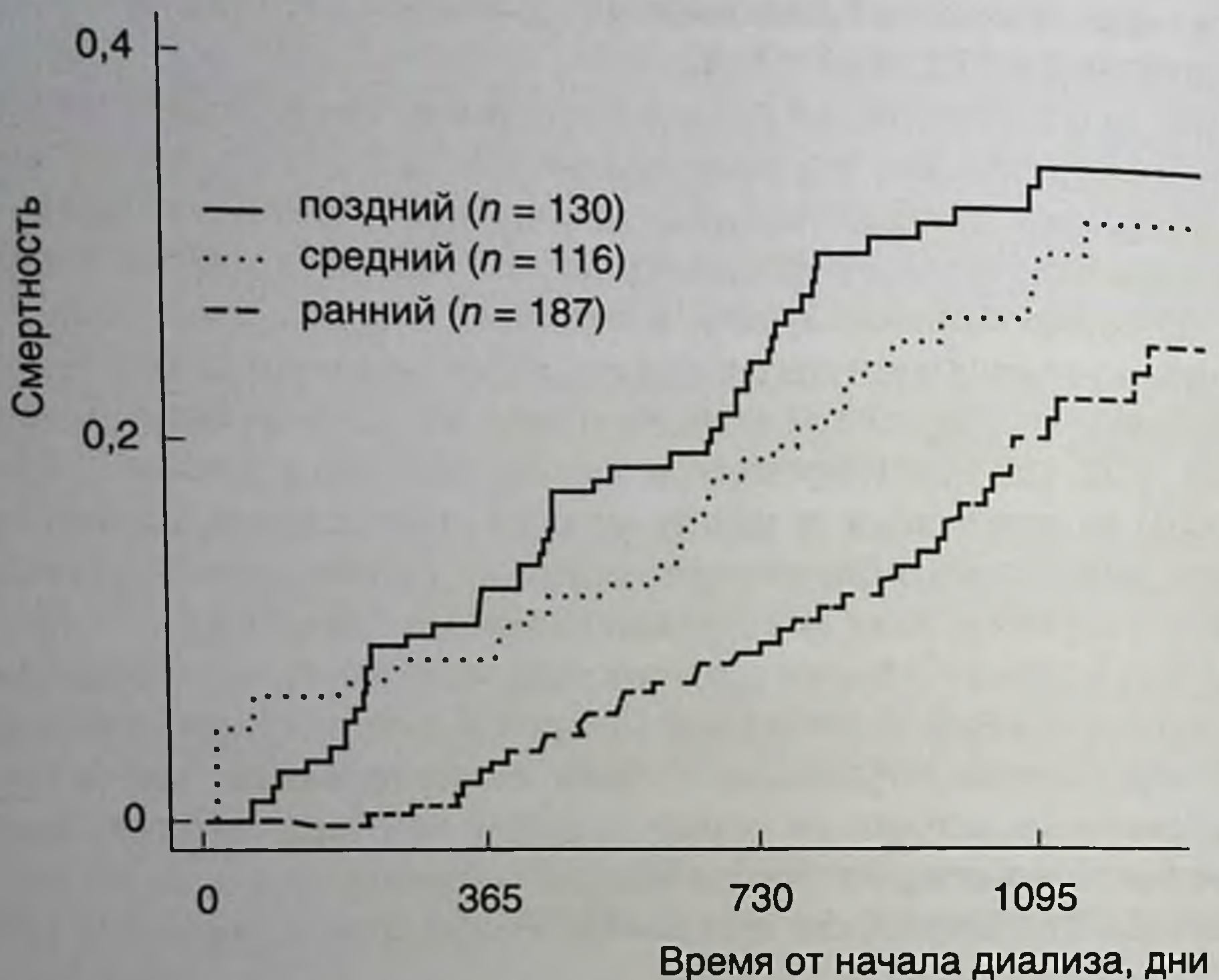
При подготовке данных для компьютерного анализа файл составляют таким образом, как это описано в разд. 1.5.6. Всем случаям присваивается признак времени исхода — например, смерти или выпадения из исследования (цензурирования). Дополнительно вводится переменная «цензурированный случай», в которой случаям с известным исходом присваивается один код (например, 0), а цензурированным — другой (например, 1). Результаты расчета в программе Statistica приведены в табл. 2.21, где цензурированные случаи помечены плюсом. Эти результаты не отличаются от наших «ручных» вычислений, но они предлагают нам также величину стандартной ошибки, расчет которой вручную, впрочем, тоже не составляет труда (см. разд. 1.4.4).

Метод Каплана–Майера в полной мере использует всю информацию о выживании каждого пациента. Его результаты в графической форме — характерные «ступеньки» — более иллюстративны, чем результаты применения актуарного метода, и лучше передают точность оценки вероятности выживания. Метод Каплана–Майера включен во все современные статистические программы. Результаты применения метода



Рис. 2.48. Кривая выживания, рассчитанная методом Каплана–Майера по данным рис. 2.47 с помощью пакета программ Statistica





**Рис. 2.49.** Кумулятивная вероятность смерти у больных сахарным диабетом с хронической почечной недостаточностью в зависимости от срока доступа к консультации нефролога (по данным Kinchen K.S. et al. *Ann. Intern. Med.* 2002. № 137. P. 479–86)

Каплана–Майера к реальным данным могут быть приведены в разном виде. На рис. 2.48 они даны в виде кумулятивной кривой выживания. Если число отслеженных людей и число наблюдаемых случаев велики, то ступеньки на кривой выживания становятся маленькими, что отражает точность оценки кривой (рис. 2.49).

### Другие возможности и ограничения анализа выживания

Компьютерные средства позволяют оценить статистическую значимость различий между сравниваемыми группами пациентов, например группами в контролируемом испытании лекарства. Для оценки статистической значимости используется вариант критерия Вилкоксона и логранговый критерий Мантеля–Хэнзеля (Mantel–Haenszel). Эти критерии оценивают не статистическую значимость различий между кривыми выживания в отдельный момент, а во всем изученном временном интервале.



Важным достоинством методов оценки выживания является возможность устранить вычислительными средствами влияние на летальность отдельных факторов, например различий возраста к моменту заболевания или дохода пациента. Эту возможность предоставляет метод Кокса, называемый *методом пропорциональных рисков* (*Cox proportional hazards model*). Функция риска, которую иногда называют «сила смертности» (см. разд. 2.6.1), представляет собою мгновенную вероятность смерти для живых в данный момент. Модель Кокса предполагает, что отношение рисков для двух сравниваемых субъектов остается постоянным во времени или пропорциональным. Относительный риск, вычисленный на основании оценок выживания, называют отношением рисков (англ. *hazard ratio*).

Методы оценки выживания исходят из предположения, что в течение периода, когда собирают данные о выживании пациентов, результаты изучаемого лечения неизменны. Если период сбора данных длителен, то это предположение может быть несправедливо. Например, с течением времени исходы операции могут по разным причинам улучшаться. Поэтому суммировать случаи, наблюдавшиеся в раннем периоде, с поздними неправильно. Целесообразно рассчитать выживание для ранних и поздних случаев отдельно и в случае наличия различий анализировать их отдельно. В другом случае операция, которая в первые годы выполнялась только тяжелым больным как экспериментальное вмешательство, с течением времени может распространяться на больных с более легкими формами (ранними стадиями) болезни. В результате даже при отсутствии реального улучшения исходов лечения может создаваться иллюзия прогресса. Такую систематическую ошибку (смещение) иногда можно выявить, если сравнить ранние и поздние случаи по какому-нибудь признаку, например по объему опухоли. Если окажется, что пациенты раннего периода отличаются от пациентов позднего периода исследования, то их следует анализировать отдельно.

Анализ выживания предполагает, что пациенты, выпадающие из наблюдения, не отличаются от остающихся под наблюдением и испытывающих изучаемые исходы (заболевающих, умирающих). Это называют «предположение о неинформативности цензурирования». Если это не так, то оценки анализа выживания могут быть смещенными. Это может происходить, если цензурирование связано с исходами. Например, при лечении алкогольной, наркотической или иной зави-



симости из наблюдения могут чаще выпадать лица, у которых вновь возникают проблемы с употреблением алкоголя. Если цензурировано много случаев, то оценка продолжительности ремиссии окажется завышенной.

### 2.6.3. Влияние болезней на смертность

Главное, очевидное и наиболее важное влияние болезней на человека — увеличение смертности. В человеческой культуре болезни, сопровождающиеся высокой смертностью, воспринимаются как опасные, тяжелые, страшные. Обычно меньше внимания привлекают заболевания пожилых людей, которые выглядят как «естественное угасание», а не как драматическая и «несвоевременная» смерть. К таким состояниям относится сердечная недостаточность. В развитых странах таких больных очень много — 1–2% населения. Их возраст в момент установления диагноза обычно старше 70 лет. Причиной состояния чаще всего являются ишемическая болезнь сердца, гипертоническая болезнь и менее 10% — клапанные пороки. В развивающихся странах

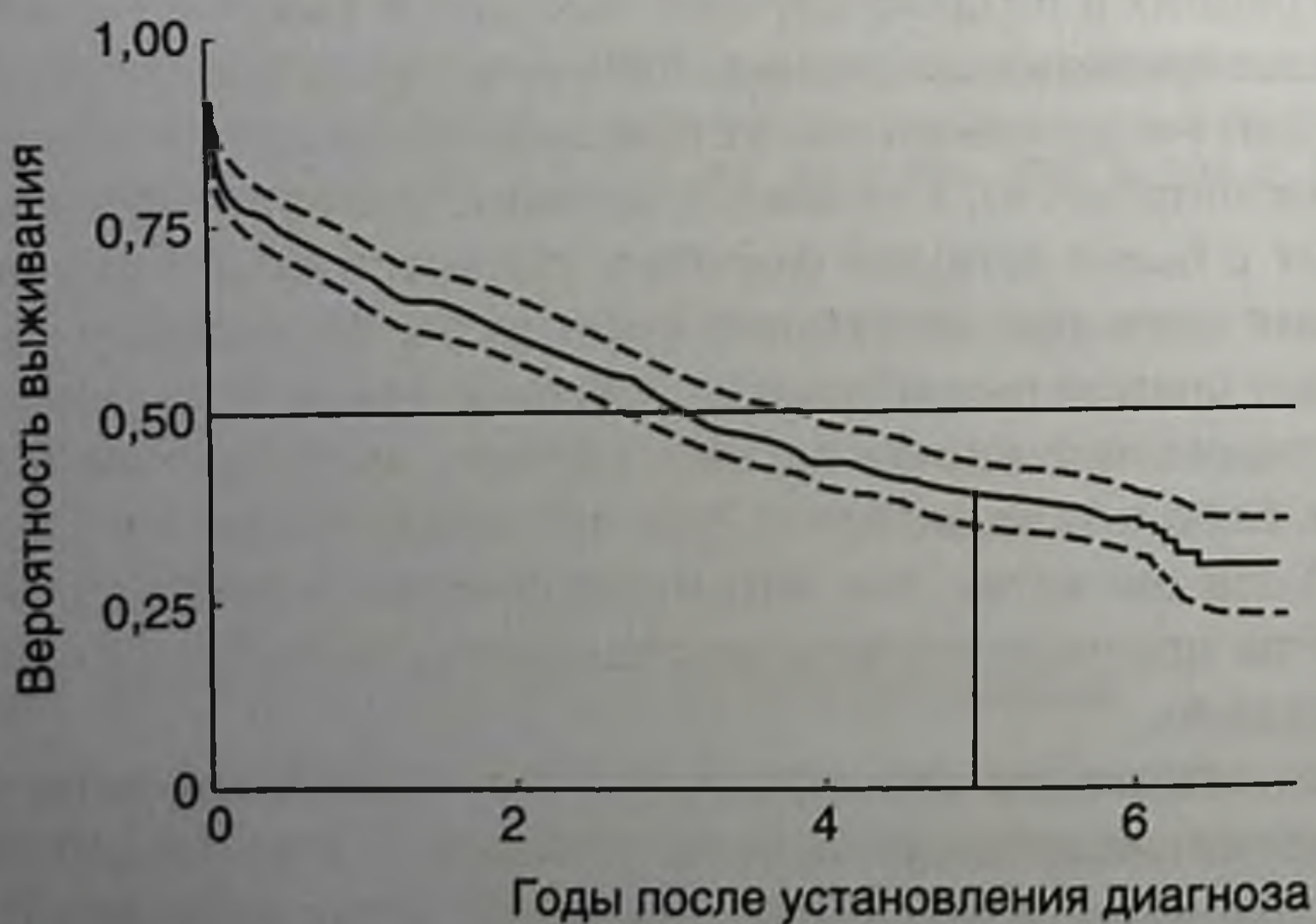


Рис. 2.50. Выживание больных с сердечной недостаточностью. По данным о 552 больных, отслеженных в лондонских исследованиях 1995–1998 гг. Кумулятивная оценка по МКМ (линия) и 95% доверительный интервал (пунктир) (по Cowie M. и Zaphiriou A. *BMJ*. 2002. № 325. P. 422–5, с разрешения издателя)



значительная доля случаев — следствие приобретенных пороков сердца. При сердечной недостаточности очень высока смертность, особенно — в первый год после развития состояния. Несмотря на интенсивные исследования последних десятилетий прогресс в лечении сердечной недостаточности невелик (рис. 2.50).

Традиционным способом выражения смертности у больных было вычисление доли выживающих в течение 5 лет. Эта цифра условна: просто это срок, дольше которого обычно не удастся наблюдать достаточно большую часть пациентов, а меньший срок представляется малым. С помощью анализа выживания можно устанавливать с большей точностью медиану выживания и 5-летнюю выживаемость (соответственно вертикальная и горизонтальная линии на рис. 2.50 пересекают кривую выживания на этих уровнях). Однако главное достоинство анализа выживания в том, что он позволяет уйти от таких «точечных» оценок и представить, как развивается смертность пациентов во всем изученном интервале времени.

## Контрольные вопросы и задачи к разделу 2.6

1. В возрасте 5–14 лет смертность от рака составляет 15% от общей, а в 35–64 года — 30% от общей. Каков относительный риск смерти от рака в 35–64 года в сравнении с возрастом 5–14 лет?
  - а) 0,5;
  - б) 2;
  - в) не может быть вычислен;
  - г) 0,66.
2. Данные о смертности в течение первых 20 лет жизни были сопоставлены с данными о курении матерей, записанными при рождении. Это исследование принадлежит к типу:
  - а) сравнения с контролем;
  - б) проспективное;
  - в) когортное с ретроспективным сбором данных;
  - г) с вмешательством.
3. Скорость роста популяции зависит от:
  - а) рождаемости;
  - б) смертности;
  - в) экономического благополучия;



- г) численности иммиграции;
  - д) численности эмиграции.
4. Выберите, какой показатель из списка характеризует приведенные ниже утверждения:
- а) частота мертворождений;
  - б) перинатальная смертность;
  - в) неонатальная смертность;
  - г) материнская смертность;
  - д) относительная частота мертворождений.
- 4.1. На 1000 рожденных живыми рождается 16 мертвых детей.
- 4.2. На 100 000 рожденных живыми умирает 20 женщин в родах.
- 4.3. В возрасте до 28 дней умирает 14 детей на 1000 рожденных живыми.
- 4.4. Среди плодов старше 28 нед и младенцев до 7 дней умирает 28 на 1000 рожденных живыми.

## 2.7. КОГОРТНЫЕ ИССЛЕДОВАНИЯ

Название когортного исследования происходит от термина «когорта», обозначающего одно из десяти подразделений древнеримского легиона. В эпидемиологии так называют группу людей, обладающих общей характеристикой, например одним годом рождения. Когортный анализ был разработан и применен в медицинских целях задолго до появления возможностей проспективных исследований. Когортный анализ основывался на использовании ретроспективных данных, доступных в национальных системах регистрации рождений и смертей, причин смерти. Одним из первых применил когортный анализ в медицинских целях Е. Галлей (E. Halley, известен трудами в математике и астрономии, комета названа его именем). Галлей применил для изучения продолжительности жизни математический метод: в 1692 г. он показал, что не существует предсказываемых астрологами критических (опасных) периодов в жизни людей. Англичанин Уильям Фар (William Farr), называемый иногда отцом количественной эпидемиологии, в середине XIX в. разработал метод когортных таблиц продолжительности жизни. Он использовал его для оценки продолжительности жизни психически больных людей и показал, что при проведении гуманного лечения продолжительность жизни больных выше, чем при «обычном» тогда лечении.



### 2.7.1. Принцип когортного анализа

*Когортой* называют группу людей, выделенную по некоторому признаку в начальный момент исследования, часто — по году рождения, но не обязательно. Когорты могут формироваться по году окончания университета или по году вступления в брак. *Когортными*, или *поколенческими*, *эффектами* называют эффекты от воздействий среды, присущие данному поколению (когорте), в том числе включающие в себя влияние на наследственность. Когортные эффекты изменяют каждого члена когорты и придают событиям, изучаемым в когорте, регулярные (в последовательности когорт) или исключительные (нерегулярные) отличия от других когорт. Так, в последовательных когортах европейцев, рожденных в XX в., наблюдается увеличение длины тела. Наоборот, война накладывает исключительный отпечаток на смертность лиц, принадлежащих к когорте, родившейся в эти годы, впрочем, как и на их образование, брачность, социальную адаптацию.

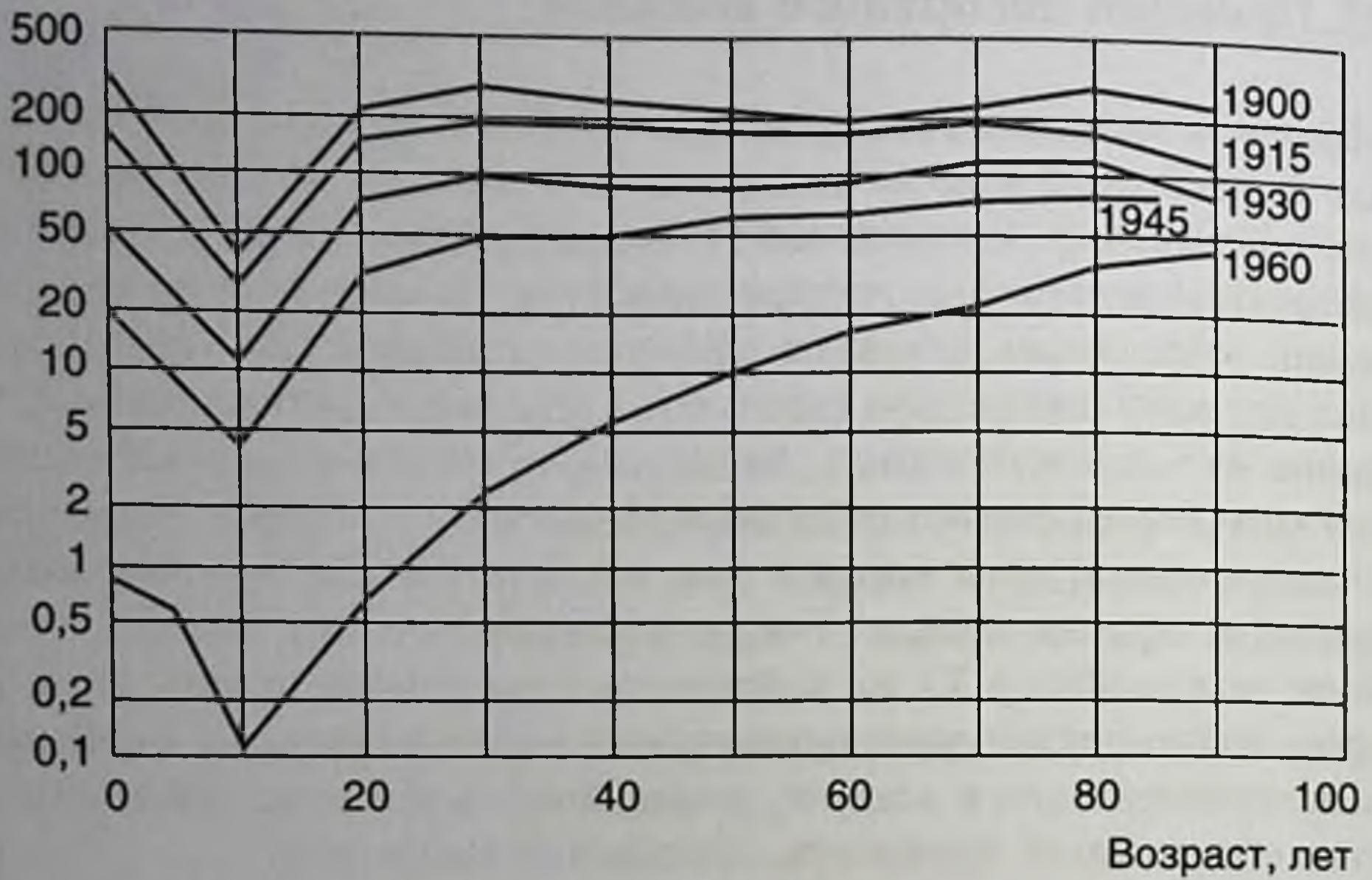
*Возрастные изменения* наблюдаются дополнительно к когортным и потому в каждой когорте можно видеть возрастные различия. Возрастные изменения определяются как внутренне присущие организмам и отражающиеся в развитии, созревании, старении. Возраст к моменту наступления события определяется интервалом между рождением и событием.

*Временные (периодические) эффекты* — это эффекты от воздействий среды, имевших место в отдельный период времени. В этот период фактор действует на людей всех когорт, но они находятся в разном состоянии (разный возраст), что обуславливает их отличия в реакции на изменение среды.

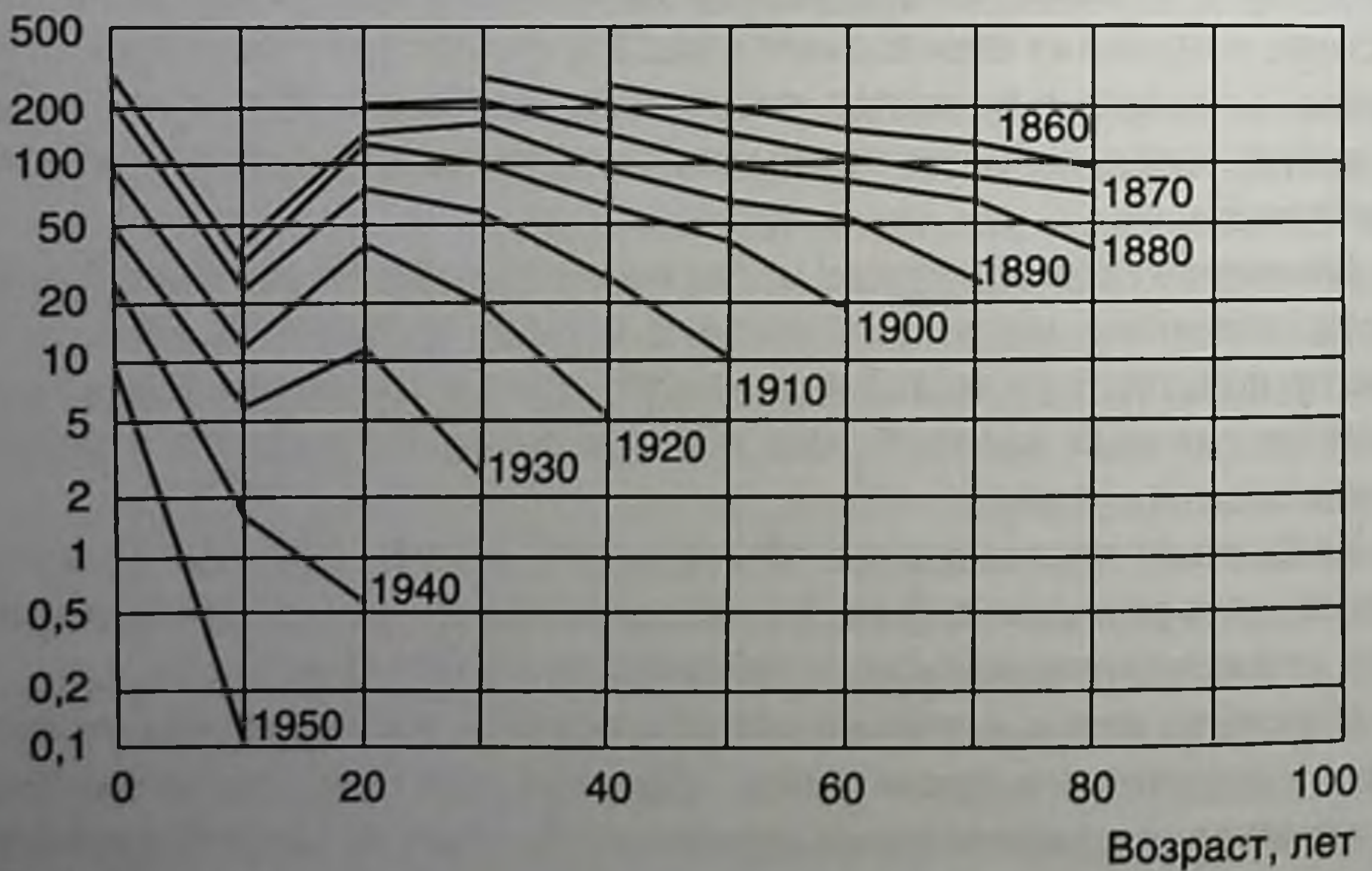
Когортные исследования обычно дают ясную картину, позволяющую понять развитие событий в отдельных поколениях. Пример — картина снижения смертности от туберкулеза в США (рис. 2.51).

Ограничением когортного анализа является взаимная связь возрастного, когортного и временного эффектов через способ вычисления. Поэтому если изменяется одновременно частота исходов (смертность) и возрастное распределение исходов, то для полноценной интерпретации изменений необходимо привлекать данные других исследований. Это ограничение когортного метода не является уникальным. Все методы исследований дают результаты, которые должны быть сопоставлены с результатами исследований других типов.





а



б

Рис. 2.51. Смертность от туберкулеза в разном возрасте (на 100 000 населения) в США в 1900–1960 гг: а — смертность в различные изученные периоды; б — в когортах 1860–1950 гг. рождения (по Dоеge Т.С. JAMA. 1963. № 192. P. 1045)



Приведенное исследование туберкулеза, в частности, дало основания полагать, что относительно невинное инфицирование туберкулезом в молодости отражается в повышенной смертности в старшем возрасте (см. рис. 2.51, смертность в 1960 г. у людей старшего возраста). Обратим внимание на медленное снижение смертности до 1945 г. и быстрое — после этого, что, видимо, связано с появлением стрептомицина.

## 2.7.2. Проспективное когортное исследование

Для того чтобы описать течение болезни, изучить вероятные ее исходы, научиться определять зависимость исходов от особенностей симптоматики, чтобы установить связь возникновения болезни с экспозицией предполагаемому причинному фактору, нужно наблюдать за развитием болезни. Если болезнь носит острый, скоротечный характер, то наблюдать ее просто. Именно поэтому в прогнозе течения острых болезней значительный прогресс был достигнут еще в XIX в. Хронические болезни развиваются медленнее и простому наблюдению почти недоступны. Многие из них просто не могут уложиться в достаточном количестве случаев в период активной работы врача. Поэтому сведения о течении хронических болезней стали накапливаться только во второй половине XX в., по мере проведения проспективных эпидемиологических исследований. С этими исследованиями связано формирование концепции «факторов риска», т.е. признаков здоровых людей, которые связаны с повышенной вероятностью заболевания. Эти исследования дали толчок не только к пониманию течения болезней, но и к разработке методов профилактики, прогнозирования развития болезней, их лечения.

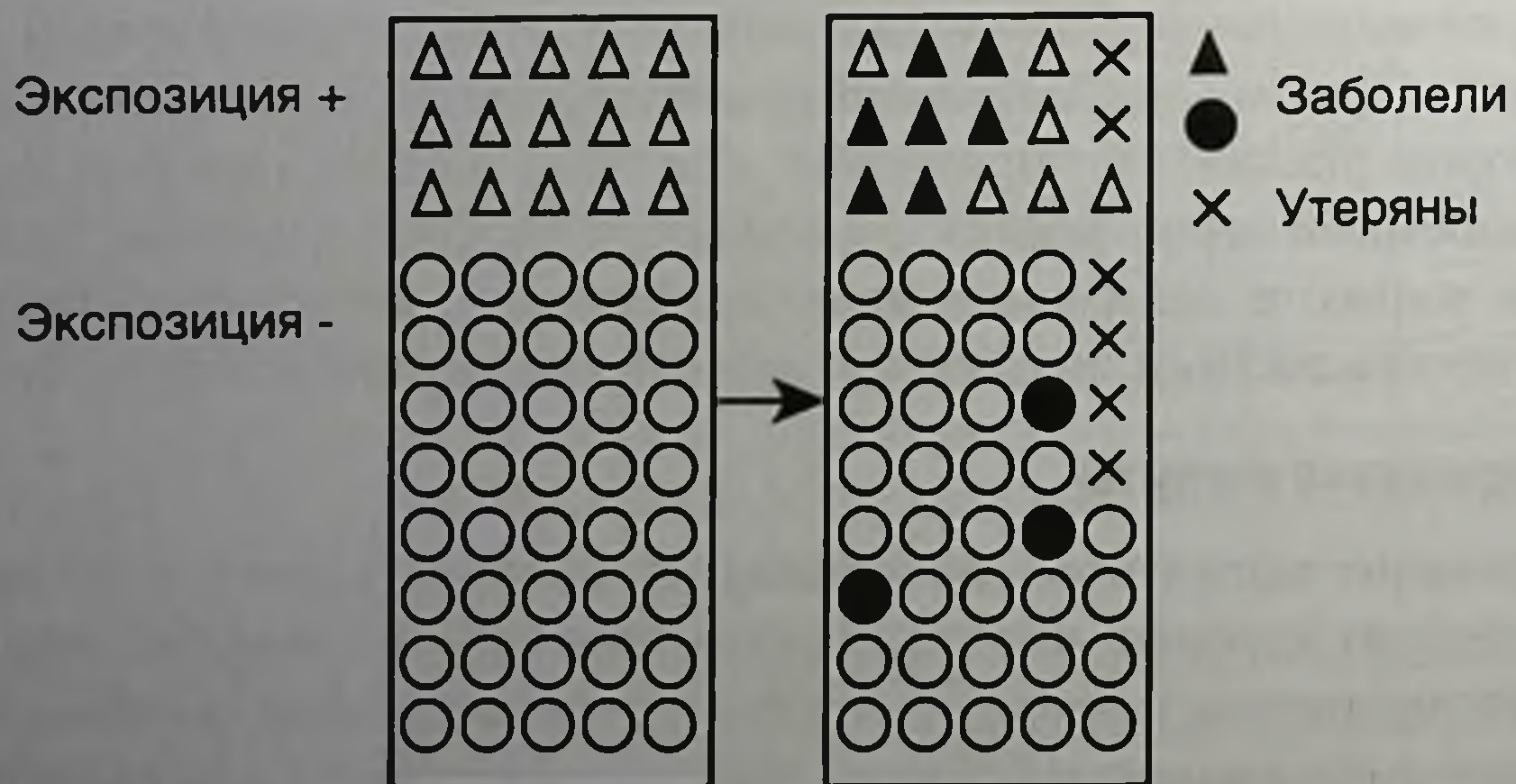
### Формирование когорты

Принцип когортного исследования состоит в изучении исходов в подгруппах когорты, выделенных по признаку экспозиции (рис. 2.52).

В медицинском и эпидемиологическом исследованиях важно, чтобы *исходная группа пациентов* (англ. *inception cohort*) была исчерпывающим образом определена. Например, в нее должны включать пациентов в детально описанной и одинаковой фазе болезни. Если не учтена фаза болезни или место проведения исследования, то нельзя будет использовать результаты исследования. Так, в исследовании исходов острого инфаркта миокарда (ОИМ) в качестве исходной группы пациентов могут быть взяты лица, находящиеся в палате интенсивной терапии. Судить



об исходах ОИМ вообще по этой группе будет неверно. В действительности летальность при ОИМ значительно выше — часть больных погибает, так и не поступив в палату интенсивной терапии. С другой стороны, во многих странах пациентов с неосложненным течением ОИМ не помещают в отделение интенсивной терапии. Исследование, ограниченное только пациентами отделения интенсивной терапии, корректно, но его результаты имеют ограниченное приложение. В случае формирования когорты больных должны быть использованы правильные критерии принятия решения о диагнозе (соответствии пациента критериям включения). Оценка стадии, тяжести и прочих особенностей течения изучаемых состояний должна быть точно описана. Желательно использование проверенных методов. Только тогда врач, читающий статью с отчетом об исследовании, сможет доверять отчету и использовать его результаты. Иногда в группу включают не всех пациентов, соответствующих критериям включения, например если таких находится на лечении с изучаемым состоянием слишком много. В таком случае в исходную группу должны включаться отобранные случайным образом пациенты, чтобы группа была репрезентативной (см. разд. 1.4.1).



**Рис. 2.52.** Структура когортного исследования при изучении причины болезни. Подбираются группы людей, подвергающихся экспозиции и не подвергающихся ей. С течением времени частота исходов измеряется в каждой группе. В группе, подвергавшейся экспозиции, инцидент = 7/13; а в группе без экспозиции инцидент = 3/36. Можно учесть утерянных участников исследования и применить соответствующий способ анализа (см. разд. 2.6.2)



Формирование исследуемой группы состоит в применении критериев включения и критериев исключения. Первые служат для очерчивания круга пациентов, которые в целом подходят для исследования, а вторые — для исключения тех, кто не сможет безопасно и достаточно длительно находиться в исследовании или, например, из-за сопутствующих болезней могут быть нетипичными для выборки. Подробнее критерии включения и исключения обсуждаются в разд. 3.3.1.

При формировании когорты возможны следующие *смещения подбора* (англ. *sampling bias*).

**А. Центропрямительное смещение** (англ. *centripetal bias*). При оказании квалифицированной помощи, в особенности при хронических и тяжелых состояниях, существует практика концентрации пациентов в специализированных центрах. В них, с одной стороны, собираются самые сложные больные, а с другой — там максимальные возможности диагностики и лечения. Поэтому полученные в таком центре результаты не вполне соответствуют общей практике.

**Б. Смещение популярности** (англ. *popularity bias*). В специализированные центры из-за их популярности и по другим причинам поступает ряд непрофильных пациентов. Им уделяют меньше внимания, в их лечении у специалистов центра меньше опыта, применяются нетипичные средства лечения.

**В. Смещение фильтрации** (англ. *referral filter bias*). По мере продвижения пациентов от звена первичной медицинской помощи к региональным больницам, большим центрам, имеющим специализированные отделения, и к национальным центрам, специализирующимся на отдельных болезнях, происходит «фильтрация» — отбор пациентов с рядом особенностей. К ним относятся не только особенности течения заболевания, но и социальные (в том числе семейные), географические, финансовые особенности и особенности ранее примененного лечения. Поэтому пациенты более «высоких» уровней оказания медицинской помощи отличаются от таких же больных, лечащихся на «низких» уровнях.

**Г. Смещение доступности диагностики** (англ. *diagnostic access bias*) — вариант смещения фильтрации. Люди, проживающие в глубинке, и малообеспеченные имеют не только меньше доступ к медицинской помощи вообще, но и при выявлении болезни имеют меньше шансов на доступ к специальным диагностическим и лечебным технологиям. Вследствие низкой доступности эффективных средств диагностики



и соответствующих особенностей лечения у таких пациентов возможно худшее течение заболеваний, более редкое их выявление.

В любом случае врачи первичного контакта (в том числе семейные врачи) наблюдают пациентов, которые имеют меньшую тяжесть болезни, и у них чаще возникает самовыздоровление (спонтанная ремиссия). При меньшей тяжести болезни они менее склонны соглашаться на травмирующее лечение и хуже выполняют предписания врачей. С другой стороны, у более легких больных труднее получить заметный эффект лечения. Соответственно, расчеты эффективности вмешательства, справедливые для специализированного центра или отделения неотложной помощи, могут быть неподходящими для общей практики. Одновременно пациенты врача общей практики имеют более низкую вероятность болезни. Это означает, что при обращении к врачу они нередко не имеют определенной болезни (их страдания носят немедицинский характер). Применять по отношению к ним то же агрессивное лечение, которое хорошо себя зарекомендовало в испытаниях, проведенных в больничных условиях, может быть ошибкой. Кроме того, пациенты врача общей практики имеют нередко размытые, несформированные, вероятно, множественные страдания. В дальнейшем в процессе развития болезни их страдание «кристаллизуется», но в приемной у врача общей практики неясных пациентов очень много. К таким пациентам следует применять те сведения, которые соответствуют этому уровню медицинской помощи. Сказанное справедливо и для других уровней оказания помощи.

### **Анализ результатов исследования**

**Полнота отслеживания** является принципиально важной для результата когортного исследования. Результат можно считать вполне соответствующим выбранной в начале группе, только если вся группа отслежена в течение необходимого срока. Достичь такого в исследовании обычно невозможно, но потеря по ходу исследования более 10% пациентов, вероятно, приведет к заметным неопределенностям в оценках, а потеря более 20% сделает оценки малопригодными для использования. Потерянные лица могут отличаться от отслеженных, поэтому связь исхода с наблюдаемой экспозицией (лечением) или исходными особенностями может быть усилена или ослаблена.

**Объективность критериев исхода** столь же важна как критерии включения в исследование. Лишь факт смертельного исхода, если он доку-



ментирован, является вполне надежным. Все остальные исходы требуют формулировки точных критериев. В том числе любая избранная причина смерти. Одно дело, если она устанавливается в клинике, другое — если по весьма приблизительному свидетельству о смерти.

Для того чтобы сделать оценку исходов более точной, надо применять **оценку вслепую**. Это необходимо для того, чтобы оценка не зависела от других особенностей пациента и его лечения. Врач может более внимательно («подозрительно») относиться к оценке состояния пациента, если ему известны изучаемые факторы риска болезни в анамнезе пациента. В результате возникает *смещение подозрения* (смещение оценки) (англ. *suspicion bias*). Врач-лаборант, патологоанатом, зная клиническую картину пациента, его анамнез, могут находиться под сильным влиянием знаний о прошлых событиях и прижизненном диагнозе. В результате возникает *смещение ожидания* (англ. *expectation bias*).

### Исследование качественных признаков

При описании результатов когортного исследования, связывающего заболевание с экспозицией (фактором риска), вычисляют относительные показатели, представленные в табл. 2.22.

Таблица 2.22. Вычисление показателей связи болезни и экспозиции

Экспозиция (признак)	Болезнь		Всего
	Есть	Нет	
Есть	A	B	A + B
Нет	C	D	C + D

Риск при наличии признака =  $A/(A + B)$ .

Риск при отсутствии признака =  $C/(C + D)$ .

Абсолютная разность рисков (АРР) (англ. *absolute risk difference*, ARD)

$АРР = A/(A + B) - C/(C + D)$ .

Относительный риск (англ. *Relative Risk*, RR)  $ОР = [A/(A + B)]/[C/(C + D)]$ .

Различие риска (снижение относительного риска) (RRR, англ. *RR reduction*)  $СОР = 1 - RR$ .

Шансы на развитие болезни при наличии фактора риска =  $A/B$ .



Шансы на развитие болезни при отсутствии фактора риска =  $C/D$ .

Отношение шансов (ОШ, отношение шансов при наличии фактора риска к шансам при его отсутствии) (англ. *odds ratio*, OR)  $OШ = (A/B)/(C/D)$ .

Статистическую значимость повышения риска в одной из групп по сравнению с другой оценивают с помощью доверительного интервала (ДИ). ДИ для ОР и других относительных показателей рассчитывается так же, как для прочих относительных величин (см. разд .4). Если ДИ не включает единицу, то ОР статистически значимо отличается от единицы. Иными словами — сравниваемые группы различаются.

### Исследование количественных признаков

В случае когда фактор риска — количественный признак, результат выглядит иначе, и оценка риска проводится по-другому. При сравнении исходных данных двух подгрупп изученной когорты выявляется количественное различие этих групп (рис. 2.53). На этом классическом примере видно, что заболевшие имели более высокую концентрацию холестерина

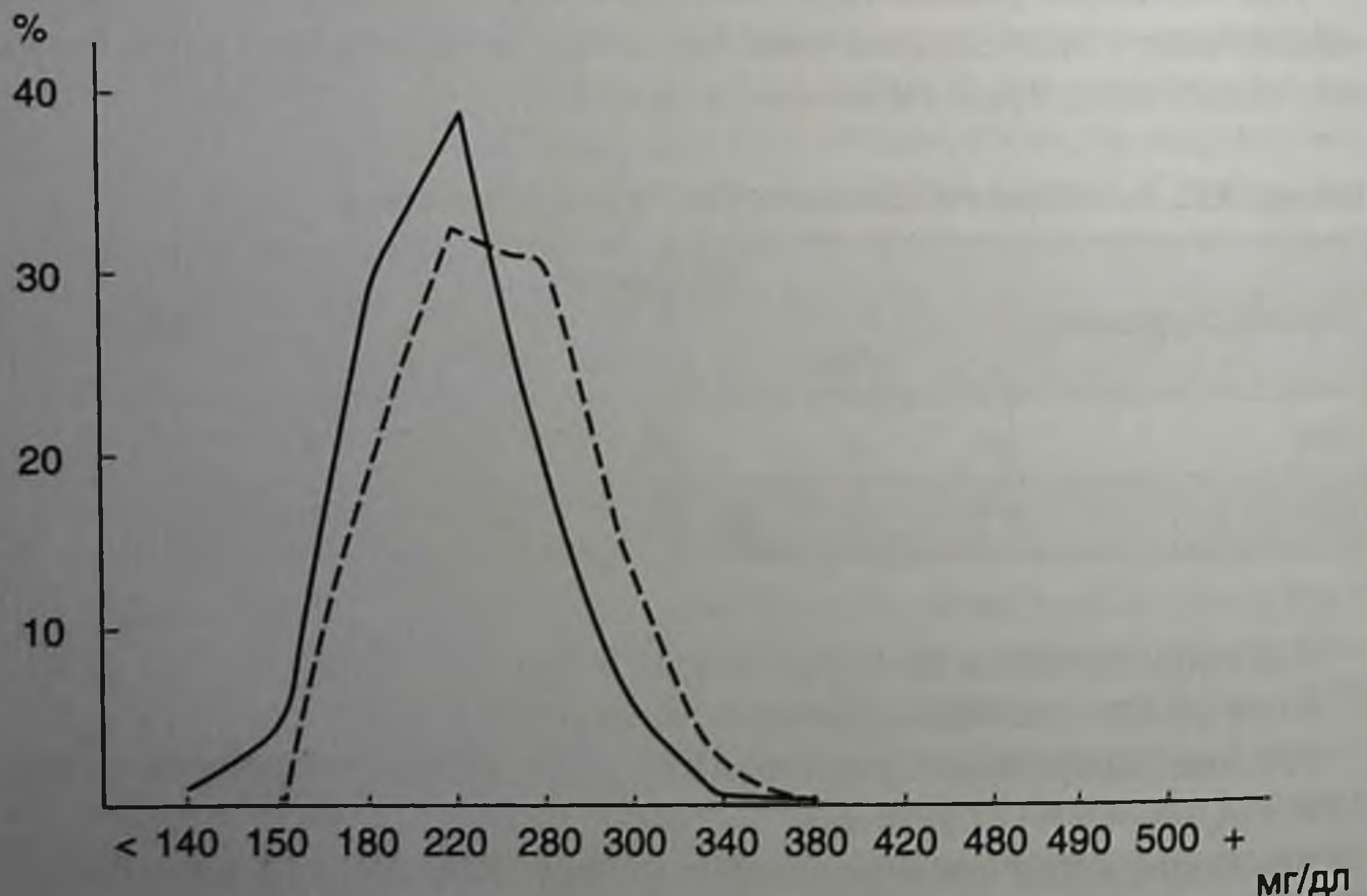


Рис. 2.53. Распределение по концентрации холестерина в плазме крови (абсцисса) в начале исследования мужчин 50–62 лет. Материалы Фремингемского исследования. Линия — лица, не заболевшие коронарной болезнью сердца, пунктир — заболевшие в течение 8 лет (по Kannel W.B. Ann. Intern. Med. 1979. Vol. 90: 85–91. P. 1985)



в плазме крови до заболевания, но эти различия нельзя назвать большими в сравнении с нормальной межиндивидуальной вариабельностью.

Если выделить разные интервалы концентрации, а затем посчитать вероятность развития болезни у людей, принадлежащих к этим интервалам, то выявляется, что чем далее к высоким концентрациям, тем больше вероятность болезни и тем чаще смерть от ишемической болезни сердца (рис. 1.5 и 2.54). Конечно, лишь небольшая доля популяции принадлежит

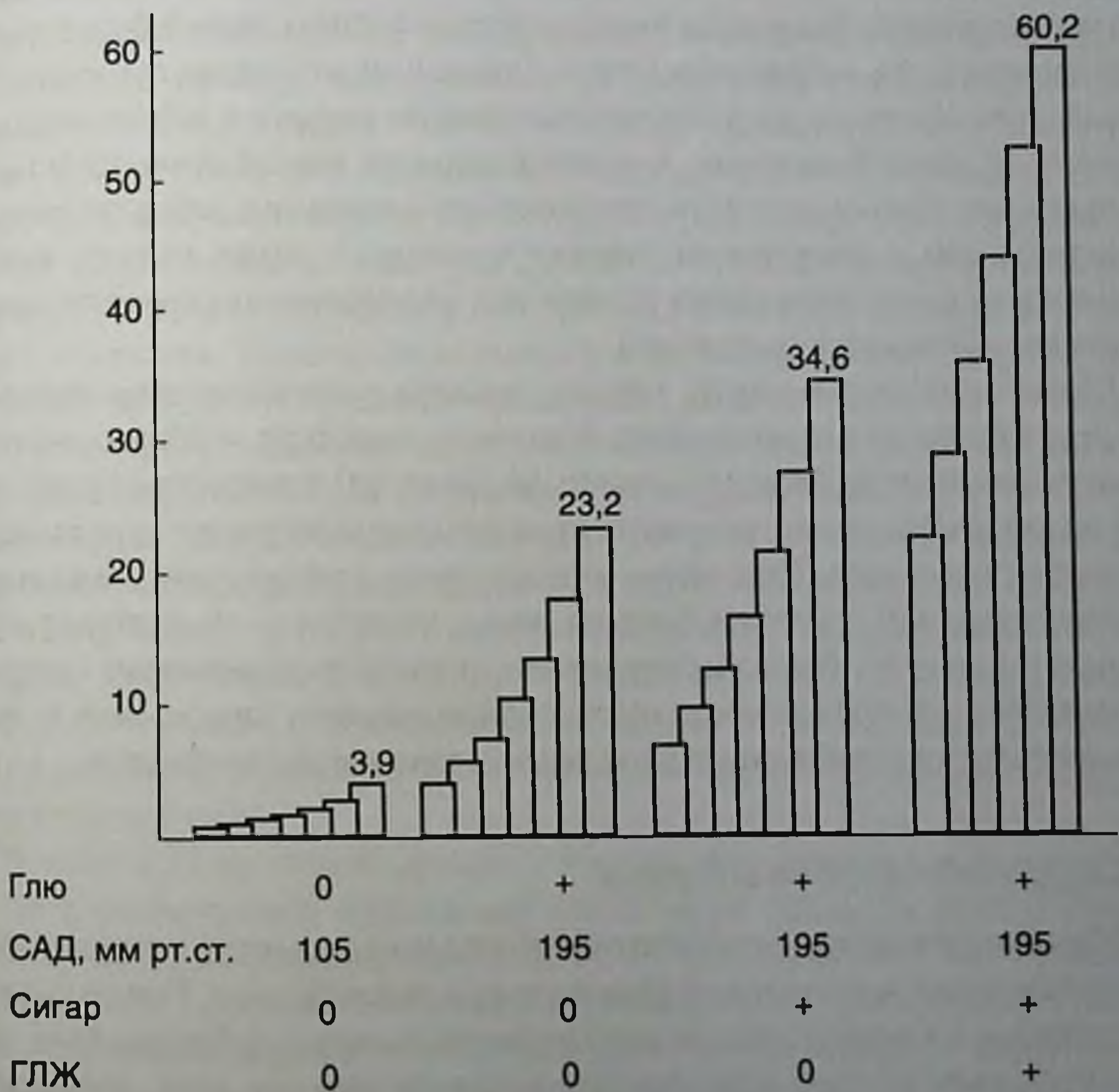


Рис. 2.54. Частота сердечно-сосудистых болезней (суммарный инцидент за 8 лет на 1000 человек, %) в зависимости от непереносимости глюкозы (Глю), величины систолического артериального давления (САД, мм рт.ст.), курения сигарет (Сигар), наличия гипертрофии левого желудочка по данным электрокардиограммы (ГЛЖ). Концентрация холестерина в крови возрастает внутри каждой из четырех групп от 185 до 335 мг/дл. Данные для мужчин 35 лет при включении во Фремингемское исследование (по Kannel W.B. Ann. Intern. Med. 1979. 90(1): 85–91. P. 1985)



к крайнему неблагоприятному диапазону, но повышение риска заболевания в этом диапазоне велико. Метод, который позволяет корректно связать количественный признак с качественными исходами, — логистическая регрессия, рассматривается в разд. 2.8.2.

Особенно резко повышается риск у лиц, имеющих несколько факторов риска одновременно (рис. 2.54). Однако с точки зрения общественного здоровья лица, у которых риск наиболее велик, представляют ограниченный интерес, поскольку они составляют очень небольшую часть популяции. Даже если вмешательство у таких людей будет очень действенным, на заболеваемости в популяции это скажется незначительно, так как вклад этих немногочисленных случаев в заболеваемость невелик. С другой стороны, в индивидуальной перспективе отдельного человека можно получить значительное снижение заболеваемости, инвалидности и смертности именно и только у таких редких людей со многими факторами риска (например, одновременно курящих, имеющих гипертензию и ожирение).

Существует упрощенный вариант анализа количественных данных. Их редуцируют до качественного, например, выделяют «гиперхолестеринемию» (концентрация выше, например 200 мг/дл) и вычисляют риск при «гиперхолестеринемии», выделяют курящих (кто выкуривает хотя бы одну сигарету ежедневно). При таком анализе часть информации, связанной с количественной стороной фактора риска, теряется — не используются данные о степени гиперхолестеринемии, о числе выкуриваемых сигарет. Соответственно, возможность исследования выявить зависимость и возможность использовать результаты исследования ограничиваются.

### 2.7.3. Дополнительный риск

Непосредственным результатом наблюдения является инцидент болезни в группе с экспозицией (фактором риска) и без нее. Разность этих инцидентов называют *абсолютной разностью рисков* (абсолютным дополнительным риском или атрибутивным риском, от англ. *attributable risk* — риск, приписываемый данной экспозиции как ее следствие). Долю случаев болезни в популяции, которую таким образом можно приписать имеющейся в популяции экспозиции, называют атрибутивной фракцией (англ. *attributable fraction*).

Чаще повышение риска вследствие экспозиции выражают отношением вероятности болезни у лиц, подвергшихся экспозиции, к ве-



роятности болезни у не подвергшихся экспозиции (*относительный риск* — ОР) (см. с. 269). Относительный риск выше единицы соответствует повышенному риску. Это превышение нередко выражают в долях от относительного риска — *относительной разностью рисков* (ОРР, иначе называемой «объяснимая доля риска», англ. *attributable fraction (exposed)* или *incidence rate fraction*).

$$\text{ОРР} = [A/(A + B) - C/(C + D)] / [A/(A + B)] \text{ или } \text{ОРР} = (\text{ОР} - 1) / \text{ОР}.$$

Использование доли ОР подчеркивает, что именно эта часть риска объясняется экспозицией (зависит от нее)<sup>1</sup>. Относительная разность рисков — это доля риска болезни, которая связана с данной экспозицией, объясняется ей и *может быть* устранена, если эту экспозицию устранить. Выражение доли риска болезни, зависящей от экспозиции, в виде доли относительного риска (ОР — 1) необходимо в основном для того, чтобы сопоставлять результаты разных исследований и разные популяции. Такая необходимость возникает потому, что в разных условиях существует разный «фоновый» риск заболеваний — инцидент при отсутствии экспозиции (C/(C + D)). Использовать абсолютную разность рисков для сравнений безотносительно этого фонового риска нельзя, так как абсолютная разность рисков обычно тем больше, чем больше фоновый риск. Даже в одной популяции равное изменение экспозиции приводит к разному снижению риска в зависимости от того, какая часть популяции, с каким уровнем риска избирается (рис. 2.55). Вместе с тем абсолютная разность рисков является *более надежной мерой важности экспозиции в конкретной популяции*, чем относительный риск.

В табл. 2.23 приведен пример того, как при равном относительном риске с увеличением преваленса экспозиции (фактора риска) повышается доля популяции, находящейся под дополнительным риском заболевания. При действии слабого фактора риска, имеющего ОР только 1,3, дополнительная доля заболевающих в популяции, может быть большой, если преваленс этого фактора риска велик.

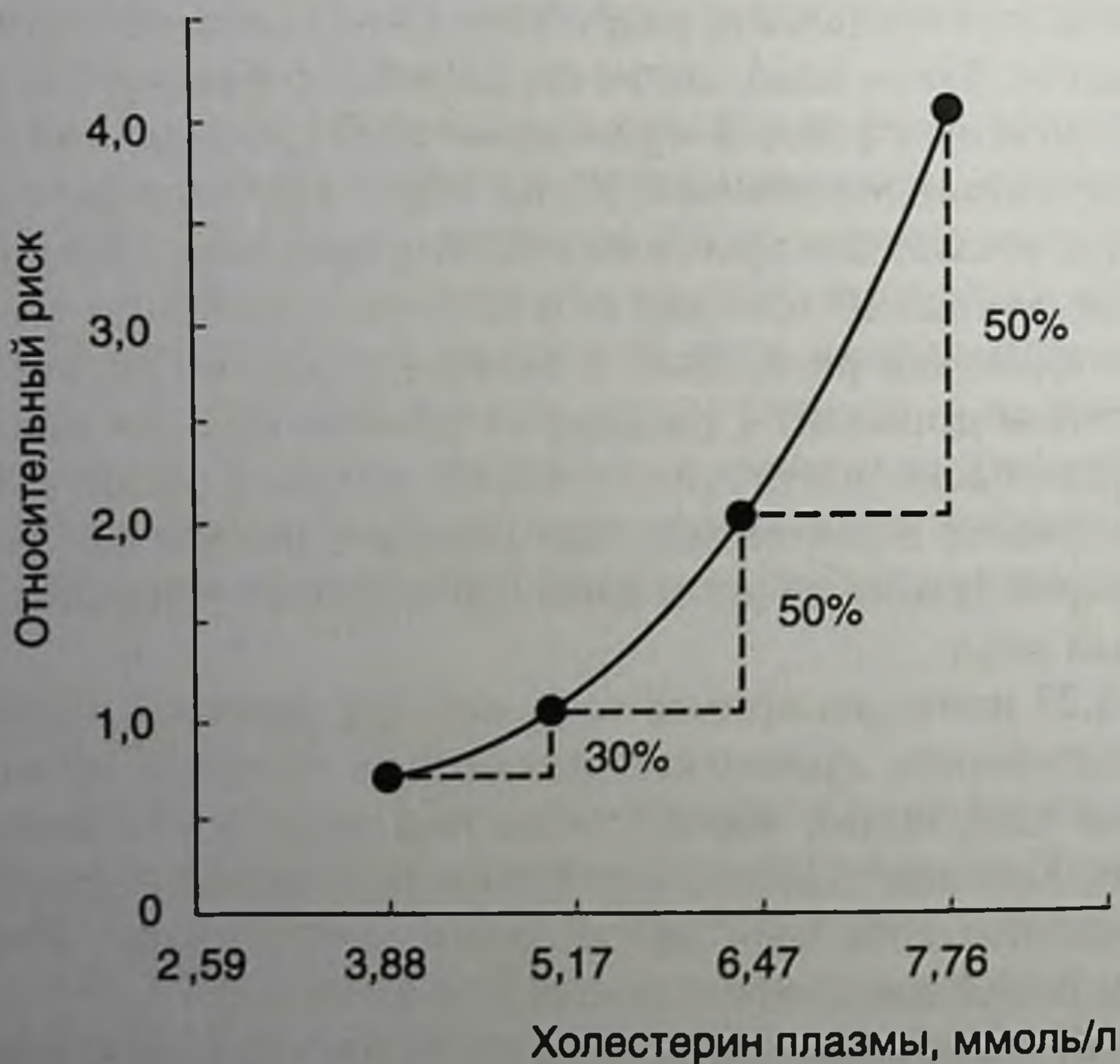
Эффект от устранения экспозиции, оцениваемый по дополнительному риску и ОР, только приблизителен. Окончательную оценку величины снижения риска при устранении экспозиции могут дать только

<sup>1</sup> Относительную разность рисков называют также по-английски *attributable fraction, etiological fraction, attributable proportion (of risk), proportion of risk reduction, частота эффекта профилактики (protective efficacy rate)*.



контролируемые экспериментальные исследования с вмешательством. Общее правило таково: устранение вредного фактора на производстве может полностью устранять связанный с ним дополнительный риск, но модификация индивидуальных факторов риска, таких как курение или гиперхолестеринемия, приводит к снижению риска меньшему, чем дополнительный риск.

Дополнительный риск не может служить для оценки доли случаев болезни, вызванных отдельным фактором. Если фактор является необходимой причиной, то он присутствует во всех случаях развития болезни (см. определения в разд. 1.1.1). Число людей, подвергающихся действию этого фактора, но не заболевших, определяется тем, как часто действуют другие факторы, которые вместе с необходимым создают комплекс достаточных условий. Поэтому невозможно определить



**Рис. 2.55.** Гипотетический размер снижения смертности от ишемической болезни сердца при снижении концентрации холестерина в плазме крови. При постоянной величине снижения концентрации холестерина (50 мг/дл или 1,29 ммоль/л) ожидаемый эффект тем больше, чем выше исходная концентрация (по данным исследования MRFIT: Grundy S.M. JAMA. 1986. № 256. P. 2849–58)



«удельное значение» отдельной причины в происхождении случаев заболевания абстрактно, вне зависимости от других причин. Даже для таких очевидных событий, как перелом ноги при спуске на лыжах с горы, наряду с падением, можно обнаружить ряд компонентных причин. Так, само занятие высокотравматичным горнолыжным спортом сопряжено с характеристиками личности и социальными условиями, а падение, в свою очередь, с употреблением алкоголя или личными проблемами, мешавшими сосредоточиться на спуске. Эти соображения справедливы и для оценки соотношения наследственных особенностей и прижизненных воздействий (см. разд. 4.4).

Таблица 2.23. Влияние преваленса фактора риска на преваленс лиц, находящихся «под риском»

Относительный риск (OR)	Преваленс фактора риска (Pr)	Доля популяции, находящаяся под дополнительным риском, %*
1,3	0,25	7
1,3	0,75	18

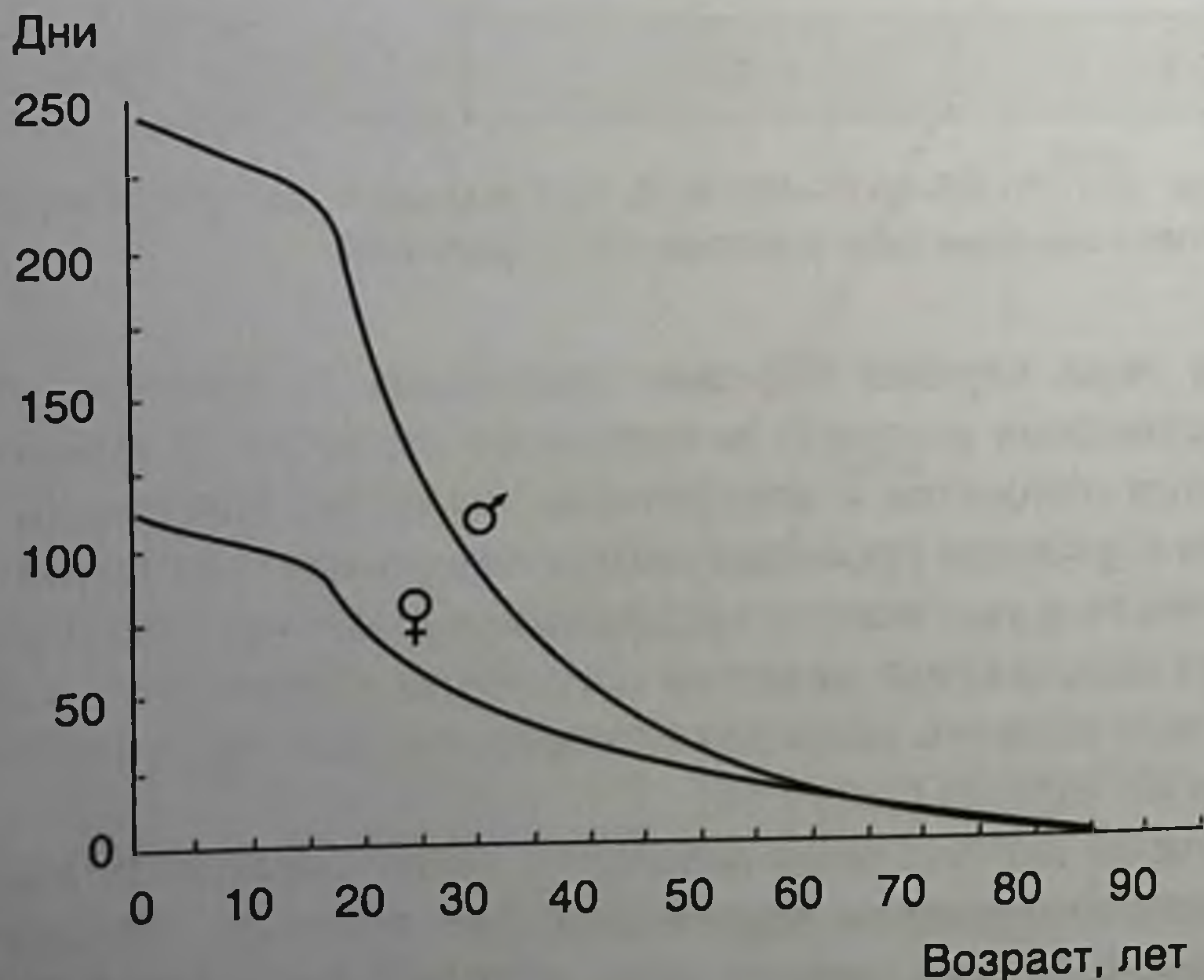
\*Рассчитывается как:  $(Pr \times (OR - 1)) / (Pr \times (OR - 1) + 1)$ , этот показатель называют также *атрибутивной фракцией случаев в популяции* (англ. *attributable fraction (population)*).

Оценка доли случаев болезни, зависящих от внешних факторов (производственных условий) методически нестрога. В лучшем случае такие оценки относятся к конкретным условиям. Например, в случае устранения дорожных происшествий можно ожидать заметного снижения смертности и увеличения продолжительности предстоящей жизни. Однако эта перспектива является абсолютно нереальной, и результат расчета может служить лишь для иллюстрации важности предотвращения гибели на дорогах (рис. 2.56).

В некоторых случаях такие оценки являются следствием ошибочных манипуляций с цифрами и ведут к неверным выводам. Например, иногда смертность от всех случаев рака, который чаще наблюдается при вредных условиях на производстве, суммируют и называют «смертностью, связанной с производственными условиями». Затем эту смертность относят ко всей смертности от опухолей. В результате получается «доля смертности от рака, вызванного производственными условиями».



Это неверная оценка, так как почти для каждого вида рака, помимо случаев, вызванных производственными условиями, существует значительное число случаев, не связанных с производственной экспозицией. Поэтому устранение экспозиции на производстве не может приводить к исчезновению всех случаев рака, отнесенного к вызываемым производственными условиями. При суммировании дополнительных рисков от действия известных факторов на вероятность возникновения одной болезни не удастся получить суммы 100%. Причина в том, что отдельные компонентные причины взаимодействуют одна с другой, поэтому *сумма случаев рака, связываемых с отдельными факторами, всегда будет более 100%*. Поэтому при изучении здоровья популяции приходится делать дополнительные поправки, позволяющие для каждого случая смерти обозначить одну причину (болезнь) и приписать возникновение болезни одному фактору риска. Такие расчеты систематически стали проводиться только в рамках проекта Глобального бремени болезней (GBD, <http://www.healthdata.org/>).



**Рис. 2.56.** Расчетное увеличение предстоящей продолжительности жизни (ордината) для лиц разного возраста при устранении смертности от дорожных травм (расчет по данным федеральной статистики США за 1983–85 гг. по Evans L. J. Gerontol. Soc. Sci. 1988. № 43. P. 186–93)



Среди всех факторов, влияющих на заболеваемость и смертность людей, есть более или менее сильные. Сила действия отдельного фактора обычно не является только его свойством, а зависит от преваленса других необходимых условий. Например, попытки оценить вклад состояния медицинской помощи в продолжительность жизни людей всегда весьма приблизительны. Даже применительно к самым действенным вмешательствам в истории медицины, таким как вакцинация, есть сомнения относительно размера эффекта от нее. Скептики полагают, что резкое снижение заболеваемости и смертности от инфекций связано с общим улучшением жизненных условий — доступностью чистой воды, теплого индивидуального жилища, питания, а не с вакцинацией. Несмотря на то что в последней трети XX в. сведения о состоянии здоровья населения в развитых странах достаточно точны, равно как и представления о распространенности факторов риска, не существует ясного представления о том, чем обусловлено снижение заболеваемости и смертности от разных причин, прежде всего от сердечно-сосудистых болезней в этот период. Преобладающая точка зрения — изменения обусловлены тем, что население активно выбирает «здоровый образ жизни». Однако многие признаки свидетельствуют, что помимо этого, по крайней мере, значительный вклад вносят повышение эффективности медицинской помощи и улучшение социальных условий. Поскольку изменения в популяции наблюдаются не в рамках контролируемого эксперимента, причины этих изменений всегда остаются не до конца понятными.

#### 2.7.4. Вмешивающиеся факторы

В выяснении причин болезней наиболее надежными из числа доступных применительно к человеку являются когортные исследования. Однако и они не абсолютно надежны. Ведь изучаемая особенность (например, обильное употребление кофе) не была привнесена исследователем случайно, как в контролируемом испытании. Она, в свою очередь, может зависеть от других, неизвестных нам особенностей. Поэтому если у мужчин, пьющих много кофе, чаще возникает инфаркт миокарда, то это может быть следствием других причин. Например, употребление кофе и подверженность инфаркту могут быть следствием некоторого общего фактора (образование, доход, профессия и т.д.). Такие *вмешивающиеся факторы* (англ. *confounding factors*,



*nuisance factors*) могут как уменьшать выраженность связи, так и увеличивать ее (рис. 2.57).

В истории эпидемиологии и медицины много примеров связей, которые считались причинными, но при детальном исследовании оказались ложными. С кофе связана еще одна гипотеза, оказавшаяся ложной. В ряде исследований типа сравнения с контролем было показано, что у женщин, беременность которых завершилась выкидышем или рождением ребенка с соматическими дефектами, в анамнезе беременности отмечается большее употребление кофе, чем в контрольных случаях благополучных беременностей. Кажется, что это достаточное основание для того, чтобы предположить вредное действие кофе. Однако позднее было установлено, что чем выраженнее побочные эффекты гормональных изменений при беременности (тошнота, рвота, нарушения аппетита), тем *ниже* вероятность выкидыша и рождения ребенка с дефектами развития. Чем выраженнее эти побочные эффекты гормональных изменений, тем *меньше пьют кофе* беременные. И наоборот, те женщины, у которых относительно мала концентрация гормонов беременности, продолжают пить относительно много кофе. Кофе оказывается не повреждающим фактором, а просто маркером низкой концентрации гормонов беременности. Окончательный ответ мог бы дать эксперимент, но использовать потенциально вредное вмешательство в эксперименте на человеке нельзя. Выход был найден в виде эксперимента по уменьшению потребления кофе беременными. Это не оказало измеримого влияния на массу тела новорожденных и срок гестации. К сожалению, мы не знаем большего, не можем описать всего причинного механизма, но у нас есть основания думать, что и в этом случае кофе оказывается напрасно заподозренным, он всего лишь сопутствует истинной причине (см. рис. 2.57, б).

Действие посторонних факторов на результат исследования неизбежно. Выбирая пациентов по стадии и форме болезни, врач обычно не может еще подбирать их и одного пола, и одного возраста, и с одинаковыми сопутствующими болезнями и т.д. Поэтому обычен компромисс: сначала больных подбирают по минимуму доступных признаков (например, по полу и возрасту), а затем из исследования исключают пациентов с крайне отличающимися характеристиками (например, лиц, страдающих тяжелыми сопутствующими болезнями). Далее при анализе данных различия между пациентами по привходящим признакам, например курению, устраняют с помощью стратификации и вычислительных методов.



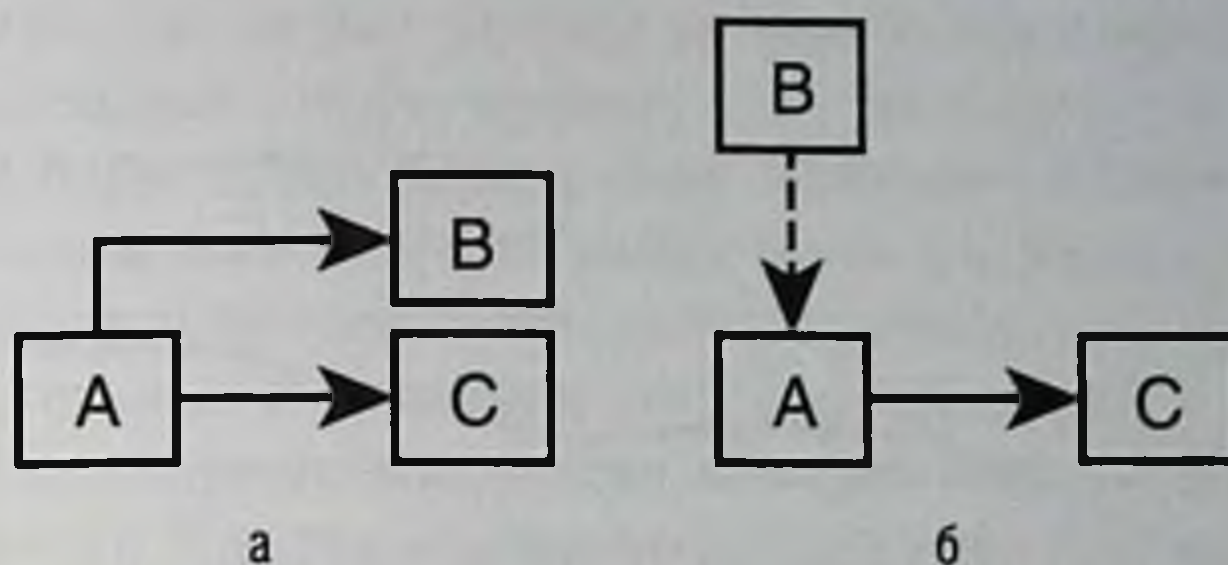


Рис. 2.57. Ложные причинные связи: а — В не является причиной для С, оба, С и В, являются следствием общей причины А; б — В не является причиной для С, а лишь сопутствует причине А

*Стратификация* в этом контексте означает разделение обследованной группы по привходящим признакам, например по курению, на подгруппы (страты, от лат. *stratum*). После этого анализ, например, связи инфицированности геликобактером с употреблением алкоголя выполняется в каждой подгруппе. Если связь сохранится в этих стратах — курящих и не курящих — значит, она не может быть объяснена большей частотой курения у употребляющих больше алкоголя. Главным *вычислительным методом* учета влияния вмешивающихся факторов является коррекция в форме стандартизации (см. разд. 1.3.8) или регрессионной коррекции (см. с. 300).

Если речь идет об изучении действия условий труда на здоровье, то врач всегда оказывается перед проблемой самоотбора работников и отбора работников работодателем. На работу принимают относительно здоровых людей, поэтому в популяции смертность всегда выше, чем среди работающих. На предприятия, где имеются вредные условия труда, людей специально отбирают по состоянию здоровья. Поэтому попытка обнаружить повышенную смертность у работников вредного производства может оказаться безуспешной. Наоборот, неоднократно показано, что у работников производств, имеющих вредные условия труда, смертность ниже, чем в популяции.

Пример одного из исследований приведен на рис. 2.58. Смертность от рака у работающих почти в 2 раза ниже, чем в популяции, хотя и приближается к популяционной с годами. Это значит, что действие первичного отбора по состоянию здоровья на смертность постепенно уменьшается. Если речь идет не о смертности, а об инциденте, то отличие работников от популяции меньше. Эта особенность инцидентса



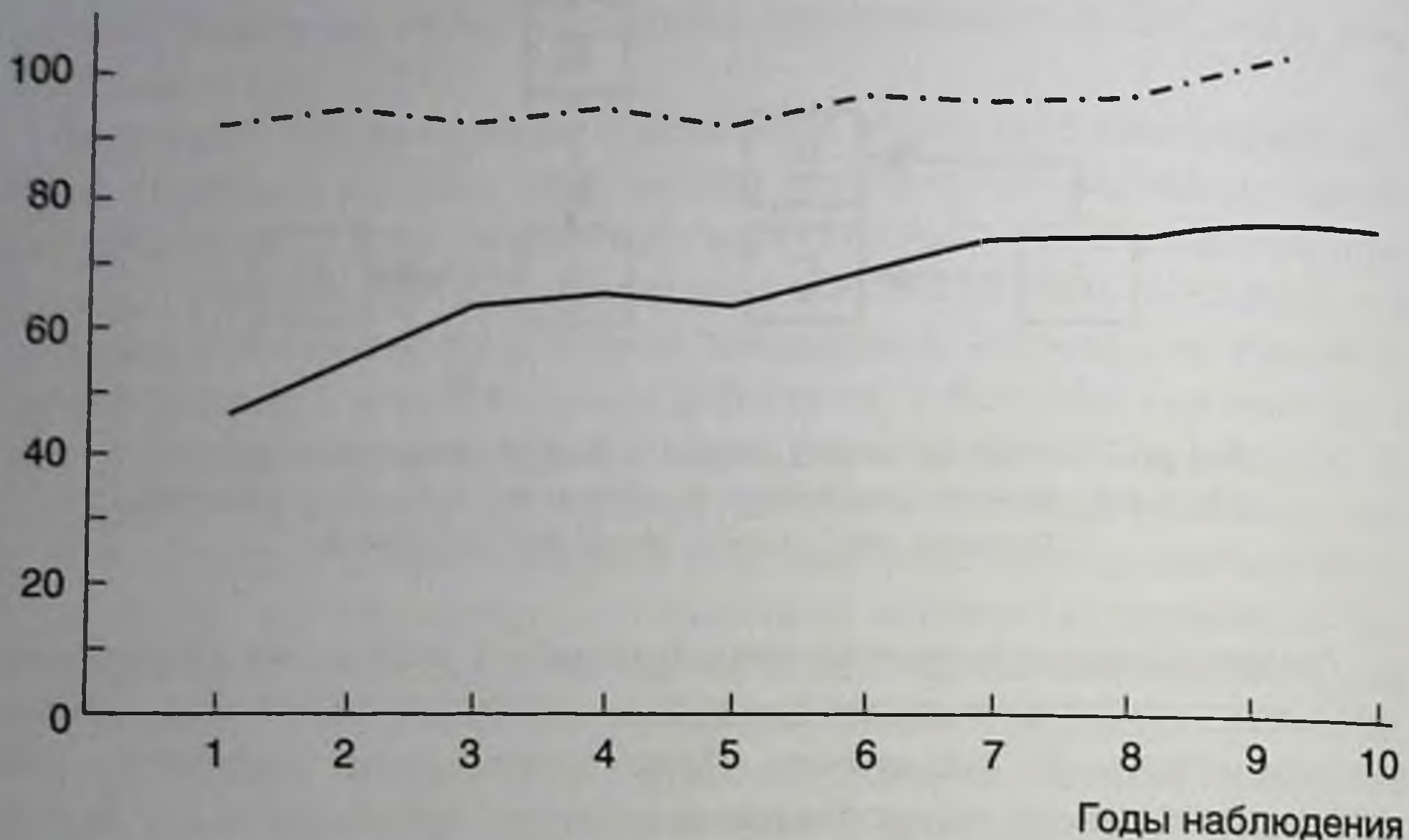


Рис. 2.58. «Эффект здорового работника». Стандартизованное отношение смертности у строительных рабочих Швеции ниже, чем в популяции (ордината, проценты от популяционных показателей). Линия — общая смертность, штрих-пунктир — инцидент рака (по данным А. Englund, G. Engholm., cit. by Breslow N.E., Day N.E. Statistical methods in cancer research. Vol. 1. IARC, Lyon, 1987)

понятна — еще не развившиеся случаи не могут влиять на самоотбор и отбор работников.

«Эффект здорового работника» нельзя измерить, и устранить его влияние на результаты продольного исследования. Если смертность у рабочих резинового или коксового производства, контактирующих на рабочем месте с ядовитыми газами, ниже, чем в популяции, то это не есть доказательство безвредности условий труда. Это означает лишь, что «эффект здорового работника» по размеру больше, чем вредное действие условий труда. Например, участники работ по ликвидации последствий Чернобыльской аварии имеют смертность ниже, чем смертность в популяции, но это не означает, что вредного воздействия на них не было. Просто к этим работам привлекались здоровые люди, в основном прошедшие отбор по состоянию здоровья.

Действие вмешивающихся факторов разнообразно. Они могут не только уменьшать размер изучаемого вреда, как описанный «эффект здорового работника», но и преувеличивать оценку и создавать види-



мость связи там, где ее нет. Одним из вариантов структуры исследования, позволяющих лучше выявить действие вредной экспозиции, является сравнение групп работников, которые подвергаются экспозиции разной величины на своих рабочих местах. Выявление большего инцидента у подвергающихся большей экспозиции людей может быть более чувствительным и доказательным исследованием.

