

Р. ФЛЕТЧЕР, С. ФЛЕТЧЕР, Э. ВАГНЕР

Клиническая эпидемиология

ОСНОВЫ ДОКАЗАТЕЛЬНОЙ МЕДИЦИНЫ

THIRD EDITION

Clinical Epidemiology

THE ESSENTIALS

Robert H. Fletcher, M.D., M.Sc.

Professor

Department of Ambulatory Care and Prevention Harvard Medical School
and Harvard Community Health Plan Department of Epidemiology, Harvard
School of Public Health Department of Medicine, Brigham and Women's
Hospital Boston, Massachusetts

Suzanne W. Fletcher, M.D., M.Sc.

Professor

Department of Ambulatory Care and Prevention Harvard Medical School
and Harvard Community Health Plan Department of Epidemiology, Harvard
School of Public Health Department of Medicine, Brigham and Women's
Hospital Boston, Massachusetts

Edward H. Wagner, M.D., M.P.H.

Director, Center for Health Studies Group Health Cooperative of Puget
Sound, Professor of Health Services University of Washington, Seattle,
Washington

Williams & Wilkins

A WAVERLY COMPANY

BALTIMORE • PHILADELPHIA • LONDON • PARIS • BANGKOK • BUENOS AIRES • HONG KONG •
MUNICH • SYDNEY • TOKYO • WROCLAW

Р. ФЛЕТЧЕР, С. ФЛЕТЧЕР, Э. ВАГНЕР

Клиническая ЭПИДЕМИОЛОГИЯ

ОСНОВЫ ДОКАЗАТЕЛЬНОЙ МЕДИЦИНЫ

**Перевод с английского под общей редакцией
канд. мед. наук С.Е. Башинского
и канд. мед. наук С.Ю. Варшавского**

Медиа Сфера

Москва 1998

УДК 616
ББК 519
Ф71

Перевод с английского:
проф. А.Д. Деев, проф. Г.С. Жуковский, канд. мед. наук С.А.Шальнова
Редакторы перевода:
доктор биол. наук Е.М.Аксель, проф. Я.В.Двойрин,
доктор биол. наук М.А.Каменская, проф. Р.Г.Оганов, М.Н.Соловова
Научный консультант:
проф. В.В.Власов

Оригинал-макет и верстка:
О.П.Пономаренко
Векторная графика:
С.М.Камышов
Дизайн обложки:
О.Б.Захарова
Корректоры:
С.М.Казинцева, Т.В.Полухина, Т.Р.Тверитнева, В.И.Федорова

Оригинал-макет подготовлен в издательстве "Медиа Сфера"
Лицензия ЛП №064159
Подписано в печать 14.08.98. Формат 60 × 90/16
Объем 22,0 усл. печ. л. Тираж 10 000. Зак. №
Отпечатано с диапозитивов издательства "Медиа Сфера"
в Gummerus Printing, Finland, FIN-40351 Jyväskylä, P.O. Box 444.
Издательство "Медиа Сфера". Москва, Дмитровское ш., 46 - 2.

Клиническая эпидемиология. Основы доказательной медицины.

Р. Флетчер, С. Флетчер, Э. Вагнер.
Пер. с англ. – М.: Медиа Сфера, 1998. – 352 с., илл.

В книге, написанной одними из основателей новой дисциплины – клинической эпидемиологии, рассматриваются основные положения этой науки и ее применение в медицинской практике: диагностике, лечении, профилактике. Большое количество клинических примеров облегчает понимание излагаемого материала. Приведен список основных терминов клинической эпидемиологии с их подробным объяснением. Для практикующих врачей, исследователей, студентов и преподавателей медицинских институтов, организаторов здравоохранения.

В книге содержатся показания к применению некоторых медицинских препаратов, их дозировки, схемы введения и сведения о побочных эффектах. Со временем эта информация может претерпевать изменения, поэтому читатель должен иметь в виду, что назначать лекарственные средства следует с использованием инструкций по применению, предоставляемых производителями упомянутых медикаментов.

ISBN 0-683-03269-0 (англ.)
ISBN 5-89084-011-8 (русск.)

© 1996 Williams & Wilkins
© Перевод на русский язык и оформление, "Медиа Сфера", 1998

СОДЕРЖАНИЕ

	Предисловие	VI
	Предисловие к русскому изданию	VIII
	Благодарности	IX
<hr/>		
Глава 1	Введение	11
Глава 2	Отклонение от нормы	32
Глава 3	Диагноз	60
Глава 4	Частота событий	98
Глава 5	Риск	121
Глава 6	Прогноз	141
Глава 7	Лечение	170
Глава 8	Профилактика	205
Глава 9	Случайность	229
Глава 10	Изучение случаев	255
Глава 11	Причина	279
Глава 12	Подводя итоги	303
<hr/>		
Приложение	Словарь терминов	330

ПРЕДИСЛОВИЕ К 3-му ИЗДАНИЮ

Со времени выхода второго издания книги в 1988 году медицина очень изменилась. Эти изменения привели к тому, что методы, которыми пользуется клиническая эпидемиология, и перспективы, которые эта наука открывает, получили общественное признание.

Государства всего мира, пытаясь обеспечить высококачественную охрану здоровья, столкнулись с нарастающими трудностями контроля за расходами на медицинскую помощь. Несоответствие между потребностью в медицинской помощи и ресурсами здравоохранения привело к пониманию необходимости использования информации о клинической эффективности различных вмешательств при выборе приоритетов.

Стало ясно, что не всякая медицинская помощь эффективна и что ее лучше всего оценивать по конечным исходам. Различия в тактике оказания медицинской помощи, которые не объясняются нуждами пациентов и не сопровождаются одинаковыми изменениями клинических исходов, поставили вопрос о том, какая помощь является оптимальной. Все это вместе привело к увеличению роли добросовестных клинических исследований и подняло на должную высоту тех, кто в состоянии выполнить такие исследования и правильно интерпретировать их результаты.

Успехи в понимании биологии болезней, особенно на молекулярном уровне, также впечатляют. Фундаментальные открытия увеличивают потребность в добросовестных исследованиях на больных. Прежде чем открытия фундаментальных наук будут признаны клинически полезными, они должны быть проверены в клинике. Таким образом фундаментальная наука и клиническая эпидемиология дополняют друг друга, не являясь взаимоисключающими или конкурирующими.

Основы медицины остаются неизменными. Как пациенты, так и врачи сталкиваются с одними и теми же вопросами о диагнозе, прогнозе и лечении и ставят прежние цели: облегчить страдания, восстановить утраченные функции и предупредить преждевременную смерть. При ответах на эти вопросы мы полагаемся на результаты основных типов клинических исследований: когортные, исследования случай–контроль, рандомизированные клинические испытания и т.п. Неопределенность, присущая всей клинической информации, даже той, которая базируется на лучших исследованиях, неизбежна.

При подготовке третьего издания мы попытались принять во внимание как радикальные перемены в медицине, так и то, что в ней незыблемо. Общая структура книги в целом сохранена, обновлены лишь примеры, поскольку появились новые заболевания (например, СПИД), представления об известных болезнях существенно углубились (например, о язвенной болезни).

Мы ориентировались на то, что наша книга – введение в клиническую эпидемиологию, и старались не углубляться в "дебри", знакомые только специалистам. Изложение сделано настолько простым, насколько позволяет тема. Тем не менее мы позволили себе углубиться в некоторые вопросы в надежде на то, что теперь читатели ожидают от книги большего, чем в период становления этой области знаний.

Данная книга может быть использована в качестве учебного пособия как клиницистами, так и начинающими исследователями, которые хотят понять, каким образом добываются и оцениваются медицинские данные. Книга будет полезна читателю независимо от уровня его клинической подготовки: будь то студент-медик или практикующий врач.

Р. Флетчер, С. Флетчер, Э. Вагнер

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

В современном обществе большинство болезней – будь то раковые заболевания или атеросклероз – хронические. Они связаны с генетическими, психологическими, социальными и природными факторами риска. Развиваются эти болезни постепенно: сначала появляются симптомы, потом – снижение функции и, наконец, наступает смерть. Врачи обычно сталкиваются с больными в тот момент, когда течение хронического заболевания приобретает драматический характер – например, хронический бронхит переходит в бронхопневмонию или течение атеросклероза осложняется развитием острого инфаркта миокарда. Однако врач может помочь больному на любом этапе – от воздействия фактора риска до наступления исхода, повлияв на естественное течение болезни.

В настоящее время стало очевидным, что исследования, посвященные диагностике, распространенности, естественному течению, лечению и профилактике болезней, основанные на эпидемиологических методах ("клиническая эпидемиология"), дают ценнейшую информацию для принятия клинических решений. Публикации результатов рандомизированных клинических испытаний, исследований случай–контроль, мета-анализов составляют существенную долю статей в наиболее влиятельных медицинских журналах во всем мире. Эти исследования дополняют традиционную "фундаментальную науку", изучающую патогенез болезней.

Термин "evidence based medicine" ("доказательная медицина") был введен для обозначения медицинской практики, основанной на данных хорошо организованных клинических исследований. Задача врача – найти эти данные (в частности, при помощи обзорных публикаций и электронных средств связи), оценить их (данная книга поможет в этом), всесторонне обдумать и применить к своему пациенту. Большинство из нас не получили этих навыков в медицинском институте и вынуждены осваивать их сейчас.

Мы рады, что "Клиническая эпидемиология" издана на русском языке. Это облегчит участие российских врачей в международном развитии доказательной медицины. Наш долг перед больными – рекомендовать им методы лечения, которые действительно приносят пользу, а не вред. Наш долг перед собой – получать интеллектуальное удовлетворение от понимания научных основ современной клинической практики.

Роберт Флетчер, Сюзан Флетчер

БЛАГОДАРНОСТИ

Мы имели удовольствие работать с удивительными людьми – основателями клинической эпидемиологии. Наши учителя – Archie Cochrane, John Cassel, Alvan Feinstein, David Sackett, Kerr White и другие – были пионерами в этой области. Эта книга отражает и то, чему мы научились у них, и наш собственный опыт.

Мы многому научились у молодых врачей, которые верили на этот путь развития здравоохранения и реализовали свою веру на практике. Эти врачи – участники различных программ, в частности International Clinical Epidemiology Network (INCLEN) и Robert Wood Johnson Clinical Scholars Program. Число этих специалистов слишком велико и перечислить здесь каждого не представляется возможным.

Мы, Роберт и Сюзан Флетчеры, особенно благодарны редакционному коллективу журнала *Annals of Internal Medicine*, где мы были главными редакторами в течение четырех лет. Статьи в рецензируемых клинических журналах – письменное свидетельство развития клинической эпидемиологии. Мы многому научились при редактировании рукописей. Мы благодарны докторам Steven Jones и Wendy Levinson – нашим хозяевам, у которых мы провели творческий отпуск в Орегоне, Tom Inui, который пригласил нас в Гарвардскую школу медицины и предоставил все возможности и ресурсы, необходимые для завершения работы над книгой.

Я, Эдвард Вагнер, благодарен за предоставленную возможность применить принципы клинической эпидемиологии при разработке и внедрении программы Group Health Cooperative of Puget Sound. Также я благодарен за поддержку своим коллегам из Университета штата Вашингтон.

Это издание отражает наш совместный опыт с того времени, когда мы вдвоем работали на факультете медицины и эпидемиологии в Университете Северной Каролины. Теперь, спустя 13 лет, мы работаем в разных местах и более чем когда-либо ценим вклад многих наших коллег, которые помогли нам лучше понять и яснее изложить клиническую эпидемиологию.

1 ВВЕДЕНИЕ

Представьте себе, что к вам обратился мужчина 51 года с жалобами на боли в грудной клетке. Прежде он чувствовал себя хорошо, но 2 нед тому назад во время подъема в гору у него появились неприятные ощущения за грудиной, прекратившиеся после 2–3 мин отдыха. С тех пор подобные ощущения повторялись несколько раз, иногда во время физической нагрузки, иногда в покое. Пациент выкуривает одну пачку сигарет в день и знает, что у него "слегка повышено" артериальное давление. Больше он ни на что не жалуется и никаких лекарств не принимает. Тем не менее он беспокоится о своем здоровье, в особенности его тревожит вопрос, нет ли у него болезни сердца? Полное физикальное исследование и регистрация ЭКГ в покое не выявили отклонений от нормы, за исключением повышения артериального давления – 150/96 мм рт. ст.

У этого пациента наверняка много вопросов к вам. "Болен ли я?" "Насколько вы уверены в этом?" "Если я болен, то какова причина болезни?" "Как болезнь повлияет на меня?" "Какие меры могут быть приняты?" "Сколько это будет стоить?"

Будучи лечащим врачом данного пациента, вы должны ответить на эти вопросы и составить соответствующий план действий. Высока ли вероятность того, что болезнь серьезна и требует немедленного лечения, а не только рекомендаций и обследования? Насколько надежны разные диагностические методы для различения таких причин болей в грудной клетке как стенокардия, спазм пищевода, миалгия, невроз и т.д.? Возможно ли, например, с помощью ЭКГ-пробы с физической нагрузкой подтвердить или исключить ишемическую болезнь сердца? Если диагностирована ишемическая болезнь сердца, то в течение какого времени она будет проявляться болевым синдромом? Существует ли опасность для жизни? Насколько вероятны осложнения – застойная сердечная недостаточность, инфаркт миокарда, атеросклероз сосудов других органов? Можно ли улучшить прогноз, устранив такие факторы риска, как курение и артериальная гипертензия? Если удастся снять боль с помощью лекарств, то понадобится ли все же операция аортокоронарного шунтирования?

На подобные вопросы врачу позволяет отвечать информация, полученная из различных источников: собственный опыт, советы коллег, медицинская литература. В принципе, прогноз для конкретного больного основывается на предшествующем опыте наблюдения аналогичных больных. От методов наблюдения и способов их анализа зависит, насколько справедливыми и, следовательно, полезными для больного окажутся сделанные заключения.

Клиническая эпидемиология

Клиническая эпидемиология (clinical epidemiology) – это наука, позволяющая осуществлять прогнозирование для каждого конкретного пациента на основании изучения клинического течения болезни в аналогичных случаях с использованием строгих научных методов изучения групп больных для обеспечения точности прогнозов. Цель клинической эпидемиологии – разработка и применение таких методов клинического наблюдения, которые дают возможность делать справедливые заключения, избегая влияния систематических и случайных ошибок. В этом заключается важнейший подход к получению информации, необходимой врачам для принятия правильных решений.

Клиническая медицина и эпидемиология

Термин "клиническая эпидемиология" произошел от названий двух "родительских" дисциплин: клинической медицины и эпидемиологии. Эта наука "клиническая", потому что стремится ответить на клинические вопросы и рекомендовать клинические решения, основанные на самых надежных фактах. Она называется "эпидемиология", поскольку многие из ее методов разработаны эпидемиологами и помощь конкретному больному здесь рассматривается в контексте большой популяции, к которой принадлежит больной.

Когда-то клиническая медицина и эпидемиология составляли одно целое [1]. Большинство основателей эпидемиологии были клиницистами. Лишь в нашем столетии две дисциплины разошлись. У каждой из них свои школы, системы подготовки специалистов, журналы и области интересов. В последнее время клиницисты и эпидемиологи стали все сильнее осознавать, что их области тесно связаны и без взаимодействия возможности каждой из них ограничены [2].

Традиционное клиническое мировоззрение

Выбор ответа на клинический вопрос определяется стоящей перед врачом задачей и его практическим опытом. Деятельность

врача – это решение проблем конкретного больного. Врачи знают всех своих больных в лицо, собирают анамнез, проводят исследования и несут персональную ответственность за каждого пациента. В результате врачи стремятся к оценке прежде всего индивидуальных особенностей каждого пациента, они с большой неохотой объединяют пациентов в группы по риску, диагнозу, методу лечения и оценивают принадлежность больного к этим группам в терминах теории вероятности.

Поскольку работа врача заключается в оказании помощи конкретным больным, клиницисты зачастую упускают из поля зрения пациентов, которые наблюдаются в других медицинских учреждениях или просто не обращаются за помощью, даже если они страдают именно той болезнью, с которой эти врачи имеют дело.

Традиционное клиническое обучение ориентировано на познание механизмов развития заболеваний на основе сведений, полученных из биохимии, анатомии, физиологии и других фундаментальных наук. Эти науки определяют научное мировоззрение студентов-медиков и основу последующих клинических исследований и публикаций. Такое образование воспитывает убеждение в том, что выяснение деталей патологического процесса у конкретного больного составляет суть медицины и, следовательно, зная механизмы заболевания, можно предсказать течение болезни и выбрать правильное лечение.

Потребность в еще одной "фундаментальной науке"

Традиционный подход в медицине "работает" при соответствующих обстоятельствах. На его основе созданы многие эффективные терапевтические средства, например вакцины, антимикробные и вазоактивные препараты, синтетические гормоны. Он оправдывает себя при коррекции нарушений кислотно-основного состояния, диагностике и лечении сдавления нервных стволов.

Однако клинические прогнозы, основанные на знании биологических механизмов болезни, следует рассматривать только как гипотезы, которые должны выдержать проверку в клинических испытаниях. Дело в том, что механизмы развития заболеваний раскрыты лишь отчасти, а на исход болезни влияют многие другие факторы (генетические, физические и социальные). Достаточно привести несколько примеров противоречий теоретическим представлениям: у больных сахарным диабетом включение в диету простых сахаров сопровождается не более серьезными нарушениями метаболизма, чем потребление сложных сахаров; некоторые антиаритмические средства сами вызывают аритмии; лекарства, улучшающие реологические свойства крови, не всегда уменьшают

частоту и степень тяжести кризов при серповидно-клеточной анемии.

Разумеется, личный опыт тоже важен для принятия клинических решений. Однако ни один врач не обладает достаточным практическим опытом, чтобы распознавать все трудноуловимые, длительно протекающие, взаимодействующие процессы, которые имеют место при большинстве хронических заболеваний (см. главу 6).

Таким образом, для врача, который хочет судить о надежности клинической информации, знания в области клинической эпидемиологии столь же необходимы, как и в области анатомии, патологии, биохимии, фармакологии. Клиническую эпидемиологию следует рассматривать как одну из фундаментальных наук, на которой зиждется здание современной медицины.

Основные положения клинической эпидемиологии

Хотя личный опыт и знание механизмов развития заболеваний безусловно важны, необходимо принимать во внимание следующее:

- в большинстве случаев диагноз, прогноз и результаты лечения для конкретного больного однозначно не определены и потому должны быть выражены через вероятности;
- эти вероятности для конкретного больного лучше всего оцениваются на основе предыдущего опыта, накопленного в отношении групп аналогичных больных;
- поскольку клинические наблюдения проводятся на свободных в своем поведении больных, и делают эти наблюдения врачи с разной квалификацией и собственным мнением, результаты могут быть подвержены систематическим ошибкам, ведущим к неверным заключениям;
- любые наблюдения, и клинические в том числе, подвержены влиянию случайности;
- чтобы избежать неверных выводов, врачи должны полагаться на исследования, основанные на строгих научных принципах, с использованием методов минимизации систематических ошибок и учета случайных ошибок.

Социальный аспект клинической эпидемиологии

Влиятельные силы современного общества ускорили признаки методов и возможностей клинической эпидемиологии. Стоимость медицинской помощи достигла такого уровня, при котором даже самые богатые группы населения не в состоянии оплатить все желаемые виды услуг. Показано, что использование новых

клинических методов совсем не обязательно сопровождается соответствующими изменениями клинических исходов; следовательно, для больного полезны далеко не все общепринятые или дорогостоящие виды лечения. Сейчас разрабатываются способы более тщательной оценки клинических данных, которыми могут пользоваться руководители здравоохранения. Сложилось единое мнение, что медицинская помощь должна основываться на результатах самих строгих исследований и оцениваться по результатам с учетом финансовых затрат, которые общество может себе позволить. Кроме того, конкретные пациенты все чаще рассматриваются в качестве составной части больших групп аналогичных больных; это помогает не только делать более точные индивидуальные прогнозы, но и выбирать наиболее целесообразный путь использования ограниченных медицинских ресурсов для оптимальной помощи возможно большему числу людей.

Основные принципы

Главная цель клинической эпидемиологии – внедрять методы клинического наблюдения и анализа данных, обеспечивающие принятие верных решений. Большинство заслуживающих доверия ответов на клинические вопросы основывается на принципах, описанных в этом разделе.

Клинические вопросы

Основные вопросы, которые ставит клиническая эпидемиология, перечислены в табл. 1.1. Это те самые вопросы, которые возникали у больного и врача в примере, приведенном в начале главы. Именно их чаще всего обсуждают между собой врачи и пациенты.

Таблица 1.1
Клинические вопросы

Предмет обсуждения	Вопрос
Отклонение от нормы	Здоров или болен пациент?
Диагноз	Насколько точны методы, использованные для диагностики заболевания?
Частота	Насколько часто встречается данное заболевание?
Риск	Какие факторы связаны с повышенным риском заболевания?
Прогноз	Каковы последствия заболевания?
Лечение	Как изменится течение заболевания при лечении?
Профилактика	Существуют ли меры предупреждения болезни у здоровых людей? Улучшается ли течение заболевания при его раннем распознавании и лечении?
Причина	Какие факторы приводят к заболеванию?
Стоимость	Каковы его патогенетические механизмы? Сколько стоит лечение данного заболевания?

Таблица 1.2
Исходы заболевания (в англоязычном варианте – пять "D")*

Смерть (Death)	Плохой исход, если смерть преждевременна
Заболевание** (Disease)	Набор симптомов, физических и лабораторных данных, отклоняющихся от нормы
Дискомфорт (Discomfort)	Такие симптомы, как боль, тошнота, одышка, зуд, шум в ушах
Инвалидизация (Disability)	Неспособность к обычной деятельности дома, на работе, во время отдыха
Неудовлетворенность (Dissatisfaction)	Эмоциональная реакция на болезнь и проводимое лечение, например тоска или гнев

* К этому списку можно еще добавить шестое "D" – финансовые затруднения (Destitution), поскольку важным следствием заболевания являются затраты средств (для самого пациента или для общества).

** Нездоровье, субъективное восприятие болезни.

Клинические исходы

Клинические явления, представляющие интерес для клинической эпидемиологии – это прежде всего исходы, имеющие наибольшее значение для больных и для медицинского персонала (табл. 1.2). Именно эти явления врачи пытаются понять, предсказать, интерпретировать и изменить при лечении больных. Важное отличие клинической эпидемиологии от других медицинских наук состоит в том, что все явления изучаются непосредственно на людях, а не на животных или элементах человеческого организма, таких как культуры тканей, клеточные мембраны, химические медиаторы, генетические последовательности нуклеиновых кислот.

Биологические явления не могут считаться эквивалентом клинических исходов, пока не получено прямого доказательства их взаимосвязи. В табл. 1.3 представлены некоторые биологические явления и клинические исходы при лечении больных ВИЧ-инфекцией. На основании наших знаний о патогенезе ВИЧ-инфекции можно предположить, что такие клинические исходы, как оппортунистические инфекции, саркома Капоши и смерть можно было бы улучшить с помощью вмешательств, препятствующих снижению числа лимфоцитов CD4⁺ и уменьшающих уровень антигена р24. Однако имеются данные, что эти маркеры не дают полного представления о прогрессировании заболевания и реакции на лечение. Наивно предполагать, что влияние вмешательства на исход заболевания может быть обусловлено исключительно воздействием на физиологические параметры, поскольку конечный результат определяется и многими другими факторами. Следовательно, клинические решения должны основываться на непосредственных данных об улучшении клинических исходов как таковых.

Таблица 1.3
Биологические явления и клинические исходы:
лечение ВИЧ-инфицированных больных

Заболевание	Вмешательство	Биологические явления	Клинические исходы
ВИЧ-инфекция	Зидовудин	Число лимфоцитов CD4+	Оппортунистические инфекции
	Дидезоксиинозин	Антигенемия p24	Качество жизни
	Дидезоксицитидин	Вирусемия	Смерть
Связь известна или предполагается?			

Количественный подход

Клиническая наука особенно убедительна, когда может обещать количественный подход при измерениях. Отчасти это обусловлено тем, что количественные результаты дают более достоверные доказательства, позволяют оценить ошибку и облегчают обмен информацией как между врачами, так и между врачом и больным. Клинические исходы, такие как смерть, болезнь или инвалидизация, можно представить в числовом выражении. Несмотря на то что качественные наблюдения в клинической медицине тоже важны, они не являются предметом рассмотрения клинической эпидемиологии.

Редко удастся точно прогнозировать тот или иной клинический исход. Скорее можно на основании результатов исследования определить вероятность того или иного исхода. Клинико-эпидемиологический подход допускает, что клинические прогнозы неопределенны, но могут быть описаны количественно в виде вероятностей: например, симптомы ишемической болезни сердца возникают у 1 из 100 мужчин среднего возраста в год; курение увеличивает в два раза риск смерти в любом возрасте; прием эстрогенов снижает в 2 раза риск переломов, обусловленных остеопорозом.

Популяции и выборки

В принципе, *популяция* (population) – это большая группа людей, проживающих в определенном географическом регионе (например, в Северной Каролине) или обладающих некоторым признаком (например, старше 65 лет). Популяция может представлять собой просто некоторую часть населения (обычно таковы популяции в эпидемиологических исследованиях причин заболеваний). Она может состоять из пациентов, госпитализированных в определенную клинику или из пациентов с определенным заболеванием (что чаще имеет место в клинических исследовани-

ях). Таким образом, можно говорить об общей популяции, госпитальной популяции или популяции пациентов с конкретным заболеванием.

Выборка (sample) – это часть популяции, полученная путем отбора. Клинические исследования обычно выполняются на выборках. Оценку характеристик популяции по практическим причинам приходится осуществлять путем оценки этих характеристик по выборке.

Систематическая ошибка

Систематическая ошибка, или *смещение* (bias) – это "систематическое (неслучайное, однонаправленное) отклонение результатов от истинных значений" [3]. Допустим, обнаружено, что препарат А действует лучше, чем препарат Б. Какого рода систематические ошибки могли привести к такому выводу, если он оказался неверным? Препарат А мог быть назначен пациентам с меньшей тяжестью заболевания; тогда результаты будут обусловлены не разной эффективностью лекарственных препаратов, а систематическим различием состояния больных в двух группах. Или же препарат А приятнее на вкус, чем Б, поэтому больные строже соблюдали схему лечения. Либо препарат А – новое, очень популярное, а Б – старое средство, поэтому исследователи и больные склонны думать, что новое лекарство непременно действует лучше. Таковы примеры возможных систематических ошибок.

Наблюдение за больными (при лечении или в исследовании) особенно подвержено систематическим ошибкам вследствие простой небрежности. Участвуя в исследовании, больные часто продолжают вести себя так, как им хочется, что подчас не отвечает условиям получения строгих научных результатов. Когда с ними пытаются провести эксперимент наподобие лабораторного, из этого часто ничего не получается. Некоторые больные отказываются участвовать, другие выбывают в процессе исследования или предпочитают сменить метод лечения. Более того, некоторые самые важные с человеческой точки зрения характеристики – эмоции, ощущение комфорта, поведение – измерить гораздо труднее, чем физические параметры, такие как артериальное давление или содержание натрия в сыворотке. К тому же клиницисты сами склонны верить в успех применяемого ими лечения (большинство больных и не захотели бы лечиться у врача, который думает иначе). Из-за этой установки, которая так важна в медицинской практике, клинические наблюдения особенно подвержены систематическим ошибкам.

Хотя существуют десятки разновидностей систематических ошибок [4], большинство из них может быть отнесено к одной из трех основных категорий (табл. 1.4).

Таблица 1.4

Систематические ошибки в клиническом наблюдении

- *Систематическая ошибка, обусловленная отбором*, возникает, когда сравниваемые группы пациентов различаются не только по изучаемому признаку, но и по другим факторам, влияющим на исход.
- *Систематическая ошибка, обусловленная измерением*, возникает, когда в сравниваемых группах больных используются разные методы измерения.
- *Систематическая ошибка, обусловленная вмешивающимися факторами*, возникает, когда один фактор связан с другим, и эффект одного искажает эффект другого.

Систематическая ошибка, обусловленная отбором (selection bias), возникает, когда сравниваемые группы пациентов различаются не только по главным изучаемым признакам, но и по другим факторам, влияющим на результат исследования. Группы пациентов часто различаются по многим параметрам – возрасту, полу, степени тяжести заболевания, сопутствующим заболеваниям, методам вмешательства. Если мы сравним данные по двум группам, которые различаются не только по специфическим интересующим нас факторам (например, метод лечения или предполагаемая причина заболевания), но и по другим признакам, от которых тоже зависит исход, то результат сравнения получится смещенным и не позволит сделать выводы о степени влияния интересующего нас фактора. В примере, приведенном выше, такая ошибка возникнет, если у пациентов, получавших препарат А, степень тяжести заболевания была меньше, чем у получавших препарат Б.

Систематическая ошибка, обусловленная измерением (measurement bias), возникает, когда в сравниваемых группах пациентов применяются неодинаковые методы оценки. К такой ошибке могло бы привести использование информации, взятой из историй болезни, для изучения риска развития тромбозов у женщин в связи с приемом пероральных контрацептивов. Предположим, что сравнивалась частота применения пероральных контрацептивов в двух группах женщин, госпитализированных по поводу флеботромбоза и по другим причинам. Легко допустить, что женщины с флеботромбозом, которые слышали о возможном влиянии эстрогенов на развитие тромбозов, скорее упомянут о приеме этих средств, чем женщины, не страдающие этим заболеванием. По тем же соображениям врачи будут более подробно расспрашивать об использовании пероральных контрацептивов именно у женщин с флеботромбозом. При таких обстоятельствах связь между при-

емом пероральных контрацептивов и развитием флеботромбоза может быть выявлена именно из-за подхода к сбору информации, а вовсе не потому, что такая связь существует в действительности.

Систематическая ошибка, обусловленная вмешивающимися факторами (confounding bias), возникает, когда два фактора взаимосвязаны ("ходят парой"), причем один из них искажает эффект другого. Это может произойти из-за систематической ошибки при отборе, под действием случайности или вследствие реально существующей взаимосвязи между факторами.

Пример. Считать ли инфекцию, вызываемую вирусом герпеса, причиной рака шейки матки? Твердо установлено, что распространенность инфицирования вирусом герпеса выше среди женщин, страдающих раком шейки матки, чем у женщин без данного заболевания. Однако и герпес, и другие инфекции, которые тоже могут быть причиной рака шейки матки, передаются половым путем. В частности, доказано, что рак шейки матки вызывается вирусом папилломы человека. Не исключено, что большая распространенность инфицирования вирусом герпеса среди больных раком шейки матки лишь косвенно связана с истинной причиной, тоже передаваемой половым путем, и является следствием более высокой половой активности (рис. 1.1). Чтобы показать, что вирус герпеса вызывает рак шейки матки вне зависимости от других факторов, необходимо выявить эффект вируса герпеса в отсутствие других факторов, связанных с повышенной половой активностью [5].

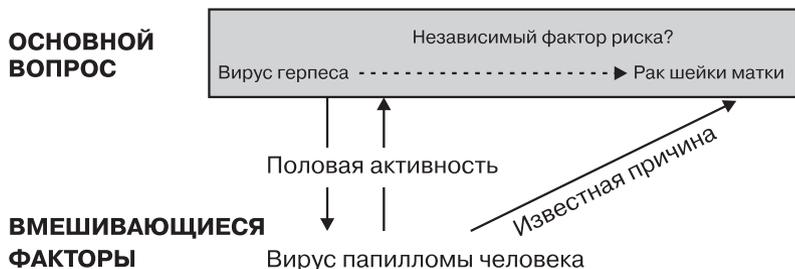


Рис. 1.1. Систематическая ошибка, обусловленная вмешивающимися факторами: может ли вирус герпеса быть причиной развития рака шейки матки? Такой вывод возможен только в том случае, если связь вируса герпеса с развитием рака шейки матки не зависит от инфицирования вирусом папилломы человека, о котором известно, что он способен вызывать это заболевание. Инфицирование обоими вирусами обусловлено повышенной половой активностью.