Универсальных правил выявления действия вмешивающихся факторов не существует. Каждое исследование строится таким образом, чтобы выявить действие изучаемых факторов и ограничить влияние прочих факторов. Наличие вмешивающегося фактора не является дефектом исследования. Вмешивающийся фактор — это истинный феномен, который обычно должен быть изучен, и его влияние на исследуемый процесс должно быть оценено. Этим влияние вмешивающегося фактора отличается от систематической ошибки (см. с. 41). *Систематическая ошибка* — это дефект в организации исследования, который привел к получению смещенной оценки, и это смещение может быть результатом того, что не был учтен вмешивающийся фактор.

Чем тщательнее план исследования, тем больше вероятность получения правильного, несмещенного результата. Однако почти никогда одно эпидемиологическое исследование не может быть проведено так, чтобы исключить решающее влияние вмешивающихся факторов на результат. Это главная причина, по которой требуется несколько исследований одной проблемы, дающих согласующиеся результаты, чтобы можно было считать размер вредного действия некоторой экспозиции установленным.

### **2.7.5. Когортное исследование по ретроспективным данным**

Когортное исследование, как и всякое другое, нацелено на реконструкцию истинной картины развития событий во времени — проспективной картины. Однако когортное исследование дорого и длительно. Поэтому интересна возможность использовать данные, имеющиеся в документации, для исследования. Когортные исследования рождаемости и смертности по записям гражданских актов являются типичными исследованиями, ретроспективными лишь по способу сбора данных.

Ярким примером когортных исследований, выполненных по ретроспективно полученным данным, являются исследования состоя-



ния здоровья людей, рожденных с разными признаками и от разных родителей. Это возможно, поскольку в документации родильных домов и в медицинских архивах содержатся сведения о роженицах и новорожденных (см. рис. 1.8). В настоящее время по ретроспективно полученным данным проведено множество исследований влияния перинатальных факторов на последующее развитие ребенка и даже на его здоровье в зрелые годы.

Когортное исследование по ретроспективно полученным данным возможно тогда, когда существуют достаточно полные сведения обо всей когорте или когда существуют сведения о репрезентативной (случайно выбранной) части когорты. Поскольку репрезентативность невозможно проверить, нужно использовать только те базы данных, которые содержат информацию об абсолютном большинстве членов когорты. Именно поэтому таких исследований мало, и чаще всего это — исследования, использующие сведения о рождении. Недостатком ретроспективных исследований является то, что они могут быть проведены только с использованием тех признаков, которые были измерены и зарегистрированы в прошлом. Достоинством является то, что можно изучить связь с любыми последующими событиями. Наоборот, в проспективном исследовании изначально можно исследовать каждый доступный исследованию признак включенных в исследование людей, но позднее связать его можно только с теми событиями, которые регистрировались в ходе исследования. Доброкачественные записи о рождении других событиях в жизни людей еще недавно были доступны только в некоторых европейских странах. По мере распространения компьютерных технологий все большие объемы разнообразных данных хранятся и при наличии доброжелательной к науке политики могут быть использованы для исследований.

## Контрольные вопросы и задачи к разделу 2.7

1. Дополнительный (атрибутивный) риск используется для оценки максимальной доли общего числа случаев заболевания, которая связана с изучаемым фактором риска. Так, до 80% случаев рака легких зависят от курения. Для конкретной болезни дополнительный риск определяется:
  - а) численностью популяции;
  - б) преваленсом данного фактора риска;



- в) длительностью болезни;
  - г) относительным риском для данного фактора риска.
2. В отношении когортных исследований справедливо:
- а) они продольны по структуре;
  - б) они проспективны по способу проведения;
  - в) субъекты включаются в них по признакам, существующим до возникновения изучаемого исхода;
  - г) это преимущественно описательные, а не аналитические исследования.
3. В когортном исследовании среди лиц, не имевших ишемической болезни сердца, за 5 лет наблюдения у 45 человек на 1000 развились симптомы ишемической болезни сердца. Это показатель:
- а) мгновенного преваленса;
  - б) кумулятивного инциденса;
  - в) стандартизованной заболеваемости;
  - г) относительного риска.
4. В том же исследовании у лиц с нарушенной толерантностью к углеводам частота стенокардии была в 1,6 раза чаще, чем ожидаемая за 10 лет. Это показатель:
- а) мгновенного преваленса;
  - б) кумулятивного инциденса;
  - в) стандартизованной заболеваемости;
  - г) относительного риска.

## 2.8. ИССЛЕДОВАНИЯ ТИПА СРАВНЕНИЯ С КОНТРОЛЕМ

Экспериментальные исследования не могут быть проведены для выяснения причинной связи заболевания человека с экспозицией. Например, недопустимы эксперименты с вдыханием паров ядовитого газа, встречающегося в атмосфере производственных помещений. Главным типом исследований этиологии болезней человека являются ИСК (англ. *case-control studies*, CCS). В основе ИСК лежит сопоставление экспозиций, которым подвергались больные с прошлыми экспозициями здоровых людей (рис. 2.59).

Из-за этого ИСК иногда называют ретроспективными исследованиями, что неверно, поскольку в ИСК реконструируется нормальное



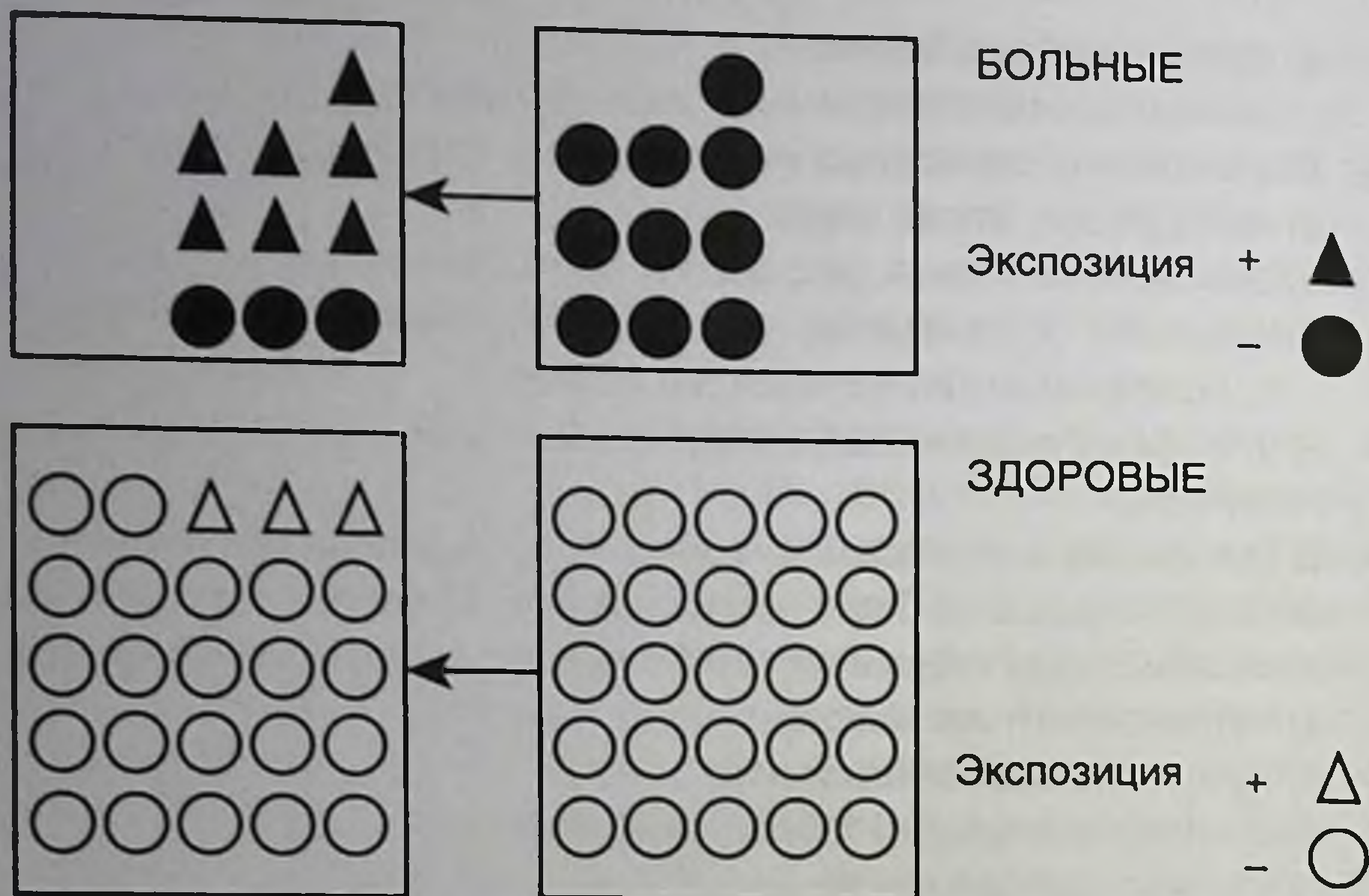


Рис. 2.59. Структура исследования типа сравнения с контролем при изучении причины болезни. Подбирают группы людей, имеющих заболевание (случаи) и здоровых (контроли). В каждой группе устанавливают, кто подвергнулся экспозиции (шансы): среди больных 7:3, среди здоровых 3:22

(проспективное) течение событий во времени на основе доступной информации. В условиях, когда экспериментальное или проспективное исследование невозможно, ИСК являются способом реконструкции развития заболеваний по данным однократного исследования. Это самый важный и сложный тип аналитических исследований, вариант обсервационного исследования (без вмешательства), нацеленный на изучение последствий вмешательства (экспозиции). ИСК являются наиболее оперативными исследованиями, экономны в реализации и при корректном проведении способны дать правильную оценку гипотезы.

В широком смысле «экспозицией» может быть любое различие между сравниваемыми группами (больными и здоровыми людьми). Важно, что между двумя группами всегда существует множество различий, так как число возможных признаков огромно. Поэтому ИСК только тогда дает результат, имеющий смысл, когда изучаются признаки, заранее запланированные, и в отношении которых заранее сформулирована гипотеза. «Забрасывать сеть» в поисках различий между группами возможно, но обнаруженные таким образом различия обычно случайны,



ничего не доказывают и могут быть лишь предпосылкой для последующих спланированных исследований.

Гипотеза, проверяемая в исследовании, может иметь разное происхождение. Знания этиологии и патогенеза болезни лежат в основе множества ИСК. Нередко гипотеза вытекает из наблюдений, указывающих на связь болезни с некоторым признаком, или из существующих представлений о причине заболевания. Примером ИСК, основанных на наблюдении, являются ИСК, выявляющие связь заболевания и предшествующей экспозиции. Например, отдельные наблюдения развития рака влагалища у девочек, матери которых получали диэтилстильбэстрол для лечения угрожающего аборта, стали основой для ИСК, подтвердившего тесную связь лечения с этим побочным эффектом. Для лечения угрожающего аборта с помощью диэтилстильбэстрола были основания лишь в виде предположений о том, что преждевременные роды могут быть связаны с «гормональной недостаточностью». Эффективность такого лечения никогда не была доказана в контролируемом эксперименте. Тем не менее применение диэтилстильбэстрола было широким, и даже после того, как вредные отдаленные последствия такого вмешательства были доказаны в ходе ИСК, еще многие годы в разных странах применяли диэтилстильбэстрол. Ни тогда, ни сейчас не существует точных знаний о том, каким образом диэтилстильбэстрол на поздних сроках беременности может оказывать свое вредное влияние на плод. Результат ИСК позднее был подтвержден в проспективных исследованиях наблюдением за девочками, родившимися от матерей, получавших диэтилстильбэстрол во время беременности. Другой пример ИСК см. на врезке на с. 25.

В исследованиях середины XX в. выкристаллизовалась парадигма эпидемиологии неинфекционных болезней, основанная на внимании к индивидуальным особенностям людей и на их изучении методом «черного ящика» (см. разд. 4.2).

Этот метод нашел свое выражение в формулировке «фактор риска» (см. разд. 1.1.4). В ИСК можно лишь выявить связь фактора риска с болезнью. Для того чтобы выявить, увеличена ли вероятность развития болезни у здоровых людей при наличии фактора риска, необходимо выполнить проспективное исследование. Для того чтобы установить, возможно ли, устранив фактор риска, предотвратить болезнь, нужен эксперимент. Таким образом, ИСК часто являются первым важным шагом в установлении истинности причинных связей.



Неверно рассматривать ИСК как дешевый и несовершенный заменитель более доказательных проспективных и экспериментальных исследований. Последние обычно невозможны на людях, а проспективные исследования так дороги, что их проведение является в истории эпидемиологии обычно результатом удачного стечения обстоятельств. Поэтому ИСК оказываются в центре эпидемиологической методологии.

### 2.8.1. Методика исследования типа сравнения с контролем

В простом ИСК сравнивают больных и здоровых лиц по признакам, которые могут отражать причину болезни и которые на этапе исследования обозначаются как факторы риска. Например, можно сравнить лиц с ОИМ и контрольных пациентов по признаку ежедневного употребления кофе. В результате получается таблица (табл. 2.24), подобная той, которая использовалась для анализа результатов когортного исследования (см. табл. 2.22).

Таблица 2.24. Способы вычисления относительных показателей связи болезни и фактора риска

Употребление кофе	Острый инфаркт миокарда		Всего
	Есть	Нет	
Есть	A	B	A + B
Нет	C	D	C + D

Очевидно, что при данной структуре исследования нельзя рассчитывать риск ОИМ при экспозиции (употреблении кофе) как  $A/(A + B)$ , поскольку не было выборки лиц, употребляющих кофе, и не было наблюдения за развитием у них инфаркта. Можно высчитать лишь, каково отношение числа лиц, употребляющих кофе, к не употребляющим кофе среди больных ОИМ и среди контрольных пациентов. Дроби  $A/C$  и  $B/D$  называются шансами, а их отношение — отношением шансов.

В ИСК больных и контрольных пациентов подбирают по их состоянию и затем уточняют имевший место ранее риск. У таких паци-



ентов соответственно это будут «шансы иметь или не иметь фактор риска при наличии болезни» (или его отсутствии). Отношение шансов  $OR = (A/C)/(B/D)$ .

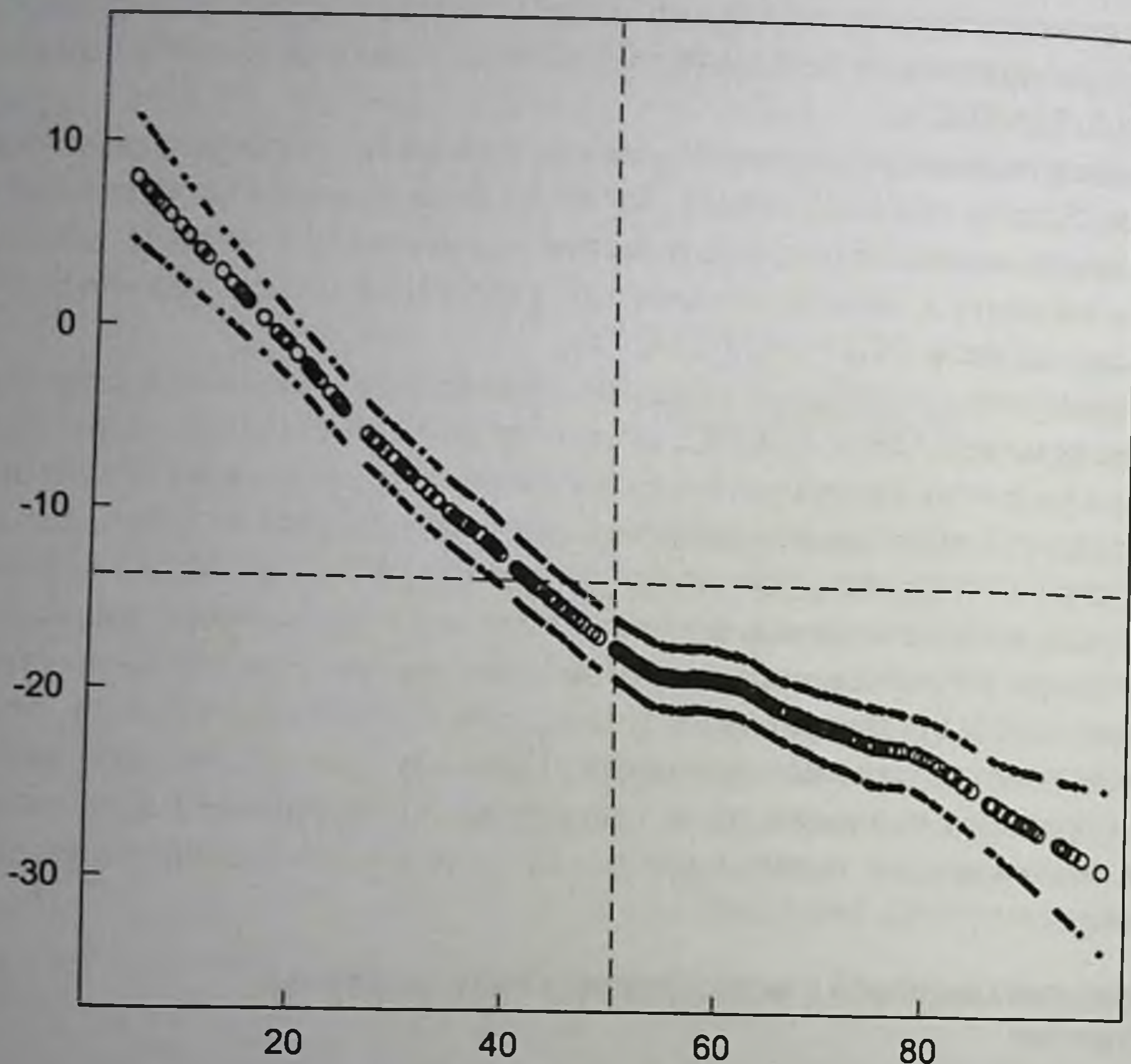
В когортном исследовании пациенты или не подбираются специально или, будучи подобранными, затем из исследования не исключаются. У таких пациентов соответственно вычисляются «шансы заболеть или не заболеть при наличии фактора риска (или его отсутствии)». Отношение шансов  $ОШ = (A/B)/(C/D)$ .

Формальные алгебраические преобразования приводят к одинаковому результату:  $OR = AD/BC$ , но это не должно скрывать принципиального различия интерпретируемых данных. В случае проспективного исследования полученные сведения дают представление о том, как часто изучаемый признак приводит к болезни. В случае ИСК полученные сведения лишь выявляют связь болезни с различиями пациентов в прошлом. Установлено, что отношение шансов, полученное таким образом, лишь приблизительно соответствует отношению риска, получаемому в когортном исследовании. Поэтому при оценке результатов ИСК обращают внимание не на статистически значимые связи, а только на *многократное* повышение риска (отношения шансов) при экспозиции.

### Уменьшение смещений в исследованиях типа сравнения с контролем

Для того чтобы ИСК давали более правильную оценку связи болезни с предшествующей экспозицией, необходимо соблюдение ряда условий. Главным из них является неискаженный сбор данных. Собирающий информацию специалист не должен знать, в каких целях, с кем он беседует или на кого делает выписки из истории болезни — на пациента основной или контрольной группы. В противном случае возможно искажение в угоду изучаемой гипотезе за счет более настойчивого и детального опроса или предвзятой оценки документов. Единообразие в сборе анамнеза преследует цель предотвратить не только ошибки интервьюера, но и ошибки пациента. На рис. 2.60 приведен пример искажения массы тела в утверждениях участниц исследования через несколько лет по сравнению с зарегистрированной величиной в проспективном исследовании. Очевидно, что в ИСК трудно даже приблизительно оценить, какого размера будет *смещение памяти* (систематическая ошибка, связанная с ошибками воспоминания).





**Рис. 2.60.** Смещение памяти: занижение массы тела при воспоминании (ордината) по сравнению с зарегистрированной фактически в подростковом возрасте (абсцисса, везде в процентилях индекса массы тела). Пунктир отмечает медиану истинной массы тела (вертикаль) и медиану смещения (горизонталь). Все участницы исследования разделены на группы по исходной массе тела; кружки — средняя, точки — верхняя и нижняя границы доверительного интервала (по Must A. et al. Am. J. Epid. 2002. № 155. P. 672)

При комплектовании группы больных и группы сравнения не должно быть излишних ограничений. Необоснованное решение об исключении каких-либо людей может привести к образованию ложных различий и ассоциаций. Например, если в группе больных не накладывать ограничений на сопутствующие болезни, а в группу сравнения не включать лиц с сахарным диабетом, то окажется, что в группе больных будут чаще присутствовать лица, принимавшие в прошлом противодиабетические препараты.



Для уменьшения вероятности ложных различий между группой больных и группой сравнения группы должны быть схожи по демографическим характеристикам (возрасту, полу, доходу). Включаемые в группу больных и в группу сравнения должны в идеале обследоваться одинаково. Это необходимо для того, чтобы убедиться, что у лиц в группе сравнения действительно нет болезни. Особенно это важно в случае болезней и состояний, которые длительно остаются невыявленными (опухоли, тугоухость, пародонтоз и т.п.). В действительности в большинстве ИСК лица, включенные в группу сравнения, не проходят такого диагностического обследования, как больные. Более того, иногда предполагаемый причинный фактор может провоцировать обследование разной глубины и соответственно разную частоту выявления болезни. Например, при употреблении кофе обычным является периодический дискомфорт в эпигастральной области. Пьющие кофе люди могут в связи с этими симптомами обследоваться, и у них повышается вероятность выявления ишемической болезни сердца и других болезней. При сопоставлении заболеваний с употреблением кофе может обнаруживаться связь, которая в действительности объясняется разницей в обследовании «больных» и группы сравнения. Такого рода систематические ошибки называют *смещениями ожидания* (см. с. 269).

В тех случаях когда известны характеристики, связанные с риском изучаемой болезни, у больных и контроля эти характеристики не должны существенно различаться. Например, при изучении связи ишемической болезни сердца с потенциальным фактором риска (например, потреблением пива) необходимо убедиться в равенстве концентрации холестерина, артериального давления и частоты курения в сравниваемых группах. В противном случае нельзя исключить, что различия между больными и контрольными пациентами будут связаны с этими побочными для данного исследования факторами.

Врачи обычно полагают, что лечат одинаково всех больных с одной болезнью. Соответственно упускается из виду возможное различие в ведении больных: в действительности спектр вмешательств и частота их применения зависят от течения болезни. Чем ярче, агрессивнее течение болезни, тем чаще применяются травматичные, но эффективные методы диагностики (рис. 2.61) и лечения.

Одним из способов уменьшения смещения является использование только полноценных документальных данных о прошлой экспозиции.



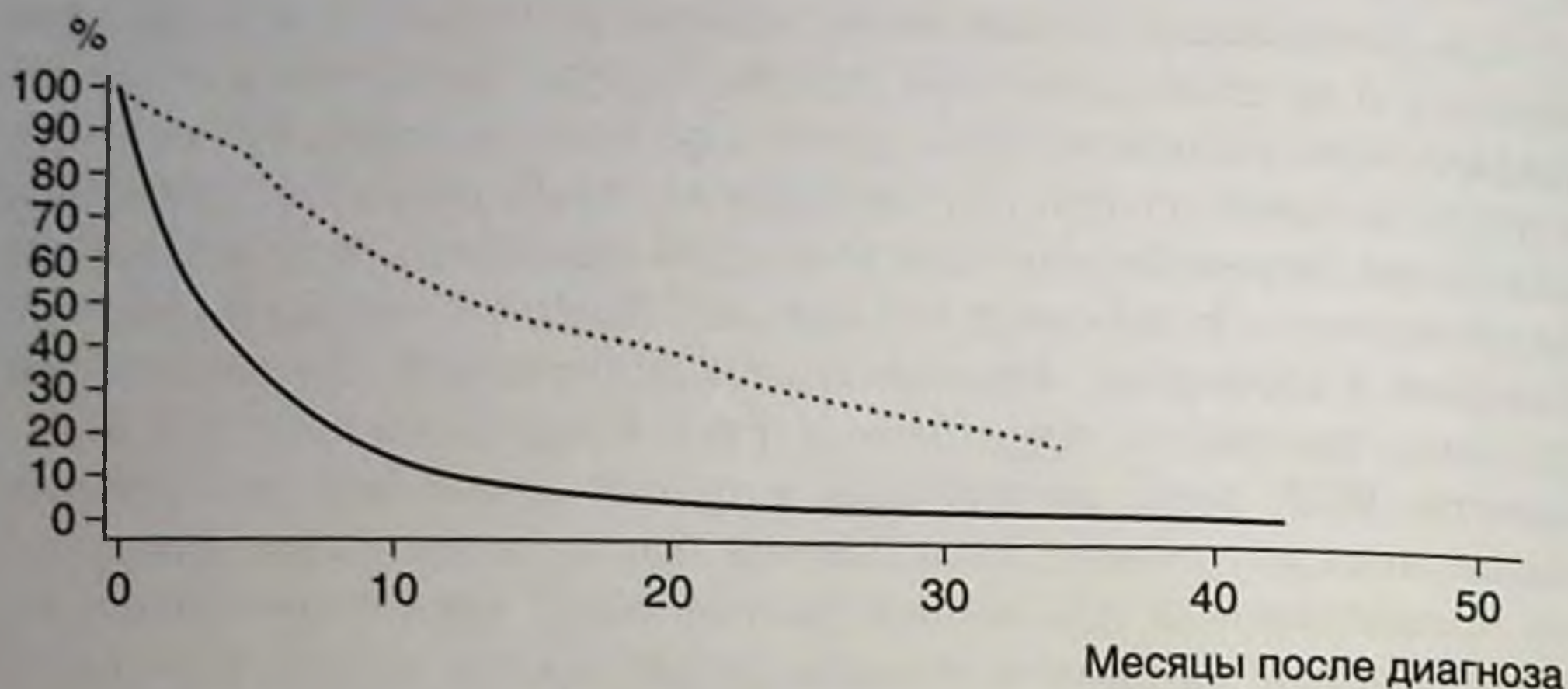


Рис. 2.61. Выживание больных раком поджелудочной железы, для которых существуют данные биопсии (линия,  $n = 142$ ) и для которых данных биопсии нет (пунктир,  $n = 979$ ). Различие значимо,  $p = 0,0001$  (по Hoppin J.A. et al. Ann. Epidemiol. 2002. № 12 (1). P. 1–6, с разрешения издателя)

Примером может служить исследование последствий Чернобыльской аварии, описанное во врезке.

### Рак щитовидной железы после Чернобыльской аварии

В течение нескольких лет после взрыва реактора у населения, подвергшегося загрязнению радиоактивными веществами территорий, обнаружили повышенный инцидент рака щитовидной железы (см. рис. 2.14). Поскольку до этого целенаправленных обследований населения не проводили, существовали сомнения: не является ли повышенный инцидент результатом более интенсивного обследования?

В 1998–2000 гг. в Гомельской области обследовали школьников, родившихся в разное время относительно взрыва 26 апреля 1986 г. Проводили ультразвуковое исследование щитовидной железы, и выполняли ряд биохимических тестов. Всех лиц с узловыми изменениями диаметром более 5 мм или нарушениями экзогенности ткани щитовидной железы подвергали аспирационной биопсии. Таким образом, не только лица, включенные в сравниваемые группы, прошли одинаковое обследование, но и особенности их экспозиции можно считать надежно установленными — в СССР система регистрации рождений была надежной.

Результаты исследования приведены в табл. 2.25 и на рис. 2.62. Уравнение логистической регрессии (принцип метода см. разд. 2.8.2), в котором группы были представлены как 0, 1 и 2, дало отношения шансов рака 11 (95% ДИ: 3–176) и 121 (95% ДИ 9–31 000) соответственно для групп 1 и 2 по отношению к группе 0,



Таблица 2.25. Частота обнаружения рака щитовидной железы в 1998–2000 гг. у детей, живущих в Гомельской области (по данным Shibata Y. Lancet. 2002. № 358. P. 1965. Частота в группе 0 рассчитана с поправкой по Ван дер Ваардену)

Обследовано	Пол	Период рождения		
		Январь 1987 г. – декабрь 1989 г.	Апрель 1986 г. – декабрь 1986 г.	Январь 1983 г. – апрель 1986 г.
		Группа 0	Группа 1	Группа 2
Всего	м	4826	1258	4810
Рак	м	0	0	7
Всего	ж	4646	1151	4910
Рак	ж	0	1	24
Частота, %	м	0,02	0,08	0,17
	ж	0,02	0,17	0,51

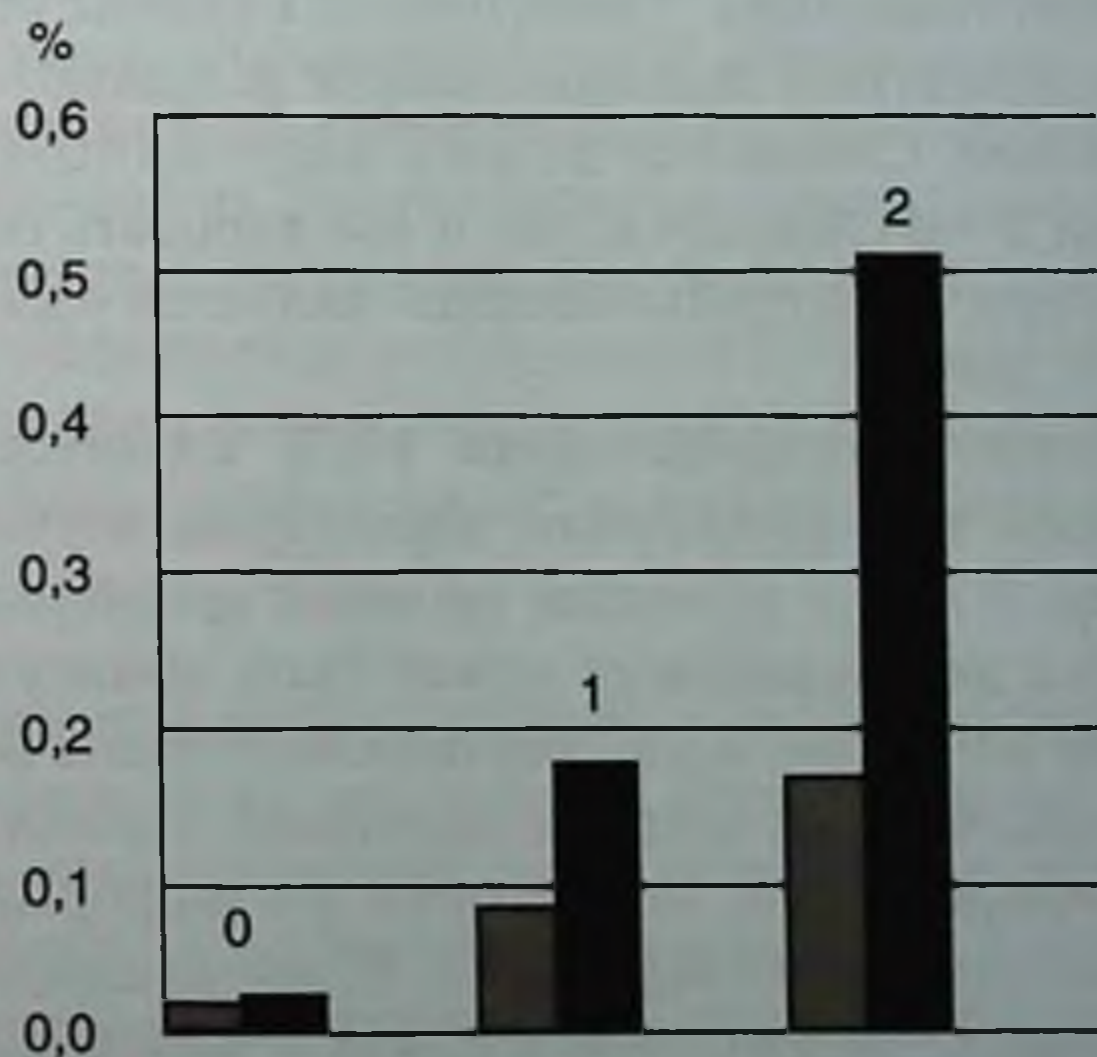


Рис. 2.62. Частота (%) рака щитовидной железы у детей, родившихся в разные годы в Гомельской области

при статистической значимости различий  $P = 0,006$ . Иными словами, риск рака щитовидной железы повышен у тех, кто облучился в первые месяцы жизни, и в меньшей степени – у облучившихся в перинатальном периоде. У родившихся



в период после декабря 1986 г., когда выброс короткоживущих изотопов йода в атмосферу резко сократился, частота рака щитовидной железы низка и, вероятно, не отличается от популяционной. В этом ИСК элемент неопределенности остается, так как неизвестно, сколько детей в каждой группе, выделенной по времени рождения, покинуло Гомельскую область. Если предположить, что избирательно покинули место постоянного проживания семьи с детьми, получившими большую дозу, то истинное отношение шансов может быть еще выше.

### **Интерпретация исследований типа сравнения с контролем**

При обнаружении связи между заболеванием и предшествующей экспозицией, например привычной диетой, результаты ИСК должны трактоваться с осторожностью. В оценке результатов ИСК обычно отношение шансов (ОШ) менее 2 рассматривается как низкое. В экспериментальном исследовании даже ОШ равное, например 1,2 и статистически значимо отличающееся от единицы будет рассматриваться как заслуживающее внимания. Такой подход к оценке результатов ИСК основывается на известной подверженности результатов ИСК разнообразным смещениям. Смещения оценки могут не только увеличивать связь экспозиции с заболеванием, но и маскировать ее. Наиболее известным смещением «восприимчивости» является «эффект здорового работника» (см. с. 280).

Другой проблемой интерпретации ИСК является неопределенность с первичным или вторичным характером изучаемого фактора риска. Например, у лиц с высоким уровнем временной нетрудоспособности в связи с заболеваниями может быть обнаружено увеличенное употребление алкоголя. Было бы опрометчивым объяснять такие случаи нетрудоспособности худшим здоровьем пьющих людей. Более правильной интерпретацией, основанной на информации, полученной за пределами такого исследования, будет предположение о том, что высокий уровень трудопотерь в связи с заболеваниями у таких работников является формой адаптации к производственным и бытовым условиям, и эта форма адаптации проявляется одновременно повышенным употреблением алкоголя. В свою очередь, проблемы, связанные со злоупотреблением алкоголем, могут оказывать влияние на трудовую адаптацию и на выбор работника в пользу отсутствия на работе «по состоянию здоровья».



На фоне сильных патогенных воздействий более слабые воздействия незаметны. Поэтому если исследователь хочет получить информацию не только об общеизвестных сильных факторах риска, то ему нужно правильно организовать исследование. Для этого нужно или исключить субъектов, подвергающихся экспозиции сильного фактора риска, или выделить их в отдельную группу. Например, при исследовании факторов риска рака легких действие курения настолько сильно, что обнаружить влияние других факторов риска на этом фоне — дело безнадежное. Необходимо исключить курящих, и в оставшейся группе действие других факторов риска может быть обнаружено.

### 2.8.2. Измерение связи

Под статистической связью понимают закономерное изменение величины одного признака при изменении другого. Обычно имеют в виду простой случай *прямой связи*, т.е. увеличения или уменьшения величины второго признака по мере увеличения первого (рис. 2.63, а, г). В абсолютном большинстве случаев речь идет о прямо *пропорциональной* (а) или обратно пропорциональной зависимости (в). Статистическая оценка связи не имеет ничего общего с причинно-следственными связями (отношениями). То, что первая величина увеличивается вместе со второй, не означает, что увеличение первой обусловлено увеличением второй. Так, обладание дорогим автомобилем положительно связано с обладанием дорогим домом, но тут нет ни причины, ни следствия: оба признака отражают третий — наличие высоких доходов.

#### Корреляция

Самым распространенным способом *измерения* связи является вычисление *коэффициентов корреляции*. Коэффициент корреляции отражает степень статистической связи между двумя признаками. Чаще всего используется коэффициент Пирсона ( $r$ , англ. *product-moment correlation*). Коэффициент равен 0 при отсутствии связи (рис. 2.63, б), +1 — при прямолинейной связи, -1 — при обратной линейной связи (рис. 2.63, а, в). Графически абсолютная величина коэффициента соответствует форме облака точек на скаттерграмме (диаграмме рассеяния): единице соответствуют точки, укладывающиеся на прямую, а нулю — образующие круглое облако (точнее — симметричное относительно вертикальной и горизонтальной осей).



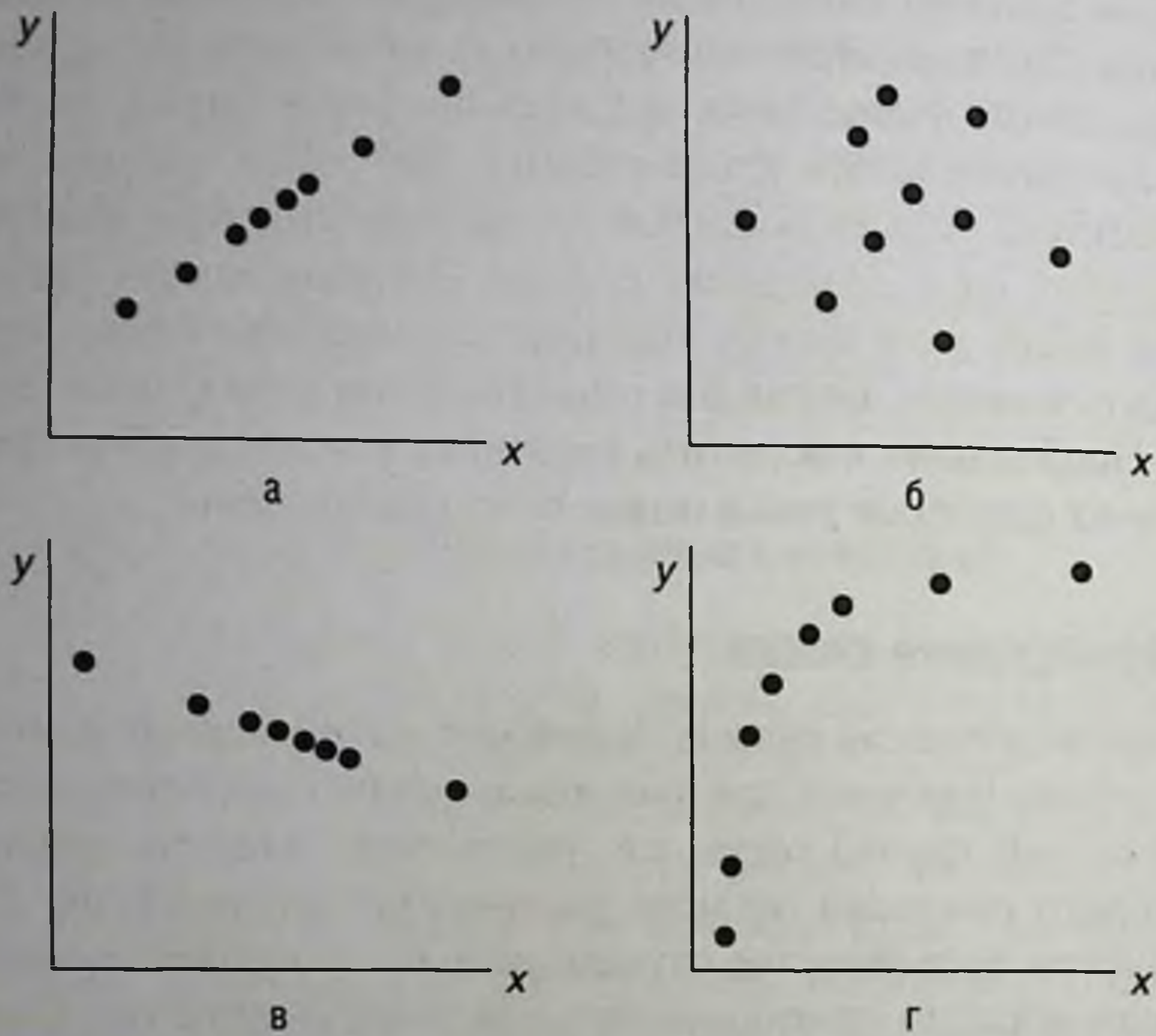


Рис. 2.53. Примеры скаттерграмм. Пояснения в тексте

Обычно связь *оценивают* качественно:

- ▶ от  $-0,25$  до  $0$  (и от  $0$  до  $+0,25$ ) как отсутствие ее или слабую;
- ▶ от  $0,26$  до  $0,50$  (и от  $-0,26$  до  $-0,50$ ) как умеренную;
- ▶ от  $0,51$  до  $0,75$  (и от  $-0,51$  до  $-0,75$ ) как среднюю (хорошую);
- ▶  $r$  с абсолютной величиной более  $0,75$  как тесную.

Квадрат коэффициента корреляции ( $r^2$ ) называют *коэффициентом детерминации*. Он отражает долю дисперсии признака, обусловленную (детерминированную) изучаемой связью. Эту долю можно выразить в процентах ( $r^2 \times 100\%$ ). Если корреляция  $x$  и  $y$  равняется  $0,5$ , то  $25\%$  дисперсии признака  $y$  связаны с дисперсией  $x$ .

Коэффициент корреляции применим к качественным переменным и легко рассчитывается по данным, сведенным в четырехпольную таблицу (табл. 2.26).

$$r = (A \times D - B \times C) / \sqrt{(A + C) \times (B + D) \times (A + B) \times (C + D)}.$$

Иными словами, коэффициент корреляции для качественных признаков равен разности между произведениями частот в ячейках с со-



впадением признаков и в ячейках с расхождением, деленной на корень квадратный из произведения крайних сумм. В нашем примере:

$$(16 \times 24 - 8 \times 7) / (23 \times 32 \times 24 \times 31)^{1/2} = \\ = (384 - 56) / 547584^{1/2} = 328 / 740 = 0,44.$$

Таблица 2.26. Пример расчета корреляции двух качественных признаков

Мокнутие	Высыпания		Всего
	Есть	Нет	
Есть	16 (A)	8 (B)	24 (A + B)
Нет	7 (C)	24 (D)	31 (C + D)
Всего	23 (A + C)	32 (B + D)	55

Обычно коэффициент корреляции используют для оценки связи количественных переменных. Его не следует применять к оценке связи количественных признаков в следующих случаях:

- ▶ дисперсии у двух изучаемых переменных заметно различаются;
- ▶ распределения несимметричны;
- ▶ связь носит нелинейный характер (см. рис. 2.63, г);
- ▶ данные представляют собой «обрезанные» распределения, например, в случае изучения корреляции психологических характеристик с массой тела в выборке лиц, в которую не входят тощие (с массой тела <80% «нормальной»);
- ▶ данные составлены из двух или более подгрупп, известных различием по одному из изучаемых признаков или по обоим; например, будет неверно объединять мужчин и женщин в одну группу в случае, если одним из изучаемых признаков является концентрация в крови тестостерона;
- ▶ изучаемые признаки являются ординальными;
- ▶ в выборке присутствуют повторные измерения, выполненные у тех же субъектов, например, при других обстоятельствах или в другое время.

Для случаев ненормального распределения, нелинейной связи и ординальных данных более адекватен коэффициент корреляции рангов Спирмена (обозначается  $\rho - \rho_0$  греческая). Существуют и другие, более



редко используемые индексы связи. Выбор статистического критерия связи важен, так как использование другого метода приносит другие результаты. Например, в случае, приведенном на рис. 2.63, г, коэффициент Спирмена будет равен точно 1,0, а пирсоновский — значительно  $< 1$ . Правда, если исходные данные заменить их рангами и считать пирсоновский коэффициент для рангов, то он будет иметь ту же величину, что и спирменовский. Интерпретация величины коэффициента Спирмена такая же, как и коэффициента Пирсона.

В отчете об исследовании связь двух признаков желательно демонстрировать скаттерграммами (см. рис. 2.64). Рассмотрение исходных данных, подобных приведенным в табл. 2.26 или на скаттерграмме, позволяет лучше оценить связь, чем только с помощью коэффициента корреляции.

В некоторых случаях исследователи для того, чтобы связь двух признаков была более наглядна, избегают представления скаттерграммы со всеми экспериментальными точками, подобной приведенной на рис. 2.64. Вместо этого они приводят средние величины для подгрупп пациентов, как это сделано на рис. 2.65 для нашего примера с артериальным давлением (см. табл. 1.7). Этим способом обычно удается ясно продемонстрировать направление и форму зависимости, но истинная сила связи, отражающаяся в облаке точек на скаттерграмме, при этом скрыта от читателя отчета. Вычислять коэффициент корреляции по таким «усредненным данным» будет ошибкой, его величина будет резко завышена. В нашем примере по «объединенным точкам»  $r = 0,974$ , что резко отличается от истинного значения.

В некоторых исследованиях применяют метод так называемой *стратифицированной выборки*. Это означает, что из популяции выбирают субъектов не просто случайным образом, а увеличивают долю в выборке тех лиц, которые обладают интересующими исследователя характеристиками, обычно экстремальными и редкими. Например, при долговременном изучении особенностей питания у лиц с гипер- и гиполиппротеидемией, можно резко сократить расходы, если уменьшить число изучаемых субъектов с нормолипидемией. Это может быть сделано, если из лиц с «нормальной» концентрацией липопротеидов высокой плотности выбирать для исследования случайным образом 1/4, а из лиц с высокими и низкими концентрациями — 1/2. В итоге сравнивать полученные три группы между собой можно будет с помощью дисперсионного анализа,



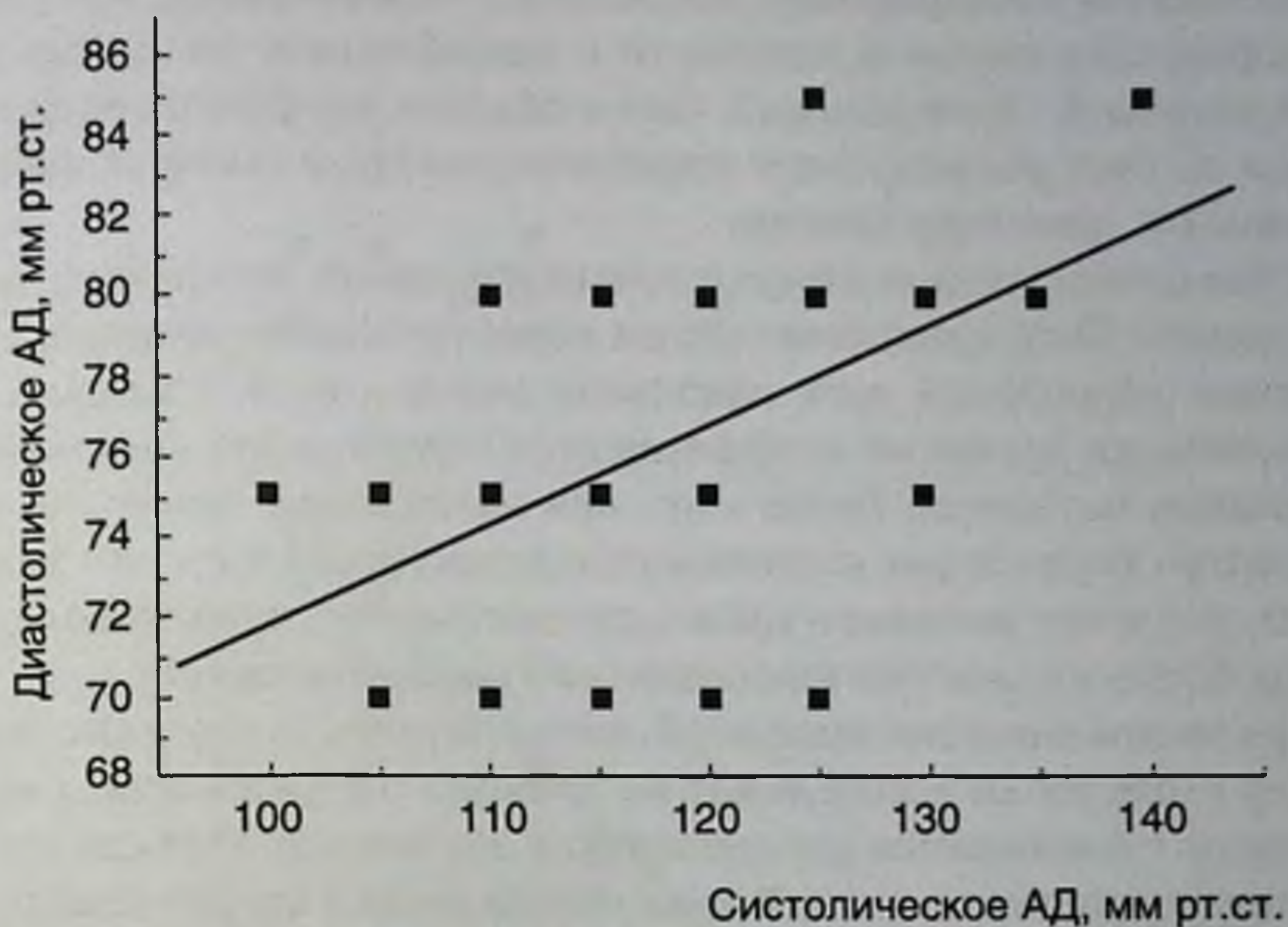


Рис. 2.64. Скаттерграмма для примера оценки корреляции между систолическим и диастолическим артериальным давлением в малой выборке.  $r = 0,535$ ,  $n = 23$ , прямая линия — линия регрессии (см. с. 300–306). Данные табл. 1.7

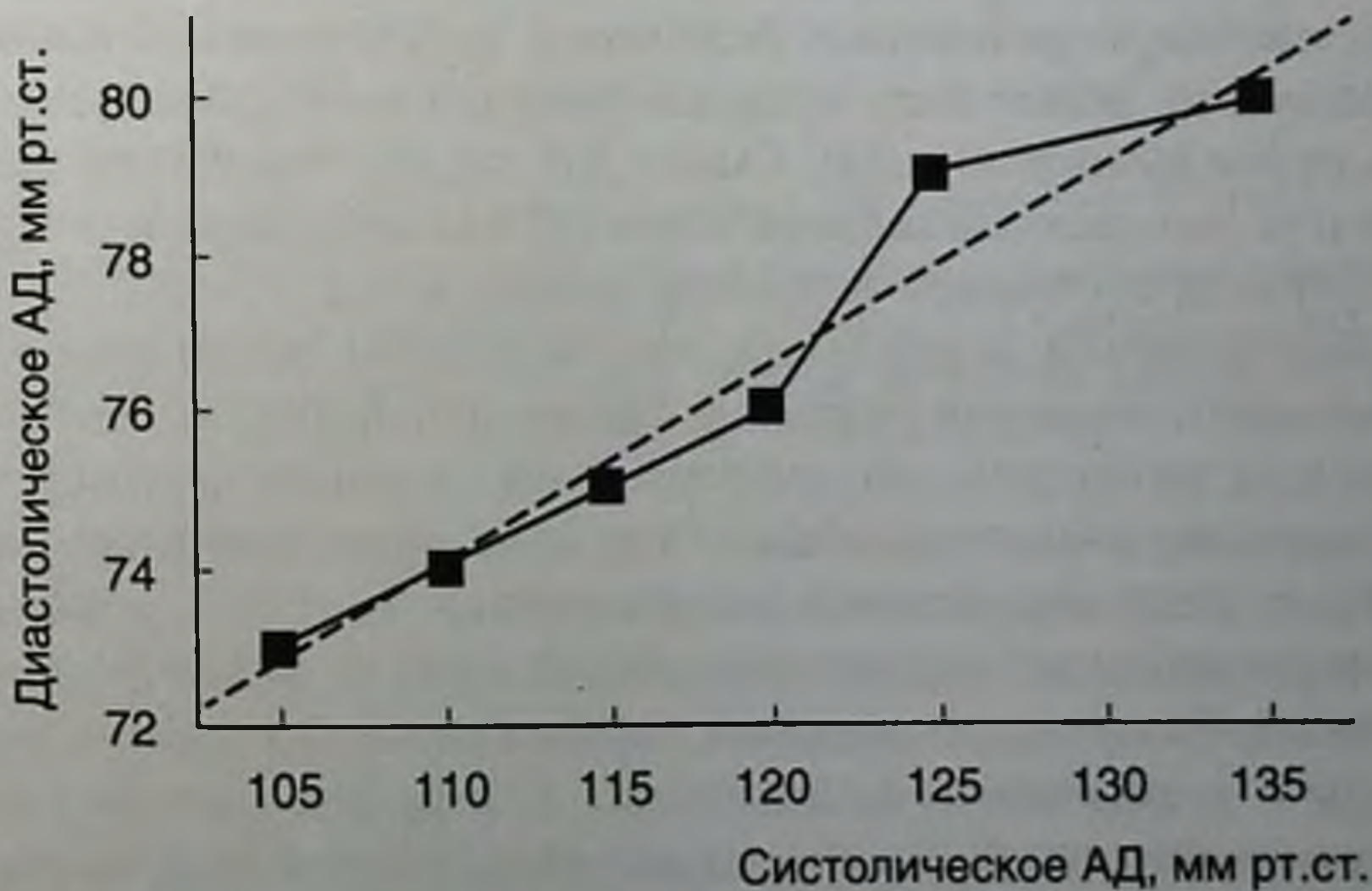


Рис. 2.65. Средние величины диастолического артериального давления в группах лиц с разным систолическим давлением в малой выборке. Прямая линия — линия регрессии (по данным табл. 1.7)



но вычислять коэффициент корреляции концентрации холестерина липопротеидов высокой плотности с потреблением животных жиров будет ошибкой. Вычисленный таким образом коэффициент будет завышен за счет увеличенного представительства в выборке субъектов с крайними характеристиками.

В том случае когда исследуют интеркорреляции нескольких признаков, должна быть приведена полная корреляционная матрица, т.е. все значения корреляций всех признаков между собой. Указание лишь статистически значимых коэффициентов ограничивает понимание исследования читателем. Более того, при вычислении множества коэффициентов корреляции увеличивается вероятность того, что какой-то из них достигнет условного уровня статистической значимости случайно (проблема множественных сравнений рассматривается в разд. 2.5.6).

При малом числе исследований легко получить ложную высокую величину  $r$  (две точки всегда лежат на прямой!). Поэтому всегда помимо величины  $r$  оценивается статистическая значимость корреляции. При большом числе исследований даже весьма низкая корреляция (например,  $r = 0,12$ ) может быть показана с высокой статистической значимостью (например, положительный знак статистической связи будет показан со статистической значимостью  $P < 0,01$ ). Однако эта значимость ничего не говорит о *силе связи*. Статистическая сила связи оценивается именно по величине  $r$ . Величина  $r$ , полученная при выборочном исследовании, может быть экстраполирована на генеральную совокупность путем вычисления ДИ. Смысл ДИ тот же, что для средней величины: с установленной вероятностью (95%) корреляция в генеральной совокупности составляет, например, от 0,01 до 0,23.

Коэффициенты корреляции предназначены для изучения связи **независимых** измерений (у разных субъектов). Коэффициент корреляции нельзя вычислять для сопоставления результатов измерения одного признака двумя способами. Это называется изучением согласия и требует других статистических приемов (см. разд. 2.2.2). Коэффициенты корреляции не следует рассчитывать для результатов повторных измерений. Часто встречающаяся ошибка — сопоставление исходных величин с результатами последующих измерений того же признака. При этом возникают большие искажения вследствие «регрессии к средней», совместно ведущие к выраженной отрицательной корреляции исходной величины и величины различия между двумя измерениями [8]. При эффективном терапевтическом вмешательстве измеряемый при-



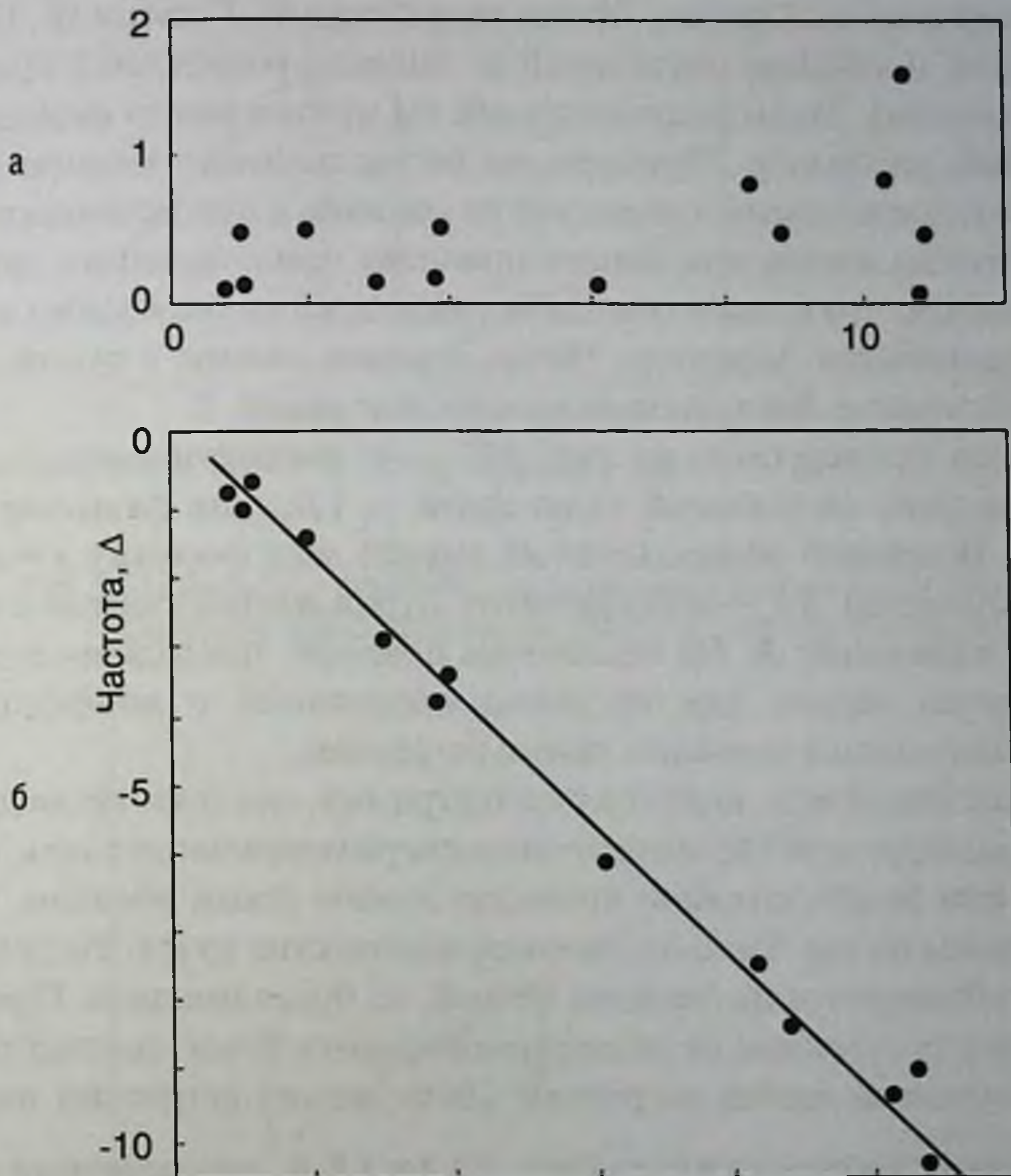


Рис. 2.66. Изменения частоты желудочковых экстрасистол при лечении этмозином. Абсцисса — исходная частота  $л \times 10^2 \text{ час}^{-1}$ . Ордината: а — конечная частота, б — разность между конечной и исходной частотой (по данным Лякишева А. и др. Терапевтический архив. 1979. № 8. С. 63–9)

знак у всех пациентов приходит к почти одинаковым («нормальным») величинам (рис. 2.66). То, что у лиц с исходно большой выраженностью симптома происходит большое изменение, а у лиц с малой выраженностью симптома — малое изменение, не является связью и не может быть описано как корреляция.

### Регрессия

Оценка коэффициента корреляции в большинстве случаев недостаточна для количественных выводов. Правильный способ описания



связи — *регрессионный анализ*. Метод разработал Ф. Гальтон (F. Galton) для изучения «возврата» роста детей от высоких родителей к среднему росту (регрессии). Затем разработанный им метод в целом назвали «регрессионным анализом». Применение регрессионного анализа полезно там, где исследователя интересует не наличие и статистическая сила связи, а степень изменения одного признака при изменении другого<sup>1</sup>. Важно помнить, что и такое описание связи тоже не доказывает ее причинно-следственный характер. Метод хорошо описан в руководствах [14, 15] на примерах биомедицинских исследований.

В случаях, приведенных на рис. 2.67 а–в, корреляционный анализ покажет наличие одинаковой силы связи — 1,0, хотя различие связи очевидно. Линейный регрессионный анализ даст формулу  $y = a + bx$ , где  $a$  — постоянная, а  $b$  — коэффициент, отражающий степень изменения  $Y$  при изменении  $X$ . На численном примере, приведенном на рисунке, хорошо видно, как величина постоянной и коэффициента влияют на положение и наклон линии регрессии.

Когда исследователь, получив скаттерграмму, «на глаз» проводит линию, проходящую через большинство экспериментальных точек, он сознательно или бессознательно проводит линию таким образом, чтобы по обе стороны от нее было одинаковое количество точек. Регрессионный анализ базируется на сходном методе, но более мощном. Программа учитывает расстояние от экспериментальных точек скаттерграммы до предполагаемой линии регрессии. Далее линия регрессии выбира-

<sup>1</sup> В отличие от проверки гипотез (разд. 1.1.2 и 1.8.3), регрессионный анализ относится к методам математического моделирования. Моделированием называют абстрактное представление взаимоотношений между логическими или эмпирическими компонентами системы. Это представление может быть количественным (формализованным), например кривая выживания или линейная регрессия. Моделирование может быть выполнено в эксперименте на животных или на примере воображаемой популяции с использованием закономерностей, изученных в прошлых эпидемиях. Так, В. Фарр еще в 1840 г. рассчитал предполагаемую смертность от оспы в будущей эпидемии.

Весь статистический анализ базируется на моделировании. Так, нормальное распределение (разд. 1.6.3) используется как модель реальных распределений. В модели различают закон, выражаемый формулой, и параметры модели (в случае нормального распределения это средняя величина, среднее квадратическое отклонение). Модели, которые не допускают отклонений предсказываемой (зависимой) величины от значения, вычисленного по значению независимых величин, называют детерминистскими. Модели, предусматривающие наличие отклонений (ошибки предсказания), называют вероятностными, или стохастическими.



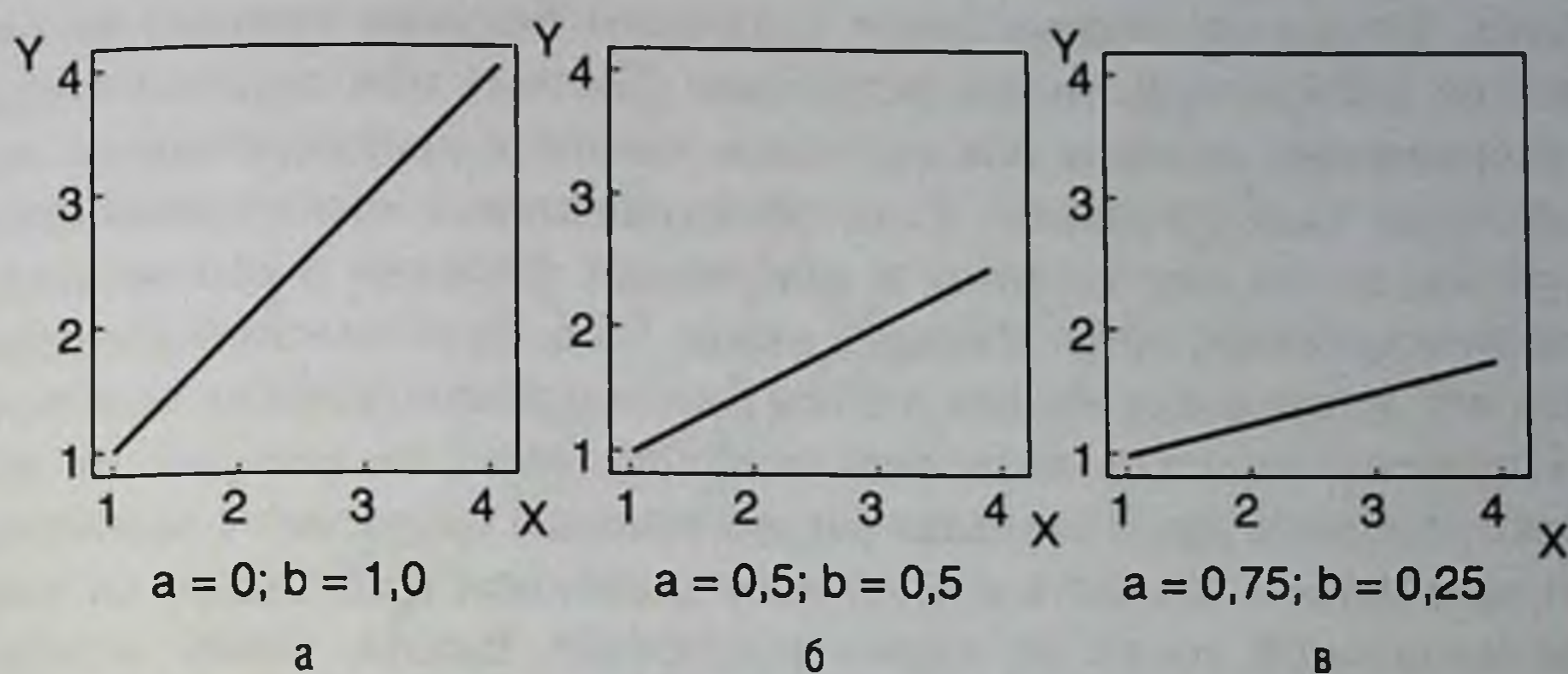


Рис. 2.67. Результаты линейного регрессионного анализа в трех случаях, когда коэффициент корреляции равен 1.0. Пояснения в тексте

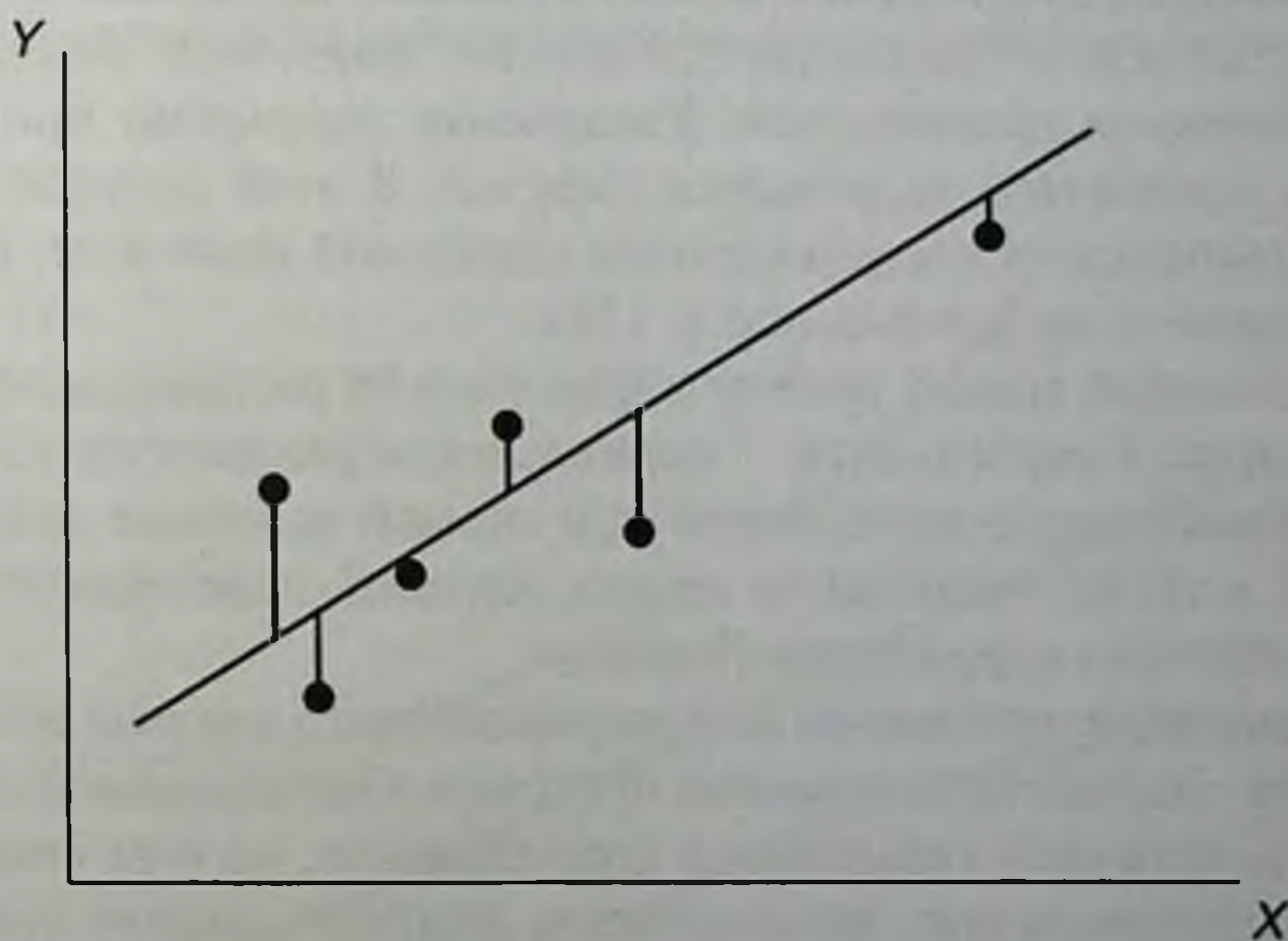


Рис. 2.68. Расстояния от экспериментальных точек до линии регрессии, минимизируемые в регрессионном анализе

ется так, чтобы сумма расстояний экспериментальных точек до линии была наименьшей (рис. 2.68). Самый распространенный метод — метод *наименьших квадратов* (англ. *least squares method*), используемый в современных компьютерных программах, выбирает линию регрессии так, чтобы наименьшей была сумма квадратов расстояний до экспериментальных точек. Для понимания результатов анализа важно, что влияние каждой точки увеличивается пропорционально квадрату расстояния.



Значит, далеко отстоящие точки оказывают большое влияние на положение избираемой линии регрессии. Поэтому при использовании регрессионного анализа обязательным является предварительное исследование скаттерграммы. Если обнаруживаются «выскакивающие» значения, то их анализируют и принимают решения в соответствии с рекомендациями, приведенными в разд. 1.5.6. Если «выскакивающих» точек нет, то проводят анализ, а затем рассматривают графики остатков.

Графиком остатков называют скаттерграммы, на которых по абсциссе отложена предсказанная регрессионным уравнением величина, а по ординате — разность с истинным значением (расстояние от экспериментальной точки до линии регрессии). Важно, чтобы остатки приблизительно симметрично располагались вокруг линии регрессии. В случае если остатки разбросаны широко у одного из флангов, можно сделать вывод о нарушении важного условия применимости регрессионного анализа — равенства дисперсий зависимой величины при разных значениях независимой. Увеличение дисперсии при высоких значениях предиктора встречается нередко. В этой ситуации следует изучить возможности трансформации зависимой величины, например использование ее логарифма (см. с. 136).

Регрессионный анализ требует нормальности распределения изучаемых признаков, в особенности — нормальности распределения предсказываемой (зависимой) переменной при каждой величине независимой (рис. 2.69), а также соблюдения других условий, перечисленных выше для коэффициента корреляции Пирсона.

Компьютерные программы для регрессионного анализа обязательно входят во все пакеты статистических программ. Они оценивают не только постоянную величину уравнения и коэффициент, но и их стандартные ошибки. Соответственно, исследователь получает оценку статистической значимости элементов уравнения регрессии. После этого можно построить ДИ (рис. 2.70). Этот интервал соответствует не области, в которой распределены экспериментальные значения, а диапазону, в котором с установленной вероятностью (например, 95%) располагается линия регрессии. На рис. 2.70 два возможных положения популяционной (истинной) линии регрессии обозначены пунктиром. ДИ, полученный на материале выборки, дает представление о том, в каких пределах лежит линия регрессии в популяции. В публикации результатов регрессионного анализа желательно видеть скаттерграмму фактических данных вместе с наложенной на них линией регрессии.



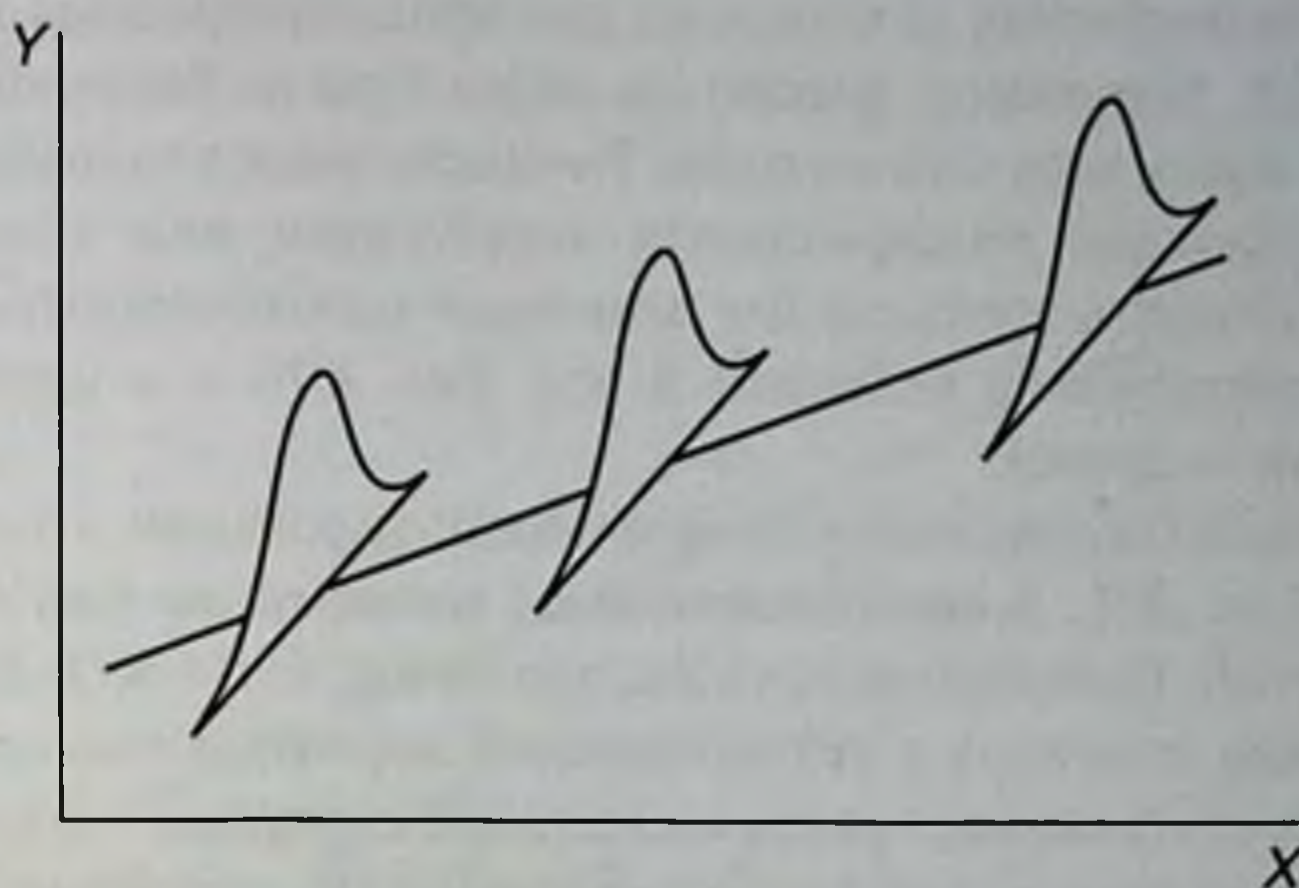


Рис. 2.69. Регрессионный анализ дает правильные результаты, если при разных величинах независимой переменной  $X$ -зависимая  $Y$  имеет нормальное распределение

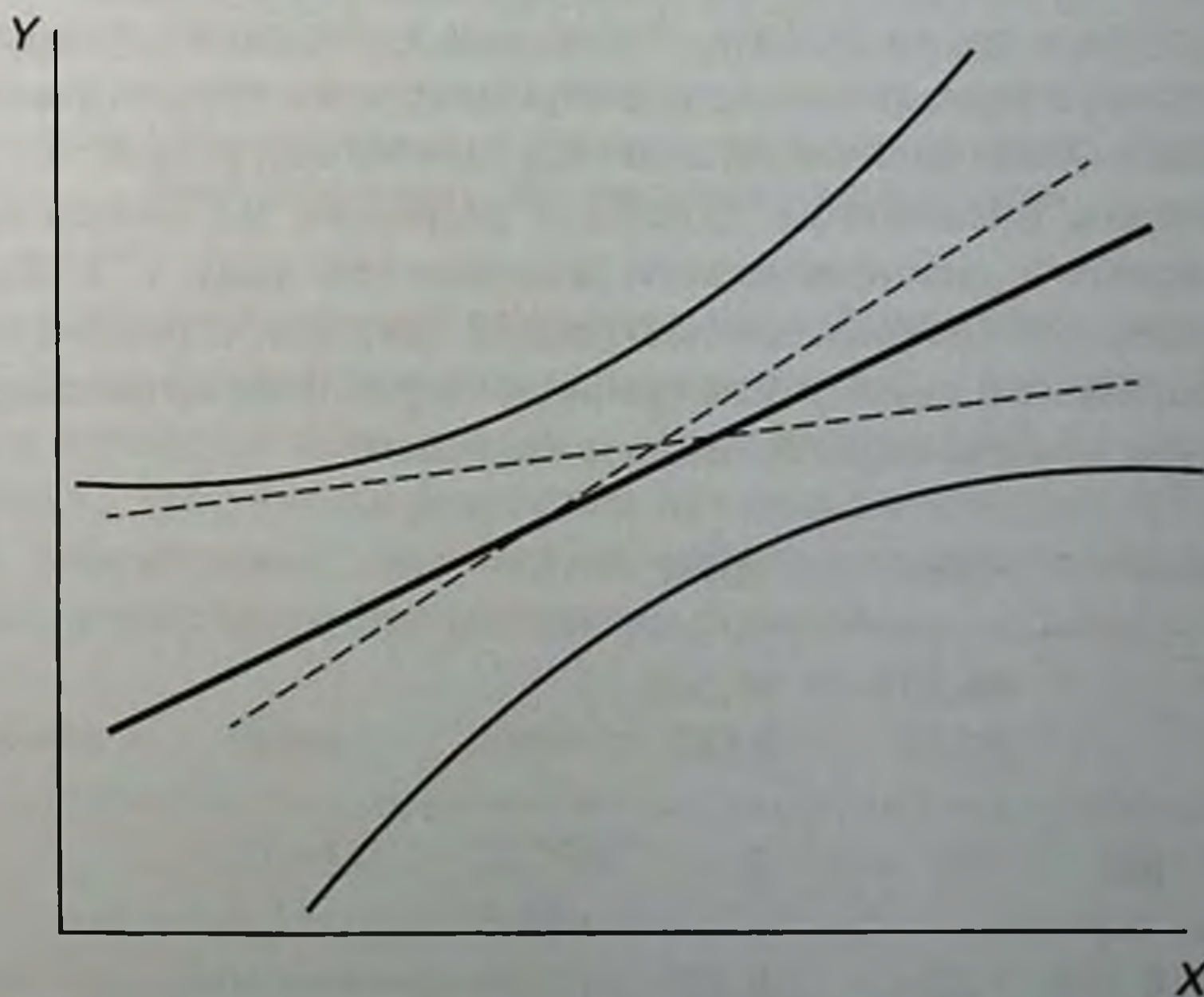


Рис. 2.70. Результаты линейного регрессионного анализа данных, представленных в виде скаттерграммы. Толстая линия — оценочная линия регрессии; изогнутые линии — границы доверительного интервала; пунктир — возможные истинные линии регрессии



Уравнения регрессии используют для прогнозирования величины  $Y$  по величине  $X$ . Например, можно по массе тела до беременности прогнозировать массу тела после родов. Точность такого прогноза не определяется величиной коэффициента корреляции или коэффициента регрессии. Точность прогноза для популяции можно оценить по ДИ для  $Y$  при соответствующей величине  $X$  (см. рис. 2.70 — в центре он уже, на периферии — шире).

Если речь идет об индивидуальном прогнозировании, то его точность описывается не ДИ, а *предсказательным интервалом* (англ. *prediction, tolerance interval*). Разница здесь та же, что между  $m$  и СКО: ДИ предсказывает, в каких границах с установленной вероятностью лежит линия регрессии для популяции; предсказательный интервал — в каких границах варьирует величина зависимого признака ( $Y$ ) при данной величине независимого признака ( $X$ ). Предсказательный интервал соответствует обычно 95% вероятности индивидуального прогноза.

Результат регрессионного анализа годится для прогнозирования  $Y$  по  $X$ , но не наоборот — для этого нужен обратный анализ. Например, если предполагается по частоте хромосомных aberrаций оценивать полученную дозу излучения, то нужно строить уравнение регрессии, в котором независимой величиной является частота aberrаций.

Рассмотрим применение линейной регрессии на нашем примере систолического и диастолического давления (см. табл. 1.7). Для получения параметров уравнения и построения графика регрессии воспользуемся программой для научной графики Origin. Выбрав систолическое давление как независимую величину, получаем:

<i>Parameter</i>	<i>Value</i>	<i>Error</i>	
A	46,373	10,358	
B	0,253	0,087	
R	SD	N	P
0,535	4,168	23	0,009

На такой распечатке результатов анализа обозначения хорошо понятны: параметры (*Parameter*) A и B — постоянная и регрессионный коэффи-



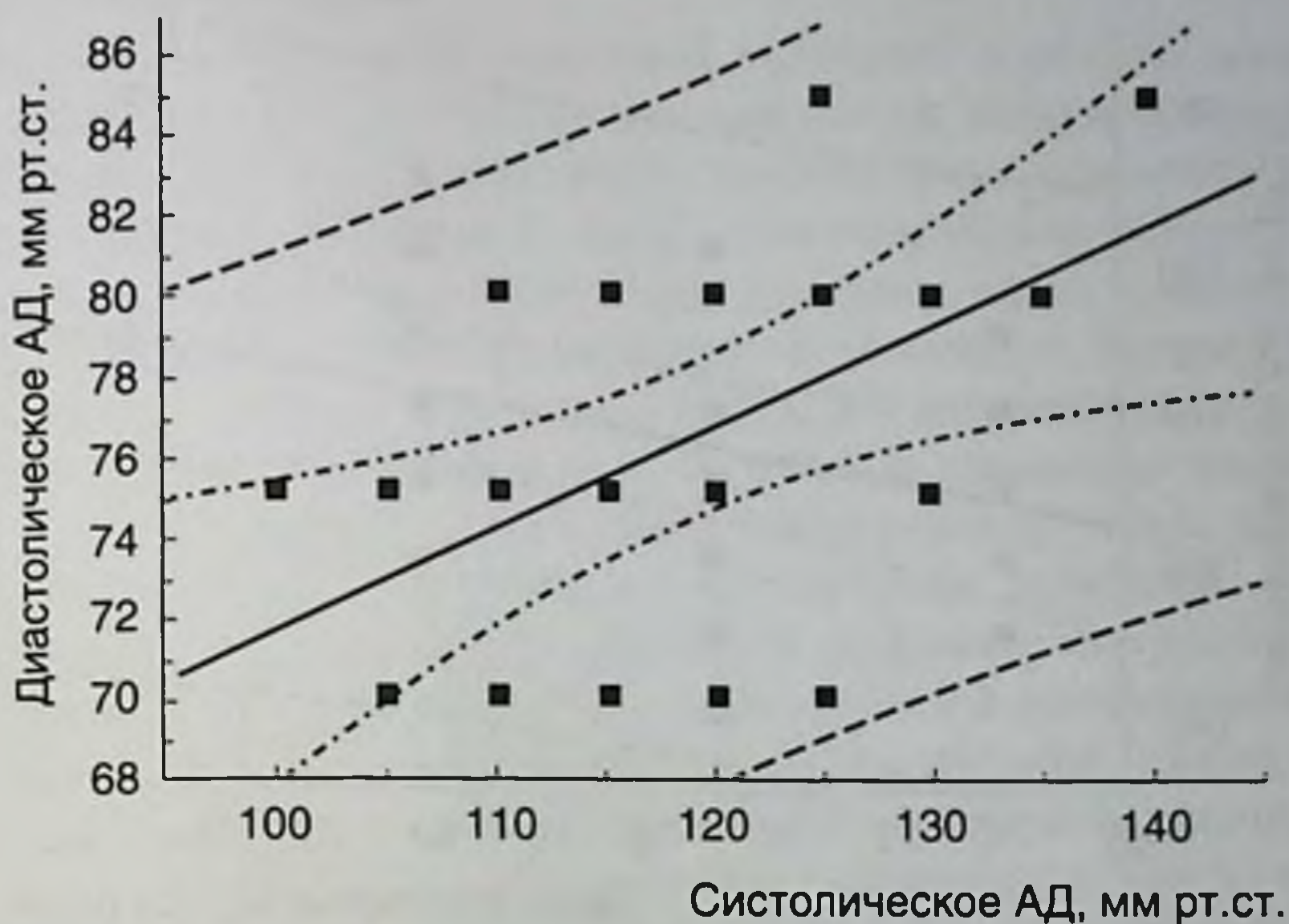


Рис. 2.71. Зависимость диастолического артериального давления от систолического. Некоторым точкам на скаттерграмме соответствует более одного случая. Центральная линия — линия регрессии, параметры которой приведены в тексте. Точечная линия ограничивает 95% доверительный интервал для линии регрессии. Пунктирная линия ограничивает 95% предсказательный интервал

циент; в столбике *Value* — их величины, в столбике *Error* — стандартные ошибки. Ниже *R* — коэффициент корреляции, *SD* — стандартное отклонение для зависимой величины, *N* — число наблюдений, *P* — величина *P* (альфа-риск). Графически результаты представлены на рис. 2.71.

Если независимой переменной избирается диастолическое АД, то мы получаем следующие параметры регрессии:

<i>Parameter</i>	<i>Value</i>	<i>Error</i>
A	32,053	29,784
B	1,13	0,39
<i>R</i>	<i>SD</i>	<i>N</i>
0,535	8,806	23
		<i>P</i>
		0,009



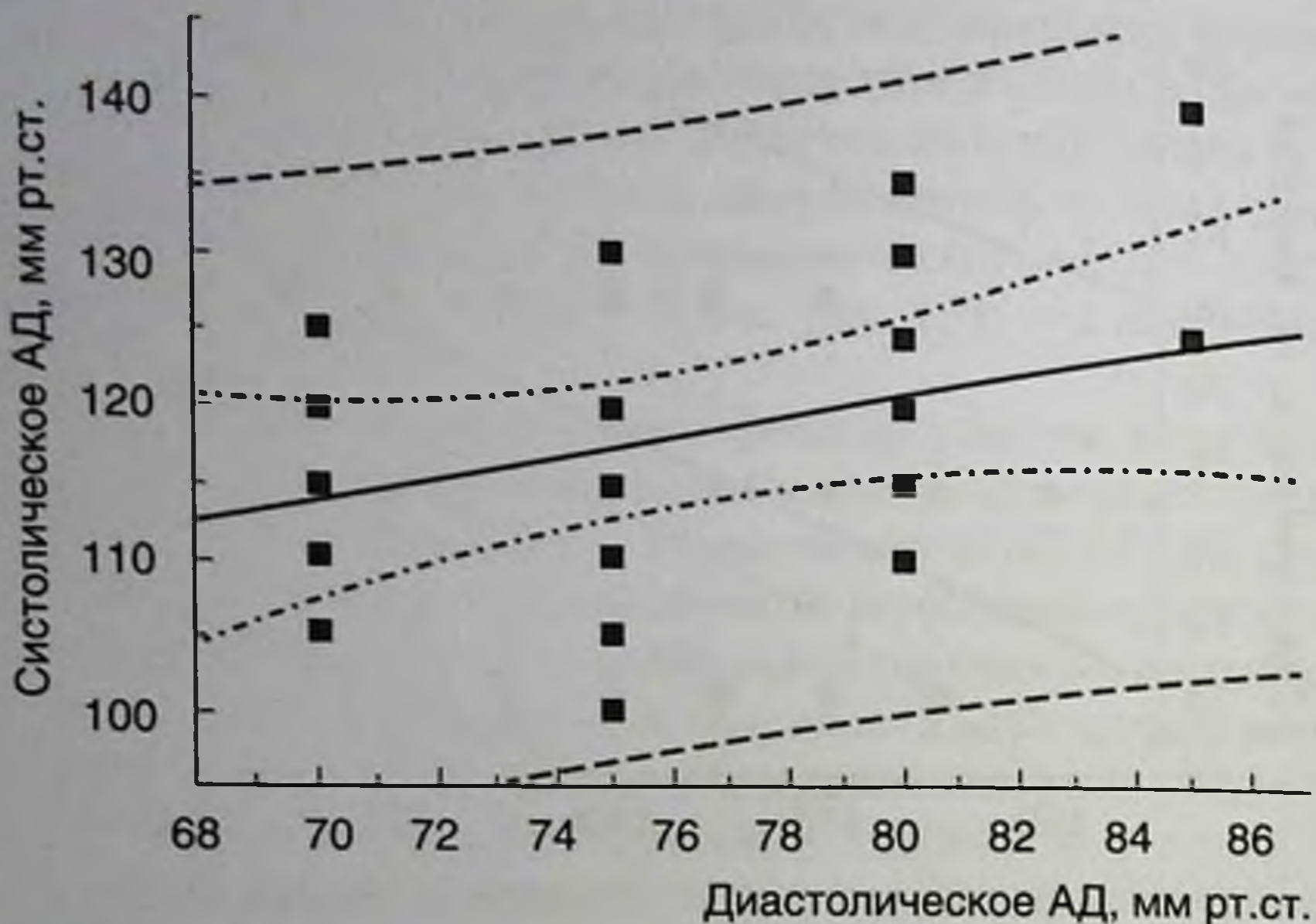


Рис. 2.72. Зависимость систолического артериального давления от диастолического. Некоторым точкам на скаттерграмме соответствует более одного случая. Обозначения как на рис. 2.71

Обратите внимание: коэффициент корреляции и величина  $R$  такие же (эти параметры «симметричны»), а параметры уравнения регрессии — иные.

На двух рисунках (2.71 и 2.72) видно, что положение линий регрессии различается. Сравните, какое диастолическое АД соответствует систолическому 100 мм рт.ст. на линии регрессии первого графика, и какому диастолическому АД соответствует систолическое 100 мм рт.ст. на втором. Если бы «точность измерений» в нашем примере была больше, то мы могли бы получить более точную оценку связи, более узкие доверительные интервалы для линий регрессии и предсказательные интервалы. Использованное в данном примере «измерение» давления с точностью до 5 мм рт.ст. привело к низкой точности модели.

В случае когда наклон линии регрессии увеличивается с увеличением независимого признака, это означает наличие более чем линейной зависимости. Обычно для описания такой зависимости используется квадратическая зависимость:  $y = a + bx + cx^2$ . В таком уравнении к обычной линейной зависимости добавляется квадрат независимой величины с коэффициентом  $c$ .



## Множественная регрессия

В исследовании связей зависимой величины с предикторами нередко возникает задача изучить одновременное действие нескольких предикторов. Для этого используется техника множественного регрессионного анализа. Например, масса тела подростков находится под влиянием множества факторов, таких как рост, возраст, доход и образование родителей, занятия спортом, калорийность питания. Путем использования *индекса массы тела* (см. с. 88) исследователь устраняет влияние роста, поскольку этот очевидный предиктор массы тела не представляет интереса для данного исследования. После этого исследуется зависимость от дохода и образования родителей, занятий спортом, пола, калорийности питания. В результате получается уравнение вида:  $y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4$ , где  $x$  и  $b$  соответственно независимые признаки и регрессионные коэффициенты. Такую линию регрессии невозможно изобразить на плоскости бумаги, а вычисления были почти недостижимы до появления компьютеров. После получения уравнения регрессии исследуют графики остатков для каждой независимой величины.

В уравнении множественной регрессии могут быть использованы производные от независимых переменных, например их квадраты. Поскольку исследователь может предполагать, что предикторы могут взаимодействовать между собой, как, например, калорийность питания и занятия спортом, он может использовать в качестве предиктора произведение этих признаков.

В клинических исследованиях нередко для небольшой группы больных бывают доступны десятки признаков, которые исследователь хотел бы проанализировать. Этого обычно не следует делать. Почему? Представим, что у нас есть признаки, каждый из которых может принимать целое значение 1, 2 или 3. При исследовании 10 пациентов по одному такому признаку больше половины пациентов будут иметь не уникальные значения. Если исследовать два таких признака, то будет  $3 \times 3 = 9$  вариантов сочетания результатов. Возможно, что только в одной «ячейке» окажутся два случая (пациента). Если исследовать три таких признака, то в 27 ячейках ( $3 \times 3 \times 3$ ) большая их часть окажутся пустыми. При дальнейшем увеличении числа признаков оказывается, что все случаи занимают уникальные ячейки. Возникает та же ситуация, что и с вычислением коэффициента корреляции для двух наблюдений: он равен 1. Поэтому применение метода множественной регрессии воз-



можно только в том случае, если число случаев превышает число изучаемых признаков многократно, как минимум в 10 раз, лучше — если в десятки раз.

При изучении множества предикторов возникает очевидная задача установить, какие из них оказывают существенное влияние на зависимую величину, а какие — нет. Решить эту задачу, рассматривая полученное уравнение множественной регрессии, нельзя. Дело в том, что в присутствии независимой переменной  $x_1$  регрессионный коэффициент для переменной  $x_2$  имеет иное значение, чем в уравнении, где  $x_1$  отсутствует. Поэтому ни величина коэффициента, ни его ДИ не могут служить ориентиром для сравнения вклада независимых переменных. Для решения этой задачи используются статистические приемы последовательного (пошагового) включения и исключения переменных. Например, сначала в уравнение включается та переменная, которая имеет наибольшую корреляцию с зависимой переменной. После включения первой переменной оценивается корреляция остатков с другими переменными, и второй кандидат на включение — переменная с максимальной корреляцией. Процесс повторяется до тех пор, пока оставшиеся переменные имеют корреляцию выше установленной исследователем величины. В качестве критерия для включения переменной может использоваться не только коэффициент корреляции, но и F-критерий.

Технология перебора признаков может быть сложнее. Например, после того как уравнение составлено методом включения, может начинаться процесс исключения переменных, дающих самый малый вклад в остаточную дисперсию зависимой переменной. Возможен еще более длительный процесс перебора признаков. Эти методы множественного регрессионного анализа доступны во всех «больших» пакетах статистических программ (SAS, Statistica, BMDP, SPSS, EpiInfo, Systat и др.).

Самое частое применение множественного регрессионного анализа — выяснение независимости предикторов. При изучении множества предикторов всегда встает вопрос: является ли изучаемый предиктор, например концентрация лейкоцитов в крови, независимым признаком утолщения выстилки сонной артерии? Возможно иное: связь сосудистых изменений с концентрацией лейкоцитов целиком объясняется курением (известно, что курение ведет к возрастанию концентрации лейкоцитов и сопровождается сосудистыми изменениями). Если оказывается, что связь изучаемого предиктора с изменениями сосудов сохраняется после статистического устранения влияния курения, то можно пола-



гать, что изучаемый предиктор независим. Уравнения множественной регрессии трудно интерпретировать. Поскольку точность измерений всегда ограничена, а в медицине чаще всего — невелика, никогда нельзя исключить, что включенные признаки коррелируют между собою и не являются действительно независимыми предикторами.

### Логистическая регрессия

Моделирование зависимости вероятности качественных событий от количественных признаков, например инциденса болезней от величины фактора риска (рис. 1.5 и 2.54), требует применения специальных методов. Необходимость в них вытекает из следующего. Если использовать для описания зависимости линейную регрессию, то смерть кодируется как 1, выживание — 0. В уравнении регрессии вычисляется доля выживающих (вероятность выживания). При равной численности пациентов в центре ряда вероятностей — от 0,2 до 0,8 — неопределенность оценки примерно одинакова. На флангах, ближе к 1,0 и к 0,0, неопределенность естественным образом уменьшается, будучи ограничена соответственно единицей и нулем (в группах с высокой или низкой вероятностью исхода). Если же использовать обычную линейную регрессию для оценки такой зависимости, то можно получить значения вероятности за пределами 0 и 1.

Для статистических манипуляций удобнее иметь равную неопределенность зависимой величины (вероятности) во всем диапазоне и прямолинейную зависимость вероятности от изучаемой переменной, например летальности от возраста. Это достигается трансформацией вероятности через *логит-преобразование*:

$$\text{logit}(p) = \log_e [p/(1 - p)],$$

где  $p$  — вероятность события. Логарифм с основанием  $e$  вычисляется от выражения  $p/(1-p)$ , которое называется шансами (см. разд. 1.3.3). Логит-преобразование (рис. 2.73) изменяет вероятность подобно трансформации в соответствии с закономерностями нормального распределения (см. рис. 1.51).

Если обозначить  $\text{logit}(p_1)$  как  $l_1$ , то при сравнении двух групп (1 и 2) разность логитов трансформируется следующим образом:

$$\begin{aligned} l_1 - l_2 &= \text{logit}(p_1) - \text{logit}(p_2) = \log[p_1/(1 - p_1)] - \log[p_2/(1 - p_2)] = \\ &= \log[p_1 \times (1 - p_2)/p_2 \times (1 - p_1)]. \end{aligned}$$



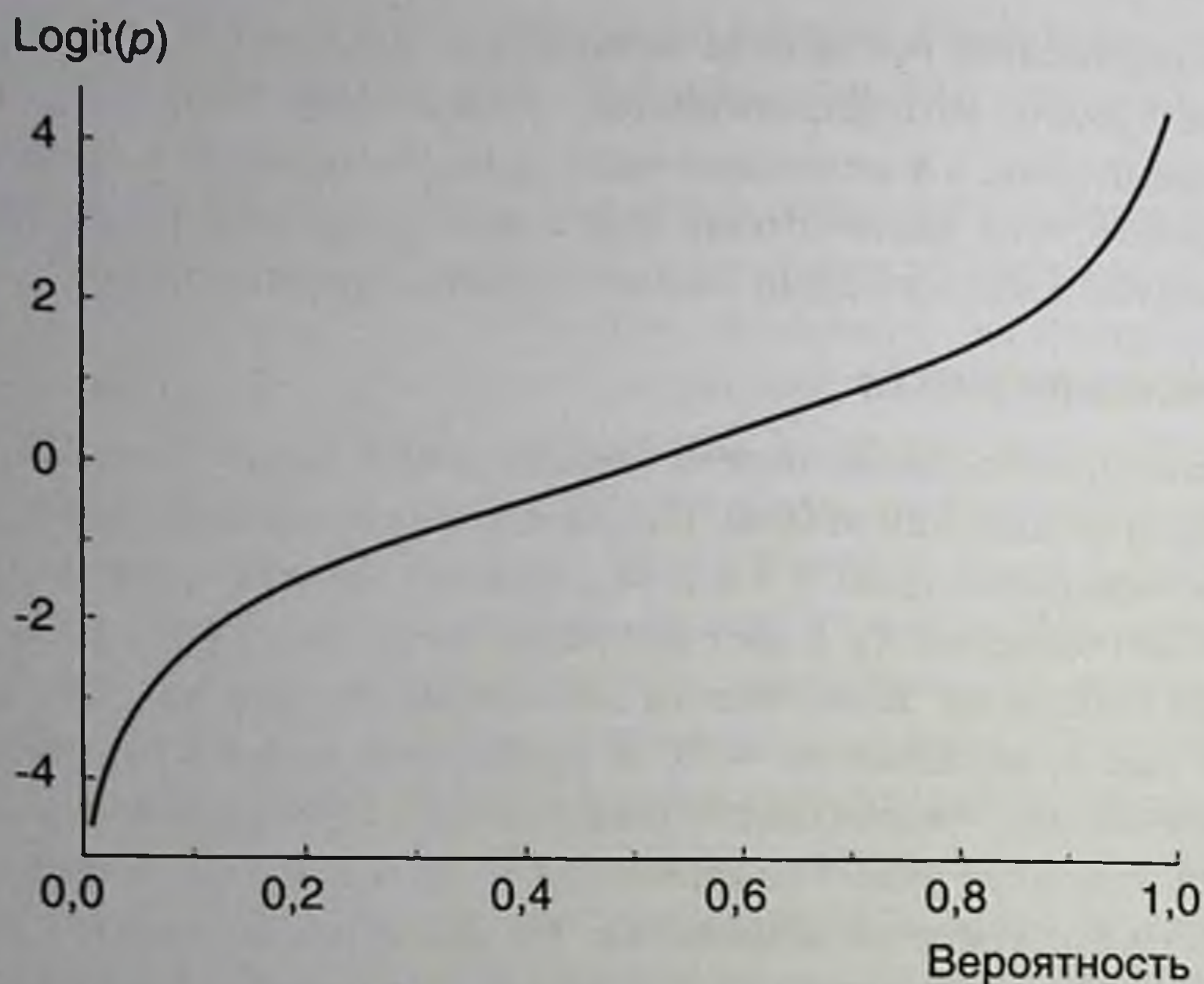


Рис. 2.73. Результат логит-преобразования вероятности

Последнее есть не что иное, как логарифм отношения шансов. Отношение шансов широко используется в эпидемиологических исследованиях (см. с. 287).

Из  $\text{logit}(p)$ , который изменяется от  $-\infty^1$  до  $+\infty$ , можно вычислить  $p$ , изменяющуюся от 0 до 1. Если  $\text{logit}(p) = 1$ , то  $e^1 = p/(1-p)$  и  $p = e^1/(1+e^1)$ .

Собственно инструмент регрессионного анализа применяется к логиту, и получается уравнение вида:

$$\text{logit}(p) = a + b_1x_1 + b_2x_2,$$

где  $a$  — постоянная,  $x_1$  и  $x_2$  — качественные или количественные признаки,  $b_1$  и  $b_2$  — коэффициенты.

Логистическая регрессия включена в наборы инструментов статистических программ. Технология вычислений отличается от применяемой во множественной регрессии. Уравнение логистической регрессии обладает привлекательным свойством:

- регрессионные коэффициенты позволяют получить отношение шансов для зависимой переменной, связанных с данной независимой переменной:  $\text{ОШ}(x_1) = e^{b_1}$ ;

<sup>1</sup>  $\infty$  — символ, применяемый для обозначения бесконечно больших величин.



- компьютерная программа предоставляет нам стандартную ошибку  $se(b)$  для коэффициента  $b$ , и соответственно можно вычислить нижнюю ( $b - 1,96 \times se(b)$ ) и верхнюю ( $b + 1,96 \times se(b)$ ) границы ДИ, которые, в свою очередь, могут быть трансформированы в ДИ для отношения шансов.

### Дискриминантный анализ

Дискриминантный анализ (*discriminatory analysis, discriminant function analysis*, от англ. *discriminate* — различать) является вариантом множественного логистического анализа. В дискриминантном анализе зависимая переменная является качественной (точнее — дискретной). Самый частый вариант применения дискриминантного анализа — разработка уравнения множественной регрессии, зависимая величина в котором будет разделять больных двумя болезнями (здоровых и больных, годных и негодных и т.д.) или людей с разным количеством детей.

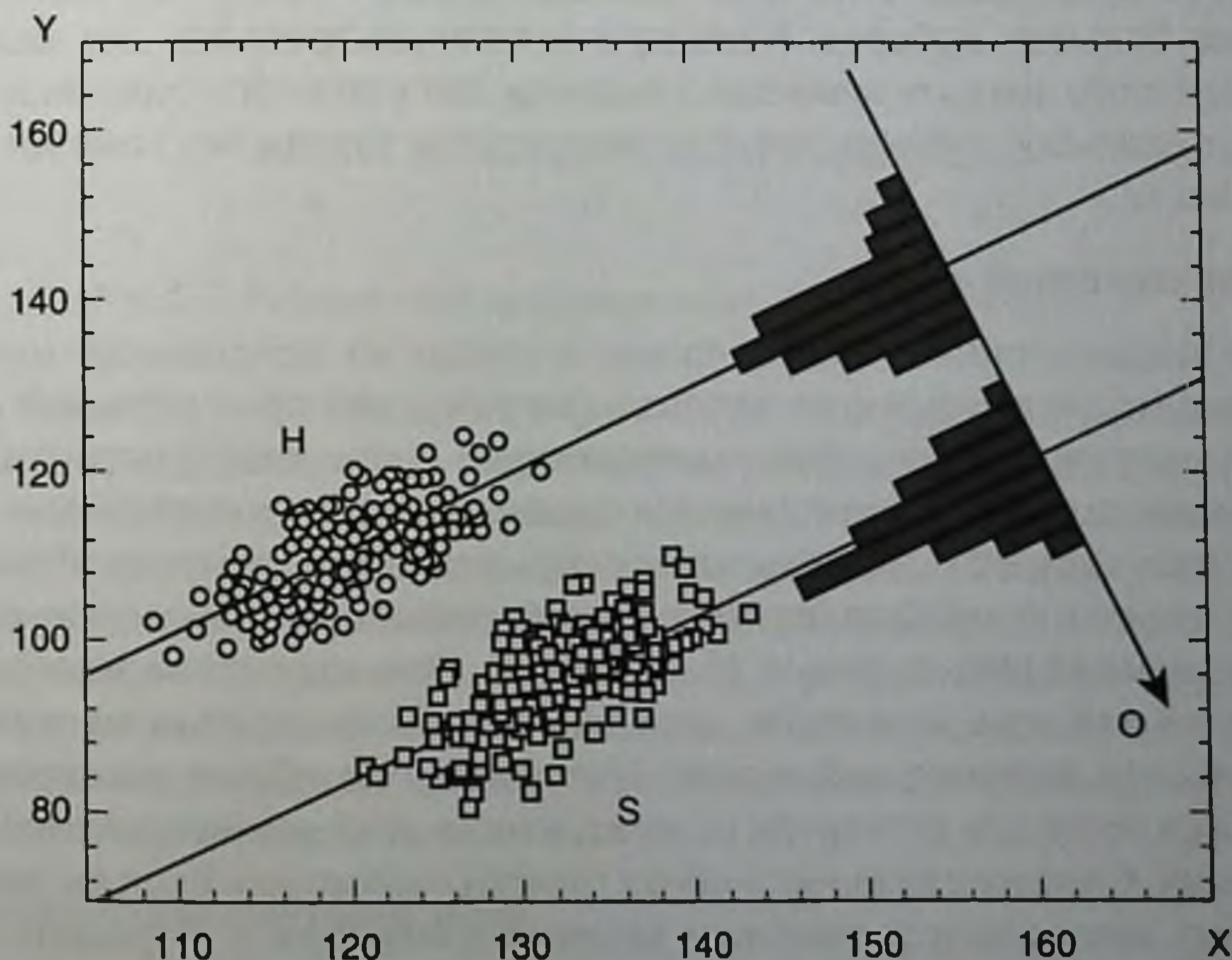


Рис. 2.74. Принцип дискриминантного анализа. В то время как каждый из имеющихся признаков ( $x, y$ ) плохо разделяют группы H и S, методами регрессии создается новая переменная O, лучше различающая изучаемые группы по линии деления AA.

Благодарность В. Леонову



Существует иная форма дискриминантного анализа, в котором возможно разделение более чем двух групп. Этот вариант дискриминантного анализа близок к множественной линейной регрессии. В то время как входящие в уравнение признаки по отдельности не позволяют эффективно разделить эти группы, комбинации отдельных признаков (дискриминантные функции или канонические переменные) могут позволить такое более эффективное разделение (рис. 2.74).

Дискриминантный анализ отступает от ограничений на регрессионный анализ (см. выше), а именно — использует в качестве зависимой переменной не количественный признак, распределенный нормально, а качественный признак, и исходит из предположения о прямолинейной связи зависимого признака и независимых, нормальном распределении всех переменных с равными дисперсиями внутри каждой из групп. Эти отступления и нестрогие предположения основываются на предположении о *робастности* (см. с. 234) регрессионного анализа. Естественно, что это предположение может быть справедливо только при использовании больших выборок. Несмотря на наличие программ для дискриминантного анализа в пакетах программ для статистического анализа, его правильное выполнение и интерпретация трудны без помощи специалиста.

### Ковариационный анализ

Ковариационный анализ является одним из приложений множественного регрессионного анализа. Он позволяет при сравнении двух и более групп между собой учесть влияние привходящих переменных. Классическим примером является сравнение эффектов внешнего воздействия в группах, различающихся по исходному состоянию. Представим, что по исходной величине изучаемого признака две сравниваемые группы различаются (рис. 2.75, абсцисса). При отсутствии изменений средние для этих двух групп должны были бы находиться на верхней косой штрихпунктирной линии. Мы видим, что облака экспериментальных точек для двух групп находятся ниже этой линии исходного состояния. Среднее снижение в обеих группах одинаково. Если бы анализ данных завершался сравнением величин изменений в результате воздействия, то вывод был бы очевиден: воздействие одинаково изменило состояние обеих групп. Регрессионный анализ показывает, что если бы группа с меньшей исходной величиной изменялась под воздействием так же, как группа с большей исходной величиной, то ее изменения



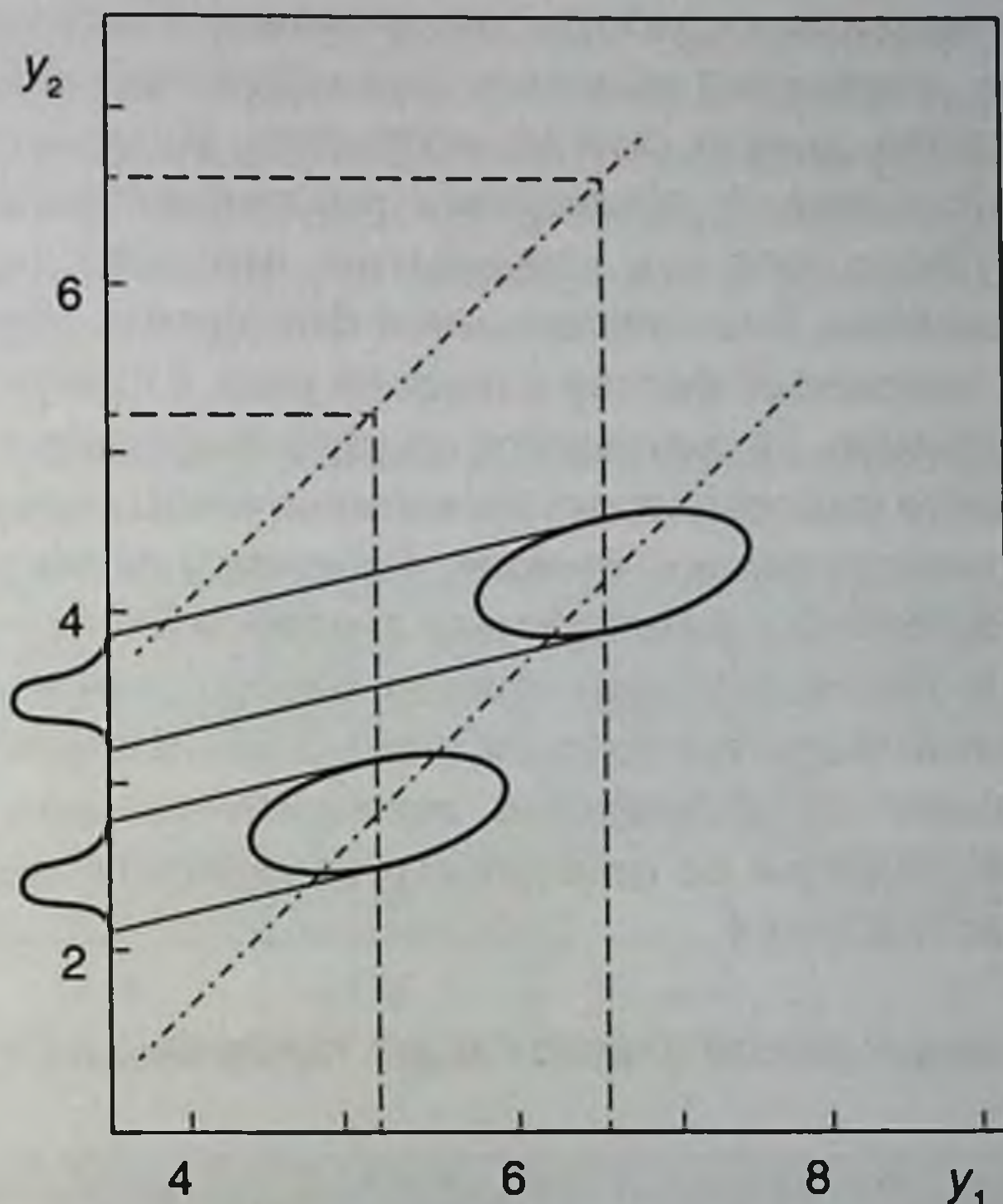


Рис. 2.75. Принцип ковариационного анализа. Пояснения в тексте

были бы почти в 2 раза меньше (рис. 2.75). Это различие выявляется сравнением средних, скорректированных на исходные различия. Графически это выглядит, как сравнение проекций на ординату, выполненных по направлению длинной оси эллипсов данных. Это направление соответствует наклону линии регрессии. Таким же образом может быть учтено влияние и других различий между группами. При увеличении числа признаков, включаемых в модель, и особенно в случае, когда в сравниваемых группах наклон линии регрессии различается, модель труднее интерпретировать.

### Взаимодействие факторов риска

Факторы риска обычно взаимодействуют в возникновении, развитии и проявлении болезни в неизвестных соотношениях. В эпидемиологическом исследовании взаимодействие факторов риска рассматривается количественно. Имеющиеся в распоряжении эпидемиолога инстру-



менты — специальные структуры исследования и методы статистического анализа, позволяют выявлять взаимодействие факторов риска. При рассмотрении данных сначала исследуют, не может ли связь экспозиции с заболеванием объясняться вмешивающимися факторами (разд. 2.7.4). После того как установлено, что связь не объясняется целиком известными вмешивающимися факторами, необходимо рассмотреть, как изучаемый фактор влияет на риск в присутствии третьих факторов. Например, увеличивается ли риск инфицированности ВИЧ при употреблении инъекционных наркотиков одинаково у гетеросексуальных и гомосексуальных субъектов. Если увеличение риска НЕодинаково, то имеет место взаимодействие факторов риска.

Рассмотрим гипотетический пример взаимодействия двух факторов во влиянии на развитие болезни (табл. 2.27). В таблице приведены неполные данные двухфакторного эксперимента: есть только данные о том, как факторы по отдельности увеличивают инцидент: фактор В на 2 и фактор А на 4.

Таблица 2.27. Инцидент болезни в группах людей, подвергавшихся воздействию двух факторов

Фактор А	Фактор В	
	–	+
–	2	4
+	6	

Если риск от действия двух факторов суммируется, то это соответствует случаю, который называется суммационным или аддитивным взаимодействием. В табл. 2.28 слева приведены величины дополнительного риска от действия факторов, а справа — итоговый инцидент, включающий «базовый» риск при отсутствии экспозиций и при их комбинациях. Всего:  $[2(\text{базовый риск})] + 2 + 4 = 8$ .

При множительном (мультипликативном) взаимодействии двух факторов суммарный риск увеличивается кратно увеличению риска от действия отдельных факторов. Соответственно, правильно описать результаты можно не разностью рисков, а их отношением (табл. 2.29). Иными словами, относительный риск — это инструмент описания мультипликативного взаимодействия факторов риска. Всего:  $[2(\text{ба-$



зовый риск)]  $\times 2 \times 3 = 12$ . Именно с таким взаимодействием связывается понятие «синергизм». Это взаимодействие наблюдается у курения и экспозиции асбесту в отношении возникновения рака легких, при взаимодействии хронического гепатита В и экспозиции афлатоксинам в отношении возникновения рака печени и во многих других случаях.

Таблица 2.28. Риск в группах людей, подвергавшихся воздействию двух факторов, аддитивная модель

Дополнительный риск			Инцидент		
Фактор А	Фактор В		Фактор А	Фактор В	
	-	+		-	+
-	0	2	-	2	4
+	4	6	+	6	8

Таблица 2.29. Риск в группах людей, подвергавшихся воздействию двух факторов, мультипликативная модель

Относительный риск			Инцидент		
Фактор А	Фактор В		Фактор А	Фактор В	
	-	+		-	+
-	1	2	-	2	4
+	3	6	+	6	12

При анализе реальных данных оценивают, существует ли статистически значимое повышение риска, а затем повышение риска сопоставляется с тем, что ожидается исходя из аддитивной модели и из мультипликативной модели. Имеет значение, какой инструмент используется для описания данных. Логистическая регрессия предусматривает множительное взаимодействие. Между тем суммирование эффектов наблюдается достаточно часто и при использовании только логистического анализа может быть не выявлено.



## Контрольные вопросы и задачи к разделу 2.8

Выберите один правильный ответ.

1. Наклоном в уравнении линейной регрессии  $y = 2,4 + 0,2x$  является:
- а) 2,4;
  - б) 0,2;
  - в) 0,8;
  - г)  $x$ .

Выберите все правильные ответы.

2. Для выяснения связи аллергических проявлений у детей первого года жизни с курением матери в период беременности опросили о курении матерей 100 детей, обратившихся за помощью в связи с данным состоянием, и 100 контрольных матерей, родивших детей в тот же день в том же стационаре, но не имеющих этого состояния у ребенка. Найдено отношение шансов 2,3 при 95% доверительном интервале от 1,2 до 5,7. Полученный результат означает:
- а) отношение шансов может быть повышено, поскольку включены дети только до 12 мес от рождения;
  - б) отношение шансов может быть повышено, поскольку в контрольной группе могут быть дети с таким же состоянием, но матери не считают это проблемой;
  - в) результаты подтверждают гипотезу, но связь слаба для уверенного суждения;
  - г) поскольку исходы оценены у детей, а экспозиция — у матерей, это не настоящее исследование типа сравнения с контролем;
  - д) шансы могут быть повышены, поскольку матери больных детей могут полнее представлять данные о курении.

Выберите один правильный ответ.

3. Для выяснения связи курения подростков с курением их родителей опросили о курении родителей 100 курящих школьников и 100 контрольных родителей подростков, обучающихся в тех же классах средней школы. Найдено отношение шансов 4,3 при 95% доверительном интервале от 1,5 до 6,2. Полученный результат означает:
- а) результат не подтверждает гипотезы о связи курения подростков с курением их родителей;
  - б) если результаты исследования правильны, то они означают, что подросток, у которого один или два родителя курят, имеет шансы в 4,3 раза больше курить, чем тот, у кого родители не курят;



в) поскольку ДИ не включает 1,  $P > 0,05$ ;

г) 90% ДИ, вероятно, включает 1,0.

Выберите все правильные ответы.

4. При выборе контрольной группы в ИСК необходимо, чтобы она:
- а) позволяла получить данные об экспозиции таким же образом, как в основной группе;
  - б) позволяла подобрать члену основной группы сходного контрольного субъекта;
  - в) была набрана из той же популяции, что и основная.

Выберите один правильный ответ.

5. Анализируя данные о 1000 последовательно выявленных больных сахарным диабетом, исследователь нашел, что 600 из них имеют избыточную массу тела, и сделал вывод о связи диабета и ожирения. Если он ошибся, то в чем может быть причина ошибки?
- а) нет группы сравнения;
  - б) нет отслеживания пациентов;
  - в) не сделана поправка на возраст;
  - г) знаменатель не определен.



## Глава 3

# ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Для использования в практике лечебных и профилактических вмешательств врачи нуждаются в сведениях о том, насколько эти вмешательства полезны, т.е. помогают достигать желанных исходов и безопасны. До XIX в. врачи руководствовались в основном умозрительными теориями болезней, проявлением чего было широкое применение кровопускания и слабительных при всех болезнях (идея очищения, изгнания болезнетворных начал). С XIX в. вмешательства вводились в практику на основании естественно-научных представлений о механизмах болезней. Поскольку представления о работе «машины» человеческого тела и ее «поломках» были еще менее совершенны, чем сейчас, не удивительно, что вмешательства нередко носили причудливый характер (кровопускание, клистир, прижигание, горчичники, кровососные банки, длительный сон, диета, обливания холодной водой и др.). Эти вмешательства обосновывались знаниями, полученными при наблюдении над больными в ходе лечения и в физиологических экспериментах. Фрагментарность знаний не останавливала врачей в формировании теорий, которые, в свою очередь, обосновывали лечение. Испытание лечения в опыте для многих казалось излишеством, поскольку хорошая теория выглядела достаточным основанием для действий.

Идея необходимости выбирать вмешательство на основании экспериментальных данных имеет, по крайней мере, 200-летнюю историю. Уже в 1840 г. парижанин Ж. Гаваре (J. Gavaret) опубликовал книгу по медицинской статистике, в которой излагалась идея медицинского эксперимента, статистического анализа данных и доверительных интервалов. Ж. Гаваре предлагал врачам отдавать предпочтение одному из методов лечения при условии, если результаты лечения лучше и если разница результатов лечения превышает величину, зависящую от числа наблюдавшихся случаев. К сожалению, практика надолго отстает от пе-



редовых идей, и современная практика терапевтического эксперимента сформировалась только во второй половине XX в., а доверительные интервалы стали широко использоваться в описании результатов исследований только в последнее десятилетие XX в.

Идея сопоставления разных методов лечения по их результатам проста, но в большинстве случаев она реализовывалась и нередко до сих пор реализуется негодными средствами. Самым распространенным из негодных способов является анализ текущей практики. При этом, например, из числа больных, находившихся на лечении, выбирают тех, кто лечился хирургическими средствами, и сравнивают их с теми, кто лечился «консервативно», т.е. терапевтически. Результат лечения через год после его завершения может быть лучше у лечившихся консервативно, но это может быть простым следствием отбора для оперативного лечения больных, до этого лечившихся консервативно без успеха. Единственным способом получить надежный, правильный ответ на вопрос о преимуществах одного вмешательства перед другим и о полезности вмешательства по сравнению с выжиданием является контролируемый терапевтический эксперимент.

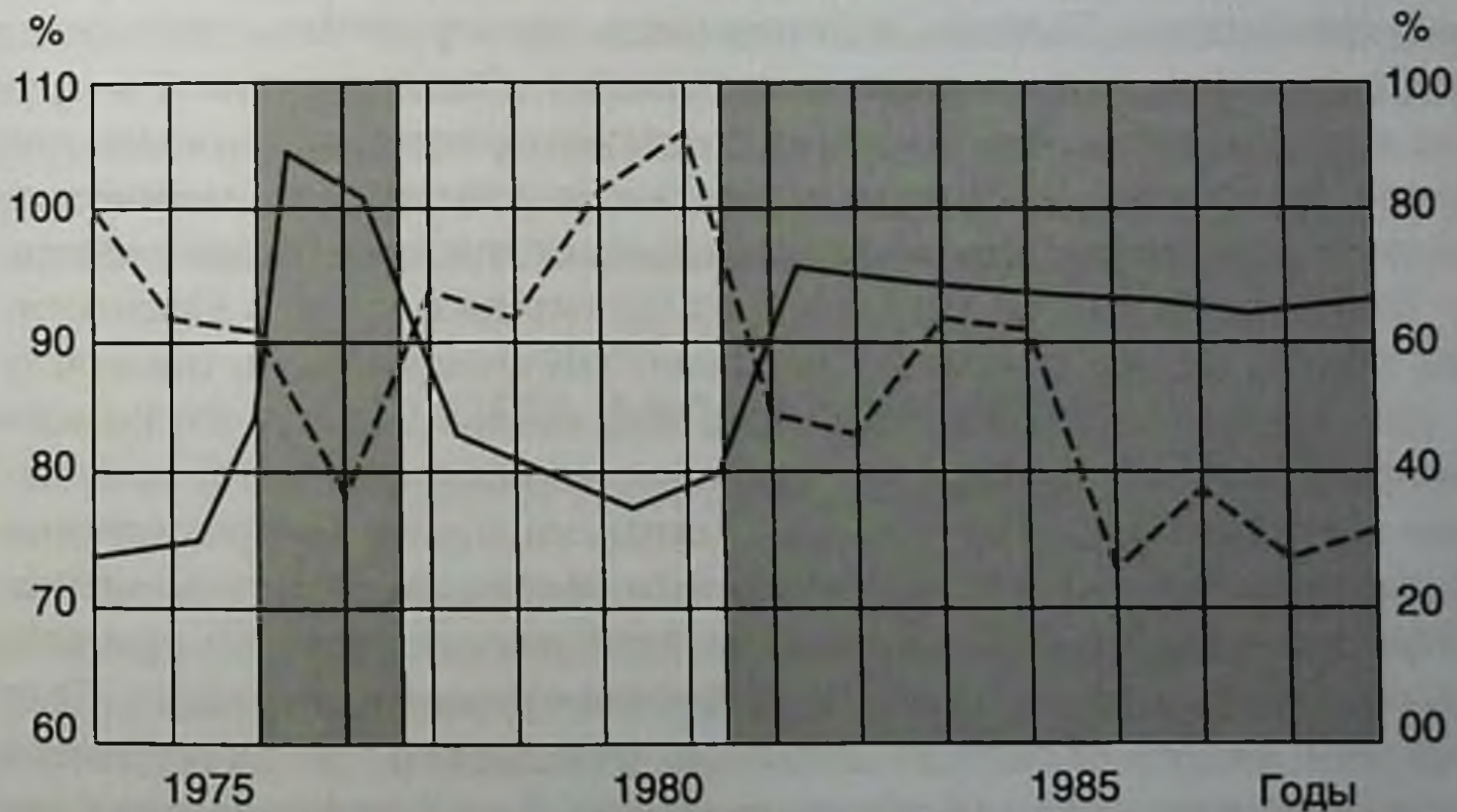


Рис. 3.1. Связь летальности в дорожных происшествиях и доли водителей, использующих ремни безопасности (пунктир, шкала справа в процентах). Линия — «индекс летальности» в процентах к уровню 1973 г. (шкала слева). Периоды, когда закон, обязывающий использовать ремни, действовал, обозначены серым, его отменяли (белый интервал) и опять вводили (пример Швейцарии, по данным Huguenin R.D., 1990, цитируется по: Evans L. Traffic Safety and the Driver. Van Nostrand Reinhold. NY, 1991)



Эксперимент является самым надежным способом испытания правильности гипотезы о причине болезни и о действенности вмешательства. Иногда в распоряжении эпидемиолога оказываются результаты «естественного эксперимента» — вмешательства, проведенного без его участия другими людьми или организациями. Поскольку такое вмешательство обычно не имеет контрольной группы, его результаты трудно объяснить — нельзя исключить, что изменения произошли по другим причинам и лишь совпали со временем вмешательства. В редких случаях удается наблюдать, как применение и неприменение вмешательства чередуются и соответственно изменяются исходы. Такого рода наблюдение над «естественным экспериментом» выглядит убедительно (рис. 3.1). Такие эксперименты по введению и отмене обязательного использования привязных ремней положили конец дискуссиям вокруг необходимости применения этого простого изобретения американского врача, десятилетиями отвергавшегося автомобилестроительными компаниями.

Применение экспериментального метода в медицинских исследованиях на людях ограничено терапевтическим (профилактическим) экспериментом. Помимо клинического эксперимента существуют неклинические медицинские исследования. Они проводятся в других целях — начиная от изучения фундаментальных закономерностей функционирования организма человека и до испытания косметических средств и средств защиты военнослужащих от действия боевых средств.

Экспериментальные исследования (испытания) средств диагностики, профилактики и лечения называют *контролируемыми*, поскольку в них используется контрольная группа, не получающая изучаемого вмешательства. Это позволяет уменьшать (контролировать) возможные случайные эффекты. Как ни сложны, ни дороги контролируемые испытания, они являются единственным надежным средством оценки эффективности вмешательства. Если эксперимент проведен правильно, то его результаты могут быть обобщены до уровня популяции. Если выбраны репрезентативные лечебные учреждения, то эксперимент в них, например, показавший, что внедрение *формуляра*<sup>1</sup> приводит к по-

---

<sup>1</sup> В эпидемиологии, в научной медицинской литературе используется термин «клинические испытания» (*clinical trials*). Клинические испытания являются одним из вариантов клинических исследований. В российской регуляторной литературе, в том числе в законодательстве, используются термины «клинические исследования», «рандомизированные клинические исследования», а экви-



вышению качества фармакотерапии, доказывает применимость этого инструмента ко всем подобным учреждениям. Если эксперимент проведен на репрезентативной группе пациентов, то его результаты могут быть обобщены для всех подобных пациентов.

Экспериментальная оценка вмешательств используется не только при терапевтических вмешательствах и не только в медицине, но во всех областях социальной жизни. Например, предполагалось, что если делинквентных подростков, т.е. совершавших мелкие правонарушения, ознакомить с теми последствиями, к которым ведет путь нарушения закона, в частности сводить их на экскурсию в тюрьму, то они будут меньше нарушать закон. Был проведен контролируемый эксперимент, который обнаружил противоположный эффект этого вмешательства — после подобных экскурсий подростки чаще оказываются в тюрьме.

Противоположностью контролируемых испытаний являются неконтролируемые. В ходе таких испытаний, например, в школах города вводится программа полового образования, и ее эффективность оценивается по изменению числа беременностей у подростков. Поскольку этот показатель сильно колеблется и находится под влиянием множества факторов, невозможно узнать, чем вызваны изменения или почему их не произошло. Всякие упрощения метода исследования приводят к ухудшению конечного результата — неверной оценке лечения, дополнительным расходам, использованию в широкой практике неэффективных и опасных средств или неиспользованию эффективных.

### 3.1. ЦЕЛЬ И КЛАССИФИКАЦИЯ КЛИНИЧЕСКИХ ИСПЫТАНИЙ

В оценке *эффективности вмешательства* в широком смысле (англ. *effectiveness*) выделяют оценку действенности и эффективности.

Под *действенностью* (англ. *efficacy*) понимают способность вмешательства предотвращать нежелательные исходы в условиях контролируемого эксперимента. Цель испытания вмешательства на действен-

---

валентом термина «лекарство» (англ. *medicine*) является термин «лекарственное средство». Терминология в этой области подвижна. Например, ранее широко использовавшееся для обозначения лекарств *drug* (англ.) используется реже, так как этот термин стал преимущественно использоваться для обозначения незаконно обращающихся веществ, наркотиков.



ность — установить, приводит ли вмешательство к тем последствиям, которые ожидаются. Для этого допустимо создавать специальные условия, которые позволят в максимальной степени проявиться ожидаемому эффекту и в которых влияние обычных сложностей клинической практики будет минимальным. Например, подбирают пациентов только с выраженной патологией, исключают широкий круг сопутствующих болезней, проверяют потенциальных участников испытания на способность к выполнению условий испытания и исключают неспособных, в том числе лиц с поведенческими проблемами.

Под собственно *эффективностью* понимают способность вмешательства достигать тех же целей в условиях обычной практики. Например, известно, что правильное использование презервативов позволяет предотвращать беременность почти в 100% случаев (действенность). Использование презервативов школьниками для предотвращения беременностей имеет значительно меньшую эффективность, так как по бытовым и психологическим причинам школьники или избегают использовать их, или делают это неправильно. Соответственно эффективность вмешательства (раздача презервативов) устанавливается в испытаниях, в которых ограничения для участников минимальны, с тем, чтобы результаты испытания можно было уверенно обобщить для условий широкой практики. Для принятия вмешательства в широкую медицинскую практику важна его комплексная оценка, включающая в себя помимо эффективности экономические достоинства и недостатки метода, широкий круг эффектов, возникающих помимо основного эффекта (см. разд. 5.1).

Применительно к лекарствам и вакцинам *клинические испытания классифицируют на 4 фазы*<sup>1</sup>. Испытания фазы I проводятся на малом числе добровольцев. Это могут быть здоровые добровольцы, но иногда — лица, страдающие болезнью, для лечения которой разработано лекарство. Таких испытуемых приглашают в случае, если препарат обладает высокой токсичностью, как, например, препараты для химиотерапии рака. Цель этих испытаний — выявление ожидаемого фармакологического действия лекарства, его токсичности и определение дозы.

---

<sup>1</sup> Фундаментальные требования к проведению исследований с участием человека сформулированы в Хельсинкской декларации Всемирной медицинской ассоциации (World medical association), а наиболее детальные — в стандарте Доброкачественной медицинской практики (Good clinical practice, GCP), поддерживаемом The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use.



Испытания фазы II проводятся на небольшом числе испытуемых с состоянием, которое должно быть объектом применения лекарства. Цель этих испытаний — выявление ожидаемого терапевтического эффекта, уточнение данных о безопасности и уточнение дозировки лекарства.

Испытания фазы III проводятся на большом числе испытуемых, в последние десятилетия — с участием нескольких медицинских центров (многоцентровые). Их целью являются выяснение размера терапевтического эффекта и уточнение данных о безопасности. Обычно на основании результатов испытаний III фазы лекарства разрешаются к применению.

Испытания фазы IV проводятся в виде наблюдения за текущей практикой, как «открытые» испытания (испытания без ослепления) (см. разд. 3.3.3). Их цель — уточнение размера эффекта от вмешательства в условиях реальной практики и уточнение данных о безопасности.

## 3.2. ВИДЫ ВМЕШАТЕЛЬСТВ

Среди множества вмешательств в первую очередь выделяют две большие группы: лечебные и профилактические. *Лечебными* называют вмешательства, применяемые по отношению к людям, у которых есть болезнь (состояние). Цель лечебных вмешательств — устранение проявлений болезни, ее излечение, но чаще — уменьшение активности процесса, прекращение фазы обострения, уменьшение страданий. Иными словами, лечение — это то, что делается в отношении установленной болезни, независимо от того, сам ли обратился пациент за помощью, или его скрытая болезнь найдена при обследовании. Например, лечение артериальной гипертензии зависит в основном от артериального давления и не зависит от того, обратился пациент к врачу с жалобами, или его гипертензия была выявлена при профилактическом осмотре.

В свою очередь, лечебные вмешательства делят на симптоматические, этиотропные, патогенетические, излечивающие и паллиативные. Такое деление в значительной степени зависит не от того, какое вмешательство использует врач, а от того, с какими намерениями он это делает. Так, врач использует бета-агонисты и кортикостероиды у больного бронхиальной астмой для того, чтобы прервать приступ, для предотвращения нового приступа и для замедления развития осложнений.

*Профилактическими* называют вмешательства, применяемые по отношению к людям, у которых нет целевого состояния, с целью предот-



вратить его развитие или при наличии состояния снизить вероятность неблагоприятных явлений. Например, использование ремней безопасности в автомобиле снижает вероятность гибели пассажиров при лобовом столкновении машин.

*Первичной профилактикой* называют вмешательства при отсутствии патологического состояния. Например, обеспечение медицинских работников одноразовым и эффективно стерилизуемым инструментом — средство первичной профилактики передачи инфекций, а использование респираторов — средство первичной профилактики болезней легких у рабочих в пыльных цехах. С точки зрения периодизации развития патологии первичная профилактика имеет место в период индукции болезни (состояния) (см. рис. 1.4).

*Вторичной профилактикой* называют вмешательства, нацеленные на замедление развития заболевания и снижение вероятности осложнений. Например, при облитерирующих поражениях сосудов, ишемической болезни сердца длительно применяют ацетилсалициловую кислоту (Аспирин), что снижает вероятность критической ишемии. После резекции молочной железы при раке возможно применение противоракового препарата тамоксифена для снижения вероятности развития рака во второй молочной железе. На этом примере хорошо видно, что одно и то же средство может применяться как лечебное и как профилактическое (противораковый препарат). Более того, при вторичной профилактике врач лишь *оценивает* свое вмешательство как профилактическое. В действительности его вмешательство может быть модифицирующим течение болезни, т.е. лечебным. Например, ацетилсалициловая кислота может угнетать воспалительные процессы в сосудах, которые провоцируют тромбоз. Тамоксифен может подавлять развитие клеток опухоли, не удаленных при операции.

*Третичной профилактикой* называют вмешательства, нацеленные на облегчение приспособления больного к вызванному болезнью дефекту, предотвращение инвалидности.

### 3.3. ОБЪЕКТ ЭКСПЕРИМЕНТА

Объектом эксперимента могут быть административные районы, больницы, семьи или отдельные субъекты (пациенты, продавцы и т.д.). Эксперимент обычно проводится не над популяцией, а над вы-



боркой из популяции. Выборка должна быть репрезентативной, чтобы результаты могли быть распространены затем на всю популяцию (см. разд. 1.4.1). Вместе с тем в исследование нельзя включать людей, не желающих в нем участвовать, тех, кому изучаемое вмешательство может повредить, а также тех, кто не может свободно проявить свое несогласие (например, дети, заключенные, военнослужащие; см. сноску на с. 322).

### 3.3.1. Формирования выборки

В клинических исследованиях обычно включают в исследование всех пациентов в течение запланированного периода или до достижения запланированного числа пациентов. Если период включения короток, то необходимо оценить, не окажутся ли полученные оценки смещенными из-за временных ограничений. Например, исследование структуры заболеваемости (или госпитализации) только в летние месяцы даст неверное представление о структуре, характерной для стационаров города, — для этого необходимо исследовать хотя бы один полный годовой цикл. Если технические возможности не позволяют включать в исследование всех больных с изучаемой патологией, то должен быть предусмотрен способ рандомизации, позволяющий включать в исследование репрезентативную часть поступающих пациентов. Нельзя, например, просто включать в исследование больных, поступающих в понедельник, поскольку характеристики больных, поступающих в разные дни недели, различаются. Различия возникают не только из-за действия неизвестных исследователю факторов, провоцирующих заражение или обострение болезни, но и из-за различий в доступе к медицинской помощи в разные дни недели.

#### Критерии включения и исключения

Ясные критерии включения пациентов (обследуемых) необходимы для описания популяции (генеральной совокупности), которой соответствуют включенные в исследование субъекты. Если исследуются пациенты с почечной недостаточностью, то должно быть решено, по какому критерию устанавливается ее наличие (например, по концентрации креатинина в крови). Если обследуются поликлиники, то необходимо уточнить, в селе или в городе, какого размера, с каким штатом врачей, с какой численностью обслуживаемого населения.



Ясные критерии исключения необходимы для создания достаточно однородной группы пациентов. Например, из числа участников терапевтического эксперимента исключают лиц с тяжелыми сопутствующими болезнями, с состояниями, мешающими выполнить условия эксперимента (например, с деменцией), тех, кому вмешательство может быть вредно. Исключением нетипичных случаев обеспечивается однородность группы пациентов, т.е. меньшая вариабельность переменных в исходном состоянии и в оценке величины эффекта вмешательства. Этим увеличивается статистическая мощность эксперимента. Например, в результате организации образовательных программ для врачей будет легче выявить изменения, если из числа экспериментальных поликлиник исключить поликлиники медицинского университета.

В исследование могут быть включены все и только те пациенты, для которых сравниваемые вмешательства могут принести сравнимую пользу, или если отличие испытываемого вмешательства от плацебо не установлено (принцип неопределенности — Р. Пито (R. Peto)). В противном случае пациент должен получать вмешательство, польза которого для него определена, доказана.

### Согласие участников

В идеале все пациенты, подходящие под критерии включения, должны быть включены в испытание. Практически же люди имеют разные ценности и по-разному представляют ценность вмешательств. Поэтому не все пациенты дают согласие на участие в испытании. Одни могут предпочитать один из испытываемых методов лечения и не хотят оставлять его выбор случаю. Другие в принципе не хотят быть объектом исследования или избирают иной способ лечения. Соответственно таких пациентов в исследование не включают, и они получают то лечение, которое им показано и которое они избрали (дали согласие). Необходимо, чтобы отклик (англ. *response rate*), т.е. доля людей, откликнувшихся на просьбу включиться в исследование, от всех соответствующих критериям включения был достаточно высок — не менее 80%. В противном случае сформированная группа, возможно, не будет соответствовать сформулированным критериям включения и исключения.

Результат испытания, от участия в котором отказались много пациентов, может быть неприложим ко всем людям (например, к тем, кто



отказывается). Так, испытывать методы лечения поясничных болей трудно, поскольку немало людей уверены в том, что им поможет только мануальная терапия (манипуляции с позвоночником, выполняемые целителем-хиропрактиком). Результаты лечения у таких людей лучше, независимо от лечения: пациенты, активно выбирающие некоторый метод лечения, лечатся старательнее, правильнее выполняют назначения. Это свойство людей называют *аккуратностью*, или *исполнительностью*, пациентов, но чаще — *комплаенс* (англ. *compliance*). Окончательного решения проблемы самоотбора пациентов не существует. В некоторых исследованиях, где ожидается множество отказов от рандомизации, применяется следующий прием. Пациентов, которые на рандомизацию не согласились, включают в исследование и предоставляют им то вмешательство, которое им показано и на которое они согласились. Затем результаты лечения у них сравнивают с результатами в группах, где вмешательство применялось на основе рандомизации.

### Планирование числа участников

Число включаемых в эксперимент пациентов (размер выборки) должно быть обосновано. Обоснование исходит из: 1) предполагаемой величины эффекта (например, снижения частоты обострений или увеличения выживания при применении нового лечения); 2) структуры исследования и 3) установленного заранее порога статистической значимости обнаружения эффекта. Исследование должно планироваться так, чтобы число пациентов было достаточным для обнаружения предполагаемого эффекта.

Методики выбора числа обследуемых (пациентов) разработаны для многих структур исследования, а не только для исследований с экспериментальным вмешательством. Вычисления относительно сложны, и их следует выполнять с помощью статистических программ<sup>1</sup>. В целом чем большая статистическая значимость различия в действительности двух

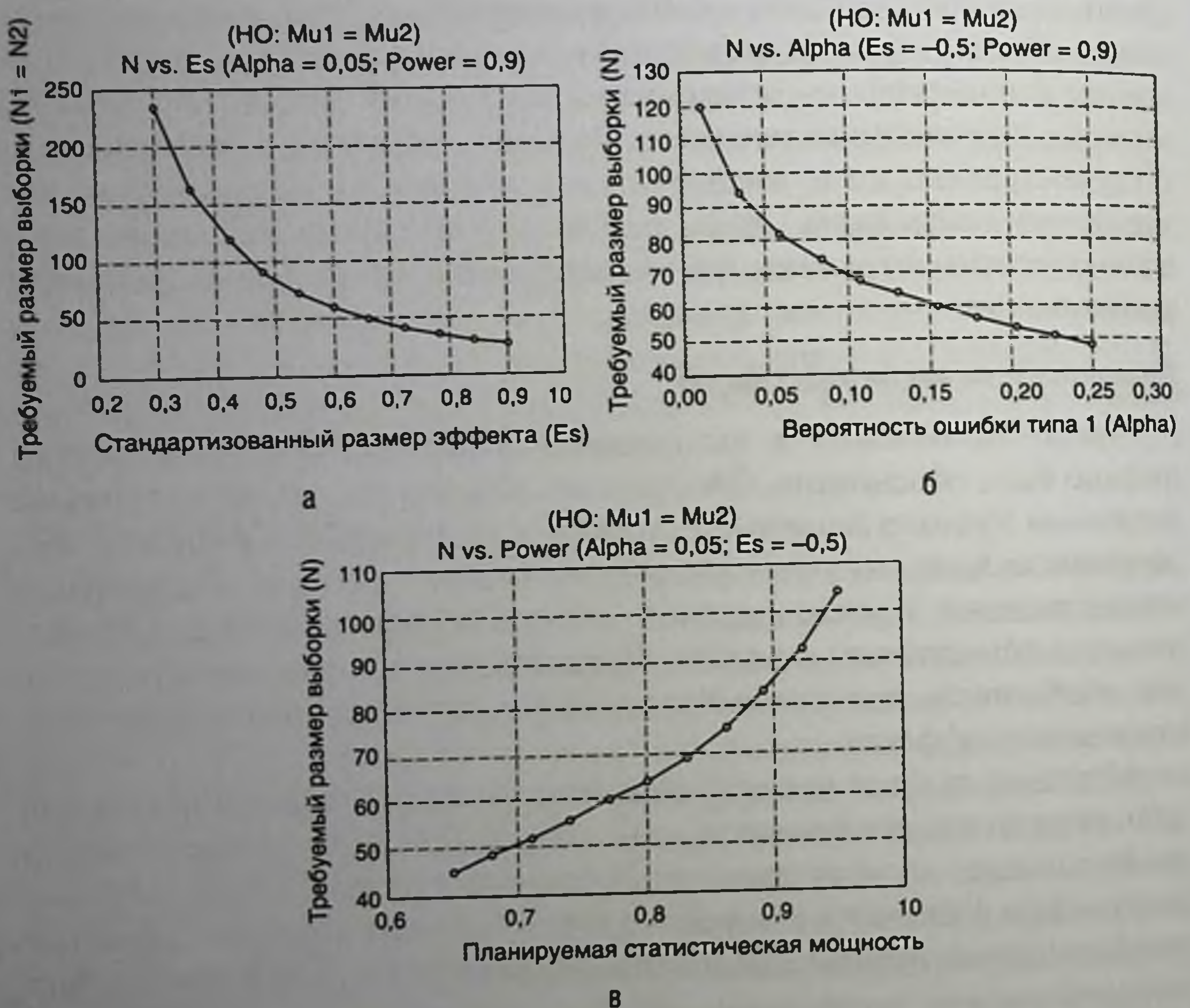
---

<sup>1</sup> Такие программы доступны бесплатно в интернете (<http://www.doctor.ru/doctor/biometr/lib/sensit.htm>; <http://www.stat.ucla.edu/~jbond/htmlpower/index.html> и <http://hedwig.mgh.harvard.edu/size.html>)), встроены в некоторые пакеты статистических программ, например программа EGRET SIZ — дополнение к биомедицинскому статистическому пакету программ EGRET или в специальные программы для эпидемиологов, доступные бесплатно: PEPI — [www.usd-inc.com/pepi.html](http://www.usd-inc.com/pepi.html), EpiSheet — <http://members.aol.com/krothman/episheet.xls> и EpiInfo (см. с. 103).



вмешательств задается и чем меньший эффект ожидается обнаружить, тем большее число пациентов необходимо включить в исследование (рис. 3.2).

При равном числе испытуемых (пациентов) в сравниваемых группах достигается максимальная статистическая мощность эксперимента. Когда возможности применить одно из вмешательств ограничены его ценой или физическими возможностями (например, количеством



**Рис. 3.2.** Планирование численности экспериментальных групп для эксперимента, в котором будут сравниваться две независимые выборки с помощью  $t$ -критерия: а — необходимый размер выборки быстро уменьшается при увеличении ожидаемого размера эффекта (стандартизованный эффект — это ожидаемая разность между средними в сравниваемых группах, деленная на стандартное отклонение); б — необходимый размер выборки быстро растет при снижении допустимого  $\alpha$ -уровня; в — необходимый размер выборки растет при снижении допустимого  $\beta$ -уровня. Для расчета использован специальный модуль программы Statistica



полученного препарата) или когда в исследовании типа сравнения с контролем число найденных случаев невелико, можно увеличить статистическую мощность исследования, увеличивая численность одной из сравниваемых групп. Этот метод малоэффективен, и увеличивать численность «доступной» группы более чем вдвое по сравнению с меньшей практически не рационально.

Проведение исследований с числом пациентов, недостаточным для получения статистически значимого результата, неэтично, поскольку в таком исследовании его субъекты будут подвергаться рискам испытания без надежды на получение нового полезного знания.

### 3.3.2. Распределение и применение вмешательств

#### Неконтролируемый эксперимент

Неконтролируемый эксперимент в виде наблюдения за результатами вмешательства является самым древним и примитивным видом эксперимента. Он дает достаточные основания для оценки вмешательства, только если размер эффекта огромен. Так, хинин при малярии и нитроглицерин при стенокардии демонстрируют замечательный эффект, который не нуждается в дальнейшей проверке. За редкими исключениями медицинские вмешательства оказывают небольшое влияние на исход болезни, и его легко не заметить, если проводить неконтролируемый эксперимент. Также легко обнаружить эффект там, где его нет.

Важнейшей причиной обнаружения эффекта там, где его нет, является волнообразность естественного течения болезни. Хроническим болезням свойственно проходить через периоды затишья (ремиссии) и периоды обострения. Больной обычно обращается к врачу в период обострения. Очевидно, что, поскольку вслед за обострением придет ремиссия, что бы ни делал врач, больному станет легче (рис. 3.3). Например, долгие годы применяли рентгенотерапию для лечения остеоартрита и болезней позвоночника (см. рис. 2.16). Больным объясняли, что им станет легче после начала лечения или чуть позднее, но может быть и временное ухудшение состояния. Это объяснение было правильным, но оно лишь соответствует естественному течению болезни и не говорит ничего об эффективности лечения.

Регрессия к средней — вторая важнейшая причина ошибок в оценке вмешательства. Регрессия к средней означает тенденцию аномальных



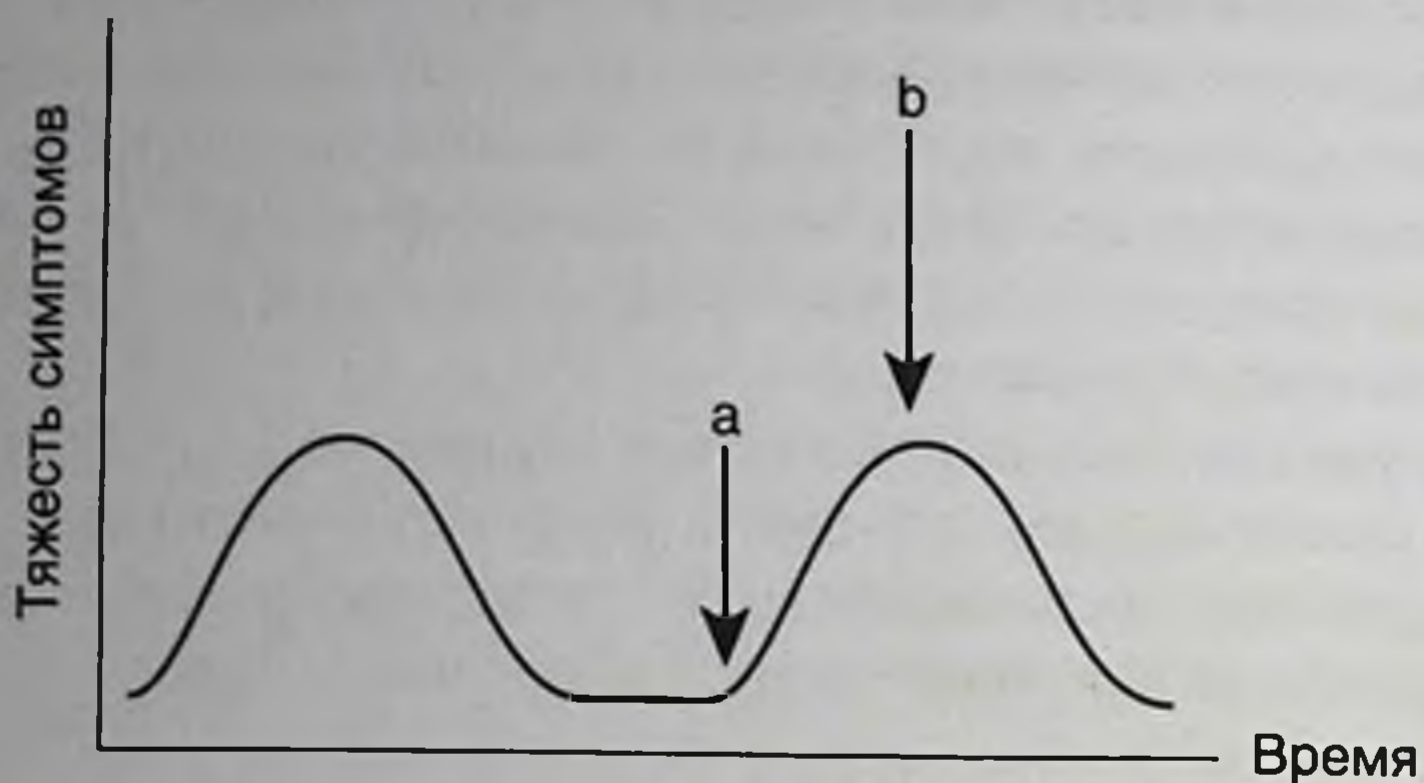


Рис. 3.3. Колебания в активности болезни. Вслед за вмешательством может наблюдаться ухудшение (после a) или улучшение (после b) при полной неэффективности вмешательства

(больших и малых) величин «нормализоваться» со временем, т.е. приближаться к средним групповым характеристикам (см. с. 94). При неэффективном воздействии или при отсутствии воздействия у людей, выбранных по признаку «высокого артериального давления», со временем произойдет некоторое снижение артериального давления. Этим обосновано правило повторного измерения любых переменных после обнаружения неожиданного отклонения, например, при лабораторном исследовании по множеству признаков. Тот же эффект имеет место вследствие ограниченной воспроизводимости клинических и инструментальных данных (см. разд. 2.2.2). Например, известно, что при гастродуоденоскопии врач может обнаруживать язвы там, где их нет, и не замечать имеющиеся язвы. В обычной практике врач лечит только тех, у кого найдены язвы. При повторном исследовании у части пациентов язвы будут не замечены. Некритически мыслящий врач будет полагать, что эти больные были излечены.

Поскольку врачи действуют для блага пациентов, им естественно верить в то, что от их действий больным становится лучше. Это называют *психологической систематической ошибкой* в оценках врачей. Врачи, размышляя над результатами лечения, находятся под впечатлением недавних или ярких случаев — вспомнить и систематизировать всех подобных пациентов невозможно. Такое смещение в оценке результатов лечения называют *смещением воспоминания* (англ. *recall bias*). Единственным способом избежать ошибок, связанных с оцен-



кой медицинских вмешательств в неконтролируемом эксперименте или по результатам текущей практики, является проведение контролируемых испытаний.

### Испытание в параллельных группах

Наиболее важным достижением клинической эпидемиологии является разработка методологии испытания медицинских вмешательств — *клинических испытаний* (КИ, англ. *clinical trials*). Простейшим случаем КИ является сравнение результатов вмешательства в двух группах — контрольной и опытной. Опытная группа получает изучаемое вмешательство, а контрольная — нет. Такую методику обозначают как проспективное исследование с вмешательством и с параллельным контролем.

**Независимые выборки.** В КИ контрольная группа должна подвергаться всем тем воздействиям, что и опытная, с одним лишь отличием — контрольная получает другое лечение. Пациентов подбирают одинаково и разделяют на опытную и контрольную группы случайным образом (рандомизация) (см. с. 335). Численность групп может быть любой, не обязательно одинаковой. Например, в группу вмешательства можно помещать в 2 раза больше пациентов, чем в группу контроля. Поскольку пациенты включаются в группы независимо друг от друга, такие группы называют *независимыми*. Полученные при таком исследовании данные также называют независимыми (несвязанными) (рис. 3.4).