Систематические ошибки, возникающие при отборе и при воздействии вмешивающихся факторов, не исключают друг друга. Однако они рассматриваются по отдельности, поскольку относятся к разным этапам клинического наблюдения или исследования.

Систематическая ошибка при отборе возникает при формировании групп пациентов для наблюдения, следовательно, об этой опасности следует помнить во время планирования исследования. Ошибка из-за вмешивающихся факторов должна учитываться в процессе анализа данных после окончания исследования.

Часто в одном и том же исследовании обнаруживается несколько видов систематических ошибок, как это показано в следующем гипотетическом примере.

Пример. В исследовании была поставлена цель определить, снижает ли регулярная физическая тренировка риск развития ишемической болезни сердца (ИБС). Программа физической тренировки проводилась среди рабочих и служащих завода. Частота клинических проявлений ИБС¹ сравнивалась в группе желающих участвовать и в группе отказавшихся от участия в программе. Клинические проявления ИБС выявлялись посредством регулярных добровольных обследований, включавших тщательный сбор анамнеза, регистрацию электрокардиограммы и общую проверку состояния здоровья. В группе лиц, проводивших физические тренировки, отмечено меньшее число клинических проявлений ИБС. Однако и курящих здесь было меньше.

В этом примере систематическая ошибка при отборе может иметь место в том случае, если у лиц, желающих участвовать в программе, исходно был меньший риск развития ИБС, например в связи с низким уровнем липидов сыворотки или благодаря неотягощенному в отношении ИБС семейному анамнезу. Систематическая ошибка при измерении могла появиться из-за того, что добровольцы, регулярно проходя обследование, имели более высокий шанс выявления ИБС. Наконец, вывод о снижении риска развития ИБС под действием физических тренировок мог быть сделан вследствие систематической ошибки из-за вмешивающихся факторов: участвовавшие в программе физические тренировки добровольцы реже курили, а как известно, курение – фактор риска развития ИБС.

Сама по себе возможность систематической ошибки еще не означает, что она обязательно присутствует в конкретном исследовании. Для того чтобы и исследователи, и читатели могли успешно бороться с систематическими ошибками, необходимо прежде всего знать, где и как их искать, и что можно сделать, чтобы нивелировать их влияние. Кроме того, необходимо уметь определять, действительно ли систематическая ошибка имеет место и настолько ли она велика, чтобы клинически значимым образом повлиять на результат исследования.

¹ Клиническое проявление ИБС (coronary event) – любое клиническое проявление или неблагоприятный исход ИБС, например стенокардия, инфаркт миокарда, нарушения ритма сердца, внезапная смерть. – *Примеч. ред.*

Случайная ошибка

Заболевания обычно изучаются на выборках пациентов, а не на общей популяции (генеральной совокупности) всех лиц с рассматриваемым состоянием. Результаты наблюдений в выборке, даже если эта выборка несмещенная, могут не отражать положения в популяции в целом из-за случайной ошибки. Однако если повторять наблюдения во многих выборках таких больных, то получаемые результаты будут колебаться около истинной величины. Отклонение результата (отдельного) наблюдения в выборке от истинного значения в популяции, обусловленное исключительно случайностью, называется *случайной вариацией*.

Все мы знакомы со случайностью на примере, когда подброшенная 100 раз монета падает орлом не точно 50 раз. Подобное явление случайной вариации относится и к обсуждавшемуся примеру с оценкой эффективности препаратов А и Б. Допустим, что в исследовании, оценивающем два вида лечения, устранены все возможные систематические ошибки. Предположим также, что в действительности эти два препарата одинаково эффективны и каждый из них вызывает улучшение примерно у половины больных. Однако в единичном исследовании с небольшим числом больных в сравниваемых группах вполне может (исключительно из-за случайности) оказаться, что прием препарата А дает улучшение в большем проценте случаев, чем препарата Б, либо наоборот.

Случайная ошибка может вмешаться на любом этапе клинического наблюдения. В сравнительной оценке препаратов А и Б случайные вариации возникают при отборе пациентов, формировании групп лечения, при проведении измерений в группах.

В отличие от систематической ошибки, которая вызывает отклонение оценки от истины либо в одну, либо в другую сторону, случайная вариация с одинаковой вероятностью приводит к завышенной и к заниженной оценке. В итоге, среднее значение результатов множества несмещенных наблюдений в выборках стремится к истинному значению в популяции, даже если результаты, полученные в отдельных небольших выборках, далеки от этого.

При анализе клинических данных вероятность случайных вариаций определяется статистическими методами. Применение статистики также помогает минимизировать случайную ошибку путем выбора оптимальных методов исследования и анализа данных. Однако случайную вариацию никогда нельзя исключить полностью и следует обязательно учитывать при оценке результатов клинических наблюдений.

Соотношения между систематической и случайной ошибкой иллюстрируются на примере измерений диастолического артери-

ального давления (АД) у одного больного (рис. 1.2). Истинное значение диастолического АД, полученное при введении внутриартериальной канюли, у данного пациента составляло 80 мм рт.ст. Однако подобный метод неприменим в качестве рутинного, и в клинической практике АД обычно измеряется непрямым способом с помощью сфигмоманометра. Этот более простой инструмент допускает ошибки – отклонения от истинных значений. Ошибка заключается в том, что все показания сфигмоманометра в данном случае сдвинуты вправо от истинного значения (см. рис. 1.2). Отклонение показаний сфигмоманометра вправо (систематическая ошибка) может объясняться разными причинами: плохо калиброванный аппарат, неподходящий размер манжеты или ослабленный слух у врача. Смещение также может зависеть от выбора тона, по которому определяется диастолическое АД. Обычно это фазы IV и V тонов Короткова, которые имеют тенденцию исчезать соответственно несколько выше и ниже истинного уровня диастолического давления, а у лиц с ожирением связь между тонами Короткова и АД вообще непредсказуема. Кроме того, отдельные показания сфигмоманометра подвержены ошибкам из-за случайных вариаций, что на рис. 1.2 отражено в виде разброса показаний сфигмоманометра около среднего значения (90 мм рт.ст.).

Два источника ошибок – смещение и случайность – не исключают друг друга. Как правило, они присутствуют одновременно. Их необходимо различать, поскольку бороться с тем и другим придется по-разному.

Теоретически систематическую ошибку можно предотвратить путем правильного проведения клинических наблюдений или коррекцией при последующем анализе данных. Внимательный читатель без труда обнаружит систематическую ошибку, если она имеется. Большая часть этой книги посвящена тому, как распознать, избежать или минимизировать систематическую ошибку. В отличие от систематической ошибки, влияние случайности нельзя устранить, но можно уменьшить с помощью правильно спланированного исследования, а оставшуюся ошибку затем оценить статистическими методами. Аналогичным образом можно устранить и влияние известных систематических ошибок. Однако никакая обработка данных не в состоянии скорректировать неизвестную систематическую ошибку. Некоторые специалисты в принципе возражают против статистической обработки данных, подверженных смещению из-за плохо составленного плана исследования, поскольку это не дает ничего, кроме ложного впечатления наукообразия работы, не заслуживающей доверия.

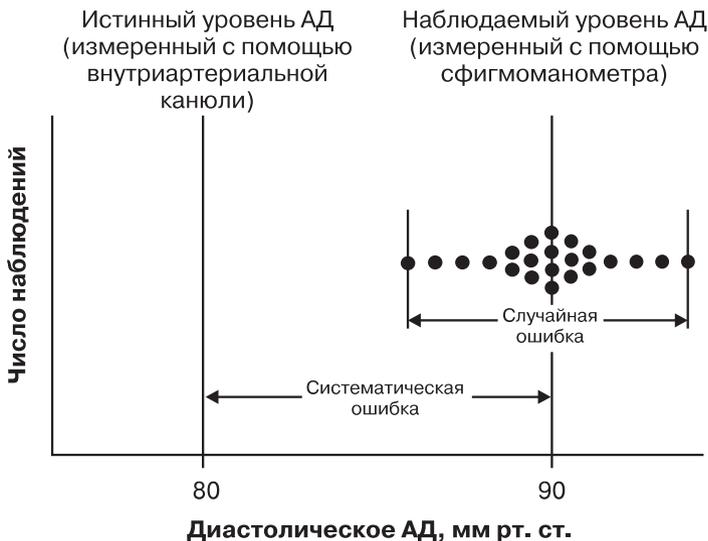


Рис. 1.2. Соотношение между систематической и случайной ошибкой: измерение артериального давления путем введения внутриартериальной канюли и с помощью сфигмоманометра.

Достоверность и обобщаемость

Когда заключения относительно популяции делаются на основании наблюдений в выборке, возникает два фундаментальных вопроса (рис. 1.3). Во-первых, справедливы ли полученные выводы для всех элементов выборки? Во-вторых, отражает ли выборка интересующую нас популяцию?

Достоверность (internal validity) исследования определяется тем, в какой мере полученные результаты справедливы в отношении данной выборки. Это внутренняя характеристика, она касается именно данной группы больных и не обязательно распространяется на другие группы. Достоверность клинического исследования определяется тем, насколько хорошо разработан план, правильно проведены сбор и анализ данных; степень достоверности определяется наличием и выраженностью систематических и случайных ошибок. Достоверность – необходимое, но не достаточное условие для того, чтобы клиническое наблюдение было полезным.

Обобщаемость (external validity, or generalizability) – внешняя характеристика, она определяется тем, в какой мере результаты данного исследования применимы к другим группам больных. Для врача это соответствует ответу на вопрос: "Если результаты

данного исследования верны, то применимы ли они к моему пациенту?" Обобщаемость отражает обоснованность допущения, что пациенты в проведенном исследовании сравнимы с другими подобными.

Результаты безусловно выполненного исследования, обладающего высокой достоверностью, могут вводить в заблуждение, если их перенести на несопоставимую группу больных.

Пример. Каков риск разрыва аневризмы брюшной аорты? Врачи, наблюдающие подобных больных, должны знать это, чтобы правильно принимать решение о необходимости хирургического вмешательства. Ответ зависит от того, о каких пациентах идет речь. Среди больных с диаметром аневризмы менее 5 см, которым обычно еще не советуют проводить операцию, частота разрыва за 5 лет наблюдения в специализированном центре оказалась в 10 раз выше, чем в общей популяции (рис. 1.4) [6]. Это может быть обусловлено тем, что в специализированные клиники больные обычно направляются уже с симптомами угрожающего разрыва. Если бы врачи общей практики при прогнозировании возможного разрыва опирались на результаты исследования больных в специализированных центрах, то риск разрыва был бы значительно завышен, что, возможно, повлекло бы за собой неверное решение о необходимости хирургического вмешательства у больных с аневризмой аорты.



Рис. 1.3. Достоверность и обобщаемость.

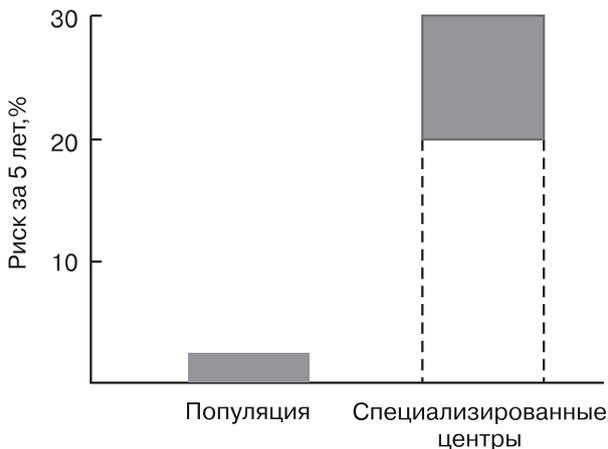


Рис. 1.4. Систематическая ошибка при отборе. Диапазон значений риска разрыва аневризмы брюшной аорты (закрашенная область) за 5 лет наблюдения больных с аневризмой (диаметром менее 5 см) в общей популяции и в специализированном центре [6].

Обобщаемость клинических исследований, даже имеющих высокую достоверность, зависит от личной точки зрения, не обязательно совпадающей у разных людей.

Пример. В Исследовании здоровья врачей (Physician's Health Study) [7] показано, что прием низких доз аспирина (по 325 мг через сутки) предупреждает развитие инфаркта миокарда у мужчин без ИБС. Среди 11 037 врачей, случайным образом отобранных и получавших аспирин, число случаев инфаркта миокарда было на 44% ниже, чем у 11 034 принимавших плацебо. Исследование проводилось тщательно, в соответствии со строгим научным планом, так что в результатах было трудно сомневаться, однако в нем участвовали только здоровые врачи-мужчины. После публикации результатов исследования клиницистам пришлось размышлять, назначать ли аспирин женщинам, лицам с множественными факторами риска и больным с диагностированной ИБС. Впоследствии анализ всех имевшихся в литературе данных показал, что аспирин эффективен и для этих групп [8].

Высокая обобщаемость редко может быть достигнута в каком-либо одном исследовании. Даже популяция, относящаяся к целому географическому региону, представляет собой смещенную выборку из еще более крупной популяции. Например, пациенты больницы – это смещенная выборка из населения района; население района – из населения города, страны и т.д. Проведение мно-

гоцентрового исследования может повысить обобщаемость, но не решит проблему окончательно.

Обычно самое большое, что может предпринять исследователь для повышения обобщаемости, это обеспечить достоверность своего исследования, подобрать популяцию, соответствующую поставленным задачам, и избегать работы с группами, значительно отличающимися от общей популяции. Дальнейшее повышение обобщаемости достигается уже за счет проведения других исследований в других условиях.

Смещенная выборка – это такая выборка, которая систематическим образом отличается от популяции, представляющей предмет исследования, или от популяции по отношению к которой должны применяться результаты исследования. Большинство клинических исследований проходит в медицинских центрах, где обычно сосредоточены больные с более тяжелой стадией заболевания; эти исследования имеют дело со смещенными выборками, в результате заболевание представляется более серьезным, чем в действительности.

Применение клинической эпидемиологии

Изучение и применение клинической эпидемиологии требуют от загруженного практической работой врача дополнительных затрат времени. Зачем ему это надо?

Понимание сильных и слабых сторон клинических данных, например в публикациях материалов исследований, дает интеллектуальное удовлетворение и ощущение уверенности вместо недоумения и разочарования. Повышается эффективность восприятия информации, поскольку врач теперь может, исходя из фундаментальных принципов, быстро разобраться, какие статьи или источники информации заслуживают доверия. Это обеспечивает достойную альтернативу тем признакам, по которым обычно судят о правоте собеседника в процессе общения с коллегами – убежденности, красноречию, профессиональному стажу или глубине специализации. Благодаря использованию принципов клинической эпидемиологии врачи разного профиля получают единую научную базу, поскольку опираются прежде всего на выводы хорошо организованных и достоверных исследований. Наконец, клиническая эпидемиология позволяет клиницисту судить о том, в какой мере его усилия в борьбе с другими факторами – биологическими, физическими, социальными, способны повлиять на исход заболевания; иными словами, врач осознает, что он может изменить и чего не может.

Исходя из перечисленного, мы считаем, что время, потраченное на изучение клинической эпидемиологии, окупится сторицей.

Информация и принятие решений

Книга посвящена прежде всего оценке качества клинической информации и ее правильному толкованию. Принятие решения – это особый вопрос. Разумеется, верные решения требуют достоверной информации; однако для них нужно нечто гораздо большее, в частности, определение цены решения, сопоставление риска и пользы.

В последние годы принятие решений в медицине стало самостоятельной дисциплиной. Эта область знаний включает в себя качественное исследование процесса принятия врачами решений, методов его оптимизации и оценку влияния на него систематических ошибок. Здесь применяются и количественные методы – анализ принятия решений и анализ эффективности затрат. Эти методы позволяют детально оценить компоненты процесса принятия решений на основе изучения различных абсолютных величин и вероятностей.

Некоторые аспекты анализа принятия решений, такие как оценка диагностических методов, вошли в эту книгу. Однако авторы не стали углубляться в эту проблему, учитывая, что клинические решения правильны в той мере, в какой верна положенная в их основу информация, а нам есть что сказать о принципах сбора и интерпретации клинических данных. Читатели, которые пожелают более глубоко изучить процесс принятия клинических решений, могут обратиться к публикациям, рекомендованным в конце этой главы.

Структура книги

Книга предназначена для врачей, которые хотят научиться определять достоверность клинических исследований, что позволит им судить о надежности собственных клинических данных или наблюдений своих коллег, а также публикаций в медицинской литературе. Мы адресуем свой труд не только тем, кто выполняет клинические исследования, но и всем использующим результаты таких исследований в своей работе. В то же время мы полагаем, что насущные потребности тех и других одинаковы.

В пособиях по клинической медицине информация о заболеваниях обычно представлена в виде ответов на традиционные вопросы: диагноз, клиническое течение, лечение и т.д. С другой стороны, в большинстве книг о клинических исследованиях за основу взята методика исследования – клинические испытания (clinical trials), обследования (surveys), исследования случай – контроль (case – control studies). Такой способ изложения рас-

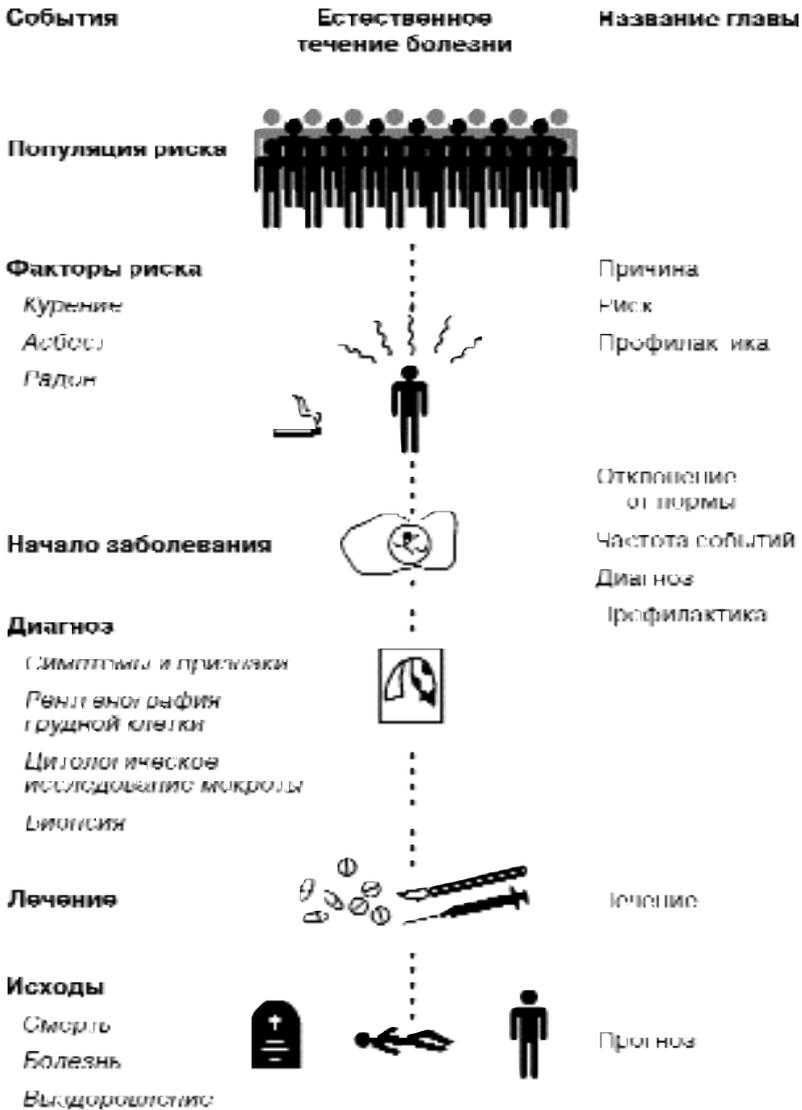


Рис. 1.5. Структура книги.

считан преимущественно на тех, кто выполняет клинические исследования, но неудобен для клиницистов.

Авторы построили свою книгу, исходя из последовательности вопросов, которые встают перед врачом в его обычной работе с больным. Так, на примере рака легкого (рис. 1.5) показано, как эти вопросы соответствуют главам книги. Вопросы охватывают всю историю естественного течения болезни, от момента, когда люди без рака легкого впервые подвергаются риску заболеть, до появления у них признаков болезни и, наконец, исхода заболевания.

В каждой главе рассматриваются типы исследований, применяемых для ответа на поставленные в этой главе клинические вопросы. Такой тип, как когортное исследование (cohort study), применим для получения ответов на различные клинические вопросы. Для удобства изложения авторы обсуждают каждый из типов исследований преимущественно в одной из глав, ссылаясь на нее в других главах, когда вновь касаются того же типа исследования.

Литература

1. White K.L. Healing the schism. Epidemiology, medicine and public health. New York: Springer-Verlag, 1991.
2. Fletcher R.H. Clinical medicine meets modern epidemiology and both profit. *Ann Epidemiol* 1992;2:325–333.
3. Murphy E.A. The logic of medicine. Baltimore: Johns Hopkins University Press, 1976.
4. Sackett D.L. Bias in analytic research. *J Chronic Dis* 1979;32:51–63.
5. Jha P.K.S., Beral V., Peto J., Hack S., Hermon C., Deacon J., Mant D., Chilvers C., Vessey M.P., Pike M.C., Muller M., Gissmann L. Antibodies to human papillomavirus and to other genital infectious agents and invasive cervical cancer risk. *Lancet* 1993;341:1116–1118.
6. Ballard D.J. Abdominal aortic aneurysm [Letter]. *N Engl J Med* 1993;329:1275.
7. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* 1989;321:129–135.
8. Antiplatelet Trialists' Collaboration. Collaborative overview of randomized trials of antiplatelet therapy—1. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. *Br Med J* 1994;308:81–106.

Рекомендуемая литература

- Andersen B. Methodological errors in medical research. Boston: Blackwell Scientific Publications, 1990.
- Eisenberg J.M. Clinical economics. A guide to the economic analysis of clinical practices. *JAMA* 1989;262:2879–2886.

Facts, figures, and fallacies series.

- Jolley T. The glitter of the t table. *Lancet* 1993; 342:27–29.
- Victoria C.G. What's the denominator? *Lancet* 1993;342:97–99.
- Grisso J.A. Making comparisons. *Lancet* 1993;342:157–160.
- Datta M. You cannot exclude the explanation you haven't considered. *Lancet* 1993;342:345–347.
- Mertens T.E. Estimating the effects of misclassification. *Lancet* 1993;342:418–421.
- Leon D. Failed or misleading adjustment for confounding. *Lancet* 1993;342:479–481.

- Sitthiamorm C., Poshachinda V. Bias. *Lancet* 1993;343:286–288.
- Glynn J.R. A question of attribution. *Lancet* 1993;342:530–532.
- Carpenter L.M. Is the study worth doing? *Lancet* 1993;343:221–223.
- Feinstein A.R. *Clinical epidemiology. The architecture of clinical research.* Philadelphia: WB Saunders, 1985.
- Feinstein A.R. *Clinimetrics.* New Haven, CT: Yale University Press, 1987.
- Friedman G.D. *Primer of epidemiology*, 4th ed. New York: McGraw-Hill, 1994.
- Gehibach S.H. *Interpreting the medical literature*, 3rd ed. New York: McGraw-Hill, 1993.
- Hennekens C.H., Buring J.E. *Epidemiology in medicine.* Boston: Little, Brown & Co, 1987.
- Hulley S.B., Cummings S.R. *Designing clinical research. An epidemiologic approach.* Baltimore: Williams & Wilkins, 1988.
- Jenicek M., Cleroux R. *Epidemiologie clinique. Clinimetrie.* St-Hyacinthe, Que., Canada: Edisem, 1985.
- Kramer M.S. *Clinical epidemiology and biostatistics. A primer for clinical investigators and decision-makers.* New York: Springer-Verlag, 1989.
- Riegelman R.K., Hirsch R.P. *Studying and study and testing a test.* 2nd Ed. Boston: Little, Brown & Co, 1989.
- Sackett D.L., Haynes R.B., Guyatt G.H., Tugwell P. *Clinical epidemiology, a basic science for clinical medicine*, 2nd ed. Boston: Little, Brown & and Co, 1991.
- Sox H.C., Blatt M.A., Higgins M.C., Marton K.I. *Medical decision making.* Stoneham, MA: Butterworth, 1988.
- Users' guide to the medical literature series*
- Guyatt G.H., Rennie D. *Users' guides to the medical literature.* *JAMA* 1993; 270:2097–2098.
- Oxman A.D., Sackett D.L., Guyatt G.H. 1. How to get started. *JAMA* 1993; 270:2093–2095.
- Guyatt G.H., Sackett D.L., Cook D.J. II. How to use, article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993;270:2598–2601.
- Guyatt G.H., Sackett D.L., Cook D.J. II. How to use, article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;271:59–63.
- Laupakis A., Wells G., Richardson S., Tugwell P. V. How to use an article about prognosis. *JAMA* 1994;272:234–237.
- Jaeschke R., Guyatt G., Sackett D.L. III. How to use, article about a diagnostic test. A. Are the results of the study valid? *JAMA* 1994;271:389–391.
- Jaeschke R., Guyatt G., Sackett D.L. III. How to use, article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA* 1994;271:703–707.
- Levine M., Walter S., Lee H., Haines T., Holbrook A., Moyer V. IV. How to use an article about harm. *JAMA* 1994; 271:1615–1619.
- Weiss N.S. *Clinical epidemiology. The study of the outcome of illness.* New York: Oxford University Press, 1986.
- White K.L. *Healing the schism. Epidemiology, medicine and the public's health.* New York: Springer-Verlag, 1991.

2

ОТКЛОНЕНИЕ ОТ НОРМЫ

Врачи затрачивают массу времени на то, чтобы отличить норму от патологии. Это нетрудно сделать при явном отклонении от нормы. Все мы знаем иллюстрации в учебнике по пропедевтике внутренних болезней, демонстрирующие резкое увеличение печени и селезенки, огромный зуб или сильную деформацию пальцев кисти при ревматоидном артрите, и не испытываем особой гордости, сумев распознать изменения такой степени. Однако гораздо чаще приходится выявлять трудноуловимые отклонения. Означает ли кратковременная боль в грудной клетке, что у пациента стенокардия? Расценивать ли мягкий систолический шум как признак заболевания клапанов сердца или это функциональное явление? Указывает ли незначительное повышение активности щелочной фосфатазы в сыворотке на заболевание печени, бессимптомную болезнь Педжета или это вообще не патология?

Сделать заключение о наличии патологии бывает особенно трудно, когда дело касается пациента из общей, неселективной популяции, как правило, вне больницы. Когда пациенты уже отобраны, например для направления в специализированный центр, в заболевании обычно уже нет сомнений. В этом случае задача состоит в том, чтобы уточнить диагноз и назначить лечение. Однако к врачу обращаются люди как с минимальными признаками патологии, так и с обычными повседневными жалобами, которые бывают у практически здорового человека. Тем не менее подобными жалобами нельзя пренебрегать. У одного пациента боли в животе могут быть обусловлены гастроэнтеритом, который пройдет самостоятельно, а у другого такие боли могут оказаться ранним проявлением аппендицита. Как правило, жалобы на боль в горле и охриплость возникают при фарингите легкой степени, но они могут быть проявлением и нередко приводящего к смерти эпиглоттита, вызванного *Haemophilus influenzae*. Эти примеры свидетельствуют о том, насколько трудно и как важно отличать одни виды патологии от других.

Главная задача при разграничении нормы и патологии заключается в том, чтобы отделить клинические случаи, требующие вмешательства, от таких, которые следует лишь принять к сведению.

териальной гипертензии, к тому же есть свои отрицательные стороны в сообщении пациенту о наличии у него болезни и в назначении лекарств. Другие сочтут диагноз оправданным, полагая, что такой уровень артериального давления связан с повышенным риском развития сердечно-сосудистых заболеваний и что лечение может уменьшить этот риск. Однако в самой общей форме представленный перечень патологических состояний служит основой для принятия решений о диагнозе, прогнозе и лечении, и эти решения должны быть приняты обязательно, путем ли назначения дополнительных диагностических тестов и лечения (активные действия) или же без каких-либо дополнительных вмешательств (пассивное наблюдение).

В этой главе описаны некоторые подходы, применяемые врачами, чтобы различать норму и патологию. Сначала мы рассмотрим, как измеряются биологические показатели, каковы их вариации и каким образом производится их обобщение. Затем можно будет обсудить использование этих данных в качестве основы для заключения о том, что называть патологией.

Типы клинических данных

Данные, полученные в результате клинических наблюдений, бывают трех типов: качественные, порядковые и количественные.

Качественные данные (nominal data) – это величины, которые нельзя расположить в естественном порядке. Примерами качественных данных служат характеристики пациента, определяющиеся малым числом генов (тканевые антигены, пол, врожденные нарушения метаболизма), или важные дискретные события (смерть, гемодиализ, хирургическое вмешательство). Такие данные могут быть легко отнесены к той или иной категории, здесь трудно ошибиться.

Качественные данные, которые могут быть отнесены только к двум категориям (присутствует – отсутствует, да – нет, жив – умер), называются дихотомическими (dichotomous data).

Порядковые данные (ordinal data) – это величины, которые могут быть расположены в естественном порядке или ранжированы, например от малого до большого или от хорошего до плохого, но размер интервала между такими категориями не может быть выражен количественно. Приведем некоторые примеры: оценка выраженности отеков нижних конечностей по шкале от + до +++++; градация сердечных шумов от 1 до 6 баллов (от прослушиваемых с трудом до слышимых без стетоскопа).

Количественные данные (interval data) – это такие величины, которым присущ естественный порядок расположения с равными

интервалами между последовательными значениями, независимо от их места на шкале. Количественные данные могут быть непрерывными и дискретными. *Непрерывные* данные могут принимать любое значение на непрерывной шкале. В качестве примеров можно привести большинство биохимических показателей крови, массу тела, кровяное давление, парциальное давление кислорода в артериальной крови.

При измерении и описании непрерывных переменных на практике можно довольствоваться некоторым количеством точек на непрерывной шкале, часто это целые числа, поскольку точность измерений ограничена и на практике дальнейшая детализация не имеет смысла. Например, уровень глюкозы крови может фактически составлять 193,2846573 мг/дл, но обычно регистрируется как 193 мг/дл. *Дискретные* данные всегда выражаются целыми числами. Примеры дискретных данных – число беременностей и число рождений живых младенцев; число судорожных припадков у больного в месяц.

При рассмотрении порядковых и количественных данных возникает вопрос: где кончается норма и где начинается патология? Например, начиная с какой величины размеры предстательной железы перестают считаться нормальными? Врач может выбрать любую точку разделения. На чем может основываться такой выбор, будет рассмотрено в этой главе.

Характеристики данных

Независимо от того, к какому из перечисленных типов относятся данные, они могут быть охарактеризованы различными способами, описанными ниже.

Достоверность измерений (validity)

Как указывалось в главе 1, достоверность показывает, в какой степени полученные данные соответствуют истинным значениям, т.е. действительно отражают то явление, оценка которого производится. Иногда достоверность называют точностью оценки (accuracy).