**Связанные выборки** (подобранные пары, англ. *matched pairs*) — это случай, когда каждому пациенту в одной группе соответствует парный ему пациент в другой. Парой может быть пациент, подобранный, напри-

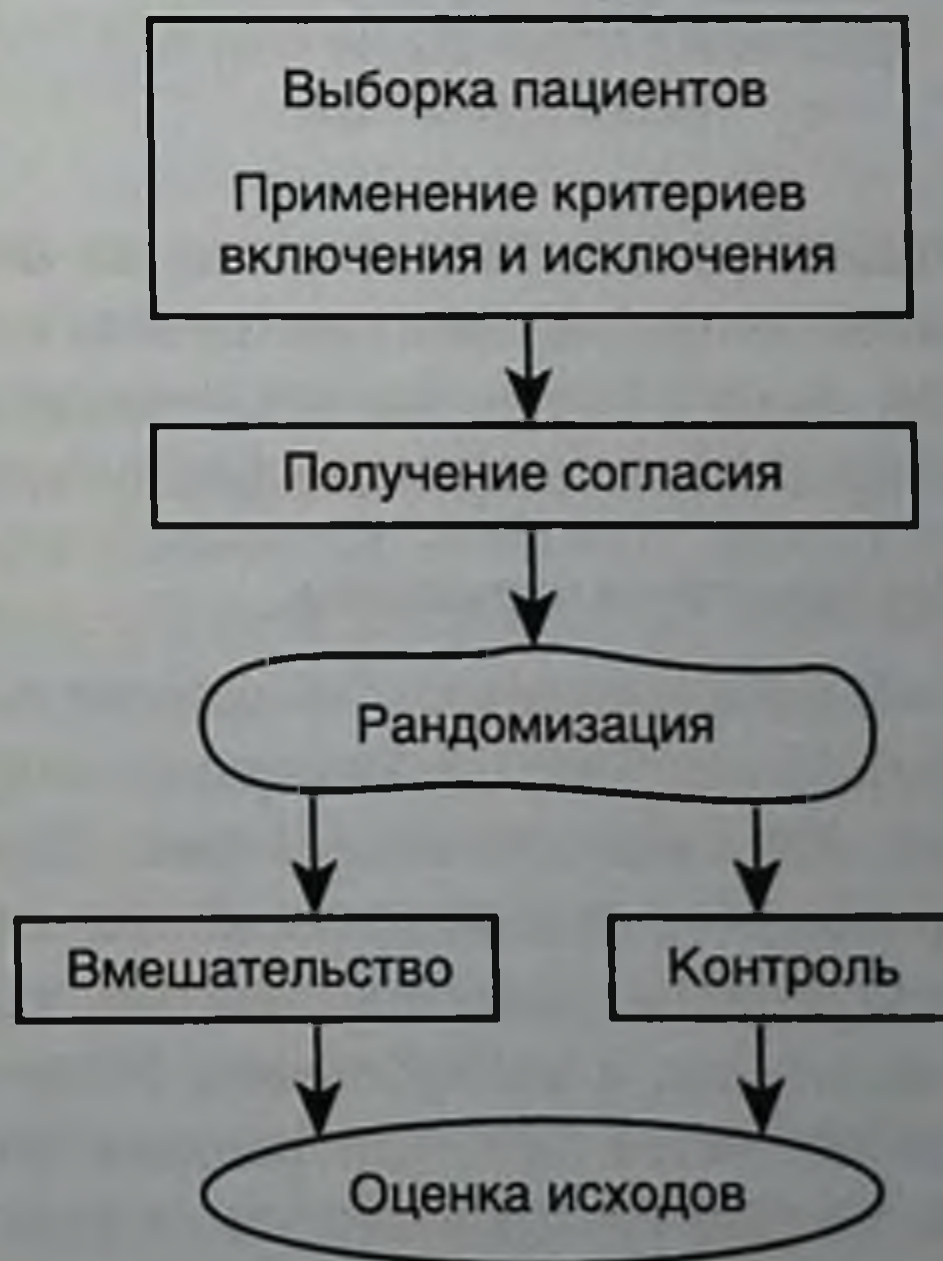


Рис. 3.4. Структура клинического испытания с параллельным контролем





Рис. 3.5. Структура клинического испытания с параллельным контролем на связанных парах

мер, по полу и равному возрасту. Когда подобрана пара, проводится рандомизация. Она определит, какое вмешательство получит один из членов пары; второй получит другое (рис. 3.5). Конечно, пациенту «опытной» группы может соответствовать не один, а два или более пациентов контрольной. При использовании более 5–6 признаков подобрать пару оказывается практически невозможным.

Подбором пар можно устранить влияние лишь известных (предполагаемых) побочных факторов. Никогда нельзя исключать действия других факторов. Поэтому одного подбора пар недостаточно, и нужно вмешательства внутри пар рандомизировать. Если пары в исследовании подо-

браны, то они уже не являются независимыми. К связанным данным должны при анализе применяться специальные статистические методы (см. разд. 1.8.2), а при планировании эксперимента должны использоваться соответствующие формулы для расчета численности групп.

### Перекрестное испытание

Чтобы максимально устранить неизбежную межиндивидуальную вариабельность, иногда применяется *перекрестная организация* испытания *КИ (cross-over design)* (рис. 3.6). При этом между основной и контрольной группами почти нет отличий: обе группы получают по очереди лекарство и плацебо. Иными словами, каждый пациент является и опытным, и контрольным. Например, в первом периоде одной группе пациентов дают контрольное лечение, а второй — изучаемое новое. Во втором периоде лечение в двух группах меняется. Сопоставляется эффект от лечения в каждой группе в каждом периоде. Такая структура испытания подходит для сравнения, например, противовоспалительных средств в терапии хронических болезней.



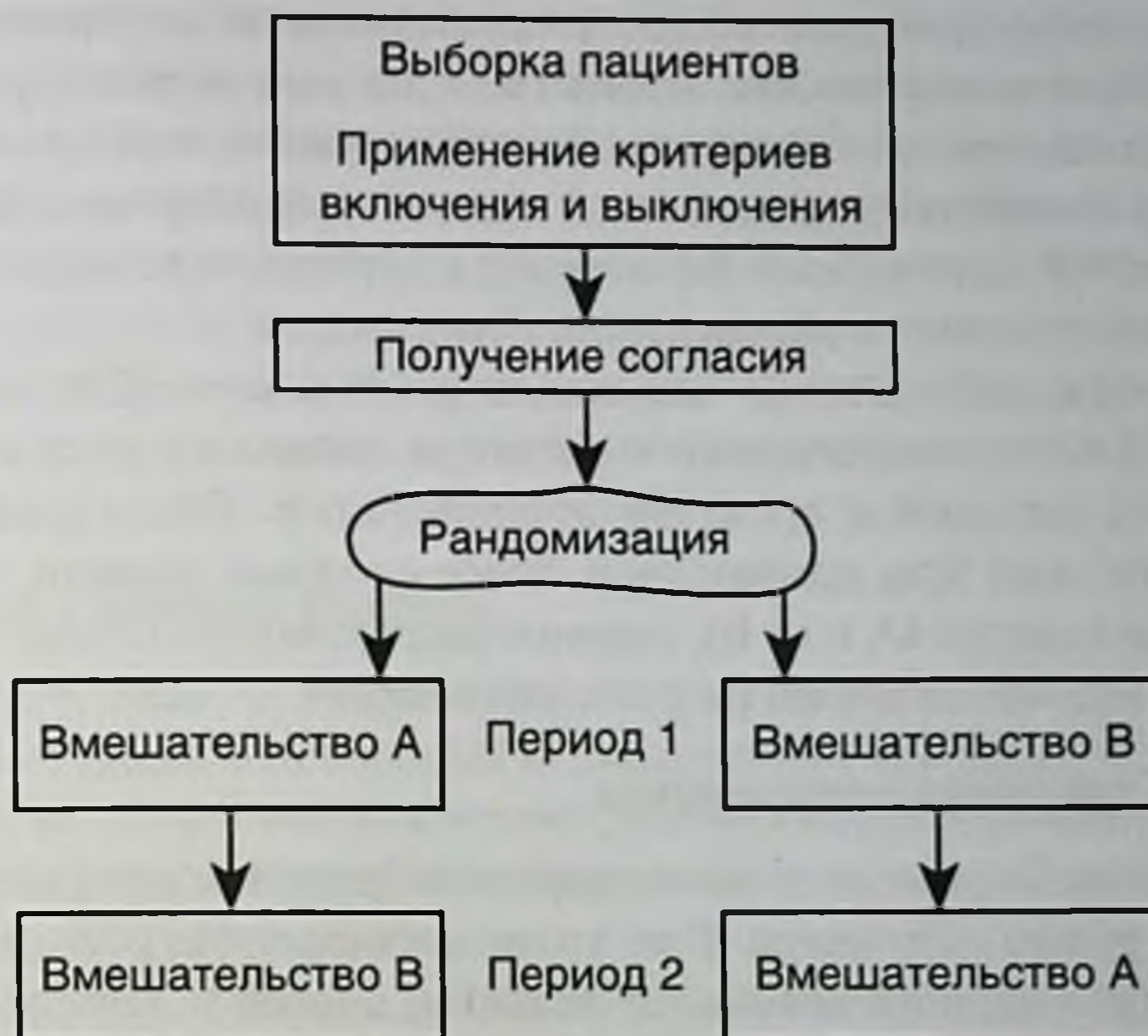


Рис. 3.6. Схема двухпериодного перекрестного испытания лекарства А в сравнении с плацебо В

Поскольку эффект может зависеть от последовательности лечения и может сохраняться какое-то время после лечения, между этапами эксперимента делается перерыв, и применяется две очередности: лечение А, затем В и лечение В, затем А. Главным достоинством такой схемы является экономичность — статистически значимый результат можно получить при числе пациентов, в 2 раза меньшем, чем при параллельном контроле. Это важно, поскольку информированное согласие на экспериментальное лечение дает не каждый пациент, и больных исследуемой болезнью может быть немного. Перекрестная схема позволяет лучше провести испытание в одной клинике и выявить более тонкие различия. Последнее возможно, поскольку точно одни и те же пациенты обследуются при новом и контрольном лечении.

Перекрестные КИ не должны применяться в случае, когда эффект одного из видов лечения очень длителен или предполагается «потеря» испытуемых, или пациенты выпадают из исследования по причине полного выздоровления (как это имеет место при острых заболеваниях, завершающихся обычно излечением). Перекрестные КИ ненадежны



в случаях, когда сравниваемые методы лечения дают внешне различающиеся фармакологические эффекты, если не удастся вернуть пациентов к исходному состоянию после предыдущего этапа лечения, если изменения состояния пациентов во времени неоднородны или эффект накапливается постепенно. Во всех этих случаях нужны не перекрестные КИ, а испытания с параллельным контролем.

Статистический анализ данных перекрестного КИ может быть простым. Для этого сопоставляют исходы, например силу кисти и болезненность суставов в процессе лечения А и В. Более плодотворный анализ возможен при применении более сложной модели, учитывающей вмешательство (А или В), период эксперимента (1 или 2) и последовательность применения (первая или вторая).

### Факторная структура эксперимента

В одном эксперименте можно сравнить более чем два вмешательства на одной группе пациентов. Для этого надо сначала рандомизировать пациентов для первого лечения и плацебо, а затем — для второго лечения и плацебо (табл. 3.1). В результате будут пациенты, которые получают вмешательство 1 и 2 (в примере их 36), которые получают вмешательство 2 (38 пациентов), которые получают вмешательство 1 и которые получают только плацебо. Конечно, получающие плацебо получают в действительности два плацебо, а получающие одно вмешательство получают вместо второго плацебо. Факторная структура неприемлема, если существует взаимодействие между изучаемыми вмешательствами.

Таблица 3.1. Распределение пациентов в эксперименте факторной структуры

Вмешательство 2	Вмешательство 1	
	Да	Нет (плацебо)
Да	36	38
Нет (плацебо)	37	40

Факторная структура исследования экономически очень эффективна в сравнении с проведением двух отдельных испытаний для оценки двух вмешательств. На одной группе испытуемых изучается два вопроса без потери статистической мощности эксперимента. Обычно факторную структуру применяют, когда запланировано испытание некото-



рого вмешательства, но появляются интерес и достаточные основания для проверки действенности еще одного вмешательства. Например, при подготовке испытания ацетилсалициловой кислоты в профилактике сердечно-сосудистых болезней в эксперимент было добавлено еще одно вмешательство — бета-каротин. Оба фактора были оценены успешно — действенность ацетилсалициловой кислоты подтверждена и эта часть эксперимента прекращена, а действенность бета-каротина выявить не удалось до запланированного окончания рандомизированного контролируемого испытания (РКИ).

### Исторический и внешний контроль

Выше речь шла о параллельном или «одновременном» (англ. *concurrent*) контроле. В некоторых исследованиях используют исторический контроль, т.е. результаты нового лечения сопоставляют с результатами другого лечения в предшествующие месяцы, годы. Так поступают, когда хотят сэкономить средства и время (все пациенты войдут в новую «опытную» группу, и она быстро будет укомплектована), когда считают невозможным плацебо-контроль или уже считают невозможным применение старого лечения.

Часто до нашего времени так поступают хирурги — на экспериментальных (опыты на животных) и теоретических основаниях они вводят новую методику операции, не проводя рандомизированного контролируемого сопоставления со старой методикой. Применительно к новым хирургическим вмешательствам показано, что впервые новые вмешательства проводят относительно тяжелым больным по очень узкому кругу показаний. Затем, если вмешательство выглядит как эффективное, круг показаний расширяют, в него включают менее тяжелых больных. Естественно, результаты лечения улучшаются. Если не учитывать изменения спектра пациентов, то можно, проводя историческое сравнение, сделать вывод о совершенствовании методики лечения и повышении его эффективности.

В случае применения исторического контроля должны быть приняты меры к тому, чтобы пациенты, включенные в группу вмешательства сегодня, не отличались от группы контроля (получавших сравниваемое лечение в прошлые годы). Это трудно, ведь со временем изменяются структура населения и его питание, квалификация врачей, структура лечебных учреждений региона, обеспеченность лекарствами, транспортная ситуация и многое другое. Поэтому необходимо собрать исчер-



пывающие сведения об «опытном» и «контрольном» периодах для того, чтобы можно было убедиться в их схожести. Из-за того, что избежать различий во времени невозможно, и поскольку новые вмешательства вводятся с энтузиазмом, в исследованиях с историческим контролем чаще обнаруживаются преимущества у нового метода лечения, чем в исследованиях с одновременным контролем.

Другим вариантом несовершенного выбора контрольной группы является сопоставление результатов изучаемого лечения с результатами лечения в другой клинике. Такой «внешний» контроль неприемлем, так как в разных клиниках пациенты, условия и методы лечения различаются.

Структуры исследований с историческим и внешним контролем являются несовершенными. Обычно последующее испытание с параллельным контролем не подтверждает многообещающих результатов методически неправильного сравнения. Оценки технологии (вмешательства) должны делаться по результатам доброкачественных испытаний с параллельным контролем.

## Рандомизация

*Случайное* (англ. *random*) распределение пациентов по группам — классическая процедура достижения неразличимости групп в КИ (см. раздел 1.4.2). Рандомизация впервые была применена в клиническом испытании только после Второй мировой войны. Распределяют пациентов по группам случайным образом для того, чтобы различия между группами были как можно меньше. Еще важнее, что различия групп, сформированных рандомизацией, носят случайный характер по всем признакам, даже по тем, о которых исследователь ничего не знает. Из принципа случайного комплектования групп исходит методология статистического анализа данных — различия групп по определению являются случайными.

Рандомизация выполняется разными средствами — от подбрасывания монеты и использования таблиц случайных чисел до компьютерных программ. Несмотря на простоту этих средств, иногда рандомизацию подменяют негодными методами. Примерами *псевдорандомизации* является распределение в группы по первой букве имени (возникают неравенства групп из-за этнических различий, связанных с именами), дате рождения, номеру медицинской карты, дню недели поступления в клинику и т.д. Псевдорандомизация опасна тем, что метод узнают третьи лица и смогут влиять на подбор пациентов. Например, сестра в при-



емном отделении сможет присваивать четные номера медицинским картам симпатичных ей пациенток. Главное неблагоприятное следствие псевдорандомизации — становится известным, кто из пациентов какое лечение получает. Этим не выполняется *главное условие рандомизации* — *сокрытие ее результатов* (англ. *concealment*), и рандомизация не может выполнить своей важнейшей функции — обеспечения слепого характера исследования (см. с. 343). В исследованиях, где меры к сокрытию результатов рандомизации не принимались или сокрытие было недостаточным, оценка эффективности вмешательства завышалась примерно на 25%. Для обеспечения сокрытия принимаются надежные технические меры. Например, после регистрации в испытании пациента, давшего согласие, его данные сообщаются в территориально удаленный центр, откуда в ответ сообщается результат рандомизации.

По экономическим соображениям и из-за того, что число вовлекаемых в эксперименты людей должно быть минимальным, всегда в эксперименте существует опасность, что вследствие обычной для рандомизации непредсказуемости число пациентов, рандомизированных в одну из групп, будет мало, и это снизит статистическую мощность эксперимента. Чтобы сбалансировать число пациентов в группах применяют прием *блочной рандомизации*. Для этого создают исчерпывающий набор *блоков*, которые затем рандомизируют. Например, для 2 вариантов вмешательства (активное лечение А и плацебо Р) есть шесть вариантов блоков размером 4: 1) ААРР; 2) РРАА; 3) АРАР; 4) РАРА; 5) АРРА; 6) РААР. Реализация каждого такого блока помещает в экспериментальные группы равное число пациентов. Блоки могут быть разной длины. Это может быть полезным для сокрытия механизма рандомизации, но может служить и иным целям. Например, в начале испытания, когда скорость включения пациентов в испытание в разных центрах неизвестна, используют минимальные блоки, чтобы в каждом центре иметь сбалансированные группы, даже если там рекрутируют мало пациентов.

Для того чтобы выборка была репрезентативной, обычно недостаточно простой рандомизации, в особенности если выбирается для исследования мало единиц, например населенных пунктов области. Так, из 20 могут быть случайным образом выбраны преимущественно крупные. В таких ситуациях обычно применяется прием *стратифицированной рандомизации*, когда из популяции выделяются населенные пункты с большой и малой численностью населения или приморские и континентальные, сельские и промышленные. После этого случайным обра-



зом из каждой группы населенных пунктов по таблице случайных чисел выбирают те, которые войдут в группу вмешательства, и те, которые войдут в контрольную группу. В клиническом испытании обычно изначально легко выделить подгруппы пациентов, у которых отклик на лечение должен различаться. Например, все гормонозависимые раки у женщин в до- и послеменопаузальном периодах текут по-разному. Соответственно, нужно стратифицировать выборку пациентов (от лат. *stratum* — слой) по признаку наступления климакса, и для каждой страты подготовить свою случайную последовательность вмешательств. Важно, что в этом случае обычно нельзя использовать простую рандомизацию внутри страты, так как в меньшей по размеру подгруппе легко возникает дисбаланс. Следует использовать вариант блочной рандомизации для каждой страты. Стратификация обычна при проведении многоцентровых клинических испытаний, поскольку всегда можно ожидать, что исходы различаются в разных лечебных учреждениях (странах, регионах).

В случае малого числа пациентов провести стратификацию затруднительно, и для устранения почти неизбежных случайных различий между группами проводят *рандомизацию с минимизацией различий*. При этом помещение очередного пациента в группу определяется случайно, но с учетом уже возникших ранее различий между группами. Так, если в группу вмешательства были помещены лица более старшего возраста, то очередной молодой пациент будет с высокой вероятностью помещен в группу вмешательства. Этот способ формирования групп схож с тем, который иногда практикуют в опытах на животных: перемещают животных из группы в группу, пока не добьются сходства групп по основным признакам. Важно, чтобы при минимизации сохранялся элемент рандомизации, иначе пропадают важные преимущества рандомизации — формирование групп, сходными по неизвестным признакам, и уменьшается возможность сокрытия вмешательства.

В отдельных случаях рандомизировать людей в группы вмешательства и контроля нельзя. Например, это невозможно при вмешательствах, нацеленных на врачей, на целую больницу или район. В таком случае единицей рандомизации становятся «группы», например районы, поликлиники. Исходы вмешательства оценивают по результатам ведения пациентов: правильности назначения гипотензивных средств (показатель процесса) или по частоте госпитализации при бронхиальной астме (показатель исхода). Такое распределение вмешательств называют «кластерная рандомизация» (англ. *cluster randomization*). Пла-



нирование и анализ данных такого исследования сложнее. Приходится по сравнению с рандомизацией отдельных пациентов заметно увеличивать число пациентов для того, чтобы выявить такой же по размеру эффект, т.е. статистическая чувствительность исследования ниже.

Конечно, рандомизация необходима не только для эксперимента на человеке в клинике, физиологической лаборатории или в исследованиях общественного здоровья; она точно так же необходима в эксперименте на животных, в исследованиях воспроизводимости лабораторных процедур и даже в повседневной работе с измерительными приборами: последовательность анализируемых проб, их размещение в кассетах для автоматического анализа должны быть случайными. Иначе возможные различия в результатах при включении прибора и при его многочасовой работе, в результатах измерения в центральных рядах и в крайних рядах ячеек могут приводить к систематическим различиям между сравниваемыми группами пациентов.

### 3.3.3. Характеристика вмешательства

Независимо от характера вмешательства (лечебное, диагностическое, профилактическое), оно должно быть ясно описано и носить стандартный, измеряемый характер. Вмешательство, технические характеристики которого неясны, невозможно воспроизвести, а потому сведения о его действенности не имеют ценности.

#### Дозиметрия и применение вмешательства

Применительно к некоторым вмешательствам дозировка не представляет проблемы: инъекция в установленное время обеспечивает введение в организм запланированного количества активного вещества. Применение пероральных форм введения уже приводит к трудностям в дозировке. Пациенты могут не полностью принимать суточную дозу, в случае ясно выраженных побочных эффектов могут уменьшать дозу для того, чтобы уменьшить побочные эффекты. Крайний вариант — вмешательства, которые трудно дозировать. К таким вмешательствам относится большинство хирургических вмешательств и манипуляций, выполняемых «руками». Обычно хирурги с трудом могут стандартизовать степень, в которой, например, затрудняется пассаж содержимого желудка после пластики. Манипуляции руками («хиропрактика») на позвоночнике практически не дозируются, так же как и воздействия



игл при иглоукалывании. Сократить вариабельность вмешательства помогает их стандартизация, например, по времени манипуляций, глубине проникновения, месту и методике внедрения иглы.

В ходе КИ обычно прекращают лечение, назначенное пациенту ранее. Срок, проходящий после прекращения прошлого лечения (период «отмывки») до начала КИ, устанавливают таким образом, чтобы действие принимавшихся ранее препаратов закончилось. Поскольку действие прошлого лечения или эффект отдачи после прекращения лечения могут быть длительными, иногда в КИ включают только пациентов, не имеющих опыта лекарственной терапии (англ. *treatment naïve*). В процессе КИ собирают информацию о том, как часто пациенты опытной и контрольной групп используют побочное, не назначенное в ходе испытания лечение. Если пациенты опытной группы используют дополнительные средства для лечения своего состояния (*коинтервенция*), то это может увеличить оценку действенности вмешательства. Если пациенты контрольной группы используют то же вмешательство, что предоставляется в экспериментальной группе (*контаминация, загрязнение*), то разница между группами уменьшается, и оценка действенности вмешательства будет занижена.

В испытании принимают меры к предотвращению приема непредписанных средств и к повышению исполнительности пациентов и персонала в выполнении предписанных протоколом действий. Одним из способов достижения комплаенса является проведение до испытания вводного этапа. На этом этапе выявляют пациентов, не соблюдающих режим, например, путем определения в моче веществ, введенных в препарат в качестве метки. Затем в испытание включают только исполнительных пациентов. Коинтервенция и контаминация почти неизбежны и должны учитываться при анализе данных.

## Плацебо

Первым вопросом относительно полезности нового вмешательства является вопрос о его действенности, т.е. способности уменьшить вероятность неблагоприятных исходов по сравнению с отсутствием вмешательства. Отсутствие вмешательства в контрольной группе может быть психологически неприемлемо для пациентов и может приводить к неисполнению ими режима испытания. Больные, которых не лечат, сами начинают лечиться. Поэтому пациентам контрольной группы дают вещество (проводят процедуры), которые неотличимы от активного вмешательства. Обычно плацебо — это лекарственная форма, лишен-



ная активного компонента, например таблетка, идентичная активной, но содержащая только наполнитель (обычно каолин, крахмал). Для полноты ослепления плацебо придают вид, максимально сходный с активным препаратом, и используют тот же путь введения.

Даже если полезность «обычного» лечения не доказана, врачи и пациенты могут с трудом воспринимать возможность замены его на плацебо. Тогда в качестве контроля применяют «обычное» лечение, стандартное в текущий момент. Считается неприемлемым применять плацебо и в некоторых специфических случаях, таких как исследования эффективности противозачаточных средств. Плацебо нельзя применять при болезнях, для которых есть эффективное лечение.

Фармацевтические компании предпочитают проводить плацебо-контролируемые КИ, поскольку для государственной регистрации действенность нового ЛС легче показать в сравнении с плацебо, чем в сравнении со стандартным лечением, даже минимально действенным (рис. 3.7). При сравнении с существующим ЛС надо или показать,

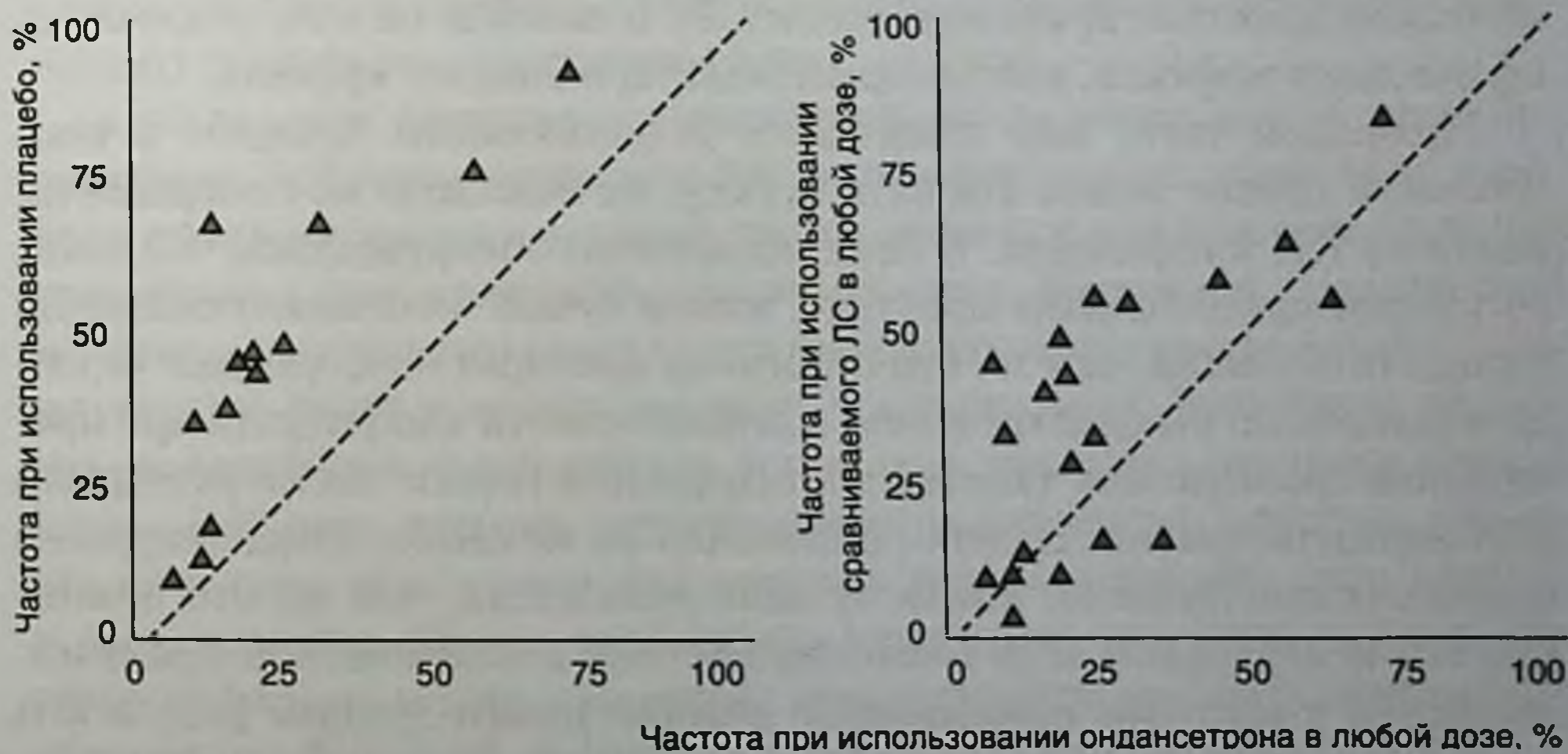


Рис. 3.7. Эффективность ондансетрона в сравнении с плацебо (а) и с любыми другими лекарствами (б) в любой дозе по показателю частоты позывов на рвоту и рвоты в течение 48 ч после операции. Диагональ — линия равной частоты. Каждая точка соответствует паре результатов в группе вмешательства и в группе сравнения. В то время как по сравнению с плацебо ондансетрон явно эффективен, у него нет преимуществ перед другими лекарствами (по данным Tramer M.R. et al. BMJ. 1998. № 317. P. 875–80. Такое представление результатов клинического испытания называют *графиком Лаббе* (L'Abbe plot))



что новое ЛС дает больший эффект, или что его эффект идентичен (не меньше) эффекту существующего ЛС. И то и другое требует резкого увеличения числа пациентов в КИ. Интерес к плацебо-контролируемым испытаниям — законный коммерческий интерес, но это объяснение не меняет негативного отношения врачей к оставлению пациентов в контрольной группе без лечения.

В соответствии с вековой традицией нередко полагают, что применение плацебо дает полезный эффект — «эффект плацебо». В действительности в систематическом обзоре тысяч РКИ показано, что плацебо — по сравнению с отсутствием вмешательства — оказывает незначительное влияние и лишь на исходы, которые отражают эмоции, субъективную картину болезни человека (качество сна, интенсивность боли). На клинически важные исходы прием плацебо не влияет (продолжительность жизни, длительность ремиссии, функциональный дефект и т.д.). Важно, что «эффект плацебо» присущ и активному ЛС, но не является специфическим эффектом ЛС, а в испытаниях изучают именно специфические эффекты вмешательств. Впоследствии в клинической практике врачи применяют ЛС в надежде на получение специфического эффекта, например антибактериального эффекта.

Примером того, как правильное использование плацебо в контрольной группе может вскрыть истину, не лежащую на поверхности, является КИ клофибрата. В этом испытании обнаружилось, что лица, регулярно принимавшие препарат, имели лучше конечные показатели липидного обмена, чем те, кто принимал препарат нерегулярно. Отсюда можно было бы сделать вывод о действенности клофибрата при правильном применении. Однако в контрольной группе также учитывали регулярность приема плацебо. Выявилось то же самое: лица, аккуратно принимавшие плацебо, имели лучшие показатели, чем те, кто принимал его неаккуратно. Это типичная картина в клинической практике: пациенты с высоким комплаенсом обычно имеют лучшие результаты лечения.

В опытах на животных возможен хирургический аналог плацебо — *ложные операции* (англ. *sham surgery*) для того, чтобы отделить воздействие операционной травмы на организм от главной манипуляции (резекции, пластики и т.п.). В исследовании на людях такой прием используют редко. В хирургических исследованиях контрольные группы часто отсутствуют или совершенно неудовлетворительны и распространен «исторический» контроль (см. выше). Соответственно, в хирургии годами сосуществуют



разные варианты операционной тактики, каждый из которых оценивается его создателями как лучший и о каждом есть более или чаще менее убедительные свидетельства испытаний с несовершенным контролем или вовсе без него. В конце XX в. слепые РКИ хирургических вмешательств стали проводиться. Так, было показано, что лапароскопическая холецистэктомия не приводит к более быстрому выздоровлению и реабилитации, чем мини-лапаротомическая холецистэктомия.

### Слепой метод

Главным условием эффективного испытания медицинского вмешательства является *ослепление участников* (англ. *blinding*), т.е. сокрытие от них, к какой группе в исследовании они принадлежат. Слепую оценку впервые применили А. Лавуазье и Б. Франклин в конце XVIII в. при экспериментальной проверке месмеризма — терапии гипнозом. Экспериментаторы тогда действительно завязывали глаза участникам, чтобы они не видели происходящего. Сначала слепая оценка применялась для выявления обмана, и лишь позднее — для предотвращения систематических ошибок в оценках. Иногда вместо термина «ослепление» применяют термин «маскирование» (англ. *masking*).

Если лечение известно исследователю, но не участнику, то такая организация КИ называется «слепое испытание» (англ. *single blind trial*). Если не замаскировать, к какой группе принадлежит пациент, он будет относиться к своему лечению по-особенному. Например, получающий «старое» лечение чаще будет им пренебрегать. Получающий новое вмешательство будет относиться к нему с энтузиазмом. Возможно, наоборот, получающие «проверенное временем» лекарство будут его ценить выше. Наконец, высока вероятность, что пациенты, которые знают, что получают неактивное или «старое» лечение, прервут свое участие в исследовании. Точно предсказать отношение участников к вмешательству нельзя, но обычно новые вмешательства рассматриваются как обещающие больший эффект.

Врачи тоже могут вносить систематическую ошибку. Например, выздоровление при новом вмешательстве может переоцениваться (рентгенолог найдет снижение плотности тени более быстрым при новом лечении). При возникновении у пациента неожиданных симптомов (головокружения, диспепсии) врач может прервать испытание у пациента, получающего новое лекарство, но может игнорировать эти симптомы у получающего плацебо. Поэтому желательно, чтобы все работающие



с пациентами не знали, кто какое лечение получает. Это называется «двойное слепое испытание» (англ. *double blind trial*). Отсутствие двойного ослепления в РКИ приводит к смещению оценки эффективности вмешательства — в среднем к завышению на 19%. При ознакомлении с результатами РКИ врач всегда должен спрашивать себя, было ли ослепление в испытании удовлетворительным.

*Тройными слепыми испытаниями* называют такие РКИ, в которых результат рандомизации скрыт не только от пациента и исследователя, но и от статистика, проводящего анализ данных. В таком случае статистик сравнивает результаты в двух группах, но их лечение открывается только по завершении анализа. Испытания, результат рандомизации в которых не скрывают, называют *открытыми РКИ*.

Ослепление врачей, оценивающих исходы, совершенно необходимо в исследовании болезней, при которых наблюдается самопроизвольное выздоровление, часты спонтанные изменения течения, обычны ремиссии, и когда оценка результатов лечения проводится по «мягким» субъективным критериям (настроение, качество сна, интенсивность изжоги, зуда, головной боли и т.п.), а не по объективным и надежным признакам (смертность, масса тела, биохимические и гематологические признаки и т.п.). Чем объективнее метод измерения эффекта, тем меньше необходимость в использовании слепой оценки. Ослепление всего персонала важно там, где велика роль сопутствующих вмешательству элементов лечения. Например, при КИ антибиотиков в лечении ран необходимо обеспечить, чтобы манипуляции над раной (обработки, перевязки) проводились одинаково всем пациентам. Например, при сравнении эндоскопической холецистэктомии и холецистэктомии через мини-разрез хирург проводил рандомизацию после ввода пациента в наркоз и обработки операционного поля, а по завершении вмешательства всем пациентам создавал идентичные наклейки, включая окрашивание кровью. Остальной персонал не знал о характере вмешательства.

В случаях когда нужно индивидуализировать лечение, например подбирать дозу антикоагулянта, один врач подбирает дозу по определенному алгоритму и соответственно этому изменяет дозу плацебо у пациентов контрольной группы. Другой врач наблюдает за пациентами, не зная о характере лечения и дозе. Такую структуру КИ называют «двойное слепое с открытым терапевтическим мониторингом».

В действительности нередко новое ЛС и контрольный препарат (традиционный или плацебо) вызывают столь различные фармаколо-



гические эффекты, что как пациенты, так и врачи догадываются о том, к какой группе принадлежит пациент. Например, если в одной из групп применяются  $\beta$ -адреноблокаторы, то брадикардия откроет врачу и пациенту характер вмешательства. Такое естественное ограничение возможностей скрыть характер вмешательства не лишает смысла ослепление, а требует еще более внимательного отношения к выбору контрольного вмешательства. Для того чтобы полнее ослепить участников такого исследования, можно применить «активное плацебо», т.е. препарат, который имитирует заметные эффекты испытуемого вмешательства, например сухость во рту. Таким образом участникам и исследователям затрудняется распознавание характера лечения, которое они получают.

Ослепление необходимо не только при испытаниях терапевтических вмешательств, но и при любых исследованиях в клинике, где дополнительное знание о пациенте может повлиять на поведение врача. Например, при испытании диагностических средств оценивающий результаты теста врач не должен иметь никакой информации, кроме результата теста. Иначе он будет сознательно или бессознательно искажать свои оценки в угоду совпадению с этой информацией.

## 3.4. ИСХОДЫ

Исходами («целевыми» признаками) называют события, по которым судят о прогрессировании болезни, эффективности лечения или других вмешательств. В медицине обычно на первом месте стоят такие исходы, как летальность (продолжительность жизни), длительность периода между обострениями (частота обострений), возникновение инвалидности (сохранение работоспособности), сохранение (возникновение) боли, уменьшение объективных проявлений болезни (отеки, ограничения подвижности, непроходимость), функциональные возможности (самообслуживание), удовлетворенность жизнью, видом своего тела. Первые исходы более традиционны для медицины, последние — привлекли больше внимания в последние десятилетия как показатели «качества жизни».

### 3.4.1. Оценка исходов

Выбор исходов болезни определяется целью и задачами исследования. Обычно в исследовании избирают самые важные исходы, характерные для



данной болезни. Например, при сердечной недостаточности это летальность и степень снижения функциональных возможностей (класс NYHA). Выявить изменения в таких исходах может быть затруднительно, так как в исследуемый интервал времени летальных исходов наблюдается мало.

Традиционным решением является замена клинически важных исходов на доступные (промежуточные, косвенные, суррогатные) (англ. *surrogate*). При сердечной недостаточности можно оценивать фракцию изгнания левого желудочка, так как уменьшение фракции изгнания отражает уменьшение функциональных возможностей пациентов. При лечении судорожных расстройств в качестве меры эффективности лечения можно использовать не частоту припадков, а изменения на электроэнцефалограмме, которые развиваются уже в ближайшие часы после начала приема препарата; при профилактике ишемической болезни сердца — вместо инцидента инфаркта, стенокардии, внезапной смерти — концентрацию холестерина липопротеидов низкой плотности в плазме. Использование таких «суррогатных» исходов может приводить к неверным результатам. Например, противоаритмические средства при инфаркте миокарда помогают нормализовать ритм, и это было основанием рекомендовать их профилактическое применение. Контролируемые испытания показали, что летальность у пациентов, получающих профилактически антиаритмические средства, даже повышена по сравнению с летальностью получающих плацебо. Только в США, где это лечение применялось наиболее активно, ежегодно погибали от этого лечения, по приблизительной оценке, 50 000 человек — больше, чем погибло американцев во вьетнамской войне.

Другой пример. При лечении остеопороза основным оцениваемым исходом является инцидент переломов, а суррогатным — минеральная плотность костей. Одни лекарства ожидаемым образом увеличивают минерализацию костей и снижают инцидент переломов (эстрогены, этидронат). Другие лишь увеличивают минеральную плотность, но увеличивают инцидент переломов (препараты фтора). Суррогатные оценки, например, минерализация костей, приемлемы только в случае, когда клинически важные исходы (частота переломов) нельзя оценить, или до тех пор, пока клинически важные исходы еще не изучены. Суррогатные оценки могут быть воспроизводимыми, но они могут неполно и даже неверно отражать основные клинически важные исходы.

Обычно исследователи стремятся к использованию «надежных» исходов, таких как смерть, смерть от определенной причины, воз-



никновение болезни, инвалидность. Такие исходы, как качество сна, выраженность дизурии, труднее использовать, поскольку по таким признакам люди оценивают себя сами и оцениваются другими неточно, с низкой воспроизводимостью. Точность и воспроизводимость «надежных» исходов значительно выше, но ее также не следует переоценивать. При установлении диагноза или определении причины смерти всегда существует неопределенность, на которую могут повлиять входящие факторы. Поэтому *оценка исходов всегда должна проводиться вслепую*. Данные диагностических исследований должны оцениваться специалистами, не знающими, к какой группе относится обследуемый, из архивных материалов данные должны выписываться вслепую, опрашиваться участники исследований также должны вслепую. Оценка изображений (рентгенограмм), гистологических препаратов должна проводиться двумя специалистами независимо и с привлечением третьего для согласования в случае расхождения оценок.

### 3.4.2. Качество жизни

Несмотря на традиционное внимание к исходам, отражающим клинически важные события, врачи никогда не игнорировали субъективные страдания больного и побочные эффекты лечения. Вместе с тем всегда существовало стремление достичь клинически важных исходов, несмотря на причиняемые пациенту страдания в ходе лечения. Это стремление было наиболее выражено у врачей, применяющих агрессивные, травматичные виды лечения, — у хирургов, онкологов. Именно в этих областях медицины острее всего была осознана необходимость наряду с клинически важными исходами оценивать и качество жизни, продлеваемой в результате лечения. Оценка качества жизни в соответствии с исходной идеей должна к оценке количества (продолжительности жизни) добавить ее качество. В экстремальном примере агрессивной противораковой химиотерапии небольшое увеличение продолжительности жизни может достигаться неприемлемыми страданиями в процессе лечения. Это не гипотетическая возможность, а реалии современной медицины. В 1990-е годы испытывался метод высокодозной химиотерапии рака молочной железы, который обещал увеличение продолжительности жизни больных. В результате испытаний выяснилось, что применение испытываемого протокола сопровождается тяжелыми побочными эффектами, и увеличение продолжительности



жизни, даже если оно возникает, покрывается увеличением летальности от побочных эффектов лечения.

Для того чтобы измерить, как болезни и медицинские вмешательства влияют на качество жизни пациентов, применяются специальные методы (шкалы): функциональные классы сердечной недостаточности NYHA, SF-36, шкала функциональных возможностей Карновского. Одни шкалы неспецифичны, и их можно применять к любому пациенту, другие приспособлены к особым дефектам, особенностям, возникающим, например, при астме, раке. Одни шкалы более нацелены на то, как пациент справляется с задачами повседневной жизни, в быту, другие более ориентированы на оценку самочувствия, удовлетворенности<sup>1</sup>.

Инструменты для оценки качества жизни обычно представляют собой сложные шкалы, итоговая оценка на которых получается в результате суммирования разнообразных сведений — об интенсивности боли, настроении, состоянии дыхания, способности помыться самостоятельно и т.д. Несмотря на то что многие шкалы на первый взгляд выглядят очень простыми, их разработка представляет собой достаточно сложный процесс. Поэтому исследователи используют стандартизованные шкалы качества жизни, специализированные для изучаемого вида патологии. По происхождению эти шкалы являются ординальными переменными. Вместе с тем число градаций на шкале обычно велико, и к оценкам качества жизни применяют те же статистические методы, что и к количественным признакам в надежде на робастность этих критериев.

В простейшем виде концепция качества жизни используется в виде вычисления среднего качества жизни, достигаемого при некотором виде лечения. Если часть больных умирает, то к ним нельзя применить оценку качества жизни. Поэтому создают интегральные оценки «количества и качества» жизни. Зарегистрированная продолжительность жизни умножается на коэффициент, отражающий ее качество. Предполагается, что есть некоторая шкала, на которой нулю соответствует смерть, а здоровая жизнь — единице. Тогда можно год жизни с параплегией приравнять, например, к 3 мес здоровой жизни. Прожитые годы умножают на коэффициент, и произведение используют как *годы жизни, скорректированные на ее качество* (QALY, англ. *quality adjusted life years*), или как *годы жизни, скорректированные на инвалидность* (DALY,

<sup>1</sup> Обширная коллекция инструментов для оценки качества жизни находится на <https://www.medicalalgorithms.com/>.



англ. *disability adjusted life years*). С такими скорректированными сроками выживания можно проводить те же вычисления, что и с годами жизни. К сожалению, применение интегральных оценок продолжительности жизни, скорректированной на ее качество, приводит к некоторым неприемлемым последствиям. Например, оказывается, что у инвалида («качество жизни»  $< 1$ ) профилактика ишемической болезни сердца или лечение ревматоидного артрита менее полезны, чем у трудоспособного человека (дают меньший прирост качества жизни).

### 3.4.3. Прекращение испытания

Длительность КИ планируется, исходя из числа участников и ожидаемого инциденса исходов, ожидаемого различия между вмешательствами (размера эффекта) и планируемой статистической значимости результата. Можно провести относительно короткое испытание, если включить много участников. Если не запланировать длительность испытания, а просто ожидать, когда различие между сравниваемыми группами станет статистически значимым, то это будет ошибкой. Повторное применение статистических критериев для проверки значимости различий приводит к увеличению вероятности ложноположительных выводов — вероятности обнаружения значимого различия там, где его нет. Это один из вариантов ошибки, возникающей при множественных сравнениях (см. разд. 2.5.6). Проведение испытания до момента, когда его результат станет статистически значимым, будет ошибкой, которую в обыденной жизни люди не допускают: никто не согласится играть на деньги до момента, когда выигрыш у противника будет больше 1000 рублей. Поэтому длительность испытания устанавливается заранее.

В длительных испытаниях устанавливают правила прекращения КИ, поскольку это может быть необходимым для обеспечения безопасности участников и при получении убедительных результатов в пользу одного из изучаемых вмешательств. Принято полагать, что после того, как преимущества одного из сравниваемых вмешательств установлены, изменяется ситуация: для участников КИ вмешательства становятся неравными, какими они были в момент начала испытания. Испытание должно быть прекращено, и всем участникам предоставлены информация о результатах и доступ к найденному более эффективному вмешательству.

Для прекращения КИ на основании достижения статистически значимой разности эффектов используются специальные статистические



критерии. Они учитывают возрастание вероятности ошибки первого типа при повторной проверке нулевой гипотезы. Это так называемые методы последовательного группового анализа. Решение о прекращении КИ выносит наблюдательный комитет, независимый от группы, проводящей испытание. Комитет оценивает промежуточные результаты в установленные моменты КИ и принимает решение об остановке испытания, если одно из вмешательств оказывается опасным или если одно из вмешательств демонстрирует значительные преимущества, превышающие по величине заранее заданные.

Заблаговременно формулируют *критерии исключения пациентов по ходу исследования*. У них могут развиваться новые заболевания, препятствующие их участию в исследовании; может обнаружиться потребность в прекращении режима испытания и даже в раскрытии характера лечения (до этого «слепого»); может обнаружиться, что пациент по каким-либо причинам не способен соблюдать установленные условия. Удаление пациентов из исследования должно быть определено заранее точными критериями и процедурой. В противном случае исключение отдельных пациентов по причинам «возникновения новых заболеваний» или «осложнений, препятствующих дальнейшему испытанию» может исказить и обесценить результат исследования.

Если по ходу исследования теряется много участников, то его конечные результаты нельзя отнести ко всем, включенным в исследование и тем более экстраполировать на популяцию. На рис. 3.8 приведены результаты гипотетического исследования по медикаментозному лечению ожирения. У лиц, длительное время участвовавших в исследовании,

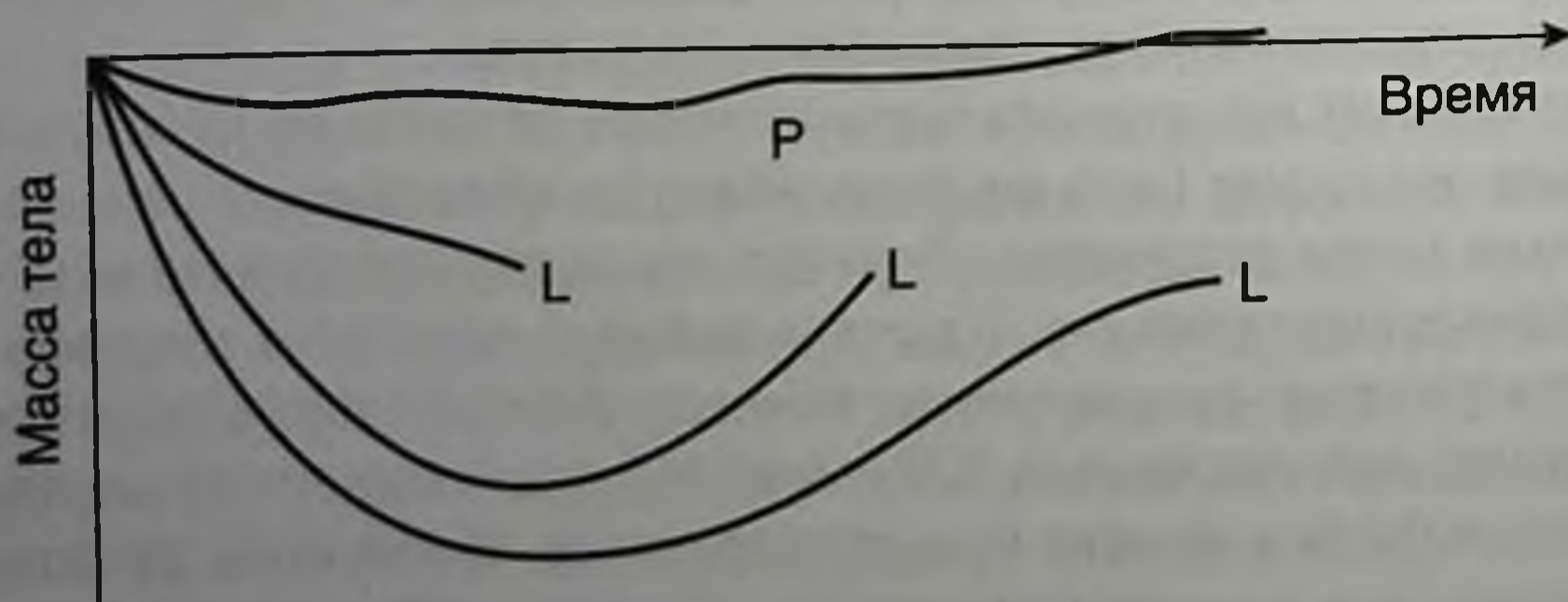


Рис. 3.8. Изменения массы тела у участников испытания, получавших плацебо (P) и испытуемое лекарство (L). Обратите внимание на различие в результатах лечения у участников, прекративших лечение в разные сроки



результат разительно отличается от тех, кто раньше прекратил участие в эксперименте. Если ситуация такова, то трудно ответить на вопрос о величине эффекта. Является ли эффект таким, как это наблюдается у участвовавших в испытании до конца? Конечно же, нет. В реальной практике он будет значительно меньше, поскольку у прекративших участие в испытании раньше эффект был меньше.

### Контрольные вопросы и задачи к разделам 3.1–3.4

1. В исследовании причин болезней главная разница между наблюдательным и экспериментальным исследованием состоит в том, что экспериментальное:
  - а) организовано проспективно;
  - б) группа вмешательства и контрольная имеют примерно равные размеры;
  - в) исследователь определяет, кто будет подвергаться изучаемому воздействию;
  - г) изучаемая и контрольная группы комплектуются по признаку наличия экспозиции изучаемому фактору в анамнезе.
2. Рандомизация при распределении в группу вмешательства и группу сравнения гарантирует, что:
  - а) распределение в группы будет случайным;
  - б) группы будут одинаковы по всем признакам, кроме вмешательства;
  - в) смещение в оценке исходов будет устранено;
  - г) эффект плацебо будет устранен.

Выберите все правильные ответы.

3. Планируется проведение исследования по оценке эффективности приема  $\beta$ -каротина для предотвращения рака легких у курильщиков. В расчете величины выборки надо принимать во внимание:
  - а) ожидаемую разницу в инциденте рака в сравниваемых группах;
  - б) инцидент рака в контрольной группе;
  - в) частоту, с которой участники будут выходить из исследования (теряться);
  - г) планируемую величину  $\alpha$ -риска;
  - д) планируемую величину  $\beta$ -риска.

Выберите одно правильное утверждение.



4. Реклама в медицинском журнале утверждает, что при применении рекламируемого лечения симптомы фарингита в течение 4 дней исчезают у 90% больных. Это утверждение об эффективности вмешательства является:
- а) правильным;
  - б) может быть неправильным, так как нужно использовать относительную частоту для описания эффективности;
  - в) может быть неправильным, так как нужно исследовать долгосрочные эффекты лечения;
  - г) может быть неправильным, так как нужно использовать сравнение с контрольной группой;
  - д) может быть неправильным, так как нужно исследовать статистическую значимость.
5. Двойной слепой эксперимент при испытании вмешательств нужен:
- а) для достижения сравнимости леченых и контрольных субъектов;
  - б) уменьшения случайной ошибки в формировании выборки;
  - в) уменьшения ошибок наблюдателей, оценивающих диагностические тесты и исходы;
  - г) уменьшения эффекта плацебо.
6. В рандомизированном испытании планируется использовать перекрестную структуру. Это:
- а) позволит устранить влияние предшествующего лечения на результат текущего вмешательства;
  - б) уменьшит влияние приема участниками испытания непредписанных препаратов;
  - в) уменьшит необходимость в контроле над приемом участниками правильной дозы предписанного препарата;
  - г) создаст необходимость учитывать остаточные эффекты от предшествующего лечения;
  - д) облегчит применение результатов в условиях различных больниц.

### 3.5. АНАЛИЗ РЕЗУЛЬТАТОВ ИСПЫТАНИЯ

Анализ результатов эксперимента зависит от его структуры и от того, как планируется учесть описанные выше сложности: потерю пациентов по ходу эксперимента, отказ их от назначенного лечения, обоснован-



ный переход к другому, не запланированному в испытании на основании рандомизации лечению.

Существуют несколько решений этой проблемы. Можно исходы у пациентов, которые не получили лечения, положенного по рандомизации, анализировать в зависимости от полученного лечения (англ. *as treated*). Это означает, что если рандомизированный в группу консервативного лечения пациент позднее подвергся хирургическому лечению, то он перемещается в группу хирургического лечения (рассматривается как получивший хирургическое лечение). Наоборот, рандомизированные в группу хирургического лечения, но не подвергавшиеся ему, анализируются в составе группы, в которой планировалось медикаментозное лечение. Недостаток этого анализа в том, что от инвазивного лечения переходить на консервативное лечение могут преимущественно пациенты с большей тяжестью болезни. Соответственно, исходы консервативного лечения могут быть ухудшены, а инвазивного — выглядеть лучше, чем они есть на самом деле. Иногда наоборот, поскольку инвазивное лечение кажется «необходимым», пациент отказывается от лечения, назначенного по результатам рандомизации. Например, поскольку высокодозная химиотерапия в середине 1990-х годов выглядела как самый эффективный способ лечения рака молочной железы, многие пациентки искали пути к ее получению, несмотря на то что испытания еще не дали убедительных результатов. В 1999 г. выяснилось, что у такого лечения — крайне тяжелого для пациенток — нет преимуществ.

Можно исключать пациентов, отклонившихся от протокола лечения, из анализа. Это называется анализом в соответствии с выполнением протокола (англ. *per protocol*). Это тоже несовершенный метод, так как, например, побочные эффекты могут быть более выражены у более тяжелых больных, и их исключение изменит исходы в опытной группе. В группе плацебо таких случаев непереносимости лечения будет обычно меньше. Результат лечения новым методом в таком случае будет завышен.

Наиболее правильным является метод анализа в соответствии с назначением по результатам рандомизации (англ. *intention to treat analysis*, ИТТ). При таком анализе исходы лечения рассматривают у больных в группах, выделенных при рандомизации, независимо от того, какое лечение они фактически получили. Главное достоинство метода в том, что он моделирует реальную практику: врач выбирает лечение сегодня,



не зная точно, как придется вести больного завтра в связи с осложнениями, особенностями прогрессирования болезни и другими событиями. Для врача желательно знать, что с точки зрения перспективы больного выбор сегодня правильный. Одновременно метод анализа в соответствии с назначенным лечением прост для понимания и дает, видимо, несколько заниженную оценку фактической разницы между сравниваемыми методами. Это важно для осторожных практических выводов из результатов испытаний.

В случае если не все пациенты подверглись лечению по программе, данные эксперимента должны быть проанализированы дважды: исходя из предположения, что все пациенты получили лечение, и только применительно к пациентам, получившим полное лечение. В свою очередь, результаты у тех, кто вышел из исследования, можно анализировать по срокам нахождения в исследовании (см. рис. 3.8). Такой анализ разными способами называют *анализом чувствительности к смещениям* (англ. *sensitivity analysis*). Он выявляет, насколько разный способ расчета, разный анализ фактических данных сказываются на результате.

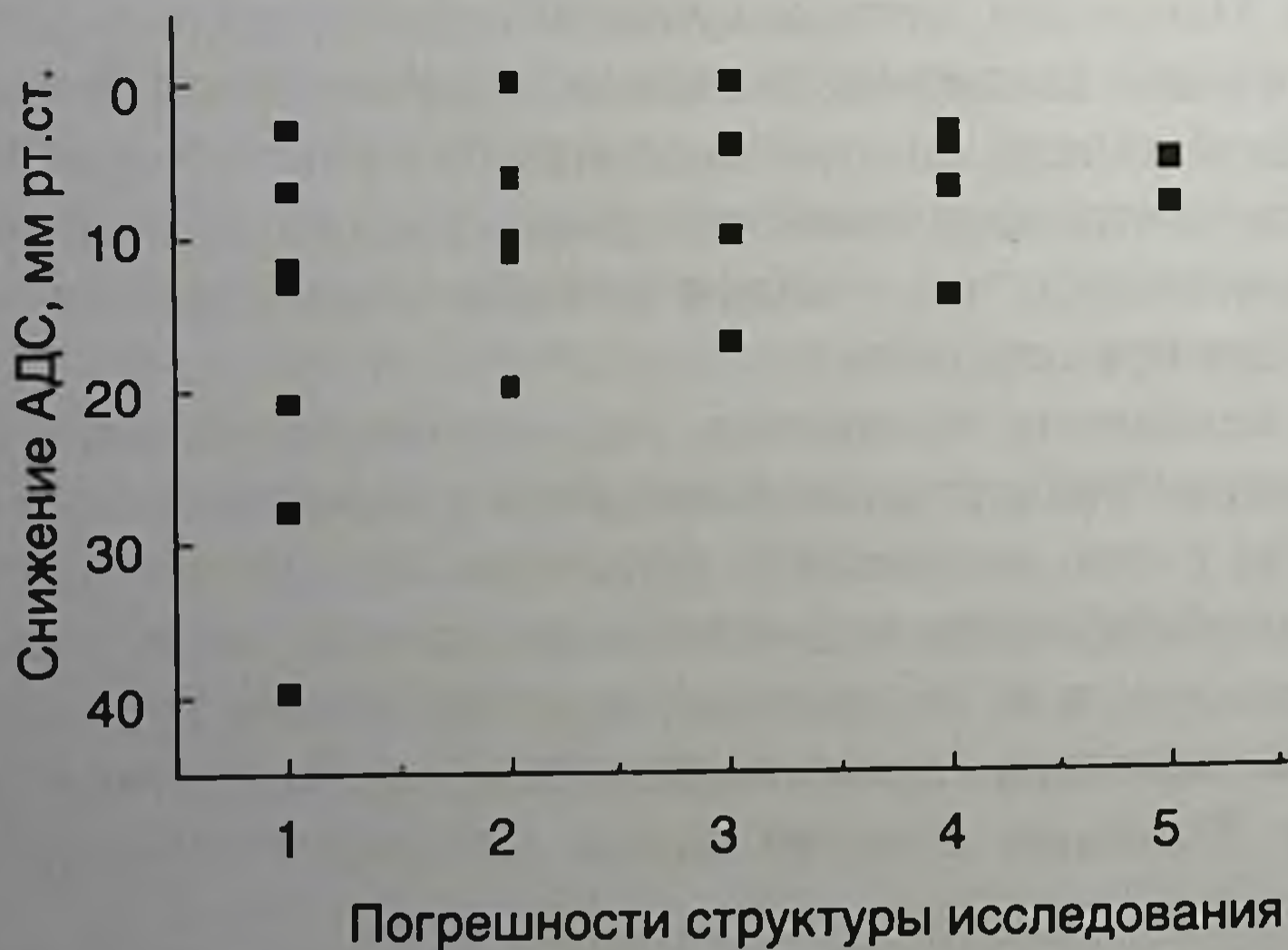


Рис. 3.9. Зависимость эффекта вмешательства от методологической строгости исследования: пример исследований снижения артериального давления при физической тренировке: 1 — исследования без контрольной группы, рандомизации и ослепления; 2 — без ослепления; 3 — без ослепления и рандомизации и с коинтервенцией; 4 — без рандомизации; 5 — без больших погрешностей (по данным Arroll B., Beaglehole R. J. Clin. Epidemiol. 1992. № 45 (5). P. 439–47)



Если результат получается одинаковым, несмотря на разные способы расчета, — значит, результат нечувствителен к влиянию привходящих факторов. Если выявляются различия, то их можно понять и объяснить на основе такого анализа.

Исследования, в которых авторы не использовали необходимых для устранения систематических ошибок приемов, могут давать как заниженные, так и завышенные оценки величины эффекта. Чаще величина эффекта завышается, в особенности, если исследование не имеет контрольной группы. Именно в таких исследованиях было обнаружено значительное снижение артериального давления при артериальной гипертензии в результате физических тренировок (рис. 3.9). Последующие обобщенные анализы РКИ только высокого качества показали, что регулярные физические нагрузки приводят лишь к небольшому (1–3 мм рт.ст.) снижению давления.

### 3.5.1. Анализ в подгруппах

Нередко результаты клинических испытаний анализируют по отдельности для подгрупп испытуемых, например отдельно для мужчин и женщин, отдельно для пожилых или отдельно для имеющих сопутствующую болезнь. Такой анализ должен предусматриваться на этапе планирования эксперимента. Если этот анализ не дает особенных результатов в подгруппах, то он лишь подтверждает, что общий результат может быть применим к таким подгруппам. Если запланированный анализ обнаруживает, например, у больных старшего возраста особенный эффект лечения, то это дает основания для особенного отношения к данной группе больных.

Проблема подобного анализа состоит в том, что его иногда применяют после того, как испытание завершено и не выявило ожидаемого эффекта. Тогда делят пациентов на разнообразные группы и проверяют, нет ли статистически значимого эффекта в какой-то из групп. Это яркий пример некорректного обращения с экспериментальными данными.

Анализ в подгруппах приемлем только, если:

- ▶ запланирован до начала эксперимента и заранее имел ясный смысл;
- ▶ численность всех экспериментальных подгрупп была достаточной;
- ▶ пациенты распределялись в подгруппы с помощью надежных критериев и по признакам, измеренным до вмешательства;



- ▶ статистический анализ выполнялся с коррекцией на множественные сравнения;
- ▶ при сравнении групп проводилась коррекция на сопутствующие различия между группами, помимо изучаемого различия.

В любом случае качественные различия между подгруппами вряд ли истинны. Под качественными понимают различия в направлении эффекта, например в одной группе снижение инциденса, а другой — повышение. Одновременно следует полагать, что между подгруппами пациентов всегда существуют количественные различия, но они обычно слишком малы для статистически значимого проявления в эксперименте. Как правило, результаты анализа в подгруппах не могут рассматриваться как свидетельства эффективности (или неэффективности) вмешательства в аналогичной части популяции, пока это не будет подтверждено в специально организованном исследовании.

Анализ в подгруппах повышает не только вероятность ошибки первого типа, но и вероятность ошибки второго типа. Если испытание выявило эффект во всей совокупности пациентов, то при анализе в подгруппах, вероятно, окажется, что в каких-то подгруппах эффект не будет статистически значимым. Такое отмечалось в реальных больших клинических испытаниях: испытание стрептокиназы для лечения инфаркта миокарда показало снижение летальности на 20%, но анализ в подгруппах не выявил снижения летальности в случае начала лечения позднее 6 ч у пациентов старше 65 лет и при заднем инфаркте. Последующие испытания не подтвердили этих результатов.

### 3.5.2. Статистическая и клиническая значимость результатов

Результаты терапевтического или хирургического эксперимента представляются в цифровом виде. Лишь иногда можно встретить сообщения, в которых авторы исследования ограничиваются заявлением о наличии блестящих результатов при применении некоторого лекарства (варианта операции). Такие сообщения следует игнорировать. Результаты клинического эксперимента даже простейшей структуры необходимо рассматривать детально, для того чтобы убедиться в том, что результат действительно существует, и что он действительно значим для практики.

*Статистически значимо* то, что действительно существует с высокой вероятностью. Обычно это обозначают  $P < 0,05$ , что означает: от-



существование различия имеет низкую вероятность — меньше 5%. Клинически значимо то, что своими размерами (например, величиной снижения смертности) убеждает врача в необходимости изменить свою практику в пользу нового образа действий. Если различия оказались статистически незначимыми, то следует проверить, достаточна ли численность обследованных.

### Показатели величины эффекта

Для выражения результата вмешательства, величины эффекта используют ряд показателей. Их вычисление иллюстрирует табл. 3.2.

Таблица 3.2. Способы вычисления относительных показателей эффективности вмешательства

Лечение	Неблагоприятный исход		Всего
	Наблюдался	Отсутствовал	
П	A	B	A + B
Л	C	D	C + D
Риск при лечении П			$P_p = A/(A + B)$
Риск при лечении Л			$P_l = C/(C + D)$
Абсолютное снижение риска (АСР; англ. <i>absolute risk reduction, ARR</i> )			$АСР = A/(A + B) - C/(C + D)$
Число пациентов, подвергаемых лечению, на один предотвращенный неблагоприятный исход (ЧПЛП; англ. <i>number needed to treat, NNT</i> )			$ЧПЛП = 1/АСР$
Относительный риск (ОР; англ. <i>relative risk, RR</i> )			$ОР = [C/(C + D)]/[A/(A + B)]$
Различие риска (снижение относительного риска СОР; RR, англ. <i>reduction, RRR</i> )			$СОР = 1 - RR$
Шансы на неблагоприятный исход при лечении П			$\Omega = A/B$
Шансы на неблагоприятный исход при лечении Л			$\Omega = C/D$
Отношение шансов (ОШ)			$ОШ = (C/D)/(A/B)$



Сопоставляя два вида лечения, сравнивают вероятность (риск) неблагоприятных исходов у больных, леченых лекарством Л и леченых плацебо П. Поскольку подразумевается, что лечение приводит к снижению риска в сравнении с плацебо, постольку разность между этими вероятностями называется «абсолютным снижением риска»<sup>1</sup>. Снижение риска лучше всего отражает ту пользу, которую дает метод в конкретной клинической ситуации и с наименьшими искажениями. Этот показатель иногда называют *терапевтической пользой* (англ. *therapeutic benefit*).

Для оценки эффективности вмешательства показатель величина снижения риска плох — он зависит от инциденса в популяции: если в популяции инцидентс высок, то размер эффекта, выраженный абсолютным снижением риска, больше. Поэтому для целей сравнения результатов нескольких КИ применяют относительные показатели — *относительный риск* или *отношение шансов*.

Статистический анализ КИ начинается с формулирования двух гипотез, одну из которых традиционно называют *нулевой* (сравниваемые вмешательства одинаково эффективны, а наблюдаемое различие есть лишь случайные вариации) (см. также разд. 1.8), а вторую — *альтернативной*. В табл. 3.3 приведены четыре ситуации, которые могут возникнуть в зависимости от того, как соотносится оценка врача с истиной.

Таблица 3.3. Два типа ошибок

Выбор врача	Истина	
	$H_1: L \neq P$	$H_0: L = P$
$H_1: L \neq P$	Нет ошибки	Ошибка типа I
$H_0: L = P$	Ошибка типа II	Нет ошибки

$H_0$  — нулевая гипотеза,  $H_1$  — альтернативная гипотеза.

Врач принимает нулевую гипотезу, и она верна: ошибки нет.

Врач считает лечение лучшим, чем плацебо, в то время, как они равноэффективны: он впадает в ошибку типа I ( $\alpha$ -риск, выражаясь статистическим языком).

<sup>1</sup> При неуспешном лечении «снижение» будет со знаком минус. Поскольку ятрогенный вред не менее важен, чем польза, для целей оценки вреда рассчитывают число пациентов, подвергаемых лечению на один случай нанесения вреда (ЧПЛВ), — аналог показателя ЧПЛП. См. с. 365.



Врач принимает нулевую гипотезу, а лечение лучше, чем плацебо: он впадает в ошибку типа II ( $\beta$ -риск).

Врач отвергает нулевую гипотезу и лечение действительно лучше, чем плацебо: ошибки нет<sup>1</sup>.

Статистический анализ результатов испытания может выполняться с помощью разных методов. В простейшем случае с помощью критерия  $\chi^2$  сопоставляются частоты в ячейках латинского квадрата, полученные в испытании (табл. 3.4) с ожидаемыми случайно. Ожидаемые случайно частоты вычисляются перемножением частот на полях таблицы. Детально ход вычислений критерия  $\chi^2$  показан в разд. 2.4.2. Вероятность ошибки типа I  $P < 0,0001$  означает, что получение такого или большего различия между вмешательствами Л и П в случае справедливости нулевой гипотезы имеет вероятность менее 0,0001.

Таблица 3.4. Вычисления относительных показателей эффекта вмешательства. В скобках — ожидаемые частоты

Лечение	Неблагоприятный исход		Всего
	Наблюдался	Отсутствовал	
Л	29 (46)	71 (54)	100
П	63 (46)	37 (54)	100
Всего	92	108	200

$$\begin{aligned} \chi^2 &= \sum[(O - E)^2 / E] = \\ &= (29 - 46)^2 / 46 + (63 - 46)^2 / 46 + (71 - 54)^2 / 54 + (37 - 54)^2 / 54 = \\ &= 6,28 + 6,28 + 5,35 + 5,35 = 23,26, \end{aligned}$$

что при числе степеней свободы 1 дает основания отклонить нулевую гипотезу с вероятностью ошибки первого типа менее 0,0001.

Снижение риска при лечении Л сравнительно с плацебо П составляет  $0,63 - 0,29 = 0,34$ . ДИ вычислим по формуле:

<sup>1</sup> Чтобы запомнить классификацию ошибок, используют мнемоническое правило OPTimist. Optimist совершает ошибку первого типа, с энтузиазмом принимая гипотезу об эффективности вмешательства. Pessimist (P стоит на втором месте) совершает ошибку второго типа, скептически отвергая альтернативную гипотезу и склоняясь к мнению о недейственности вмешательства.



$$ACR \pm 2 \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}};$$

$$0,34 + 0,07 = 0,41,$$

$$0,34 - 0,07 = 0,27.$$

То есть 95% ДИ для абсолютного снижения риска (АСР) простирается от 0,27 до 0,41.

Относительный риск для нашего примера составляет 0,46; 95% ДИ 0,39–0,53. Иными словами, с вероятностью 95% при лечении Л по сравнению с плацебо П риск неблагоприятного исхода снижается на 47–61%.

### Оценка клинической значимости

Отсутствие статистической значимости различия в исходах на общепринятом уровне 0,05 не означает, что мы обязаны исключить вероятность различия. Равным образом наличие статистически значимого различия не обязывает принимать различие. Лишь для статистического анализа данного гипотетического испытания вероятность ошибки первого типа менее 0,0001 дает окончательный ответ. Врач подходит к результатам нового испытания иначе, он спрашивает себя: какова вероятность недейственности испытанного вмешательства теперь, когда я прочитал этот отчет? Если врач имел основания сомневаться в эффективности вмешательства, например, лечения близорукости гомеопатическими средствами, то вряд ли результаты одного испытания, даже статистически значимые, радикально изменят его взгляды (см. также врезку на с. 167).

В случае если неблагоприятный исход наблюдается редко и лечение приводит к существенному снижению риска по сравнению с контролем, показатели отношения шансов (ОШ) и относительного риска (ОР) имеют примерно одинаковые величины. Если же исходы наблюдаются часто, чаще 0,1, то ОШ и ОР различаются. ОШ и ОР — вторичные относительные показатели, используемые для того, чтобы сделать оценку эффекта вмешательства менее зависимой от условий испытания, от исходного риска, на фоне которого проводится вмешательство (риска в контрольной группе). В относительных показателях результат обычно более впечатляющий.

Без учета частоты исходов в контрольной группе мы не можем себе представить, насколько велика польза для наших больных, используя



только показатели ОШ и ОР. Например, в многоцентровом испытании, известном под названием *Higulog*, было показано, что применение бивалирудина в сравнении с гепарином приводит к снижению частоты повторных инфарктов в течение первых 96 ч на треть (ОР = 0,7 [95% CI 0,56–0,87],  $p = 0,001$ ). Если же обратиться к исходным данным, то оказывается, что в группе, леченной гепарином, повторный инфаркт в эти сроки наблюдался у 2,3% пациентов, а в группе леченных бивалирудином — 1,6%. Аналогичный по размерам результат обнаружен в *Primaгу Prevention Project*: применение ацетилсалициловой кислоты с профилактическими целями дало снижение смертности на 44% по сравнению с плацебо. В показателях абсолютного риска применение ацетилсалициловой кислоты выглядит не столь привлекательным: при смертности в контроле 1,4% прием ацетилсалициловой кислоты позволил снизить ее до 0,8%. Малая абсолютная величина эффекта (малая терапевтическая польза) не означает, что вмешательство должно быть обязательно отклонено. Если оно безвредно и экономически доступно, то даже небольшое терапевтическое преимущество может быть привлекательным.

### **Сколько больных надо лечить, чтобы получить один желаемый исход?**

С точки зрения перспективы применения метода лечения удобен показатель ЧПЛП. Он переводит относительные величины в число пациентов, которое надо лечить, чтобы предотвратить один неблагоприятный исход. Это самый наглядный показатель, рекомендуемый для использования врачам.

Для методов лечения в клинике, которые принято считать стоящими использования, этот показатель составляет десятки пациентов (которых необходимо лечить, чтобы избежать одного неблагоприятного исхода или осложнения). При сопоставлении исходов применения двух вмешательств ЧПЛП выражается числом пациентов, которых надо лечить, чтобы получить один дополнительный благоприятный исход по сравнению с контрольным (старым) лечением. Поскольку при вычислении ЧПЛП учитывается не только разница инцидентов между двумя группами, но и частота исходов в контрольной группе, ЧПЛП будет всегда больше для более редких исходов.

При какой величине ЧПЛП можно принимать решение о целесообразности лечения? Этот порог зависит от многих переменных: опасности осложнений, которые наступят, если не лечить, стоимости лечения



этих осложнений, тяжести и частоты осложнений самого лечения, от отношения пациентов к осложнениям данного вида. Если речь идет о снижении частоты кожного зуда у пациентов с сахарным диабетом, то 10% уменьшение частоты — не слишком впечатляющий результат. Если же речь идет о снижении частоты ампутаций у таких больных, то это замечательный результат.

### Прагматические испытания

Повседневная практика отличается от «стерилизованных» условий РКИ. Поэтому после испытания лечения на относительно однородной совокупности пациентов на действенность (это иногда называют «объяснительное испытание», т.е. проведенное для обнаружения лечебного эффекта) обычно проводят испытание на группах пациентов, которых уже не подбирают строго и в условиях менее строгих — иногда даже без «ослепления», без «чистого» применения изолированного испытуемого метода. Зато может испытываться вся стратегия применения метода, включая другие сопутствующие манипуляции. Этот вид испытаний называют «прагматическим», подчеркивая, что его результаты лучше отражают то, что получается в обычной практике и нацелено на уточнение рекомендаций для клинической практики.

Проведение РКИ в продуманных и ясно описанных условиях является основанием для того, чтобы результаты РКИ можно было переносить на условия другой практики, например, в другой стране, в больнице другого типа. Важно, что не существует ни одного примера, когда результаты РКИ качественно отличались бы от результатов применения метода в обычной практике. Возможны лишь количественные различия. *Общее правило для решения вопроса о применимости результатов РКИ гласит: только в случае ясного и значительного отличия условий конкретной практики от условий клинического испытания следует считать его результаты неприменимыми.*

### 3.5.3. Ограничения на испытание вмешательств

Исчерпывающего доказательного основания для всей медицинской практики (каждого вмешательства) не может быть. В некоторых случаях пациенты просто не хотят рандомизации (предотвращение беременности, вскармливание детей). В других случаях болезнь протекает в несколько различающихся стадий и имеет несколько вариантов, и нереалистично



предполагать, что в обозримое время могут быть испытаны все применяемые в данной области и новые вмешательства. В третьих случаях болезнь встречается редко, и не существует достаточного числа пациентов, чтобы провести испытание. Для таких болезней разрабатывается мало лекарств, поскольку их невозможно продавать в достаточных для окупаемости количествах. Чтобы стимулировать разработку и производство таких лекарств-сирот (англ. *orphan drugs*), специальными законами созданы исключительно благоприятные условия, облегчающие их регистрацию.

Применительно к распространенным и опасным болезням производители разрабатывают все новые и новые лекарства. Многие из них изначально не обещают большей эффективности, чем существующие. Введение в практику еще одного подобного лекарства законно. Однако врач не имеет морального права приглашать к участию в РКИ пациентов, если ответ на поставленный вопрос уже ясен. РКИ оправдано только тогда, когда есть два или более различающихся способа ведения больных, и преимущества и недостатки их неясны. Врач, приглашенный для участия в проведении РКИ, может не знать технических деталей, например статистического анализа данного РКИ, но он должен быть уверен, что фармацевтическая компания или правительственное агентство вручили проведение РКИ квалифицированному управляющему комитету, который обладает всеми данными и полнотой власти оперативно опубликовать результаты. К сожалению, производители лекарств систематически публикуют результаты только тех РКИ, которые выгодны для маркетинга.

Чтобы помешать сокрытию испытаний, создаются специальные регистры КИ (например, <https://clinicaltrials.gov/>). В будущем полная регистрация всех КИ до их начала и регистрация результатов должны ослабить публикационное смещение, ослабить зависимость медицинского сообщества от опубликования результатов. Результаты зарегистрированного испытания так или иначе станут достоянием общества. В Европейском союзе появилась возможность получать доступ к данным клинических испытаний, что в прецеденте с препаратом Тамифлю позволило установить его крайне невысокую эффективность<sup>1</sup>.

---

<sup>1</sup> В результате селективного предоставления информации из клинических испытаний спонсора оселтамивир занял исключительное место в ряду противовирусных препаратов, что обеспечило рекордные многолетние продажи, в том числе в государственные резервы на случай эпидемии. История расследования и анализа данных опубликована: <https://www.cebm.net/2016/06/timeline-events-since-publication-cochrane-review-neuraminidase-inhibitors/>.



### 3.5.4. Изучение побочных эффектов вмешательства

Под *побочным эффектом* (англ. *side effect*) понимается любой эффект, который не является целенаправленно используемым эффектом данного лекарства. Под *неблагоприятным эффектом* понимается любой нежелательный эпизод во время или вследствие применения лекарства. Неблагоприятные эффекты классифицируют следующим образом: тип А — эффект, являющийся предсказуемым продолжением известного фармакологического действия лекарства; тип В — эффект, который не типичен для действия лекарства. Серьезными (тяжелыми) неблагоприятными эффектами называют ведущие к смерти, угрожающие жизни состояния или состояния, ведущие к госпитализации или продлению госпитализации, продолжительной или значительной инвалидности или возникновению врожденной аномалии. Частота неблагоприятных эффектов от медицинских вмешательств и, в частности, от фармакотерапии значительна и быстро растет во всех странах (рис. 3.10).