Для клинических параметров, поддающихся непосредственной оценке, достоверность оценивается относительно просто. Наблюдаемый результат измерения сравнивается с некоторым принятым стандартом. Например, прибор для определения уровня натрия в сыворотке постоянно калибруют с помощью раствора натрия известной концентрации. Достоверность лабораторных измерений обычно подвергается строгим повторным проверкам. Так, в США действует национальный стандарт оценки concentra-

ции глюкозы в крови, согласно которому лаборант производит контроль точности измерения в сравнении с высоким и низким стандартом в начале каждого рабочего дня, а также после любого изменения процедуры измерений (например, перехода к новому флакону с реактивом или замены батарейки в приборе). Аналогичным образом, достоверность результатов физикального исследования оценивается по результатам хирургического вмешательства или вскрытия.

Другие клинические симптомы, такие как боль, тошнота, одышка, депрессия или страх невозможно проверить физическими методами. Информация о такого рода явлениях получают путем сбора анамнеза заболевания. В клинических исследованиях используются более формальные и стандартизированные подходы в виде опросников или анкет.

Существуют три способа обеспечения достоверности данных:

- метод измерения должен охватывать все аспекты изучаемого явления и ничего кроме них (например, опросник по оценке боли должен включать в себя такие характеристики, как тупая, пульсирующая, жгучая, жалящая боль, но не вопросы относительно кровяного давления, зуда, тошноты, звона в ушах и т.п.);

- метод измерения должен отражать, в какой мере результат отдельного теста согласуется с другими оценками того же явления (например, интенсивность боли, определенная по опроснику, должна соответствовать другим проявлениям этой боли, таким как испарина, стоны, судороги, просьба дать обезболивающие лекарства);

- метод измерения должен показывать, насколько предсказуемы были данные, полученные в ходе измерений (например, можно проверить, соответствуют ли ответы на опросник по оценке боли явлениям, вызывающим боль известной интенсивности: слабая боль от минимальной царапины, умеренная боль при обычной головной боли или пептической язве, сильная боль при почечной колике).

О достоверности нельзя сказать, что она есть или что ее нет, как это часто думают. Напротив, с помощью описанных подходов можно выдвигать аргументы за или против достоверности той или иной шкалы измерений в зависимости от условий ее применения, чтобы убедить собеседника в том, что эта шкала более или менее достоверна.

В силу особенностей профессионального отбора и полученного образования врачи предпочитают использовать точные методы измерений, предлагаемые естественными – в том числе био-

логическими – науками, избегая или игнорируя другие подходы, особенно в исследовательской работе. Между тем ослабление выраженности симптомов, повышение удовлетворенности и ощущение благополучия – одни из важнейших результатов медицинской помощи с точки зрения как пациента, так и врача. В клинических исследованиях должны изучаться именно эти исходы, помогающие выработать правильное клиническое решение.

По этому поводу Файнштейн писал следующее [1]:

"Понятие "строгие" обычно относится к данным, которые имеют точное количественное выражение (т.е. лабораторные данные, демографические показатели, финансовые затраты); в то же время клинические проявления, ощущения комфорта, ожидания, обстановка в семье – это "неопределенные", субъективные данные. Они зависят от личных впечатлений наблюдателя и наблюдаемого и обычно выражаются не цифрами, а словами.

Для того чтобы избежать использования таких неопределенных данных, результаты лечения представляются в виде лабораторных показателей, которые могут быть объективными, количественными, надежными однако при всем этом они лишены "человеческого лица". Когда нам говорят, что уровень холестерина сыворотки 230 мг/дл, что рентгенологическое исследование выявляет кардиомегалию или что на электрокардиограмме имеются зубцы Q, мы не знаем, к какому объекту это относится – человеку или, допустим, собаке. Если же нам сказано, что восстановилась работоспособность, что схема лечения легко выполнима и лекарство приятно на вкус, что семья довольна результатами, мы сразу понимаем: речь идет о человеке".

Воспроизводимость (reliability)

Под воспроизводимостью понимается вероятность того, что при повторных измерениях некоего устойчивого явления, сделанных разными людьми, на разных приборах, в разное время и в разных местах, будет получен один и тот же результат.

Воспроизводимость лабораторных показателей устанавливается путем повторных измерений, например, на одном и том же образце сыворотки или ткани, производимых разными специалистами или на разных приборах. Воспроизводимость находок подтверждается, если их одинаково описывают разные наблюдатели при различных обстоятельствах.

На рис. 2.1 показано соотношение между воспроизводимостью и достоверностью. Инструмент измерений (лабораторный прибор или опросник) может в среднем обладать высокой достоверностью (точностью), но иметь низкую воспроизводимость, т. е. давать широкий разброс получаемых результатов вокруг истинного значения. С другой стороны, иногда инструмент имеет высокую воспроизводимость, но дает систематическое отклонение показа-

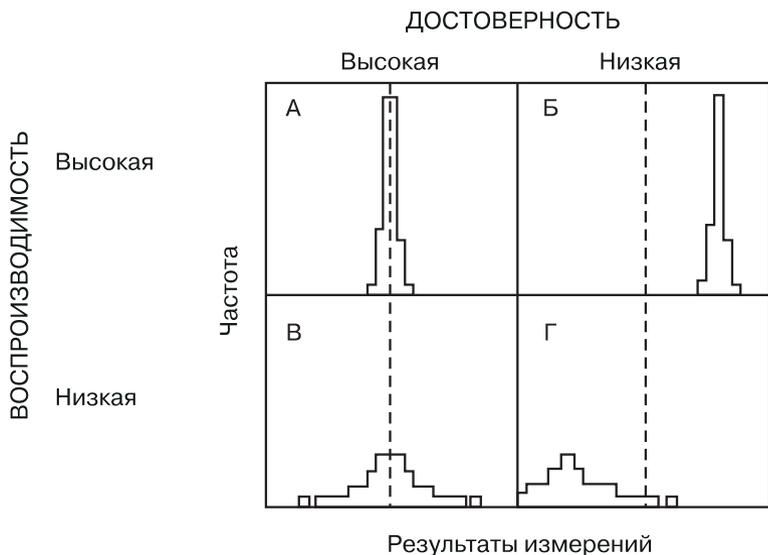


Рис. 2.1. Достоверность и воспроизводимость. А. Высокая достоверность и высокая воспроизводимость. Б. Низкая достоверность и высокая воспроизводимость. В. Высокая достоверность и низкая воспроизводимость. Г. Низкая достоверность и низкая воспроизводимость. Прерывистые линии показывают истинные значения.

ний (неточен). Однократное измерение, сделанное на приборе с низкой воспроизводимостью, имеет и низкую достоверность, потому что высока вероятность ошибочного результата просто в силу случайности.

Диапазон измеряемых значений

Инструмент может не позволять регистрировать очень низкие или очень высокие значения измеряемого параметра, ограничивая получаемую исследователем информацию. Так, поначалу метод измерения в сыворотке уровня тиреотропного гормона (ТТГ) не был пригоден для диагностики гипертиреоза и титрования дозы тироксина, поскольку не давал возможности определять низкие уровни ТТГ. Аналогично опросник "Шкала повседневной жизненной активности" (с помощью которой оценивают способность пациента принимать пищу, и регулировать отправление естественных потребностей: передвигаться, пользоваться туалетом, одеваться и умываться) не учитывает способности читать, писать или играть на фортепьяно – виды деятельности, которые могут быть очень важны для некоторых пациентов.

Способность реагировать на изменения измеряемого параметра (responsiveness)

Способность улавливать изменения измеряемого параметра – важная характеристика инструмента. Например, шкала сердечной недостаточности Нью-Йоркской ассоциации кардиологов, имеющая классы тяжести от I до IV (отсутствие симптомов, симптомы при слабой и умеренной нагрузке, симптомы в покое), не позволяет выявить незначительные изменения степени сердечной недостаточности, которые ощущают сами больные, тогда как измерение фракции выброса дает возможность определить даже незначительные изменения, не замечаемые больными.

Интерпретируемость (interpretability)

Недостаток оценочных шкал, используемых в опросниках, заключается в том, что получаемые оценки, в отличие от результатов физикального исследования, не всегда понятны врачам и пациентам. Например, хорошо или плохо иметь 72 балла по шкале депрессии Занга? Для того чтобы преодолеть этот недостаток, исследователи "привязывают" значения шкалы к знакомым явлениям: указывают, что балл ниже 50 рассматривается как норма, а от 70 и выше – как состояние тяжелой или крайне тяжелой депрессии, требующее немедленной медицинской помощи.

Вариация (variation)

Результаты клинических измерений одного и того же явления могут быть распределены в некотором диапазоне значений, в зависимости от условий проведения этих измерений. Для того чтобы избежать ошибочных выводов из таких результатов, врач должен учитывать возможные причины их вариации в конкретной ситуации и знать вклад каждой из этих причин в общую вариацию.

Общая вариация – это сумма вариаций, связанных с процессом измерения, биологическими изменениями, происходящими в организме в течение времени, и биологическими различиями между разными индивидуумами (табл. 2.2).

Таблица 2.2
Источники вариации

Источник	Определение
<i>Метод измерения</i>	
Инструмент	Средства измерения
Наблюдатель	Лица, производящие измерения
<i>Биологические причины</i>	
В пределах одного индивидуума	Изменения в организме в зависимости от времени и внешних условий
Между индивидуумами	Биологические различия между индивидуумами

Вариация измерений

Любые наблюдения подвержены вариациям, обусловленным работой приборов и действиями людей, производящих измерения. Условия измерения могут привести к смещенному результату (низкая достоверность) или к случайной ошибке (низкая воспроизводимость). Эти источники вариаций можно свести к минимуму, производя измерения тщательно и следуя стандартному протоколу. Однако, если результаты измерений оцениваются человеком, а не машиной, вариация может стать очень существенной и ее будет трудно контролировать.

Пример. Частота сердечных сокращений (ЧСС) у плода часто оценивается с помощью аускультации, которая подвержена ошибке. Электронный прибор показывает истинную ЧСС. Необычно высокая или низкая ЧСС указывает на гипоксию плода и, возможно, на необходимость более раннего родоразрешения.

Day B. et al. [2] сравнивали ЧСС плода, оцененную с помощью аускультации и при электронном мониторинге (рис. 2.2). Когда истинная частота сокращений находилась в пределах нормы, ее значения при аускультации равномерно распределялись около истинного значения, т.е. присутствовала только случайная ошибка. Однако при патологически высокой или низкой истинной ЧСС ее значения при аускультации оказывались ближе к норме. При низкой ЧСС имела тенденция к завышению, а при высокой – к занижению истинного значения ЧСС.

Это исследование иллюстрирует как случайную, так и систематическую ошибку в клинических наблюдениях. В данном случае смещение в сторону нормальных значений могло возникнуть из-за склонности медицинского персонала надеяться на то, что с плодом все в порядке и незачем предпринимать активное вмешательство, основываясь только на данных о слишком низкой или высокой ЧСС.

Вариации, обусловленные измерением, возникают также потому, что измерения проводятся на выборке, которая может быть неадекватна общей популяции. Часто включенная в выборку доля общей популяции очень невелика. Например, образец ткани, полученный при биопсии печени, соответствует только 1/100 000 части всей печени. При исследовании такой малой части целого вероятность отличия одной пробы (выборки) от другой весьма существенна.

Если измерения выполняются в разных условиях (разные лаборатории, персонал или методы), то некоторые результаты могут быть невоспроизводимы и/или систематически отличаться от истинных значений, способствуя разбросу получаемых данных.

Биологическая вариация

Вариация возникает также вследствие биологических изменений в организме, которые происходят с течением времени. Большинство биологических параметров изменяется во времени. Измерение, проведенное в определенный момент времени, представляет собой выборку из всех возможных измерений в течение данного периода и может давать результат, не соответствующий обычному значению показателя.

Пример. Оценка частоты желудочковых экстрасистол помогает решить вопрос о необходимости лечения и судить о его эффективности. По практическим соображениям врачи часто производят эту оценку при относительно коротком наблюдении, например, опреде-

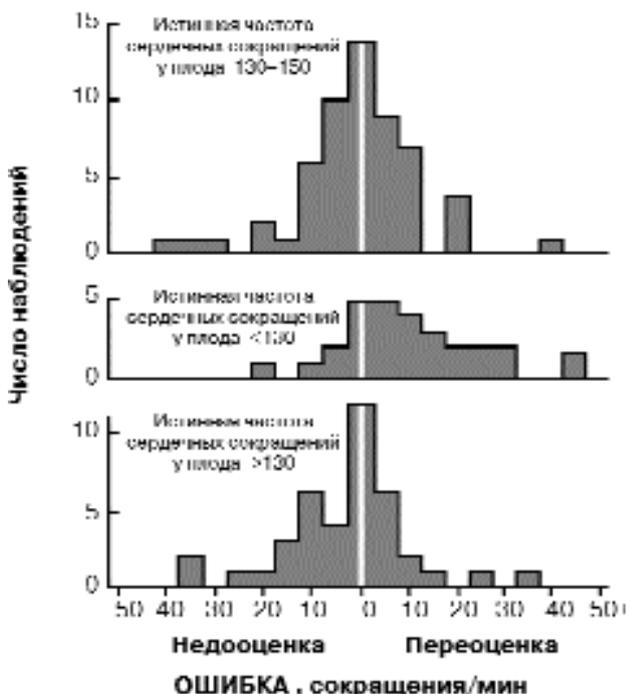


Рис. 2.2. Вариация результатов, связанная с наблюдателем. В зависимости от истинной ЧСС у плода, установленной с помощью электронного монитора, оценка ЧСС при аускультации оказывается в пределах нормы, заниженной или завышенной.

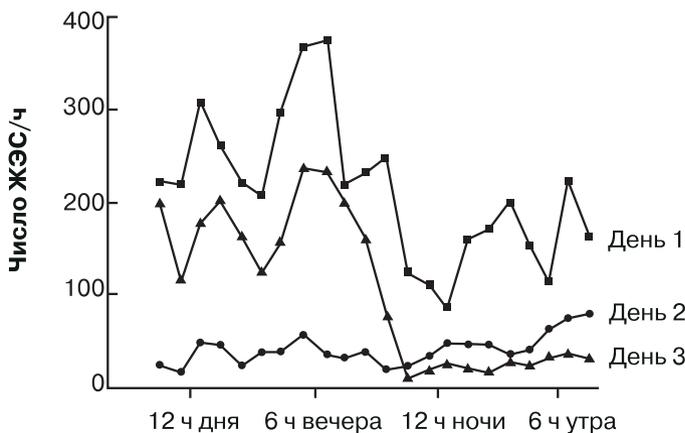


Рис. 2.3. Биологическая вариация. Число желудочковых экстрасистол (ЖЭС) у нелеченого больного при холтеровском мониторировании в течение трех последовательных суток.

ляя пульс за 1 мин или просматривая ЭКГ (запись в течение 10 с). Однако у одного и того же больного частота экстрасистол варьирует в течение времени. Для того чтобы получить более представительную выборку частоты экстрасистол, иногда используют холтеровское мониторирование ЭКГ. Тем не менее даже продолжительный мониторинг может давать ложные результаты. На рис. 2.3 показан типичный пример результатов мониторирования у пациента с желудочковыми экстрасистолами [3]. В течение 3 сут частота экстрасистол варьировала от менее чем 20 в час до 380 в час в разные дни и в разное время суток. Авторы сделали вывод: "Чтобы при сравнении только двух мониторинговых записей ЭКГ, сделанных в течение 24 ч, отличить снижение частоты желудочковых экстрасистол, обусловленное лечением, от биологических или спонтанных вариаций, снижение частоты должно превышать 83%".

Вариация возникает также из-за имеющихся индивидуальных различий между людьми. Биологические различия играют существенную роль во многих ситуациях. Например, в нескольких исследованиях показано, что регистрируемый при однократном случайном измерении высокий уровень артериального давления, хотя и может быть результатом вариации, однако все же связан с последующим развитием сердечно-сосудистого заболевания.

Общая вариация

Эффекты различных источников вариации накапливаются. На рис. 2.4 это продемонстрировано для результатов измерения

артериального давления. Вариация, обусловленная измерением, здесь относительно невелика, хотя диапазон различий оценки давления разными наблюдателями достигает 12 мм рт.ст. С другой стороны, у каждого пациента артериальное давление сильно меняется в течение дня, так что результат, полученный при любом однократном измерении может не отражать обычный его уровень. Эта вариация носит по большей части неслучайный характер: давление повышается в период бодрствования, при возбуждении, во

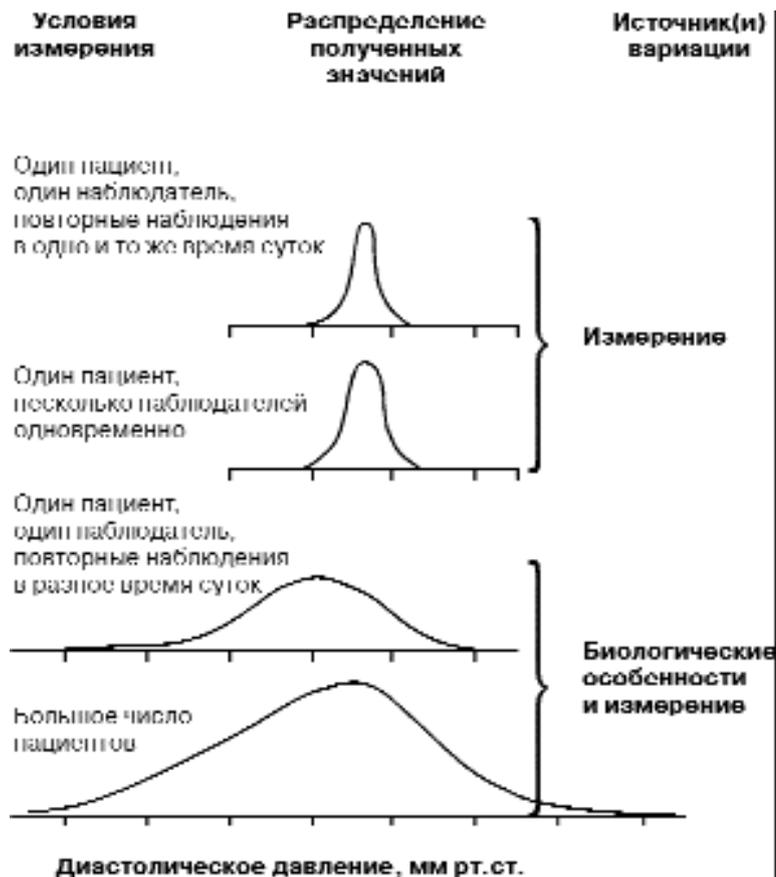


Рис. 2.4. Источники вариации. Измерение диастолического (фаза V) артериального давления [Fletcher R.H., Fletcher S.W., Boe J. et al. The blood pssessure in a population (Special issue). Acta med Scand 1957; 321:5 – 313].

время посещения врача. Конечно, в первую очередь нас интересует, как значения артериального давления у данного пациента соотносятся с уровнем давления у лиц, составляющих контрольную группу, особенно, если речь идет об осложнениях артериальной гипертензии или оценке эффективности лечения.

Влияние вариации на результат измерений

Другая особенность вариации касается ее влияния на достоверность и воспроизводимость измерений и того, каким образом это влияние можно контролировать.

Случайные вариации, например из-за нестабильности работы прибора или участия разных наблюдателей, приводят к разнонаправленным смещениям, которые имеют тенденцию нивелировать друг друга; в итоге, если произведена серия измерений, то истинное значение измеряемого параметра не искажается. Однако индивидуальные измерения могут быть ошибочны. Неточность, обусловленную случайной вариацией, можно уменьшить, увеличив выборку: подсчитав большее число клеток в мазке крови, увеличив число просматриваемых полей при исследовании осадка мочи, обследуя большее число пациентов. Кроме того, случайную вариацию можно оценить статистическими методами (см. главу 9).

С другой стороны, смещенные результаты систематически отличаются от истинных значений, независимо от числа повторных измерений. Допустим, что мы обнаружили у больного повышение активности щелочной фосфатазы в сыворотке и подозреваем у него наличие объемного образования в печени. В зависимости от распространенности патологического процесса, мы можем получить при биопсии результат, который приведет нас к ошибочному заключению о том, что заболевание отсутствует. Так, если у больного имеется метастаз в левой доле печени, то при выполнении биопсии в стандартном месте (правая доля) поражение будет обязательно пропущено. В то же время, при милиарном туберкулезе, когда вся печень поражена множеством мелких гранулем, биопсия может дать отрицательный результат только в силу случайной вариации. Еще один пример случайной вариации был представлен ранее (см. рис. 2.3, стр. 42): все высокие значения частоты желудочковых экстрасистол были зарегистрированы в 1-е сутки, а низкие – в основном на 3-и сутки. Смещенные оценки частоты экстрасистол в разные дни объясняются случайными вариациями показателей.

Распределения случайных величин

Количественные данные часто представляются в виде графика, называемого частотным распределением и показывающего

число (или долю) лиц в определенной группе с разными величинами измеряемого показателя (рис. 2.5). Такая форма представления данных обеспечивает достаточно подробную информацию.

Таблица 2.3
Способы представления центральной тенденции и дисперсии

Способ представления	Определение	Преимущества	Недостатки
Центральная тенденция			
Среднее	Сумма полученных значений, деленная на число наблюдений	Удобна для математической обработки	Подвержена влиянию крайних значений
Медиана	Точка, по обе стороны которой находится равное число наблюдений	Не подвержена влиянию крайних значений	Не слишком удобна для математической обработки
Мода	Наиболее часто встречающееся значение	Проста для понимания	Иногда мода отсутствует, а иногда их бывает несколько
Характеристики рассеяния			
Диапазон значений	Рассеяние от минимального до максимального значения в распределении	Охватывает все значения	Сильно подвержен влиянию крайних значений
Стандартное отклонение*	Абсолютная величина среднего арифметического отклонения индивидуальных значений от среднего значения	Удобно для математической обработки	Для негауссова (ненормального) распределения не описывает известную долю наблюдений (не всегда пригодно для описания негауссовых распределений)
Процентиля, децили, квартили и т.д.	Доля наблюдений, находящихся в определенной части распределения	Характеризуют "необычность" значения независимо от типа распределения	Неудобны для статистических расчетов

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

где x_i – каждое единичное наблюдение, \bar{x} – среднее из всех наблюдений, n – число наблюдений.

Описание распределений

Для описания распределений случайных величин используются два основных свойства распределения: центральная тенденция (середины распределения) и рассеяние. Некоторые способы выражения центральной тенденции и рассеяния, а также их преимущества и недостатки представлены в табл. 2.3 и на рис. 2.5.

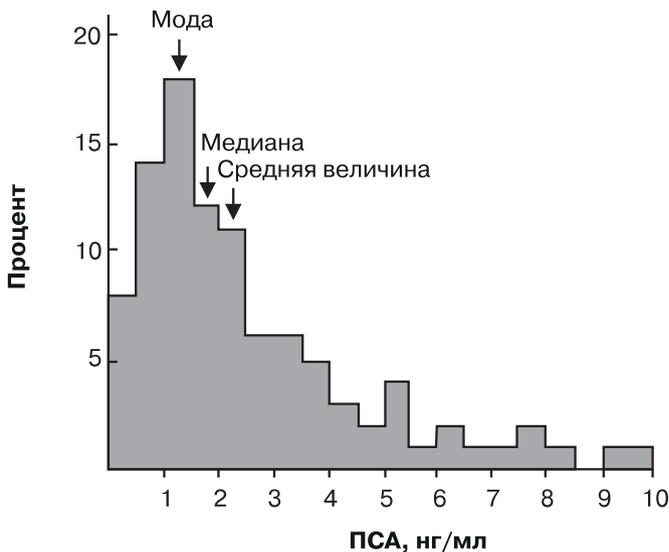


Рис. 2.5. Оценки центральной тенденции и рассеяния. Распределение уровней простатспецифического антигена (ПСА) у предположительно здоровых мужчин [Kane R.A., Littrup P.J., Babaian K. e.a. Prostate-specific antigen levels in 1695 men without evidence of prostate cancer. *Cancer* 1992;69: 1201–1207].

Фактические распределения

На рис. 2.6 показаны частотные распределения для четырех рутинных тестов показателей крови (калий, щелочная фосфатаза, глюкоза и гемоглобин). Большинство значений группируется ближе к середине, так что кроме центральной части на кривой нет "горбов" или неровностей. Участки кривой, которые соответствуют высоким и низким значениям показателей, вытягиваются в виде хвостов, причем с одной стороны "хвост" часто бывает длиннее, чем с другой (т.е. кривые в одной части имеют менее пологую форму). Некоторые кривые имеют менее пологую форму в области

высоких значений, другие – в области низких. Другими словами, все эти распределения унимодальны (имеют один пик), по форме колоколообразны и не обязательно симметричны; во всех других отношениях они не похожи друг на друга.

Распределение значений многих лабораторных показателей зависит от таких характеристик пациентов, как возраст, пол, раса, характер питания. На рис. 2.7 показано, как изменяется с возрастом распределение одного из таких показателей – концентрации азота мочевины в крови. Содержание азота мочевины, равное 25 мг/дл, считается слишком высоким для лиц молодого возраста, но является практически нормальным для пожилых.

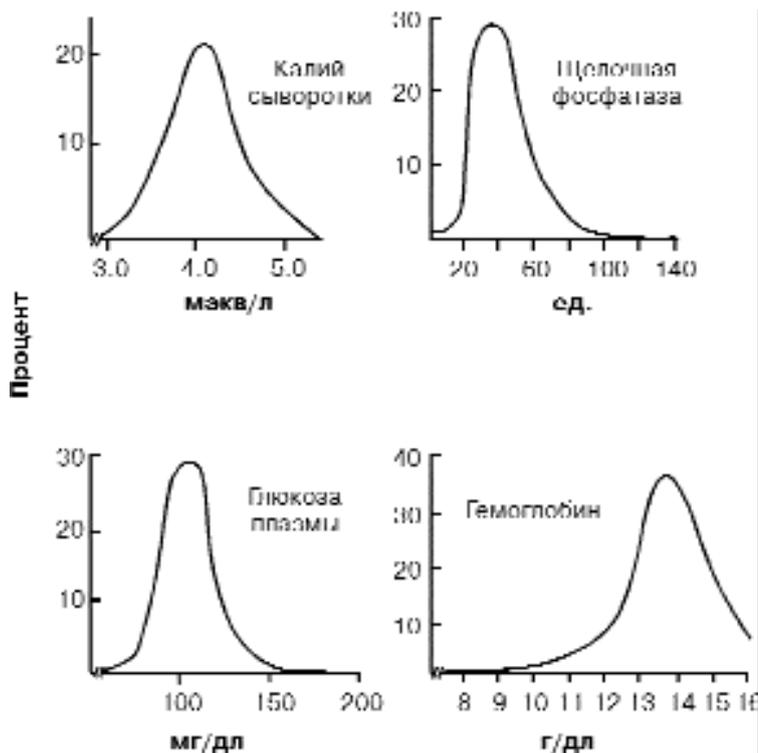


Рис. 2.6. Фактические распределения в клинической медицине [Martin H.F., Gudzinowicz B.J., Fanger H. Normal values in clinical chemistry. New York: Marcel Dekker, 1975].

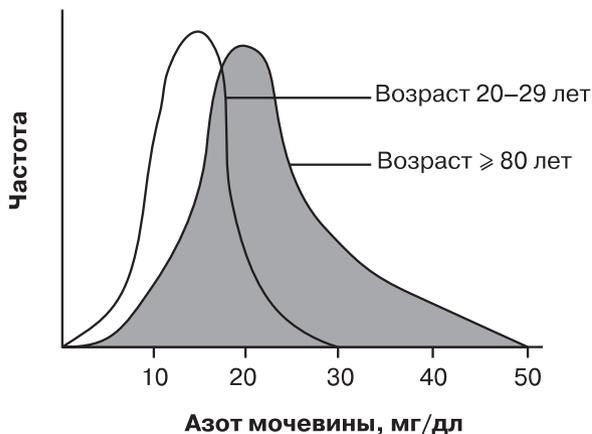


Рис. 2.7. Распределения переменных в клинической медицине в зависимости от возраста: азот мочевины у людей в возрасте 20 – 29 лет и старше 80 лет [Martin H.F., Gudzinowicz B.J., Fanger H. Normal values in clinical chemistry. New York: Marcel Dekker, 1975].

Нормальное распределение

Другой тип распределения, называемый нормальным, или гауссовым, иногда используется для аппроксимации действительно существующего распределения значений показателя, несмотря на то что нормальное распределение основывается на статистической теории и необязательно отражает естественное распределение. Кривая нормального распределения описывает частотное распределение результатов повторных измерений одного и того же физического объекта полученных одним и тем же методом. Разброс значений отражает только случайную вариацию. Нормальная кривая показана на рис. 2.8. Она симметрична и имеет форму колокола. Ее математическое свойство таково, что примерно $\frac{2}{3}$ результатов измерений лежат в пределах одного стандартного отклонения и около 95% – в пределах двух стандартных отклонений.

Хотя распределения, с которыми имеет дело клиническая медицина, часто напоминают нормальные, это сходство поверхностное. Как писал один специалист по статистике [4]:

”Согласно экспериментальным данным, для большинства физиологических переменных частотные распределения представляют собой гладкие унимодальные асимметричные кривые, причем пло-

щадь под кривой, ограниченная средней величиной ± 2 стандартных отклонения, не включает желаемых 95% значений. Математических, статистических или каких-либо других теорем, которые позволяли бы нам предсказать форму распределения результатов физиологических измерений, не существует."

В то время как нормальное распределение основано на математической теории и отражает только случайные вариации, свой вклад в распределение результатов клинических измерений вносит множество других источников вариаций, особенно биологические различия между индивидуумами. Следовательно, сходство реальных распределений с кривой нормального распределения в клинической медицине, как правило, случайно. Несмотря на это, ради удобства (поскольку средние величины и стандартные отклонения относительно просто рассчитываются) часто делается допущение, что результаты клинических измерений распределены "нормально".

Критерии отклонения от нормы

Согласитесь, было бы чрезвычайно удобно, если бы частотные распределения одних и тех же признаков были настолько разными у здоровых и у больных людей, что это позволяло бы нам

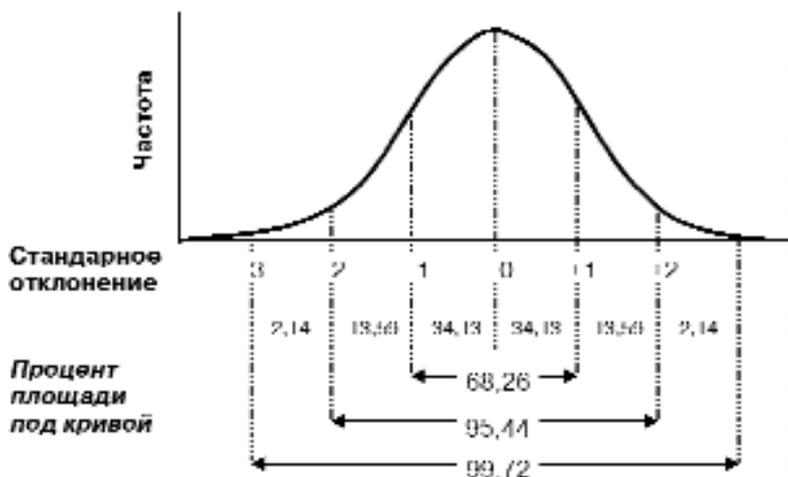


Рис. 2.8. Нормальное (гауссово) распределение.

уверенно различать такие популяции. Подобная ситуация имеет место, например, для специфических ДНК- и РНК-последовательностей и для антигенов (рис. 2.9,А), которые могут либо присутствовать, либо отсутствовать, не давая выраженных клинических проявлений.

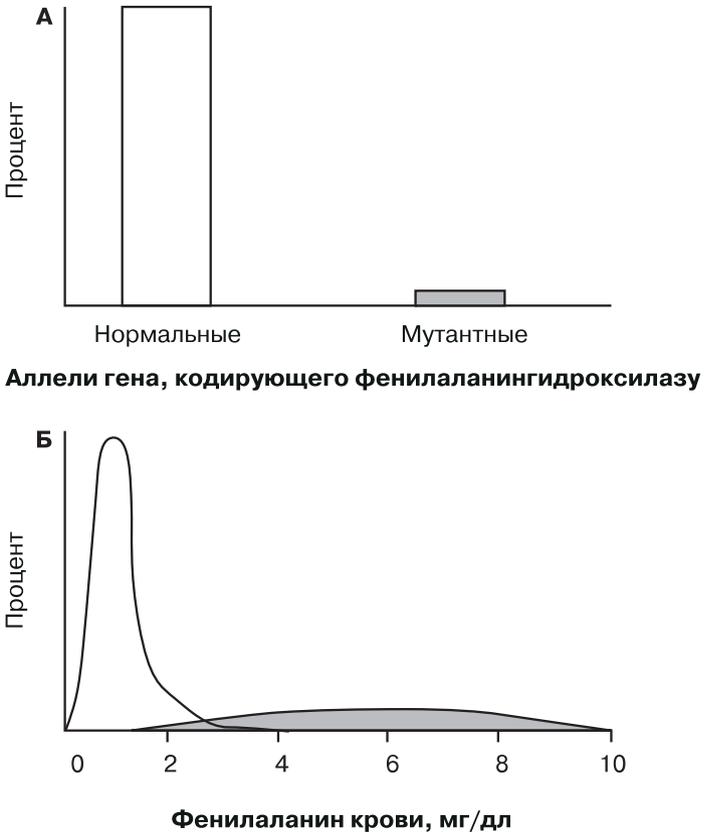


Рис. 2.9. Скрининг на фенилкетонурию (ФКУ) у младенцев: дихотомические и перекрывающиеся распределения в норме и при патологии. А – аллели гена, кодирующего фенилаланингидроксилазу, либо нормальные, либо мутантные; Б – кривые распределения новорожденных с ФКУ и без нее по уровню фенилаланина в крови резко различаются по амплитуде и частично перекрываются. (Распространенность ФКУ в действительности составляет примерно 1 на 10 000 и завышена на рисунке для наглядности.)

Однако большинство распределений переменных, встречающихся в клинической практике, непросто разделить на "норму" и "патологию", поскольку эти распределения по природе своей не дихотомические и не имеют отчетливых разрывов или двух различных пиков, из которых один соответствовал бы нормальному результату, а другой – патологическому. И тому есть несколько причин.

Во-первых, разделение популяций по многим лабораторным показателям на больных и здоровых невозможно даже с теоретической точки зрения. Заболевание может развиваться незаметно, проявляясь постепенным переходом от низких значений показателя к высоким по мере нарастания дисфункции. Именно так ведут себя лабораторные показатели, отражающие поражение определенного органа, например креатинин сыворотки при развитии почечной недостаточности.

Во-вторых, здоровые и больные фактически принадлежат к двум разным популяциям, но когда эти две популяции перемешаны, распознать каждую из них в общей массе практически невозможно, поскольку у разных больных один и тот же показатель может принимать различные значения, перекрывая значения этого показателя у здоровых; кроме того, число больных в общей популяции невелико.

Пример. Фенилкетонурия (ФКУ) – заболевание, характеризующееся прогрессирующей умственной отсталостью в детском возрасте. Наличие ряда мутантных аллелей гена, кодирующего фенилаланингидроксилазу, обуславливает отсутствие этого фермента, что при обычной диете приводит к накоплению в организме фенилаланина. Диагноз, который становится очевидным уже в первый год жизни, подтверждается устойчиво высоким уровнем (в несколько раз превышающим норму) фенилаланина и низким уровнем тирозина в крови.

Применяемый скрининг новорожденных на ФКУ по анализу крови на фенилаланин через несколько дней после рождения предназначен для своевременного начала лечения и предупреждения необратимых последствий. Однако в некоторых случаях результаты теста оказываются неправильными, так как распределения концентраций фенилаланина сыворотки у больных и здоровых новорожденных перекрываются, к тому же среди всех проходящих скрининг доля младенцев с ФКУ очень невелика, примерно 1 на 10 000 (рис. 2.9,Б). У некоторых новорожденных с ФКУ уровень фенилаланина находится в пределах нормы либо в связи с тем, что они еще не потребляют достаточного количества белка, либо потому, что они имеют такое сочетание аллелей, при котором заболевание протекает мягко. Вместе с тем у некоторых детей, не predisposed к развитию ФКУ, регистрируются относительно вы-

сокие уровни фенилаланина, например, из-за нарушений метаболизма у матери. Результат теста принято считать положительным уже в нижней области патологических значений, несмотря на то что она перекрывается с областью нормальных показателей; сделано это для того, чтобы выявить максимально возможное число младенцев с ФКУ, даже если среди пяти детей с положительным результатом скрининга болезнь разовьется лишь у одного.

В неселективных популяциях больные часто не выявляются, потому что их доля среди здоровых относительно мала, а также из-за того, что результаты лабораторных тестов, полученные в популяциях больных и здоровых людей, перекрываются. Кривая распределения признака, полученная для больных, "поглощается" большей кривой, описывающей распределение у здоровых. Если же больные и здоровые находятся в популяции в более равных пропорциях, как это бывает при отборе для исследования людей с явно высокой вероятностью заболевания, тогда итоговое распределение может быть истинно бимодальным (т.е. иметь два пика). Но даже в такой ситуации невозможно выбрать значения показателей, которые позволили бы однозначно разделить больных и здоровых (см. главу 3).

Итак, если между нормой и патологией не существует четкой границы, и врач выбирает ее по своему усмотрению, то какими основными правилами ему следует руководствоваться в случае принятия решения? Доказано, что целесообразно использование трех критериев оценки состояния как патологического: состояние должно быть необычным, проявляться болезнью и улучшаться при лечении. Эти три критерия не связаны между собой, поэтому, оценивая конкретный показатель, мы можем обнаружить, что по одним критериям он должен рассматриваться как патологический, а по другим – как нормальный.

Патология – необычное состояние

Норму принято рассматривать как наиболее часто встречающееся, или обычное, состояние. Все, что случается часто, считается нормальным и, наоборот, редкие события рассматриваются как патологические. Это статистическое определение, основанное на частоте изучаемого признака в определенной популяции. Как правило, контрольная популяция состоит из лиц без заболевания, но это необязательно. Например, мы можем сказать, что вполне естественно ощущать боль после операции или зуд при экземе.

Заманчиво было бы дать определение необычного в математических терминах. Стандартный прием, используемый для разграничения нормы и патологии, состоит в том, чтобы договорить-

ся (в некотором смысле произвольно), что все значения признака, выходящие за два стандартных отклонения от средней величины, считаются патологическими. Если предположить, что рассматриваемое распределение приближается к нормальному (гауссову), то по 2,5% наблюдений, расположенных на каждом из концов кривой распределения, следует рассматривать как патологические.

Конечно, как уже указывалось, большинство биологических процессов описывается распределениями, отличными от нормального. Поэтому лучше представить необычные величины как часть (или процентиль) фактического распределения. Подобный подход позволяет непосредственно оценить, насколько редко встречается тот или иной признак, не вдаваясь в предположения о форме его распределения.

Определение статистическими методами того, что есть норма, является общепринятым, однако, по ряду причин, такое определение может давать нечеткие или ошибочные результаты.

Во-первых, если все значения, находящиеся за произвольным статистическим пределом (допустим, за 95-м процентилем), рассматривались как патологические, то распространенность всех болезней должна быть одинаковой и составлять 5%. Это несовместимо с нашими обычными представлениями о частоте заболеваний.

Во-вторых, не существует соответствия между степенью статистической "необычности" и болезнью вообще. Такое соответствие может быть специфичным только для каждого конкретного заболевания. Для некоторых показателей отклонение от обычного уровня может быть принято как определенное проявление болезни, если их значения находятся в самых крайних участках кривой распределения – после 95-го или даже после 99-го перцентиля.

Пример. По мнению экспертов Всемирной организации здравоохранения (ВОЗ), у взрослых небеременных женщин уровень гемоглобина (Hb) ниже 12 г/дл означает анемию. В Британском обследовании у 11% из 920 небеременных женщин в возрасте 20 – 64 лет уровень Hb был ниже 12 г/дл. Это в два раза чаще ожидаемой распространенности, рассчитанной исходя из того, что ненормальной является величина, которая выходит за пределы двух стандартных отклонений [5]. Но были ли эти женщины с уровнем Hb ниже 12 г/дл "больны" из-за относительно низкого уровня Hb? Возможны два варианта: низкий уровень Hb может либо вызывать симптомы анемии, либо быть проявлением какого-то скрытого серьезного заболевания.

Такие симптомы анемии, как усталость, головокружение и раздражительность, не были связаны с уровнем Hb, по крайней мере, когда он был выше 8,0 г/дл. Более того, у женщин с уровнем Hb от 8,0 до 12,0 г/дл назначение внутрь препаратов железа привело к по-

вышению уровня гемоглобина в среднем на 2,30 г/дл, но не ослабилась выраженность симптомов по сравнению с их выраженностью у женщин, принимавших плацебо. Что касается скрытого серьезного заболевания, то низкий уровень Hb может быть проявлением злокачественного новообразования, хронической инфекции или ревматического процесса. Но эти состояния наблюдаются лишь у очень небольшой доли женщин с низким уровнем Hb.

Таким образом, только при уровне Hb менее 8,0 г/дл, который был обнаружен менее чем у 1% женщин в обследованной популяции, анемия могла быть проявлением серьезной патологии.

В-третьих, риск заболевания коррелирует с показателями многих лабораторных тестов по всему диапазону их возможных значений, снизу доверху.

Для "высоких нормальных" и "низких нормальных" значений уровня холестерина сыворотки, риск заболевания различается примерно в три раза.

В-четвертых, некоторые крайние, явно необычные значения, на самом деле предпочтительнее обычных. Это в первую очередь относится к величинам, находящимся вблизи нижней границы некоторых распределений. Вероятно, многие сочтут за благо иметь уровень креатинина сыворотки 0,4 мг/дл или систолическое артериальное давление 105 мм рт.ст. Оба значения необычно низкие, но они свидетельствуют лишь о том, что состояние здоровья у обследуемого выше, а риск заболеть – ниже среднего.

И наконец, иногда пациенты могут быть явно больны, несмотря на отсутствие отклонений результатов лабораторных диагностических тестов от обычных значений. Примерами служат гидроцефалия с низким внутричерепным давлением, глаукома без повышения внутриглазного давления, гиперпаратиреоз при нормокальциемии.

Патология – это болезнь

Более строгий подход к разграничению нормы и патологии состоит в том, чтобы называть патологическими те признаки, которые обычно приводят к болезни, инвалидности или смерти, т.е. проявляются клинически значимыми отклонениями от здорового состояния.

Пример. Потребление какого количества алкоголя (этилового спирта) следует считать "нормальным"? Некоторые исследования показали U-образную зависимость между потреблением алкоголя и смертностью: высокая смертность среди воздерживающихся, низкая – среди потребляющих умеренно и высокая – среди потребляющих неумеренно (рис. 2.10). Было сделано предположение, что

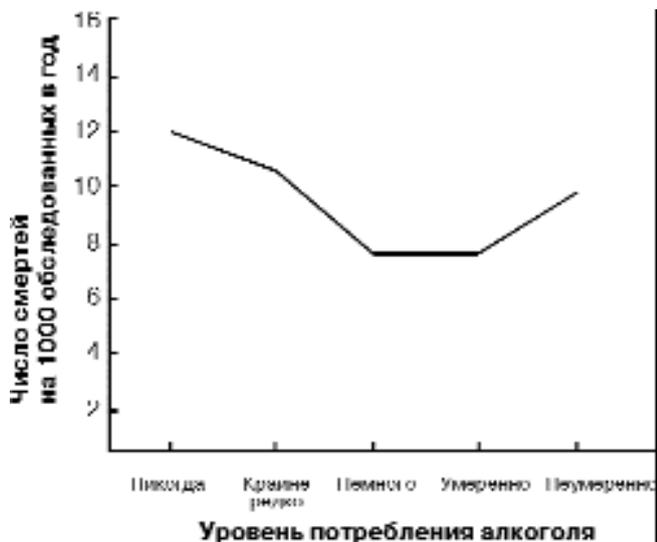


Рис. 2.10. Связь патологии с заболеванием. Соотношение между потреблением алкоголя и смертностью [Из: Shaper A.G., Wannamethee O., Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. Lancet 1988;2:1267–1273].

снижение смертности при умеренном потреблении алкоголя (нижний участок кривой) обусловлено повышением уровня липопротеидов высокой плотности, оказывающих антиатерогенное действие. С другой стороны, когда люди заболевают, они сокращают потребление алкоголя, что объясняет высокую смертность при низком потреблении алкоголя [6]. Что касается высокой смертности при значительном потреблении алкоголя, то здесь рассуждение проще: алкоголь вызывает ряд смертельных заболеваний (болезни сердца, злокачественные новообразования, инсульт). Определение причин U-образности этой кривой позволит решить вопрос о том, действительно ли полный отказ от потребления алкоголя столь же ненормальное явление, как пьянство.

Патология – значит поддающееся лечению

При некоторых состояниях, особенно таких, которые не сопровождаются жалобами (т.е. протекают бессимптомно), тот или иной признак следует считать патологическим только в случае, если лечение по поводу состояния, связанного с наличием этого признака, улучшает исход. Не любое состояние, связанное с повышенным риском негативного исхода, успешно поддается лечению:

ликвидация некоего состояния может не приводить к снижению этого риска, либо в силу того, что само по себе данное состояние является не причиной болезни, а лишь связанным с ней фактором, либо из-за необратимости уже происшедших изменений. Следует также учитывать, что сообщение пациенту о наличии у него болезни, может дать неблагоприятный психологический эффект, неоправданный в том случае, если лечение не может улучшить ситуацию.

Представления о том, что считать излечимым, меняются со временем. В идеале, врачебные решения основываются на результатах корректно выполненных клинических испытаний (см. главу 8). По мере проведения новых испытаний и, следовательно, появления новых сведений, меняется отношение к вопросу о стадии болезни, при которой лечение считается эффективным. Например, накопленные данные о лечении больных артериальной гипертензией изменили представления о том, при каком давлении следует проводить терапию. И по мере проведения исследований, оказалось, что чем ближе значения диастолического давления к норме, тем лучше эффект лечения.

Смещение к среднему

Когда результат теста слишком сильно отличается от нормы, врач склонен повторить анализ. Часто повторный результат оказывается ближе к норме. Почему это происходит? И следует ли этому доверять?

У пациентов, отобранных по крайним значениям в распределении, при последующих измерениях можно в среднем ожидать значения, отклоняющиеся от нормы в меньшей степени. Это обусловлено чисто статистическими причинами, а не улучшением состояния. Явление называется *смещение к среднему* (regression to the mean).

Феномен смещения к среднему можно проиллюстрировать следующим образом (рис. 2.11). Пациентов сначала отбирают для включения в исследование или для дальнейшей диагностики и лечения из-за того, что результат первого тестирования показателем оказался выше произвольно выбранной точки разделения на кривой распределения некоего параметра, построенной для всех обследованных пациентов. Для некоторых из этих пациентов результат останется по другую сторону точки разделения и при последующих измерениях, потому что для них истинные значения действительно выше среднего. Однако для других, у которых при первоначальном обследовании были найдены повышенные зна-



Рис. 2.11. Смещение к среднему.

чения, на самом деле характерны более низкие значения. Эти лица были отобраны только в силу случайной вариации, из-за которой в момент первого измерения у них были зарегистрированы высокие значения. При повторном измерении у этих пациентов определяются значения показателя более низкие, чем во время первого тестирования. Это явление определяет тенденцию к смещению среднего значения для всей подгруппы лиц, имевших при первом измерении значение признака выше точки разделения в сторону средней величины для всей популяции.

Таким образом, у пациентов, отобранных по необычно низкому или высокому результату лабораторного теста, при повторном тестировании следует в среднем ожидать сдвига результатов к центру распределения. Более того, результаты повторных измерений будут все больше приближаться к истинному значению признака, а именно к тому, которое можно было бы получить при многократном повторении измерения у одного и того же пациента. Таким образом, проверенную временем практику повторения лабораторных тестов, результаты которых оказались патологическими, и использования зачастую нормального результата повторного теста в качестве верного никак нельзя считать надуманной. Для этого есть как теоретические, так и эмпирические осно-

вания. Например, показано, что среди всех лиц, у которых при скрининге выявляется патологический уровень тироксина (Т4) в сыворотке, при повторном измерении у половины уровень тироксина оказывается нормальным [7]. Однако чем сильнее первоначальные значения отклоняются от нормы, тем выше вероятность обнаружения патологии и при повторном тестировании.

Резюме

Клинические данные могут быть качественными, порядковыми или количественными. Хотя многие клинические данные выражаются непрерывным рядом значений, по практическим соображениям они часто упрощаются до дихотомических (норма – патология) категорий. Оценки клинических показателей подвержены вариациям из-за ошибок измерения, а также различий в пределах одного или между несколькими индивидуумами. Метод измерения характеризуется достоверностью (действительно ли измеряется то, что предполагалось измерить?), воспроизводимостью (дают ли одинаковый результат повторные измерения одного и того же показателя?), диапазоном измеряемых значений, способностью реагировать на изменения измеряемого параметра и интерпретируемостью.

Частотные распределения для клинических показателей имеют разную форму и характеризуются центральной тенденцией и рассеянием.

Значения лабораторных показателей для здоровых и больных часто перекрываются; вследствие этого, а также из-за относительно низкой распространенности патологии в общей популяции часто бывает невозможно четко разграничить две группы пациентов по результатам однократного тестирования. Точка разделения, на которой заканчивается норма и начинается патология, выбирается произвольно и обычно связана с одним из трех критериев патологии: это состояние необычное с точки зрения статистики, оно проявляется в виде болезни и поддается лечению. Если отобраны пациенты с крайними результатами теста, то при повторении измерения результат чаще располагается ближе к центральной (нормальной со статистической точки зрения) области частотного распределения – феномен, называемый смещением к среднему.

Литература

1. Feinstein A.R. The need for humanized science in evaluating medication. *Lancet* 1972;2:421–423.
2. Day E., Maddem L., Wood C. Auscultation of foetal heart rate: an assessment of its error and significance. *Br Med J* 1968;4:422–424.

3. Morganroth J., Michelson E.L., Horowitz L.N., Josephson M.E., Pearlman A.S., Dunkman W.B. Limitations of routine long-term electrocardiographic monitoring to assess ventricular ectopic frequency. *Circulation* 1978;58:408–414.
4. Elveback L.R., Guillier C.L., Keating F.R. Health, normality, and the ghost of Gauss. *JAMA* 1970;211:69–75.
5. Elwood P.C., Waters W.E., Greene W.J.W., Sweetnam P. Symptoms and circulating hemoglobin level. *J Chron Dis* 1969;21:615–628.
6. Shaper A.G., Wannamethee G., Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. *Lancet* 1988;2:1267–1273.
7. Epstein K.A., Schneiderman L.J., Bush J.W., Zettner A. The "abnormal" screening serum thyroxine (T4): analysis of physician response, outcome, cost and health and effectiveness. *J Chron Dis* 1981;34:175–190.

Рекомендуемая литература

Department of Clinical Epidemiology and Biostatistics, McMaster University. Clinical disagreement I. How often it occurs and why. *Can Med Assoc J* 1980;123:499–504.

Department of Clinical Epidemiology and Biostatistics, McMaster University. Clinical disagreement II. How to avoid it and how to learn from one's mistakes. *Can Med Assoc J* 1980;123:613–617.

Feinstein A.R. *Clinical judgment*. Baltimore: Williams & Wilkins, 1967.

Feinstein A.R. *Problems in measurement, clinical biostatistics*. St. Louis: CV Mosby, 1977.

Feinstein A.R. *Clinimetrics*. New Haven, CT: Yale University Press, 1987.

Koran L.M. The reliability of clinical methods, data and judgment. *N Engl J Med* 1975;293:642–646,695–701.

Mainland D. Remarks on clinical "norms". *Clin Chem* 1971;17:267–274.

Murphy E.A. *The logic of medicine*. Baltimore: Johns Hopkins University Press, 1976.

Guyatt G.H., Feeny D.H., Patrick D.L. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622–629.

3 ДИАГНОЗ

Для разума при восприятии окружающего возможны четыре ситуации: явление существует и воспринимается так, как оно есть; не существует и не воспринимается; существует, но не воспринимается; не существует, но кажется существующим. Правильно распознать каждый случай — задача, достойная мудреца.

Эпиктет (греческий философ-стоик, II век н.э.)

Врачи тратят массу времени на то, чтобы, основываясь на объективных данных и жалобах больных, поставить диагноз. При этом используются разнообразные диагностические тесты. Опытным врачам помогают здравый смысл, данные литературы и оперативная оценка получаемой информации. Однако для правильной интерпретации результатов диагностических исследований врачу следует придерживаться некоторых фундаментальных принципов, которым и посвящена эта глава.

Под диагностическим тестом обычно понимают лабораторное исследование. Тем не менее обсуждаемые далее принципы применимы также к данным анамнеза, физикального и инструментального исследования, а также к ситуациям, когда диагностическим тестом служит целый комплекс показателей. Можно, например, говорить о диагностической значимости продромальных неврологических симптомов (головная боль, тошнота, рвота) при диагностике классической мигрени или кровохарканья и похудаания – в качестве признаков рака легкого у курильщиков.

Упрощение данных

В главе 2 отмечалось, что клинические показатели, включая данные диагностических тестов, бывают разных типов: качественные, порядковые и количественные. Независимо от характера данных, клиницисты обычно приводят их к более простой форме, чтобы ими было легче оперировать. Большинство порядковых шкал, по сути, отражает именно такой процесс упрощения.

Пример. Сердечный шум может варьировать от очень сильного до практически неслышимого. В процессе диагностического поиска эти тонкие оттенки интенсивности шума необязательно описывать словами. Вместо этого удобнее взять простую порядковую шкалу, допустим от 1 до 6 баллов. Еще чаще сложные данные упрощаются до дихотомии: наличие–отсутствие, патология–норма, болен–здоров. Именно такой подход применяется, когда по результатам теста назначают лечение. В определенный момент решение носит характер "или–или": назначить лечение или отменить его.

Примером упрощения информации для клинических целей может служить решение вопроса о назначении лечения в зависимости от уровня артериального давления. Обычно давление измеряют с точностью до 2 мм рт.ст., т. е. количественно. Тем не менее клинические рекомендации, в частности рекомендации Объединенного Национального Комитета по выявлению, оценке и лечению артериальной гипертензии [1], указывают определенный уровень (например, диастолическое давление 95 мм рт.ст.), при котором следует назначать медикаментозное лечение. Поступая подобным образом, клиницисты переходят от количественной шкалы к качественной (в данном случае – дихотомической). Продолжая этот пример, заметим, что Объединенный Национальный Комитет рекомендует врачам определять конкретный план терапии, исходя из того, является ли диастолическое давление "слабо повышенным" (90–94 мм рт.ст.), "умеренно повышенным" (95–114 мм рт.ст.), или "резко повышенным" (115 мм рт.ст. и более), т.е. пользуясь типичной порядковой шкалой.

Точность результатов теста

Процесс установления диагноза несовершенен – в итоге, мы можем лишь предполагать, что диагноз верен, нежели утверждать это со всей определенностью. Раньше врачи выражали степень уверенности в клиническом диагнозе, предваряя его формулировку словами "исключается..." или "возможно...". Современные клиницисты все чаще выражают свою уверенность в диагнозе через вероятности. Следовательно, врач должен понимать математическую суть диагностической ценности тестов в различных ситуациях. Как правило, это помогает врачу уменьшить степень неопределенности того или иного диагностического теста; в ряде случаев – точнее оценить эту неопределенность, а иногда – лишь осознать степень своей неуверенности в диагнозе.

Соотношение между результатами тестирования и верным диагнозом схематически изображено на рис. 3.1. Результат теста может быть либо положительным (патология), либо отрицательным (норма), а заболевание может либо присутствовать, либо отсутствовать. Возможны четыре варианта толкования результатов теста – два истинных и два ложных. Правильный ответ – это по-

ложительный результат при наличии заболевания или отрицательный при его отсутствии. Напротив, ответ ошибочный, если результат теста положителен (ложноположительный), хотя человек здоров, или отрицателен (ложноотрицательный), хотя человек болен.

		БОЛЕЗНЬ	
		Присутствует	Отсутствует
ТЕСТ	Положительный	Истинно положительный a	Ложно-положительный b
	Отрицательный	Ложно-отрицательный c	Истинно отрицательный d

Рис. 3.1. Соотношение между результатами диагностического теста и наличием заболевания. Существует два варианта правильного результата теста (истинно положительный и истинно отрицательный результат) и два варианта ошибочного результата (ложноположительный и ложноотрицательный результат).

"Золотой стандарт"

Оценка точности теста основана на сравнении его результата с результатом некоторого точного способа определения болезни, т.е. с заслуживающим доверия показателем истины – так называемым "золотым стандартом" (референтным, или эталонным методом). В действительности "золотой стандарт" тоже нельзя считать абсолютным. Иногда эталоном точности является относительно простой и недорогой тест, например посев мазка из зева для идентификации бета-гемолитического стрептококка группы А при фолликулярной ангине или серологическое исследование для выявления ВИЧ-инфекции. Однако гораздо чаще для ответа на вопрос о наличии заболевания нужны исследования достаточно трудоемкие, дорогие или сопряженные с риском – биопсия, хирургическое вмешательство и, в конце концов, аутопсия.

По отношению к заболеваниям, признаки которых становятся явными иногда лишь спустя несколько лет после появления первых симптомов, в качестве "золотого стандарта" могут использоваться результаты длительного наблюдения. В эту категорию

попадает большинство видов злокачественных новообразований, хронических и дегенеративных заболеваний. Такие болезни диагностируются несмотря на невозможность сравнения применяемых косвенных методов диагностики с "золотым стандартом", т.е. с результатами прямых тестов, ввиду небезопасности, сложности и высокой стоимости последних. Необходимо правильно рассчитать продолжительность наблюдения, учитывая, что этот период должен быть достаточно долгим, чтобы обнаружить болезнь у тех пациентов, у которых она была, но не проявлялась, однако все же не таким длинным, чтобы болезнь возникла у тех, у кого ее не было при первичном тестировании.

Поскольку точные способы установления диагноза почти всегда сопряжены с достаточно высоким риском и затратами средств, врачи и пациенты предпочитают "золотому стандарту" более простые тесты, по крайней мере на первых этапах. Так, для определения этиологии пневмонии сначала производят рентгенографию грудной клетки и анализ мокроты, а не биопсию легкого. Аналогично при подозрении на инфаркт миокарда регистрируют электрокардиограмму (ЭКГ) и проводят анализ ферментов сыворотки, не торопясь сразу прибегнуть к катетеризации сердца и визуализирующим методам. Простые тесты можно использовать для диагностики вместо более сложных, но более точных, если мы ясно сознаем, что идем на риск постановки ошибочного диагноза. Этот риск есть плата за безопасность и удобство выполнения более простого теста, что оправдано только в тех случаях, когда риск ошибки известен и допустимо низок. Для этого необходимо иметь достоверные данные о сравнении точности простых тестов с соответствующим "золотым стандартом".

Отсутствие информации об отрицательных результатах диагностических тестов

Целью всех клинических исследований, посвященных оценке диагностических тестов, является получение данных, достаточных для заполнения всех четырех полей таблицы (см. рис. 3.1). Без этих данных невозможно оценить риск ошибочной диагностики, что является важнейшей характеристикой каждого теста. Однако заполнить все четыре поля таблицы подчас бывает очень трудно, ведь объективные и надежные способы установления диагноза, даже если они теоретически существуют, иногда недоступны по этическим или практическим причинам. Рассмотрим ситуацию, при которой максимально возможная информация о диагностических тестах имеется, причем данные были получены преимущественно в клинических, а не в экспериментальных услови-

ях. Врачи исходят из того, что в интересах больного применение более подробного обследования оправданно только в случае, если предварительные тесты дали положительный результат. Они с явной неохотой идут на проведение углубленного обследования, связанного с большими расходами и риском для пациента, если результат предварительного диагностического теста был отрицательным. Вот почему в медицинской литературе так мало информации об отрицательных результатах тестирования, будь то истинно отрицательные или ложноотрицательные результаты.

В качестве примера рассмотрим результаты скрининга на наличие рака предстательной железы по уровню простатспецифического антигена (ПСА) [2]. Биопсия назначалась только при уровне ПСА выше критического. Понятно, что авторы публикации избегали проведения без веских оснований этой весьма неприятной для пациента процедуры. В итоге частота ложноотрицательных результатов при скрининге с применением ПСА осталась неизвестной.

Отсутствие информации о результатах тестирования у лиц, не имеющих заболевания

Как только что упоминалось, врачи, по вполне объяснимым причинам, избегают проведения сложных тестов без убедительных к тому показаний. Между тем оценка диагностических характеристик теста может оказаться ошибочной, если применять его только у пациентов с изучаемым заболеванием.

Пример. Магнитно-резонансная томография (МРТ) поясничного отдела позвоночника при обследовании пациентов с болями в спине во многих случаях выявляет грыжи межпозвоночных дисков. Это позволяет объяснить причину болей и назначить лечение. Но вот МРТ была проведена 98 добровольцам, у которых симптомы заболевания отсутствовали [3]. Результаты анализировались рентгенологами, не имевшими информации о наличии или отсутствии симптомов у этих лиц. Выпячивание дисков было обнаружено почти у $2/3$ обследованных, а это не на много реже, чем у пациентов с болями в спине. Авторы сделали вывод, что сочетание боли с грыжей межпозвоночного диска часто может быть случайным.

Отсутствие объективных критериев диагностики заболеваний

Для некоторых заболеваний строгие критерии диагностики вообще отсутствуют. Ярким примером такого заболевания может служить стенокардия. Ее клинические проявления описаны почти сто лет назад, однако до сих пор не найдено лучшего способа диагностики, чем тщательное изучение анамнеза. Разумеется, этот клинический синдром обусловлен множеством причин, поддаю-

щихся объективной оценке, например, стенозом коронарных артерий, выявляемым при ангиографии, при скинтиграфии с таллием во время нагрузочной пробы и по характерным изменениям на ЭКГ в покое и при физической нагрузке. Все эти объективные признаки наблюдаются у пациентов с подозрением на стенокардию, однако ни один из указанных симптомов не является патогномоничным, т.е. таким, по которому можно однозначно судить о наличии или отсутствии стенокардии.

Иногда стремление к "строгой истине" заводит врача в порочный круг. О достоверности лабораторного теста судят путем сравнения его результатов с клиническим диагнозом, основанным на тщательном изучении анамнеза и данных физикального исследования. Затем этот тест используется для подтверждения клинического диагноза, установленного по анамнезу и данным физикального исследования! Пример такого подхода: использование манометрии для "подтверждения" диагноза синдрома раздраженной кишки. Дело в том, что в свое время определенная форма манометрической кривой сокращений кишечника была описана как характерная для синдрома раздраженной кишки именно вследствие того, что она выявлялась у тех больных, которым этот диагноз был поставлен по клинической картине.