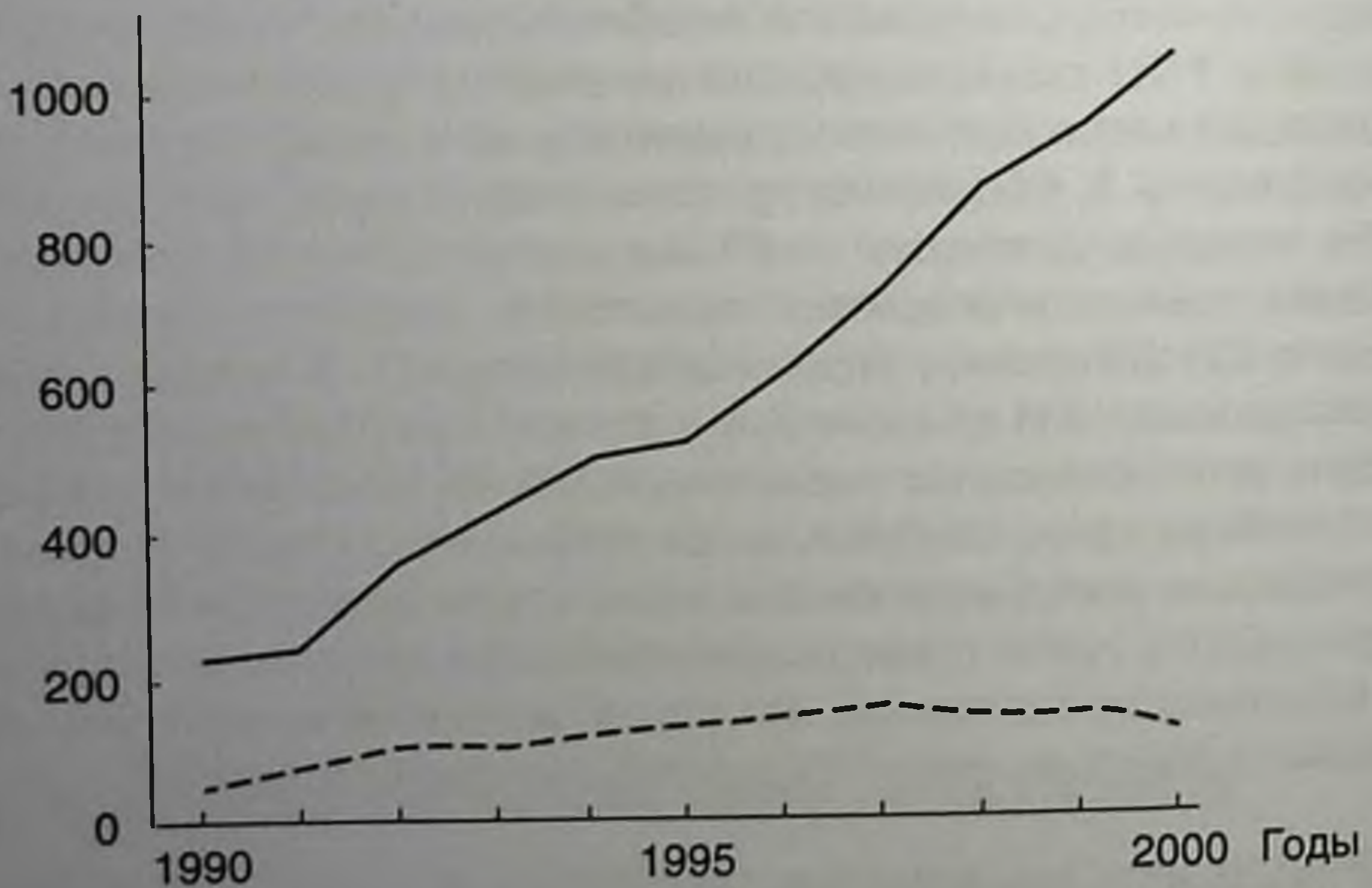


Рис. 3.10. Число смертей в Англии и Уэльсе в 1990–2000 гг. вследствие ошибок в использовании лекарств (назначение неправильного лекарства или неправильной дозы лекарства, пунктир) и вследствие непредсказуемых неблагоприятных эффектов (линия). Источник: Audit Commission, NHS, UK



Для выражения частоты осложнений и неблагоприятных побочных эффектов, возникающих при лечении, используют показатель ЧПЛВ, или число пациентов, подвергаемых лечению на один вредный исход (табл. 3.5).

Таблица 3.5. Вычисления относительных показателей эффекта вмешательства

Лечение	Неблагоприятный исход		Всего
	Наблюдался	Отсутствовал	
П	A	B	A + B
Л	C	D	C + D
Абсолютное повышение риска (АПР) (англ. <i>absolute risk increase, ARI</i> )		$APR = C/(C + D) - A/(A + B)$	
Число пациентов, подвергаемых лечению, на один вредный исход (ЧПЛВ) (англ. <i>number needed to harm, NNH</i> )		$ЧПЛВ = 1/APR$	

В данном примере предполагается, что осложнения (вред) возникают от активного лечения Л. Например, при профилактике инфаркта миокарда у больных ишемической болезнью сердца ацетилсалициловой кислотой можно рассчитать ЧПЛП для предотвращения одного случая инфаркта и число пациентов, подвергаемых лечению, на один дополнительный вредный исход (ЧПЛВ) для возникновения одного случая кровотечения.

Побочные эффекты могут быть изучены, оценены по тяжести и частоте одновременно с испытанием лечебного эффекта. Обычно побочные эффекты и осложнения у вводимых в практику лекарств редки. Соответственно эти осложнения трудно выявить в обычных клинических испытаниях. Чтобы выявить осложнение, встречающееся у одного пациента из 400, недостаточно провести исследование на 500 пациентах — вероятность выявления одного случая осложнения будет только 30% (примерно). Поэтому уже после одобрения нового лекарства и его разрешения к применению продолжают наблюдения за осложнениями при применении лекарства, поддерживаются регистры всех осложнений, при возникновении подозрения проводятся исследования для проверки гипотезы о возможной связи осложнения с применением лекарства.



Самыми слабыми свидетельствами возможного вреда являются описания отдельных случаев, надежнее — ИСК, лучше — когортные исследования, самые надежные — контролируемые испытания. Примеры того, как обсервационные исследования были использованы для изучения неблагоприятного побочного эффекта лекарства, приведены во врезке на с. 29.

Основным способом изучения побочных и неблагоприятных эффектов являются ИСК. Чтобы сделать результаты изучения побочных эффектов лекарств с помощью ИСК более надежными, в ИСК следует использовать данные из надежных регистров осложнений. Создание таких регистров — трудное дело, так как осложнения при медицинских вмешательствах связаны с ответственностью за них медицинского работника. Неудивительно, что медицинские работники не полностью сообщают об осложнениях. В конце XX в. в США, как показывают исследования, основанные на учете осложнений, ятрогенные осложнения возникают не реже чем в 2–3% случаев медицинских вмешательств. Это означает, что от осложнений, связанных с медицинскими вмешательствами, погибает больше людей, чем в дорожных авариях. Более пессимистичные оценки получены при непосредственном наблюдении за медицинской практикой. Выявляется, что в хирургических отделениях частота осложнений превышает 40% госпитализируемых, из которых половина — серьезные. Большая часть осложнений является не ошибками хирурга, а следствием несовершенства системы организации помощи.

### **3.5.5. Приложение результатов испытаний к отдельному пациенту**

Для многих начинающих врачей составляет проблему сам принцип переноса «статистических данных» на отдельного пациента. Кажется, если лечение приводит к выздоровлению так же, как и плацебо, в 60% случаев, то: 1) лечение полезно для 6 пациентов из 10 и 2) плацебо полезно такому же числу пациентов. Это — грубая ошибка. В действительности такие сведения ясно свидетельствуют, что данное лечение бесполезно, а плацебо бесполезно по определению (если не считать обман пользой).

Второе схожее заблуждение состоит в том, что даже если польза вмешательства мала, то отказ от него — это оставление пациента без лечения. Например, в испытаниях скрининга на рак молочной железы (РМЖ) (см. подробнее разд. 5.2.3) было показано некоторое снижение смерт-



ности от РМЖ при скрининге только в группах женщин старшего возраста. Соответственно американские профессиональные организации рекомендовали профилактическую маммографию только женщинам старше 50 лет. Если процедура не рекомендована, то в США это означает, что ее проведение не будет оплачиваться страховой компанией. Для неспециалиста ситуация выглядит так, как будто имеет место «экономия на здоровье женщин». Якобы из-за такой экономии на основании статистических данных у «отдельной конкретной женщины» не будет проведена маммография и не будет своевременно выявлен рак. На основании подобных соображений Конгресс США оказал давление на онкологическое сообщество, и рекомендации были изменены — возраст снижен до 40 лет. В 2016 г. решение, принятое по ненаучным основаниям, было исправлено. Это яркий пример того, что политика в системе здравоохранения формируется не только на основе научных данных.

Главная ошибка таких рассуждений состоит в следующем. Если врач считает, что «данная конкретная женщина» не отличается существенно от женщин, на которых проводили испытания маммографического скрининга, то на нее распространяются все сведения о полезности (бесполезности) процедуры и ее вреде. Других оснований думать, что она получит от маммографии пользу, нет. Более того, поскольку испытания показали, что для таких женщин пользы нет, то это означает, что у данной женщины *нет оснований ожидать пользы независимо от результатов маммографии*. Вред же от медицинских вмешательств всегда реален и ощутим (см. рис. 2.24, 3.10). Основной вред скрининга на РМЖ заключается в огромном числе биопсий, выполняемых вслед за маммографией. При этом чаще всего РМЖ не подтверждается, но болезненная процедура и напряженное ожидание результатов гистологического анализа на время делают жизнь женщины ужасной. Подчеркнем, что в нашем примере речь идет о скрининге на РМЖ, а не о процедурах диагностики у женщин, в молочной железе которых обнаружен узел, и не о скрининге женщин, принадлежащих к группе высокого риска (например, вследствие семейной предрасположенности).

### **Контрольные вопросы и задачи к разделу 3.5**

1. Результаты рандомизированного двойного слепого испытания показали, что малые дозы пероральных кортикостероидов дают лучший результат в смягчении симптомов ревматоидного артри-



- та, чем нестероидные противовоспалительные средства. Этот результат может быть ложноположительным, так как:
- а) критерии включения больных были слишком мягкими;
  - б) при оценке воспалительной реакции не учитывалось число пораженных суставов;
  - в) разница в исходах может быть случайной;
  - г) размер выборки слишком мал.
2. В ряде контролируемых испытаний, изучавших связь между приемом эстрогенов и раком эндометрия, использовался односторонний статистический критерий. Основным соображением, на котором базировалось предпочтение перед двухсторонним критерием, было:
- а) распределение интенсивности экспозиции нормально в сравниваемых группах;
  - б) исходно были основания предполагать, что инцидент может повышаться при приеме эстрогенов;
  - в) существует прямая зависимость доза—эффект;
  - г) в основном в таких исследованиях представляет интерес ошибка первого типа.
3. В рандомизированном контролируемом испытании обнаружено различие действенности двух вмешательств со статистической значимостью  $P < 0,05$ . Предположим, что на самом деле вмешательства не различаются по действенности. Эта ситуация является примером:
- а) ошибки типа 1 ( $\alpha$ );
  - б) ошибки типа 2 ( $\beta$ );
  - в)  $1 - \alpha$ ;
  - г)  $1 - \beta$ .
4. В контролируемом рандомизированном испытании в группе вмешательства болевой синдром был устранен в течение суток у 8 из 10 пациентов, а в контрольной группе — у 4 из 11. Рассчитайте показатель ЧПЛП (число больных, подвергаемых лечению для получения одного полезного исхода).
5. В контролируемом испытании, когда сравниваемые группы создаются способом рандомизации, между ними возникают различия случайного происхождения. И следует:
- а) проверить, значимы ли они статистически;
  - б) при статистическом анализе различий в эффекте вмешательства выполнить коррекцию на величину этих различий;
  - в) игнорировать их как случайные;
  - г) увеличить численность выборки.



## Глава 4

# СПЕЦИАЛЬНЫЕ РАЗДЕЛЫ ЭПИДЕМИОЛОГИИ

Современная эпидемиология далеко шагнула за пределы материнской дисциплины — эпидемиологии инфекционных болезней. В настоящем разделе рассматриваются основные особенности важнейших областей применения эпидемиологических методов. Эти разделы эпидемиологии благодаря специфике изучаемого объекта (субстрата), особенностям применяемых вмешательств и влиянию предшествующего опыта существенно отличаются друг от друга и используют некоторые специальные методы дополнительно к методам общей эпидемиологии. Инфекционная эпидемиология является частью курса инфекционных болезней, но во многих российских вузах преподается как отдельный курс эпидемиологии, преимущественно инфекционной. Основы генетической эпидемиологии рассматриваются в общей генетике и в курсе наследственных болезней, эпидемиология профессиональных болезней — в курсах профессиональных болезней и гигиены труда, элементы эпидемиологии хронических болезней и социальной эпидемиологии — при изучении почти всех медицинских дисциплин, элементы *фармакоэпидемиологии* — в курсе клинической фармакологии (объектом исследований фармакоэпидемиологии является распространенность использования лекарств и связанных с этим эффектов).

### 4.1. ЭПИДЕМИОЛОГИЯ ИНФЕКЦИОННЫХ БОЛЕЗНЕЙ

*Инфекционная эпидемиология* — раздел эпидемиологии, выделенный по субстратному признаку. Инфекционная эпидемиология изучает распространение инфекционных болезней и определяется как раздел инфектологии (науки об инфекциях, включая изучение возбудителя



и клинической картины инфекционных болезней). Несмотря на исключительное значение, которое имела эпидемиология инфекционных болезней в истории медицины и эпидемиологии, в частности, ныне инфекционная эпидемиология представляет собой такой же раздел частной эпидемиологии, как эпидемиология рака или генетическая эпидемиология.

Эпидемиология инфекционных болезней сложнее, так как в эпидемическом процессе случаи заболевания не являются только исходами экспозиции, а сами становятся источником инфекции (экспозицией) для других людей. Сложнее и периодизация заболевания. При инфекционных болезнях говорят об *инкубационном периоде* (от заражения до появления симптомов) и о *латентном периоде*, за которым следует *заразный период*, т.е. период, в котором возможна передача инфекции (рис. 4.1). Применительно к инфекционным болезням действует зависимость вероятности заболевания от дозы инфекционного агента, лучше всего изученная применительно к действию лекарств. Чем выше доза, измеряемая числом бактериальных тел, тем больше вероятность заболевания. При любой дозе часть людей заболевает, часть — нет, одни болеют тяжело, другие — легко. В различных последствиях одинакового воздействия проявляются индивидуальные особенности резистентности людей (устойчивости к воздействию). Именно при инфекционных болезнях высокая резистентность отдельных людей (иммунитет) впервые привлекла внимание исследователей и стала основой великих открытий.

Эпидемиология инфекционных болезней продолжает сохранять ведущее положение в изучении заболеваемости человека в связи с сохраняющейся и постоянно изменяющейся угрозой новых инфекционных болезней. Более того, для многих болезней, еще в середине XX в.

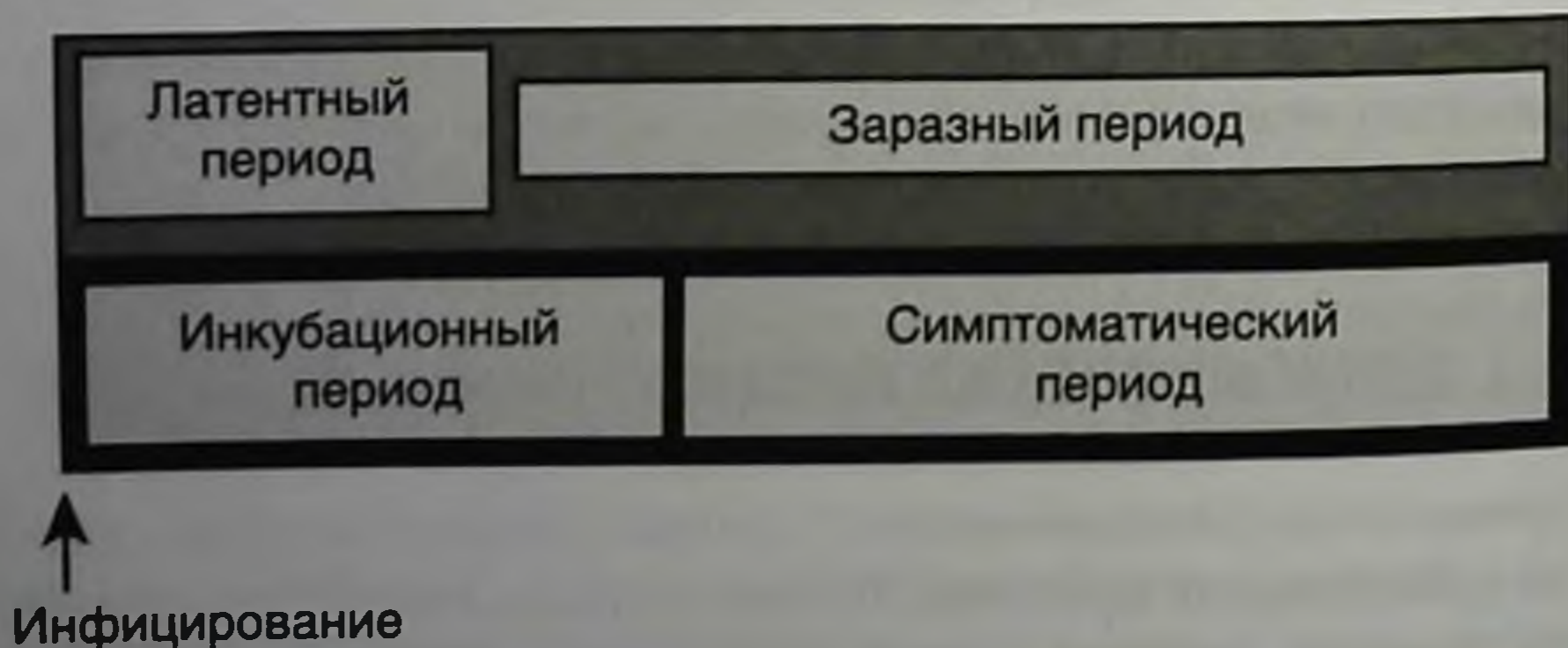


Рис. 4.1. Периодизация инфекционного заболевания



считавшихся неинфекционными, связь с инфекцией или доказана (язвенная болезнь, некоторые виды рака), или в пользу участия инфекции в этиологии существуют веские аргументы (атеросклероз, болезнь Крейтцфельда—Якоба).

Для инфекционных болезней характерно внезапное учащение заболеваний, становящихся массовыми, — *эпидемия*. Обычно число заболевших в начальной стадии эпидемии увеличивается экспоненциально в связи с тем, что каждый заболевший становится источником заражения новых случаев. Если эпидемия захватывает территории континентов, то говорят о *пандемии*. В случае «классических» инфекционных болезней иммунитет стоек. Поэтому переболевшие люди невосприимчивы к инфекции более или менее длительное время, и на годы число заболевающих в местности, где была эпидемия, остается незначительным. Поэтому нередко вместо термина «эпидемия» употребляют термин «*вспышка*». Чаще «вспышкой» обозначают связанные общим происхождением заболевания, не затрагивающие большей части популяции, например групповые заболевания дизентерией вследствие употребления в пищу загрязненных продуктов. Отдельные (*спорадические*) случаи инфекционной болезни не приводят к эпидемии по причине невосприимчивости популяции — имеет место популяционный или коллективный иммунитет (англ. *herd immunity*).

По ряду причин болезни, в том числе инфекционные, различаются в разных странах, регионах. В местностях, где в воде недостаточно йода, распространен зоб. Исключительно в Центральной Африке, где обитает муха цеце (род *Glossina*), передающая *Trypanosoma brucei*, люди страдают от сонной болезни (африканского *трипаносомоза*). На территориях России, где обитают иксодовые клещи, ими распространяется клещевой энцефалит. В свою очередь, ареал распространения насекомого-переносчика (*вектора*) зависит от температурных, влажностных, геохимических особенностей местности. Такие большие регионы с *эндемической* заболеваемостью называют глобальными *нозоареалами*. Особенности природных условий, в том числе наличие природных очагов инфекции и наличие специфических переносчиков, выражаются в концепции *природной очаговости болезней человека*. В противоположность эндемической выделяют *экзотическую* заболеваемость, т.е. заболеваемость инфекциями, не свойственными данной местности.

Инфекционные болезни классифицируют по различным признакам. По механизму сохранения инфекции в природе и механизму ее переда-



чи выделяют антропонозы и зоонозы. *Антропонозы* — болезни человека, передающиеся от человека и не опасные для животных. В случае отсутствия больных людей инфекция исчезает — так программа всемирной вакцинации против оспы привела к ее исчезновению<sup>1</sup>. *Зоонозы* — болезни человека и животных, передающиеся обычно от животных человеку. Передача от человека к человеку может быть ограничена, как, например, при кожной форме сибирской язвы и туляремии, или может быть частой (при легочной чуме). Зоонозы всегда угрожают людям, поскольку в естественных условиях в популяциях животных (в природных очагах) инфекция сохраняется. Она всегда может быть передана человеку и дать начало эпидемии у людей. Так, в 2019 г. новый коронавирус (SARS-CoV-2) проник в человеческую популяцию и вызвал первую пандемию 21-го века, нарушившую весь образ жизни современного человечества. Эпидемиология инфекционных болезней во многих отношениях уникальна, отлична от прочих разделов эпидемиологии, так как в основе ее лежат представления о передаче инфекционного агента и взаимной адаптации макро- и микроорганизмов.

Многие эпидемиологические концепции, лежащие в основе эпидемиологии инфекционных болезней, сформировались, когда представления об «инфекции» в современном понимании отсутствовали. Тогда возникновение вспышек и концентрация заболеваний в неблагоприятных условиях рассматривались в свете теории миазмы, т.е. вредоносного испарения (эманации). Эти представления санитарного периода эпидемиологии применялись успешно на практике в борьбе с массовыми заболеваниями путем очистки от нечистот, строительства канализации и водопровода. Ошибаясь в понимании происхождения болезней, деятели санитарного периода правильно указывали направление поиска причинного агента и способы его поиска.

Достижения Я. Генле (1840 г., он вслед за Дж. Фракасторо на более современном уровне сформулировал гипотезу о микроскопических возбудителях болезней), Джона Сноу (1854 г., объяснившего эпидемию холеры заражением питьевой воды), Луи Пастера (1865 г., показавшего,

---

<sup>1</sup> Вирус сохраняется в двух лабораториях — в России и в США, и еще в 2000 г. обсуждалась перспектива полного уничтожения этих «коллекционных» вирусов. После сентября 2001 г. опасения использования биологического оружия террористами заставили отказаться от уничтожения вируса. Возобновились вакцинации против оспы медицинских работников и работников спецслужб. Это яркий пример того, как человек подошел вплотную к существенному вмешательству в природу, но сильнее его оказались человеческие, политические проблемы.



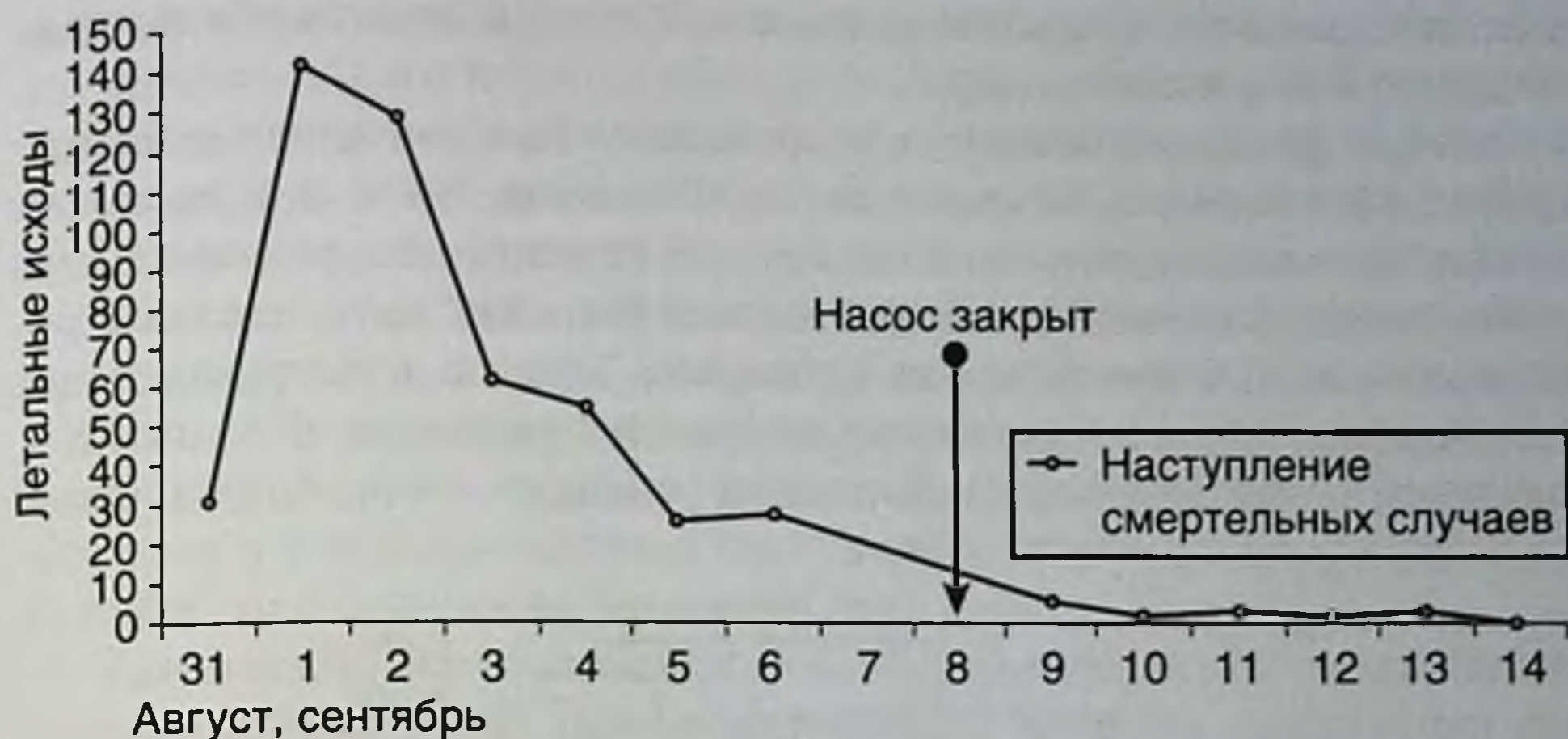


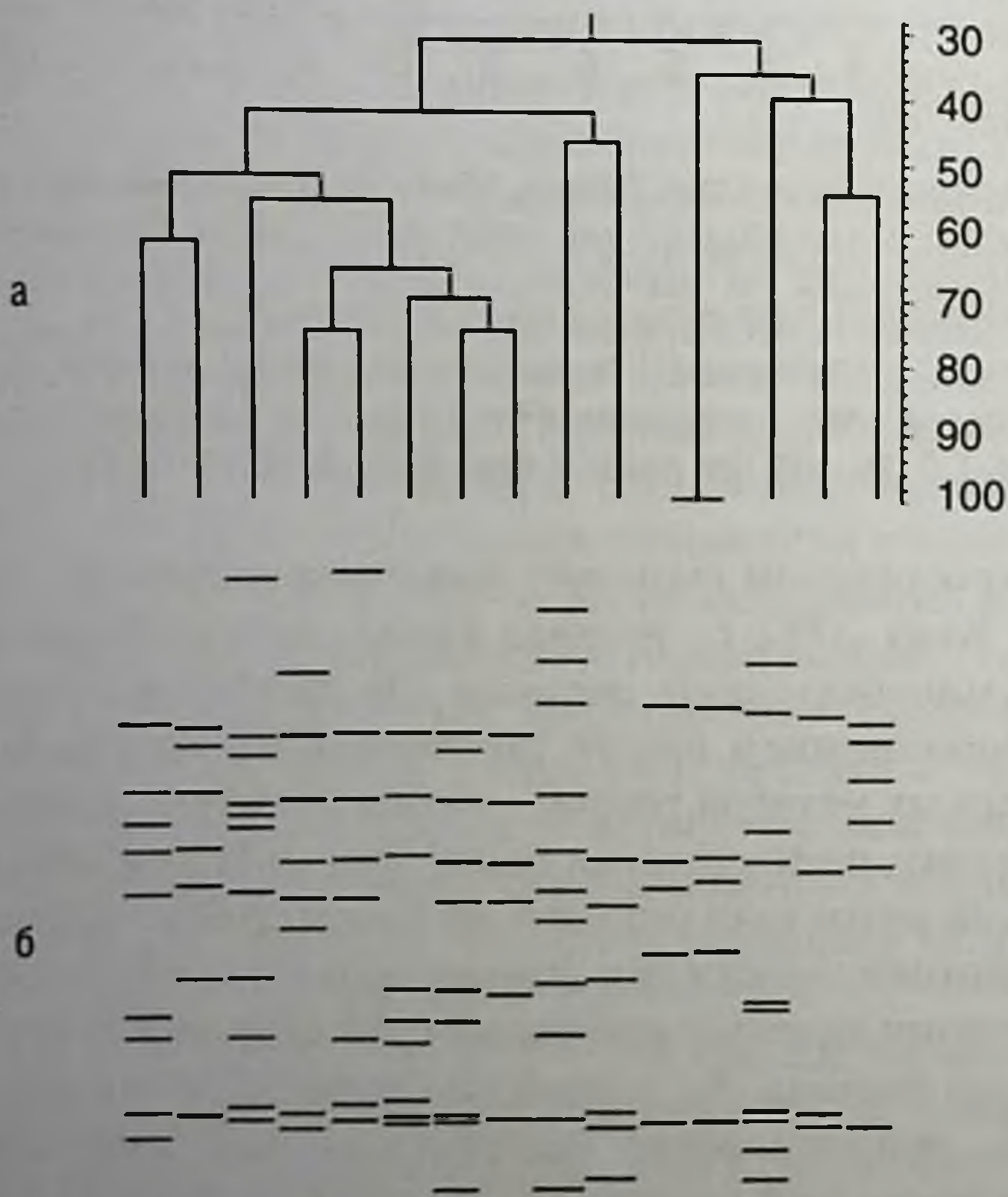
Рис. 4.2. Вспышка холеры на Брод-стрит, Лондон, 1854 г. Водоразборная колонка отключена (вертикальная стрелка), когда вспышка уже почти прекратилась. Возможно, этим была предотвращена новая вспышка, но главной заслугой Дж. Сноу является объяснение причины, ставшее достоянием науки. Проанализировано 668 смертей, для 127 дата начала заболевания неизвестна (по Whitehead H. Remarks on the outbreak of cholera in Broad Street, Golden Square, London, in 1854. Transactions of the Epidemiological Society of London. 1876. № 3. P. 99–104. Цит по Int. J. Epid. 2001. № 30. P. 1–11)

что живой микроорганизм вызывает эпидемию у тутового шелкопряда), и Роберта Коха (1882 г., ученика Генле, установившего причину туберкулеза — микобактерию), открыли эру инфекционной эпидемиологии. С возникновением нового понимания инфекционных болезней и лабораторных методов теория миазмы устарела подобно теории флогистона. Успехи инфектологии были столь значительны, что эпидемиологический метод стал отходить на второй план. Новые вакцины и новые антибиотики, успехи экономического и социального развития привели к быстрому падению инфекционной заболеваемости и смертности в развитых странах. Во второй половине XX в. на первое место стала выходить эпидемиология неинфекционных болезней. Вместе с тем эпидемиология инфекционных болезней постепенно трансформировалась. Для лучших ее представителей предстоящие угрозы новых инфекций не были тайной (Р. Дюбо, В. Беляков). Благодаря этому сформировались современные концепции инфекционной эпидемиологии как экологической дисциплины. Потребность в предотвращении инфекций, передаваемых с кровью, оказывает заметное влияние на технику хирургов и на методы ведения оперируемых больных (терпимость



к умеренной послеоперационной анемии, особые технологии заготовки крови и ее компонентов).

Успехи фундаментальных наук оказывают большое влияние на прогресс инфекционной эпидемиологии. В течение XX в. эпидемиологи довольствовались делением возбудителей болезни на серотипы. Число различных разновидностей возбудителя было так мало, что было невозможно проследить источник заражения. Удавалось, например, лишь подтвердить, что все выявленные больны туберкулезом. В конце XX в. появился метод полимеразной цепной реакции — тест, позволяющий



**Рис. 4.3.** Результаты генетического анализа микобактерий отдельных больных: а — древовидная диаграмма результатов анализа подобия культур микобактерий, выделенных у отдельных больных. Чем ближе генетическая структура сравниваемых микобактерий, тем ниже они соединены между собой (шкала подобия справа); б — «отпечатки пальцев», схематически изображенные результаты электрофореза фрагментов ДНК. (по Kulaga S. et al. Amer. J. Epidemiol. 2004. № 159. P. 507)



с высокой точностью выявить микроорганизм, вызывающий болезнь, по фрагментам его дезоксирибонуклеиновой кислоты (ДНК) в пробах тканей. Это привело к радикальному повышению точности диагноза при инфекционных заболеваниях. Наконец, для некоторых микроорганизмов изучение их генетики позволило выявлять большое число разновидностей. Стало возможным среди множества больных туберкулезом выявлять группы, имеющие одну разновидность возбудителя, и таким образом ретроспективно выявлять случаи реактивации старой инфекции и случаи свежего группового инфицирования. Стало возможным в отдельных случаях отслеживать цепочки заразившихся туберкулезом и другими инфекциями (рис. 4.3).

В некоторых аспектах развитие эпидемиологии и инфектологии сохраняет удивительную преемственность. С момента изобретения эффективной вакцины против оспы и введения обязательной вакцинации в Англии в 1840 г. вакцинация продолжает оставаться самым эффективным способом предотвращения многих инфекционных болезней. Всякий раз, когда по каким-то причинам процедуры вакцинации прерываются, обычно в надежде на «окончательную победу над инфекцией», инфекция напоминает о себе (см. рис. 1.9). Тем не менее до настоящего времени сохранение и развитие системы вакцинопрофилактики проходят в постоянной борьбе с группами граждан, противодействующими вакцинации. В середине XIX в. эта борьба с современной точки зрения проходила на рациональной основе сопротивления жесткому закону о вакцинации, не оставлявшему места для самоопределения граждан. Однако при ближайшем рассмотрении видно, что как 150 лет назад, так и сейчас в сопротивлении вакцинации было и остается много иррационального (рис. 4.4).

Несмотря на то что эпидемиология конца XX в. стала эпидемиологией всех болезней и даже преимущественно эпидемиологией хронических болезней, именно во второй половине XX в. были сформированы наиболее эффективные системы наблюдения за инфекционной заболеваемостью. Они позволили выявить СПИД и беспрецедентно быстро — в масштабах человеческой истории — реагировать на другие возникающие угрозы новых инфекций, изучать новые инфекции и новых возбудителей (прионы, микоплазмы). К исходу XX в. стало ясно, что представления о том, что инфекции находятся «под контролем», преувеличены. Они никогда не были под контролем в развивающихся странах и легко выходят из-под контроля в развитых странах, как показал опыт с туберкулезом, ящуром, сибирской язвой, COVID-19.





**Рис. 4.4.** Карикатура периода сопротивления вакцинации в Великобритании. Закон о вакцинации сделал в 1840 г. оспопрививание бесплатным для бедных и запретил вариоляцию. В 1853 г. вакцинация стала обязательной для младенцев «простых людей», и родители подвергались преследованию за непрививание новорожденных. В 1867 г. обязательное оспопрививание было распространено на лиц в возрасте до 14 лет. Эти законы были первыми, внедрившимися во имя общественного здоровья в область, ранее защищенную традиционными гражданскими свободами. Сопротивление вакцинации было значительным, с насильственными выступлениями во многих городах Англии и Ирландии. Законодательство было изменено в пользу смягчения наказаний только в 1898 г.

Благодаря наличию молекулярно-биологических методов в 2020 г. в течение нескольких недель была изучена молекулярная структура нового коронавируса, что сделало возможным почти немедленное изготовление диагностикумов и начало разработки вакцин. Важно, однако, что и после года развития эпидемии специфического противовирусного лекарства нет. Ее сдерживание основывается на применении эпидемиологических методов социального разобщения людей и выявления и изоляции больных и контактировавших с ними лиц. Будущие эпидемии неизбежно потребуют от человечества использования в первую очередь этих эпидемиологических средств, к чему общество готовится, в том числе путем использования новых информационных технологий.



## 4.2. ЭПИДЕМИОЛОГИЯ ХРОНИЧЕСКИХ БОЛЕЗНЕЙ

Современная эпидемиология, начиная с санитарного движения XIX в., всегда была озабочена проблемами общественного здоровья. Это направление мышления стало преобладающим во второй половине XX в., когда крупные достижения были связаны с исследованиями эпидемиологии сердечно-сосудистых болезней (Фремингемское исследование, В. Каннел) и рака (Р. Долл, А.Б. Хилл). Важнейшим теоретическим достижением эпидемиологии неинфекционных болезней было исследование взаимодействия факторов среды и индивидуальных особенностей людей (врожденных и приобретенных) в генезе заболевания. Впервые эта идея была сформулирована в начале века, но свое окончательное выражение получила лишь в 1950-е годы.

Развитие неинфекционной эпидемиологии проходило в условиях, когда среди инструментов медицины уже были контролируемый эксперимент с рандомизацией и быстро развивающийся статистический анализ, включая сложные вычислительные методы, применимые только на компьютерах. Эпидемиология обогатила свой инструментарий эффективными наблюдательными исследованиями — когортными и типа сравнения с контролем. Кибернетика сформулировала принцип анализа явлений, основанный на идее «черного ящика» — целого объекта, структура которого неизвестна, и изучению доступны только вход и выход. Этот принцип был успешно использован по отношению к неинфекционным болезням. В соответствии с этим принципом необязательно изобретать объяснительную патогенетическую гипотезу. Можно исследовать связь явлений — исходов с факторами риска (см. разд. 1.1.4).

Самым ярким примером успеха нового эпидемиологического принципа стало выявление связи рака легких с курением. Тесная статистическая связь, выявленная в наблюдательных исследованиях, ни изначально, ни теперь не основывается на стройной биологической теории. Число объясняющих гипотез огромно, но они вторичны по отношению к выявленной эпидемиологически связи. Вмешательства в интересах общественного здоровья на основе эпидемиологических данных резко изменили отношение к курению в развитых странах. Катастрофический рост числа курящих, характерный для середины XX в., там был переломлен. Вместе с эпидемиологическими данными о факторах риска сердечно-сосудистых заболеваний сведения о курении изменили образ жизни населения развитых стран, структуру расходов граждан и государств.



Примером относительно быстрого успеха применения принципа «черного ящика» вплоть до выявления причины болезни и ее эффективного предотвращения является исследование дефектов развития нервной трубки у плода. Старые сведения показывали, что эти дефекты встречаются с разной частотой в зависимости от социального класса, этнической принадлежности, экономического состояния общества. Эти дефекты чаще отмечаются у бедных, у представителей национальных меньшинств, в периоды экономического спада. Более детальные исследования врожденных дефектов показали, что их возникновение связано с голоданием в ранние сроки беременности, и что хорошее питание и витаминизация позволяют сократить риск. Наконец, эксперименты на животных показали, а последующие испытания на людях подтвердили, что прием фолиевой кислоты в период, предшествующий зачатию, и в ближайшее время после зачатия позволяет предотвратить значительную часть дефектов. В короткий исторический период — менее чем за 20 лет — был пройден весь цикл изучения патологии и эффективного решения проблемы, которая существовала всегда в человеческой истории. Однако во многих странах, в том числе и в России, обогащение пищи фолиевой кислотой не проводится и через 40 лет после получения доказательств эффективности профилактики.

Новые технологии изучения генетического аппарата человека позволили надежно установить многие причинные связи, о которых ранее можно было только подозревать. Саркома Капоши теперь имеет ясное происхождение от инфекции вирусом герпеса. Установлено, что некоторая часть случаев рака молочной железы имеют наследственное происхождение — ген предрасположенности передается в семье. Новые методы получения изображений внутренних структур человеческого тела дали толчок к установившемуся пониманию шизофрении не как функциональной болезни, а как болезни, имеющей материальный субстрат, и впереди большие открытия в ее изучении. Технологии, делающие видимыми функционирование человеческих тканей, далеко ушли от визуализации начала XX в., когда У. Кеннон делал движения желудка видимыми под рентгеном, прикрепляя к желудку собаки ряды кнопок. С другой стороны, картирование человеческого генома обещает в ближайшем будущем резкое изменение структуры задач, решаемых эпидемиологией. Огромный список систематизированных генетических признаков ждет своего сопоставления с заболеваниями.



### 4.3. ЭКОЛОГИЧЕСКАЯ ЭПИДЕМИОЛОГИЯ<sup>1</sup>

Экологическая эпидемиология (*environmental epidemiology*) изучает последствия воздействия на человека неблагоприятных факторов окружающей среды. При изучении воздействия неблагоприятных факторов окружающей среды на состояние здоровья выявляют не только явно выраженные эффекты и заболевания, но и первичные функциональные изменения в организме. Соотношение между клиническими, субклиническими и бессимптомными стадиями заболеваний колеблется в широких пределах в зависимости от специфических особенностей факторов окружающей среды, интенсивности, длительности и механизмов их воздействия. Субклинические сдвиги и физиологические изменения неясного значения определяются только в специальных исследованиях. В последнее время для этой цели широкое распространение получили биологические маркеры, свидетельствующие о наличии предвестников заболеваний.

По характеру своего действия на организм человека неблагоприятные факторы окружающей среды группируют по основным вредным последствиям: вызывающие злокачественные новообразования (канцерогены); вызывающие нарушения внутриутробного развития (тератогены), имеющие фетотоксическое, мутагенное и иное действие. Эпидемиологическими методами изучается воздействие на здоровье человека различных неблагоприятных факторов, присутствующих в разных средах: загрязнения атмосферного воздуха, питьевой воды, почвы, пищевых продуктов, шума, естественной радиации, электромагнитных полей, внутреннего воздуха помещений, а также более специфических агентов, например излучения персональных электронных устройств; электромагнитных полей линий электропередач.

Эпидемиологические исследования позволяют выявить последствия загрязнения (изменения) окружающей среды на состояние здоровья населения, определить количественную величину изучаемых воздействий, установить причинно-следственные отношения между неблагоприятными факторами среды обитания человека и показателями здоровья. В связи со сложной, многофакторной природой хронических неинфекционных заболеваний доказать причинную связь между развившимся у человека заболеванием и предшествующей экспозицией трудно. Это возможно в правильно спланированных эпидемиологических исследо-

<sup>1</sup> Автор Б.А. Ревич.



ваниях. Фактор окружающей среды рассматривается как фактор риска, т.е. такой компонент этиологии, который хотя и важен для развития и прогрессирования заболевания, но сам по себе при отсутствии других условий (например, генетической предрасположенности, измененного иммунного или эндокринного статуса организма) обычно не способен вызвать заболевание у конкретного человека. Другими словами, факторами риска называются внешние воздействия или особенности организма, приводящие к увеличению вероятности возникновения неблагоприятных эффектов.

Методы экологической эпидемиологии разработаны с учетом того, что все внешние факторы действуют при участии других — вмешивающихся (*confounding*) факторов, которые сами оказывают влияние на здоровье. В первую очередь это факторы образа жизни человека: курение, употребление алкоголя, наркотиков, малая физическая активность, особенности питания, контакт с вредными производственными факторами, жилищные условия. Действие всех этих факторов реализуется на фоне значительного генетического разнообразия популяции и в условиях различной доступности медицинской помощи. Доля влияния неблагоприятных факторов окружающей среды в заболеваемости и смертности населения в целом по стране не превышает нескольких процентов, но на наиболее экологически неблагополучных территориях может быть значительно выше.

В качестве примера проекта с использованием метода исследования сравнения с контролем (ИСК) приведем результаты работы в г. Чапаевске Самарской области, где на протяжении десятилетий среда загрязнялась выбросами химического завода, содержащими хлорорганические вещества, в том числе и высокотоксичные диоксины. Эти вещества обладают канцерогенными свойствами и являются псевдогормонами. Была поставлена задача оценить воздействие загрязнений как фактора риска развития рака молочной железы (РМЖ). При этом учитывались и вмешивающиеся факторы, которые также влияют на развитие этого заболевания<sup>1</sup>. В исследовании оценивали такие факторы риска, как работа на химическом производстве, использование местных продуктов питания, в том числе рыбы из загрязненных водоемов, использование контрацептивов, кормление грудью и другие факторы, которые могли исказить показатель силы связи между воздействием и эффектом. Осо-

<sup>1</sup> Ревич Б.А., Ушакова Т.И., Сергеев О.В., Зейлерт В.Ю. Рак молочной железы в Чапаевске. Гигиена и санитария. 2005. № 1. С. 18–21.



бое внимание уделялось особенностям питания. Это связано с тем, что основная доля диоксинов поступает в организм человека с продуктами питания, содержащими жир, такими как мясо, яйца, рыба. Чем больше жира в том или ином продукте, тем больше вероятность накопления в нем диоксинов. Обнаружено, что больные РМЖ женщины чаще используют мясо и жир свиней, выращенных в Чапаевске, чем женщины контрольной группы. Для статистической оценки факторов риска РМЖ использовали отношение шансов (*odds ratio* — OR), указывающее на большую частоту наличия фактора в основной группе по отношению к группе контроля. Значение отношения шансов более 1 указывает на повышенный риск (табл. 4.1). На статистическую значимость обнаруженного повышения риска указывает нижняя граница 95% ДИ выше 1,0.

Таблица 4.1. Основные факторы риска рака молочной железы в г. Чапаевске

Фактор риска	Отношение шансов (OR)	95% доверительный интервал
РМЖ у родственников	9,0	1,85–43,6
Использование местной свинины в рационе питания чаще чем в 50%	5,7	1,3–25,5
Использование оральных контрацептивов	2,4	1,03–5,5
Использование местной рыбы в рационе питания чаще чем в 50%	2,3	1,13–4,80
Работа на химическом заводе	2,1	0,95–4,68
Работа родителей на химическом заводе	1,8	0,80–4,00

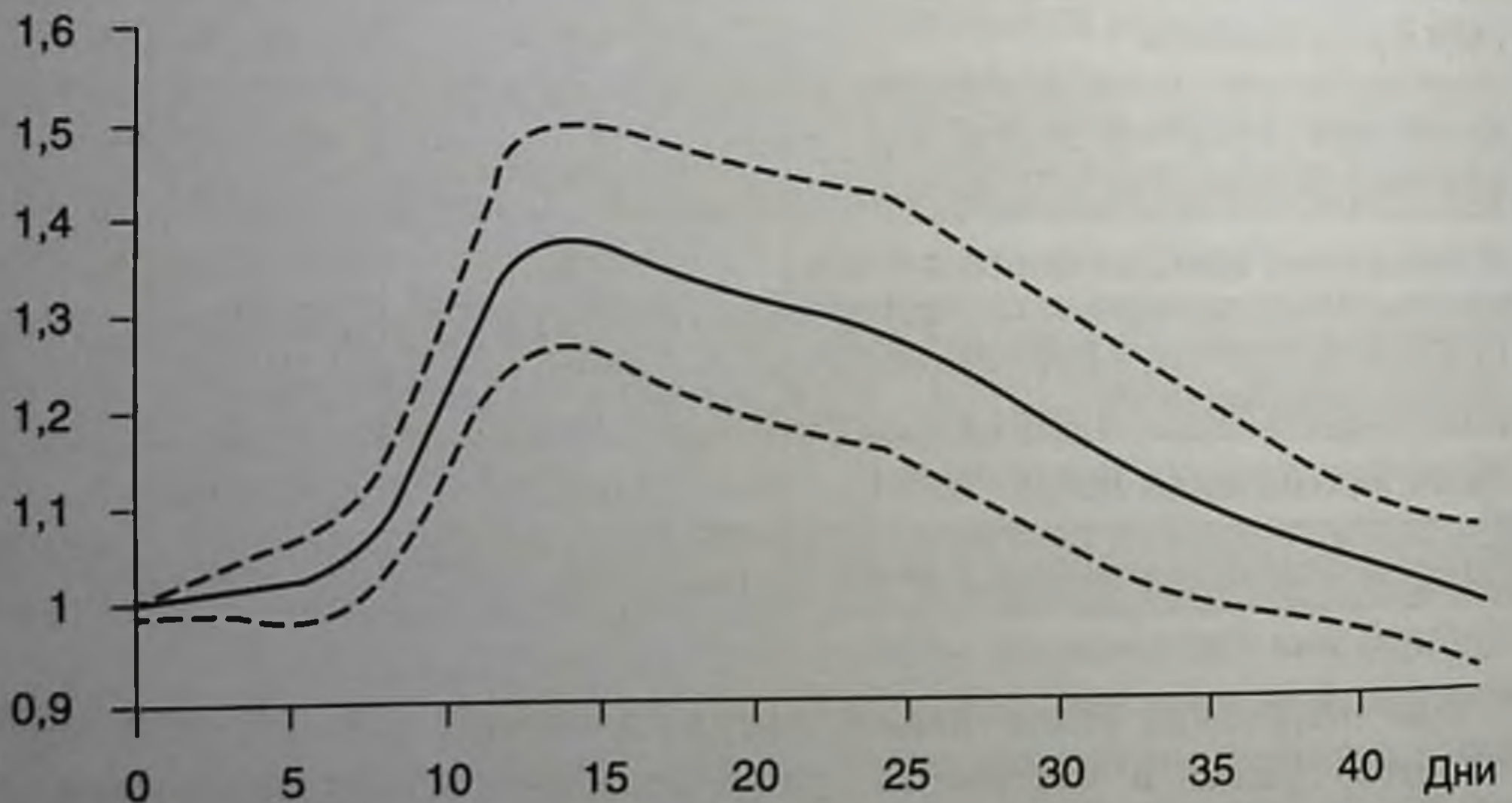
Для получения более точной оценки влияния загрязненной окружающей среды, в частности использования местных загрязненных продуктов питания, на формирование данной патологии, необходимо устранить влияние других факторов риска. В этом случае они рассматриваются как вмешивающиеся, и для этого требуется более сложный статистический анализ полученных данных.

Второй пример показывает воздействие волн жары на смертность населения. Смертность увеличивается при низких и при высоких температурах (см. рис. 1.45). При высокой температуре воздуха в городе



увеличивается загрязнение воздуха микрочастицами и газами, среди которых лучше изучен озон (см. рис. 2.31). Такая продолжительная волна жары (44 дня) имела место в Москве летом 2010 г.<sup>1</sup>. На рис. 4.5 видно, как быстро увеличилась общая смертность и сохранялась повышенной в течение месяца.

На основании данных о ежедневной смертности от основных причин, о температуре и загрязнении воздуха была разработана математическая модель. Она показала, что основной вредный фактор — высокая температура (суммарная дополнительная смертность 4300 случаев, 40% общей дополнительной смертности в 11 000 случаев). Загрязнения воздуха, сопутствующие жаре, добавляют 9% (1000 случаев, в том числе влияние взвешенных частиц 5% и озона 4%). Совместное действие высокой температуры и загрязнения воздуха дает 20% (2200 случаев смерти). Это означает, что если бы население в этот период находилось только под воздействием эквивалентных концентраций смога при температуре ниже 18°C, то дополнительная смертность снизилась бы на 91%.



**Рис. 4.5.** Повышение смертности (кроме насильственных смертей) в Москве в период 6 июля — 18 августа 2010 г. Линия — относительный риск в сравнении с многолетней средней, пунктир — 95% доверительный интервал. Абсцисса: дни от начала волны жары (по Shaposhnikov D. et al. *Epidemiology*. 2014. № 25 (3). P. 359–64)

<sup>1</sup> Shaposhnikov D., Revich B., Bellander T. et al. Mortality related to air pollution with the moscow heat wave and wildfire of 2010. *Epidemiology*. 2014. № 25 (3). P. 359–64.



Если бы загрязнение атмосферного воздуха оставалось на обычном фоновом уровне, то смертность в жару снизилась бы на 29%.

На основании результатов этой работы были разработаны предложения по количественным значениям температуры и загрязнения атмосферного воздуха, при которых необходимы профилактические действия со стороны местных властей, метеослужбы, самих жителей города. Выполнение подобных исследований приводит в самых важных случаях к уточнению международного регулирования технической политики. Так, по результатам расследования отравлений метилртутью (см. с. 228) была заключена Минаматская конвенция, ограничивающая производство и использование ртути. Стокгольмская конвенция направлена на ограничение производства и применения хлорорганических веществ, в том числе таких известных, как дихлордифенилтрихлорметилметан и диоксины. Основную массу заболеваний, связанных с неблагоприятным действием факторов окружающей среды, вызывают не высокие концентрации, приводящие к катастрофическим последствиям, а более низкие уровни загрязнения, характерные для производственных помещений, атмосферы городов. Они приводят преимущественно к неспецифическим последствиям, увеличивая заболеваемость и смертность людей.

## 4.4. ГЕНЕТИЧЕСКАЯ ЭПИДЕМИОЛОГИЯ

Молекулярная эпидемиология возникла на основе успехов в изучении молекулярных основ патологии. Она вытекает из клинической традиции и является торжеством способа рассуждения «от частного к общему». Выявление молекулярных механизмов, ответственных за патогенез или за повышение индивидуального риска, занимает ключевое положение в понимании развития болезней. На этой основе возможна разработка наиболее эффективного лечения и профилактики. В отдельных случаях обосновано выявление уязвимых субъектов и их защита от вредного для них воздействия. Наиболее известный пример — фенилкетонурия, состояние редкое (1:10 000 родов), ведущее к тяжелому дефекту развития мозга в раннем возрасте. Своевременное выявление состояния у новорожденных и специальное питание (ограничение фенилаланина и добавление тирозина) позволяют людям с таким дефектом метаболизма жить долго и почти без интеллектуального дефекта.



В то время как Дж. Сноу и его коллеги делали самые известные шаги в истории эпидемиологии (1854), монах Грегор Мендель работал над статистическими законами расщепления свойств в последующих поколениях организмов. Результаты исследований были опубликованы им в незамеченной долгое время статье 1865 г. Поскольку наследственные признаки мало доступны исследованию, эпидемиология долгое время занималась почти исключительно изучением внешних признаков и факторов среды, влияющих на развитие заболеваний. Даже после открытия носителя наследственности — ДНК, ее исследования долгое время были недоступными, а опыты по скрещиванию на человека остаются неприемлемыми. Соответственно при изучении генетики человека доступны только неэкспериментальные исследования, без возможности узнать, каков механизм связи признака и исхода, т.е. эпидемиологические методы. Результаты эпидемиологических исследований дополняются исследованиями на модельных животных, на которых возможно экспериментирование. *Генетической эпидемиологией* называют науку, изучающую роль генетических факторов в возникновении болезней в группах родственников, в человеческих популяциях, а также взаимодействие генетических факторов с факторами окружающей среды.

Среди методов выявления наследственного характера болезни на первом месте стоят два относительно простых: выявление связи болезни с другой болезнью, наследственная природа которой хорошо известна, и выявление связи с другим признаком (*генетическим маркером*), который передается по наследству. Например, у детей с синдромом Дауна повышена вероятность развития лейкоза. Поскольку природа синдрома Дауна хорошо известна, эта связь стала одним из доказательств наследственного компонента в развитии лейкозов.

Г. Мендель изучал признаки, которые было легко определять — цвет и форму семян и цветов гороха. В генетической эпидемиологии человека большая часть признаков описывается нечетко и большая часть болезней зависит более чем от одного гена (полигенное наследование). Развитие химии и генетики позволило вместо формы ушных раковин, отпечатков пальцев и цвета глаз изучать многие тысячи признаков — белки тканей. Все эти признаки являются генетическими маркерами, т.е. позволяют изучать по фенотипическим признакам соотношение между генами и болезнями (состояниями) у человека. Классическим примером *молекулярной генетической эпидемиологии* являются иссле-



дования связи групп крови АВ0 с болезнями. Установлено, что холеру хуже переносят лица с группой крови 0, а от хронических болезней выше смертность у лиц с группой АВ. Ограниченный круг таких маркеров накладывал ограничения и на перспективы исследований. Ранее для выявления наследственного характера болезней использовали доступные изучению антигены, прежде всего антигены главного комплекса гистосовместимости. Было показано, что сахарный диабет 1-го и 2-го типа — разные болезни, а не варианты одной болезни.

Теперь появляется возможность исследовать гены людей, включенных в эпидемиологическое исследование. Можно сопоставлять исходы и экспозицию не только с маркерами, количество которых ограничено, но далее уточнять связь с генами, и не только применительно к моногенным болезням, но и к полигенным. Для этого ДНК может быть получена из клеток крови в ходе клинического обследования или из эпителиальных клеток, например слизистой оболочки щеки. Образцы могут храниться неограниченно долго и обеспечивать большое количество анализов. Уже создаются большие коллекции ДНК длительного хранения, которые могут быть использованы для проверки гипотез исследователей относительно связи болезней с генами.

Структура исследования для выявления связи болезни с генетическим маркером или другой болезнью — ИСК: больных изучаемой болезнью и контрольных людей исследуют на предмет наличия у них исследуемого признака. Если у больных признак выявляется чаще, чем ожидается, то это является доказательством связи. Так же как и в ИСК, существует проблема выбора контрольной группы: исследователь обычно не может получить репрезентативную выборку из популяции. Лица, которые одновременно с изучаемыми больными находятся на стационарном лечении, очевидно, отличаются от популяции. Точно так же сильно отличаются от популяции призывники и доноры крови. Поэтому наличие связи болезни и маркера, как это обычно для ИСК, считается доказанным, только если выявляется сильная связь.

Вторым традиционным методом изучения наследственного компонента в этиологии болезней является изучение частоты болезни у близких родственников больного. Обычно речь идет о братьях, сестрах, матери, отце и детях, поскольку уже у двоюродных братьев (сестер) различия наследственности столь велики, что не отличают их от других членов популяции. Метод анализа родословных впервые был применен



к родословным королевских семей, где передача наследственных болезней была хорошо документирована в хрониках и потребовала лишь правильной интерпретации в XX в.

Для того чтобы установить долю наследственности в возникновении болезней, сопоставляют заболевания у однояйцовых и двуяйцовых близнецов, так как по определению у первых наследственные данные абсолютно идентичны. Соответственно болезни, абсолютно передаваемые по наследству, должны возникать у обоих близнецов или не возникать у обоих (конкордантное проявление). Относительная частота расхождений в диагнозе изучаемой болезни у однояйцовых близнецов (частота дискордантности) отражает долю ненаследственных (средовых) факторов в возникновении болезни. Если сравнить конкордантность у двуяйцовых близнецов с конкордантностью у однояйцовых, то большая конкордантность у однояйцовых свидетельствует о значимом наследственном компоненте в этиологии болезни.

У близнецового метода есть важные ограничения. Например, сам факт «однойяйцовости» близнецов (референтный признак) традиционно устанавливался путем изучения фенотипических признаков. Совпадающие по ряду традиционно изучаемых признаков близнецы классифицировались как однояйцовые. Этот метод дает небольшое число ошибок, как в виде признания однояйцовых близнецов разнояйцовыми, так и наоборот. Эти ошибки классификации, в свою очередь, влияют на точность оценки доли наследственной предрасположенности в развитии заболевания. Лишь в конце XX в. зиготность близнецов в исследованиях стали проверять, сопоставляя их ДНК.

Оценка доли наследственности в генезе болезни обычно не является самоцелью. Главная цель генетических исследований — выяснение наследственного механизма, знание которого могло бы предотвращать, лечить болезнь или рано ее выявлять. Ограничения оценки доли наследственного механизма вытекают из общего ограничения для эпидемиологических исследований: изучая события в популяции, мы получаем знания, соответствующие ей и не обязательно справедливые для всех возможных сочетаний условий. Поясним это примером. Порода кур КК страдает от наследственной непереносимости красной кукурузы. Если у фермера А есть куры разных пород, то, начав кормить их завезенной красной кукурузой, он обнаруживает заболевание у кур породы КК и делает вывод о наследственной причине болезни. У фермера Б есть только куры породы КК. Он начинает их кормить красной кукурузой



зой и, обнаружив возникшую у всех болезнь, обоснованно делает вывод о ее внешнем происхождении, об особой пище как причине. В реальных условиях соотношения между преваленсом внешнего фактора риска и преваленсом чувствительных субъектов могут быть не столь контрастными, и это еще более затрудняет выяснение причинно-следственных отношений.

Другими классическими методами генетической эпидемиологии являются *семейные исследования* и *исследования приемных детей*. В семейном исследовании изучают в основном болезни с высокой *пенетрантностью*. Так называют болезни, которые при наличии генетической предрасположенности проявляются с высокой частотой. Например, болезнь Хантингтона — один из вариантов быстро прогрессирующей деменции, возникающей у мужчин среднего возраста, — передается по мужской линии и проявляется почти в 100% случаев. Тем не менее и при таких состояниях, если человек умирает раньше от других причин, то болезнь не проявится. Если пенетрантность низка (не у всех носителей гена развивается болезнь), то предрасположенность к болезни с трудом может быть изучена по семейным данным.

Вариантом семейных исследований является изучение супружеских пар. Обычно супруги не имеют общей наследственности, точнее, они обычно различаются в генетическом отношении так же, как случайно выбранные члены популяции. С другой стороны, в течение долгих лет супруги подвергаются многим общим экспозициям: сходно питаются, живут в одном доме, придерживаются одной религии и т.д. В результате при изучении супружеских пар возникает возможность оценить, как часто по сравнению со случайным имеет место совпадение болезней. Далее можно выделить супружеские пары, разделяющие некоторую экспозицию. Например, можно оценить частоту болезней легких у жен курящих мужчин и сравнить ее с частотой этих болезней у женщин, живущих с некурящими мужчинами. На исследованиях такого типа в значительной степени базируются представления о вреде пассивного курения.

Исследования приемных детей также позволяют разделить наследственную предрасположенность и влияние среды. Приемные дети не разделяют генетической предрасположенности родителей, но подвергаются тем же экспозициям, что и приемные родители, братья и сестры. Например, показано, что у больных шизофренией, воспитанных в приемных семьях, приемные родители крайне редко страдают шизоф-



ренной. Наоборот, у больных шизофренией, воспитанных биологическими родителями, шизофрения родителей встречается в несколько раз чаще, чем у родителей здоровых детей. Значит, шизофрения не связана с условиями воспитания, взросления в семье, а имеет наследственное происхождение. Конечно, такого рода исследования доказательны, только если исследуются приемные дети, отлученные от биологических родителей рано, при рождении, поскольку влияние среды может реализовываться очень рано в индивидуальном развитии. Среда оказывает общее влияние на близнецов и в пренатальном периоде, что накладывает естественное ограничение на возможности разделения биологических и средовых факторов.

Другим вариантом таких исследований генетической предрасположенности и влияний среды является изучение близнецов, выросших в разлуке. По сравнению с близнецами, выросшими вместе, такие разлученные близнецы имеют меньше конкордантность по связанным с внешней средой болезням, но такую же — по болезням, преимущественно наследственным.

Классические генетические эпидемиологические исследования для повышения их эффективности используют методы, сходные с методами клинической эпидемиологии и экспериментальной медицины. Например, для выявления редких вариантов генов (аллелей), предрасполагающих к болезни, исследуют изолированные популяции островных жителей. В таких популяциях генетическое разнообразие людей ниже, а некоторые изучаемые аллели встречаются чаще, чем в гетерогенных континентальных популяциях. Аналогом исследования однородных популяций является использование линейных животных в экспериментах и клинически однородных групп больных в клинических испытаниях. В результате использования такого приема уменьшается общая вариабельность, не зависящая от изучаемого признака, и относительно этой общей вариабельности увеличивается вариабельность, связанная с изучаемым фактором (геном, вмешательством). Аналогом проведения генетических исследований в популяциях с высокой частотой изучаемого признака (гена, болезни) является проведение клинических испытаний в группах больных с достаточно высокой тяжестью болезни и достаточно высокой частотой изучаемого исхода (обострения). В таких популяциях и в таких группах больных увеличивается возможность выявления статистически значимого эффекта.



## 4.5. СОЦИАЛЬНАЯ ЭПИДЕМИОЛОГИЯ

В то время как эпидемиология зарождалась в виде санитарии в середине XIX в., ее ведущие деятели, такие как Э. Чадвик, обосновывая необходимость благоустройства городских трущоб, предполагали, что бедность будет побеждена так же, как побеждаются болезни и высокая смертность, так как она сама есть проявление болезней<sup>1</sup>. Санитарная статистика того времени начиналась как исследование общественной жизни и сбор данных о ней. Санитарное движение дало мощный толчок социалистическому движению во всем мире и питало народническое движение в России. Однако уже в санитарной эпидемиологии очевидные социальные факторы не признавались за болезнетворные, а постулаты Генле—Коха почти исключили их из рассмотрения медицинской наукой (социальные условия трактовались как способствующие действию материальных причин). В XX в. эпидемиология неинфекционных болезней обратилась к популяции и междисциплинарному анализу, но целью ее стали индивидуумы с их особенностями, индивидуальные риски в поведении человека и окружающей его среде.

Между тем эпидемиологические исследования показали, что многое в распространении инфекций зависит от социальных факторов, от поведения людей. Ярким примером является история заболеваемости гонореей в Европе в XX в. (рис. 4.6). Пик после Второй мировой войны связан с раскрепощением освободившихся от нацистского гнета людей. Инцидент быстро снижается благодаря появлению эффективных средств лечения и профилактики. Второй пик связан с «сексуальной революцией», и его преодоление преимущественно обеспечено эффективными социальными вмешательствами.

Известно, что влияние социального положения на заболеваемость и смертность очень велико. В большинстве случаев влияние социального положения на смертность сильнее, чем влияние «основных» факторов риска (рис. 4.7). Тем не менее социальное положение обычно рассматривается лишь как вмешивающийся фактор, т.е. фактор, затрудняющий исследование «основных» (биологических, физических) факторов. Социальная эпидемиология определяется как исследование

---

<sup>1</sup> Исторический анализ здоровья населения, в том числе в свете социального развития см. Johan P. Mackenbach. A History of Population Health: Rise and Fall of Disease in Europe. Leiden – Boston: Brill-Rodopi, 2020. 430 pp. ISBN 978-90-04-42582-8 (<https://brill.com/view/title/57111>).



влияния социальных факторов на заболеваемость и смертность. Исследования тесно связаны с социологией, экономикой, антропологией. Возникновение социальной эпидемиологии как самостоятельного научного направления относится к последней четверти XX в.

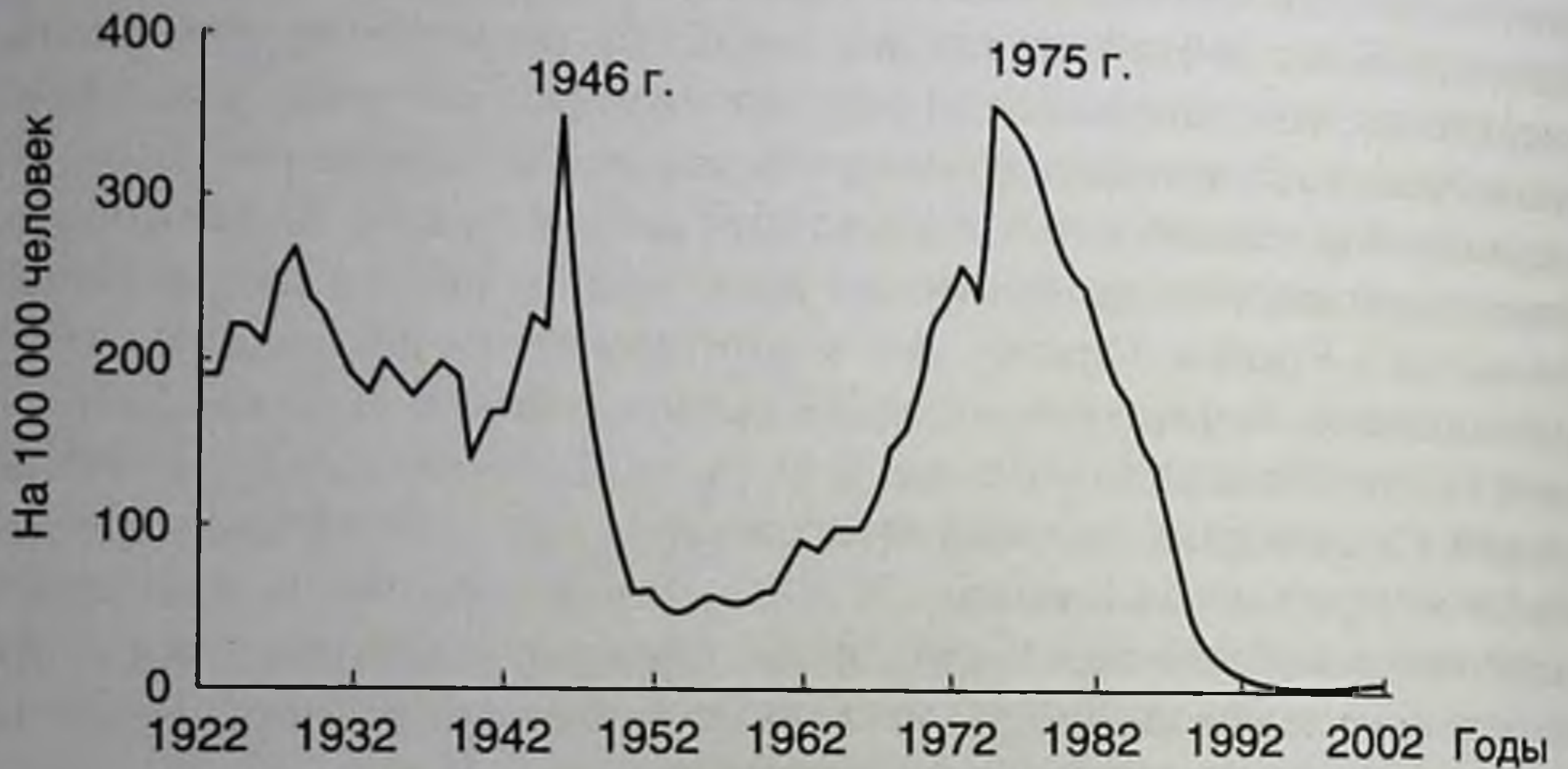


Рис. 4.6. Заболеваемость гонореей в Норвегии

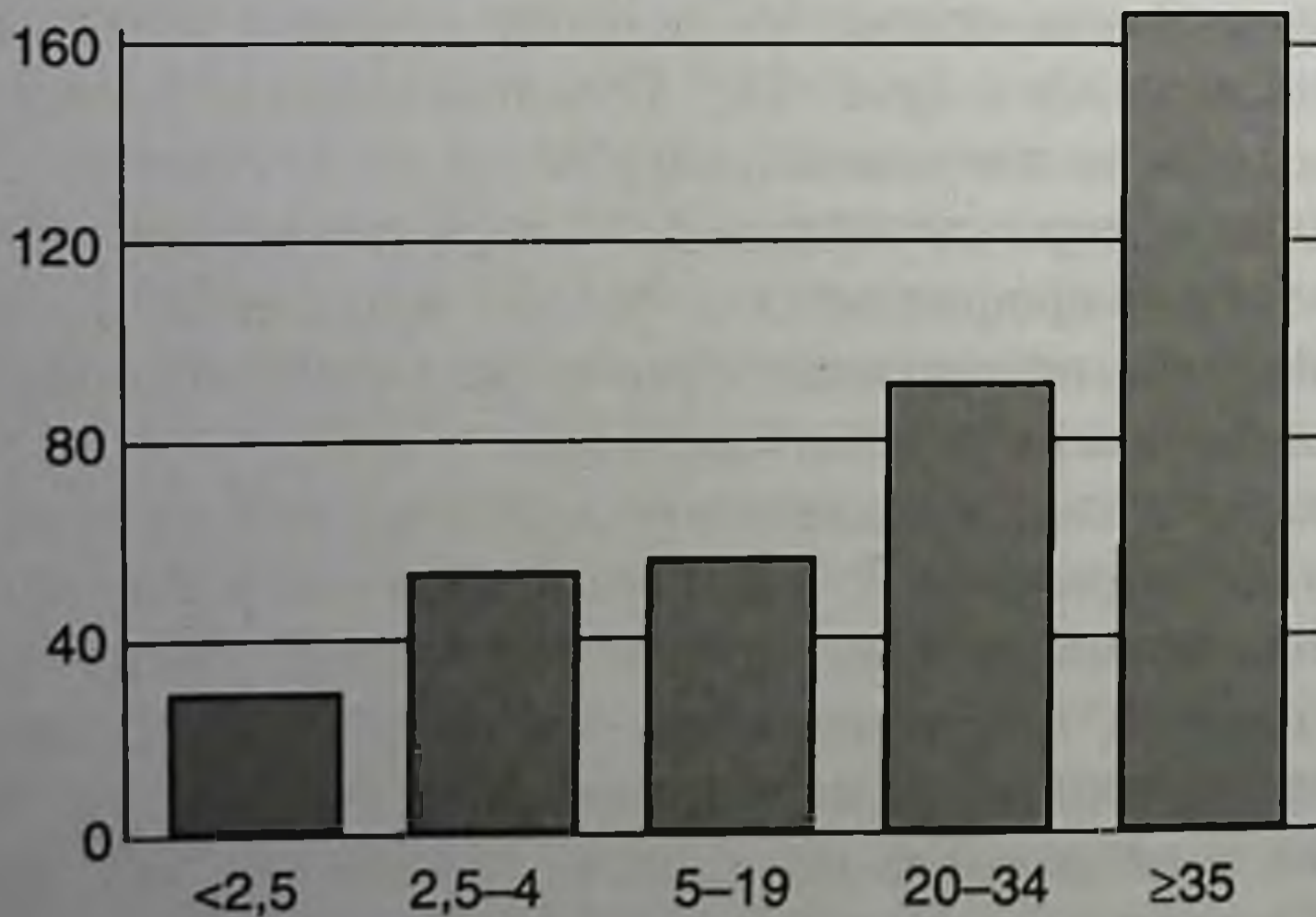


Рис. 4.7. Зависимость детской смертности (в возрасте до 5 лет, на 1000 родившихся живыми) от пропорции популяции, страдающей от недостатка питания (абсцисса). По отчету Международной продовольственной организации ООН (UN FAO annual report «The State of Food Insecurity in the World 2002»; <http://www.fao.org/DOCREP/005/Y7352e/Y7352e00.HTM>)



Некоторые социальные факторы, влияющие на заболеваемость, ясны — это бедность и неравенство. Трудность интеграции исследования социальных факторов в эпидемиологию в том, что способы повлиять на них малопонятны, и научная традиция восстает против включения «политики» в контекст исследования. Опыт эпидемиологии XX в. показал, что влияние на индивидуальные факторы риска у членов общества имеет ограниченную перспективу. Социальная динамика, поведение и факторы риска, видимо, более перспективны, но понятны значительно хуже, чем индивидуальные факторы риска.

Социальная эпидемиология расширяет рассмотрение медицинских проблем до их популяционного, общественного уровня. Курение, таким образом, предстает не только проблемой канцерогенеза и благоприобретенной зависимости, но также и проблемой экономики, последствий запретительных мер в развитых странах и перемещения потребления табака в развивающиеся страны. Факторы риска, присущие популяциям, не обязательно действуют внутри популяции. Индивидуальные факторы риска могут быть приложимы к популяциям, но популяционные — не обязательно приложимы к отдельным субъектам. В этом смысле экологический анализ неложный, но неверным является экстраполяция его результатов на индивидуальный уровень (см. разд. 2.5, 4.3). Анализ на многих уровнях подчеркивает и необходимость учета временных изменений. Не только инфекции приходят и уходят, точно так же возрастает и снижается заболеваемость язвенной болезнью, ишемическая болезнь сердца, сначала бывшая болезнью обеспеченных людей, в конце XX в. стала болезнью, более свойственной бедным. В многомерном анализе вторичный фактор может стать в некотором сочетании условий самым важным и определить направление действий.

Социальные науки в XX в. прошли путь, сходный с развитием эпидемиологии. От «опросов» и наблюдения они пришли к эффективным наблюдательным исследованиям и к модели экспериментирования, схожей с экспериментированием в медицине, — с планированием эксперимента, информированным согласием, рандомизацией, сокрытием вмешательства. Так же как в медицине, формируются базы данных, позволяющие с большей эффективностью исследовать социальные процессы. Эти сведения могут быть использованы для новых, более эффективных эпидемиологических исследований в интересах общественного здоровья. Вмешательства в обществе, нацеленные на повышение безопасности использования транспортных средств, профилактику преступлений или по-



вышение эффективности обучения в школе, ранее основывались лишь на идеях. Теперь они тоже проверяются в хорошо организованных экспериментах. Как данные контролируемых испытаний в медицине синтезируются с помощью систематических обзоров и метаанализа (см. разд. 5.1.2), так и данные социальных и психологических исследований обобщаются этими методами и в рамках международного сотрудничества<sup>1</sup>.

На рис. 2.48 приведены сведения из национального исследования в США об огромных различиях в летальности больных сахарным диабетом с почечной недостаточностью в зависимости от того, как скоро они получают доступ к нефрологу. Через 2 года от момента консультации летальность почти в 3 раза выше у получивших доступ к специалисту поздно (менее чем за 4 мес до начала диализа). Кто они, пациенты, получающие доступ к специализированной помощи позже? Это представители национальных меньшинств, лица, не имеющие страховки и имеющие тяжелые сопутствующие болезни. Сопоставляя характеристики пациентов, авторы учли многие биологические и демографические переменные, но на первом месте остались очевидные корреляты затрудненного доступа к специализированной помощи — принадлежность к афроамериканцам и бедность. Вероятно, истинный размер влияния препятствий к получению специализированной помощи на смертность еще больше, так как при прочих равных условиях люди, ранее получающие помощь, обычно имеют более тяжелое заболевание. Например, у получающих помощь раньше — быстро прогрессирующий рак (см. рис. 2.61).