Последствия применения несовершенных стандартов

Из-за перечисленных трудностей врачи иногда не могут найти информацию о том, насколько используемые ими на практике диагностические тесты соответствуют абсолютно надежному стандарту. В качестве такого стандарта они вынуждены выбирать какой-либо тест, который, безусловно, далек от совершенства, но считается лучшим из существующих. В итоге сравниваются два сомнительных теста, один из которых принят в качестве стандарта либо потому, что он больше применяется, либо, исходя из согласованного мнения экспертов. Так можно дойти до парадокса: при сопоставлении нового теста с давно применяемым (но неточным) может показаться, что изучаемый тест хуже старого, даже если на самом деле он лучше. Например, если новый тест чувствительнее, чем стандартный, то дополнительно выявленные с его помощью случаи будут считаться ложноположительными по отношению к старому тесту. Именно так случилось, например, при сравнении ультразвукового исследования и холецистографии для диагностики желчнокаменной болезни [4]. При ультразвуковом исследовании у 5 больных были обнаружены камни, не выявленные на холецистограмме. Двоим из этих больных впоследствии была проведена хирургическая операция и камни желчного пузыря действи-

тельно были найдены. Следовательно, по крайней мере для этих двух больных стандартная холецистография оказалась менее точной, чем более новый ультразвуковой метод. Аналогичным образом, если новый тест чаще дает отрицательные результаты у лиц без соответствующего заболевания, то его результаты по отношению к данным старого теста будут рассматриваться как ложноотрицательные. Итак, при сравнении с несовершенным стандартом новый тест может показаться менее эффективным, несмотря на то, что в действительности дает результаты, более близкие к истине.

Чувствительность и специфичность

Взаимоотношения между результатами диагностического теста и реальным наличием или отсутствием заболевания продемонстрированы на рис. 3.2. Это такая же четырехпольная таблица,

		БОЛЕЗнь			
		Присутствует	Отсутствует		
ТЕСТ	Положительный	a	b	a + b	+ PV = $\frac{a}{a+b}$
	Отрицательный	c	d	c + d	- PV = $\frac{c}{c+d}$
		a + c	b + d	a + b + c + d	

$$Se = \frac{a}{a+c} \quad Sp = \frac{d}{b+d} \quad P = \frac{a+c}{a+b+c+d}$$

$$LR+ = \frac{\frac{a}{a+c}}{\frac{b}{b+d}} \quad LR- = \frac{\frac{c}{a+c}}{\frac{d}{b+d}}$$

Рис. 3.2. Характеристики диагностического теста и их определения. Se (sensitivity) – чувствительность; Sp (specificity) – специфичность; P (prevalence) – распространенность заболевания; -PV (negative predictive value) – прогностическая ценность отрицательного результата теста; +PV (positive predictive value) – прогностическая ценность положительного результата теста; LR+ (positive likelihood ratio) – отношение правдоподобия положительного результата теста; LR- (negative likelihood ratio) – отношение правдоподобия отрицательного результата теста.

как и на рис. 3.1, но с добавлением некоторых важных понятий. В остальном эта глава посвящена детальному рассмотрению указанных взаимоотношений, что иллюстрируется конкретным клиническим примером на рис. 3.3. Здесь решается вопрос о том, с какой точностью врач на основании клинических данных может определить, вызвана ангина гемолитическим стрептококком группы А или нет. В качестве "золотого стандарта" использован метод посева мазка из зева.

Определения

Как показано на рис. 3.2, *чувствительность* определяется как доля лиц с положительным результатом теста в популяции с изучаемым заболеванием. Чувствительный тест редко пропускает пациентов, у которых имеется болезнь. *Специфичность* – это доля

		Бета-гемолитический стрептококк в посеве мазка из зева		
		Присутствует	Отсутствует	
Клинический диагноз стрептококковой ангины	Да	27	35	62 + PV = $\frac{27}{62} = 44\%$
	Нет	10	77	87 - PV = $\frac{77}{87} = 88\%$
		37	112	149

$$Se = \frac{27}{37} = 73\% \quad Sp = \frac{77}{112} = 69\% \quad P = \frac{37}{149} = 25\%$$

$$LR+ = \frac{\frac{27}{27+10}}{\frac{35}{35+77}} = 2,3 \quad LR- = \frac{\frac{10}{10+27}}{\frac{77}{77+35}} = 0,39$$

Рис. 3.3. Точность диагноза стрептококковой ангины, установленного на основе клинических наблюдений, по сравнению с результатами посева мазка из зева [Fletcher S.W., Hamann C. J Comm Health 1976;1:196–204].

лиц с отрицательным результатом теста в популяции без изучаемой болезни. Специфичный тест, как правило, не относит здоровых к категории больных.

Применив эти определения к случаю с диагностикой ангины (см. рис. 3.3), мы увидим, что результат посева мазка был положительным у 37 из 149 пациентов, а врач общей практики правильно диагностировал 27 случаев, т.е. чувствительность клинического метода составила 73%. С другой стороны, у 112 пациентов результат посева был отрицательным, и в 77 случаях врач общей практики справедливо отказался от применения антибиотиков, следовательно, специфичность клинического метода равна 69%.

Использование чувствительных тестов

При выборе диагностического теста врач должен принимать во внимание его чувствительность и специфичность. Чувствительный тест (т.е. такой, который при наличии болезни обычно дает положительный результат) следует выбрать, если есть риск пропустить опасную, но излечимую болезнь – туберкулез, сифилис, лимфогранулематоз. Чувствительные тесты рекомендуются также на ранних стадиях диагностического поиска для сужения его рамок, когда возможных вариантов много и диагностические тесты позволяют исключить некоторые из них, т.е. сделать вывод, что эти заболевания маловероятны. Например, при инфильтратах в легких, сочетающихся с похуданием, в качестве раннего диагностического теста целесообразен тест на антитела к ВИЧ для исключения оппортунистических инфекций, связанных со СПИДом. Подводя итог, подчеркнем, что для врача чувствительный тест особенно информативен в том случае, когда он дает отрицательный результат.

Использование специфичных тестов

Специфичные тесты нужны для подтверждения (установления) диагноза, предложенного на основании других данных. В самом деле, высокоспецифичный тест не должен дать положительный результат в отсутствие заболевания (т.е. редко бывает ложноположительным). Высокоспецифичные тесты особенно необходимы, если ложноположительный результат может нанести пациенту вред – физический, эмоциональный или финансовый. Например, прежде чем назначать больному со злокачественным новообразованием химиотерапию, сопряженную с риском осложнения, эмоциональной травмой и затратой средств, требуется морфологическое подтверждение диагноза, так как результаты менее специфичных тестов недостаточны. Итак, специфичный тест диагностически наиболее эффективен, когда дает положительный результат.

Компромисс между чувствительностью и специфичностью

Желательно, чтобы диагностический тест одновременно был и высокочувствительным, и высокоспецифичным. К сожалению, этого, как правило, достичь не удастся, но возможен компромисс между этими двумя показателями, по крайней мере в тех случаях, когда клинические данные принимают значения в некотором интервале. В подобных ситуациях положение точки разделения (cut-off point) на непрерывном переходе (континууме) между нормой и патологией устанавливается произвольно. Следовательно, для любого конкретного результата, выраженного количественно непрерывной величиной, одна характеристика (допустим, чувствительность) может быть усилена не иначе как за счет другой (специфичности). Данные табл. 3.1 иллюстрируют эти взаимоотношения на примере диагностики сахарного диабета. Если принять за диагностический критерий болезни уровень глюкозы крови более 9,9 ммоль/л (180 мг/дл) через 2 ч после нагрузочной пробы с глюкозой, то поставленный диагноз, несомненно, будет во всех случаях верен, однако из-за слишком высокого его значения многие больные окажутся невыявленными. Тест будет очень специфичным в ущерб чувствительности. С другой стороны, если считать больными сахарным диабетом всех, у кого уровень глюкозы крови более 3,85 ммоль/л (70 мг/дл), то вряд ли кто-то из больных будет

Таблица 3.1

Компромисс между чувствительностью и специфичностью при диагностике сахарного диабета*

Глюкоза крови через 2 ч после нагрузки, ммоль/л (мг/дл)	Чувствительность, %	Специфичность, %
3,85 (70)	98,6	8,8
4,4 (80)	97,1	25,5
4,95 (90)	94,3	47,6
5,5 (100)	88,6	69,8
6,05 (110)	85,7	84,1
6,6 (120)	71,4	92,5
7,15 (130)	64,3	96,9
7,7 (140)	57,1	99,4
8,25 (150)	50,0	99,6
8,8 (160)	47,1	99,8
9,35 (170)	42,9	100,0
9,9 (180)	38,6	100,0
10,45 (190)	34,3	100,0
11,0 (200)	27,1	100,0

* [Public Health Service. Diabetes program guide. Publication no.506. Washington. DC. US. Government Printing Office, 1960].

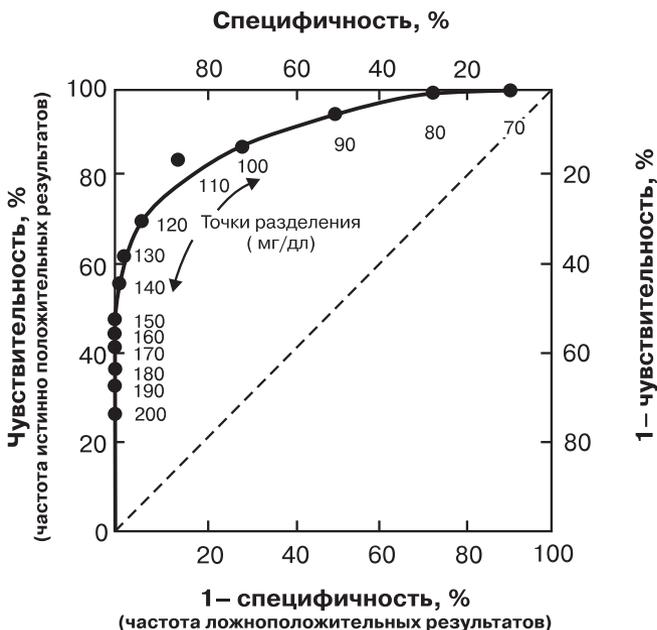


Рис. 3.4. Характеристическая кривая. Определение уровня глюкозы крови (через 2 ч после нагрузочной пробы с глюкозой) в качестве диагностического теста на наличие сахарного диабета [Public Health Service. Publication no. 506. Washington, DC: U.S. Government Printing Office, 1960].

пропущен, но в то же время будет много случаев гипердиагностики. Значит, при этих условиях тест чувствителен, но не специфичен. Пользуясь лишь одним показателем – уровнем глюкозы крови при обычных условиях, невозможно улучшить одновременно как чувствительность, так и специфичность теста.

Другой способ выразить соотношение между чувствительностью и специфичностью теста – построение так называемой характеристической кривой (ROC-curve, receiver-operator characteristic curve). На рис. 3.4 показана характеристическая кривая для диагностики сахарного диабета по уровню глюкозы в крови. По оси ординат обозначена частота истинно положительных результатов (чувствительность), по оси абсцисс – частота ложноположительных результатов (1 минус специфичность) по всему диапазону точек разделения. Значения по осям соответствуют вероятностям от 0 до 1 (т.е. от 0 до 100%). Полученный график иллюстрирует дилемму, возникающую при попытке достижения компро-

мисса между чувствительностью и специфичностью. Если в качестве точки разделения принять уровень глюкозы крови 5,5 ммоль/л (100 мг/дл), то будет пропущено только 11% больных сахарным диабетом, но 30% здоровых обследуемых будут встревожены из-за ложноположительного результата тестирования. Повысив порог (точку разделения) до 6,6 ммоль/л (120 мг/дл), мы уменьшим число ложноположительных ответов до менее чем 10%, однако это произойдет ценой пропуска почти 30% больных.

Тесты, обеспечивающие четкое разграничение больных и здоровых, имеют характеристическую кривую, вытянутую в сторону верхнего левого угла графика чувствительность/специфичность. Для них повышение чувствительности (снижение точки разделения) практически не сопровождается убыванием специфичности вплоть до момента, когда будет достигнут очень высокий уровень чувствительности. Для менее диагностически эффективных тестов характеристическая кривая располагается ближе к диагонали, соединяющей левый нижний и правый верхний углы. Диагональ отражает соотношение частот истинно и ложноположительных результатов для абсолютного неинформативного теста, например, когда выбор между диагнозами врач делает путем подбрасывания монеты.

Характеристическая кривая показывает, насколько сложен компромисс между чувствительностью и специфичностью теста. С помощью этой кривой можно определить оптимальное значение точки разделения. В принципе, точка разделения должна находиться вблизи "плеча" кривой, если только нет медицинских причин для минимизации вероятности либо ложноотрицательных, либо ложноположительных результатов.

Построение характеристических кривых особенно ценно в качестве способа сравнения альтернативных тестов для одного и того же диагноза. Общая точность теста может быть представлена в виде площади под характеристической кривой: чем больше эта площадь, тем эффективнее тест. На рис. 3.5 сравниваются характеристические кривые для двух диагностических опросников, используемых при скрининге на алкоголизм у пожилых людей – CAGE и MAST (Мичиганский тест для скрининга на алкоголизм) [5]. Для CAGE, который более чувствителен и более специфичен, чем MAST, площадь под кривой гораздо больше.

Естественно, что тесты, одновременно обладающие высокой чувствительностью и специфичностью, очень ценны и популярны. Однако на практике врачам редко удастся работать с такими тестами, так что приходится применять различные способы нахождения компромисса между чувствительностью и специфично-

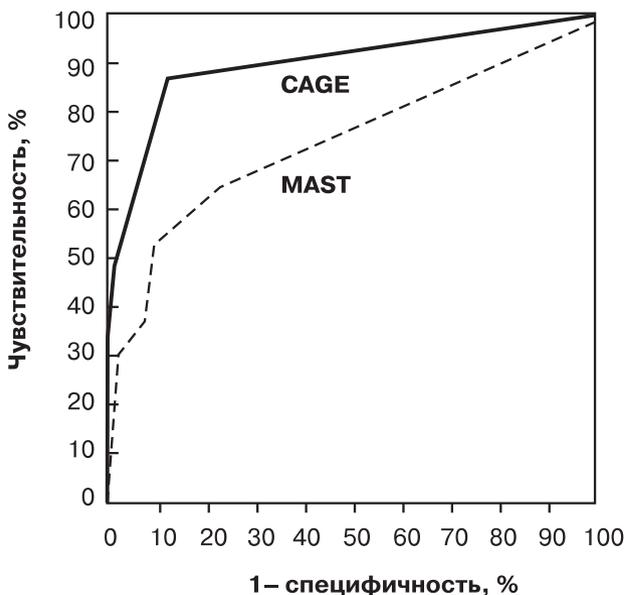


Рис. 3.5. Характеристические кривые по результатам применения опросников CAGE и MAST у пожилых людей, страдающих и не страдающих алкоголизмом [Jones T.V. et al. J Gen Intern Med 1993;8:674–678].

стью. Наиболее распространенный из них состоит в одновременном использовании нескольких тестов (см. ниже).

Факторы, определяющие чувствительность и специфичность

Нередко новый диагностический тест, который поначалу вызывает восторг, по мере накопления опыта оказывается далеко не таким эффективным, как это казалось. Достаточно вспомнить, как угас энтузиазм по поводу клинической ценности определения сывороточного карциноэмбрионального антигена (КЭА). Сразу после открытия КЭА, его определение рассматривали как многообещающий метод диагностики рака толстой кишки. Однако позднее оказалось, что уровень КЭА повышен при многих других заболеваниях, а также примерно у 20% курящих людей без злокачественных новообразований. Подобное разочарование вслед за энтузиазмом порождается не недобросовестностью авторов, выполнивших начальные исследования, и не последующим несправедливым скептицизмом медицинской общественности. Скорее, это обус-

ловлено ограниченными возможностями методов, с помощью которых оценивались характеристики нового теста. В простейшем случае, характеристики диагностического теста (например, чувствительности и специфичности) могут быть оценены неточно по причине выбора неверного стандарта для сравнения, что обсуждалось выше. Следует назвать еще три фактора, существенно влияющих на оценку как чувствительности, так и специфичности. Это разброс характеристик пациентов, которым проводился тест, систематическая ошибка при оценке характеристик теста и, наконец, случайная ошибка.

Характеристики пациентов

Трудности возникают, если характеристики пациентов, отобранных для участия в исследовании по оценке теста, отличаются от характеристик пациентов той группы, в которой этот тест будет применяться в клинической практике. Раньше тесты часто оценивались путем сравнения результатов, полученных на явно больных и на заведомо здоровых людях (например, на добровольцах, отобранных из студентов-медиков). Понятно, что изучаемый тест прекрасно дифференцирует эти крайние группы.

Однако следует учитывать, что даже пациенты с заболеванием, для диагностики которого тест предназначен, различаются по тяжести, стадии или длительности заболевания, а чувствительность теста выше у более тяжелых больных.

Пример. На рис. 3.6 показано, что эффективность теста КЭА неодинакова при разных стадиях рака толстой кишки. Тест КЭА высокоэффективен при метастазировании опухоли, но не в начальных стадиях заболевания. Таким образом, чувствительность теста для диагностики рака толстой кишки зависит от состава больных с различными стадиями заболевания в той группе, в которой производилась оценка характеристик теста, а стабильность этой оценки повышается, если в группу вошли больные с одной и той же стадией болезни [6].

Точно так же, по каким-то причинам тест может оказаться положительным у лиц, у которых лишь предполагается интересующее нас заболевание, но в действительности оно отсутствует. Это приведет к увеличению частоты ложноположительных результатов и снижению специфичности теста. Например, уровень КЭА повышен у многих больных язвенным колитом или циррозом печени. Если при оценке эффективности теста КЭА для выявления рака толстой кишки такие больные будут включены в группу "без заболевания", то увеличится число ложноположительных заключений и специфичность теста снизится.

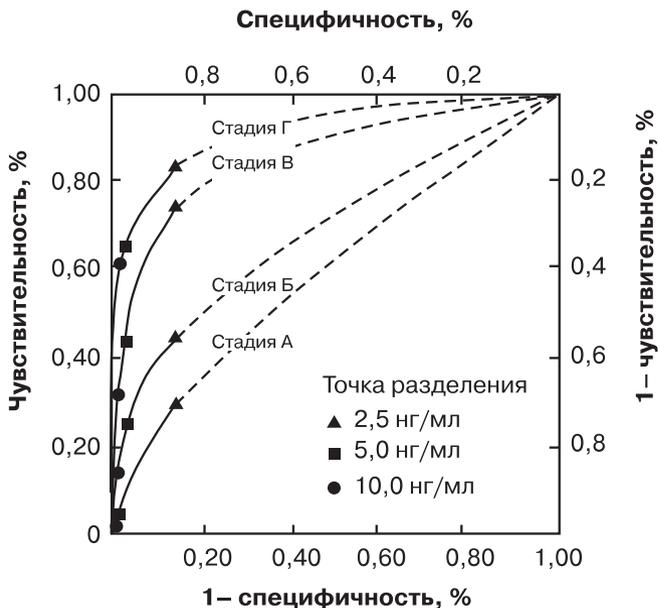


Рис. 3.6. Характеристическая кривая теста КЭА для диагностики рака толстой кишки на разных стадиях заболевания. Чувствительность и специфичность теста меняются в зависимости от стадии заболевания [Fletcher R.H. Ann Intern Med 1986;104:66 –73].

Теоретически чувствительность и специфичность не должны зависеть от распространенности заболевания в выборке, в которой производится оценка теста (см. рис. 3.2). Однако на практике происходит иначе: как уже отмечалось, чувствительность и специфичность теста могут быть различными, в зависимости от стадии и степени тяжести заболевания у больных, входящих в группу, по которой оценивается эффективность теста. Встречаемость же больных с разной выраженностью заболевания в выборке может быть различной. Это подтверждается при использовании скрининговых тестов (более подробно скрининг обсуждается в главе 8). Скрининг представляет собой обследование популяции, в которой отсутствуют проявления заболевания; распространенность заболевания обычно низкая и у пациентов в основном наблюдаются ранние стадии и менее тяжелые формы болезни. В подобных условиях наблюдается тенденция к более низкой чувствительности и более высокой специфичности, чем в случае применения того же теста у предположительно больных, поскольку у многих из них болезнь значительно выражена.

Систематическая ошибка

Иногда чувствительность и специфичность теста оцениваются с учетом данных, которые были использованы для постановки истинного диагноза, что приводит к смещенным оценкам характеристик теста. Этому может быть несколько причин. Как уже указывалось, если тест оценивается по данным, полученным в ходе клинического обследования пациентов с подозрением на изучаемое заболевание, то положительный результат теста может побудить врача к продолжению диагностического поиска, что увеличивает возможность выявления заболевания. И наоборот, отрицательный результат может привести к отказу от дальнейшего обследования, что увеличивает вероятность невыявления заболевания, если даже оно имеется.

В других случаях результат теста может стать частью информации, используемой для постановки диагноза, либо наоборот, результаты теста интерпретируются с учетом окончательного диагноза или других клинических данных. Такого рода систематические ошибки часто возникают при интерпретации рентгенограмм, поскольку этот процесс во многом субъективен и, вследствие этого, подвержен влиянию дополнительной клинической информации. Всем врачам знакома ситуация, когда с учетом клинических данных ошибочно выявляется патология на рентгенограмме или же когда при обращении к старым рентгенограммам на них находят то, что ранее пропустили, поскольку в отсутствие соответствующих клинических признаков не обратили должного внимания на определенную область снимка. Во избежание таких смещенных оценок некоторые рентгенологи предпочитают изучать снимок дважды: первый раз – не имея клинической информации, второй раз – располагая ею. Все эти смещения оценки приводят к повышению соответствия между исследуемым и стандартными тестами, т.е. тест кажется более информативным, чем на самом деле. Примером служит приведенная в начале главы ситуация с диагностикой грыжи межпозвоночного диска с помощью МРТ у пациентов с жалобами на боли в спине.

Случайная ошибка

Значения чувствительности и специфичности (или отношения правдоподобия – другой характеристики диагностических тестов, рассматриваемой ниже) обычно оцениваются путем наблюдений в сравнительно малых группах лиц с изучаемым заболеванием и без него. Вследствие случайных вариаций в любой выборке, особенно малой, чувствительность и специфичность теста могут исказиться, даже если в исследовании отсутствует системати-

ческая ошибка. Наблюдаемые значения соответствуют истинным значениям в диапазоне, обозначаемом обычно как 95% доверительный интервал (см. главу 9). Ширина этого диапазона определяет степень точности оценок чувствительности и специфичности. Следовательно, если обследованные группы пациентов малы, то оцененным по ним значениям чувствительности и специфичности не следует слишком доверять.

На рис. 3.7 показано, как возрастает точность оценки чувствительности с увеличением числа обследованных. В данном примере наблюдаемая чувствительность диагностического теста равна 75%, причем, если эта оценка основана на наблюдении всего 10 пациентов, то истинная чувствительность просто из-за случайности может колебаться от 45 почти до 100%. Если в исследование включить больше пациентов, то 95% доверительный интервал сужается, т.е. точность оценки возрастает.

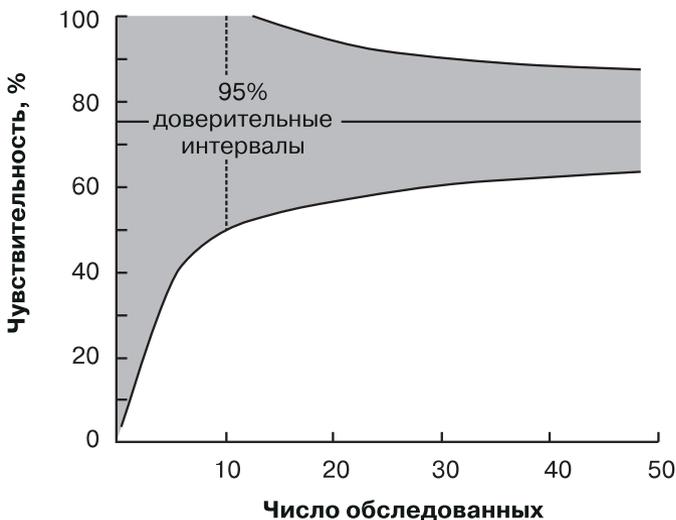


Рис. 3.7. Точность оценки чувствительности. Значения 95% доверительного интервала для чувствительности, равной 75%, в зависимости от размера выборки.

Прогностическая ценность теста

Как уже отмечалось, чувствительность и специфичность — это характеристики теста, которые необходимо учитывать при решении вопроса о том, следует ли использовать данный тест. Но

как только тест применен и получены результаты (положительные или отрицательные), понятия чувствительности и специфичности теряют смысл, поскольку теперь диагноз известен. Но ведь если бы диагноз был известен заранее, незачем было бы назначать тест! Задача, стоящая перед врачом, заключается в том, чтобы, зная результат теста, сделать вывод, болен человек или нет.

Определения

Вероятность наличия заболевания при условии известного результата теста называется *прогностической ценностью* (predictive value) теста (см. рис. 3.2). *Прогностическая ценность положительного результата* – это вероятность заболевания при положительном (патологическом) результате теста; *прогностическая ценность отрицательного результата* – вероятность отсутствия заболевания при отрицательном (нормальном) результате теста. Знание прогностической ценности теста позволяет врачу ответить на вопрос: "Каковы шансы, что данный пациент страдает (не страдает) определенным заболеванием, если у него результат теста положительный (отрицательный)"? Прогностическую ценность иногда называют *апостериорной* (или *посттестовой*) *вероятностью*, т.е. вероятностью выявления болезни после того, как стали известны результаты теста (см. рис. 3.3). Среди пациентов, которым врачи общей практики поставили по клинической картине диагноз стрептококковой ангины, менее половины (44%) действительно имели это заболевание в соответствии с результатом посева мазка (прогностическая ценность положительного результата 44%). Прогностическая ценность отрицательного заключения врача общей практики оказалась выше: из 87 пациентов, которым диагноз стрептококковой ангины не был поставлен, у 77 (88%) диагноз также был исключен по результатам посева мазка.

Ранее были объяснены показатели, используемые для оценки эффективности диагностического теста в целом. Один из таких показателей – *точность* – соответствует доле правильных результатов теста в общем количестве результатов как положительных, так и отрицательных. (Для представленного на рис. 3.3 примера с диагностикой ангины точность диагноза, поставленного врачом общей практики по клинической картине, равнялась 70%.) Другой обобщенный показатель при известных результатах применения теста – это *площадь под характеристической кривой*. Однако такие суммарные показатели слишком грубы и не могут быть полезны, поскольку специфическая информация об их составляющих – чувствительности, специфичности и прогностической ценности для конкретных точек разделения, теряется при их объединении.

Факторы, определяющие прогностическую ценность теста

Прогностическая ценность – это характеристика не только самого теста. Она зависит как от чувствительности и специфичности теста, так и от *распространенности заболевания* в исследуемой популяции, т.е. доли лиц с изучаемым заболеванием в определенной популяции в данный момент времени. Распространенность называется также *априорной* (или *претестовой*) *вероятностью*, т.е. это вероятность выявления болезни до того, как стали известны результаты теста. (Подробнее о распространенности см. главу 4.)

Математическая формула, связывающая чувствительность, специфичность теста и распространенность заболевания с предсказательной ценностью положительного результата, выводится из теоремы Байеса для условных вероятностей:

$$\text{Прогностическая ценность положительного результата} = \frac{(\text{чувствительность}) \times (\text{распространенность})}{(\text{чувствительность} \times \text{распространенность}) + (1 - \text{чувствительность}) \times (1 - \text{распространенность})}$$

Чем чувствительнее тест, тем выше прогностическая ценность его отрицательного результата (т.е. возрастает уверенность врача в том, что отрицательные результаты теста отвергают наличие заболевания). Наоборот, чем специфичнее тест, тем выше прогностическая ценность его положительного результата (т.е. врач может с большей уверенностью считать, что положительные результаты теста подтверждают предполагаемый диагноз). Поскольку распространенность заболевания влияет на прогностическую ценность теста, последняя неизбежно зависит и от условий его применения. Если положительные результаты даже высокоспецифичного теста получены в популяции с низкой вероятностью заболевания, то они окажутся преимущественно ложноположительными. Аналогично, отрицательные результаты высокочувствительного теста, полученные в популяции с высокими шансами наличия заболевания, скорее всего будут ложноотрицательными. Таким образом, интерпретация прогностической ценности положительного или отрицательного результата диагностического теста меняется в зависимости от распространенности заболевания.

Казалось бы, какое отношение имеет распространенность заболевания к отдельному пациенту? Для тех, кто настроен скептически, продемонстрируем, как будет работать тест при крайних значениях распространенности заболевания. Вспомним, что независимо от того, насколько чувствителен и специфичен тест (всегда далекий от совершенства), небольшая доля пациентов немину-

емо классифицируется ошибочно. Представим себе популяцию, в которой ни у кого нет рассматриваемого заболевания. В такой группе все положительные результаты, даже при очень специфичном тесте, будут ложноположительными. Следовательно, когда распространенность заболевания стремится к нулю, прогностическая ценность положительного результата теста тоже стремится к нулю. Наоборот, если данная болезнь есть у каждого в исследуемой популяции, все отрицательные результаты даже высокочувствительного теста окажутся ложноотрицательными. Когда распространенность стремится к 100%, прогностическая ценность отрицательного результата теста стремится к нулю. Убедить скептика в существовании перечисленных связей поможет также анализ рис. 3.2: фиксируя чувствительность и специфичность на постоянном уровне и изменяя распространенность, попробуйте вычислить прогностическую ценность теста.

Влияние распространенности заболевания на прогностическую ценность положительного результата теста для различных высоких уровней чувствительности и специфичности иллюстрирует рис. 3.8. Когда распространенность болезни в исследуемой популяции относительно высока (несколько процентов и более) тест работает хорошо. Однако при низких значениях распространенности прогностическая ценность положительного результата падает почти до нуля и тест практически непригоден. По мере снижения чувствительности и специфичности зависимость прогностической ценности теста от распространенности заболевания усливается.

Пример. Прогностическая ценность теста с ПСА для диагностики рака предстательной железы исследовалась при разных значениях распространенности заболевания. Среди пожилых мужчин, у которых отсутствовали какие-либо симптомы, а распространенность рака предстательной железы оценивалась в 6–12%, лишь приблизительно у 15% с уровнем ПСА 4 мг/дл и более действительно был рак предстательной железы. В группе более высокого риска (с симптомами или вызывающими подозрение результатами пальцевого ректального исследования), где распространенность составляла 26%, рак был обнаружен у 40% мужчин с положительным результатом теста [7]. Иными словами, если бы тест использовался для скрининга в группе мужчин без симптомов заболевания, то на один случай истинного выявления рака пришлось бы 5–6 случаев ошибочного выявления заболевания, причем этих пациентов пришлось бы подвергнуть дополнительным исследованиям, часто включая биопсию. В то же время в группе с выраженной симптоматикой рака предстательной железы, заболевание подтверждается почти у 50% больных, имеющих положительный результат теста при скрининге по уровню ПСА.

Другой пример влияния распространенности болезни на прогностическую ценность положительного результата теста относится к принимаемым в наше время мерам профилактики передачи вируса иммунодефицита человека через препараты крови.

Пример. Донорская кровь подвергается скринингу на антитела к ВИЧ. При определенной точке разделения чувствительность теста равна 97,8%, а специфичность – 90,4%. В 1985 г. прогностическая ценность положительного результата теста была рассчитана исходя из того, что вероятность обнаружения ВИЧ-инфицированной донорской крови не превышает 1 на 10 000. При этих условиях на каждый истинно положительный результат должно быть 9250 ложноположительных [8]. Для того чтобы предотвратить один случай переливания зараженной крови, пришлось бы выбраковывать или подвергать дальнейшему анализу почти 10 000 порций донорской крови! Авторы [8] сделали вывод, что в данном случае "тщательное соблюдение принципов оценки эффективности диагностического теста позволит избавиться от нереалистичных ожиданий".

Однако положение менялось. По мере увеличения распространенности ВИЧ-инфекции в общей популяции повысилась и прогностическая ценность положительного результата теста при скрининге. Через год после выхода в свет упомянутой статьи были опубликованы данные о том, что распространенность инфицированных порций крови составляет 25 на 10 000 среди 67 190 обследованных, следовательно, при тех же значениях чувствительности и специфичности прогностическая ценность положительного результата теста равна 2,5%, т.е. стала гораздо выше, чем несколькими годами раньше [9].

Оценка априорной вероятности заболевания

Как оценить априорную вероятность заболевания у пациента, чтобы вычислить прогностическую ценность того или иного результата теста? Существует несколько источников информации: медицинская литература, архивы медицинских учреждений, мнение специалистов. Хотя итоговая оценка априорной вероятности заболевания редко бывает точной, ошибка, как правило, не настолько велика, чтобы изменить клинические суждения, базирующиеся на этих оценках. Во всяком случае, врачебные решения, безусловно, становятся точнее, если они основаны на формальных оценках, а не на общих представлениях.

В принципе, при определении прогностической ценности результата теста априорная вероятность заболевания более важна, нежели чувствительность и специфичность (см. рис. 3.8). Одна из причин этого заключается в том, что вероятность заболевания обычно варьирует в широких пределах, изменяясь от долей процента до значительных величин в зависимости от возраста, пола,

факторов риска и клинических особенностей пациентов. Для примера достаточно взять различия в вероятности наличия болезни печени у здорового молодого человека, не злоупотребляющего наркотиками или алкоголем, и у желтушного наркомана, применяющего внутривенные наркотики. В настоящее время врачам не нужны тесты, чувствительность и специфичность которых существенно ниже 50%, а если оба эти показателя равны 99%, то тест считается отличным. Другими словами, на практике чувствительность и специфичность редко варьируют более чем в двукратных пределах.

Как повысить априорную вероятность заболевания

Поскольку прогностическая ценность результата теста зависит от распространенности заболевания, становится понятным, почему врачи стремятся применять диагностические тесты в группах больных с повышенной распространенностью подозреваемой болезни. В действительности (см. рис. 3.8) диагностические тесты наиболее эффективны, когда распространенность заболевания не слишком высока и не слишком низка.

Существуют различные способы повышения вероятности заболевания в обследуемой группе до назначения диагностического теста.

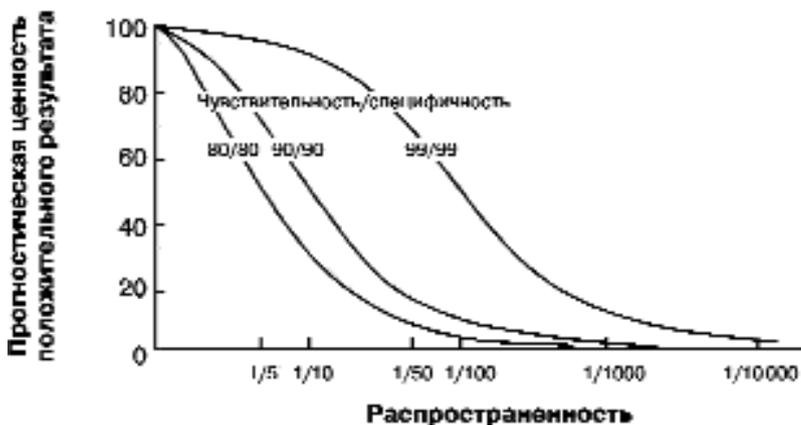


Рис. 3.8. Прогностическая ценность положительного результата теста в зависимости от чувствительности, специфичности и распространенности заболевания.

Процесс предварительного отбора

Хорошо известен такой способ повышения вероятности болезни в обследуемой группе, как направление пациентов на обследование в клинику или отделение неотложной медицинской помощи. Заключение специалистов увеличивает шансы того, что за жалобами пациента действительно кроется серьезное заболевание, что оправдывает более активный подход к проведению диагностических тестов. В практике первичной медицинской помощи, особенно в отсутствие у пациента жалоб, шансы обнаружения болезни значительно ниже, поэтому тесты следует назначать с большей осторожностью.

Пример. Работая в военном госпитале, один из авторов этой книги наблюдал сотни людей с головной болью. Он редко назначал диагностические тесты и не встретил ни одного случая, когда головная боль была обусловлена серьезными причинами. Автор полагает, что серьезные заболевания не были пропущены, поскольку пациенты этой категории постоянно лечились в одном госпитале и долгое время находились под наблюдением. Однако после возвращения на работу в районную больницу автор в течение первой же недели столкнулся со случаем, когда у пациента, обратившегося в отделение неотложной медицинской помощи по поводу головной боли, был обнаружен абсцесс мозжечка.

Поскольку на протяжении своей клинической практики врач порой оказывается в условиях, когда распространенность того или иного заболевания может быть крайне низкой или крайне высокой, он должен иметь в виду, что интенсивность диагностических мероприятий должна соответствовать конкретной ситуации.

Демографические группы

Работая с определенным контингентом больных, врач может повысить эффективность диагностических тестов путем их проведения в демографических группах с повышенным риском заболевания. Так, у 65-летнего мужчины вероятность ишемической болезни сердца в качестве причины атипичной боли в грудной клетке будет в 15 раз выше, чем у 30-летней женщины. Следовательно, ЭКГ-проба с физической нагрузкой (специальный тест для диагностики ИБС) менее эффективна для подтверждения диагноза у молодой женщины, нежели у пожилого мужчины [10]. Аналогичным образом прогностическая ценность положительного результата теста на серповидно-клеточную анемию выше у представителей негроидной расы, чем у европеоидов.

Особенности клинической ситуации

Решение о назначении диагностического теста в существенной мере зависит от особенностей клинической ситуации. Вероятность выявления болезни повышается или снижается в зависимости от наличия или отсутствия симптомов и факторов риска. Например, диагноз ИБС у женщины с болями в грудной клетке более вероятен, если пациентка страдает типичной стенокардией, она курит и у нее повышено артериальное давление. В подобных случаях проведение ЭКГ-пробы с физической нагрузкой выявит ИБС с большей вероятностью, чем у лиц с жалобами на неспецифические боли в грудной клетке, и без факторов риска ИБС.

Большинство врачей интуитивно осознают, что ценность диагностического теста повышается, если вероятность данной болезни у пациента велика. Однако при возрастающей доступности диагностических тестов соблазнительнее менее избирательный подход к их назначению. Нужно помнить, что чем менее избирательно врач подходит к назначению теста, тем ниже и распространенность заболевания, и прогностическая ценность положительного результата теста.

Роль этого эффекта может быть гораздо важнее, чем полагают многие.

Пример. На рис. 3.9 изображено влияние некоторых факторов на интерпретацию результатов ЭКГ-теста с физической нагрузкой. Прогностическая ценность положительного результата этого теста для диагностики ИБС может изменяться от 1,7 до 99,8% в зависимости от возраста, симптомов и выраженности изменений на ЭКГ. Так, снижение сегмента ST на 1 мм во время нагрузки у 35-летнего пациента, не имеющего симптомов, будет ложноположительным в 98% случаев и более. Такие же результаты теста у 60-летнего пациента с типичной стенокардией в анамнезе будут означать наличие ИБС в 90% случаев и более [10].

Учитывая возможность такого эффекта, врачи должны по-разному интерпретировать сходные результаты тестов в различных ситуациях. Отрицательный результат при пробе с физической нагрузкой у 35-летнего пациента без каких-либо симптомов только подтверждает и без того малую вероятность ИБС, а положительный результат у такого обследуемого (бегуна, пилота, бизнесмена), наоборот, введет в заблуждение. Совсем иначе обстоит дело при обследовании 65-летнего мужчины с типичной стенокардией, когда тест скорее поможет подтвердить, а не отвергнуть болезнь. Тест полезнее всего использовать в промежуточных ситуациях, когда априорная вероятность заболевания не слишком высока и не слишком низка.

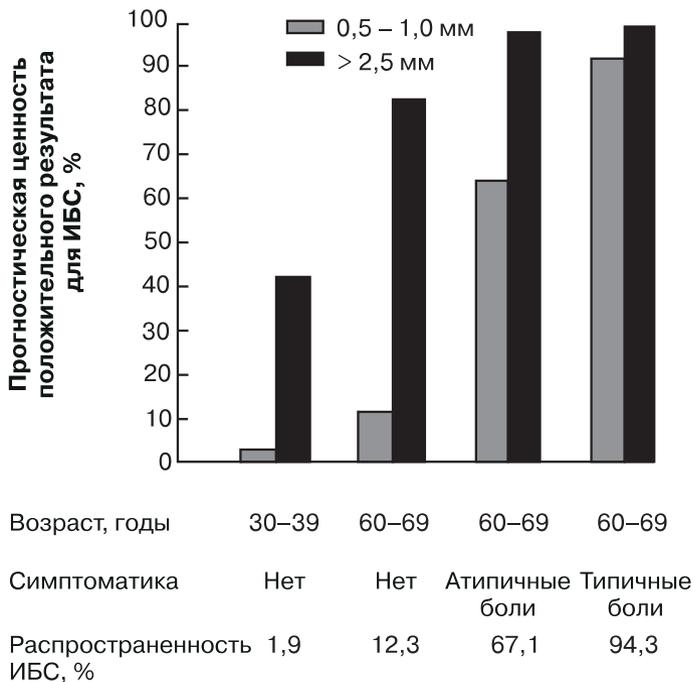


Рис. 3.9. Влияние распространенности заболевания на прогностическую ценность положительного результата теста. Вероятность наличия ишемической болезни сердца (ИБС) у мужчин в зависимости от возраста, симптомов и снижения сегмента *ST* на ЭКГ при пробе с физической нагрузкой [Diamond G.A., Forrester J.S. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. N Engl J Med 1979;300:1350–1358].

Например, у 60-летнего мужчины с атипичной болью в области сердца вероятность ИБС составляет 67% до проведения пробы с физической нагрузкой и 99% после пробы, показавшей снижение сегмента *ST* на 2 мм и более.

Поскольку диагностическая эффективность теста сильно зависит от распространенности заболевания, врач должен оценить вероятность болезни до назначения теста. До недавнего времени при оценке априорной вероятности заболевания врачи опирались лишь на клинические наблюдения и собственный опыт. Теперь благодаря наличию больших электронных банков клинических данных есть возможность количественно оценивать вероятность заболевания при различных сочетаниях клинических признаков [11].

Замечания по поводу медицинской литературы

Публикуемые описания диагностических тестов часто включают в себя не только показатели чувствительности и специфичности, но и интерпретацию положительных и отрицательных результатов теста, т.е. прогностическую ценность. Это следует приветствовать, поскольку такая информация действительно нужна клиницистам. Однако подобные исследования чаще всего проводятся в университетских клиниках, среди пациентов, у которых распространенность серьезных заболеваний относительно высока. Вследствие этого имеющиеся в медицинской литературе выводы о прогностической ценности тестов могут оказаться ошибочными, если применить тот же тест в менее рафинированных условиях. Еще хуже то, что некоторые авторы оценивают эффективность теста, проводя его в группе заведомо больных и в равной ей по численности группе заведомо здоровых. Это неплохой способ определения чувствительности и специфичности. Однако прогностическая ценность положительного результата, определенная в таких исследованиях, ни о чем не говорит, поскольку она определена в популяции пациентов с искусственно заданной авторами распространенностью заболевания на уровне 50%.

Отношения правдоподобия

Эффективность диагностического теста можно характеризовать через отношения правдоподобия, которые обобщают ту же информацию, что и показатели чувствительности и специфичности, и могут использоваться для вычисления вероятности болезни на основании положительного или отрицательного результата теста.

Шансы

Поскольку использование отношений правдоподобия связано с понятием шансов, необходимо сначала выяснить, чем шансы отличаются от вероятности. Вероятность, используемая для выражения чувствительности, специфичности и прогностической ценности результатов теста, соответствует доле испытуемых, обладающих некоторой характеристикой, например положительным результатом теста. В свою очередь шансы – это отношение двух вероятностей. Шансы и вероятность содержат одну и ту же информацию, но по-разному выражают ее. Одно может быть преобразовано в другое с помощью двух простых формул:

$$\text{Шансы события} = (\text{вероятность события}) / (1 - \text{вероятность события})$$

$$\text{Вероятность события} = (\text{шансы события}) / (1 + \text{шансы события}).$$

Эти понятия должны быть знакомы большинству читателей, поскольку ими пользуются в обыденных ситуациях. Например, можно сказать, что шансы баскетбольной команды "Сиетл Суперсоник" выиграть сегодняшний матч составляют 4:1, или же, что вероятность их выигрыша 80%.

Определения понятий

Отношением правдоподобия (likelihood ratio) для конкретного результата диагностического теста называется отношение вероятности данного результата у лиц с заболеванием к вероятности этого же результата у лиц без заболевания. Отношение правдоподобия показывает, во сколько раз выше (или ниже) вероятность получить данный результат теста у больных, нежели у здоровых. Если тест дихотомический (положительный–отрицательный), то его способность различать больных и здоровых соответствует двум типам отношений правдоподобия: один тип связан с положительным результатом теста, другой – с отрицательным (см. рис. 3.2).

В примере со стрептококковой ангиной (см. рис. 3.3) можно рассчитать отношения правдоподобия для стрептококковой ангины при положительном и отрицательном результате клинического обследования. Положительный результат клинического обследования примерно в 2,5 раза более вероятен при наличии стрептококковой ангины, чем при ее отсутствии. Отношение правдоподобия для отрицательного результата (вывода об отсутствии болезни на основании клинической картины) равно 0,39. Соответственно шансы поставить отрицательный клинический диагноз при наличии стрептококковой ангины и при ее отсутствии составляют 1 : 2,6.

Использование отношений правдоподобия

Претестовую, или априорную, вероятность (распространенность) можно преобразовать в претестовые шансы, применив формулу, приведенную выше. Затем, используя отношения правдоподобия, можно преобразовать претестовые шансы в посттестовые (апостериорные) с помощью следующей формулы:

$$\text{Претестовые шансы} \times \text{отношение правдоподобия} = \text{посттестовые шансы.}$$

Посттестовые шансы в свою очередь можно снова преобразовать в вероятности с помощью формулы, уже приводившейся в этой главе. В этих соотношениях претестовые шансы содержат ту же информацию, что и априорная вероятность (распространен-

ность), отношения правдоподобия – ту же информацию, что и чувствительность и специфичность, а посттестовые шансы – ту же информацию, что и прогностическая ценность положительного результата теста (посттестовая, или апостериорная, вероятность).

Главное преимущество отношений правдоподобия состоит в том, что они помогают выйти за рамки грубой классификации результатов теста (либо норма, либо патология), с которой мы сталкиваемся, если описываем точность диагностического теста только в терминах чувствительности и специфичности при единственной точке разделения. Очевидно, наличие заболевания более вероятно при крайнем отклонении результата теста от нормы, чем в случае результата, близкого к границе нормы. Отношения правдоподобия позволяют обобщать информацию, содержащуюся в различных результатах теста. Эти отношения можно определять для любого количества результатов теста по всему диапазону допустимых значений. При таком подходе мы получаем информацию о степени отклонения от нормы, а не только о факте наличия или отсутствия болезни. При вычислении отношений правдоподобия внутри некоторого диапазона значений результатов теста под чувствительностью понимается наша способность использовать конкретный результат теста для идентификации лиц с заболеванием, а не с той или иной степенью отклонения от нормы. То же относится и к специфичности.

Таким образом, показатель отношения правдоподобия отвечает разумной клинической практике, когда при оценке вероятности (или шансов) изучаемого заболевания больший вес придается крайне высоким (или низким) результатам теста, а не пограничным между нормой и патологией.

Пример. Насколько точна диагностика гипотиреоза на основании единственного теста – измерения уровня тироксина (T_4) сыворотки? Такой вопрос был поставлен при обследовании 120 амбулаторных пациентов с подозрением на гипотиреоз [12]. Диагноз гипотиреоза ставился, если уровень тиреотропного гормона (ТТГ) в сыворотке был повышен и если этот диагноз подтверждался данными последующего наблюдения с использованием других методов изучения функции щитовидной железы и результатами лечения. Сравнения исходный уровень T_4 у 27 пациентов с гипотиреозом и у 93 пациентов без гипотиреоза, авторы определяли может ли этот простой метод служить самостоятельным тестом для диагностики гипотиреоза.

Как и ожидалось, отношения правдоподобия для гипотиреоза были наибольшими при низких уровнях T_4 и наименьшими – при высоких (табл. 3.2). Наименьшие значения T_4 (менее 4,0 мкг/дл¹) обнаружены только у больных гипотиреозом, т.е. такие находки под-

¹ В Международной системе единиц (СИ) нмоль/л, коэффициент пересчета 12,871.

тверждают диагноз наверняка. Самые высокие значения T_4 (более 8,0 мкг/дл) у больных гипотиреозом вообще не наблюдались, т.е. они отвергают заболевание.

Авторы сделали вывод, что "можно достичь значительной экономии средств без потери диагностической точности, если применять для первичной оценки пациентов с подозрением на гипотиреоз только измерение уровня T_4 в сыворотке."

Отношение правдоподобия в качестве характеристики эффективности теста имеет еще несколько преимуществ перед чувствительностью и специфичностью. Оно позволяет выразить получаемую информацию одним числом вместо двух и облегчает расчет посттестовых шансов на основе претестовых. Особенно удобно применять отношения правдоподобия для описания общей вероятности заболевания, когда диагностические тесты применяются последовательно (см. далее).

Таблица 3.2

Распределение концентраций тироксина в сыворотке у пациентов с гипотиреозом и без него*

Общий тироксин сыворотки, мкг/дл	Результаты теста		Отношение правдоподобия
	число больных, %	число здоровых, %	
<1,1	2 (7,4)		↑
1,1 – 2,0	3 (11,1)		↑ Диагноз подтвержден
2,1 – 3,0	1 (3,7)		↓
3,1 – 4,0	8 (29,6)		↓
4,1 – 5,0	4 (14,8)	1 (1,1)	13,8
5,1 – 6,0	4 (14,8)	6 (6,5)	2,3
6,1 – 7,0	3 (11,1)	11 (11,8)	0,9
7,1 – 8,0	2 (7,4)	19 (20,4)	0,4
8,1 – 9,0	–	17 (18,3)	↑
9,1 – 10,0	–	20 (21,5)	↑
10,1 – 11,0	–	11 (11,8)	↑ Диагноз исключен
11,1 – 12,0	–	4 (4,3)	↓
>12	–	4 (4,3)	↓
Всего	27 (100)	93 (100)	↓

* [Goldstein B.J., Mushin A.I. Use of a single thyroxine test to evaluate ambulatory medical patients for suspected hypothyroidism. J Gen Intern Med 1987;2:20–24].

Понятие "отношение правдоподобия" имеет и свои недостатки. Оно заставляет пользоваться шансами вместо более привычных для нас вероятностей. Кроме того, преобразование вероятностей в шансы и наоборот требует математических расчетов или использования номограмм, что нарушает относительную простоту определения посттестовых шансов по отношениям правдоподобия. Наконец, при тестах с большим разбросом результатов расчет отношений правдоподобия основан на показателях чувствительности и специфичности, отличающихся от обычных.

Применение нескольких диагностических тестов

Поскольку врачи обычно применяют несовершенные тесты, чувствительность и специфичность которых ниже 100%, вероятность заболеть при использовании только одного теста часто определяется как не очень высокая и не очень низкая, предположим, между 10 и 90%. Как правило, получив такой результат, мы не можем прекратить диагностический поиск. Ведь ни врача, ни пациента не удовлетворит заключение о 20% вероятности рака толстой кишки или о наличии 42% вероятности наличия ИБС у 35-летнего больного без каких-либо симптомов, но со снижением сегмента *ST* на 2,5 мм при пробе с физической нагрузкой (см. рис. 3.9). Даже для менее опасных заболеваний, например гипотиреоза, промежуточные значения посттестовых вероятностей не слишком информативны. В подобных ситуациях врач старается значимо повысить или понизить вероятность заболевания, разумеется, за исключением ряда ситуаций: когда нет диагностических возможностей для уточнения результата, когда больному независимо от результата невозможно помочь или когда продолжение диагностического поиска связано с дополнительным риском. Во всех остальных случаях врач будет стремиться продолжить обследование.

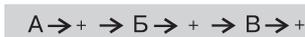
Когда использовано несколько тестов, и все они положительны или все отрицательны, их смысл очевиден. Гораздо чаще бывает так, что результаты одних тестов положительны, а других – отрицательны. Тогда толкование усложняется. В этом разделе мы рассмотрим принципы назначения и интерпретации результатов нескольких тестов.

Существует два способа применения нескольких тестов (рис. 3.10): *параллельный* (несколько тестов одновременно, причем положительный результат любого из них рассматривается в пользу наличия болезни) и *последовательный* с учетом результатов предыдущего теста. При последовательном подходе для установления диагноза все тесты должны дать положительный ре-

Тактика	Последовательность применения тестов	Результаты
----------------	---	-------------------

Последовательное применение тестов

Результаты теста А и теста Б и теста В положительны



↓ Чувствительность
↑ Специфичность

Параллельное применение тестов

Результаты теста А или теста Б или теста В положительны



↑ Чувствительность
↓ Специфичность

Рис. 3.10. Последовательное и параллельное применение нескольких диагностических тестов.

зультат, поскольку в случае отрицательного результата диагностический поиск прекращается.

Параллельное применение нескольких диагностических тестов

Несколько тестов назначаются параллельно, когда необходима быстрая оценка состояния, например у госпитализированных больных, при неотложных состояниях или же у амбулаторных пациентов, прибывших для обследования на короткое время.

Несколько тестов, проведенных параллельно, обеспечивают, как правило, более высокую чувствительность, а следовательно, и большую прогностическую ценность отрицательного результата при данной распространенности заболевания, чем каждый тест в отдельности. С другой стороны, снижаются специфичность и прогностическая ценность положительного результата теста. Таким образом, уменьшается вероятность того, что заболевание будет пропущено (применение параллельного тестирования – одна из возможных причин, в силу которой в специализированных центрах часто диагностируют заболевания, не выявленные врачами в

поликлинике), но одновременно возрастает вероятность ложноположительных диагнозов (в специализированных центрах отмечается некоторая тенденция к гипердиагностике). Степень повышения чувствительности и прогностической ценности отрицательного результата теста зависят от того, в какой мере больные, не выявленные с помощью одних тестов, будут выявлены с помощью других. Например, при использовании двух тестов с чувствительностью 60 и 80% соответственно, общая чувствительность двух тестов, использованных параллельно, составит 80%, если более чувствительный тест идентифицирует все случаи, выявляемые и менее чувствительным тестом. Если же каждый из двух тестов обнаруживает все случаи, пропускаемые другим, то чувствительность параллельного использования обоих тестов равна 100%. В случае полной взаимной независимости двух тестов чувствительность параллельного использования двух тестов составит 92%.

Параллельное использование нескольких тестов особенно полезно в ситуации, когда нужен очень чувствительный тест, но реально вместо этого имеется несколько относительно малочувствительных тестов, оценивающих разные клинические феномены. Благодаря параллельному использованию нескольких тестов повышается общая чувствительность диагностического процесса. Платой за такое повышение чувствительности является обследование или лечение некоторого числа пациентов, у которых изучаемое заболевание отсутствует.

Пример. Определение уровня ПСА и пальцевое ректальное исследование – не очень чувствительные тесты для диагностики рака предстательной железы [7]. В табл. 3.3 показаны их чувствительность и специфичность, а также прогностическая ценность положительного результата при скрининге (т.е. в популяции мужчин, у которых отсутствуют симптомы заболевания). При параллельном применении обоих тестов повышается чувствительность, но снижается специфичность; прогностическая ценность положительного результата в этом случае ниже, чем при диагностике только по уровню ПСА.

Последовательное применение нескольких диагностических тестов

Такая тактика обычно предпочтительна в клинических ситуациях, не требующих быстрой оценки состояния больного, например в амбулаторной практике. Кроме того, последовательное применение диагностических тестов целесообразно, если встает вопрос о дорогостоящем или рискованном исследовании – такой тест обычно назначается только после положительных результатов применения более простых и неинвазивных методов исследо-

Таблица 3.3
Характеристики теста определения ПСА
и пальцевого ректального исследования*

Тест	Чувствительность	Специфичность	Прогностическая ценность положительного результата
ПСА 4,0 мкг/мл	0,67	0,97	0,43
Патология при пальцевом ректальном исследовании	0,50	0,94	0,24
Повышенный уровень ПСА или патология при пальцевом ректальном исследовании	0,84	0,92	0,28
Повышенный уровень ПСА и патология при пальцевом ректальном исследовании	0,34	0,995	0,49

* ПСА и пальцевое ректальное исследование применялись раздельно и в сочетании (параллельно и последовательно) для диагностики рака предстательной железы [Kramer B.S. et al. Ann Int Med 1993; 119:914–923].

вания. Например, риск рождения ребенка с синдромом Дауна сначала оценивается по возрасту матери и результатам анализов крови на альфа-фетопротеин, хорионический гонадотропин и эстриол и лишь затем беременным из группы высокого риска предлагается амниоцентез [13]. Последовательное применение тестов по сравнению с параллельным уменьшает объем лабораторной работы – ведь каждый последующий тест учитывает результаты предыдущего. Вместе с тем последовательное применение тестов требует больше времени, так как очередная диагностическая процедура назначается только после получения результатов предыдущей.

При последовательном применении тестов повышаются специфичность и прогностическая ценность положительного результата, но снижаются чувствительность и прогностическая ценность отрицательного результата (см. табл. 3.3). В итоге возрастает уверенность врача в том, что положительный результат теста подтверждает наличие предполагаемой болезни, но одновременно возрастает риск пропустить заболевание. Последовательное применение тестов особенно полезно, когда ни один из имеющихся в распоряжении методов диагностики не обладает высокой специфичностью.

Если врач собирается применить два теста последовательно, то эффективнее назначить сначала тест с большей специфичностью. В табл. 3.4 показано, какую роль играет последовательность применения двух тестов. Тест А специфичнее, чем тест Б, тогда как Б – чувствительнее, чем А. Использование теста А первым позволит подвергнуть обоим тестам меньшее число пациентов, хотя в итоге число пациентов с установленным диагнозом будет одним и

тем же, независимо от последовательности применения тестов. Однако, если один тест значительно дешевле другого или сопряжен с меньшим риском для пациента, то вероятно, разумнее вначале назначить именно его.

Предположение о независимости результатов нескольких диагностических тестов

При использовании нескольких тестов, что было описано выше, точность результата определяется тем, в какой мере данные, получаемые при последующей диагностической процедуре, не зависят от результатов предыдущих тестов (последующие тесты не должны дублировать уже известную информацию). Эта предпосылка лежит в основе рассмотренного ранее подхода к определе-

Таблица 3.4

Эффект последовательности применения тестов: "А после Б" и "Б после А" *

<i>Распространенность заболевания</i>									
Число пациентов, прошедших обследование		1000							
Число больных		200 (распространенность 20%)							
<i>Чувствительность и специфичность тестов</i>									
Тест	Чувствительность				Специфичность				
А	80				90				
Б	90				80				
<i>Последовательность применения тестов</i>									
тест А – первый					тест Б – первый				
Заболевание					Заболевание				
+					+				
–					–				
А	+	160	80	240	В	+	180	160	340
	–	40	720	760		–	20	640	660
		200	800	1000			200	800	1000
240 пациентам выполнен второй тест Б					340 пациентам выполнен второй тест А				
+					+				
–					–				
Б	+	144	16	160	А	+	144	16	160
	–	16	64	80		–	46	144	180
		160	80	240			180	160	340

* Обратите внимание на то, что независимо от последовательности проведения тестов А и Б было получено одно и то же число пациентов с диагностированным заболеванием (160), равно как и число истинно положительных результатов оказалось одинаковым (144). Однако, когда тест А (более специфичный) использовался первым, второй тест проводился меньшему числу пациентов; более низкая чувствительность теста А не ухудшила конечный результат.

нию прогностической ценности теста. Однако для большинства заболеваний полная взаимная независимость тестов маловероятна. Если же эта предпосылка неверна, то при вычислении вероятности заболевания по результатам нескольких тестов возникает тенденция к завышению их прогностической ценности.

Отношения правдоподобия при последовательном применении нескольких тестов

При последовательном применении нескольких тестов можно рассчитать общую вероятность заболевания на основании отношений правдоподобия для результата каждого теста, как это показано на рис. 3.11. Сначала исходный показатель распространенности заболевания преобразуется в претестовые шансы. По мере выполнения каждого очередного теста его посттестовые шансы становятся претестовыми шансами для следующего теста. В итоге вычисляется новый показатель вероятности заболевания – апостериорная вероятность, которая учитывает информацию, полученную во всех последовательных тестах.

Способность теста улавливать изменения измеряемого параметра

Состояние больных постоянно изменяется под действием лечения, вследствие заболевания или старения. Врачи регулярно сталкиваются с вопросом: "Улучшилось или ухудшилось состояние моего пациента?" Тесты, применяемые для наблюдения за течением болезни (например, для оценки изменений тяжести симптомов, функционального состояния), часто весьма отличны от используемых в целях диагностики, но подходы к определению эффективности и тех, и других практически одинаковы.



Рис. 3.11. Использование отношений правдоподобия (ОП) при последовательном применении тестов.

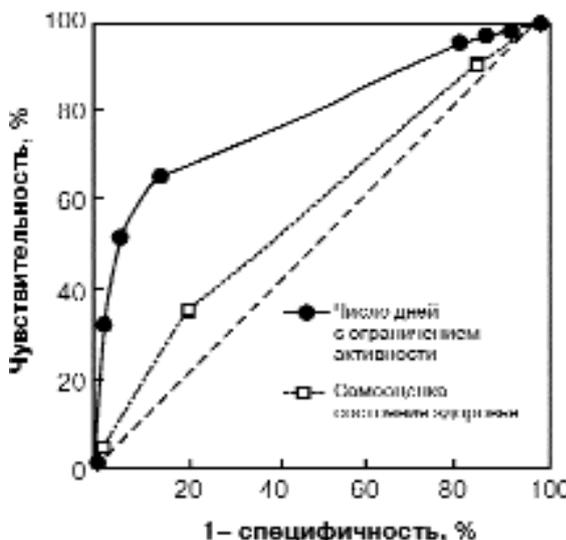


Рис. 3.12. Характеристики двух опросников по оценке состояния здоровья. Способность различать пациентов с серьезными заболеваниями и без них у пожилых пациентов [Wagner E.H. et al. J Am Geriatr Soc 1993;41:241–248].

Способность теста правильно отражать динамику состояния концептуально связано с представлением о достоверности диагностического теста, с той только разницей, что в этом случае в качестве "золотого стандарта" принимается наличие или отсутствие выраженных изменений клинического состояния, а не наличие или отсутствие заболевания. Способность теста реагировать на изменение изучаемого состояния выражается через чувствительность, специфичность и прогностическую ценность или вычисляется как площадь под характеристической кривой.

Пример. Для оценки здоровья населения и эффектов лечения обычно используются различные опросники, которые заполняются респондентами самостоятельно. В одном из таких опросников спрашивается о числе дней, в которые нормальная жизнедеятельность респондента была ограничена болезнью или травмой. В другом — напротив, респондента просят оценить в целом состояние своего здоровья по сравнению со здоровьем других людей того же возраста, выбрав ответ на шкале от "отлично" до "плохо". Характеристики этих тестов при опросах с интервалом в 1 год определялись путем сопоставления изменений каждого из указанных показателей в двух группах пациентов — перенесших и не перенесших в течение этого года тяжелые заболевания [14]. Характеристические кривые на рис. 3.12 показывают, что изменения самооценки состояния здоро-

вья для тяжело болевших почти не отличаются от случайных, тогда как изменение числа дней с ограничением активности – гораздо более полезный тест, так как площадь под его характеристической кривой составляет 80%.