Одно из первых направлений развития социальной эпидемиологии — развитие психосоциальной теории, объяснявшей влияние социальных факторов через влияние на психику человека, что в конечном итоге ведет к снижению резистентности человека к повреждающим факторам среды. Основными «вредными» факторами социальной среды виделись подчиненное положение в социальной иерархии, разрушение социальной организации, быстрые социальные изменения, маргинальное положение в обществе, включая социальную изоляцию, вдовство. Действием таких факторов пытались объяснить, почему отдельные социальные группы непропорционально подвержены туберкулезу, раку, психозам, самоубийствам. Эпидемиологические исследования показывали, что одним из средств повышения резистентности к действию этих факторов является развитие социальной поддержки.

<sup>1</sup> Кэмпбелловское сотрудничество — The Campbell Collaboration, см. <http://www.campbellcollaboration.org>.



Важным для социальной эпидемиологии подходом служит изучение состояния здоровья человека как следствия всей его предшествующей жизни, жизни его семьи (англ. *life course approach*). В частности, период внутриутробного развития стал рассматриваться как влияющий на состояние здоровья взрослого (англ. *Barker hypothesis*). Удивительно, что алкогольный синдром плода был описан только в 1973 г., но еще до 1989 г. алкогольные напитки не маркировались как опасные для беременных.

Социально-экономическая идея в социальной эпидемиологии приглашает исследовать, каким образом происходит распределение ресурсов в обществе, и как это сказывается на благополучии людей, какую политику проводит государство и в чью пользу. Основная гипотеза состоит в том, что социальные и экономические институты, которые ответственны за возникновение, сохранение и развитие неравенства и привилегий, являются причинами социального неравенства в состоянии здоровья.

Экосоциальная теория (середина 1990-х годов) предлагает взгляд на распределение болезней как на следствие динамического взаимодействия множества факторов. Это взаимодействие не сводится к линейной цепи причинных факторов или к плоской паутине взаимодействующих факторов. Взаимодействие является изменяющимся во времени и многомерным. Например, известно по крайней мере 100 лет, что ранние беременности уменьшают риск рака молочной железы, а поздние — увеличивают его. Первоначально этот факт рассматривался как свидетельство участия эндокринных факторов в канцерогенезе (беременность — экспозиция, влияющая на риск). В течение XX в. женщины в развитых странах стали позднее рожать детей. Этот сдвиг связан с изменением роли женщины в обществе, улучшением образования женщин. В рамках более широкой концепции беременность перестает быть изучаемой «экспозицией», вызывающей рак, а становится только лишь биологическим процессом, который изменяет чувствительность к канцерогенам. Экосоциальный подход создает возможность систематической интеграции знаний, полученных на разных уровнях изучения заболеваемости.

## Контрольные вопросы и задачи к главе 4

1. Выберите, какие из показателей а–г описывают приведенные ниже утверждения.
  - ▶ 1. Частота вторичного инфицирования.
  - ▶ 2. Летальность.



- ▶ 3. Частота заболеваний.
- ▶ 4. Стандартизованная смертность.
- ▶ 5. Общая смертность.
  - а) смертельным исходом заканчивается 8% случаев менингита;
  - б) в России в 1995 г. умирало 15 человек на 1000 живых в начале года;
  - в) в семьях, контактировавших с ребенком, больным ветрянкой, в 70% заболевает ребенок, если он восприимчив;
  - г) у детей 1–5 лет от роду в год наблюдается в среднем 8 вирусных инфекций.
- 2. При исследовании инфекционного заболевания были зарегистрированы все случаи. Сравнение заболевших и контрольных лиц показало, что относительный риск для связи заболевания с получением крови или продуктов ее переработки (фактор риска)  $\geq 1,0$ . Это означает:
  - а) нет связи между этим фактором и заболеванием;
  - б) фактор защищает людей от заболевания;
  - в) рандомизация или подбор пар были неудовлетворительными;
  - г) группа сравнения была неподходящей;
  - д) между фактором и заболеванием связи нет или она отрицательна.
- 3. Если у монозиготных близнецов конкордантность по некоторой болезни выше, чем у дизиготных, то это означает, что:
  - а) болезнь связана с действием внешних факторов;
  - б) болезнь врожденная;
  - в) различия объясняются полом;
  - г) действие внешних факторов имеет место наряду с действием генетических;
  - д) врожденные особенности играют основную роль, но имеет место и действие среды.

Выберите все правильные ответы.

- 4. При изучении инциденса болезни у приемных детей в сравнении с инцидентом у биологических родственников и у приемных родственников рассматривают признаки:
  - а) возраст заболевания;
  - б) наличие контакта у приемного ребенка с его биологическими родителями;
  - в) семейный статус биологических родителей;
  - г) фактор селекции (кого усыновили, и кого нет).



## Глава 5

# ЭПИДЕМИОЛОГИЯ И ОБЩЕСТВЕННОЕ ЗДОРОВЬЕ

Эпидемиологические методы являются естественным и лучшим инструментом для оценки не только отдельных медицинских вмешательств в клинической практике, но полезности этих вмешательств в широком социальном контексте, любых программ. *Общественным здоровьем* (англ. *public health*) называют науку и практику снижения смертности, заболеваемости и продления жизни. Для достижения этих целей существует широкий спектр методов. Их обоснованное применение до сих пор не стало правилом в деятельности медицинских и социальных служб так же, как не стало правилом использование только проверенных в испытаниях методов диагностики и лечения в клинической практике.

Эпидемиологическое исследование этиологии и патогенеза болезни нацелено на выявление связи экспозиции к потенциально вредному условию (фактору риска) с заболеванием. Это исследование проводится в условиях реально действующей системы оказания медицинской и социальной помощи, в условиях, когда действуют такие важные факторы,

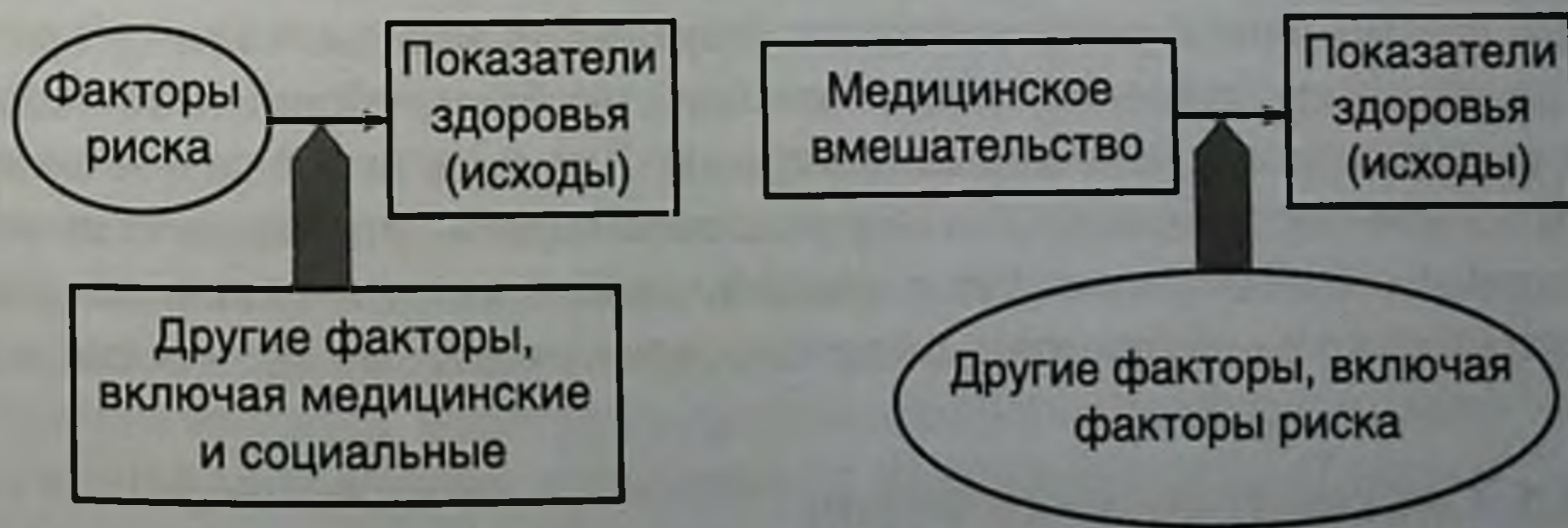


Рис. 5.1. Различия в эпидемиологическом исследовании влияния факторов риска на здоровье (слева) и исследовании влияния медицинских и социальных вмешательств



как ограниченный рабочий день, семейная поддержка и многие другие. Исследование строится таким образом, чтобы исключить влияние этих привходящих факторов и выявить связь, если таковая существует (рис. 5.1). Эпидемиологическое исследование эффектов оказания медицинской помощи или социальных вмешательств, наоборот, рассматривает факторы среды, в том числе потенциально вредные факторы производственной среды, другие факторы риска как вмешивающиеся. Соответственно исследование строится таким образом, чтобы уменьшить их влияние и выделить влияние изучаемого медицинского вмешательства на исходы болезней.

Анализ эффективности, продуктивности, затратной эффективности, социальной приемлемости медицинского вмешательства называют *оценкой медицинской технологии* (см. разд. 5.1.3). Такая оценка базируется на эпидемиологических данных, в том числе фармакоэпидемиологических, результатах контролируемых испытаний.

## 5.1. ОБОБЩЕНИЕ РЕЗУЛЬТАТОВ ОТДЕЛЬНЫХ ИССЛЕДОВАНИЙ

О преваленсе болезней обычно имеются сведения из разных источников. Существуют разные оценки частоты вредных эффектов при экспозиции вредным факторам на производстве. Испытания эффективности медицинских вмешательств проводятся повторно. Даже если исследования проводятся в сходных условиях, они дают различающиеся результаты, поскольку экспериментальные оценки имеют ограниченную точность, подвержены действию случайной и систематических ошибок. При рассмотрении различающихся результатов исследований у врача и организатора здравоохранения возникает проблема обобщенной оценки результатов нескольких исследований. Для этих целей разработаны технологии систематического обзора и метаанализа. Эти технологии исследования называют *синтетическими исследованиями*, так как они нацелены на синтез информации, содержащейся в отдельных исследованиях.

### 5.1.1. Систематический обзор

*Систематическим обзором* (англ. *systematic review*) называют обобщение результатов исследований по проблеме, отвечающее двум важ-



нейшим условиям. Первое: должен проводиться интенсивный поиск исследований по проблеме, позволяющий найти все или почти все исследования. Второе: найденные исследования критически оцениваются, и обобщаются только результаты высококачественных исследований. Если эти исследования достаточно однородны по своим характеристикам, то проводится *метаанализ* (англ. *meta-analysis*), т.е. количественное объединение результатов исследований для получения обобщенной оценки.

Поиск данных о проведенных исследованиях осуществляется по различным источникам — журналам, каталогам библиотек, коллекциям специалистов. Наиболее эффективным и быстрым является поиск информации в электронных библиографических базах данных. Самой известной и общедоступной является база данных MEDLINE<sup>1</sup>. Несмотря на высокое качество этой базы данных и ее оперативное пополнение, ограничиваться поиском в MEDLINE нельзя. Исследования показывают, что поиск в MEDLINE позволяет найти только примерно 50% существующих публикаций по отдельным проблемам. Это связано с объяснимой неполнотой коллекции библиотеки, преимущественным отражением в MEDLINE англоязычной литературы и трудностями описания журнальных статей ключевыми словами.

Не все журнальные статьи имеют такие заглавия и ключевые слова, какие кажутся необходимыми врачу, разыскивающему результаты исследований. Поэтому важна правильная формулировка запросов, позволяющая найти в базах данных максимальное количество сообщений (чувствительность поиска) и сделать это с достаточной точностью (получить в результатах поиска меньше статей, не соответствующих изучаемой проблеме). Этим задачам соответствуют стратегии поиска, разработанные специалистами, объединенными в международную организацию Кокрейн (Cochrane, ранее «Кокрейнское сотрудничество», названо по имени А. Кокрейна, идеолога обобщения результатов исследований для врачей). Стратегии доступны на интернет-сайте организации и в электронной Кокрейнской библиотеке<sup>2</sup>. Библиотека существует в электронном виде и состоит из нескольких разделов. Главным является коллекция систематических обзоров, составленных по единой методологии и называемых Кокрейнскими обзорами. Эти

<sup>1</sup> Кэмпбелловское сотрудничество — The Campbell Collaboration, см. <http://www.campbellcollaboration.org>.

<sup>2</sup> Кокрейнское сотрудничество, см. <http://www.cochrane.org>.



обзоры являются наиболее надежными обобщениями данных об эффектах медицинских вмешательств (пример см. рис. 5.2). Тип графика, на котором относительно линии отсутствия эффекта (различий) изображаются результаты отдельных испытаний, называют *лесным* (англ. *forest plot*).

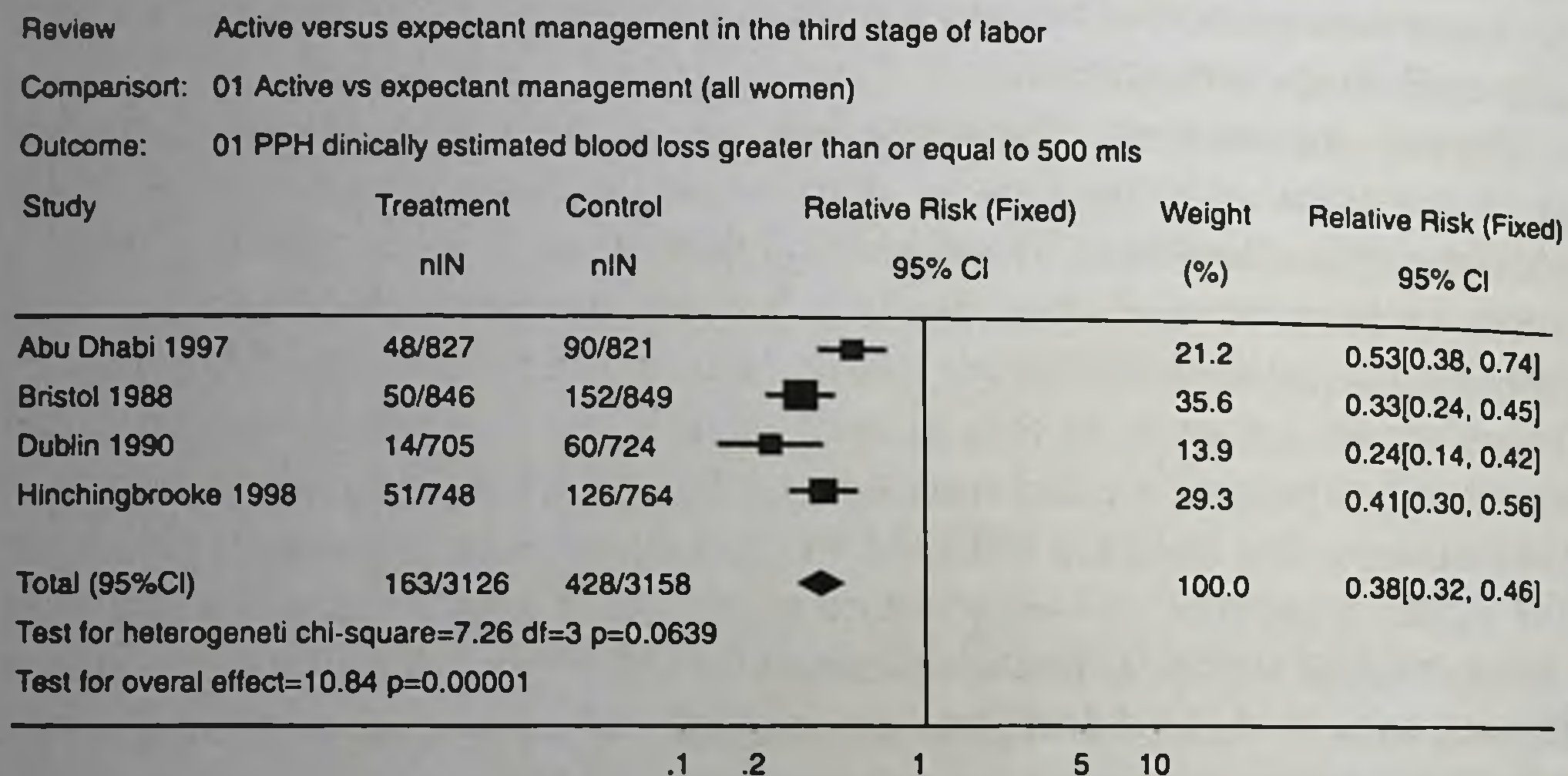


Рис. 5.2. Обобщающее представление данных Кокрейновского обзора, сравнивающего активное и выжидательное ведение родов в третьей стадии (Prendiville W.J. et al. Active versus expectant management in the third stage of labour. In: The Cochrane Library, Issue 4, 2003. Oxford: Update Software). На рисунке сравнивается вероятность кровопотери 500 мл и более при двух вмешательствах. Обобщаются данные четырех испытаний (фамилии авторов слева). Результаты представлены в виде относительного риска (ОР, Relative Risk, в центре, шкала внизу). ОР=1,0 означает отсутствие различий в исходах сравниваемых вмешательств. Квадрат — оценка ОР в отдельном испытании вмешательства, горизонтальная линия — доверительный интервал (ДИ), размер квадрата соответствует условному весу, отражающему точность оценки (Weight, чем больше, тем точнее оценка). Хорошо видно, что суммарная оценка (ромб) меньше единицы, и что ДИ (длина ромба) для суммарной оценки не включает единицу. Значит, активное ведение родов ведет к меньшей кровопотере с 95% ДИ 0,32–0,48

После того как статьи (отчеты об испытаниях или других исследованиях) найдены, должно быть оценено их качество, т.е. вероятность смещения. Оценка качества рассматривается в каждом из разделов этого учебника, посвященных отдельным структурам исследований. Например, для испытаний лечебных вмешательств обязательными критериями качества являются: проспективный характер испытания, на-



личие контрольной группы, распределение участников в сравниваемые группы с использованием рандомизации, сокрытие результатов рандомизации, слепая оценка исхода лечения. Для того чтобы обеспечить объективность оценки качества отчетов об испытании, оценку качества проводят два специалиста независимо друг от друга, с последующим обсуждением расхождений. Отобранные для анализа исследования, отвечающие критериям качества, сводят в таблицы. В случае если исследования проведены на различающихся пациентах или в различающихся условиях, то применить к ним математические методы объединения результатов нельзя. В случае если исследования достаточно однородны, выполняется метаанализ.

### 5.1.2. Метаанализ

*Метаанализом* называют математическое обобщение результатов нескольких исследований. Метод может быть применен не только к результатам клинических испытаний, но и к результатам обсервационных исследований. В основе метаанализа лежит простая идея — получить более точную оценку эффекта вмешательства, объединяя данные многих небольших исследований, каждое из которых в силу малого числа участников дает неточную оценку.

Целью метаанализа является получение неискаженной оценки эффекта вмешательства. Качество метаанализа обеспечивается не применением сложных расчетов, а исчерпывающим поиском всех исследований, в том числе неопубликованных. Метаанализ является вариантом описательного исследования. Его объекты — не проявления болезни в группе больных, а оценки эффективности вмешательства, полученные в группе исследований.

Наиболее важной проблемой для метаанализа является *публикационное смещение*. Публикационным смещением называют преимущественное обнародование исследователями результатов исследований, принесших статистически значимый результат, подтвердивших гипотезу, продемонстрировавших эффективность вмешательств, выполненных на большом количестве пациентов. Ответственность за выборочную публикацию несут не только исследователи, но и редакции журналов, чаще отвергающие статьи с «отрицательным» результатом, и диссертационные советы, неблагоприятно относящиеся к диссертациям, содержащим «отрицательные» результаты.



## Метод Пито

Рассмотрим выполнение метаанализа двух гипотетических контролируемых испытаний с помощью метода Пито (табл. 5.1). Формулы для расчета приведены в табл. 5.2. Расчеты для примера — в табл. 5.3. Этот метод расчета суммарного отношения шансов для качественных исходов используется в Кокрейновских обзорах.

Таблица 5.1. Данные двух испытаний терапевтического вмешательства

	Вмешательство	Плацебо	Всего
Исследование № 1			
Умерли	18 (a)	30 (b)	48 (g)
Выжили	100 (c)	100 (d)	200 (h)
Всего	118 (e)	130 (f)	248 (n)
Отношение шансов = 0,60			
Исследование № 2			
Умерли	35	200	235
Выжили	130	500	630
Всего	165	700	865
Отношение шансов = 0,67			

Таблица 5.2. Формулы для расчета суммарного отношения шансов по методу Пито и 95% доверительного интервала

Суммарное отношение шансов	$ОШ_p = e^{\Sigma(O_i - E_i) / \text{sum\_variance}_i}$
То же, поэтапный расчет	$\ln ОШ_p = \Sigma(O_i - E_i) / \text{sum\_variance}_i$
	$E_i = (e_i \times g_i) / n_i$
	$\text{Variance} = (E_i \times f_i \times h_i) / [n_i \times (n_i - 1)]$
Доверительный интервал, 95%	$e^{\ln ОШ_p \pm 2 / \sqrt{\text{sum\_variance}_i}}$



Таблица 5.3. Расчеты суммарного отношения шансов и 95% доверительного интервала по методу Пито применительно к данным табл. 5.1

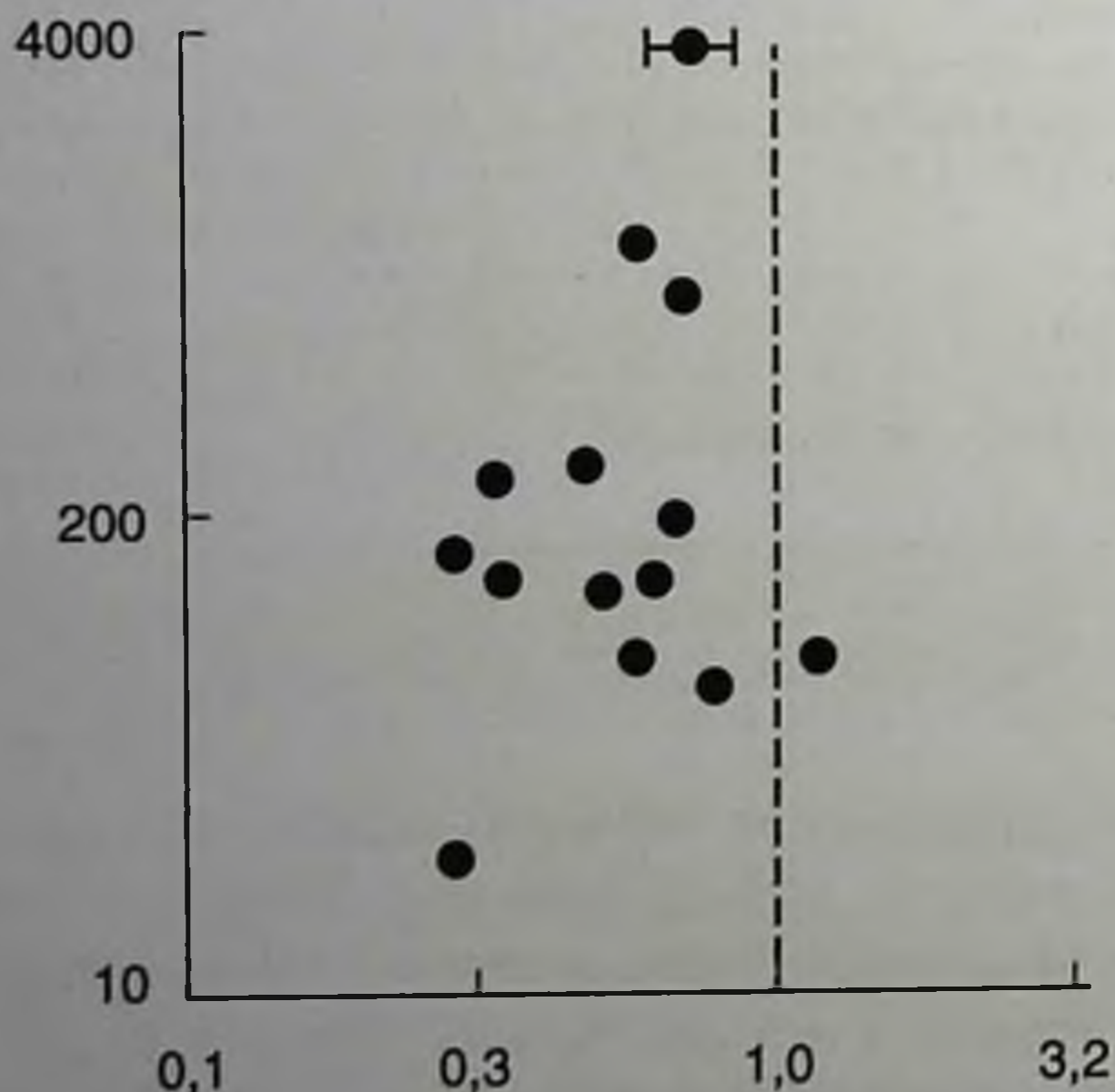
Ожидаемая частота $E_i = (e_i \times g_i) / n_i$ для исследования № 1	$118 \times 48 / 248 = 22,8$
Для исследования № 2	$165 \times 235 / 865 = 44,8$
Разность наблюдаемой и ожидаемой частот для исследования № 1 ( $O_1 - E_1$ )	$18 - 22,8 = -4,8$
Для исследования № 2 ( $O_2 - E_2$ )	$35 - 44,8 = -9,8$
Дисперсия для разности наблюдаемой и ожидаемой частот ( $E_i \times f_i \times h_i / [n_i \times (n_i - 1)]$ ) для исследования № 1	$(22,8 \times 130 \times 200) / (248 \times 247) = 9,7$
Для исследования № 2	$(44,8 \times 700 \times 630) / (865 \times 864) = 26,4$
Сумма разностей наблюдаемых и ожидаемых частот $\Sigma(O_i - E_i)$	$-4,8 + -9,8 = -14,6$
Суммарная дисперсия $\text{sum\_variance}_i$	$9,7 + 26,4 = 36,1$
Логарифм ОШ <sub>p</sub>	$-14,6 / 36,1 = -0,4$
Суммарное ОШ <sub>p</sub> = $e^{-0,4}$	0,67
95% доверительный интервал, верхняя граница = $e^{-0,67+0,33}$	0,71
Нижняя граница = $e^{-0,67-0,33}$	0,36

Метод Пито, применение которого рассмотрено во врезке, представляет собой вариант метода Мантеля–Хэнзеля и требует, чтобы для каждого исследования были доступны полные таблицы данных, т.е. данные об исходах в каждой группе. Существует ряд других методов метаанализа, имеющих свои достоинства. С помощью этих методов можно обобщать не только качественные исходы, но и количественные данные и данные типа времени жизни. Для выполнения расчетов и графического анализа данных созданы компьютерные программы, из которых чаще используются EPIMETA и RevMan. Последняя программа бесплатна для участников Кокрейн ([www.cochrane.org](http://www.cochrane.org)).



На рис. 5.3 приведены данные метаанализа испытаний кортикостероидов для предотвращения смерти у младенцев, рожденных в преждевременных родах. Этот тип графика, на котором результаты отдельных исследований размещаются относительно шкалы шансов, а по ординате откладывается показатель точности оценки эффекта, называется *вороночным* (англ. *funnel plot*). При отсутствии публикационного смещения воронка должна быть почти симметричной, сужающейся кверху, где находятся точки, принадлежащие самым точным оценкам, полученным в самых больших по численности участников исследованиях. При публикационном смещении «воронка» оказывается несимметричной из-за отсутствия на графике точек, соответствующих малому эффекту вмешательства.

Метаанализ иногда выявляет, что, несмотря на то, что решение проблемы уже найдено, действенность вмешательства и размер эффекта установлены, — продолжают проводиться все новые исследования. Это может быть объяснимо неосведомленностью исследователей о том, что имеющиеся данные уже позволяют оценить размер эффекта и уве-



**Рис. 5.3.** Результаты метаанализа испытаний инъекции кортикостероидов при преждевременных родах для предотвращения смерти младенца. Абсцисса — отношение шансов, ордината — число пациентов в испытании; обратите внимание на логарифмический характер обеих шкал. Вверху — суммарное отношение шансов и его доверительный интервал. Пунктир — уровень отсутствия эффекта (по данным Crowley P. Am. J. Obstetr. Gynecol. 1995. № 173. P. 322–35)



ренно применять вмешательство. Для того чтобы избежать ненужных исследований, государственные органы, финансирующие медицинские исследования в ряде европейских стран стали предоставлять средства только при условии, если заявители с помощью систематического обзора показали, что решения проблемы еще не найдено, например преимущества вмешательства не установлены. Повторное выполнение испытаний вмешательств, полезность которых установлена, может объясняться финансовыми интересами производителей лекарственного препарата — для привлечения внимания врачей к своему продукту.

Метаанализ результатов отдельных небольших исследований может давать убедительные оценки тогда, когда неформальное освоение результатов проведенных исследований врачами еще не указывает на достижение надежного знания. Например, как показал *ретроспективный* анализ, выполненный в 1992 г., данных об эффективности тромболиза при инфаркте миокарда было достаточно уже в 1975 г., но учебники и обзоры стали рекомендовать это лечение только после 1985 г. Наличие убедительных научных данных, обобщенных в метаанализе, и официальные рекомендации руководителей здравоохранения и медицинских про-

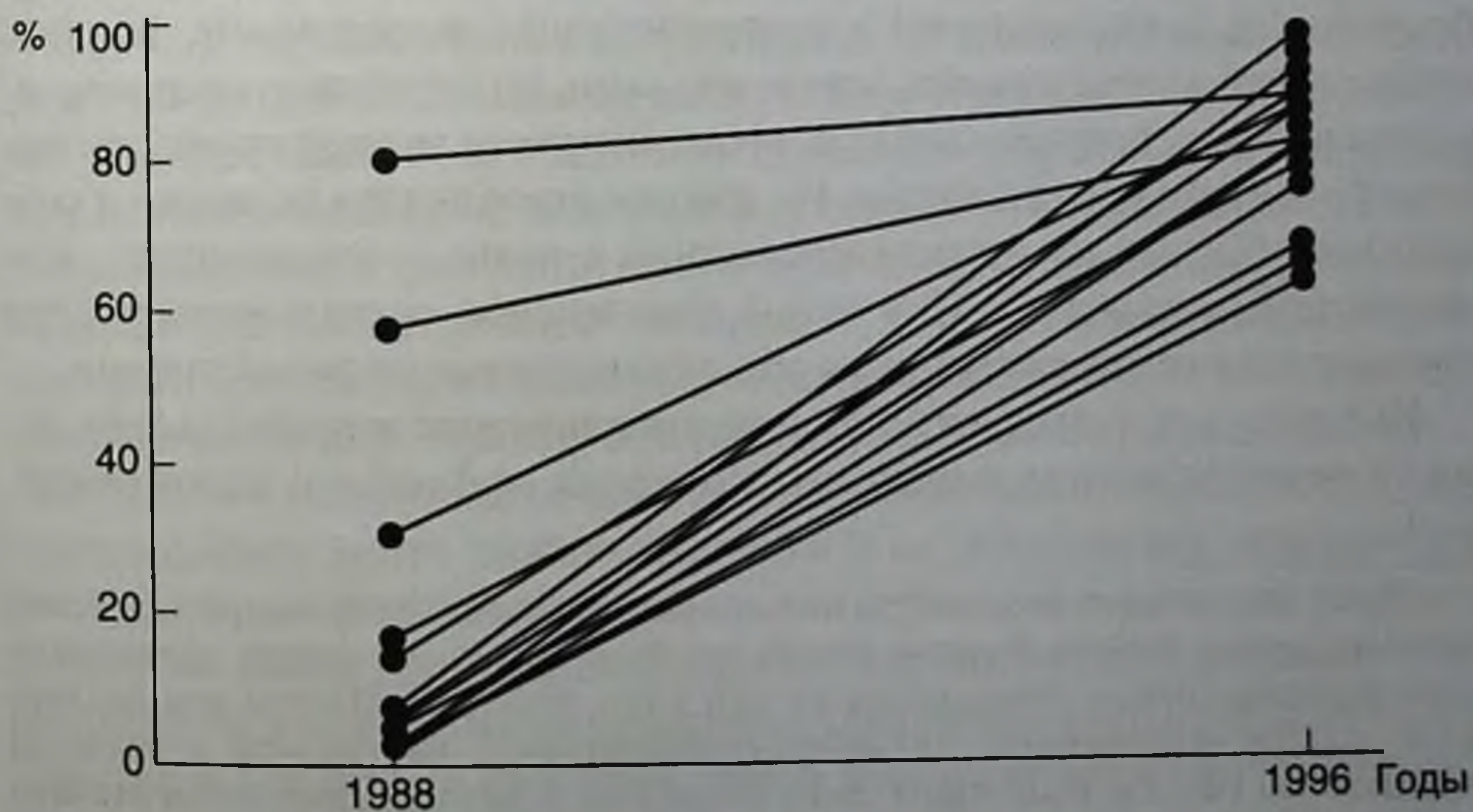


Рис. 5.4. Изменения в частоте соответствия практики ведения преждевременных родов стандартам аудита в Англии и Уэльсе. Включены в анализ только случаи родов, проходивших в обстановке, позволявшей по времени использовать препарат. Каждая линия соединяет результаты аудита в одном учреждении в разные годы (по Lilford R. J. *Epidemiol. Commun. Health*. 2000. № 54. P. 731–739; с разрешения издателя)



фессиональных организаций способны изменить практику, подвигнуть врачей к использованию эффективных вмешательств. Например, в Великобритании некоторые акушеры использовали кортикостероиды при преждевременных родах и в середине 1980-х годов (рис. 5.4). Однако только после упомянутого метаанализа (см. рис. 5.3) и выхода рекомендаций Коллегии акушеров такое лечение стали получать абсолютное большинство нуждающихся в нем рожениц.

В результате систематического обзора и метаанализа уточняются представления о приложимости результатов испытаний в изученной популяции и в других популяциях. Это называют *обобщаемостью*, или *генерализацией результатов* (англ. generalisability)<sup>1</sup>.

### 5.1.3. Комплексная оценка медицинских вмешательств

В оценке медицинской технологии (процедуры, метода, программы) можно условно выделить три этапа. Первый этап — сугубо технологический. На нем оценивают технические характеристики — точность и воспроизводимость измерения, действенность и токсичность лекарства. На втором этапе оценивают диагностическую или терапевтическую эффективность (действенность) в контролируемом эксперименте. На этом этапе оценка ограничивается определенными, но не вполне совершенными «конечными точками» анализа, вроде интервала до первого обострения или продолжительности жизни. На третьем этапе анализ включает в себя возможно полный комплекс естественных оценок — клинических, экономических, социальных. Все этапы комплексной оценки выполняются с помощью методов, являющихся объектами изучения эпидемиологии.

Решение о том, следует ли применять некоторое вмешательство, зависит от соотношения ожидаемых полезных эффектов и осложнений.

---

<sup>1</sup> Врач, изучающий результаты испытания вмешательства, например испытания лекарств, должен ответить на два вопроса: можно ли верить результатам этого исследования, и приложимы ли они в его практике. Иногда эти две стороны оценки исследования называют соответственно внутренней и внешней валидностью (В). По аналогии с лабораторными и психологическими тестами В определяется как соответствие результатов измерения тому, что предполагается измерять. Внутренняя В — соответствие результатов испытания поставленному вопросу и выборке из популяции, в которой проведен эксперимент, т.е. отсутствие систематических ошибок. Внешняя В — приложимость результатов эксперимента к подобным пациентам в иных условиях — в других странах, на пациентах другого пола и возраста, с сопутствующими заболеваниями и т.д.



Например, частота кровотечения — известного осложнения приема нестероидных противовоспалительных средств — низка. Поэтому осложнения, редко тяжелые, не препятствуют применению этих лекарств в случаях, когда их применение показано. Лишь у детей при приеме ацетилсалициловой кислоты может очень редко возникать тяжелое, угрожающее жизни осложнение — синдром Рейе, и поэтому в некоторых странах существует рекомендация не применять его у детей до 12 лет.

В случае если польза вмешательства не столь велика или вред достаточно велик, необходимо соотнести пользу и вред. Решение должно основываться на специальном количественном анализе *соотношения пользы и вреда*<sup>1</sup>. Сущность этого анализа выражается правилом: лечение должно приводить к сохранению большего числа жизней (большого числа лет жизни), чем его альтернатива, например выжидательная тактика, при учете всех возможных осложнений. Кроме этих медицинских оснований для предоставления населению доступа к вмешательству, есть еще немедицинские факторы (условия). Важнейшие из них — экономические — могут быть включены в комплексную оценку.

### Принципы оценки

Подход, основанный на взвешивании полезных и вредных последствий вмешательства, используется во всех областях социальной жизни, выходит далеко за пределы медицины, общественного здоровья и социальной практики. Использование сверхзвуковых самолетов вместо дозвуковых, строительство атомных электростанций в сравнении со строительством тепловых — все может, должно быть и действительно становится объектом анализа соотношения пользы и вреда. Исследования в этой области, помимо объективных данных исследований, включают в себя учет субъективных представлений людей об опасности. Эти представления могут сильно отличаться от объективных показателей риска, например от показателей смертности в связи с использованием отдельных видов транспорта. С другой стороны, анализ рисков учитывает политические аспекты проблемы. Например, в США сравнительно с другими развитыми странами много людей гибнет от огнестрельных ран, поскольку многие граждане имеют в своем распоряжении оружие. Вероятно, простой запрет на обладание оружием мог бы быстро при-

<sup>1</sup> Аудит — проверка деятельности независимыми экспертами на соответствие установленным критериям. Применяется в различных областях деятельности, прежде всего — в финансах. В медицине впервые применен Ф. Найтингейл.



вести к снижению смертности от огнестрельных ран, но в этой стране большая часть населения полагает, что нуждается в нем для самозащиты и считает право на обладание оружием важнейшей гарантией свободы.

В контролируемом испытании дается ответ лишь на первый и важнейший вопрос — о действенности метода или процедуры. Дальнейшее лежит уже частично за пределами контролируемого эксперимента: приносит ли процедура в действительности такие результаты, как планировалось, приемлема ли она для тех, кому предназначалась? Это вопрос об эффективности и полезности. Вопрос третий: достигает ли процедура тех людей, которые могут от нее получить пользу? Это вопрос о доступности. Вопросы о целесообразности использования именно этого, а не другого метода диагностики или лечения, предпочтение программы профилактики, а не программы развития методов лечения, предпочтение финансирования данного метода профилактики, а не другого, должны рассматриваться на основе сопоставления всех аспектов их эффективности. Очевидно, что должно отдаваться предпочтение тому методу, той программе, которая при одинаковых затратах и приемлемости дает больше пользы, или той, которая приносит большую пользу при одинаковых затратах. Проблема состоит в том, что представления о пользе различны, и измерение пользы является не простой задачей.

Экономический анализ может проводиться только как сравнительное исследование. Изучение расходов на отдельное вмешательство лишено смысла. Эффективность сравниваемых вмешательств и их цена могут рассматриваться как взаимно перпендикулярные (рис. 5.5). В случае если новое вмешательство приносит больший эффект и цена его ниже (квадрант 1), решение о переходе на такой способ действий очевидно. Столь же проста ситуация, когда эффект меньше, а цена — выше (квадрант 3). Это вмешательство неперспективно. Большая часть медицинских вмешательств не обладают очевидными достоинствами или недостатками, т.е. принадлежат 2-му и 4-му квадрантам на схеме. Они при более высокой эффективности имеют и более высокую цену или при меньшей цене имеют и меньшую эффективность. Более высокая цена может быть приемлемой платой за высокую эффективность, а дешевизна может искупать некоторое снижение эффективности. Для ответа на вопрос, приемлема ли плата, искупается ли дешевизной снижение эффективности, необходимо выполнять количественный анализ.

При проведении экономического анализа должна быть ясной перспектива анализа: чьи интересы, чьи исходы ставятся во главу угла.



Эффект	Больше
1	2
Ниже	Цена Выше
4	3
	Меньше

Рис. 5.5. Принципиальная схема экономического анализа.  
Пояснения в тексте

Обычно интересы специфических групп граждан, например, больных сахарным диабетом, лучше учитываются, если рассматриваются варианты вмешательства применительно к этим больным. Если сопоставляются перспективы популяционной профилактики, то интересы отдельных групп могут оказаться незамеченными. В таких анализах берет верх *общественная перспектива*.

Возможно проведение анализа, ограниченного результатами вмешательств и затратами в рамках отдельной системы здравоохранения или отдельной страховой организации либо только стоимостью фармакотерапии. Результат такого анализа не может быть распространен на национальную систему и на процесс оказания всесторонней помощи, включая стационарное лечение. Еще более узкие по учету расходов и последствий анализы могут рассматривать перспективу плательщика, например государства или страховщика, и тогда прямые затраты самого пациента могут выпасть из поля зрения.

### Варианты оценки

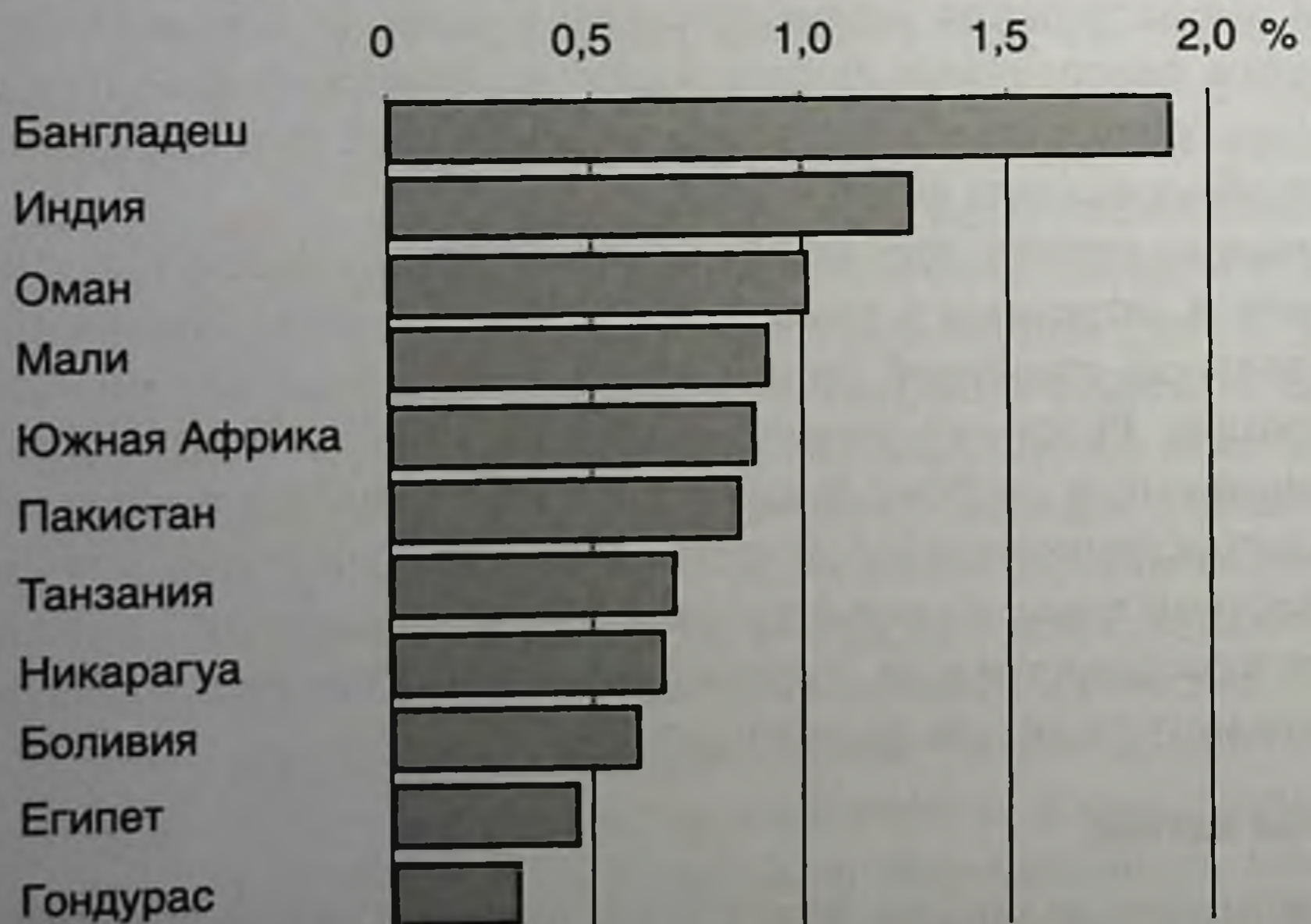
Экономическая оценка может быть полной и частичной. Частичной является оценка только одного компонента — пользы или затрат. Пример оценки затрат — оценка общественных потерь в связи с анемией и связанными с ней болезнями, нарушениями индивидуального развития, нетрудоспособностью (рис. 5.6). Пример оценки пользы — метаанализ



результатов испытаний лекарств для лечения артериальной гипертензии. Полная оценка имеет место, когда взвешиваются альтернативы и рассматриваются затраты и последствия вмешательств. Так же как клиническое испытание невозможно без сравнения (контроля), точно так же невозможно изучать соотношение затрат и пользы для отдельного вмешательства.

В полной экономической оценке выделяют 4 вида анализа (табл. 5.4).

**А. Минимизация затрат** (англ. *cost minimization analysis*). При сопоставлении медицинских процедур (программ), имеющих одинаковую действенность, можно сравнить затраты на них. В результате можно сделать выбор в пользу более экономной практики. Такой анализ проводят после завершения клинических испытаний методов лечения или диагностики, если эти испытания не выявляют существенных различий сравниваемых методов. Этот вид анализа практически идентичен анализу затрат и встречается редко. Может быть применен в специального вида испытаниях, в которых при равных бюджетах сравниваются исходы лечения (англ. *outcome maximization design*).



**Рис. 5.6.** Экономические последствия анемии в процентах от совокупного национального продукта в группе стран. По отчету Международной продовольственной организации ООН (UN FAO annual report «The State of Food Insecurity in the World 2002»); <http://www.fao.org/DOCREP/005/Y7352e/Y7352e00.HTM>



Таблица 5.4. Измерения последствий в комплексной оценке альтернативных процедур

Тип оценки	Последствия	Измерение последствий, единица
А. Минимизация затрат	Идентичны во всех важных аспектах	Нет
Б. Затрат и эффективности	Единый эффект, общий для альтернатив, но достигаемый в разной степени	Естественные единицы — годы добавленной жизни, дни трудоспособности, снижение артериального давления в мм рт.ст. и т.д.
В. Затрат и выгоды	Единый или многие эффекты, не обязательно общие для альтернатив	Денежные единицы
Г. Затрат и полезности	То же	Дни здоровой жизни, годы доброкачественной жизни

**Б. Анализ соотношения затрат и эффективности** (англ. *cost-efficiency analysis*). Если самая дешевая программа не является одновременно самой эффективной, то возникает проблема сопоставления эффекта и необходимых затрат. Эта проблема решается отношением цены к единице эффекта (например, к сохраненному году жизни). Метод, которым прирост года жизни пациентов достигается большими расходами, хуже. Если существует некоторый бюджет, то может быть использована обратная единица — число сохраненных лет жизни на затраченный 1 млн рублей. Таким образом можно сравнить между собой вмешательства и целые программы не качественно, как это представлено на рис. 5.5, а количественно.

В рамках анализа затрат и эффективности возможно *сравнение любых процедур*. Можно сопоставить борьбу с курением и флюорографию грудной клетки, пересадку почек и обязательное использование ремней безопасности в автомобиле. Возможно это потому, что используются общая единица измерения последствий и общая единица цен, а также потому, что влияние процедуры на смертность (инвалидность, длительность нетрудоспособности) не обязательно должно быть прямым (диагностические процедуры, например, влияют косвенно).

**В. Анализ соотношения затрат и выгоды** (англ. *cost-benefit analysis*). Оценка действенности вмешательства, например, в виде контролиру-



емого испытания может выявить сдвиги только в продолжительности выживания или других прямо наблюдаемых исходов. В действительности каждая медицинская процедура, будучи введенной в практику, вызывает изменения во всех аспектах расходов на медицинскую помощь и во всех аспектах результатов этой помощи. Поэтому при сопоставлении альтернативных программ возникает необходимость сопоставлять, с одной стороны, увеличение продолжительности жизни, а с другой, например, увеличение продолжительности трудоспособности до наступления инвалидности плюс снижение временной нетрудоспособности.

Исходы можно сравнить в денежных единицах, как это делают в страховании. Страховщики выражают медицинские и социальные последствия (например, потерю трудоспособности) в денежных единицах, и общество благосклонно принимает такую практику. Результаты подобного анализа выражаются как результаты любой деятельности — в рублях выгоды и потерь или в отношении выгоды к потерям. Эти результаты можно сравнить у двух и более программ и выбрать лучшую. Одной из сравниваемых программ может быть стратегия «ничего не делать» — это стратегия экономии всего, что входит в цену. При ее реализации выгода равна экономии. Обычно ресурсы здравоохранения фиксированы, и речь идет лишь об оптимальном их использовании, например для сохранения дорогостоящей системы туберкулезных диспансеров или для обучения врачей общей практики.

Г. *Анализ соотношения затрат и полезности* (англ. *cost-utility analysis*). Если сравниваемые вмешательства приводят к изменению разных исходов, то их прямое сравнение по исходам невозможно. Однако возможно применение интегральной оценки типа QALY (продолжительность жизни, скорректированная на качество жизни) или DALY (продолжительность жизни, скорректированная на инвалидность) (см. разд. 3.4.2). После того как различные исходы приведены к единому показателю, можно сравнивать цену одного года жизни, скорректированного на качество жизни, при различных вмешательствах.

Применение экономических оценок в виде отношения затраты/польза может приводить к несправедливости. Лечение одинаковой болезни сердца в кардиологическом центре столицы обойдется дешевле для столичного жителя, чем для жителя Камчатки. Получается, что можно в рамках бюджета помочь большему числу людей, если лечить только столичных жителей. Лечение одинаковой болезни сердца приведет к большему приросту «пользы», если оперировать работоспособного



мужчину среднего возраста, а не пенсионера или инвалида. Соотнесение расходов с годами жизни есть лишь математическая операция и всегда следует задаваться вопросом, почему процедура расчета является именно такой, а не более сложной, и правильно ли была выбрана перспектива анализа. Экономическая оценка не является волшебным средством. Она не может исправить ошибок клинического испытания. Это лишь метод критического осмысления медицинских процедур в реалиях жизни. Ни один из описанных выше методов не является лучшим и исчерпывающим. Во многих исследованиях экономическую оценку проводят несколькими методами.

Комплексная оценка медицинских технологий наряду с доказательной медициной является вариантом трансляционной науки, трансляционных исследований (англ. *translational research*). Этим термином обозначают исследования, направленные на создание новых, практически важных вмешательств для медицины на основе достижений фундаментальной науки и на их продвижение в практику.

#### 5.1.4. Анализ принятия решений

После того как получены данные об отдельных вмешательствах и оценены экономические аспекты вмешательств, возможно проведение суммарной оценки процесса оказания помощи — *анализ принятия решений* (англ. *decision analysis*). На рис. 5.7 приведен пример оценки последствий применительно к решению, ревакцинировать или нет детей в школе, привитых ранее однократно, после возникновения серии новых случаев кори. Решение, принимаемое ответственным специалистом, обозначается на дереве решений прямоугольником. В нашем примере оно одно. Все остальные последствия наблюдаются или не наблюдаются вероятностно, что обозначается кружками. У привитых и непривитых вероятность экспозиции к инфекции (Э) одинакова — 0,2. Вероятность противоположного события не обозначается, так как она дополнительна к этой, т.е. 0,8. У непривитых и привитых вероятность заболевания вследствие контакта различается, что отражено на соответствующих ветвях как вероятность К. Вероятности смерти (М) от кори у заболевших и выживания (Ж) обозначена на последнем ветвлении.

Анализ дерева решений начинается справа от самых последних событий, где отражены интересующие исследователя исходы. Последовательно используя вероятности, реализующиеся на каждом ветвлении,



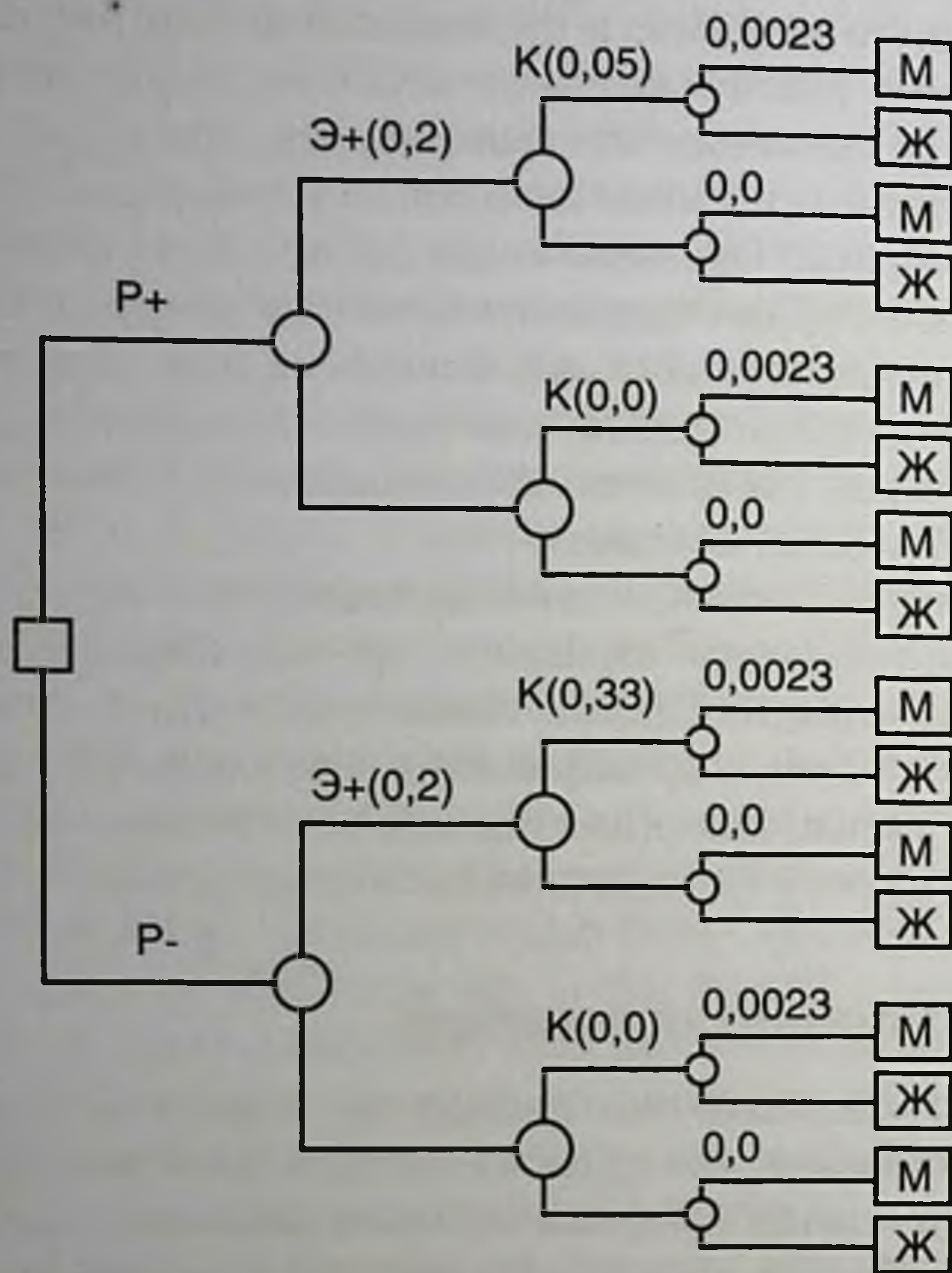


Рис. 5.7. Пример анализа последствий принятия решения о ревакцинации ( $P_+$ ) против кори. Пояснения в тексте

врач приближается к первому главному ветвлению, соответствующему поставленному перед анализом вопросу, а именно: каков выигрыш от ревакцинации всех детей в числе смертельных исходов.

Выполнение такого анализа возможно вручную, он успешно может быть реализован средствами, предоставляемыми электронными таблицами, но удобнее всего его выполнять с помощью специализированных компьютерных программ, например TreeAge (TreeAge Software). Специализированные программы не только позволяют гибко изменять структуру дерева, приближая ее к реальной задаче, но и позволяют включить в рассмотрение длительности отдельных этапов лечения, стоимость отдельных вариантов вмешательств, приемлемость для больных отдельных исходов (качество жизни).



## 5.2. ОЦЕНКА ДЕЯТЕЛЬНОСТИ МЕДИЦИНСКИХ УЧРЕЖДЕНИЙ И ПРОГРАММ

Во всем мире работа врачей, медицинских и социальных учреждений подвергается периодической оценке. Эта оценка является объективно необходимой, так как медицинские и социальные программы обладают высокой значимостью для общества и поглощают огромные ресурсы нации.

Наибольшее влияние на современные представления о том, как функционирует система оказания медицинской помощи, оказали взгляды А. Донабедеяна (A. Donabedian), классифицировавшего происходящее в этой системе на структуры, процессы и исходы.

Таблица 5.5. Компоненты оцениваемой клинической практики

Клиническая практика		
Структура	Процессы	Исходы
Подразделения, оборудование, квалификация персонала	Технические (тестирование, физиологический мониторинг, назначение лекарств, выполнение других процедур), личностные (обучение пациента)	По здоровью пациента (смертность, инвалидность, симптомы, функциональное и психологическое благополучие)

Под структурами понимается сочетание компонентов организации, в которой предоставляется медицинская помощь. Структуры относительно стабильны и очевидным образом влияют на то, какая помощь предоставляется.

Широкий спектр активности в отношении медицинской помощи называется «процессами» — от решений о финансировании до взаимоотношений между врачами и диагностическими лабораториями. Представления о том, каковы «правильные» процессы, в медицине традиционно базировались на мнениях экспертов. Поэтому врачи XVIII в. «отворяли кровь», а в 1940–1950 гг. при выхаживании незрелых младенцев следили за тем, чтобы в кувезах поддерживалась атмосфера с содержанием кислорода, близким к 100%. И та и другая практики были вредными, что стало известно после многих лет их применения. Важнейшими документами, регламентирующими процессы оказания меди-



цинской помощи, являются клинические рекомендации (стандарты). Современные варианты этих документов базируются на результатах доказательных исследований.

Показатель «исходы» кажется очевидным для оценки медицинской практики и социальных вмешательств. В действительности в этой области существуют противоречия. Акцентируя внимание на исходах у отдельных пациентов, врачи приходят к одним выводам, а обращаясь к исходам в *группах* пациентов (обобщенным), — к другим. Отсюда вытекают противоречия, например, между обоснованием расходов на дорогие, но редкие лечебные вмешательства для больных, и дешевые, но массовые вмешательства.

Современная медицина, называемая *доказательной медициной*, ставит во главу угла исходы медицинских вмешательств, и провозглашает, что в отношении отдельного пациента должны применяться только вмешательства, эффективность которых доказана в доброкачественных исследованиях. Соответственно, *показатели процесса* оказания медицинской помощи должны устанавливаться, исходя из того, насколько полезны эти процессы для достижения важных исходов.

Деятельность в области общественного здоровья, которая признает возможным применение только тех вмешательств, полезность которых доказана, называют *доказательным общественным здоровьем* (англ. *evidence based public health*). Способ финансирования вмешательств является ключевым в определении стоимости помощи, ее доступности и структуры расходов. В этой области сформировался подход, называемый *доказательным финансированием* (англ. *evidence based purchasing*). Он заключается в оплате в рамках страховой или государственной программы только тех вмешательств, полезность которых доказана в доброкачественных исследованиях.

Подход, основанный на оценке исходов, кажется интуитивно наиболее важным. К несчастью, оценка исходов даже применительно к однородной группе больных не всегда выявляет надежную связь между вмешательством и исходом. На популяционном уровне отдельные медицинские вмешательства, как правило, не могут быть надежно связаны с изменениями в состоянии здоровья населения. Например, несмотря на то что хирургическое лечение инфаркта миокарда на уровне группы пациентов эффективно, остается загадкой, какой вклад в снижение смертности от сердечно-сосудистых болезней в последней четверти XX в. внесли хирургические вмешательства. Однако в отдельных



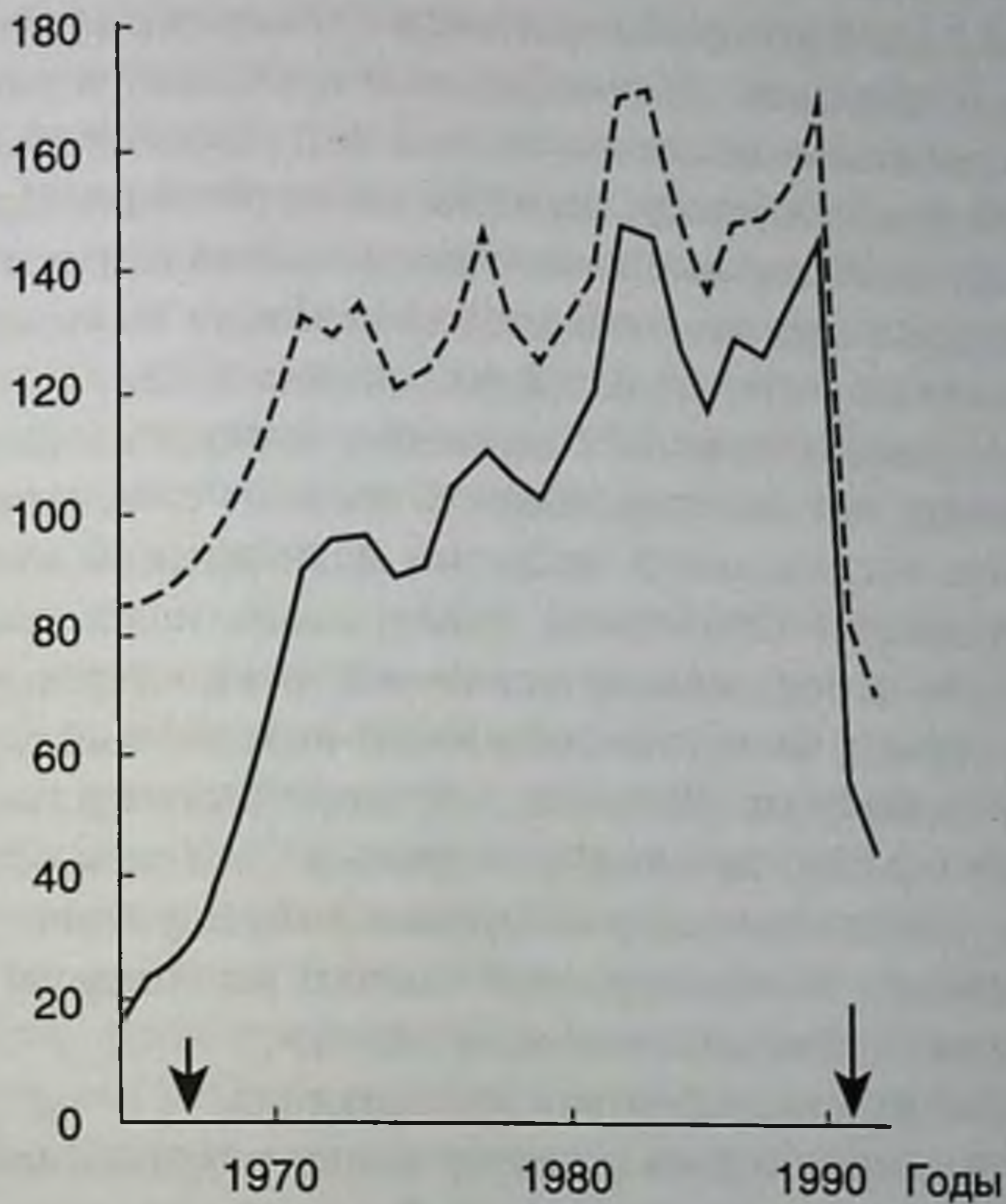


Рис. 5.8. Материнская смертность (ордината, на 100 000 живорождений) в Румынии в 1965–1992 гг. Аборты были запрещены в 1966 г. и разрешены в 1991-м. Верхняя кривая — общая материнская смертность, нижняя — связанная с абортами. Данные Министерства здравоохранения Румынии

случаях даже официальная статистика, инструмент, подверженный систематическим ошибкам, дает убедительные аргументы в пользу конкретных медико-социальных решений (рис. 5.8).

Из-за трудностей в оценке исходов менеджеры обычно анализируют процессы и структуры, но не исходы. *На уровне отдельного пациента* исходы неопределенны, поскольку зависят от привходящих факторов, например от возраста или от сопутствующих болезней. Врач ориентируется на лучшие исходы, доступные при данной болезни при применении адекватных средств, но из-за того, что исход зависит от многих факторов, критерием правильности действий врача является соответствие профессиональным стандартам, а не исход лечения.

*На уровне популяции, групп населения* исходы также неопределенны, хотя в исследовательских целях возможно выявление статистических



связей. Важно, что популяционные цели оказывают влияние на ведение отдельного пациента. Применительно к отдельному пациенту лечение может оказаться неполным, так как оно угрожает «благополучию группы населения», например, через большие расходы. Прирост совокупных показателей здоровья населения является показателем объективным. Одновременно этот показатель отражает интересы, которые могут противостоять интересам отдельного пациента.

Поскольку применительно к отдельному врачу и к отдельной больнице применение показателей исходов связано с большими трудностями, в основе исследований оказания медицинской помощи лежат показатели процесса. Примерами таких показателей (индикаторов) являются частота использования врачами кортикостероидов при преждевременных родах, частота назначения  $\beta$ -адреноблокаторов пациентам, перенесшим инфаркт миокарда, частота назначения пожизненной гипотензивной терапии при гипертонической болезни. Обычным является использование показателей процесса наряду с доступными показателями исходов. Например, при оценке успешности программы вакцинации такими показателями могут быть:

- ▶ доля популяции, подвергнутая вакцинации;
- ▶ доля групп высокого риска, подвергнутых вакцинации;
- ▶ доля вакцинированных, у которых наблюдался серологический ответ;
- ▶ доля заболевших в период после вакцинации;
- ▶ доля заболевших в период после вакцинации среди всех вакцинированных.

### 5.2.1. Качество медицинской помощи

Организаторы здравоохранения уделяют оценке качества медицинской помощи серьезное внимание. Качество — это совокупность свойств и характеристик товара или услуги, которые определяют ее способность удовлетворять установленным или предполагаемым требованиям. Высококачественная помощь — это помощь, которая максимизирует благо пациента, принимая во внимание баланс ожидаемых плюсов и минусов для процесса лечения на всех его этапах. Американская медицинская ассоциация в более понятных словах определяет доброкачественную помощь как повышающую или поддерживающую качество или продолжительность жизни; одновременно имеет значение, соответствуют ли действия врачей современным знаниям.



*Контроль качества* — это проверка соответствия индикаторов стандартам. *Индикаторами* называют избранные показатели, отражающие оказание медицинской помощи наилучшим образом. Индикаторы выбираются, поскольку не все показатели доступны, и не все доступные надежно отражают качество медицинской помощи. *Управление качеством* включает в себя оценку качества, контроль качества и обеспечение качества. *Обеспечить качество*, значит, создать необходимые условия для качественного оказания медицинской помощи. Сюда относится не только обеспечение необходимыми ресурсами, но и правильная организация работы.

На групповом и экологическом (территориальном) уровнях также используется подход к оценке качества работы медицинских и социальных служб, называемый *индикаторным* (англ. *health indicators*). При его реализации предполагают, что несколько избранных болезней (состояний) отражают общий уровень здоровья населения и оказания медицинской помощи. Такими показателями могут быть младенческая смертность, больничная летальность при инфаркте миокарда или другие показатели. Если сравнивать разные регионы страны, сопоставлять индикаторы и условия оказания помощи, то можно делать выводы о том, чем обусловлены эти изменения и различия.

Страховые компании, профессиональные организации врачей, органы государственного управления, лечебные учреждения различных форм собственности и частнопрактикующие врачи, пациенты и потребительские организации заинтересованы в высоком качестве медицинской помощи. Они используют разные влияния на качество медицинской помощи, в том числе создают стандарты оказания помощи в виде детальных протоколов, действующих в медицинской организации или сети медицинских организаций. Протоколы основываются на совокупности знаний об оптимальном оказании помощи, обобщенных в клинических рекомендациях (рекомендации). Рекомендации для клинической практики (англ. *clinical practice guidelines*) — это документы, содержащие рекомендации, нацеленные на оптимизацию ведения пациентов, основанные на систематическом обзоре научных доказательств и оценке пользы и вреда альтернативных способов ведения пациентов. Современные рекомендации для клинической практики не могут быть исчерпывающим образом основаны только на надежных доказательствах, полученных в контролируемых клинических испытаниях, но при должной процедуре разработки и от-



сутствии недопустимого влияния коммерческих интересов являются наилучшими документами, предписывающими оптимальное ведение пациента или профилактические вмешательства. Важнейшие положения рекомендаций для клинической практики обосновываются систематическими обзорами и комплексной оценкой медицинской технологии.

### 5.2.2. Изучение исходов

*Изучением исходов* (англ. *outcomes research*) называется разновидность эпидемиологических исследований в интересах оценки медицинских вмешательств, выполняемая с использованием существующих статистических демографических, экономических данных и сведений об исходах лечения и показателях здоровья людей. Особенностью таких исследований является использование для анализа преимущественно существующих вторичных документов, суммирующих сведения для различных целей. По мере развития системы компьютерной регистрации данных стало возможным создание баз данных о расходах, процессах и исходах заболеваний для работников отдельных предприятий, для застрахованных в государственной или в частной системе страхования, для получателей пособий и других больших категорий граждан.

С конца XX в. в развитых странах некоторые данные о работе больниц, в том числе данные об исходах основных вмешательств, открыто публикуются и доступны в интернете. Везде предпринимаются периодические оценки результатов применения дорогостоящих технологий, например искусственного оплодотворения. В результате выясняется, каков прогресс и насколько стабильны показатели эффективности. Например, в 2016 г. в Великобритании для женщин в возрасте до 38 лет успешными были 27% циклов экстракорпорального оплодотворения по сравнению с 22% в 1998 г. При этом, однако, показатели клиник широко различаются: от 10 до 46%. Это с очевидностью свидетельствует о проблемах в применении технологии.

Недостатком использования существующих данных является ограниченность исследования только имеющимися (зарегистрированными) переменными. Этот набор переменных соответствует тем задачам, для которых он создавался, например финансовым. Процедуры получения данных о включенных в базы данных гражданах могут отличаться



от принятых в эпидемиологических исследованиях, а классификации могут быть несопоставимыми. Даже собственно медицинские базы данных, используемые для обсервационных исследований, обычно не могут быть основаниями для надежных выводов. Например, одним из процессов, который лучше всего документируется, является течение беременности и родов. Более 100 лет принято считать, что беременные должны наблюдаться в течение беременности, и произвольно установлена периодичность их визитов к врачу. Используя всю совокупность данных о визитах беременных к врачу и исходах беременности, нельзя получить информацию о том, нужны ли эти визиты. Если своевременно «вставшие на учет» беременные завершают беременность с лучшими показателями, то это может быть лишь следствием их особенностей: известно, что женщины необразованные, живущие в плохих условиях, злоупотребляющие алкоголем и наркотиками позднее встают на учет. С другой стороны, у тех, кто учтен только в III триместре, может быть меньше абортс уже потому, что у таких женщин не может быть абортс ранее этого срока (он уже прошел).

Для оценки результатов применения некоторой медицинской технологии или способа организации медицинской помощи (способа оплаты и т.д.) наилучшей организацией исследования является рандомизированное контролируемое испытание (см. разд. 3.1). В случае когда РКИ не может быть проведено по каким-либо причинам, возможны несколько вариантов исследования, которые несовершенны, дают смещенные оценки, но при отсутствии возможности исследования более высокого качества могут быть приняты к рассмотрению.

### **Сравнение получающих и не получающих вмешательство**

По разным причинам всегда есть пациенты, которые получают некоторое вмешательство (коронаропластику при инфаркте миокарда) и не получающие его. Возможно сравнение исходов у таких пациентов. К сожалению, обнаружение различий не дает почти ничего для оценки полезности вмешательства. Пациенты, получающие вмешательство, могут отличаться вследствие самоотбора и вследствие отбора врачами для применения вмешательства. В результате различие в исходах может объясняться более тем, что это разные пациенты, чем тем, что их по-разному лечили. Выявив различия сравниваемых пациентов по известным признакам, например по возрасту, тяжести инфаркта, времени от инфаркта до госпитализации и до тромболизиса, можно



учесть эти различия статистическими средствами. К сожалению, поскольку группы формируются случайным путем, исследователь не может быть уверен, что он таким образом устранил влияние привходящих факторов.

Вариантом сравнения получающих и не получающих лечение является сравнение тех, кому оно недоступно, с теми, кто его получил. Такой вариант исследования возможен в странах, где доступ к медицинским вмешательствам регламентируется эксплицитными (открыто и ясно сформулированными) критериями. Например, если вмешательство входит в список оплачиваемых по одному страховому плану, но не оплачивается по другому, то можно сравнить исходы при одной болезни у пациентов, имеющих разные страховки. Этой структуре исследования свойственны те же недостатки — пациенты, имеющие разные страховки и проживающие на разных территориях, могут отличаться по многим признакам, и результаты такого исследования в лучшем случае могут служить для выдвижения гипотезы о преимуществах или недостатках данного способа оказания помощи.

### **Историческое сравнение**

Отдельные новации, особенно управленческие, вносятся в оказание медицинской помощи без достаточных научных обоснований. После введения в практику нового вмешательства, новой организации процесса оказания медицинской помощи, новой системы платежей сохраняется возможность сравнить новые результаты лечения с результатами, имевшими место до этого. Несмотря на принципиальную возможность такого исследования, реализовать его приемлемым образом крайне трудно (см. также с. 341). Прежде всего данные, собиравшиеся до внедрения вмешательства, отличаются от собираемых после. Например, данные, собираемые в целях исследования, собираются персоналом, который знает об исследовании, собираемые сведения более детальны, более стандартизованы, чем регистрировавшиеся до этого.

Даже если имеют место изменения, невозможно узнать, являются ли они результатом нововведений или они лишь сопутствуют им, являясь в действительности результатом действия привходящих факторов. Даже в рандомизированных испытаниях исторические изменения могут быть столь значительны, что скрывают эффект от испытываемого вмешательства. Так произошло в широкомасштабном испытании многофактор-



ной профилактики сердечно-сосудистых болезней MRFIT в США. Оно по времени совпало с быстрым снижением смертности от сердечно-сосудистых болезней. Это снижение было значительным не только в группе вмешательства, но и в группе контроля.

Явления селекции тоже непостоянны. Пациенты, которые посещают клинику после нововведений, могут быть иными пациентами, чем те, кто лечился в ней ранее. Это может быть следствием новой системы платежей, нового вида хирургических вмешательств и по иным причинам. Таким образом, исследования с историческим контролем могут быть использованы только как предварительные исследования, результат которых может дать или не дать основание для проведения рандомизированного контролируемого испытания.

Организационные новации и вмешательства, которые представляются неизбежными, например вакцинация новой вакциной в период эпидемии, могут быть оценены по эффективности. Для этого существует лестничный дизайн РКИ (англ. *stepped wedge design*). В нем единицами рандомизации являются больницы или регионы страны. Рандомизируется время введения новации. В итоге новация вводится во всех регионах, но в разное время. Поэтому есть возможность оценить изменения, наступившие в регионах, уже реализовавших новую систему, с изменениями в регионах, еще работающих по-старому. При отсутствии рандомизации сравнение регионов и больниц невозможно, поскольку новации могут вводиться в первую очередь в самых благополучных, или, наоборот, самые бедные регионы могут осваивать, например, особую форму кредитовыми помощи.

## Контрольные вопросы и задачи к разделам 5.1 и 5.2

1. При анализе принятия решений важными шагами не являются:
  - а) вычисление необходимого размера выборки;
  - б) построение дерева решений;
  - в) систематический поиск данных о вероятностях исходов, ценах и полезности исходов;
  - г) оценка статистической значимости различий в частоте исходов;
  - д) анализ чувствительности к смещениям.
2. Когда экспериментальное исследование с вмешательством не может быть использовано, для суждения о причинной связи события



с предшествующей экспозицией судят по данным обсервационных исследований. Из этих данных важны:

- а) статистическая сила связи;
  - б) повторное наблюдение связи в независимых исследованиях;
  - в) наличие зависимости частоты эффекта от дозы экспозиции;
  - г) наличие объяснений биологического механизма изучаемого эффекта.
3. В национальный гематологический центр в 1998–2001 гг. поступили больные острым лейкозом в количестве соответственно: 120, 136, 154, 183. Эти данные говорят:
- а) об изменении порядка госпитализации;
  - б) увеличении преваленса лейкозов в стране;
  - в) увеличении инциденса лейкозов у детей;
  - г) улучшении финансирования центра;
  - д) ни о чем определенно.

Выберите все правильные ответы.

4. При обобщении данных когортных исследований влияния фактора производственной среды (электромагнитного излучения) для проведения метаанализа необходимо, чтобы:
- а) все исследованные группы в каждом исследовании имели равную численность;
  - б) интенсивности воздействия были близкими;
  - в) сравниваемые группы в разных исследованиях были схожи по приводящим признакам;
  - г) в начале исследования сравниваемые группы были схожи по оцениваемому признаку (исходу).

Выберите один правильный ответ.

5. По обобщенной оценке 80% случаев рака легких возникают вследствие курения. Такое представление данных является:
- а) инцидентом;
  - б) дополнительным риском;
  - в) относительным риском;
  - г) преваленсом;
  - д) пропорциональным отношением смертности.

Выберите все правильные ответы.