Резюме

Для оценки эффективности диагностического теста строят четырехпольную таблицу и сравнивают результаты теста при наличии и в отсутствие заболевания. Все четыре поля таблицы должны быть заполнены. Когда чувствительность и специфичность нового диагностического теста оцениваются по опубликованным данным, должен указываться "золотой стандарт", относительно которого определяется точность теста. Испытуемые как с заболеванием, так и без него, отбираемые для оценки эффективности теста, не должны отличаться от пациентов, у которых этот тест будет применяться на практике. Кроме того, знание окончательного диагноза не должно влиять на интерпретацию результатов теста и, наоборот. При изменении точки разделения между нормой и патологией изменяются чувствительность и специфичность. Дополнительным способом описания точности диагностического теста служит отношение правдоподобия.

Прогностическая ценность теста – наиболее адекватная характеристика для интерпретации его результатов. Она определяется не только чувствительностью и специфичностью теста, но и распространенностью заболевания, которая неодинакова в разных условиях. Обычно, для того чтобы получить достаточно надежный диагноз, приходится использовать несколько тестов параллельно или последовательно. Способность теста обнаруживать изменения клинического состояния оценивается с помощью такой же четырехпольной таблицы.

Литература

1. Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure. The fifth report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (JNC V). Arch Intern Med 1993;153:154–183.
2. Catalona W.J. et al. Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med 1991;324:1156–1161.
3. Jensen M.C., Brant-Zawadzki M.N., Obuchowski N., Modic M.T., Malkasian D., Ross J.S. Magnetic resonance imaging of the lumbar spine in people without back pain. N Engl J Med 1994;331:69–73.
4. Bartrum R.J. Jr., Crow H.C., Foote S.R. Ultrasonic and radiographic cholecystography. N Engl J Med 1977;296:538–541.
5. Jones T.V., Lindsey B.A., Yount P., Soltys R., Farani-Enayat B. Alcoholism screening questionnaires: are they valid in elderly medical outpatients? J Gen Intern Med 1993;8:674–678.

6. Fletcher R.H. Carcinoembryonic antigen. *Ann Intern Med* 1986; 104:66–73.
7. Voss J.D. Prostate cancer, screening and prostate-specific antigen: promise or peril? *J Gen Intern Med* 1994;9:468–474.
8. Barry M.J., Mulley A.G., Singer D.E. Screening for HTLV III antibodies: the relation between prevalence and positive predictive value and its social consequences. *JAMA* 1985;253:3395.
9. Ward J.W., Grindon A.J., Feorino P.M., Schable C., Parvin M., Alien J.R. Laboratory and epidemiologic evaluation of an enzyme immunoassay for antibodies to HTLV III. *JAMA* 1986;256:357–361.
10. Diamond G.A., Forrester J.S. Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. *N Engl J Med* 1979;300:1350–1358.
11. Tiemey W.M., McDonald C.J. Practice databases and their uses in clinical research. *Stat Med* 1991;10:541–557.
12. Goldstein B.J., Mushlin A.I. Use of a single thyroxine test to evaluate ambulatory medical patients for suspected hypothyroidism. *J Gen Intern Med* 1987;2:20–24.
13. Haddow J.E., Palomaki G.E., Knight G.J., Williams J., Pulkkinen A., Canick J.A., Sailer D.N. Jr., Bowers G.B. Prenatal screening for Down's syndrome with use of maternal serum markers. *N Engl J Med* 1992;327:588–593.
14. Wagner E.H., LaCroix A.Z., Grothaus L.C., Hecht J.A. Responsiveness of health status measures to change among older adults. *J Am Geriatr Soc* 1993;41:241–248.

Рекомендуемая литература

- Cebul R.D., Beck L.H. Teaching clinical decision making. New York: Praeger 1985.
- Department of Clinical Epidemiology and Biostatistics, McMaster University. Interpretation of diagnostic data. V. How to do it with simple math. *Can Med Assoc J* 1983;129:22–29. (Reprinted in Sackett D.L., Haynes R.B., Tugwell P. eds. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little, Brown 1985.)
- Pagan T.J. Nomogram for Bayes' theorem (Letter). *N Engl J Med* 1975;293:257.
- Griner P.F., Mayewski R.J., Mushlin A.I., Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med* 1981;94:557–600.
- Griner P.F., Panzer R.J., Greenland P. Clinical diagnosis and the laboratory: logical strategies for common medical problems. Chicago: Year Book, 1986.
- McNeil B.J., Abrams H.L. eds. Brigham and Women's Hospital handbook of diagnostic imaging. Boston: Little, Brown, 1986.
- Pauker S.G., Kassirer J.P. Clinical application of decision analysis: a detailed illustration. *Semin Nucl Med* 1978;8:324–335.
- Sheps S.B., Schechter M.T. The assessment of diagnostic tests. A survey of current medical research. *JAMA* 1984;252:2418–2422.
- Sox H.C. Jr. Probability theory in the use of diagnostic tests. An introduction to critical study of the literature. *Ann Intern Med* 1986;104:60–66.
- Sox H.C. Jr., Blatt M., Higgins M.C., Marton K.I. Medical decision making. Kent, UK: Butterworth, 1987.
- Wasson J.H., Sox H.C. Jr., Neff R.K., Goldman L. Clinical prediction rules, applications and methodological standards. *N Engl J Med* 1985;313:793–799.
- Weinstein M.C., Fineberg H.V., Elstein A.S., Frazier H.S., Neuhauser D., Neutra R.R., McNeil B.J. Clinical decision analysis. Philadelphia: WB Saunders 1980.

4

ЧАСТОТА СОБЫТИЙ

В главе 1 был очерчен круг основных вопросов, встающих перед врачом в процессе его практической работы. В этой главе будут даны определения и описания количественных показателей, которыми врачи руководствуются, принимая диагностические и терапевтические решения.

Пример. Мужчина 72 лет жалуется на постепенно нарастающее учащение мочеиспускания, вялую и прерывистую струю мочи. При пальцевом ректальном исследовании выявлено симметричное увеличение предстательной железы. Диурез снижен. Уровень простатспецифического антигена в сыворотке находится в пределах нормы. Поставлен диагноз аденомы предстательной железы. Следующий вопрос – какое лечение выбрать: консервативное или оперативное? Врач и больной должны взвесить достоинства и недостатки каждого из этих методов, принимая во внимание, что при консервативном лечении имеется риск ухудшения течения заболевания или развития обструктивной нефропатии, а в случае оперативного вмешательства – риск смерти или нарушения половой функции.

Традиционно решения принимаются на основе "клинического опыта" – навыка, приобретаемого в процессе практической деятельности. В последние годы в медицине стали применяться количественные методы принятия клинических решений. Выбор медицинской тактики по сути основан на применении методов анализа принятия решений, а также методов анализа экономической целесообразности и затрат. При применении этих методов используют количественные данные о частоте клинических событий и их последствий для больного. Методы, более подробно описанные в конце этой главы, имеют ограниченное применение и могут использоваться лишь для оценки вероятностей или частот клинических исходов, описанных в этой книге.

Что касается больного с аденомой предстательной железы, то для обоснованного выбора лечения требуется точная информация, с одной стороны, о вероятности ухудшения состояния, развития острой задержки мочи или нефропатии при консервативном лечении, с другой стороны – о вероятности улучшения состояния и развития осложнений (импотенции, ретроградной эякуляции,

смерти) при хирургическом вмешательстве. Именно таков характер сведений, необходимых для получения ответа на большинство клинических вопросов. Решения определяются вероятностью тех или иных исходов при различных обстоятельствах, например, при положительном или отрицательном результате диагностического теста или в случае применения разных способов лечения. Поскольку вероятность болезни, улучшения или ухудшения состояния, излечения или смерти составляет основу для ответов на большинство клинических вопросов, в этой главе будут рассмотрены методы оценки частоты клинических состояний и исходов.

Количественное выражение вероятностных утверждений

Врачи предпочитают выражать вероятность не числами, а словами – “обычно”, “иногда”, “редко” и т.п. Слова понятнее, чем числа, кроме того, они позволяют избежать точных утверждений в тех случаях, когда вероятность трудно определить однозначно. Однако общеупотребительным словам, обозначающим частоту событий, сплошь и рядом придается совершенно разный смысл.

Пример. Врачей попросили оценить вероятность заболевания, соответствующую каждому из 30 слов, выбранных при просмотре рентгенологических заключений и результатов лабораторных исследований. Мнения относительно каждого слова сильно различались. Так, выражение “возможно, что” соответствовало вероятности от 0,18 до 0,98; “маловероятно” – от 0,01 до 0,93. Авторы этого исследования сделали вывод, что “различие мнений разных врачей относительно тактики лечения отражает различие в том, какой смысл они придают словам, определяющим вероятность” [1].

Больные в свою очередь тоже расходятся в терминах, которыми они обозначают вероятность или частоту отдельных симптомов. В другом исследовании при опросе лиц, занимающихся интеллектуальным трудом, оказалось, что они приписывают слову “обычно” вероятность от 0,35 до 1,0 (± 2 стандартных отклонения), а слову “редко” – от 0 до 0,15 [2].

Итак, замена числовых значений словами затрудняет передачу информации. Авторы данной книги выступают за использование чисел везде, где это возможно.

Восприятие частоты

Восприятие вероятности клинических событий и исходов зависит от личного опыта каждого врача. Оцениваемая врачом веро-

ятность конкретного исхода повышается, если подобный случай недавно был в его практике. И наоборот, частота событий недооценивается, если они редко встречаются в практике врача или же больные предпочитают не говорить о них. Например, результаты систематических опросов больных после трансуретральной резекции предстательной железы показывают, что более 50% больных страдают ретроградной эякуляцией [3]. Большинство урологов оценивают этот симптом как гораздо более редкий, поскольку многие больные неохотно обсуждают свои сексуальные проблемы.

Распространенность и частота новых случаев

В принципе клинически важным показателем частоты, или вероятности, события служит дробь, где в числителе – число пациентов, у которых наблюдается интересующее нас состояние или определенный исход, а в знаменателе – число всех обследованных.

По сути такая дробь является пропорцией, но обычно ее называют частотой. Клиницисты имеют дело с двумя показателями частоты – распространенностью и частотой новых случаев. *Распространенность* (prevalence) – это отношение числа лиц, у которых наблюдается изучаемое состояние, ко всем обследованным за определенный промежуток времени. Распространенность может оцениваться как число лиц, имеющих определенное заболевание на момент обследования, или как число всех случаев выявления заболевания за какой-то период времени, например, в течение одного года.

Частота новых случаев (incidence) – это отношение числа лиц, у которых в течение определенного времени развилось интересующее нас состояние, ко всем обследованным в группе, где исходно этого состояния никто не имел. Частота новых случаев болезни в популяции, где это заболевание исходно отсутствовало, называется *заболеваемостью*. Частота новых случаев исходов при определенном заболевании указывает соответственно на инвалидность или летальность при этом заболевании. Как будет описано далее в этой главе и более подробно в главе 5, для оценки частоты новых случаев отбирают группу лиц без заболевания (или изучаемого исхода) и периодически обследуют ее, подсчитывая число новых случаев, появившихся за этот период.

Пример. Продемонстрировать различия между распространенностью и заболеваемостью поможет рис. 4.1, где отмечены случаи появления болезни в группе из 100 человек за 3 года (1992–1994). С течением времени у обследуемых возникает заболевание и это со-

стояние сохраняется до тех пор, пока не наступит выздоровление или смерть. В течение 3 лет болезнь появилась у 16 человек, а у 4 она имела место ранее. Остальные 80 человек не заболели, поэтому на рисунке они не обозначены.

В начале 1992 г. зарегистрировано 4 случая заболевания, так что распространенность в этот момент времени составила 4 на 100. Если все 100 человек, включая заболевших, обследовать в начале каждого года, можно рассчитать распространенность на соответствующий момент. В начале 1993 г. распространенность составила 5 на 100, так как 2 случая из предшествующего 1992 г. перешли в 1993 г., а 2 случая из развившихся в 1992 г. закончились (будем надеяться, что больные успешно вылечились) прежде чем началось обследование в 1993 г. Можно рассчитать распространенность и для каждого из двух последующих ежегодных обследований; если предположить, что ни один из 100 исходных пациентов не умер, не переехал и не отказался от обследования, то распространенность составит 7 на 100 человек в начале 1994 г. и 5 на 100 – в начале 1995 г.

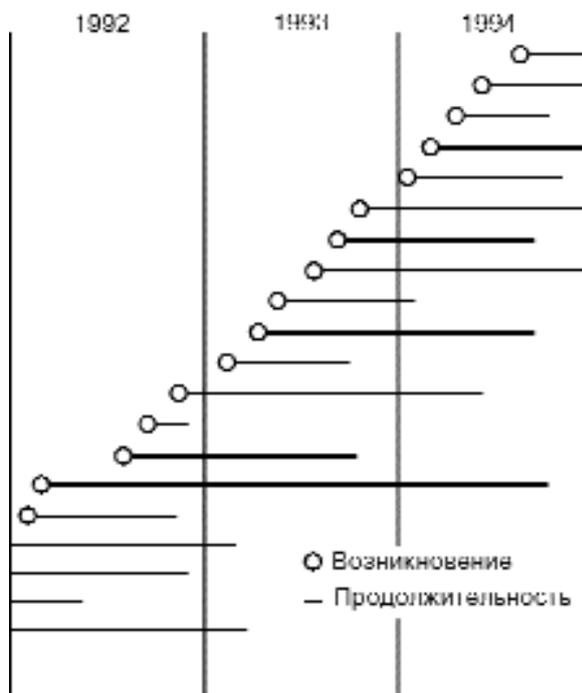


Рис. 4.1. Возникновение заболевания в обследуемой группе риска из 100 человек в период с 1992 по 1994 г.

Для того чтобы определить число новых случаев в группе, мы берем только 96 обследуемых, не имевших заболевания в начале 1992 г., и рассматриваем, что случилось с ними в течение последующих 3 лет. Пять новых случаев выявлены в 1992 г., 6 – в 1993 г. и еще пять – в 1994 г. Трехлетняя заболеваемость – это общее число новых случаев, развившихся за 3 года [16], деленное на число пациентов в группе к началу периода наблюдения (96). Итак, 16 на 96 за 3 года. А какова ежегодная заболеваемость за 1992, 1993 и 1994 гг.? Не забывая о том, что необходимо вычитать из знаменателя случаи, возникшие в предшествующие годы, мы можем рассчитать ежегодную заболеваемость, которая составляет 5 на 96 человек за 1992 г., 6 на 91 – за 1993 г. и 5 на 85 – за 1994 г.

Любая оценка частоты заболевания обязательно содержит некое указание на время. При оценке распространенности болезни время рассматривается как одно мгновение – наподобие отдельного кадра в кинофильме. Этот момент считается одним и тем же для всех пациентов в группе, хотя в действительности для обследования группы могло потребоваться несколько недель или месяцев.

При оценке частоты новых случаев устанавливаемый период времени имеет существенное значение, поскольку это тот срок, в течение которого продолжается наблюдение за пациентами группы риска для выявления возникновения изучаемого события.

В табл. 4.1. обобщены характеристики понятий частоты новых случаев и распространенности. Хотя отличия между ними кажутся очевидными, в литературе эти термины, особенно понятие частоты новых случаев, нередко используются неправильно [4].

Таблица 4.1
Различия между показателями частоты новых случаев и распространенности

Характеристика	Частота новых случаев	Распространенность
Числитель	Новые случаи за определенный срок среди лиц, не имевших заболевания	Все случаи, выявленные при однократном обследовании группы
Знаменатель	Общее число лиц в группе к началу периода наблюдения	Общее число обследованных, включая лиц с заболеванием и без него
Время	Период определенной продолжительности	Один момент времени
Способ оценки	Когортное исследование (см. главу 5)	Одномоментное исследование

Почему так важно понимать различие между этими двумя понятиями? Потому, что распространенность и частота новых случаев – это две разные характеристики популяции: распространенность показывает, какая доля популяции имеет изучаемое состояние, а частота новых случаев – каков темп появления новых случаев в популяции за определенное время. Один показатель нельзя рассчитать, зная другой.

Оценка распространенности и частоты новых случаев

Исследования распространенности

Распространенность болезни определяется путем наблюдения за группой лиц, часть которых в данный момент имеет, а другая не имеет изучаемого заболевания. Доля больных в исследуемой группе определяет *распространенность заболевания* (рис. 4.2).

Такие однократные обследования популяций, включающих в себя и больных, и здоровых, называются *исследованиями распространенности*, или *одномоментными исследованиями*, так как обследование популяции проводится в определенный момент времени. Это один из наиболее часто встречающихся в медицинской литературе методов исследования. Приведем характерный пример.

Какова распространенность деменции в общей популяции пожилых людей? Для ответа на этот вопрос в Англии (Кембридж) было обследовано 1968 человек в возрасте старше 75 лет. С помощью краткого опросника определяли когнитивный статус каждого участника. Затем применяли стратегию последовательного тестирования (см. главу 3): лица, получившие по результатам опроса



Рис. 4.2. Методика исследования распространенности.

25 и менее баллов, проходили стандартное обследование у психиатра, который и устанавливал окончательный диагноз. Распространенность слабоумия составила около 10% в общей популяции, причем в каждом следующем пятилетнем возрастном интервале этот показатель удваивался [5].

Исследования частоты новых случаев

Для оценки частоты новых случаев, в отличие от оценки распространенности, сначала выделяют группу лиц, не имеющих изучаемого заболевания, и затем в течение определенного срока периодически проводят обследования, выявляя новые случаи заболевания. Такая группа, или когорта, может состоять из здоровых людей, наблюдаемых вплоть до начала болезни, или из больных, наблюдаемых до наступления исхода. Этот процесс, называемый также когортным исследованием, будет подробнее обсуждаться в главе 5.

До сих пор частотой новых случаев мы называли темп появления новых случаев изучаемого события в наблюдаемой в течение заданного периода времени группе с определенной численностью. Это – *накопленная плотность случаев* (cumulative density), так как новые случаи с течением времени прибавляются к уже ранее возникшим.

Пример. Для того чтобы изучить заболеваемость деменцией, исследователи из Кембриджа выделили когорту, исключив из нее тех пациентов, у которых диагноз деменции уже был установлен в одномоментном исследовании, описанном ранее [6]. В когорту вошли 1778 человек. Из них 305 умерли, 190 отказались от дальнейшего обследования, с 88 была потеряна связь или же тяжесть их состояния не допускала дальнейшего обследования. Остальные 1195 человек обследовались повторно в среднем через 2,5 года после первоначального обследования. Общая ежегодная заболеваемость деменцией составила 4,3%, а в подгруппе лиц, возраст которых на момент проведения исследования распространенности был старше 85 лет, заболеваемость превышала 8% в год.

Другой подход к оценке заболеваемости состоит в подсчете числа новых случаев болезни в постоянно меняющейся группе, где время наблюдения неодинаково для разных участников. Показатель, полученный в таких исследованиях, иногда называют *плотностью заболеваемости* (или *плотностью новых случаев* – incidence density).

Типичные примеры подобных исследований – клинические испытания длительного лечения хронически больных, когда набор группы проводится в течение нескольких лет, так что пациен-

ты, включенные в начале исследования, подвергаются лечению и наблюдению более длительное время, чем те, кто был включен позднее. Для того чтобы вклад разных пациентов был соразмерен со сроком их наблюдения, в качестве знаменателя в формуле расчета плотности новых случаев используют не число пациентов, включенных в испытание за определенный период времени, а время наблюдения каждого пациента до возникновения изучаемого события (исхода). Так, вклад больного, который наблюдался в течение 10 лет и у которого не развился изучаемый исход, составляет 10, а пациента, наблюдаемого 1 год, – только 1. Плотность заболеваемости представляет собой дробь, в числителе которой находится число возникших в ходе наблюдения исходов, а в знаменателе – суммарное число человеко-лет.

Подход с применением человеко-лет полезен также при оценке заболеваемости в больших популяциях известной численности, когда возможен точный подсчет новых случаев и известна группа риска, например, при использовании популяционного регистра онкологических заболеваний.

Недостаток подхода с применением плотности заболеваемости состоит в том, что при этом смешиваются пациенты с разной продолжительностью наблюдения. При расчетах малое число пациентов, наблюдаемых длительное время, может внести в знаменатель такой же вклад, как и большое число пациентов с коротким сроком наблюдения. Если пациенты, наблюдаемые в течение длительного времени систематически отличаются от больных, наблюдаемых короткое время, то результирующий показатель частоты новых случаев окажется смещенным.

Интерпретация частотных показателей в клинике

Для того чтобы понять, что означает та или иная величина распространенности или частоты новых случаев, необходимо прежде всего четко определить числитель и знаменатель при расчете этих показателей.

Определение числителя, или что есть "случай"?

До сих пор в нашем изложении термин "случай" означал отдельного взятого пациента с конкретным изучаемым заболеванием или исходом. В классической эпидемиологии термин "случай" обозначает пациента, имеющего некое заболевание, а под распространенностью и частотой новых случаев понимают частоту случаев в популяции, например, среди населения некоторой территории. Однако клинические решения часто основываются на

информации о частоте клинических проявлений заболевания (патологических симптомов или лабораторных показателей), либо о частоте исходов заболевания (смерть, инвалидность, выздоровление). Таким образом, в клинической практике понятие "случай" часто относится к тем больным, которые имеют определенные клинические проявления или исходы заболевания.

Рассуждая о частоте, необходимо дать точное определение понятию случая, так как частота может меняться в зависимости от того, что мы понимаем под этим словом.

Пример. Простой способ выявления случая – проведение опроса для выявления лиц с определенным состоянием. Как этот подход соотносится с более строгими методами? В исследовании Комиссии по хроническим заболеваниям распространенность различных состояний, выявленных посредством опросов в домашних условиях, сравнивалась с распространенностью, оцененной при клиническом обследовании тех же лиц. На рис. 4.3 показано, какова распространенность различных состояний, полученная путем опроса и при клиническом обследовании. Видно, что при двух разных подходах к выявлению случая оценка распространенности может существенно изменяться как в ту, так и в другую сторону в зависимости от состояния [7].

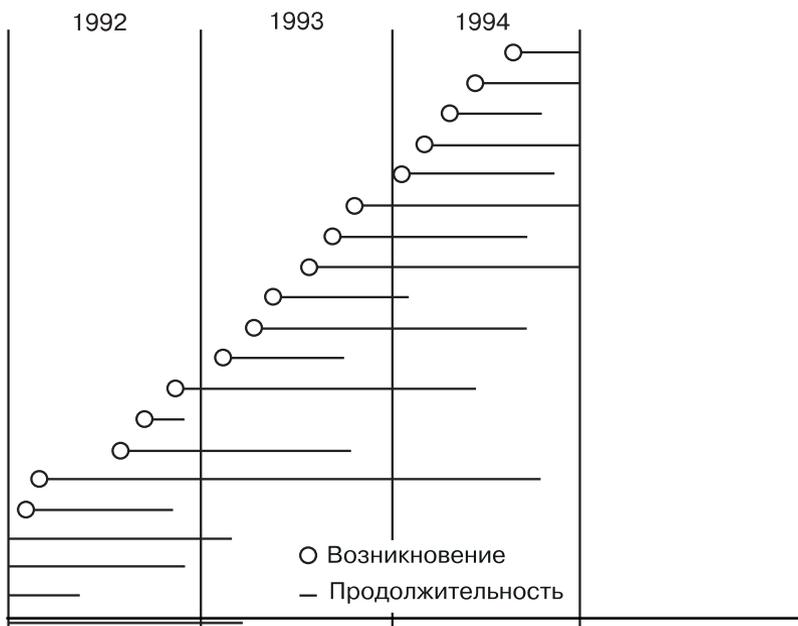


Рис. 4.3. Зависимость распространенности от метода выявления случаев. Определение распространенности заболеваний в общей популяции на основе опроса и клинического обследования.

Для некоторых состояний существуют общепринятые точные диагностические критерии. Примером могут служить критерии Центра профилактики и контроля заболеваемости США, принятые для установления диагноза лаймской болезни (табл. 4.2) [8]. Эти критерии обеспечивают специфичность, необходимую для надежного выявления такой общественно значимой болезни. Они также демонстрируют компромисс между строгой теорией и клинической реальностью. Если бы в расчет принимались только доказанные случаи, то большинство лиц, обычно считающихся больными, были бы приняты за здоровых. С другой стороны, включение в расчеты всех вероятных случаев болезни привело бы к завышению истинной частоты заболевания.

Таблица 4.2
Критерии диагностики лаймской болезни*

Определение: диагноз считается подтвержденным при наличии следующих признаков:

Хроническая мигрирующая эритема или
по крайней мере одно позднее проявление и лабораторное подтверждение инфекции

Поздние проявления (когда не найдено других причин)

Костно-мышечная система:

Кратковременные рецидивирующие отеки суставов

Нервная система (любые из следующих симптомов):

Лимфоцитарный менингит

Неврит черепных нервов (в частности, парез лицевого нерва)

Энцефаломиелит с наличием антител в цереброспинальной жидкости

Сердечно-сосудистая система:

Остро начинающаяся преходящая атриовентрикулярная блокада II или III степени

Лабораторное подтверждение (любые из следующих находок)

Выделение спирохеты *Borrelia burgdorferi*

Диагностические уровни антител IgM и IgG к спирохете в сыворотке или в цереброспинальной жидкости

Значимые различия серологических реакций между парными образцами сыворотки в острую фазу болезни и в фазу выздоровления

* Критерии Центра контроля и профилактики болезней США [US Department of Health and Human Services. Case definitions for public health surveillance. MMWR 1990;394:19–20].

Пример. В графстве Ольмстед (штат Миннесота, США) изучалась заболеваемость лаймской болезнью [9]. За период с 1980 по 1990 г. среди населения было клинически диагностировано 68 случаев заболевания. Из них 17 (25%) удовлетворяли критериям Центра контроля и профилактики болезней. В штате Миннесота обязательно сообщать о случаях лаймской болезни в официальные органы здравоохранения, однако это было сделано только в 7 случаях, из которых 4 удовлетворяли критериям Центра контроля и профилактики болезней. Пример иллюстрирует трудности точной оценки частоты болезней, диагноз которых основывается на множественных клинических критериях.

Определение знаменателя, или что есть популяция?

Показатель частоты полезен лишь в том случае, если практикующий врач понимает, к каким категориям больных он применим. Иными словами, должны быть известны численность и характеристики группы, из которой отбирались случаи для оценки частоты.

Группу, численность которой представляет собой знаменатель показателя частоты, называют популяцией, а точнее – популяцией риска, причем под риском понимается потенциальная возможность развития той болезни или того исхода, число случаев которых стоит в числителе этой дроби. Например, показатели заболеваемости или распространенности рака шейки матки будут занижены, если в популяцию включить мужчин, или, скажем, женщин, у которых удалена матка.

Характеристики популяции, численность которой стоит в знаменателе, или репрезентативной выборки из этой популяции зависят от того, на какой вопрос мы хотим получить ответ, а формирование этой популяции определяется целями исследования. Например, если мы хотим знать истинную распространенность ревматоидного артрита среди американцев, мы предпочтем включить в знаменатель случайную выборку из всего населения США. Но если нужно выяснить распространенность ревматоидного артрита в медицинской практике, например для планирования организации помощи, то в качестве знаменателя правильнее будет использовать число больных, обратившихся к врачу, а не численность всего населения. В одном исследовании было показано, что только 25% взрослых больных с симптомами артрита или ревматизма (не обязательно ревматоидного артрита) получают соответствующую помощь от специалистов или в медицинских учреждениях [10].

Эпидемиологи привыкли понимать под популяцией совокупность индивидуумов, проживающих в определенном географическом регионе. Это оправданно при исследованиях источников и эффектов различных воздействий в общей популяции. Но при решении клинических вопросов популяция должна состоять из больных, которые имеют определенные заболевания или симптомы и наблюдаются в таких же условиях, в которых эта информация в дальнейшем будет использоваться. В основном эти больные собраны в университетских и институтских клиниках и, следовательно, такие популяции могут отражать особенности небольших и своеобразных подмножеств всех пациентов с данными симптомами в определенном географическом регионе. Популяции, наблюдаемые в таких клиниках, могут являть собой необычную группу для общей медицинской практики.

Каким же образом выбор той или иной популяции влияет на оценку частоты событий? Для ответа на этот вопрос необходимо обратиться к понятию обобщаемости наблюдаемых частот.

Выборка

Редко удается обследовать все население, которое имеет или может иметь изучаемое состояние. Обычно формируется такая выборка, которую можно реально обследовать. Возникает вопрос: репрезентативна (представительна) ли она по отношению ко всей популяции?

Существует два способа получения представительной выборки. При *случайной выборке* (random sample) вероятность попасть в группу обследования для каждого индивидуума в популяции одинаковая. Более общий способ – *вероятностная выборка* (probability sample) используется, когда каждый индивидуум имеет известную (не обязательно равную) вероятность включения в группу обследования. Часто требуется, чтобы исследуемая выборка включала достаточное число представителей определенных подгрупп, например этнических меньшинств. Если эти подгруппы малы, то простая случайная выборка из всей популяции может не включать достаточного количества их представителей. Для исправления ситуации из таких подгрупп отбирается случайным образом более значительный, чем из остальной популяции, процент лиц. Окончательная выборка остается репрезентативной по отношению к общей популяции, если при анализе данных учитывается способ формирования выборки. В среднем характеристики пациентов в вероятностных выборках такие же, как в той популяции, из которой они отбирались, особенно при достаточно большом размере выборки.

При других методах формирования выборки последняя может быть значительно смещена по отношению к исходной популяции. Большинство групп больных, описываемых в медицинской литературе и соответствующих практике большинства врачей, представляют собой такие смещенные выборки. Как правило, больные включаются в исследование потому, что они находятся на лечении в университетской клинике, готовы и желают участвовать в обследовании и, кроме того, представляют собой интересный и/или тяжелый случай. В этом нет ничего плохого, если существует четкое понимание, по отношению к каким больным можно применять (или не применять) полученные результаты.

Соотношение между заболеваемостью, распространенностью и длительностью заболевания

Любой фактор, который увеличивает длительность заболевания или его клинических проявлений, повышает вероятность того, что больной будет выявлен при одномоментном исследовании (см. рис. 4.1). Соотношение между заболеваемостью, распространенностью и длительностью заболевания при их постоянстве (т.е. когда ни один из этих параметров не изменяется существенно во времени) может быть приблизительно выражено следующим уравнением:

$$\text{Распространенность} = \text{заболеваемость} \times \text{средняя длительность заболевания.}$$

Пример. В табл. 4.3 показаны приблизительная ежегодная заболеваемость и распространенность бронхиальной астмы. Заболеваемость снижается с увеличением возраста, иллюстрируя тот факт, что заболевание чаще впервые выявляется у детей. Однако распространенность остается весьма стабильной в течение всей жизни пациентов, что соответствует переходу бронхиальной астмы в хроническую форму, особенно у лиц старшего возраста. Кроме того, распространенность не увеличивается, поскольку по мере выздоровления некоторых пациентов появляется примерно такое же число вновь заболевших.