6. Показателями процесса оказания помощи являются:
- а) доля пациентов, у которых измерили артериальное давление;



- б) доля пациентов с инфарктом миокарда, которым при выписке назначили ацетилсалициловую кислоту;
- в) доля пациентов, госпитализированных повторно в течение месяца;
- г) доля пациентов, у которых измерили массу тела.

### **5.3. ОБЕСПЕЧЕНИЕ ГОТОВНОСТИ ОБЩЕСТВА К ПРИРОДНЫМ И СОЦИАЛЬНЫМ КАТАСТРОФАМ**

Не только собственно эпидемии инфекционных болезней, но и любые другие массовые страдания, связанные с природными и социальными катастрофами, являются объектом эпидемиологии. В законодательстве развитых стран существует развитая основа для реализации мер защиты населения от инфекций и многих других неблагоприятных факторов, способных негативно повлиять на общественное здоровье. Естественным образом эти подготовительные меры опираются на имеющиеся знания о прошлых эпидемиях, катастрофах. В период холодной войны большое внимание уделялось средствам защиты от последствий применения химического, ядерного или биологического оружия в военном конфликте. На фоне эпидемиологического перехода основное внимание стало уделяться неинфекционным болезням, а разрядка международной напряженности и затем исчезновение коммунистического лагеря резко снизили опасность большого военного противостояния. В большинстве развитых стран стали снижаться расходы не только на оборону, но и на поддержание системы готовности защиты населения.

Аналогичные процессы протекали и в системах здравоохранения. В стремлении удержать рост затрат на здравоохранение все страны сокращали количество коек в больницах и интенсифицировали процессы оказания помощи. Системы здравоохранения развитых стран, что лучше всего известно на примере Великобритании, оказались лишенными резервных мощностей в такой степени, что обычное повышение заболеваемости зимой стало приводить к коллапсу больничной помощи.

Два события подтолкнули человечество к развертыванию программ готовности к эпидемиям: короткая эпидемия коронавируса SARS-CoV-1, вызывавшего острый тяжелый респираторный синдром (2003), и заболевания «птичьим гриппом» H5N1 (2005). Всемирная организация здравоохранения призвала страны-участницы разрабатывать про-



граммы подготовки национальных систем здравоохранения, экономик и инфраструктуры к возможной пандемии новой инфекции, передаваемой респираторно. Акцент делался на модели тяжелого гриппа, который в связи с высокой контагиозностью способен в течение короткого срока, нескольких дней, поразить все континенты<sup>1</sup>. Большие компании начали разрабатывать планы по сохранению работников и производственных процессов в период эпидемии. Для систем здравоохранения разрабатывались алгоритмы сохранения работников. Например, одна из схем предусматривала изоляцию части персонала в начальный период. Другая часть персонала начинает работать и через некоторое время заболевает и становится неспособной оказывать помощь. Их сменяет находившаяся в изоляции часть коллектива. Поскольку в эти годы единственным практически применимым препаратом, эффективным против гриппа, считался Тамифлю, многие страны сделали стратегические его запасы. В целях длительного хранения закладывалась чистая сухая субстанция. Предполагалось, что при возникновении эпидемии будут изготавливаться растворы, и они будут раздаваться гражданам, изолированным в своих домах, квартирах. Для бедных стран Юго-восточной Азии, где наиболее вероятно возникновение нового варианта гриппа, ВОЗ создала местные запасы Тамифлю. В некоторых странах были увеличены мощности инфекционных стационаров, имеющих койки в палатах с отрицательным атмосферным давлением, финансировались программы изготовления стратегического запаса аппаратов для искусственной вентиляции легких.

В дальнейшем ВОЗ расширила подход к повышению готовности к катастрофам<sup>2</sup>. В нем ключевое положение занимает стратегическое планирование, в том числе поддержание системы здравоохранения в состоянии готовности к быстрому и эффективному ответу. Пандемия COVID-19 показала, что акцент на подготовке к пандемии гриппа ограничен. Складирование противовирусных препаратов имеет ограниченную ценность не только потому, что оселтамивир малоэффективен, но и потому, что новая пандемия может быть вызвана не вирусом гриппа. Необходим акцент на подготовке к реализации именно неспецифических эпидемиологических мер по сдерживанию распространения

<sup>1</sup> Influenza. Pandemic preparedness. <https://www.who.int/influenza/preparedness/pandemic/en/>.

<sup>2</sup> A strategic framework for emergency preparedness. Geneva, World health organization. 2017. 17 p.



инфекции, отслеживанию контактных лиц с помощью новых информационных технологий, сохранению жизнеспособности экономики в целом и работающей системы здравоохранения в частности.

## 5.4. СКРИНИНГ

Вмешательства в интересах здоровья популяции могут быть безадресными (фторирование водопроводной воды) или направленными на отдельные группы людей (беременных, бездомных, школьников). Несмотря на направленность отдельных вмешательств на специфические группы (бездомные, наркоманы, сироты), реализация таких вмешательств обычно приводит к благу для всей популяции. Ввиду огромной потенциальной эффективности и общественной значимости ряд социальных и медицинских вмешательств регламентируется национальными законами, например в России Законом об иммунопрофилактике, Законом о санитарно-эпидемиологическом благополучии населения.

Выделение популяций с высоким риском заболеваний, выяснение причин высокого риска и анализ затрат и последствий проведения профилактических программ базируется на эпидемиологических методах. Среди медицинских программ, ориентированных на популяцию, наиболее важными являются программы скрининга, так как они включают большие массы людей, основаны на применении медицинских вмешательств и нацелены на снижение распространенности болезней.

*Скринингом* называют обследование людей, не считающих себя больными и, соответственно, не обращающихся к врачам с целью выявления среди них больных для проведения раннего лечения или для иного предотвращения нежелательных событий. Например, у супругов, гетерозиготных по гену талассемии (тяжелой болезни, связанной с дефектом гемоглобина), возможно выявить зачатие ребенка гомозиготного, т.е. обреченного на это состояние, лечение которого в большинстве стран недоступно. При проведении такой программы можно резко сократить рождение больных детей (рис. 5.9).

Для скрининга обычно применяют специальные тесты — недорогие, нетравматичные и не обязательно с высокими чувствительностью и специфичностью. Предполагается, что при скрининге выявляются люди с повышенной вероятностью выявляемого состояния, которые могут быть затем глубже обследованы.



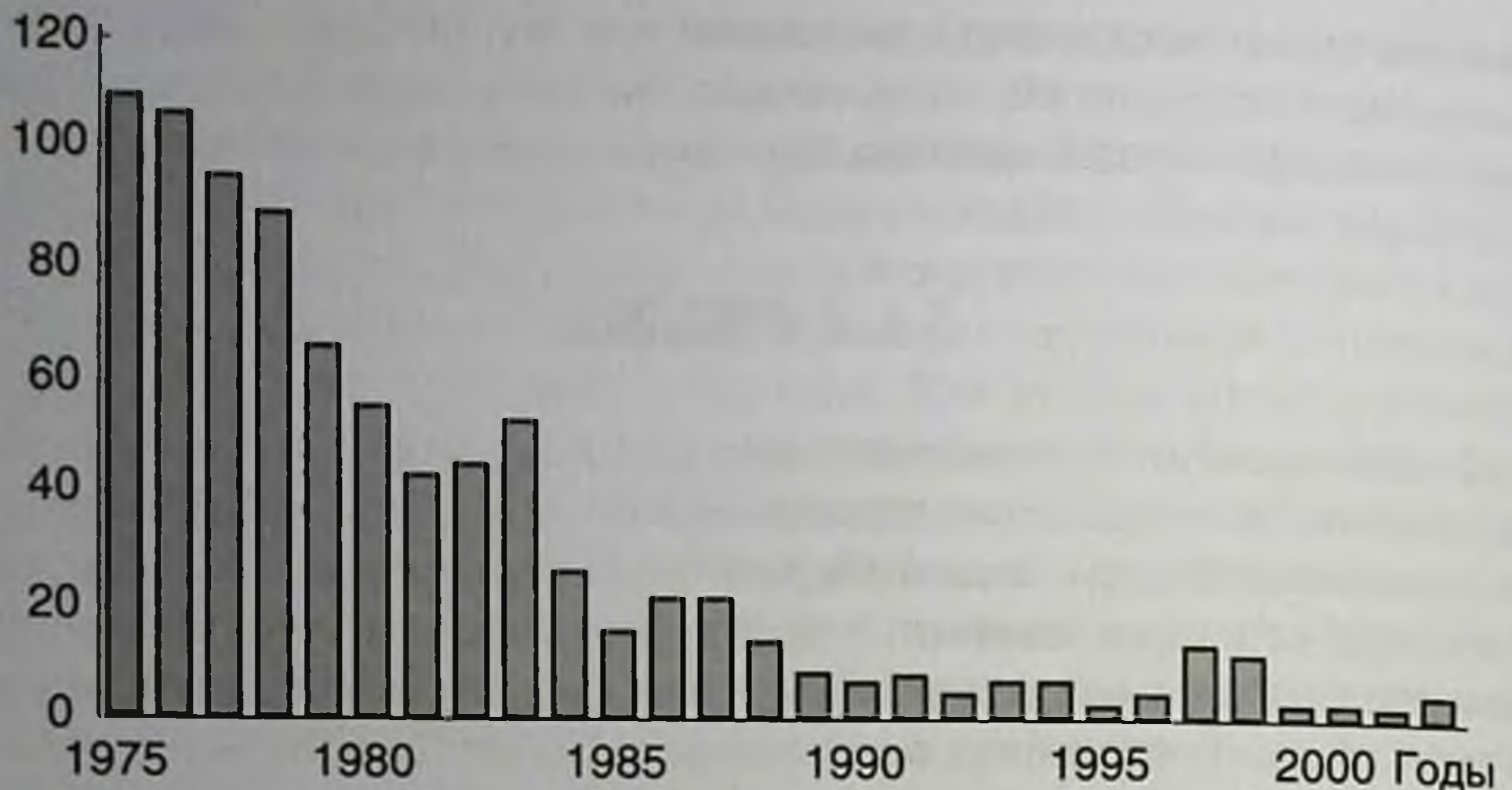


Рис. 5.9. Снижение числа рождений детей, гомозиготных по  $\beta$ -талассемии, в год в Сардинии после введения скрининга в 1975 г. (по Cao A., Galanello R. *New Engl. J. Med.* 2002. № 347. P. 1200–1202)

Скрининг на хронические заболевания основывается на двух положениях. Первое: некоторые люди обращаются за помощью поздно, т.е. тогда, когда болезнь имеющимися методами уже неизлечима. Прежде всего это относится к раку. Опухоли малого размера чаще поддаются радикальной резекции. Вместе с тем пока опухоль не становится достаточно большой, человек не обнаруживает ее сам. Поэтому существует стремление выявить опухоль тогда, когда она еще не выявляется самим больным, т.е. на доклинической стадии. Второе положение кажется очевидным, но пока не было убедительно подтверждено в экспериментах. Это предположение о том, что если выявлять болезни раньше, до того как они станут клинически явными, то их можно более эффективно лечить. Например, применительно к раку молочной железы предполагается, что можно чаще выполнять экономные, косметически приемлемые и одновременно *излечивающие* резекции, без применения травматичных видов лечения — облучения и химиотерапии. В обычной практике большая часть операций являются уродующими и незначительно продлевают жизнь, несмотря на применение токсичной химиотерапии и рентгенотерапии. Не случайно перспектива скрининга описывается обычно в «хирургическом» аспекте. Сама идея скрининга связана с представлениями об отдельности болезни и человека, возможности «изъятия» болезни из организма, излечения.



Вариантом скрининга являются программы периодических профилактических обследований. Чаще всего речь идет о ежегодных обследованиях. Обычно периодические профилактические обследования не предназначены только для выявления скрытых состояний (скрининг), но нацелены также на оценку течения и правильности лечения болезней. Такого рода периодические обследования не имеют надлежащего теоретического и экспериментального обоснования, но практикуются в США и в некоторых странах Европы. В России они называются *диспансеризация*. Основа диспансеризации — периодический врачебный профилактический осмотр — не приносит пользы по показателям здоровья<sup>1</sup>. Анализ существующих доказательных исследований позволяет рекомендовать проведение профилактических вмешательств (рекомендация прекратить курение, измерение артериального давления, напоминание о ревакцинации, цервикальный мазок и др.) в ходе встречи врача и больного в связи с острым или хроническим заболеванием, для этого нет необходимости организовывать специальные посещения [17]. Эффективность комплексных профилактических программ трудно исследовать, и, вероятно, полных данных об их эффективности никогда не будет получено. Отдельные программы скрининга исследованы хорошо и имеют известную эффективность и безопасность.

### 5.4.1. Теория скрининга

В теории скрининга используется представление о критических точках в развитии болезни. Например, такой точкой является момент появления региональных метастазов опухоли. Иными словами, критическая точка разделяет периоды развития болезни, различающиеся по перспективе использования лечебных вмешательств. Невозможно определить, пройдена ли такая точка в развитии болезни отдельного пациента, но предполагается, что скрининг дает возможность вмешаться ранее, чем развитие болезни пройдет такую критическую точку у большого числа пациентов.

Еще одно важное теоретическое предположение скрининга — расчет на то, что болезнь, единожды начавшись, развивается в клиническую форму, проходя через выявляемую скринингом скрытую (локальную) стадию. Применительно к раку — это предположение о том, что мест-

<sup>1</sup> Krogsboll L.T. et al. (2019). General health checks in adults for reducing morbidity and mortality from disease. Cochrane Database Syst Rev 1: CD009009.



ный рак является предшественником метастазирующего и что его можно выявить. В действительности при многих болезнях и, в частности, применительно ко многим опухолям показано, что часть случаев локализованного рака претерпевает обратное развитие (рис. 5.10). Естественно, что выявление таких случаев и лечение приносят только вред — неизбежный вред противораковой терапии и хирургических вмешательств. Применительно к некоторым опухолям уже известно, что они неоднородны по происхождению, например отдельные формы их связаны с разными вирусами, а потому текут несколько различно, но эти знания еще не дали врачам инструмента для различения опухолей. Для некоторых опухолей показано, что вероятность их метастазирования не зависит от размера в момент диагностики.

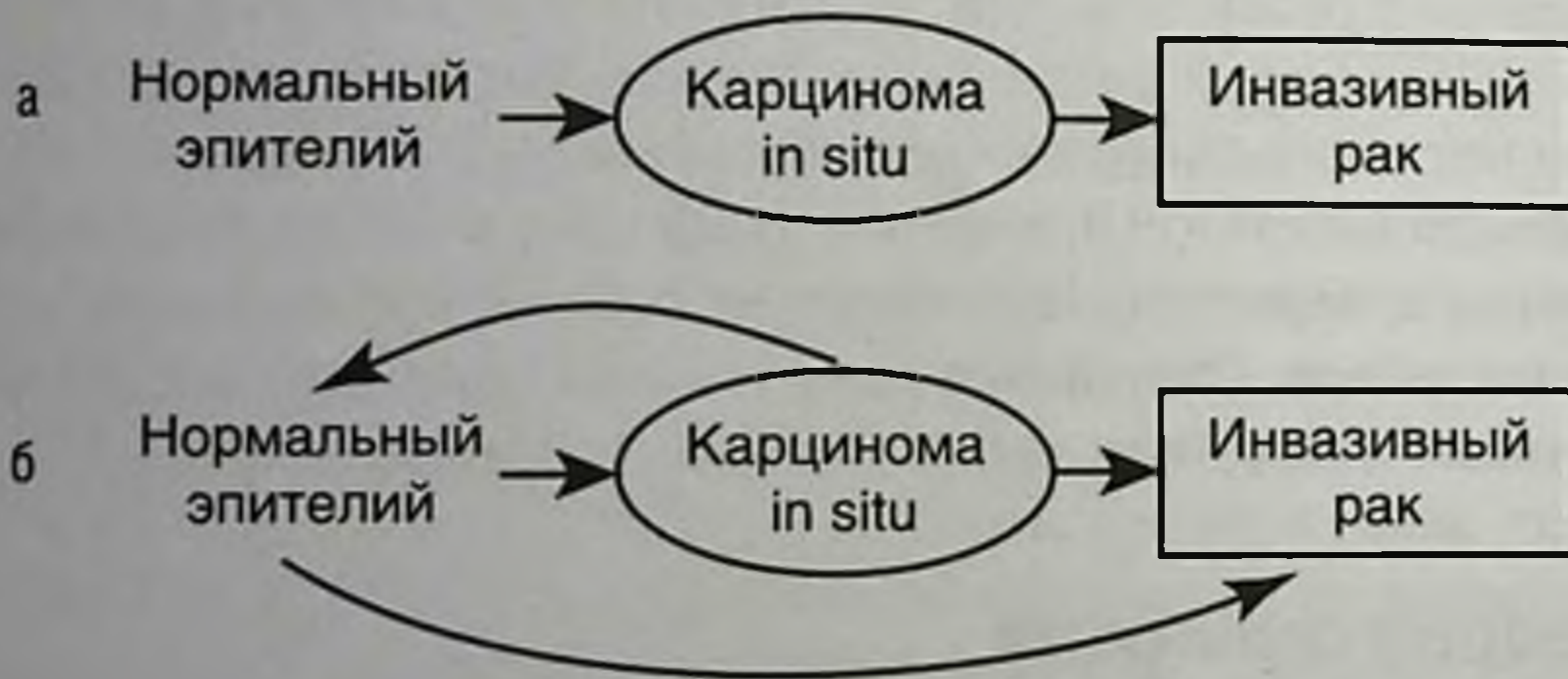


Рис. 5.10. Естественное развитие рака: а — модель линейного развития; б — возможные варианты развития

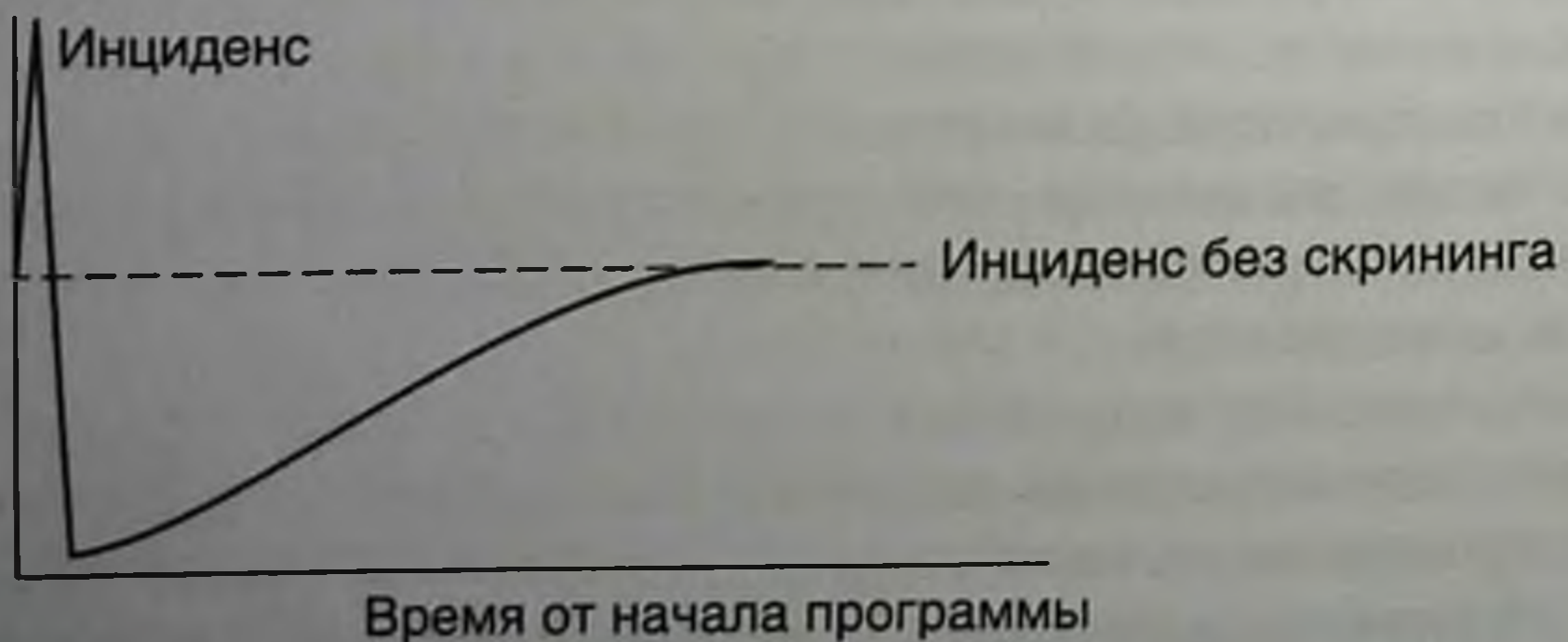


Рис. 5.11. Изменения в инциденте болезни в результате однократного скрининга





Рис. 5.12. Кумулятивный инцидент болезни при отсутствии скрининга (пунктир) и при проведении однократного обследования

С началом скрининга должно резко возрасть число выявляемых случаев (инцидент), преимущественно за счет ранних случаев болезни у людей, не обследовавшихся до этого. Если повторных обследований не будет, то в результате того, что скрытые случаи были обнаружены, вслед за однократным скринингом в течение некоторого времени инцидент должен быть сниженным (рис. 5.11). Однократное обследование или первое обследование в долгосрочной программе скрининга выявляет относительно большое число новых случаев, соответствующее преваленсу в популяции той скрытой формы болезни, которая выявляется принятым в скрининге методом. Если больше скрининг не проводить, то в течение некоторого времени новых случаев *обычными методами* обнаруживается мало. При проведении повторных обследований инцидент должен оставаться относительно повышенным, поскольку ранняя диагностика всегда сопряжена с относительно высокой частотой ложноположительных результатов и выявлением случаев, которые так никогда и не стали бы клинически значимыми (рис. 5.12). Это постоянное повышение инцидента при систематическом проведении скрининга по сравнению с оказанием помощи без скрининга составляет 10–30%.

При проведении скрининга возможно увеличение смертности в результате побочных эффектов массовых вмешательств у большого числа новых пациентов. Вслед за тем должно наступить снижение смертности,





**Рис. 5.13.** Предполагаемые изменения смертности после однократного скрининга в популяции. Сверху — отложенная смертность, снизу — устраненная (излечение)

если выявляемые заболевания более эффективно лечатся. В результате скрининга может возникать уменьшение смертности в виде устранения смертности от болезни у части случаев (излечение) и в виде «откладывания» смертности (замедление развития болезни) (рис. 5.13). Оба варианта ценны для участников программы, но при продолжающемся скрининге его результаты могут выглядеть как незначительные, если излечение у выявленных случаев не наступает.



### 5.4.2. Методологические проблемы скрининга

Совокупность идей, связанных со скринингом, столь привлекательна, что в мире реализуется множество программ скрининга. Например, почти повсеместно проводится скрининг на рак шейки матки, во многих экономически благополучных странах реализуются программы скрининга на рак молочной железы. Между тем исчерпывающих оснований для такой деятельности нет. Почему нельзя просто по результатам скрининга понять, полезен он или нет?

Как только идея скрининга формулируется и скрининг вводится в ограниченной части популяции, каждый новый случай выявления скрытой болезни служит на первый взгляд доказательством полезности скрининга. Пропорционально числу обнаруженных случаев растет число вмешательств (рис. 5.14). Этим как бы подтверждается полезность скрининга. Однако само число выявленных случаев в популяции прямо зависит от объемов скрининга, и число консервативных резекций увеличивается более значительно, чем снижается число мастэктомий и случаев поздней диагностики (рис. 5.15).

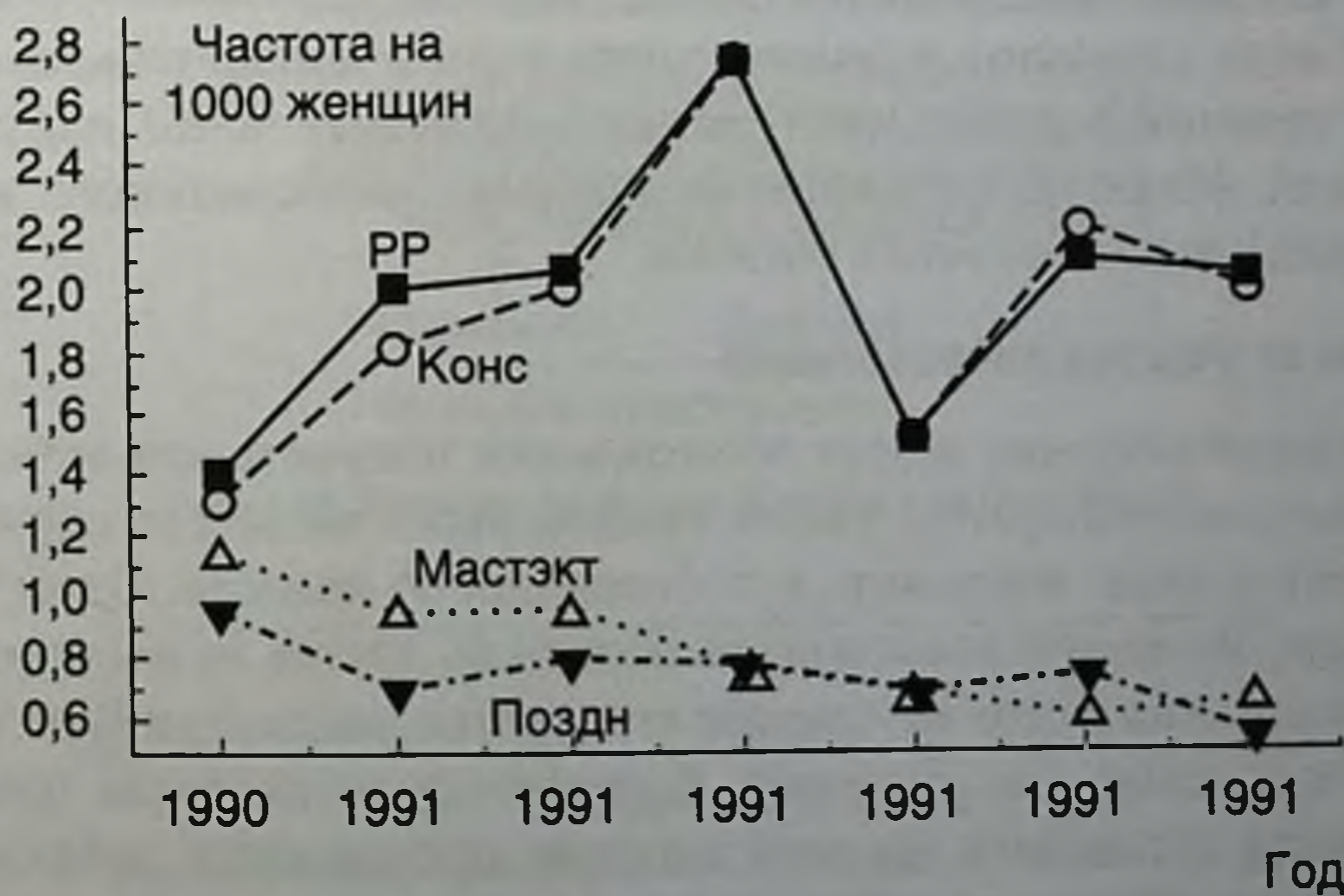


Рис. 5.14. Изменения инциденса и характера оперативных вмешательств после начала скрининга в популяции. Программа скрининга на рак молочной железы во Флоренции. PP — ранние формы рака (<2 см и карцинома *in situ*); Конс — консервативные резекции; Мастэкт — мастэктомии; Поздн — поздние случаи рака (T2, по классификации метастазирования опухолей) (по данным Paci E. et al. *BMJ*. 2002. № 325. P. 418)



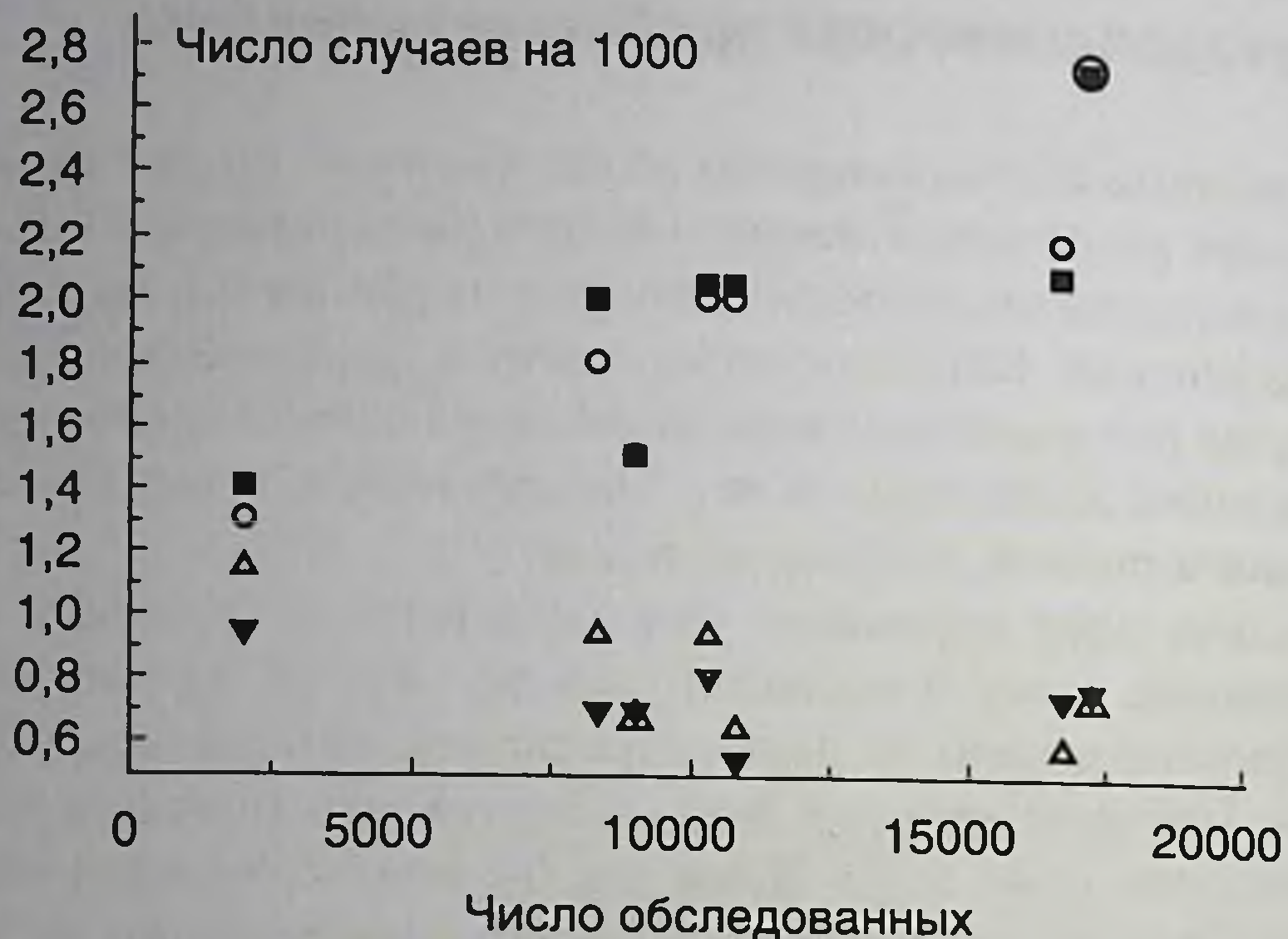


Рис. 5.15. Результаты (число выявленных случаев и число операций) в зависимости от объема скрининга. Программа скрининга на рак молочной железы во Флоренции. Обозначения как на рис. 5.14 (по данным Paci E. et al. BMJ. 2002. № 325. P. 418)

Даже если скрининг и последующее раннее вмешательство не изменяют течения болезни, они создают иллюзию более длительного выживания. Иллюзия возникает вследствие согласованного влияния нескольких систематических ошибок.

### Смещение от участия добровольцев

При предоставлении людям возможности получить что-либо, всегда часть людей не пользуются своим правом. Даже не всегда обращаются за деньгами в виде, например, освобождения от налогов. Одни не считают предоставленную возможность полезной, другие не воспринимают приглашения, как будто не слышат его, третьи рассматривают предложение как опасное, а не полезное. В случае предоставления гражданам возможности скрининга на приглашение откликаются добровольцы, которые более образованы, более последовательны в выполнении врачебных рекомендаций, имеют выше доход, реже курят и злоупотребляют алкоголем, т.е. отличаются рядом признаков, сопутствующих здоровью и более быстрому выздоровлению в случае болезни. Соответственно, если сравнить результаты лечения у больных, выявленных среди добровольцев, они окажутся лучше, чем у прочих больных, обратившихся



в клинической стадии болезни. Поэтому для испытания эффективности программы скрининга необходимо проведение рандомизированных контролируемых испытаний.

### Смещение от раннего диагноза

Скрининг осуществляется в надежде на то, что по сравнению с диагностикой болезни в симптоматический период, можно раньше выявлять болезнь за счет обнаружения ранних симптомов симптоматического периода, еще не приведших человека к врачу, и за счет скрытых симптомов латентного периода, которые могут быть обнаружены инструментальными средствами (рис. 5.16). Диагноз при скрининге ставится раньше на интервал, называемый *интервалом опережения* (англ. *lead time*).

Рассмотрим пример, приведенный на рис. 5.17. Представим болезнь, возникшую в 1990 г., приведшую пациента к врачу в 1997 г. и закончившуюся смертью в 2002 г. Зарегистрированная продолжительность

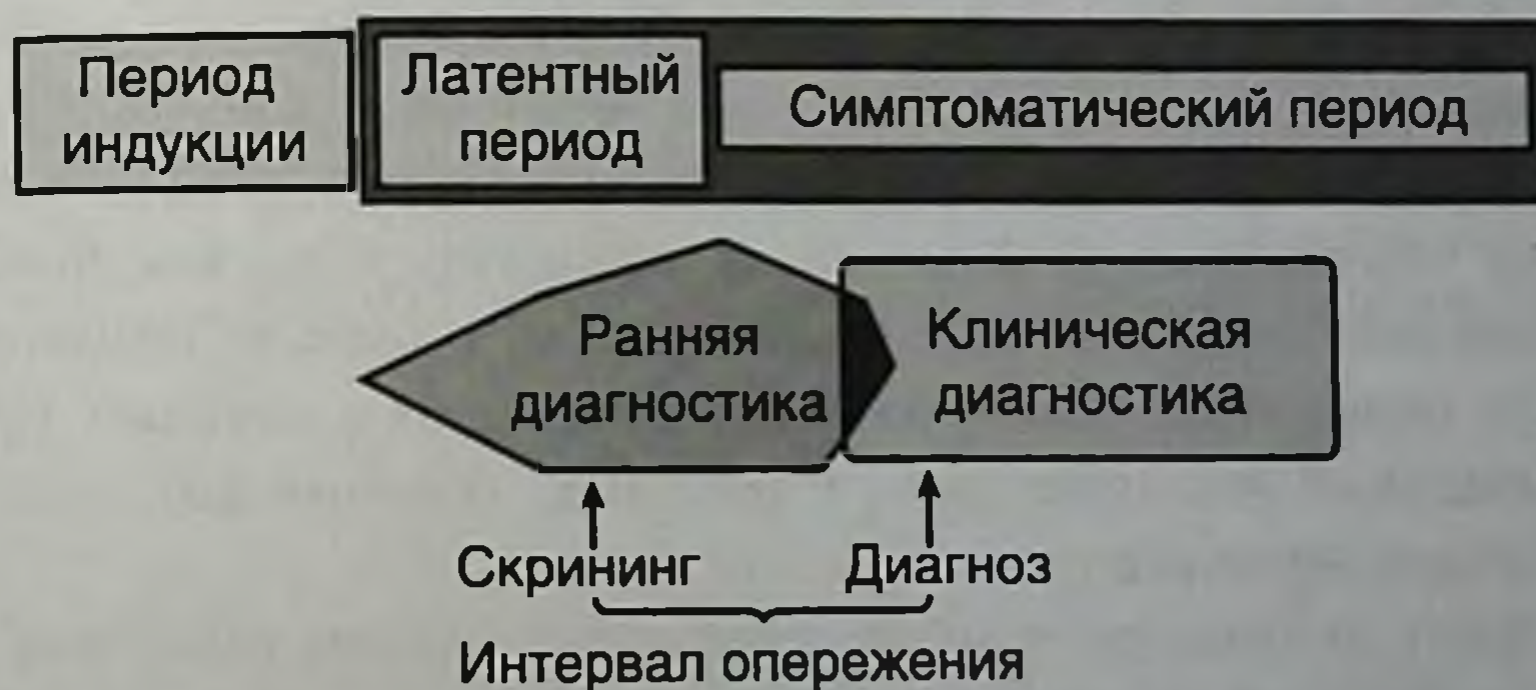


Рис. 5.16. Периоды развития болезни и ее выявления

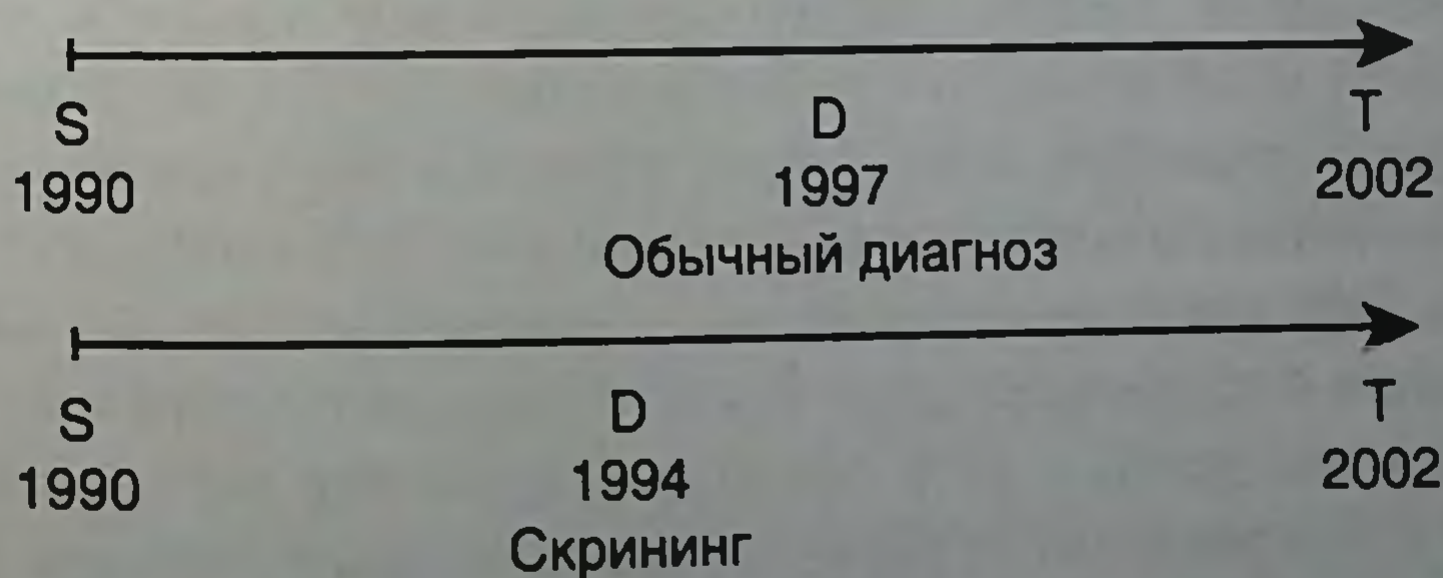


Рис. 5.17. Время выживания при естественном течении болезни и при раннем ее выявлении в процессе скрининга. S — момент возникновения болезни, D — момент выявления болезни, T — смертельный исход



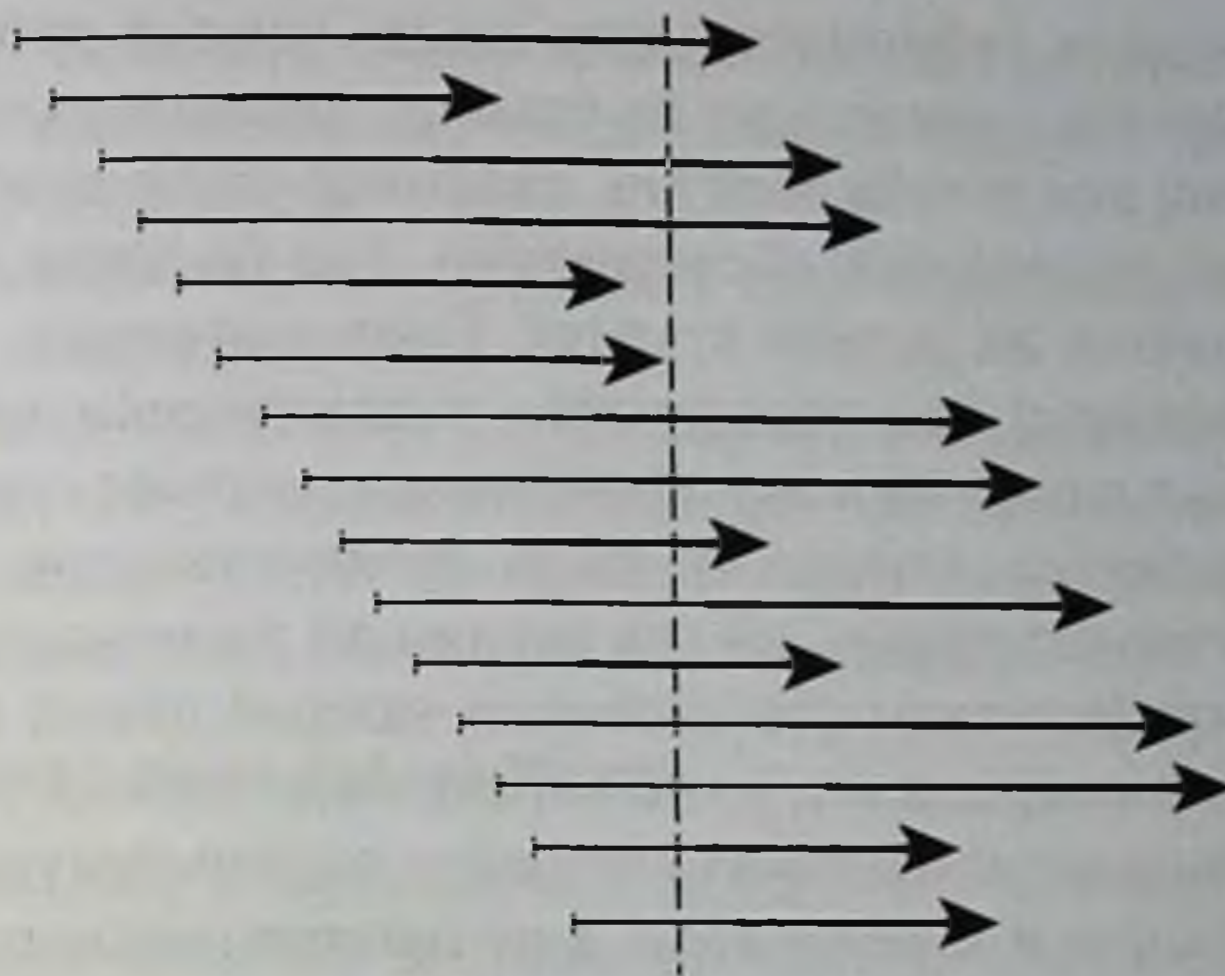
жизни больного с момента выявления болезни составляет 5 лет. С началом программы скрининга болезнь могла быть обнаружена у такого больного раньше, уже в 1994 г. Представим, что лечение не изменило продолжительности болезни, и смерть наступила в 2002 г. Зарегистрированная продолжительность жизни больного с момента выявления болезни составляет 8 лет. При неизменной продолжительности жизни этот пациент предстает для врачей как больной, получивший пользу в виде дополнительных 3 лет жизни. В действительности такой человек *потерял 3 года здоровой жизни* при неизменной общей продолжительности жизни. Поэтому увеличение продолжительности жизни после раннего диагноза не является доказательством полезности раннего диагноза.

### **Смещение от выявления медленно развивающихся случаев**

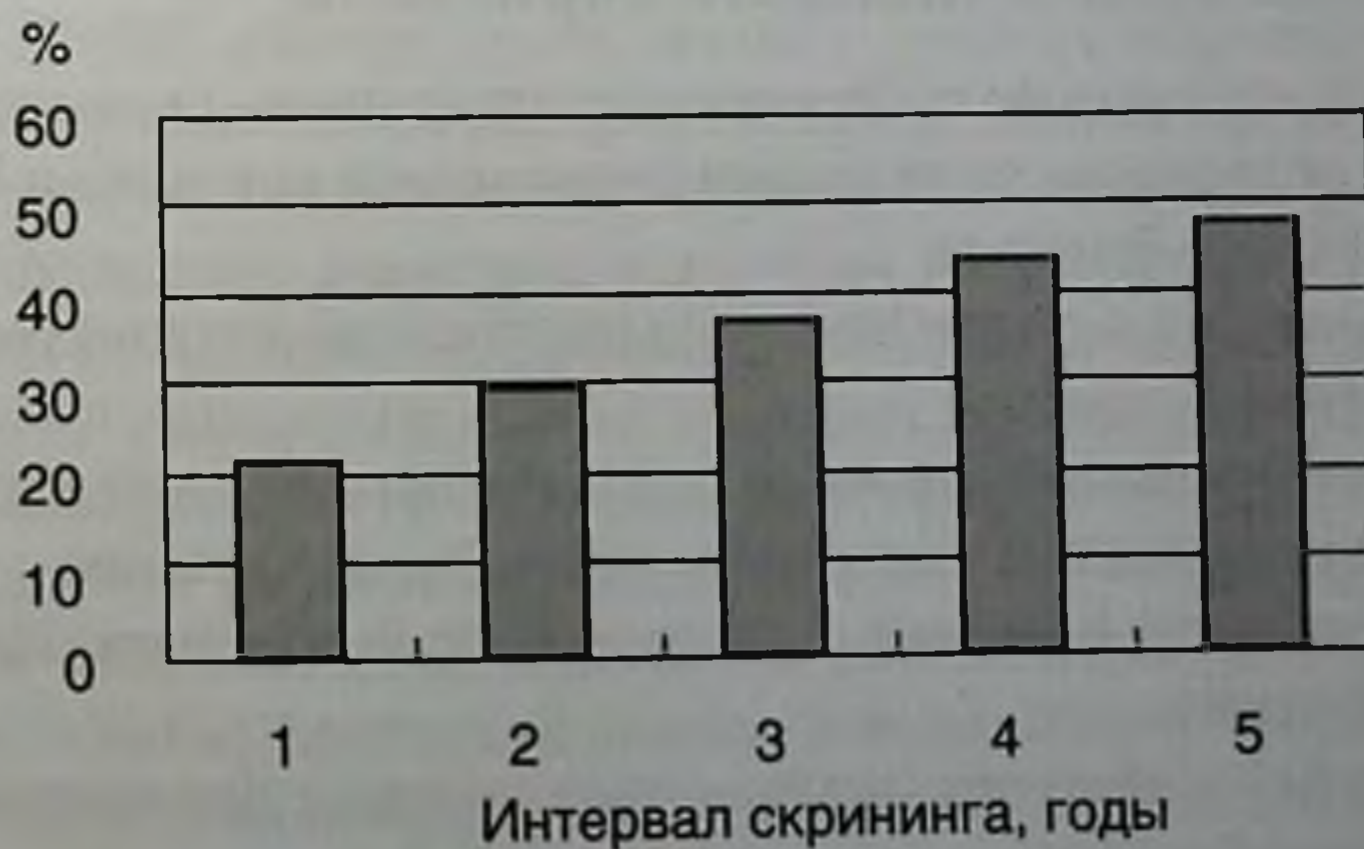
При скрининге более вероятно выявление медленно развивающихся случаев болезни (рис. 5.18). Применительно к раку известно, что скорость, с которой растет опухоль, сильно различается у разных людей. Это справедливо и для других болезней. Чем медленнее развивается заболевание на начальной стадии, тем медленнее оно прогрессирует и на заключительной стадии. Быстро развивающиеся случаи рано становятся клинически заметными, а медленно развивающиеся долго не проявляют себя и имеют больше шансов быть обнаруженными в процессе скрининга. Такие случаи, естественно, создают иллюзию более длительной выживаемости у больных, лечение которых начато рано благодаря скринингу.

Более того, возможно, что при скрининге обнаруживаются случаи болезни, которые, будучи предоставленными естественному течению, закончились бы самоизлечением или никогда не развились бы до клинически заметной стадии. Например, применительно к скринингу на рак шейки матки с использованием мазков по Папаниколау показано, что из числа женщин, отказавшихся после обнаружения метаплазии от вмешательства, позднее при обследовании у 36% получены нормальные мазки. При гистологическом исследовании молочных желез женщин, умерших в возрасте старше 40 лет от других причин и не знавших о наличии у них рака, у 40% обнаруживается рак молочной железы. Ничего не известно о естественном развитии таких случаев. Поскольку преваленс столь велик, можно думать, что значительная доля таких случаев не развиваются до клинически явной стадии в пределах жизни человека.





**Рис. 5.18.** Зависимость выявляемости случаев болезни от особенностей их течения. Отдельные случаи болезни ранжированы по моменту их возникновения. Представлены два варианта течения — быстрый и медленный (длинные стрелки). Период возможного выявления выделен более толстой линией. Более короткие линии — быстро текущие случаи болезни — имеют меньше шансов быть пересеченными вертикальным пунктиром, т.е. быть обнаруженными при скрининге. В данном примере выявлено лишь 2 быстротекущих случая из 7. Остальные или уже завершились, или еще не диагностируемы. Медленно текущих случаев выявлено 5 из 8



**Рис. 5.19.** Увеличение доли случаев рака молочной железы, выявляемых в интервале между обследованиями, по мере урежения обследований. Включены только женщины, выполнявшие все условия программы скрининга (по данным Day N.E. In: Screening for Breast Cancer. UICC. Hans Huber Publ. 1986. P. 108



В ходе программ скрининга почти всегда удается добиться увеличения доли случаев, выявляемых на стадиях, которые считаются более благоприятными для исхода болезни, например опухоли меньшего размера. Чем чаще проводятся обследования, тем большая доля случаев болезни выявляется на ранних стадиях. Соответственно, меньше случаев обнаруживаются «по обращению» в клинически явной стадии. Вместе с тем ни в одной из испытывавшихся программ скрининга доля таких «пропущенных» случаев не была незначительной. В наиболее изученном случае скрининга на рак молочной железы даже при ежегодном маммографическом обследовании каждый пятый случай обнаруживается не при скрининге, а «по обращению» (рис. 5.19). Наиболее вероятно, что большинство таких случаев — случаи быстрого развития болезни, возникшие в течение года, а не пропущенные при скрининге. Неудивительно поэтому, что такие случаи хуже поддаются лечению, чем найденные при скрининге. Летальность в выявленных при скрининге случаях 13%, а в выявленных в интервалах между обследованиями — 38%. Единственным способом правильно измерить полезность скрининга являются проспективные контролируемые испытания с параллельными группами, в которых измеряется не летальность больных, а смертность в популяции.

### 5.4.3. Оценка эффективности скрининга

Решение о необходимости проведения программы скрининга, ее итоговая оценка складываются на основе соответствия критериям Вильсона—Джаннера (Wilson—Jungner):

1. Состояние, на которое нацелена программа, должно быть важной проблемой здоровья.
2. Развитие болезни (состояния) должно быть хорошо изученным.
3. У состояния должна быть выявляемая ранняя стадия.
4. Вмешательство в ранней стадии должно быть более эффективно, чем в поздних.
5. Должен быть доступен диагностический тест для выявления ранней стадии болезни.
6. Тест для выявления ранней стадии должен быть приемлемым.
7. Должны быть определены интервалы между повторными обследованиями.



8. Должны быть обеспечены дополнительные диагностические и лечебные вмешательства, потребность в которых возникает в результате скрининга.
9. Физический и психологический вред должны быть меньше, чем польза от программы.
10. Стоимость программы в отношении к ее полезности должна быть не выше, чем у прочих медицинских вмешательств (программ).

Несоответствие хотя бы одному критерию делает внедрение программы скрининга сомнительным предприятием.

### **Диагностические тесты для скрининга**

На первом месте в скрининге стоит обследование людей, считающих себя здоровыми. Соответственно его можно провести, если:

- ▶ у болезни (состояния) есть скрытая стадия;
- ▶ имеются соответствующие диагностические тесты;
- ▶ решена проблема ложноположительных результатов.

Скрытая стадия существует у многих болезней. У хронических болезней она, вероятно, существует, но обычно неизвестно, как ее выявлять и использовать. Например, при гипертонической болезни диагноз ставится с использованием диагностического критерия — наличия артериального давления выше 140/80 мм рт.ст. Очевидно, что это — условная граница. Казалось бы, можно при меньших величинах давления (130/80) обнаруживать людей с «ранней гипертензией». Собственно такие попытки предпринимаются, например рекомендации Американской медицинской ассоциации в 2003 г. относят к повышенному артериальному давлению величины выше 120/80. Совокупность существующих данных свидетельствует, что установление диагноза «артериальная гипертензия» при меньших величинах приводит к распространению лечения на значительную часть популяции и ухудшению баланса пользы и вреда от вмешательств.

Главная теоретическая проблема диагностического компонента скрининга — выбор диагностического метода. В простейшем случае для ранней диагностики используется тот же метод, что и для клинической диагностики. Это сомнительная практика, так как хорошо известно, что в процессе развития заболевания его симптомы изменяются. Поэтому использование тестов, эффективность которых показана в клинической практике, для скрининга основывается на предположении, что на доклинической стадии симптомы те же, только менее выражены.



Даже если это предположение справедливо, то меньшая выраженность симптома соответствует меньшей различимости «болезни» и «здоровья». Так, уменьшение размеров тени от опухоли на рентгенограмме приводит к затруднению ее обнаружения на фоне нормальных теней. Основным результатом применения менее эффективного диагностического теста является увеличение числа ложноположительных результатов. Например, патологи, оценивающие мазки по Папаниколау, обязательно выносят ложноположительные решения, и число таких заключений тем больше, чем больше энтузиазм в программе скрининга. В результате группа больных с истинными злокачественными изменениями «разбавляется» ложноположительными случаями. Естественно, показатели летальности в такой смешанной группе будут лучше, чем в группе, где диагноз ставился на основании жалоб, клинического обследования и гистологического исследования биоптата.

Вторая теоретическая проблема диагностического компонента скрининга — «исчезновение нормы» при скрининге по множеству признаков. Если применять в лабораторном скрининге референтные интервалы для здоровых людей, включающие 95% величин здоровых людей, то, по определению, не менее 5% здоровых людей в популяции будут выявлены как имеющие «отклонения». Если тестов будет 2, то не имеющими отклонений будут только  $0,95 \times 0,95 = 0,9$  популяции. При дальнейшем увеличении числа *независимых* тестов число лиц, не имеющих отклонений, должно быстро уменьшаться. При 10 тестах — 0,6; при 20 — 0,36. В этом смысле здоровых людей нет, а есть только недообследованные.

Для того чтобы соответствовать целям скрининга, диагностические тесты должны быть эффективны. Это требование очевидно и рассматривалось применительно к клинической диагностике (разд. 2.3). Поскольку скрининг почти всегда направлен на обнаружение редких и мало выраженных состояний, важнейшая проблема — проблема ложноположительных результатов. Их очень много во всех испытывавшихся программах скрининга. Например, при маммографическом скрининге на рак молочной железы за 10 лет участия в программе женщина, у которой не выявлено рака, с вероятностью 0,3 будет после получения положительного результата подвергнута дополнительным обследованиям и с вероятностью 0,2 только после биопсии удастся исключить рак.

В табл. 5.6 приведен пример применения одного теста в диагностике и в скрининге. В клинической практике 10 ложноположительных ре-



зультатов на 2 истинно положительных — не редкость и не трагедия. Больные, пришедшие за помощью, с пониманием относятся к назначению новых исследований, все более травматичных, в надежде на уточнение диагноза и лечение болезни, которая их беспокоит. В скрининге ситуация отличается радикально. На 900 выявленных больных (число больных  $\times 0,9$ ) почти 50 тыс. должны пройти дополнительные исследования (число здоровых  $\times (1-0,95)$ ). Даже если используется тест со специфичностью 99%, число ложноположительных результатов составит 10 тыс. Представим, что речь идет о дефекте развития плода, и ультразвуковое исследование дает такие результаты. Если семья принимает решение рожать ребенка, то в ней месяцы беременности превращаются в кошмар. Если семья принимает решение делать аборт, то на один погибший дефектный плод приходится 55 здоровых. Даже в обществе, с пониманием относящемся к абортам, такая перспектива не будет приемлемой.

Таблица 5.6. Результаты применения теста с чувствительностью 90% и специфичностью 95% в диагностике и в скрининге

Ситуация	Преваленс	Число обследованных	Больных	Истинно положительных	Здоровых	Ложноположительных
Диагностика	0,01	200	2	2	198	10
Скрининг	0,001	1 000 000	1000	900	999 000	49 950

На этом же примере рассмотрим перспективу дообследования. Предположим, что после ультразвукового обследования в случае обнаружения специфического симптома программа скрининга предусматривает амниоцентез для уточнения диагноза. Это возможно, если дефект зависит от хромосомной аномалии, как, например, синдром Дауна. Исследование клеток на наличие хромосомной аномалии может иметь очень высокие операционные характеристики, однако, сам амниоцентез является травматичной процедурой для беременной и в значительном числе случаев приводит к аборту (обычно до 1% случаев) и другим осложнениям. Таким образом, даже если программа скрининга почти исключает ненужные вмешательства при ложноположительных результатах, сами обследования могут наносить вред.



## Испытания программ скрининга

Целью скрининга *не* является раннее обнаружение болезней или предшествующих им состояний (например, цитологическое обнаружение дисплазии клеток шеечного эпителия — мазки по Папаниколау). Цель скрининга — предотвращение нежелательных исходов. Для подтверждения эффективности *диагностического исследования* достаточно провести диагностический эксперимент (см. разд. 2.2). Для подтверждения эффективности программы скрининга надо показать, что в результате раннего выявления болезней важные для больных и для общества исходы будут наблюдаться реже. Например, если речь идет о ранней диагностике рака, то надо показать, что больные, у которых его рано выявляют, имеют лучшие исходы, чем те, кто получает помощь обычным порядком. Надо еще показать, что в обществе в целом не только исходы при данной болезни становятся лучше, но и важные глобальные исходы улучшаются. Если скрининг действительно сокращает смертность от рака, то должно иметь место и снижение общей смертности. В противном случае можно думать, что скрининг — не лучше рака.

Скрининг выходит за рамки обычной медицинской практики. В результате в сферу медицинских вмешательств вовлекаются люди, в обычных условиях не являющиеся пациентами врача. Поэтому для подтверждения эффективности скрининга необходимы полномасштабные испытания, подтверждающие эффективность всей программы. Требование рандомизированного слепого испытания скрининга не избыточно, а представляет собой только применение к дорогому, массовому и травматичному вмешательству тех же стандартов научной доказательности, которые используются применительно к введению в практику лекарств.

Испытание эффективности скрининга проводится по той же схеме, что и испытание других вмешательств, включающих в себя полный спектр действий системы медицинской помощи — от организации профилактических обследований до лечения в специализированных клиниках. Эта схема обычно не может быть реализована применительно к отдельным пациентам и даже применительно к отдельным медицинским организациям. Если рандомизировать отдельных пациентов, то в рамках одной организации не удастся реализовать разные способы обследования и ведения пациентов в группах вмешательства и контрольных. Медицинские работники станут работать стереотипно, пациенты опытной и контрольной групп будут обследоваться и лечиться



примерно одинаково. В результате возможное различие в исходах между группами будет скрыто. Поэтому чаще прибегают к рандомизации целых районов, из которых одни становятся районами экспериментального вмешательства, а другие — контрольными (рис. 5.20) (кластерная рандомизация) (см. разд. 3.3.2).

В испытании должно быть показано, что скрининг действительно приводит к благоприятному изменению важных исходов. Такими исходами обычно являются снижение специфической смертности и общей смертности. Эти изменения не всегда однозначны. Например, в испытаниях скрининга на рак молочной железы показано небольшое снижение смертности от этой болезни, но снижения общей смертности в результате скрининга не найдено.

Несмотря на то что для скрининга избираются обычно дешевые тесты вследствие больших объемов обследования и большого числа ложноположительных результатов, нуждающихся в дообследовании, а также резкого увеличения числа вмешательств (см. рис. 5.14), стоимость программ скрининга очень велика. Она оказывается особенно большой в расчете на один выявленный случай болезни и на один предотвращенный неблагоприятный исход.

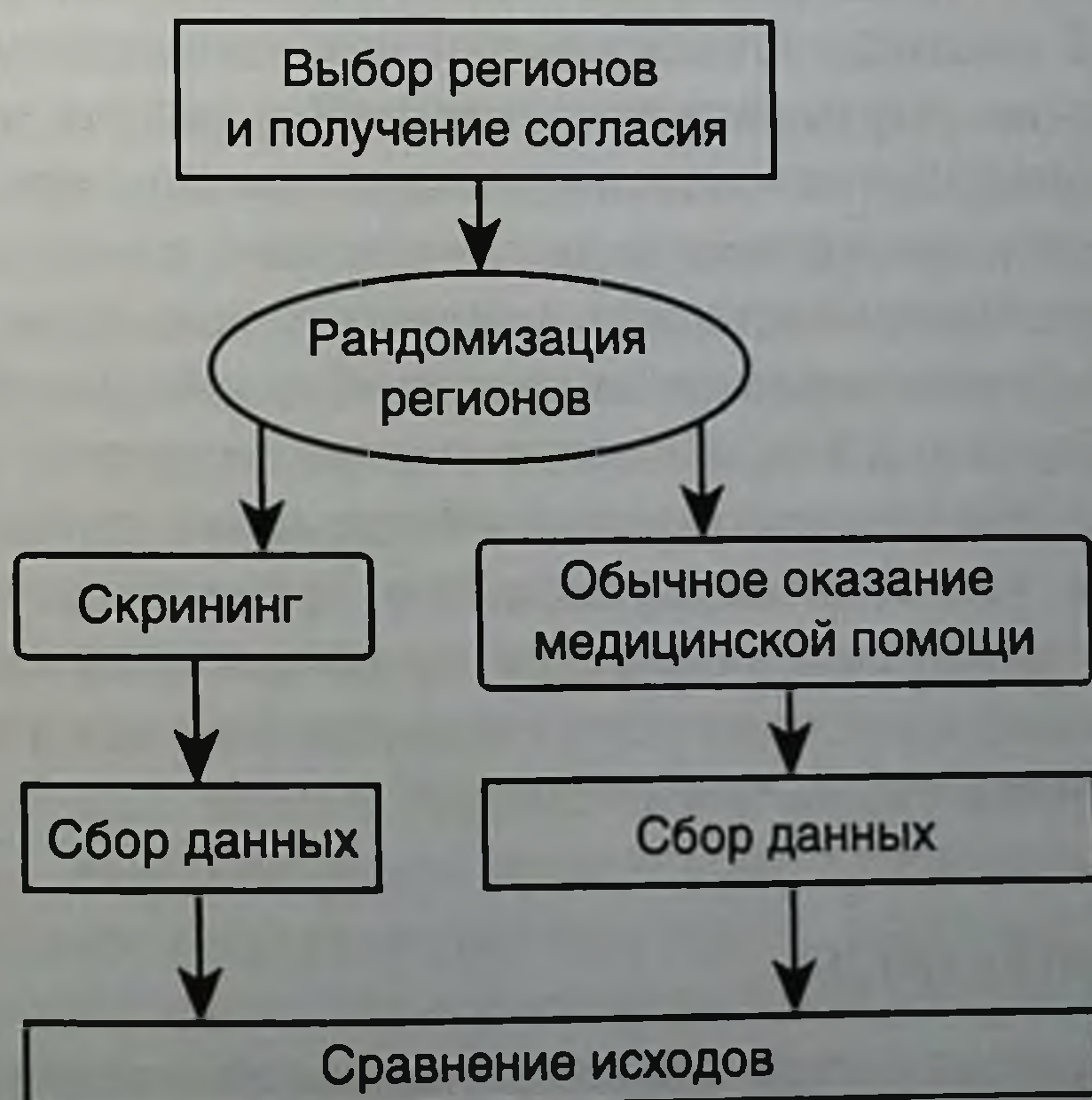


Рис. 5.20. Схема эксперимента для оценки эффективности программы скрининга



В оценке программ скрининга *показатели качества процесса*, в том числе достижение целей по скринингу популяции, полнота дообследования и лечения занимают вторичное положение. На первом месте должны стоять *показатели исходов*, т.е. результаты испытания программы скрининга на действенность. Лишь после того как действенность программы скрининга будет показана в контролируемом испытании, ее можно реализовывать в широких масштабах. Только тогда показатели структуры и качества процесса могут занять свое место в обеспечении качества наряду с исходами. Во всех профилактических программах важная проблема состоит в том, что труднее всего включить в программу людей, принадлежащих к группам самого высокого риска. Между тем именно они нуждаются во вмешательстве более всего. Поэтому нужно установить, насколько программа скрининга служит всему обществу, а не группам населения, имеющим наилучший доступ к медицинской помощи.

## 5.5. ЭПИДЕМИОЛОГИЯ В ФОРМИРОВАНИИ СОЦИАЛЬНОЙ ПОЛИТИКИ

Опираясь на данные различных исследований и будучи методологией научного анализа, эпидемиология является естественным средством обобщения результатов исследований и выхода их в практику. Связь с практикой настолько тесна, что практическое применение эпидемиологических знаний иногда даже включают в определение науки эпидемиологии. Исторически эпидемиологические исследования инициировались именно возникновением проблем большой социальной значимости. Лишь в XX в. общество смогло позволить себе роскошь финансировать эпидемиологические исследования по более широкому кругу вопросов. Тем не менее и в настоящее время эпидемиологическое наблюдение и исследования нередко не получают достаточной поддержки, поскольку в условиях относительного благополучия необходимость таких усилий не кажется очевидной.

### 5.5.1. Профилактика

Профилактические меры являются вариантом медицинского вмешательства, но обычно выходят за рамки мер, применяемых собственно представителями медицинской профессии. Уже поэтому роль обще-



ственного мнения, средств массовой информации, принципиальности законодателей, влияния лоббистских структур оказывается очень большой. Даже в тех случаях, когда эпидемиологические данные достаточны, например данные о вреде курения, законодатели могут не налагать ограничений на рекламу сигарет и не ограничивать курение в общественных местах.

Профилактические вмешательства делят на специфические и неспецифические. Специфическими являются вмешательства, нацеленные на профилактику отдельного состояния (болезни). Обычно под специфической профилактикой понимают вакцинацию. Специфичность вмешательства всегда ограничена. Так, противотуберкулезная вакцина БЦЖ используется с успехом для лечения рака мочевого пузыря, а йод не только для профилактики эндемического зоба, но и для профилактики поражения щитовидной железы в условиях выброса радиоактивного йода в атмосферу.

В реализации профилактических вмешательств существуют два принципиально разных подхода: профилактика в группах высокого риска и профилактика во всей популяции. Профилактика в группах риска интуитивно представляется более правильным подходом, поскольку вмешательство применяется только в отношении лиц, у которых выявлен значительно повышенный риск в ходе скрининга. Таким образом сужается круг лиц, подвергающихся воздействию, им можно уделить больше внимания, достигая выполнения ими профилактических рекомендаций, снижается стоимость программы, уменьшается число побочных эффектов, которые существуют для любого вмешательства. При индивидуальной профилактике в группе повышенного риска снижение риска (разность между вероятностью заболевания без вмешательства и с вмешательством) обычно очень значительно.

Противоположностью является популяционная профилактика, когда объектом вмешательства становится вся популяция, независимо от риска для отдельного субъекта. Этот подход рассматривается по следующим причинам. Дело в том, что для всех факторов риска сердечно-сосудистых болезней имеет место плавное увеличение риска почти во всем диапазоне величины фактора риска, например артериального давления или концентрации холестерина в плазме крови (см. рис. 1.5, 1.6, 2.55). Поэтому выделение лиц только с очень высоким риском приводит к тому, что в этой малой группе может быть предотвращена лишь небольшая часть от всех случаев в популяции (см. табл. 2.5). Выделе-



ние большей доли популяции, например половины популяции, дает возможность предотвратить большую часть случаев болезни, но индивидуальный риск для включаемых в программу лиц приближается к среднему. Обнаружение лиц с высоким риском обычно требует затрат. Отказ от выявления лиц с повышенным риском приводит к экономии всех или почти всех средств, затрачиваемых на оценку риска. Однако при популяционном вмешательстве снижение риска для отдельного человека незначительно, и велики суммарная стоимость вмешательства и общее число побочных эффектов.

Политическим решением во многих странах был выбор комбинированной стратегии профилактики сердечно-сосудистых болезней, в которой сочетаются два подхода — адресованный лицам с высоким риском и адресованный популяции в целом. Вмешательства, направленные на популяцию, имеют очень низкий риск побочных эффектов, приемлемы и дешевы. Вмешательства для лиц с высоким риском могут быть приняты только людьми, мотивированными к этому. У таких людей допустимы некоторая частота побочных эффектов, инвазивные вмешательства, ежедневный прием профилактических лекарств.

### 5.5.2. Оценка и модификация рисков

Эпидемиологические данные, обосновывающие проведение профилактической программы, служат первым и важным, но не решающим фактором для начала реализации такой программы. Решающими факторами являются ценности общества и наличие достаточных средств для проведения программы. Например, в современном русском обществе пренебрежение «малыми» рисками для здоровья приводит к терпимости граждан к курильщикам, повсеместной продаже табака и его рекламе. Точно так же относятся к предотвращению инфицирования гриппом в период вспышек.

В оценке рисков на первом месте стоят исследовательские программы. В них изучают связи общественно значимых исходов с факторами риска (экспозициями), моделируют зависимости и выясняют зависимость частоты исходов от экспозиции, например оценивают вероятность эффекта при малых интенсивностях воздействия (см. рис. 2.23 и 2.24), оценивают экспозицию в популяции, на рабочих местах, в жилищах граждан. Для получения этих данных используются все доступные источники.



Применительно к макроэкспозиции, т.е. экспозиции всей популяции, используются естественно-научные данные (состав поверхностных пород, преобладающие строительные материалы, высота над поверхностью океана, объем осадков, очистка сточных вод и т.д.) и административные базы данных.

Применительно к отдельным людям такими источниками являются:

- ▶ врачебные записи в медицинских документах;
- ▶ больничные документы;
- ▶ данные регистров и специализированных учреждений (диспансеров);
- ▶ свидетельства о смерти;
- ▶ данные работодателей о месте и условиях труда, длительности работы;
- ▶ данные трудового анамнеза и анамнеза жизни, полученные у обследуемых и у их родственников (представителей).

На основании данных, полученных в исследованиях, проводят суммарную оценку, которая позволяет ответить на вопрос, каково влияние фактора риска на популяцию в показателях инцидента неблагоприятных эффектов. Для анализа макроэкспозиции подходит экологический анализ, т.е. сравнение регионов. Для анализа микроэкспозиции необходимо применять весь спектр методов аналитической эпидемиологии (см. разд. 4.3).

Суммарная оценка риска становится основой для разработки методов регулирования риска. Такими методами могут быть законодательные запреты (например, запрет на использование асбеста в строительстве жилых и общественных помещений или полный запрет, как это имеет место в большинстве развитых стран), рекомендации, в том числе позитивные экономические стимулы к замене технологий, просветительская деятельность.

В анализе рисков законодателями и государственными агентствами риск является производным от вероятности неблагоприятных эффектов и их тяжести. Вмешательства для уменьшения риска рассматриваются в сочетании с последствиями, к которым они могут привести на национальном, региональном и индивидуальном уровне. Например, Канада многие годы производит асбест высокого качества и в большом количестве. Запрет на его использование во многих странах привел к сокращению потребления, но запрет производства в Канаде привел бы к исчезновению всей промышленности с очевидными негативными по-



следствиями для добывающих асбест регионов страны. На основании предложений государственные агентства принимают решения, которые в большей или меньшей степени отражают первичные предложения эпидемиологов. Конечный результат на уровне административных и политических решений лишь отчасти зависит от научных, в том числе эпидемиологических, данных. Однако чем точнее эпидемиологические данные о риске, тем больше окончательное решение зависит от них.

Одной из самых важных проблем в принятии решений об устранении вредных экспозиций является принятие решений о запретном уровне этих экспозиций (предельно допустимом уровне, предельно допустимой концентрации). Экономически невозможно устанавливать правила, полностью устраняющие вредную экспозицию. Такие правила будут или катастрофическими для экономики страны, или не будут выполняться, как это было в СССР с весьма строгими стандартами воздуха рабочей зоны, загрязнения городской среды и сточных вод. Установление «приемлемых» предельно допустимых уровней воздействия означает, что у части работников будут наблюдаться негативные последствия вредной экспозиции в связи с их индивидуально высокой чувствительностью, которую обычно невозможно предугадать. Поэтому модификация производственных, экологических рисков силами государственных агентств или силами законодателей всегда предполагает компромисс между научными данными и требованиями, проистекающими из области за пределами науки. В идеале ответственный законодатель, вводя профилактическое вмешательство в интересах популяции, устанавливает справедливый порядок компенсации ущерба людям, пострадавшим от вмешательства, например пострадавшим от осложнений вакцинации.

В последние десятилетия XX в. эпидемиологические данные приобрели такой вес, что стали довольно широко использоваться в судах как способ доказать причинную связь болезни с экспозицией. Наиболее важными были решения американского суда о компенсации табачными компаниями вреда, нанесенного их продукцией гражданам. Сформировалось даже правило, что дополнительный риск более 50% может быть доказательством причинной связи наблюдаемого случая болезни с экспозицией, если экспозиция доказана. Это правило не идеально, но суд и должен использовать относительно простые правила. Ранее иногда в суде использовали как доказательство причинности наличие статистически значимой связи, т.е. связи, вероятность ошибочности



которой менее 5%, что, конечно же, неверно, так как этот показатель не свидетельствует о силе связи. Еще раньше суды основывались в основном на свидетельствах экспертов, но этот подход, практиковавшийся применительно к любому научному свидетельству, скрывал от суда саму научную аргументацию. В 2000 г. Национальный центр юстиции США издал для судей справочное пособие, излагающее основы использования научных свидетельств, в том числе эпидемиологических данных. Можно думать, что вскоре суды и в других странах также станут интенсивнее использовать эпидемиологические данные для вынесения решений.

### Контрольные вопросы и задачи к разделам 5.3 и 5.4

1. Скрининг одинаковым тестом проведен в двух популяциях, но доля ложноположительных результатов среди положительных в популяции А оказалась в 3 раза выше, чем в популяции В. В чем может быть причина?
  - а) это нельзя установить;
  - б) специфичность теста в популяции А выше;
  - в) специфичность теста в популяции А ниже;
  - г) преваленс состояния в популяции А выше;
  - д) преваленс состояния в популяции А ниже.
2. Для скрининга на рак толстой кишки применяются тесты на скрытую кровь в кале, которые имеют весьма умеренные операционные характеристики. Если чувствительность такого теста 70% и специфичность 70%, то в популяции с преваленсом рака толстой кишки 12 на 1000 в возрасте старше 50 лет какой будет прогностичность положительного результата:
  - а) 84%;
  - б) 26%;
  - в) 24%.
3. В случае когда операционные характеристики применяемых тестов низки, тесты иногда применяют «в пакете». Применительно к скринингу на рак толстой кишки применяется следующая стратегия: лицам, у которых тест на скрытую кровь положителен, его повторяют с помощью подобного теста на скрытую кровь и проводят колоноскопию только тем, у кого второй тест положителен. При таком последовательном применении двух тестов на скрытую кровь:



- а) повышается общая чувствительность и специфичность;
  - б) повышается общая чувствительность;
  - в) общая чувствительность снижается, но повышается специфичность;
  - г) общая чувствительность повышается, а специфичность снижается.
4. При оценке эффективности популяционного вмешательства методом исторического контроля по снизившемуся показателю летальности будет правильным предположить, что:
- а) снизившаяся летальность является следствием снижения преваленса болезни;
  - б) снизившаяся летальность является следствием снижения инциденса болезни;
  - в) снизившаяся летальность является следствием внедрения нового вмешательства;
  - г) снизившаяся летальность является следствием действия факторов, не связанных с вмешательством.
5. Скрининг должен предприниматься в отношении болезней:
- а) имеющих низкий инцидент;
  - б) имеющих низкую летальность;
  - в) течение которых может быть изменено лечением;
  - г) которые хорошо выявляются и эффективно лечатся.



## Глава 6

# ЭТИЧЕСКИЕ ПРОБЛЕМЫ ЭПИДЕМИОЛОГИИ

Экспериментальные исследования с участием человека — отдельных людей или коллективов — должны проводиться только с соблюдением этических правил и после одобрения протокола (плана) исследования этическим комитетом. Этические требования к исследованиям сформулированы в Нюрнбергском кодексе<sup>1</sup>, созданной на его основе Хельсинкской декларации Всемирной медицинской ассоциации (WMA, <http://www.wma.org>), детализированы в Рекомендациях Совета научных медицинских обществ (CIOMS [18, 19]), закреплены в Конвенции Совета Европы [20] и законодательстве Российской Федерации [21, 22]. Приняты специальные международные правила, регламентирующие проведение эпидемиологических исследований [18, 23]. В том числе предусматривается ограничение публикации статей, освещающих результаты неэтичных и жестоких исследований. Такого рода ограничение должно предотвратить интерес исследователей к неэтичным экспериментам, так как их результаты нельзя будет опубликовать.

Этические и профессиональные проблемы эпидемиологии сходны с этическими проблемами других медицинских дисциплин, но заметно шире их, поскольку эпидемиология как базовая наука общественного здоровья своими результатами непосредственно выходит в практику и оказывает прямое влияние на большие группы людей. Кроме того, большинство эпидемиологических исследований, в отличие от физиологических, медицинских и многих других, финансируется из общественных средств, что налагает на эпидемиологов дополнительную

---

<sup>1</sup> Кодекс был сформулирован в 1947 г. международным трибуналом в решении по делу немецких врачей, выполнявших преступные эксперименты на людях. В кодексе впервые декларированы условия приемлемого экспериментирования на людях.



ответственность. Наконец, люди — участники эпидемиологических исследований, обычно не получают прямой пользы от участия в исследовании, в отличие от терапевтического эксперимента, где они заведомо в любом случае получают доступ к помощи более высокого качества, чем обычно.

## 6.1. ОБЯЗАТЕЛЬСТВА ПЕРЕД УЧАСТНИКАМИ ИССЛЕДОВАНИЯ

### Согласие

Главным условием этичности исследования является *добровольность участия* в нем людей. Добровольность обеспечивается получением от участника информированного и свободного согласия. Этим достигается уважение к самоопределению людей и предотвращается нанесение им неприемлемого для них вреда. В отличие от клинических исследований, в эпидемиологических исследованиях чаще возникает ситуация, когда полное информирование предполагаемых участников исследования невозможно, поскольку предоставление полной информации привело бы к изменению ими образа жизни или изменило бы их ответы на вопросы. Эта проблема разрешается на основе общих принципов: сокрытие информации возможно только в том случае, если это не сопровождается нанесением субъекту даже минимального вреда и если по окончании исследования участники получают полную информацию.

### Сообщение информации

*Сообщение информации* по результатам исследования является естественным выражением благодарности участникам за потраченное ими время и понесенные неудобства. В ряде случаев такая информация может быть бесполезна для участников исследования, например, по причине невозможности противопоставить что-нибудь обнаруженному фактору риска. Маловероятно, что участники исследования способны изменить свое социальное положение или избежать развития болезни, против которой нет средств профилактики и нет эффективного лечения. Это соображение заставляет некоторых эпидемиологов задавать предполагаемым участникам исследования вопрос, желают ли они получить информацию по окончании исследования. Отказ от по-



лучения информации является законным и должен быть уважаем, но лишение участников исследования информации по решению исследователя или спонсора неприемлемо, так как такой патерналистский акт лишает граждан возможности самостоятельно принимать решения на основе относящейся к ним информации.

### Конфиденциальность

Важнейшим условием получения согласия является гарантия *сохранения в тайне (конфиденциальности)* сообщенных субъектом сведений. Технически конфиденциальность обеспечивается обезличиванием всех записей и предоставлением результатов для анализа только в виде записей с идентификационным кодом. Все участники исследовательского проекта должны дать письменное обязательство не разглашать сведения об участниках исследования. Практически существуют пределы обеспечения конфиденциальности. Например, решение суда о доступе к данным не может быть в большинстве случаев не исполнено. В ходе исследования эпидемиолог может получить информацию о том, что зараженный СПИДом субъект был донором крови, и было бы безответственным перед обществом не предотвратить опасные последствия.

В медицине сохранение сведений о пациенте в тайне является традицией. С конца XX в. и в особенности в эпидемиологии проблема конфиденциальности обостряется в связи с введением электронных записей о пациентах и участниках исследований. Технически доступ к электронным данным о больших группах граждан является настолько простым, что получение согласия граждан на это оказывается основной работой, которую должны выполнить исследователи. В обычных медицинских исследованиях, наоборот, получение согласия является относительно простой однократной процедурой по сравнению, например, с длительным, дорогим и травматичным клиническим испытанием. Поэтому существует правило предоставлять доступ исследователю к электронным записям только после получения согласия участников исследования, только к записям на давших согласие и только в обезличенной форме, т.е. в такой форме, в которой идентификация субъекта невозможна или, по крайней мере, затруднена.

К сожалению, выполнение этого правила в полном объеме делает многие исследования невыполнимыми. Например, пациенты могут умереть к моменту проведения исследования и не могут дать согласия; объем базы данных о населении региона может быть очень большим,



и необходимо именно первичное изучение базы данных для того, чтобы найти небольшую группу людей, отвечающих признакам включения в исследование. Наконец, о лицах, не давших согласия, нельзя извлечь информацию, позволяющую уточнить их отличия от согласившихся участвовать в исследовании.

Практическое решение, уравнивающее реалистичные требования к доступу к данным с необходимостью защиты конфиденциальности, состоит в следующем:

1. Информированное согласие обязательно для всех элементов исследования, кроме извлечения сведений из существующих баз данных.
2. Все полученные данные сохраняются «под замком», в том числе не размещаются в доступных для других лиц компьютерных файлах.
3. На всех исследовательских формах присутствуют только идентифицирующие участников исследования коды, а ключ (документ) к кодам существует в единственном экземпляре, хранится отдельно от данных под замком. Этот ключ не создается в электронном виде, не размещается на компьютере, не включается в таблицы и распечатки данных.
4. Идентифицирующий ключ уничтожается по завершении сбора данных, если его сохранение не одобрено специально этическим контрольным комитетом.
5. Все данные публикуются (обнародуются) только в групповой форме, и никто из участников исследования не идентифицируется.

## 6.2. ОБЯЗАТЕЛЬСТВА ПЕРЕД ОБЩЕСТВОМ

В отличие от медицины, где врач может избежать конфликта, признавая только интересы конкретного пациента, эпидемиология изначально действует в интересах как отдельных граждан, так и в интересах общества. В таком случае возникают специфические проблемы.

### Групповое согласие и предотвращение вреда

Эпидемиологические исследования, будучи обращенными к особенностям групп людей (населения города, учеников отдельной школы, работников предприятия и т.д.), приобретают обязательства перед эти-



ми коллективными субъектами. Эпидемиологические исследования, обращенные к группам, равно как и вмешательства на популяционном уровне, требуют получения *согласия групп* или *их законных представителей*. Последнего может быть недостаточно, поскольку интересы власти могут не совпадать полностью с интересами населения, и обязательным условием является информирование населения о планируемом исследовании. Соответственно при проведении популяционных вмешательств принято не только получать согласие приглашаемой для исследования группы, но и признавать другие обязательства перед группой, аналогичные обязательствам перед отдельным субъектом, например сохранение конфиденциальности данных. В противном случае, например, обнаружение информации о заболеваемости может нанести вред населению региона. Даже если вред в материальном плане маловероятен, необходимо учитывать культурные, национальные и религиозные предпочтения населения.

### **Сохранение конфиденциальности**

Сохранение конфиденциальности данных необходимо потому, что разглашение их *может* нанести вред. Например, выявление в отдельном населенном пункте высокой зараженности населения болезнями, передаваемыми половым путем, или иных особенностей может сделать населенный пункт непривлекательным для инвестиций и туризма. Получение подобной информации возможно, по крайней мере иногда, независимо от проведения исследования, но это не отменяет обязательств для исследователей.

Это обязательство вытекает из общего принципа медицинской этики — *не наносить вред*, т.е. не совершать действия, нацеленного на причинение вреда или несущего больше вреда, чем пользы, или не несущего определенной пользы. Эти действия могут быть совершены другими людьми, но такие действия запрещены для представителей медицинской профессии.

Обязательства перед обществом существуют у отдельного врача, но занимают вторичное положение по отношению к обязательствам перед пациентом, и многими врачами обоснованно игнорируются. Эпидемиология обращена к особенностям групп людей, к групповым проблемам и групповым вмешательствам. Поэтому в эпидемиологии особенно легко интересы отдельного участника исследования, отдельного гражданина могут стать жертвой предпочтения интересов группы,



государства. По определению, свобода и интересы отдельных граждан нарушаются чаще всего тогда, когда преследуются общегосударственные интересы.

### **Проблема расы**

В эпидемиологии признак «раса» используется для распределения участников исследования в подгруппы, например, при сравнении заболеваемости представителей рас. Чаще раса используется для устранения разницы между группами, например, работников разных производств, по признаку «раса» (стандартизации). Термин «раса» тесно связан с произвольной системой классификации людей по внешним признакам (прежде всего — по цвету кожи), распространенной в XIX в. Например, синдром Дауна имел название «монголизм», так как предполагалось, что при этом наблюдаются проявления примитивной «монголоидной расы». Не существует ясной границы между людьми, которых в соответствии с расовой концепцией следовало бы отнести к разным расам. Наоборот, доказано, что существует непрерывный спектр генетического разнообразия человека. В исследовательской и социальной практике в XX в. термин «раса» используется для обозначения культурной самоидентификации и определяется по тому, как сам человек себя обозначает. Чаще используется термин «этническая» или «национальная» принадлежность для обозначения группы людей, к которым себя относит участник исследования. Этническая принадлежность является в большей степени культурной и языковой, чем биологической.

Использование национальной принадлежности как переменной в эпидемиологических исследованиях не должно заменять такие прямо определяемые переменные, как вероисповедание, особенности диеты, другие поведенческие признаки. Связи национальной принадлежности с образованием или делинквентностью не должны заменять связи этих признаков с конкретными предполагаемыми причинными факторами (вакцинация, образование, доход родителей и т.п.). Признак национальной принадлежности должен использоваться в эпидемиологических исследованиях, как и всякий другой признак, только в случае наличия обоснованной гипотезы.

### **Повторный анализ данных**

По завершении исследования его исходные материалы должны быть сохранены для возможного повторного анализа, по крайней мере,



в течение 5 лет после обнародования его результатов. В медицине, где исследования проводятся с участием пациентов, бескорыстное предоставление последними своего тела, своего времени исследователю делает полученные сведения общественным достоянием. В некоторых странах закон обязывает сохранять сведения, полученные в ходе исследований, выполненных на средства бюджета и обеспечивать к этим материалам доступ другим исследователям. В результате такой практики материалы многих исследований повторно используются для проверки новых гипотез и позволяют получить новые знания без проведения новых дорогостоящих исследований. Ведущие международные журналы требуют от авторов поступивших статей подтверждения, что исходные материалы исследования сохранены и могут быть предъявлены для контрольного анализа.

Результаты эпидемиологических исследований подвергаются повторному анализу значительно чаще, чем результаты клинических исследований. Чем более важные результаты получены и чем болезненнее проблема, тем выше вероятность проведения нового анализа. При решении вопроса о доступе к данным других исследователей принято руководствоваться следующим.

1. Другие исследователи получают доступ к данным тогда, когда *основной* анализ авторами проведен, и опубликование результатов является признанием авторами того, что этот анализ проведен.
2. Процедура доступа к данным зависит от правил, принятых в конкретной стране, и, как правило, данные, выполненные на государственные средства или на средства государственных университетов, являются открытыми для новых анализов, а в случае финансирования исследований частными предприятиями доступ к данным осуществляется с их согласия.
3. Точка зрения исследователя, запрашивающего данные для нового анализа, не должна влиять на решение о доступе к данным, а интересы исследователя, проведшего исследование, защищены тем, что он уже выполнил анализ данных.
4. Ключи для идентификации участников исследования, как правило, не должны предоставляться, даже если они сохранены.

### Публикация

Обнародование результатов исследования является обязанностью исследователя перед участниками исследования, финансирующими



организациями (государством, университетом) и обществом. Анализ данных может занимать длительное время, и результаты анализа могут публиковаться по мере завершения отдельных этапов работы. Тем не менее исследователь не вправе публиковать фрагменты анализа до выполнения общего анализа. В противном случае результаты последующих этапов анализа могут вступить в противоречие с опубликованным ранее. Этим может быть нанесен вред как участникам исследования, так и учреждению, в котором работает исследователь.

Автор должен раскрывать свои финансовые интересы, указывая, кто является его работодателем (например, университет) и кто финансировал исследование. Иначе читатель (потребитель) научного сообщения не сможет установить, присутствует ли конфликт интересов. Если авторы выполняют работу на средства ассоциации производителей кофе и публикуют работу, показывающую, что потребление кофе не связано с риском рака поджелудочной железы, это не обязательно означает, что исследование искажено в угоду финансирующей стороне. Даже если нет конфликта интересов, *перспектива анализа* может быть различной, и в зависимости от этого он может привести к разным выводам. Например, скрининг с равной вероятностью выявления и лечения рака прямой кишки может выглядеть по-разному, если его рассматривать в перспективе полезности отдельному пациенту и в перспективе полезности обществу. Поэтому сокрытие потенциального конфликта интересов считается в международном медицинском сообществе грубым *нарушением правил профессионального поведения в науке* (англ. *research misconduct*), сопоставимым с фальсификацией данных.

Исследователь несет ответственность за качество исследования и за обнародование его результатов. Результаты исследования являются интеллектуальной собственностью исследователя, а не финансирующей исследование организации, государственной или коммерческой. Всякие попытки финансирующей исследование организации или других заинтересованных сторон задержать обнародование результатов или повлиять на содержание отчета должны категорически отвергаться.

Публикация данных эпидемиологического исследования, как и всякого другого исследования, должна начинаться с представления сообщений в профессиональные журналы, где эти сообщения станут предметом рецензирования. При подготовке публикации в журнале автор имеет возможность на основании рецензий уточнить анализ и способ представления данных. В результате в печати появляется от-



ветственное, взвешенное сообщение, несущее выверенную информацию. Лишь после этого результаты исследования могут обсуждаться с журналистами, представляться на пресс-конференциях, публиковаться в популярных журналах. Исключением являются лишь исследования, которые выявляют немедленную и значительную опасность для людей, о которой необходимо немедленно сообщить.

К сожалению, некоторые исследователи не выполняют этого правила и в стремлении к привлечению внимания к себе обращаются в первую очередь к журналистам. В результате граждане знакомятся с сообщениями о том, что в действительности не является важным или не проверено в надежных исследованиях. Ярким примером является нескончаемое обсуждение в прессе так и не обнаруженного вреда здоровью от сотовых телефонов и крайне редких случаев нападения акул. Одновременно признается возможной реклама курения, хотя *огромный* вред здоровью от курения убедительно доказан. Большая часть искажений реальных проблем создана не врачами, но врачи вносят свой вклад в деформацию правильной картины, сообщая неверную информацию или неправильным образом. Более того, врачи и эпидемиологи участвуют в формировании местной и государственной политики в вопросах, касающихся общественного здоровья, и несут ответственность за донесение до законодателя научных оснований для принимаемых решений.

## Контрольные вопросы и задачи к главе 6

1. Фторирование водопроводной воды является эффективным способом профилактики кариеса, но сопряжено с некоторым риском нежелательных эффектов (флюороза), поэтому принятие решения о фторировании воды должно:
  - а) быть волевым для большего блага граждан;
  - б) принято ответственными исполнительными органами;
  - в) широко пропагандироваться в печати;
  - г) не привлекать внимания;
  - д) стать предметом предварительного общественного обсуждения.
2. При планировании исследования число пациентов должно быть:
  - а) минимальным, чтобы уменьшить возможный вред;
  - б) большим, чтобы получить возможно более точную оценку полезности и опасности вмешательства;



- в) достаточным для выявления ожидаемого эффекта при заданной вероятности ошибок типа 1 и 2;
  - г) зависящим от того, насколько опасно изучаемое состояние.
3. При получении согласия на участие в эксперименте приглашаемым пациентам должна быть предоставлена информация:
- а) любая, какую они запросят;
  - б) строго в рамках протокола;
  - в) минимальная и не травмирующая;
  - г) заверяющая их в отсутствии вреда.
4. Результаты популяционных исследований должны обнародоваться в соответствии с принципом:
- а) поощрить обследованные популяции к устранению обнаруженных факторов риска;
  - б) наказать лиц, виновных в возникновении вреда здоровью граждан;
  - в) не нанести вреда участникам исследования и популяции в целом;
  - г) научной объективности и полноты;
  - д) независимости от интересов финансирующих организаций.



# ОТВЕТЫ НА КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

## Разделы 1.1, 1.2

1: б. Возникновение инфаркта не определяется полностью изменениями, возникающими в течение часов перед развитием клинических проявлений. Действительно, часть пациентов умирают, не будучи доставленными в стационар, а часть переживают бессимптомный инфаркт, но нельзя заранее сказать, какое влияние на предпринятое сравнение это оказывает. Перенесшие клинически выраженный или бессимптомный инфаркт, вероятно, ближе по своим характеристикам к основной группе (с инфарктом), и их исключение только увеличило бы различие с контрольной группой.

2: д. Связь можно с большими основаниями интерпретировать как причинно-следственную, если соблюдаются условия а–г; выполнение исследования фармацевтической компанией или другими, связанными с промышленностью исследователями, дает результаты, смещенные в сторону усиления привлекательной связи, но не может быть основанием для исключения причинного характера связи.

3: в. В эксперименте изучать подобного рода связи неприемлемо, и исследование типа сравнения с контролем — соответствующая структура исследования. Нам не известна статистическая значимость различия, но контрольная группа (группа сравнения) есть: для женатых (замужних) — это не состоящие в браке. В исследовании типа сравнения с контролем важна правильная комплектация групп, иначе может оказаться, например, что не состоящие в браке — это молодые неквалифицированные работники.

4: в. Э не изучает болезней других животных — этим занимается другая наука (зоологическая). Начавшись как изучение распространения инфекционных болезней, Э ныне охватывает проблемы общественного здоровья, в том числе исследование таких вмешательств, как терапевтические и диагностические, и способы организации помощи как частные вопросы.

5: б. В этом исследовании не было спланированного врачом вмешательства. Это проспективное исследование. В исследовании типа сравнения с контролем проводилось бы сравнение матерей у детей с разной массой при рождении. Действительно, часть курящих может называть себя некурящими или неверно сообщать о выкуриваемой дозе, и объек-



тивизация лабораторными методами может быть полезной. Отсутствие объективизации может ослабить выявленную связь курения и массы тела новорожденных. Ограничение исследования только родами позднее 20 нед является обоснованным ограничением. Представляет интерес не только завершение беременности аборт, но и другие осложнения у новорожденного и беременной, но никакое исследование не может быть всеобъемлющим.

б: г.

### Раздел 1.3

1: а)  $50/200\,000\,000 = 0,000\,000\,25 = 25$  на 100 млн или 0,25 на 1 млн;  
б)  $3/50 = 6\%$ ; в) 1,5 на 100 млн.

2: а. Коррекция на возраст позволяет устранить различия в распределении по возрасту в сравниваемых популяциях, но только в случае, если сравниваются полные данные или репрезентативные выборки; смещение в формировании выборки вследствие низкого качества свидетельств о смерти или неполноты данных скорректировать нельзя.

3: б. У детей выше инцидент и соответственно факторы риска должны встречаться чаще; общая смертность обычно не может скрывать многократные различия в смертности от отдельных причин; отсутствие различий в преваленсе при больших различиях в инциденте может наблюдаться, когда высокий инцидент отражает развитие в группе особой агрессивной формы болезни, отличающейся высокой смертностью.

4: г. Вероятность К 0,05, П 0,1; совместная вероятность КП  $0,05 \times 0,1 = 0,005$ , т.е. ожидаемое число случаев КП — 5; фактически имеет место  $150 - 120 = 30$  случаев, т.е. К и П имеют тенденцию наблюдаться совместно.

5: г. Всего инфузий за 2 года 6; вероятность заражения при каждой инъекции не зависит от предшествующих и последующих событий, но зараженный однажды уже не заразится, если не иметь в виду инфицирование другим вирусом; поэтому *вероятность не заразиться* составляет  $0,95 \times 0,95$  и т.д.; соответственно, вероятность заражения составляет  $1 - 0,95^6 = 1 - 0,74 = 0,26$ .

### Раздел 1.4

1: в.

2: б—г. Стандартизация условий взятия проб уменьшает внутрииндивидуальные вариации, но не влияет на аналитические.



3: а, б, в.

4: б. Этот вариант используется для вычисления плотности инцидента.

### Разделы 1.5, 1.6

1: г.

2: б.

### Раздел 1.7, 1.8

2: в. В проверке гипотез исходят из предположения о справедливости нулевой гипотезы, т.е. гипотезы об отсутствии различий, поэтому ее вероятность оценить нельзя, а можно лишь отклонить ее, если  $P$  достаточно мало; организация исследования и метод формирования выборки влияют на способ вычисления  $P$ .

2: в, г.

3: а, б, в, г. Наличие статистической значимости не исключает случайного происхождения различия, а лишь указывает на низкую вероятность этого; такое различие может быть любого происхождения — это зависит уже от интерпретации, а не от статистической значимости.

4: г. Секторальную диаграмму обычно не следует применять для количественных признаков, а столбиковая должна обычно применяться к номинативным признакам, соединять экспериментальные точки линией не следует, так как результаты получены на разных пациентах; скаттерграмма наилучшим образом отражает соотношение результатов для двух количественных признаков.

### Разделы 2.1, 2.2

1: д. При исследовании только больных можно определить только чувствительность, но не специфичность и не прогностичность, и этот способ определения чувствительности подвержен смещениям, поскольку пациенты не обследуются вслепую.

2: б. При низком преваленсе прогностичность положительного результата уменьшается, а отрицательного — увеличивается за счет относительно большого числа ложноположительных результатов.

3: г.

4: б.



**Разделы 2.3, 2.4**

1: б. Выявление источника основано на сопоставлении частоты заболевания — разница между евшими и не евшими подозрительной пищи должна быть наибольшей.

2: г.

3: в.

4: в.

**Раздел 2.5**

А: б.

Б: б.

В: а—в.

Г: а.

**Раздел 2.6**

1: в. Приведенные показатели характеризуют структуру смертности — пропорциональное отношение смертности; для вычисления относительного риска необходимо иметь показатели смертности в виде (число смертей от рака)/(число людей этого возраста), их отношение и будет относительным риском.

2: в.

3: а, б, г, д.

4.1.: а. 4.2.: г. 4.3.: в. 4.4.: б.

**Раздел 2.7**

1: б, г.

2: а, в. Когортные исследования могут быть выполнены при ретроспективном сборе данных; основная цель когортных исследований всегда формулируется как аналитическая.

3: б.

4: в.

**Раздел 2.8**

1: б.

2: в, д. Это исследование типа сравнения с контролем, и для ИСК найденное повышение шансов невелико, оно вполне может быть следствием смещения вследствие смещения памяти — более полного упоминания экспозиции теми участниками, кто имеет изучаемое пора-



жение; более полное включение случаев, развивающихся в более позднем возрасте, или слабо выраженных случаев не обязательно приводит к увеличению ОШ.

3: б. В исследовании типа сравнения с контролем повышение шансов в 4 раза — заметное повышение; то, что 95% ДИ не включает 1, означает, что  $P < 0,05$ , а 90% ДИ будет уже; связь может быть интерпретирована: подросток, у которого один или два родителя курят, имеет шансы в 4,3 раза большие курить, чем тот, у кого родители не курят.

3: а, б, в.

4: а.

### Разделы 3.1–3.4

1: в. Экспериментальное и обсервационное исследования могут быть проспективными, иметь любые размеры, но только в экспериментальном — экспозиция определяется исследователем.

2: а. Рандомизация позволяет сделать назначение вмешательства или плацебо случайным, но она не может гарантировать, что группы будут одинаковы, так как различия между ними обязательно возникают случайным образом и тем больше по размерам, чем меньше группы; рандомизация помогает в ослеплении и таким образом косвенно помогает объективно оценивать исходы, но не гарантирует этого; эффект плацебо возникает во всех группах, где есть вмешательство.

3: а, б, в, г, д.

4: г.

5: в.

6: г.

### Раздел 3.5

1: г.

2: б.

3: а.

4: 3. Группа лечения:  $2/10 = 0,2$ ; контрольная  $6/11 = 0,54$ ; абсолютное снижение риска  $0,54 - 0,2 = 0,34$ ; ЧПЛП =  $1/АСР = 1/0,34 = 2,9$ ; округляем в сторону увеличения — 3 пациента.

5: б. Поскольку группы созданы рандомизацией, их различия случайны, и проверять статистическую значимость не следует, а учесть различия при анализе необходимо.



**Глава 4**

- а: 2.
- б: 5.
- в: 1.
- г: 3.
- 2: д.
- 3: д.
- 4: а, б, г.

**Разделы 5.1, 5.2**

- 1: а, г.
- 2: все приведенные важны.

3: д. Национальные центры обычно оказывают помощь лишь части больных страны, поэтому поступление больных в эти центры не отражает инциденса и преваленса, а зависит от многих факторов, среди которых, конечно, порядок госпитализации и состояние финансирования, но влияние других факторов так велико, что и эти два важных фактора не влияют на число госпитализируемых однозначно.

- 4: б, в, г.
- 5: б.
- 6: а, б, г.

**Разделы 5.3, 5.4**

- 1: д.
- 2: в.
- 3: в.
- 4: г.
- 5: в.

**Глава 6**

- 1: д.
- 2: в.
- 3: а.
- 4: а, в, г, д.



## СЛОВАРЬ ОСНОВНЫХ ЭПИДЕМИОЛОГИЧЕСКИХ ТЕРМИНОВ

Более полный Эпидемиологический словарь под ред. Дж. Ласта в русском переводе с указателем англо-русских соответствий свободно доступен в интернете (<http://osdm.org/blog/2014/04/24/epidemiologicheskij-slovar/>).

**Анализ выживания** — метод статистического анализа данных о выживании больных, позволяющий оценить динамику смертности и использовать данные обо всех пациентах, в том числе о тех, чей статус известен только в течение короткого периода времени. С помощью анализа выживания можно оценивать эффекты вредных воздействий и влияние лечения на выживание. Анализ выживания применим к анализу любых данных типа времени жизни (время до выписки из больницы, до первого обострения заболевания).

**Анализ чувствительности к смещениям** — повторный анализ данных с использованием различных значений привходящих факторов (*преваленс*, летальность, *комплаенс* и др.) с целью выяснить, насколько результат зависит от этих факторов, например, каких результатов можно ожидать от лечебного вмешательства в разных условиях.

**Аналитические исследования** — исследования зависимостей между признаками, между *факторами риска* (*экспозицией*) и *заболеваемостью*. Отличаются тем, что в аналитических исследованиях проверяется заранее сформулированная гипотеза, например, о зависимости преваленса болезни от способа обработки мяса, принятого в стране. К аналитическим исследованиям относятся *поперечные*, *когортные*, *исследования типа сравнения с контролем* и *экспериментальные*. Ср. *описательные* и *поисковые исследования*.

**Вариабельность** — степень различий между повторными измерениями признака у одного субъекта (*внутрииндивидуальная В*) или у разных лиц (*межиндивидуальная В*). Если сопоставляются результаты отдельных исследований, то большие различия называют *гетерогенностью* результатов.

**Вероятность** (относительная частота) — число событий, отнесенное к численности популяции или выборки из нее, например 2 заболевания на 36 наблюдавшихся детей. Обычно выражается в долях единицы (0,06), процентах (6%) или на стандартное число людей (55 на 1000).



**Вмешивающиеся факторы** — факторы (условия), которые могут оказывать влияние на результат исследования помимо основных изучаемых факторов. *Структура исследования*, способы формирования групп и способ анализа данных избираются таким образом, чтобы исключить влияние вмешивающихся факторов или учесть его в ходе анализа данных.

**Выборка** — группа людей, избранных из *популяции* для изучения закономерностей, имеющих место в популяции. Размер  $N$  планируется таким, чтобы исследование имело достаточную *чувствительность*, а способ получения выборки с помощью *рандомизации* позволяет уменьшить вероятность *систематической ошибки*.

**Генеральная совокупность** — *популяция*, которую отражает *выборка*, например совокупность больных диабетом второго типа.

**Гетерогенность** — высокая степень различия результатов измерений или результатов исследований, обычно означает невозможность получения обобщенной оценки.

**Действенность вмешательства** — степень, в которой вмешательство уменьшает *вероятность* или выраженность неблагоприятных исходов по сравнению с другим вмешательством (невмешательством), например снижение смертности, инвалидности. Оценивается в контролируемых условиях *клинических испытаний* (ср. *эффективность вмешательства*).

**Доверительный интервал** — интервал величин, в котором с установленной *вероятностью* (например, 95%) лежит величина *популяционной* оценки. Рассчитывается при статистическом анализе *выборочного* исследования.

**Доказательная медицина** — способ медицинской практики, отличающийся сознательным и последовательным применением только вмешательств, *действенность* которых доказана в доброкачественных исследованиях.

**Значимость клиническая** — степень, в которой обнаруженная в исследовании *действенность* вмешательства убеждает врача и больного в целесообразности его использования.

**Значимость статистическая** — степень, в которой результат исследования, например снижение летальности, реально существует, т.е. не может быть объяснено случайностью. Статистические методы позволяют оценить, насколько вероятно обнаружение между вмешательством и исходом связи такой же или большей величины, чем



наблюдалась в исследовании случайным образом (когда справедлива нулевая гипотеза). Традиционным порогом значимости является эта вероятность, равная 0,05.

**Инцидент** — частота новых случаев в популяции.

**Исследования типа сравнения с контролем (ИСК)** — вид медицинского исследования, в котором сравниваются люди, имеющие и не имеющие изучаемое состояние по их настоящим или прошлым признакам. Применяется в основном для выявления связи изучаемого состояния (болезни) с предшествующей *экспозицией* — возможной причиной развития состояния.

**Исход** — событие, по которому судят о возникновении или прогрессировании болезни, ее излечении или других процессах. Различают клинически важные (основные) исходы — летальность, длительность ремиссии, качество жизни; и косвенные (суррогатные) — артериальное давление, плотность костной ткани.

**Клинические испытания (КИ)** — вид медицинских исследований, в которых изучается *действенность* медицинских вмешательств. Выделяют контролируемые КИ, в которых действенность вмешательства сравнивается с другим вмешательством (невмешательством) и *рандомизированные КИ*, которые дают наиболее надежную оценку *действенности*.

**Когортные исследования** — см. *Продольные исследования*.

**Кокрейн** (ранее — Кокрейновское сотрудничество, Cochrane Collaboration) — международная организация, объединяющая специалистов и потребителей, подготавливающих систематические обзоры по эффективности медицинских вмешательств; обзоры объединены в электронную «Кокрейновскую библиотеку».

**Маскировка** — см. *Ослепление*.

**Метаанализ** — статистический метод обобщения результатов исследований.

**Обобщаемость или генерализация** — возможность перенесения результатов выборочного исследования на популяцию или другие группы больных с данным заболеванием или группы, подвергающиеся воздействию аналогичных условий труда.

**Обсервационные исследования** — исследования, в которых наблюдают за естественным развитием событий (без вмешательства исследователя). К обсервационным исследованиям относятся *когортные, поперечные и исследования типа сравнения с контролем*.



**Обследование** — вид исследования, в котором все члены *популяции* (сплошное О.) или *выборка* из нее (выборочное О.) подвергаются измерениям (*опрашиваются*); вариант *поперечного исследования*.

**Объяснимый риск** (атрибутивный риск) — часть риска у подвергшихся экспозиции лиц, которая связана с экспозицией. Измеряется как разность вероятности неблагоприятного события (заболевания) в группе, подвергшейся экспозиции (например, вредным условиям труда) и в контрольной группе.

**Операционные характеристики диагностического теста** — см. *Чувствительность, специфичность*.

**Описательные исследования** — исследования, в которых описывается случай заболевания, совокупность случаев, или признаки *популяции*. К описательным исследованиям относятся исследования типа *обследование* и *регистр*.

**Опрос** — вид исследования, в котором его участники отвечают на вопросы, предъявляемые интервьюером или самостоятельно (отвечая на вопросы, собранные в опросник-листочку или предъявляемые в компьютерной программе).

**Ослепление** (маскирование) — вариант исследования, в котором его участники (слепое) или участники и исследователи (двойное слепое) не осведомлены, к какой группе принадлежит субъект, каковы результаты проведенных диагностических тестов, подвергался ли он *экспозиции*. В *клинических испытаниях* основной целью является маскировка результатов *рандомизации*. Применяется в исследованиях разной *структуры* для уменьшения *систематических ошибок*.

**Отклик** — доля больных или членов иной *популяции*, согласившаяся на участие в исследовании, например вернувших заполненные опросники. Низкий отклик может означать повышенную вероятность *систематических ошибок* в связи с включением в исследование только небольшой части людей, отобранных помимо воли исследователя.

**Плацебо** — недействующее вещество (вмешательство), которое предоставляется участникам *клинического испытания*, отнесенным к контрольной группе, для того чтобы *маскировать* отсутствие вмешательства.

**Подобранные пары** — вариант подбора участников исследования, в котором каждому субъекту, помещенному в основную группу, подбирается один или более парных субъектов с аналогичными характеристиками, например, того же пола и возраста, помещаемых в группу сравнения (контрольную группу). Используется в исследованиях раз-



ной структуры, прежде всего в *клинических испытаниях* и *исследованиях типа сравнения с контролем*.

**Поисковые исследования** — вид исследований, в которых не существует проверяемой гипотезы, а рассматривается множество признаков для выявления возможных связей.

**Поперечные исследования (одномоментные)** — вид исследований, в которых участники подвергаются однократному обследованию, обычно в течение относительно короткого периода времени, что позволяет получить как бы поперечный срез через ряд характеристик *популяции*.

**Популяция** — в эпидемиологии вся совокупность людей, проживающих на территории, или вся совокупность больных некоторым заболеванием, т.е. та совокупность, на которую предполагается обобщать данные исследования. Соответственно, *выборка* из П для проведения выборочного исследования планируется таким образом, чтобы она отражала существенные свойства П.

**Преваленс** — относительная частота признака (болезни, состояния, *экспозиции*) в популяции.

**Прогностичность диагностическая** — относительная частота искомого заболевания (состояния) у пациента с известным результатом диагностического теста.

**Продольные исследования (когортные)** — вид исследований, в которых участники подвергаются повторным обследованиям. В результате удается установить истинные временные отношения. Могут быть выполнены не только *проспективно*, но и по архивным данным.

**Проспективные исследования** — вид *продольных исследований*, в которых участники планомерно включаются в исследование, и за ними наблюдают, выявляя изменения в состоянии здоровья и других признаках.

**Рандомизация** — процесс случайного выбора людей из популяции или случайного размещения людей в экспериментальные группы. В *выборочном* исследовании — важнейший способ обеспечения *репрезентативности* выборки, а в клиническом испытании — обеспечения сравнимости групп и *маскировки* вмешательств. Выполняется с помощью специальных методов, обеспечивающих случайные числа (таблицы случайных чисел, бросание монеты и др.).

**Репрезентативность** — при проведении выборочного исследования формирование *выборки* таким образом, чтобы ее свойства отражали свойства *популяции*. Важнейший способ достижения *обобщаемости* результатов исследования.



**Систематическая ошибка** — отклонение результатов измерения или оценки, полученной в исследовании, от истинного значения. Например, завышение действенности изучаемого вмешательства вследствие включения в основную группу пациентов с более легкой формой болезни, чем в группе сравнения. Систематическая ошибка обычно является следствием действия *вмешивающегося фактора*, действие которого не удалось устранить (учесть в анализе).

**Смещение** — см. Систематическая ошибка.

**Скрининг** — обследование людей, не имеющих симптомов заболевания, для его раннего выявления и лечения.

**Специфичность диагностическая** — доля людей, не имеющих заболевания, у которых диагностический тест дает отрицательный результат.

**Структура исследования (дизайн)** — организация (архитектура, план) исследования, который включает выбор участников исследования, их разделение на подгруппы, наличие воздействия на участников, порядок измерений.

**Характеристическая кривая** — кривая, отражающая возможности метода разделять две группы объектов. Для диагностического теста — зависимость чувствительности от специфичности при изменении строгости подхода к оценке изображения или при сдвиге точки разделения в оценке результатов измерения.

**Чувствительность диагностическая** — доля людей, имеющих заболевание, у которых диагностический тест дает положительный результат.

**Чувствительность статистическая (мощность)** — способность исследования выявить различия между группами участников, зависимости между признаками и другие закономерности. Чувствительность тем выше, чем совершеннее структура исследования и выше численность участников.

**Экологические исследования** — вид исследований, в которых сравниваются характеристики популяций, проживающих в разных местностях. В результате удается сопоставить, например, заболеваемость и особенности питания населения или локальные загрязнения воздушной среды.

**Экономический анализ** — вид исследования, в котором сопоставляются последствия заболеваний и медицинских вмешательств с затратами.

**Экспериментальные исследования** — вид исследований, в которых участники подвергаются воздействию и повторным обследованиям для выявления эффектов воздействия.



**Экспозиция** — контакт человека с фактором, который предположительно вызывает заболевание, например, с ионизирующим излучением или курением. Термин «экспозиция» относится также к количественной характеристике внешнего воздействия. Экспозицией может быть не только вредной, но и полезной, например экспозиция к вакцине.

**Эпидемиологический переход** — переход популяции от высокой смертности и высокой рождаемости к низкой смертности и низкой рождаемости и от смертности от инфекций к смертности от хронических болезней.

**Эффект неблагоприятный** — *побочный эффект* вмешательства, который пациент и врач рассматривают как нежелательный или вредный.

**Эффект побочный** — эффект медицинского вмешательства, возникающий без намерения врача, непредсказуемо или как продолжение известного эффекта вмешательства.

**Эффективность вмешательства** — степень, в которой вмешательство уменьшает *вероятность* неблагоприятных исходов по сравнению с другим вмешательством (невмешательством), например снижение смертности, инвалидности в обычных условиях (ср. *действенность вмешательства*).



## СПИСОК ЛИТЕРАТУРЫ

1. Эпидемиологический словарь (перевод Dictionary of Epidemiology, ed. J. Last). 4 ed. Власов В.В., ed. М.: Открытый институт здоровья, 2009. 316 с.
2. Давыдовский И.В. О проблеме причинности в медицине (Этиология). М.: Медицина, 1965. С. 1–184.
3. Kuhn T. (Кун Т.) Структура научных революций. М.: АСТ, 2001. 606 с.
4. Glantz S.A. (Гланц С.) Медико-биологическая статистика. М.: Практика, 1999. 459 с.
5. Реброва О.Ю. Статистический анализ медицинских данных: применение пакета прикладных программ STATISTICA. М.: МедиаСфера, 2002. 305 с.
6. Petrie A., Sabin C. (Петри А., Сэбин К.) Наглядная статистика в медицине. М.: ГЭОТАР-Мед, 2003. 143 с.
7. Де Пой Э., Гитлин Л.Н. Методы научных исследований в медицине и здравоохранении. М.: ГЭОТАР-Мед, 2017.
8. Власов В.В. Реакция организма на внешние воздействия : Общие закономерности развития и методические проблемы изучения. Иркутск: Иркутский университет, 1994. 344 с.
9. Покровский В.И. Руководство по эпидемиологии инфекционных болезней. Под ред.: Покровский В.И. М.: Медицина, 1993. С. 5–25.
10. Власов В.В. Эффективность диагностических исследований. М.: Медицина, 1988. 1–245 с.
11. Johnson W.O., Ward E.B., Gillen D.L. Chapter 15. Bayesian Methods in Public Health. In: Srinivasa Rao ASR, Pyne S., Rao C.R., ed. Handbook of Statistics. Elsevier. 2017. P. 407–442.
12. Fletcher R.H., Fletcher S.W., Wagner E.H. (Флетчер Р., Флетчер С., Вагнер Э.) Клиническая эпидемиология : основы доказательной медицины. М.: МедиаСфера, 1998. 347 р.
13. Fleiss J.L. (Флейсс Дж.) Статистические методы для изучения таблиц долей и пропорций. М.: Финансы и статистика, 1989. 319 р.
14. Draper N.R., Smith H. (Дрейпер Н., Смит Г.) Прикладной регрессионный анализ. М.: Финансы и статистика, 1986. Т. 1, 366-Т. 2, 51 р.



15. Afifi A.A., Azen S.P. (Афифи А., Эйзен С.) Статистический анализ : подход с использованием ЭВМ. М.: Мир, 1982. 488 р.
16. Федеральное руководство для врачей по использованию лекарственных средств (формулярная система). М.: ГЭОТАР-Мед, 2014.
17. Руководство по профилактической медицине (пер. с англ). М.: Новая слобода, 1993. 160 с.
18. International ethical guidelines for epidemiological studies. Geneva: CIOMS, 2009. 1–130 р.
19. International ethical guidelines for biomedical research involving human subjects. Geneva: CIOMS, 2002. 63 р.
20. Конвенция о защите прав и достоинства человека в связи с применением достижений биологии и медицины : Конвенция о правах человека и биомедицине, СЕД № 164. 1997.
21. Об основах охраны здоровья граждан в Российской Федерации от 21.11.2011 № 323-ФЗ. М., 2011.
22. Об обращении лекарственных средств. 61-ФЗ (2010).
23. Pharmacoepidemiology ISf. Guidelines for Good Pharmacoepidemiology Practices (GPP): International Society for Pharmacoepidemiology; 1996. Available from: <https://www.pharmacoepi.org/resources/policies/guidelines-08027/>.



## ПРИЛОЖЕНИЯ

### Приложение 1. Величины $t$ -распределения при 5–200 степеней свободы для использования в вычислении доверительных интервалов

Рассчитаны с помощью программы Excel.

Число степеней свободы	Ширина доверительного интервала			Число степеней свободы	Ширина доверительного интервала		
	90%	95%	99%		90%	95%	99%
5	2,02	2,57	4,03	105	1,66	1,98	2,62
10	1,81	2,23	3,17	110	1,66	1,98	2,62
15	1,75	2,13	2,95	115	1,66	1,98	2,62
20	1,72	2,09	2,85	120	1,66	1,98	2,62
25	1,71	2,06	2,79	125	1,66	1,98	2,62
30	1,70	2,04	2,75	130	1,66	1,98	2,61
35	1,69	2,03	2,72	135	1,66	1,98	2,61
40	1,68	2,02	2,70	140	1,66	1,98	2,61
45	1,68	2,01	2,69	145	1,66	1,98	2,61
50	1,68	2,01	2,68	150	1,66	1,98	2,61
55	1,67	2,00	2,67	155	1,65	1,98	2,61
60	1,67	2,00	2,66	160	1,65	1,97	2,61
65	1,67	2,00	2,65	165	1,65	1,97	2,61
70	1,67	1,99	2,65	170	1,65	1,97	2,61
75	1,67	1,99	2,64	175	1,65	1,97	2,60
80	1,66	1,99	2,64	180	1,65	1,97	2,60
85	1,66	1,99	2,63	185	1,65	1,97	2,60
90	1,66	1,99	2,63	190	1,65	1,97	2,60
95	1,66	1,99	2,63	195	1,65	1,97	2,60
100	1,66	1,98	2,63	200	1,65	1,97	2,60



## Приложение 2. Критические значения F-критерия

N <sub>1</sub>	P	N <sub>2</sub>									
		1	2	3	4	6	8	10	15	20	∞
1	0,05	161	199	215	224	234	239	242	246	248	254,3
	0,01	4051	4999	5403	5624	5859	5981	6056	6158	6209	6365,9
2	0,05	18,51	19	19,16	19,25	19,33	19,37	19,40	19,43	19,45	19,50
	0,01	98,50	99	99,17	99,25	99,33	99,75	99,80	99,87	99,90	99,50
3	0,05	10,13	9,55	9,28	9,12	8,94	8,85	8,79	8,70	8,66	8,53
	0,01	34,11	30,82	29,46	28,71	27,91	27,49	27,23	26,87	26,69	26,13
4	0,05	7,71	6,94	6,59	6,39	6,16	6,04	5,96	5,86	5,80	5,63
	0,01	21,20	18	16,69	15,98	15,21	14,80	14,55	14,20	14,02	13,46
6	0,05	5,99	5,14	4,76	4,53	4,28	4,15	4,06	3,94	3,87	3,67
	0,01	13,74	10,92	9,78	9,15	8,47	8,10	7,87	7,56	7,40	6,88
8	0,05	5,32	4,46	4,07	3,84	3,58	3,44	3,35	3,22	3,15	2,93
	0,01	11,26	8,65	7,59	7,01	6,37	6,03	5,81	5,52	5,36	4,86
10	0,05	4,96	4,1	3,71	3,48	3,22	3,07	2,98	2,84	2,77	2,54
	0,01	10,04	7,56	6,55	5,99	5,39	5,06	4,85	4,56	4,41	3,91
15	0,05	4,54	3,68	3,29	3,06	2,79	2,64	2,54	2,40	2,33	2,07
	0,01	8,68	6,36	5,42	4,89	4,32	4	3,80	3,52	3,37	2,87
20	0,05	4,35	3,49	3,10	2,87	2,60	2,45	2,35	2,20	2,12	1,84
	0,01	8,10	5,85	4,94	4,43	3,87	3,56	3,37	3,09	2,94	2,42
40	0,05	4,08	3,23	2,84	2,61	2,34	2,18	2,08	1,92	1,84	1,51
	0,01	7,31	5,18	4,31	3,83	3,29	2,99	2,80	2,52	2,37	1,80
60	0,05	4	3,15	2,76	2,53	2,25	2,10	1,99	1,84	1,75	1,39
	0,01	7,08	4,98	4,13	3,65	3,12	2,82	2,63	2,35	2,20	1,60
120	0,05	3,92	3,07	2,68	2,45	2,18	2,02	1,91	1,75	1,66	1,25
	0,01	6,85	4,79	3,95	3,48	2,96	2,66	2,47	2,19	2,03	1,38
∞	0,05	3,84	3	2,60	2,37	2,10	1,94	1,83	1,67	1,57	1,17
	0,01	6,63	4,61	3,78	3,32	2,80	2,51	2,32	2,04	1,88	1,24



### Приложение 3. Оптимальные статистические критерии для сравнения групп при разных типах данных и разной структуре исследования

(по Evans S.J.W. Brit. J. Clin. Pharmac. 1983. № 15. P. 629–648 с изменениями)

Число и тип выборок	Тип данных		
	Номинативные	Ординальные	Количественные
Одна	А. Биномиальный критерий. Б. $\chi^2$ для одной выборки на полноту соответствия	Колмогорова–Смирнова для одной выборки	Неприложимо. <b>Примечание:</b> случай с одной выборкой редко встречается в медицинской литературе
Две независимых	А. $\chi^2$ . Б. Точный критерий Фишера (ТКФ)	А. Манна–Уитни U. Б. Ранговый критерий Вилкоксона	А. Критерий Стьюдента <i>t</i> . Б. Критерий Манна–Уитни U
Две связанных	Тесты Мак-Нимара и Мантеля–Хэнзеля (варианты $\chi^2$ )	А. Критерий знаков. Б. Ранговый критерий Вилкоксона для разностей рангов	<i>t</i> -Тест для связанных выборок
3 и более независимых выборок	$\chi^2$ для K независимых выборок. <b>Примечание:</b> не может быть использован, если >20% ячеек имеют ожидаемую частоту <5 или если хотя бы одна ячейка имеет ожидаемую частоту <1	Тест Краскела–Уэллиса (одномерный дисперсионный анализ, ANOVA)	Одномерный дисперсионный анализ (ANOVA) <b>Примечание:</b> 1. Неприемлемо использовать множество <i>t</i> -тестов для сравнения 3 и более групп. 2. Если найдено различие между 3 и более группами существуют критерии для локализации различия



Число и тип выборок	Тип данных		
	Номинативные	Ординальные	Количественные
3 и более связанных выборок	Тест Кокрейна Q	Двумерный дисперсионный анализ по Фридману (ANOVA)	<p>Многофакторный (для повторных измерений) дисперсионный анализ (MANOVA).</p> <p><b>Примечание:</b></p> <p>1. Неприемлемо использовать множество парных критериев для сравнения 3 и более повторных измерений.</p> <p>2. Если найдено различие между 3 и более группами, существуют критерии для локализации различия</p>
Меры корреляции (согласия)	Козэффициенты сопряженности	<p>А. Козэффициент корреляции рангов Спирмена (<math>r_s</math> или <math>r</math>).</p> <p>Б. Козэффициент корреляции рангов Кенделла (<math>T</math>).</p> <p>В. Козэффициент конкордантности Кенделла (<math>w</math>)</p>	<p>Козэффициент корреляции Пирсона (<math>r</math>)</p> <p><b>Примечание:</b></p> <p>Линейная регрессия и корреляция — это разные процедуры</p>



## Приложение 4. Оптимальные статистические критерии для сопоставления разных типов данных в медицинском исследовании

(по Evans S.J.W. Brit. J. Clin. Pharmac. 1983. № 15. P. 629–648 с изменениями)

Тип зависимой переменной (ответ, результат)	Тип независимой переменной (объяснительной, предиктора)			
	Качественный	Номинативный	Ординальный	Количественный
Качественный	$\chi^2$	$\chi^2$	Логлинейные модели	Логистическая регрессия
Номинативный	$\chi^2$	$\chi^2$	Логлинейные модели	Логистическая регрессия
Ординальный	Манна–Уитни U	Краскела–Уэллеса	Кенделла или корреляции рангов Спирмена	Кенделла или корреляции рангов Спирмена
Количественный	t-Критерий	ANOVA	Кенделла или корреляции рангов Спирмена	Регрессионный



## Приложение 5. Выбор непараметрического статистического критерия для анализа данных

(по Krauth J. Distribution-free statistics — an application-oriented approach. Amsterdam etc.: Elsevier, 1988, с изменениями)

Случай двух выборок				
Тип данных	Оценка различий			Оценка зависимостей
	Вид выборки			
	Независимые выборки	Связанные выборки	Подобранные выборки	
Интервальные	Фишера–Питмана	Рандомизации	ТКФ	Питмана
Ординальные	Вилкоксона; ранжированных категорий; Гехана для цензурированных данных	Суммы рангов	Критерий знаков	Спирмена; ранжированных категорий; цензурированных данных
Номинативные	Фишера для четырехпольной таблицы или Фишера для таблицы сопряженности 3 и более категорий	Мак Нимара для двух категорий; Лемахера для 3 и более категорий	Баукера	Фишера для четырехпольной таблицы или Фишера для таблицы сопряженности 3 и более категорий
Случай трех и более выборок, критерии на значимость различий				
Интервальные	Рандомизации	Рандомизации	Питмана–Уэлча	—
Ординальные	Краскела–Уэллиса; ранжированных категорий; Шемпера для цензурированных данных; критерий тренда; Пателя–Хоеля на взаимодействие	Суммы рангов	Фридмана	—
Номинативные	Фишера для таблицы сопряженности	Тест для таблицы сопряженности	Уэллса	—



## Приложение 6. Распределение $\chi^2$

Число степеней свободы	Двухсторонняя вероятность (P)					
	0,2	0,1	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,268
4	5,989	7,779	9,488	11,668	13,277	18,465
5	7,289	9,236	11,070	13,388	15,086	20,517
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19,679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,873	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45,315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,314	52,620



В таблице приведены величины  $\chi^2$ , соответствующие вероятности ошибки первого рода ( $P$ ) при соответствующем числе степеней свободы. Предполагается двухсторонний тест. Если величина  $\chi^2$  больше приведенной в таблице, то  $P$  меньше соответствующей величины в заголовке колонки таблицы. Пример:  $\chi^2 = 8,06$ ; при числе степеней свободы 2 вероятность  $P$  менее 0,02, но более 0,01.

### Приложение 7. Критические значения для критерия знаков при уровне значимости $P = 0,05$

N	Критическая величина		N	Критическая величина	
	Односторонний	Двухсторонний		Односторонний	Двухсторонний
5	0	–	18	5	4
6	0	0	19	5	4
7	0	0	20	5	5
8	1	0	21	6	5
9	1	1	22	6	5
10	1	1	23	7	6
11	2	1	24	7	6
12	2	2	25	7	7
13	3	2	26	8	7
14	3	2	27	8	7
15	3	3	28	9	8
16	4	3	29	9	8
17	4	4	30	10	9



## Приложение 8. Критические значения критерия знаковых рангов Вилкоксона

N	Двухсторонняя вероятность (P)			
	0,2	0,1	0,05	0,01
5	2	–	–	–
6	3	2	–	–
7	5	3	2	–
8	8	5	3	–
9	10	8	5	1
10	14	10	8	3
11	17	13	10	5
12	21	17	13	7
13	26	21	17	9
14	31	25	21	12
15	36	30	25	15
16	42	35	29	19
17	48	41	34	23
18	55	47	40	27
19	62	53	46	32
20	69	60	52	37
21	77	67	58	42
22	86	75	66	48
23	95	83	73	54
24	104	91	81	61
25	114	100	89	68

Для оценки статистики критерия (меньшей суммы рангов) сопоставляем ее с критическими значениями в таблице. В строке, соответствующей количеству ненулевых разностей  $N$ , движемся слева направо. Последняя колонка, в которой значение больше полученной суммы, дает вероятность  $P$ .



**Приложение 9. Критические значения критерия Манна-Уитни U при уровне значимости 0,05**

Односторонний тест										
Объем выборки большого размера	Объем выборки меньшего размера									
	2	3	4	5	6	7	8	9	10	11
3	-	0								
4	-	0	1							
5	0	1	2	4						
6	0	2	3	5	7					
7	0	2	4	6	8	11				
8	1	3	5	8	10	13	15			
9	1	3	6	9	12	15	18	21		
10	1	4	7	11	14	17	20	24	27	
11	1	5	8	12	16	19	23	27	31	34
12	2	5	9	13	17	21	26	30	34	38
13	2	6	10	15	19	24	28	33	37	42
Двусторонний тест										
	2	3	4	5	6	7	8	9	10	11
3	-	-								
4	-	-	0							
5	-	0	1	2						
6	-	1	2	3	5					
7	-	1	3	5	6	8				
8	0	2	4	6	8	10	13			
9	0	2	4	7	10	12	15	17		
10	0	3	5	8	11	14	17	20	23	
11	0	3	6	9	13	16	19	23	26	30
12	1	4	7	11	14	18	22	26	29	33
13	1	4	8	12	16	20	24	28	33	37



## ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

### А

- Аборт 33, 285, 415, 439
  - Акушер 50, 404
  - Алкоголь 32, 99, 132, 233, 279, 293, 380, 419
  - Альтернативная гипотеза 140, 211, 358
  - Альфа-риск 82, 122, 141
  - Анализ
    - автокорреляционный 222
    - в подгруппах 57, 355, 356
    - выживания 250, 255, 258
      - актуарный 255
      - Каплана–Майера 255
      - Кокса 259
      - множительный 312
      - пропорциональных рисков 259
    - данных повторный 454
    - ковариационный 312
    - принятия решений 411
    - соотношения затрат и выгоды 409
    - соотношения затрат и полезности 410
    - соотношения затрат и эффективности 409
    - соотношения затрат и пользы 408
    - спектральный 222
    - чувствительности к смещениям 354
  - Анемия 374, 407
  - Антропоморфизм 150
  - Антропонозы 372
  - Артериальное давление 109, 180
  - Афлатоксин 34, 315
- ### Б
- Базы данных 102
  - Байесовский подход 168
  - Бедность 389
  - Беляков В. 373
  - Беременность 278, 393
  - Берксон Дж. 190

- Бета-риск 143, 359
- Бивалирудин 361
- Биопсия 153
- Близнецы 386, 389
- Болезнь
  - базедова 153
  - естественное течение 329
  - Крейтцфельда–Якоба 51, 371
  - Минамата 227
  - номиналистическая концепция 150
  - объективистская концепция 149
  - определение 149, 174
  - природно-очаговые 371
  - тяжесть 44, 237
  - эссенциалистическая концепция 150
- Больной 80, 150, 329
- Бонферрони поправка 242
- Бэкон Ф. 39

### В

- Вакцинация 443, 416
- Валидность 404
- Вариабельность 87
  - аналитическая 91
  - внутрииндивидуальная 91
  - межиндивидуальная 91, 332
  - описание 115
- Вариационный ряд 107, 114
- Вероятность 56, 168, 170
- ВИЧ 214, 314
- Вмешательства 339
  - действенность 402
  - доступность 414
  - лечебные 96, 323
  - описание 339
  - профилактические 324
  - эффективность 209, 320
- Вмешивающийся фактор 23, 143, 277
- ВОЗ 67, 424
- Возраст 63, 74, 82, 87, 169, 260, 307



Временные ряды 70, 188, 21  
Вспышка заболеваний 191, 371

Выборки

- независимые 331
- размер 327
- связанные 331

## Г

Галлей Е. 262

Гальтон Ф. 300

Гаусс К. 116

Гемоглобин гликозилированный 173

Генеральная совокупность 78, 122

Генле Я. 27, 372

Гепарин 361

Гипноз 40, 343

Гипотеза 285

– нулевая 140, 194, 467

– проверка 139

Глюкоза крови 171

Гомеопатия 40

Госпитализация 69

График

- вороночный 402
- выживания 253
- гистограмма 113, 129
- кокскомб 134
- Лаббе 341
- лесной 398
- линейный 133
- остатков 302
- секторальный 134
- скаттерграмма 131
- стебель и листья 211
- столбиковый 113
- ящичный 133

Группы крови 385

## Д

Данные 41

- источники
  - вторичные 100
  - первичные 99

- клинические 153, 156
- параклинические 156
- подготовка для анализа 102
- публикация 455
- ретроспективный сбор 46
- физикальные 156
- формат записи 102
- хранение 454

Дедукция 167

Делинквентность 454

Демографический переход 206

Демография 243

Дети

- обучение 35, 391
- травматизм 222

Дженнер Э. 39

Диабет 153, 173

– диагноз 173

– смертность 206, 225

Диагноз 153

- клинический 155
- морфологический 157
- параклинический 157
- ранний 159
- референтный 158
- симптоматический 153
- синдромный 153
- эпидемиологический 155

Дискриминантный анализ 311

Диспансеризация 427

Дисперсионный анализ 238

Дисперсия 118, 239

Дифтерия 182

Диэтилстильбэстрол 33, 285

Добровольность 450

Доза-эффект, зависимость 32, 201, 205

Доказательная медицина 414

Долл Р. 232, 377

Донабедян А. 413

Дорожные происшествия 319

Доступ к медицинской помощи 442

Дюбо Р. 373



- З**  
 Заболеваемости переход 183  
 Заболеваемость 17  
 – экзотическая 371  
 – эндемическая 371  
 Здоровье 67, 151  
 Значимость 67  
 – клиническая 356  
 – статистическая 143, 356  
 Зоб 371  
 Зоонозы 372
- И**  
 Измерение  
 – воспроизводимость 96, 98  
 – методы 95, 98  
 – ошибка 95  
 – правильность 96  
 – сходимость 96  
 – точность 97  
 Изображения оценка 98  
 Индекс 88  
 – Кетле 87  
 – массы тела 88  
 – тяжести состояния 183  
 Индуктивизм 39  
 Иницирующий фактор 23  
 Инсульт 54, 206  
 Интеллектуальная собственность 456  
 Интенсивные показатели 68  
 Интервал  
 – доверительный 82, 122  
 – референтный 179  
 Инфекции 369  
 – карантинные 68  
 Информация 42  
 Инцидент 64  
 – кумулятивный 64  
 – плотность 66  
 Ионизирующие излучения 199  
 Иридодиагностика 163  
 Исследования  
 – аналитические 44, 50, 284  
 – выборочные 77  
 – диагностические 165, 347, 440  
 – для поиска сопутствующих заболеваний 165  
 – исходов 418  
 – когортные 45, 50, 149, 262  
 – контролируемые 48, 273, 320  
 – медицинские 16  
 – обсервационные 48  
 – одномоментные 44, 188  
 – описательные 50  
 – планирование 83, 145  
 – поисковые 42  
 – поперечные 44, 149, 188  
 – популяционные 45, 233  
 – приемных детей 387  
 – причинных факторов 45  
 – проверяющие гипотезу 43  
 – продольные 45  
 – проспективные 45, 48  
 – публикация результатов 455  
 – распространенности 51  
 – с участием человека 449  
 – семейные 387  
 – слепые 337  
 – сравнения с контролем 38, 47, 377, 380  
 – сравнительные 406  
 – исторические 420  
 – экологические 233  
 – экспериментальные 48, 318  
 Исходная группа 265  
 Исходы 345  
 Йетса поправка 195
- К**  
 Каннел В. 377  
 Каноническая переменная 312  
 Канцерогенез 391, 393  
 Карантинные инфекции  
 Качество жизни 347  
 Качество медицинской помощи 416



- Кеннон У. 378  
Кибернетика 15, 377  
Классификация 60  
– болезней 149, 151  
Клещи 371  
Клинические испытания 322  
– анализ в соответствии с  
– выполнением протокола 353  
– полученным лечением 353  
– объяснительные 362  
– перекрестные 333  
– период отмывки 340  
– прагматические 362  
– слепые 343  
– факторная структура 334  
Клинические рекомендации 414  
Клофибрат 342  
Ковариационный анализ 312  
Коинтервенция 340  
Кокрейн А. 397  
Кокрейнская библиотека 397  
Комплаенс 327, 465  
Конкордантность 386  
Конструкты 88  
Контаминация 340  
Контроль  
– исторический 335  
– одновременный 335  
Конфиденциальность 451  
Конфликт интересов 456  
Кортикостероиды 323, 404  
Косайнор-анализ 222  
Косвенный показатель 68  
Кофе 277, 278, 286, 289, 456  
Кох Р. 28, 373  
Коэффициент  
– вариации 121  
– детерминации 294  
– корреляции 294-300  
Критерии  
– Вильсона–Джаннера 436  
– включения пациентов 325  
– диагностические 180, 182  
– исключения пациентов 350  
– непараметрические 213  
– статистические  
– F-критерий 121  
– выбор 146  
– двухсторонние 197  
– знаков Вилкоксона 208  
– Манна–Уитни 208, 211  
– суммы рангов Вилкоксона 212  
– точный Фишера 193, 196  
– хи-квадрат 193  
Курение 23, 33, 67, 99, 274, 308
- Л**  
Лавуазье А. 343  
Латинский квадрат 194  
Легионеллез 192  
Лейкоз 384  
Летальность 65  
Линней К. 150  
Липопротеиды крови 45  
Логит-преобразование 309
- М**  
Маммография 367  
Маркер генетический 384  
Маскирование 343  
Медиана 108, 114  
Мезотелиома 37  
Мендель Г. 384  
Месмеризм 343  
Метаанализ 397, 399  
Миазма 13, 372  
Мигрирующая эритема 220  
Минимизация затрат 408  
Множественные сравнения 241  
Мода 86, 108, 113  
Моделирование 300  
Модель 300  
– детерминистская 300  
– стохастическая 300  
Мониторинг 166



- Н**
- Наблюдения 91
  - Надзор эпидемиологический 52, 182
  - Наименьших квадратов метод 301
  - Найтингейл Ф. 134, 405
  - Нарушения питания 22
  - Наследственные заболевания 226, 369, 384
  - Независимая переменная 239
  - Неравенство 390
  - Новорожденные
    - зрелость 89
    - масса тела 278, 460
  - Норма в диагностике 177
- О**
- Обзор систематический 396
  - Обобщаемость 404
  - Образ жизни 38
  - Обращаемость 62
  - Общественное здоровье 395, 423
  - Ожидаемая частота 195
  - Ожирение 207, 272
    - диагноз 87
    - лечение 29, 350
    - преваленс 63, 207
    - терапия 350
  - Ондансетрон 341
  - Операционные характеристики 162, 164, 165, 439
  - Описательная статистика 107, 137
  - Опрос 287, 391
  - Опросник 99
  - Ослепление 343
  - Оспа 39, 300
  - Остеосаркома 199
  - Отклик 326
  - Отношение 58
    - правдоподобия 169–171
    - рисков 259
  - Оценка
    - вмешательства
      - комплексная 404
      - экономическая 407
    - деятельности медицинских учреждений 413
    - параметров распределения 86, 98
  - Ошибка
    - систематическая 96, 99, 114, 190
    - случайная 96
    - стандартная 122
- П**
- Павлов И. 42
  - Пандемия 371
  - Парадигма 41
  - Параметры 86
  - Пастер Л. 372
  - Патогенез 25
  - Пациент 78, 89
  - Переносчик 228, 371
  - Перепись населения 53
  - Период
    - инаппарантный 34
    - индукции 34
    - инкубационный 370
    - латентный 34
  - Пешеходы
    - возраст 69
    - смертность 69
    - травматизм 222
  - Плацебо 340, 341
  - Пневмония 192
  - Побочные эффекты 347, 353, 365
    - канцерогенез 24, 390
    - поражения клапанов сердца 29, 55
    - рвота 278
  - Повторные измерение 296
  - Подобранные пары 146
  - Познание 40, 167
  - Показатели
    - интенсивные 68
    - наглядности 70
    - структуры 63, 68



- экстенсивные 68
- Поппер К. 40
- Популяционные обследования 45
- Популяция 16, 61, 186
- Почечная недостаточность 258, 325
- Правдоподобия отношение 169
- Преваленс 55, 61, 155, 165
- Предиктор 239, 307
- Преобразование Фурье 222
- Признаки
  - альтернативные 84
  - дихотомические 84
  - интервальные 86
  - качественные 84
  - количественные 86
  - номинативные 85
  - ординальные 85
- Причина 24
  - достаточная 27
  - компонентные 22
- Причинная связь 26, 32
- Пробит-анализ 137
- Прогноз 183
- Прогностичность диагностическая 166
- Продолжительность жизни 243, 244
- Промотор 24
- Пропорция 58
- Профилактика 442, 443
- Процентили 115, 116
- Псевдорандомизация 79
- Психосоматика (Психосоматическая болезнь 11, 19, 231)
- Р**
- Рак
  - владалища 33, 285
  - желудка 183
  - легких 23, 153, 205, 377
  - молочной железы 231, 426
  - печени 34
  - поджелудочной железы 204
  - смертность 232
  - шейки матки 431, 434
  - щитовидной железы 192, 290
- Рандомизация 336
  - блочная 337
  - кластерная 338
  - с минимизацией различий 338
  - сокрытие 337
  - стратифицированная 337
- Ранжирование 85, 178
- Распределение
  - биномиальное 125
  - кумулятивное 117, 201
  - логнормальное 124
  - нормализация 111, 121, 125
  - нормальное 116
  - Пуассона 127
  - симметрия 108
  - скошенность 111
- Регистр 54
- Регрессионный анализ 300
  - логистическая регрессия 309
  - множественный 307
  - предсказательный интервал 304
  - прогнозирование 304
- Регрессия к средней 94, 329
- Резистентность 370
- Ремиссия 329
  - спонтанная 268
- Рентгенография (Рентгенограммы 98, 175)
- Репрезентативность 77
- Референтные величины 178
- Референтный метод 95
- Риск
  - абсолютное повышение 365
  - абсолютное снижение 358
  - дополнительный 272
  - модификация 444
  - относительный 358
    - повышение 272
    - снижение 357
  - оценка 444



- Робастность 234, 312
- Роды
- кесарево сечение 63
  - преждевременные 285, 402–404
- Ротман К. 21, 206
- Ртуть (ртутьсодержащие 227)
- С**
- Санитарное движение 13, 377, 389
- Сглаживание 222
- Секулярный тренд 70
- Сердечно-сосудистые заболевания 16
- смертность 26, 35, 206, 231
  - факторы риска 37, 377, 443
- Сиденгам Т. 150
- Сила связи 32, 139, 298
- Симптом
- патогномоничный 162
  - ятротропный 156
- Синергизм 315
- Скользящая средняя 222
- Скрининг 425
- диагностические тесты 157, 165, 437
- Случаи заболевания
- спорадические 371
  - стандартное определение 80, 182
- Смертности 65, 206, 243
- плотность 66
  - сила 66
- Смертность
- детская 22, 206
  - материнская 249
  - младенческая 249
  - производственная 57
- Смещение
- Берксона 190
  - воспоминания 330
  - восприимчивости 292
  - доступности диагностики 267
  - к круглым величинам 114
  - к предполагаемому диагнозу 163
  - миграции между стадиями 184
  - ожидания 269
  - от выявления медленно развивающихся случаев 434
  - от раннего диагноза 433
  - от участия добровольцев 432
  - памяти 287
  - подбора 267
  - подозрения 269
  - популярности 267
  - публикационное 399
  - регрессии к средней 94
  - центростремительное 267
- Сноу Дж. 12, 372, 384
- События
- взаимоисключающие 60
  - независимые 61, 193
- Согласие
- измерение 161
  - индекс каппа 161
  - информированное 333, 450, 452
  - пациентов 326
- Специфичность 162
- диагностическая 162, 166
- СПИД 30, 217, 375
- Среднее квадратическое отклонение 116
- Средняя арифметическая 108, 117
- Стандартизация 71
- непрямой метод 72
  - прямой метод 74
- Стандартизованное отношение смертности 72
- Стандартная ошибка 81
- относительной величины 81
  - средней 122
- Стандартное определение случая 182
- Статистика 17
- аналитическая 138
  - медицинская 62
  - описательная 107
- Статистическая значимость 143, 298
- Статистические выводы 43, 138



Статистический анализ 137  
Статистическое наблюдение 82  
Стенокардия 329, 346  
Стратификация 279  
Структура исследования 42, 281, 385

**Т**

Таблицы  
— аналитические 102  
— сопряженности (сопряжения) 195  
— электронные 102  
Тамоксифен 324  
Точка разделения 171  
Точность  
— измерения 95  
— оценки параметров  
распределения 86  
— представления данных 119  
Трипаносомоз 371  
Туберкулез 21, 175, 265

**У**

Укачивание 111  
Участники исследования 450  
— добровольность 450  
— информирование о результатах 450  
— конфиденциальность 451

**Ф**

Файнстайн А. 184  
Фактор 22  
Факторы риска 33  
— взаимодействие 313  
— специфичность 37  
Фальсифицируемость теории 40  
Фармакоэпидемиология 369  
Фенфлурамин 29, 55  
Фертильности переход 249  
Фертильность 206, 249  
Фишер Р. 41, 193  
Флюорография 165  
Фолиевая кислота 378  
Формуляр 320

Фракасторо Дж. 372  
Франклин Б. 343  
Фремингемское исследование 377  
Фтор 346  
Функциональный класс 348

**Х**

Характеристическая кривая 176  
Хельсинкская декларация 449  
Хилл А. 29, 30, 377  
Хиропрактика 339  
Холера 12  
Холестерин 35, 186, 270, 346

**Ц**

Цензурированные случаи 250  
Цикл  
— годовой 134  
— циркадианный 222

**Ч**

Чадвик Э. 389  
Частота 57  
Чернобыльская авария 280, 290  
«Черный ящик» 14, 15  
Чжень-цзю терапия 40  
Число степеней свободы 118, 195  
ЧПЛВ 358, 365  
ЧПЛП 358, 361  
Чувствительность 143, 162  
— диагностическая 162  
— статистическая 82, 83  
— статистического критерия  
— эксперимента 196  
Чума 52, 372

**Ш**

Шансы 58, 270  
— отношение 287, 310, 360, 381, 400  
Шизофрения 11, 387  
— диагноз 388  
Шкала 136  
— аналоговая 89



- Алгар 88
- дискретная 86
- логарифмическая 214, 215
- непрерывная 86
- нормальная 137
- отношений 86
- трансформация 136

**Э**

- Экология 226
- Эпидемиологический переход 206
- Эпидемиология 150
  - аналитическая 149
  - генетическая 370, 383, 384
  - инфекционных заболеваний 369
  - молекулярная 383
  - общественное здоровье 395
  - социальная 389
  - экологическая 379
- Эпидемия 11, 190, 371
- Эпистемология 40
- Эстрогены 346
- Этидронат 346
- Этиология 24
- Этические принципы 43
- Эффект
  - здорового работника 280
  - неблагоприятный 364
  - статистическая оценка 293

**Я**

- Язвенная болезнь 19
  - инфекция 391
  - история 19, 229
  - факторы риска 231

**А**

- Absolute risk reduction. См. Риск «абсолютное снижение»
- Accuracy. См. Измерение точность
- ANCOVA. См. Дисперсионный анализ
- ANOVA. См. Дисперсионный анализ

- APACHE. См. Индекс тяжести состояния
- As treated. См. Клинические испытания «анализ в соответствии с выполнением протокола»
- Attributable fraction. См. Риск объяснимый
- Attributable proportion. См. Риск объяснимый
- Attributable risk fraction. См. Риск дополнительный

**В**

- Bias. См. Смещение
- Blinding. См. Ослепление
- Box-whisker plot. См. График Ящичный

**С**

- Case finding. См. Инфекции: карантинные
- Case-control studies. См. Исследования сравнения с контролем
- Case-fatality. См. Летальность
- Censored. См. Цензурированные случаи
- Centripetal bias. См. Смещение центростремительное
- Chances. См. Шансы
- Clinical trials. См. Клинические испытания
- Cluster randomization. См. Рандомизация кластерная
- Coherence. См. Причинная связь
- Compliance. См. Комплайенс
- Concealment. См. Рандомизация «сокрытие»
- Confidence interval. См. Интервал доверительный
- Confirmatory studies. См. Исследования, проверяющие гипотезу



Confounding. См. Вмешивающийся фактор

Cost minimization analysis.

См. Минимизация затрат

Cost-benefit analysis. См. Анализ соотношения затрат и выгоды

Cost-efficiency analysis.

См. Анализ соотношения затрат и эффективности

Cost-utility analysis. См. Анализ соотношения затрат и полезности

Cross-over design. См. Клинические испытания «анализ в соответствии с выполнением протокола»

Cross-sectional study.

См. Исследования поперечные

Cumulative incidence. См. Инцидент Кумулятивный

## D

Decision analysis. См. Анализ принятия решений

Descriptive studies. См. Исследования описательные

Diagnostic access bias. См. Смещение доступности диагностики

Disability adjusted life years, DALY.

См. Качество жизни

Disease frequency survey.

См. Исследование распространенности

## E

Effectiveness. См. Вмешательства эффективность

Efficacy. См. Вмешательства действенность

Epidemiological transition.

См. Эпидемиологический переход

EpiInfo 103

Etiological fraction. См. Риск объяснимый

Expectation bias. См. Смещение ожидания

Exploratory. См. Исследования поисковые

## F

Force of morbidity. См. Инцидент плотность

Forest plot. См. График лесной

Funnel plot. См. График вороночный

## G

Generalisability. См. Обобщаемость

## H

Hazard rate. См. Инцидент плотность

Hazard ratio. См. Отношение рисков

## I

Inception cohort. См. Исходная группа

Incidence density. См. Инцидент плотность

Incidence proportion. См. Инцидент кумулятивный

Incidence rate fraction. См. Риск дополнительный

Incidence. См. Инцидент

Inferential statistics.

См. Статистические выводы

Instantaneous incidence rate.

См. Инцидент плотность

## K

Kaplan–Meier Method. См. Анализ выживания Каплана–Майера

## L

L'Abbe plot. См. График Лаббе

Least squares method.

См. Наименьших квадратов метод

Likelihood ratio. См. Правдоподобия отношение



Longitudinal. См. Исследования продольные

## М

Masking. См. Ослепление

Matched pairs. См. Подобранные пары

Meta-analysis. См. Мета-анализ

Mortality, all causes. См. Смертность общая

## Н

NHANES

— Исследования популяционные.

См. NNH. См. ЧПЛВ

Nuisance factors.

См. Вмешивающийся фактор

## О

Odds Ratio. См. Шансы, отношение

Outcomes research. См. Исследования исходов

## Р

Per protocol. См. Клинические испытания «анализ в соответствии с выполнением протокола»

Person-time incidence rate.

См. Инцидентная плотность

Pie chart. См. График секторальный

Poll. См. Опрос

Popularity bias. См. Смещение популярности

Power. См. Чувствительность статистическая

Precision. См. Измерение «сходимость»

Prediction interval.

См. Регрессионный анализ

«предсказательный интервал»

Predictive value. См. Прогностичность диагностическая

Prevalence study. См. Исследование распространенности

Product-moment correlation.

См. Коэффициент корреляции

Product-moment method. См. Анализ выживания Каплана–Майера

Proportional hazard. См. Анализ выживания пропорциональных рисков

Protective efficacy rate. См. Риск объяснимый

Ргоху. См. Косвенный показатель

## Q

QALY, Quality adjusted life years.

См. Качество жизни

## R

Randomization. См. Рандомизация

Rate. См. Частота

Recall bias. См. Смещение воспоминания

Referral filter bias. См. Смещение фильтрации

Research structure. См. Структура исследования

Response rate. См. Отклик

Reproducibility. См. Измерение воспроизводимости

Robustness. См. Робастность

ROC-анализ.

См. Характеристическая кривая

## S

Sampling bias. См. Смещение подбора

Screening. См. Скрининг

Sensitivity analysis. См. Анализ чувствительности к смещениям

Sensitivity. См. Чувствительность диагностическая

Sham surgery. См. Ложная операция

Side effect. См. Побочные эффекты



Skewness. См. Распределение «скошенность»  
SMR. См. Стандартизованное отношение смертности  
Specificity. См. Специфичность диагностическая  
Stem and leaf plot. См. График «стебель и листья»  
Study design. См. Структура исследования  
Survey. См. Исследования поперечные  
Survival analysis. См. Анализ выживания  
Survival curve. См. График выживания

Suspicion bias. См. Смещение подозрения  
Systematic review. См. Обзор систематический

**T**

Therapeutic benefit. См. Риск абсолютное снижение  
Two-tailed test. См. Критерии статистические двухсторонние

**V**

Variable. См. Признаки



## **ПРИГЛАШЕНИЕ К СОТРУДНИЧЕСТВУ**

Издательская группа «ГЭОТАР-Медиа» приглашает к сотрудничеству авторов и редакторов медицинской литературы.

**ИЗДАТЕЛЬСТВО СПЕЦИАЛИЗИРУЕТСЯ НА ВЫПУСКЕ** учебной литературы для вузов и колледжей, атласов, руководств для врачей, переводных изданий.

По вопросам издания рукописей обращайтесь в отдел по работе с авторами.  
Тел 8 (495) 921-39-07

*Учебное издание*

**Власов Василий Викторович**

# **ЭПИДЕМИОЛОГИЯ**

3-е издание, переработанное и дополненное

Зав. редакцией *А.В. Андреева*  
Менеджер проекта *И.А. Телегина*  
Выпускающий редактор *М.Ф. Смагина*  
Корректоры *М.Ю. Никитина, Е.И. Макеева*  
Компьютерная верстка *О.А. Кравченко*  
Дизайн обложки *Д.Т. Халмурзина*  
Главный технолог *О.А. Ильина*

Подписано в печать 15.02.2021. Формат 60×90 1/16.  
Бумага офсетная. Печать офсетная. Объем 31 усл. печ. л.  
Тираж 1000 экз. Заказ № К-475

ООО Издательская группа «ГЭОТАР-Медиа».  
115035, Москва, ул. Садовническая, д. 11, стр. 12.  
Тел.: 8 (495) 921-39-07.  
E-mail: [info@geotar.ru](mailto:info@geotar.ru), <http://www.geotar.ru>.

Отпечатано в АО «ИПК «Чувашия».  
428019, г. Чебоксары, пр-т И. Яковлева, д. 13.

ISBN 978-5-9704-6189-1



9 785970 461891 >