С помощью приведенной ниже формулы нетрудно определить, что средняя длительность бронхиальной астмы составляет 10 лет:

$$\text{Средняя длительность заболевания} = \frac{\text{распространенность}}{\text{заболеваемость}}$$

Таблица 4.3

Соотношения между заболеваемостью, распространенностью и длительностью заболевания бронхиальной астмой в США*

Возраст	Заболеваемость за год на 1000 населения	Распространенность на 1000 населения	Длительность заболевания ** (годы)
0 – 5	6/1000	29/1000	4,8
6 – 16	3/1000	32/1000	10,7
17 – 44	2/1000	26/1000	13,0
45 – 64	1/1000	33/1000	33,0
65 и старше	0	36/1000	33,0
Всего	3/1000	30/1000	10,0

* Приблизительные данные из нескольких источников.

** Рассчитанная, как $\frac{\text{распространенность}}{\text{заболеваемость}}$.

Когда длительность бронхиальной астмы рассчитывается для каждой возрастной группы путем деления распространенности на заболеваемость, становится очевидным, что длительность увеличивается с возрастом. Это отражает клинические наблюдения, согласно которым бронхиальная астма у детей часто проходит с возрастом, в то время как у взрослых это заболевание имеет тенденцию к хронизации.

Систематическая ошибка в одномоментных исследованиях

Одномоментные исследования могут использоваться для выявления возможных причинно-следственных связей между факторами риска и заболеванием или прогностическими факторами и исходом. В такой ситуации исследование распространенности – более быстрая, но не лучшая альтернатива когортным исследованиям. У одномоментных исследований имеется два недостатка: неопределенность последовательности событий во времени и систематические ошибки, обусловленные включением "старых" случаев, т.е. возникших до начала исследования.

Неопределенность последовательности событий во времени

В одномоментных исследованиях и болезнь, и факторы, ее вызывающие, оцениваются одновременно, поэтому подчас бывает неясно, что было раньше, а что потом. Последовательность событий неизвестна, а она бывает весьма важна для интерпретации результатов. Когда некий фактор явно предшествует началу заболевания или исходу (например, семейный анамнез или генетический маркер), причинно-следственная связь более или менее очевидна. Если же этот фактор может быть проявлением заболевания или исхода (например, патологический лабораторный показатель или психическое состояние), то судить о последовательности событий бывает гораздо труднее. Когортные исследования, напротив, имеют встроенную систему учета последовательности событий, так как возможные причины заболевания оцениваются до начала исследования (рис. 4.4).

Систематическая ошибка при включении "старых" случаев

Различия между случаями, стоящими в числителе при расчете показателей заболеваемости и распространенности, приведены на рис. 4.5. В когортном исследовании все случаи новые, причем при тщательном наблюдении факт возникновения болезни в группе риска может быть установлен. Напротив, одномоментное исследование включает смесь "старых" и "новых" случаев, выявлен-

Возможные причины

Болезнь или исход



Рис. 4.4. Временные соотношения между возможными этиологическими факторами и заболеванием в одномоментных и в когортных исследованиях.

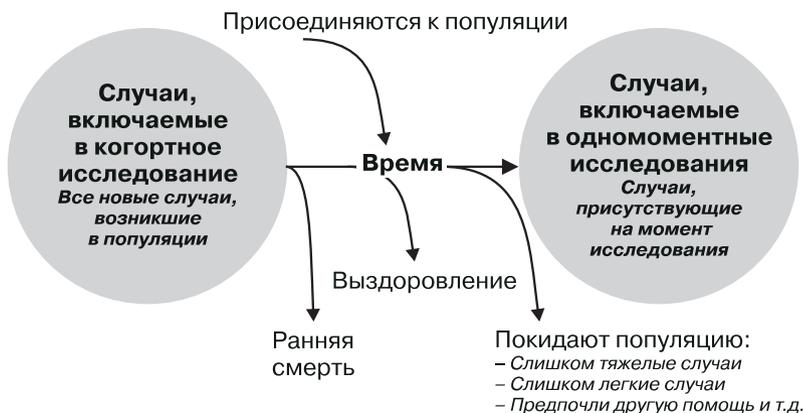


Рис. 4.5. Различия между случаями, включаемыми в одномоментные и когортные исследования.

ных во время однократного обследования. Иными словами, такое исследование выявляет все случаи, когда болезнь присутствует (т.е. может быть диагностирована), и на момент обследования еще не привела к смерти. Естественно, что распространенность будет определяться теми пациентами, которые остаются в живых и при этом не утрачивают проявлений заболевания. Различия между случаями, подставляемыми в числитель при расчетах заболеваемости и распространенности, могут влиять на интерпретацию этих показателей.

На распространенность влияет средняя длительность заболевания. Случаи заболеваний, протекающих быстро и заканчивающихся фатально, будут учтены в когортном исследовании, но, как правило, упущены в одномоментном исследовании. Так, 25 – 40% больных ИБС, ранее не имевших признаков заболевания, умирают в течение 24 ч после начала сердечного приступа. Поэтому одномоментные исследования должны занижать число случаев ИБС. С другой стороны, длительно протекающие болезни оказываются хорошо представленными в одномоментных исследованиях, даже если ежегодная заболеваемость этими болезнями невелика. Например, хотя ежегодная заболеваемость болезнью Крона составляет всего лишь от 2 до 7 на 100 000, ее распространенность превышает 100 на 100 000, что отражает хроническую природу заболевания [11].

Показатель распространенности также выборочно включает в себя более выраженные случаи несмертельных заболеваний. Например, больные с бессимптомным ревматоидным артритом могут не выявляться в исследовании, основанном на изучении текущих симптомов и физикальных данных. Аналогичным образом больные с рецидивирующими, но хорошо поддающимися лечению заболеваниями, такими как сердечная недостаточность или депрессия, могут чувствовать себя удовлетворительно в момент обследования и поэтому не будут выявлены. С другой стороны, для болезней, протекающих без ремиссий, вероятность пропуска меньше, и вследствие этого в одномоментном исследовании эти заболевания будут представлены непропорционально большим числом случаев.

Использование показателей частоты новых случаев и распространенности

Врачи применяют показатели частоты новых случаев и распространенности для трех основных целей: 1) определение прогноза для конкретного больного; 2) оценка вероятности наличия у больного определенного заболевания; 3) проведение сравнительного анализа.

Прогнозирование

Частота новых случаев – это частота, с которой болезнь или ее исход возникает за некоторый период времени в группе лиц, не имеющих данного состояния к началу наблюдения. Показатель может быть использован для прогнозирования вероятности того, что у лиц с аналогичными характеристиками такое же событие произойдет в будущем.

Пример. Для решения вопроса о наиболее приемлемом методе лечения при аденоме предстательной железы в большом когортном исследовании пожилых мужчин в Новой Англии, описанном в начале этой главы [12], оценивалась вероятность различных исходов операции трансуретральной резекции предстательной железы. Опрос более 300 мужчин, перенесших трансуретральную резекцию, показал, что эффективность операции различается в зависимости от степени тяжести симптомов до операции: после операции зарегистрировано улучшение у 93% мужчин с исходно выраженными симптомами, в то время как у больных с умеренными симптомами улучшение наступило только в 79% случаев.

Одномоментные исследования не дают надежных оснований для прогноза. Если в таком исследовании у 30% больных с инсультом выявлена депрессия, это еще не значит, что у 30% больных с инсультом без депрессии она наступит в будущем. Это может означать, что депрессия предрасполагает к развитию инсульта или же инсульт способствует депрессии, либо что больные с инсультом без депрессии выздоравливают быстрее. Для того чтобы установить долю больных с инсультом, у которых разовьется депрессия, необходимо периодически наблюдать заболевших инсультом для выявления у них симптомов депрессии.

Оценка вероятности наличия у больного определенного заболевания

Одномоментные исследования особенно полезны для принятия решения о диагнозе и лечении. Как указывалось в главе 3, знание вероятности наличия у пациента определенного заболевания при конкретном сочетании демографических и клинических характеристик влияет на применение и интерпретацию диагностических тестов. От этого может зависеть и выбор тактики лечения.

Влияние распространенности, или априорной вероятности, на подходы к решению клинических проблем видно на примере с выбором лечения при ангине.

Пример. Было проведено сравнительное исследование трех подходов к лечению ангины. Каждый подход оценивался путем сопос-

тавления потенциальной эффективности профилактики ревматизма и вероятностью развития аллергии к пенициллину. Эти три подхода таковы: 1) сделать посев мазка из зева и лечить пенициллином только тех, у кого выделен бета-гемолитический стрептококк группы А; 2) лечить всех больных, не делая посева мазка; 3) никому не делать посев мазка и никого не лечить.

Анализ данных показал, что оптимальный выбор зависит от вероятности положительного результата при посеве мазка, которая может быть оценена исходя из распространенности стрептококковой инфекции в регионе в данный момент, а также наличия или отсутствия у больного лихорадки. Авторы пришли к заключению, что если у больного вероятность положительного результата посева мазка выше 20%, то этого больного необходимо лечить, не делая посев; если ниже 5%, то можно не делать посев и не лечить; если же вероятность от 5 до 20%, то нужно делать посев мазка и затем назначать лечение в зависимости от результата посева [13].

Это исследование демонстрирует рациональный подход к использованию распространенности в качестве показателя индивидуальной вероятности заболевания в случае принятия клинического решения.

Проведение сравнений

Хотя показатели частоты новых случаев и распространенности информативны сами по себе, их роль значительно возрастает при сравнительном использовании для принятия решений. Именно сравнение частот случаев заболевания между группами обследуемых, имеющих и не имеющих определенные характеристики, обеспечивает самые строгие доказательства. Например, риск развития рака легкого (заболеваемость) у интенсивно курящих мужчин составляет около 0,17% в год, что едва ли можно рассматривать как распространенное явление. Лишь при сопоставлении этого показателя с заболеваемостью среди некурящих (примерно 0,007% в год) вред курения становится очевиден. Клиницисты используют анализ частоты как один из подходов для сравнительной оценки связи между фактором риска и заболеванием или его исходом. Методы сравнения частот более детально описаны в главе 5.

Анализ клинических решений

Количественные подходы помогают при поиске наиболее эффективных и целесообразных путей решения специфических проблем у конкретных пациентов (*клиническая политика*) или при распределении ресурсов в системе здравоохранения между регионами и группами населения (*общественная политика*).

При *анализе решений* врач намечает альтернативные пути (например, хирургическое или терапевтическое лечение при аденоме предстательной железы; посев мазка с последующим назначением антибиотика или лечение антибиотиками всех больных при стрептококковой ангине), а затем рассчитывает, какой метод даст наилучший результат, оценивая для каждой ветви последовательности событий частоту и клиническую значимость возможных исходов. Основные этапы этого процесса были представлены ранее [14] и здесь приводится лишь их краткое описание.

1. Построение "*дерева решений*". Анализ клинического решения начинается с дилеммы по отношению к конкретному больному: какое из возможных действий следует предпринять? Эти альтернативные решения образуют начало "дерева", которое затем разветвляется, включая все важные последствия этих решений, и заканчивается клинически важными исходами. Места разветвлений обозначают либо выбор того или иного решения относительно лечения пациента (квадраты на рис. 4.6), либо спонтанные события (кружки на рис. 4.6). Число последовательностей событий и исходов бесконечно, однако, как правило, только некоторые из них действительно важны и могут произойти. Для того чтобы анализ был управляемым, необходимо "подрезать" дерево, оставив для рассмотрения лишь несколько наиболее важных его ветвей.
2. *Оценка вероятности спонтанных событий*. Обычно эти вероятности соответствуют оценкам частоты клинических событий, взятым из медицинской литературы.
3. *Оценка клинической значимости исходов*. Лучше всего оценивать значимость исходов при беседе с больным, который не согласен с решением врача. Единицы оценки значимости произвольны, но она должна быть выражена в интервальной шкале, например от 1 до 100. Числовое выражение относительной значимости исходов (смерть, болезнь, инвалидность) – не простое дело, особенно когда различные исходы измеряются в разных единицах, например продолжительность и качество жизни. Однако больные все равно оценивают каждый исход, так что цифры только делают такую оценку наглядной.
4. *Расчет ожидаемой клинической значимости альтернативных решений*. Клиническую значимость исхода, указанную в прямоугольнике на конце каждой ветви справа (см. рис. 4.6), умножаем на вероятность события, указанную рядом с каждой ветвью в скобках. Затем суммируем все значения, полученные для каждой ветви, исходящей из общего узла, и определяем клиническую значимость для этого разветвления (указана в кружке).

Действуя таким образом, постепенно движемся справа налево, пока не получим клиническую значимость альтернативных клинических решений.

5. *Выбор решения с наибольшей ожидаемой клинической значимостью.*
6. *Анализ чувствительности.* Исходно оценки вероятностей и клинической значимости альтернативных решений неизвестны. Конечный этап анализа решения состоит в том, чтобы посмотреть, как зависят результаты анализа от изменений этих оценок в диапазоне допустимых значений. Другими словами, нужно выяснить, насколько решение "чувствительно" к неточности оценок. Анализ чувствительности показывает, какая именно последовательность событий в наибольшей степени влияет на принимаемое решение и какова величина этого влияния.

Пример. Был проведен анализ принятия решений [15] при выборе метода лечения пожилого больного с симптомами аденомы предстательной железы II степени (см. начало главы). До того, как появление лекарственных препаратов и лазерной простатэктомии усложнило принятие решения, возможными подходами были хирургическое вмешательство (трансуретральная простатэктомия) и вы-



Рис. 4.6. Дерево решений. Лечение 70-летнего больного с умеренно выраженными симптомами аденомы предстательной железы.

жидательная тактика. Для оценки возможного выбора использовалось "дерево решений" (см. рис. 4.6). Показатели частоты разных исходов были получены из исследования заболеваемости в Новой Англии, описанного ранее в этой главе [12], и из других публикаций [15]. Следует отметить, что оптимальным решением в этом случае является хирургическое вмешательство (клиническая значимость решения 0,94). В таком случае наилучший метод лечения – трансуретральная простатэктомия, поскольку операционная летальность крайне низка, а показатель значимости осложнений операции (недержание мочи или импотенция) такой же, как при жизни со стабильными умеренно выраженными урологическими симптомами (0,89). Если жизнь со стабильными умеренно выраженными симптомами для пациента предпочтительнее, чем недержание мочи или импотенция, то этот баланс нарушается.

Резюме

На большинство вопросов, встающих перед клиницистами, можно получить ответ исходя из частоты событий при тех или иных обстоятельствах. Частота клинических событий выражается вероятностью, или отношением, где в числителе – число событий, а в знаменателе – численность группы, в которой они выявляются.

Существует два показателя частоты: распространенность и частота новых случаев. Распространенность – это доля пациентов, имеющих изучаемый исход (заболевание) в данный момент времени. Частота новых случаев (заболеваемость) – это доля лиц в группе обследования, у которых возникли новые случаи изучаемого исхода (заболевания) в течение установленного срока.

Распространенность определяется путем однократного обследования группы лиц, включающей больных и здоровых, в то время как оценка частоты новых случаев требует проведения нескольких обследований за определенный срок в группе лиц, исходно не имеющих изучаемого состояния (заболевания или исхода). Таким образом, одномоментные исследования выявляют только пациентов, которые живы и у которых диагноз установлен на момент обследования, тогда как когортные исследования выявляют все новые случаи исхода (заболевания). Между тем выявленные при однократном обследовании случаи могут оказаться смещенным подмножеством всех случаев, поскольку они не учитывают тех, кто уже умер или выздоровел. Кроме того, одномоментные исследования часто не обеспечивают четкого представления о временном соотношении между причинным фактором и болезнью.

Для того чтобы разобраться в смысле показателей частоты новых случаев и распространенности, врач должен знать, на каком основании устанавливался диагноз заболевания и каковы харак-

теристики популяции, представленной в знаменателе. Последнее особенно важно при попытке решить вопрос о том, применимы ли конкретные значения распространенности и частоты новых случаев к пациентам, с которыми врач имеет дело в собственной практике.

Частота новых случаев считается наиболее подходящей оценкой для прогнозирования. Распространенность служит количественным выражением вероятности того, что пациент с некоторыми характеристиками имеет заболевание в данный момент времени, и поэтому используется для принятия решений при диагностике и скрининге. Однако наиболее важная область применения показателей частоты новых случаев и распространенности – сравнительный анализ альтернативных клинических решений.

Оценка болезней и их исходов через показатель частоты новых случаев – необходимая составляющая количественного подхода к принятию решений. Анализ принятия решений позволяет определить альтернативные клинические подходы, а затем оценить их количественно путем сравнения ожидаемых результатов по частоте и клинической значимости основных исходов.

Постскриптум

Подсчет клинических событий, описанный в этой главе, может показаться занудством, поскольку кажется очевидным, что количественное изучение частоты событий в различных условиях служит основой клинической науки. Между тем количественный метод оценки лечения был предложен менее 200 лет тому назад: Пьер Луи имел дерзость подсчитывать число случаев смерти и выздоровления при лихорадке в зависимости от применения кровопускания. Его осуждали за то, что он позволял себе на основе безжизненных цифр сеять сомнения в целительной силе пиявок, многократно подтверждавшейся на протяжении десятилетий пристальных клинических наблюдений.

Литература

1. Bryant G.D., Norman G.R. Expressions of probability: words and numbers. *N Engl J Med* 1980;302:411.
2. Toogood J.H. What do we mean by "usually"? *Lancet* 1980;1:1094.
3. McConnell J.D., Barry M.J., Bruskevitz R.C. Benign prostatic hyperplasia: diagnosis and treatment. *Clin Pract Guide Quick Ref Guide Clin* 1994;8:1–17.
4. Friedman G.D. Medical usage and abuse, "prevalence" and "incidence." *Ann Intern Med* 1976; 84:502–503.
5. O'Connor D.W., Politt P.A., Hyde J.B., Fellows J.L., Miller N.D., Brook C.P.B., Reiss B.B., Roth M. The prevalence of dementia as measured by the Cambridge Mental Disorders of the Elderly Examination. *Acta Psychiat Scand* 1989;79:190–198.

6. Paykel E.S., Brayne C., Huppert F.A., Gill C., Barkley C., Gehihaar E., Beardsall L., Girling D.M., Pollitt P., O'Connor D. Incidence of dementia in a population older than 75 years in the United Kingdom. *Arch Gen Psychiat* 1994;51:325–332.
7. Sanders B.S. Have morbidity surveys been oversold? *Am J Publ Hlth* 1962;52:1648–1659.
8. Case definitions for public health surveillance. *MMRW* 1990;(RR-13):19–21.
9. Matteson E.L., Beckett V.L., O'Fallon W.M., Melton L.J. III., Duffy J. Epidemiology of Lyme disease in Olmsted County, MN, 1975–1990. *J Rheumatol* 1992;19:1743–1745.
10. Spitzer W.O., Harth M., Goldsmith C.H., Norman G.R., Dickie G.L., Bass M.J., Newell J.P. The arthritic complaint in primary care: prevalence, related disability, and costs. *J Rheumatol* 1976;3:88–99.
11. Sedlack R.E., Whisnant J., Elveback L.R., Kurland L.T. Incidence of Crohn's disease in Olmsted County, Minnesota, 1935–1975. *Am J Epidemiol* 1980;112:759–763.
12. Fowler F.J., Wennberg J.E., Timothy R.P., Barry M.J., Mulley A.G., Hanley D. Symptom status and quality of life following prostatectomy. *JAMA* 1988;259:3018–3022.
13. Tompkins R.K., Bumens D.C., Cable W.E. An analysis of the cost-effectiveness of pharyngitis management and acute rheumatic fever prevention. *Ann Intern Med* 1977;86:481–492.
14. Sox H.C., Blatt M.A., Higgins M.C., Marton K.I. *Medical decision making*. Stoneham, MA: Butterworth, 1988.
15. Barry M.J., Mulley A.G., Fowler F.J., Wennberg J.W. Watchful waiting vs. immediate transurethral resection for symptomatic prostatism. *JAMA* 1988;259:3010–3017.

Рекомендуемая литература

- Ellenberg J.H., Nelson K.B. Sample selection and the natural history of disease: studies of febrile seizures. *JAMA* 1980;243: (13): 1337–1340.
- Friedman G.D. Medical usage and abuse, "prevalence" and "incidence." *Ann Intern Med* 1976;84:502–503.
- Morgenstern H., Kleinbaum D.G., Kupper L.L. Measures of disease incidence used in epidemiologic research. *Int J Epidemiol* 1980;9:97–104.

5 РИСК

Под риском обычно понимают вероятность какого-то неблагоприятного события. В этой главе термин "риск" используется в более узком смысле – как вероятность того, что у лиц, подвергшихся влиянию некоторых факторов ("факторов риска"), возникнет определенное заболевание.

Практически каждого человека волнует, насколько велик риск возникновения у него того или иного заболевания. Эта озабоченность обуславливает появление множества популярных книг и газетных статей о риске развития рака молочной железы при воздействии токсичных химических веществ, риске заражения СПИДом при переливании крови или риске развития рака предстательной железы после операции вазэктомии, а также о путях снижения риска этих и других заболеваний.

В этой главе описываются методы оценки риска по соотношению между воздействием возможных факторов и последующей заболеваемостью. Здесь также обсуждаются некоторые способы сравнения рисков, которым подвергаются отдельные индивидуумы и популяции.

Факторы риска

Факторами риска (risk factors) называются особенности организма или внешние воздействия, приводящие к увеличению риска возникновения заболевания. Некоторые факторы риска – наследственные. Например, у носителей гаплотипа HLA-B27 значительно повышен риск заболевания спондилоартритом. Работа по программе "Геном человека" позволила идентифицировать ряд других заболеваний со специфическими генами в качестве факторов риска: в частности, рак толстой кишки, остеопороз, боковой амиотрофический склероз. Другие факторы риска, такие как возбудители инфекций, фармакологические препараты и токсины, находятся в окружающей среде. Некоторые факторы являются частью социального окружения. Показано, например, что эмоциональная нагрузка в связи с утратой супруга, изменениями повседневной жизни или скученностью проживания повышают частоту заболеваний, причем не только психических, но и соматических. некото-

рые из наиболее значимых факторов риска относятся к сфере поведения человека: курение, избыточное потребление алкоголя, пренебрежение ремнями безопасности при езде на автомобиле, беспорядочная половая жизнь.

Воздействие факторов риска всегда предшествует развитию заболевания. Воздействие может быть однократным, например, облучение населения при аварии на атомной станции. Чаще контакт с факторами риска, вызывающими хронические заболевания, происходит в течение некоторого периода времени (курение, артериальная гипертензия, беспорядочные половые связи, инсоляция).

Существует несколько способов описания воздействия факторов риска: наличие факта воздействия, действующая доза, максимальная доза, общая накопленная доза, продолжительность воздействия в годах, время, прошедшее после первого контакта с фактором риска и т.д. [1]. Несмотря на то что многие из этих количественных показателей могут быть вычислены один из другого, некоторые из них демонстрируют связь между дозой и эффектом, в то время как для других эта связь отсутствует.

Так, общая накопленная доза солнечной радиации является фактором риска развития немеланомного рака кожи, а сильные солнечные ожоги больше предрасполагают к развитию меланомы. Выбор адекватного показателя интенсивности воздействия фактора риска обычно основывается на полном знании биологических эффектов воздействия и патогенеза болезни.

Выявление факторов риска

Если некоторое воздействие быстро и с высокой вероятностью приводит к развитию определенного заболевания, то нетрудно догадаться, что это воздействие является фактором риска для данного заболевания. В частности, оценка причин таких состояний, как ветряная оспа, солнечный ожог или передозировка аспирина не составляет труда, поскольку соответствующие заболевания следуют относительно быстро за воздействием очевидных факторов риска. Однако заболеваемость и смертность в большей степени обусловлены хроническими болезнями, для которых связь между воздействием фактора риска и развитием заболевания носит не столь явный характер. Поэтому врач, каким бы проныцательным он ни был, не может оценить риск на основании своей личной практики. Почему?

Продолжительный латентный период

Многие болезни имеют продолжительный латентный период, т.е. период между воздействием фактора риска и первыми прояв-

лениями патологии. Это, в частности, справедливо для некоторых видов злокачественных новообразований, таких как рак щитовидной железы у взрослых после лучевой терапии по поводу тонзиллита в детстве. Когда заболевание возникает через много лет после воздействия, исходный фактор риска может быть просто забыт и связь между ним и заболеванием останется невыявленной.

Высокая распространенность факторов риска

Многие факторы риска, такие как курение или потребление пищи с высоким содержанием холестерина и насыщенных жиров настолько распространены в нашем обществе, что в течение многих лет не казались опасными. Только при сравнении заболеваемости людей с этими факторами риска и без них или путем исследования особых подгрупп, например мормонов (которые не курят) или вегетарианцев (которые потребляют пищу с низким содержанием холестерина), удалось понять значимость этих факторов риска.

Низкая заболеваемость

Многие заболевания, которые считаются "обычными", на самом деле довольно редки. Так, хотя рак легких составляет наиболее частую причину смерти от онкологических заболеваний среди американцев, заболеваемость раком этой локализации даже среди заядлых курильщиков ниже чем 2 на 1000 в год. В жизни врача общей практики могут пройти годы между выявлением двух новых случаев рака легкого. Относительно таких редких событий врачу трудно делать какие-либо выводы.

Низкий риск

Если при воздействии некоего фактора риска вероятность развития заболевания низкая, то необходимо наблюдать большое число людей, чтобы выявить различия заболеваемости у лиц, подвергавшихся и не подвергавшихся воздействию этого фактора. Это справедливо, даже если фактор риска и заболевание встречаются относительно часто. Например, до сих пор нет твердой уверенности в том, повышается ли риск развития рака молочной железы вследствие применения пероральных контрацептивов, поскольку по всем имеющимся оценкам риск этот очень мал и может быть вообще результатом систематической или случайной ошибки. Наоборот, не вызывает сомнений, что вирус гепатита В является фактором риска развития печеночно-клеточного рака, поскольку инфицированные вирусом гепатита В заболевают раком печени в сотни раз чаще, чем те, кто не является носителем этого вируса.

Высокая распространенность заболевания

Если заболевание распространенное (болезни сердца, злокачественные новообразования или инсульт) и для него уже установлен ряд факторов риска, то бывает трудно отличить новый фактор от известных. Например, синдром внезапной смерти – достаточно распространенная причина летального исхода у взрослых. Часто внезапная смерть связана, по-видимому, с ишемической болезнью сердца. Однако не исключены и другие важные причины, пока не распознанные вследствие того, что для большинства случаев существует подходящее объяснение.

С другой стороны, редкие заболевания и необычные клинические события побуждают к усилиям по раскрытию их причин. СПИД был настолько необычным синдромом, что появление всего нескольких случаев заставило искать новый возбудитель (как оказалось, ретровирус). Аналогичным образом врачи быстро обратили внимание на появление нескольких случаев рака влагалища – весьма редкого заболевания. Были проведены тщательные исследования и найдено объяснение – матери пациенток применяли во время беременности диэтилstilбэстрол.

Множественные причины и эффекты

Обычно тесная однозначная связь между фактором риска и конкретным заболеванием отсутствует. В качестве примера на рис. 5.1 представлено соотношение между артериальной гипертензией (АГ) и сердечной недостаточностью. У некоторых людей с АГ развивается сердечная недостаточность, а у других этого не происходит. Кроме того, сердечная недостаточность возникает и у многих людей без АГ, поскольку существуют иные причинные факторы. К тому же связь затушевывается тем обстоятельством, что АГ способствует развитию разных заболеваний, а не только сердечной недостаточности. Таким образом, хотя у пациентов с АГ сердечная недостаточность наблюдается примерно в 3 раза чаще, чем у лиц без нее (причем в подобных случаях АГ признается ведущей причиной этого заболевания), врачи не были уверены в наличии такой связи вплоть до 70-х годов нашего столетия, когда, наконец, после тщательного многолетнего исследования больших групп людей появились данные, внушающие доверие.

В силу всех этих обстоятельств, основываясь на личном опыте, врач вряд ли в состоянии доказать связь между воздействием фактора риска и заболеванием, хотя может подозревать ее наличие. Для того чтобы получить точную информацию, врач должен обратиться к медицинской литературе, прежде всего к исследованиям, которые тщательно спланированы и включают большое число пациентов.



Рис. 5.1. Соотношение между факторами риска и заболеванием: артериальная гипертонзия (АГ) и сердечная недостаточность. АГ вызывает развитие многих болезней, включая сердечную недостаточность, а сердечная недостаточность имеет разные причины, в том числе и АГ.

Применение показателей риска

Прогнозирование

Знание факторов риска используется прежде всего для прогнозирования заболевания. В самом деле, уже по определению факторы риска предсказывают некоторые будущие события. Наиболее полную информацию для прогнозирования болезни у индивидуума можно получить, обобщив фактический материал, полученный при изучении большого числа людей со сходным фактором риска. Качество таких прогнозов зависит от сходства между пациентами, на изучении которых основан прогноз, и индивидуумом, для которого прогноз строится.

Важно иметь в виду, что присутствие даже очень сильного фактора риска еще не означает, что человек обязательно заболеет. Например, исследования показали, что у заядлого курильщика риск развития рака легкого в 20 раз выше, чем у некурящего, и тем не менее его шансы заболеть раком легкого в течение ближайших 10 лет составляют 1 на 100.

Существует принципиальное несоответствие между заболеваемостью в популяции и шансом заболеть у индивидуума. Естественно, что и больные, и врачи хотели бы иметь как можно более точный ответ на вопрос о вероятности развития болезни в будущем. Их не устраивает оценка вероятности в виде шанса заболеть раком легкого или инсультом в течение ближайших 5 лет. В конце концов через 5 лет у любого человека данное заболевание либо разовьется, либо нет. Так что в некотором смысле усреднение всегда дает ошибочный результат, поскольку усредняемые понятия относятся к разным категориям: с одной стороны – вероятность события, а с другой – наличие или отсутствие болезни. Однако оценка вероятности может направлять процесс клинического решения. Даже если прогноз ошибочен на уровне индивидуума, он обычно оказывается справедливым для множества аналогичных пациентов.

Причина

Хотя предсказать заболевание по наличию тех или иных факторов риска можно, они не всегда служат его причиной. Фактор риска может оказаться косвенным маркером исхода благодаря связи с одним или несколькими причинными факторами, т.е. он может лишь находиться под влиянием причинного фактора. Например, недостаточный культурный уровень матери – это фактор риска рождения младенца с низкой массой тела. Вместе с тем с культурным уровнем матери связаны и более прямые причины низкой массы тела новорожденного – неправильный образ жизни и плохое питание в период беременности, курение и т.д. Фактор риска, не являющийся причиной заболевания, называется маркером, поскольку он служит "меткой" (marker) повышенной вероятности развития заболевания. Несмотря на то что фактор риска – не причина, это не исключает его значимости для прогнозирования вероятности развития заболевания. Однако из этого не следует, что при устранении самого фактора исчезает связанный с ним повышенный риск. Например, как отмечалось в главе 1, папилломавирус человека – это фактор риска развития рака шейки матки; вместе с тем в развитии рака шейки матки роль других заболеваний, передающихся половым путем, таких как генитальный герпес или хламидиоз, остается не столь ясной. У больных с этой формой рака повышен уровень антител к перечисленным вирусам, однако эти факторы могут быть маркерами риска развития рака, а не его причиной. Если это так, то лечение этих инфекций не обязательно предотвратит развитие рака шейки матки. С другой стороны, упорядочение половой жизни послужит мерой профилактики как

приобретения причинного фактора для развития рака шейки матки, так и других инфекций, передающихся половым путем [2].

Для выяснения, является ли фактор риска причиной или просто маркером заболевания, существует несколько способов. Они рассматриваются в главе 11.

Диагностика

Знание риска может использоваться в диагностическом процессе, поскольку наличие фактора риска увеличивает вероятность заболевания, а это один из способов повышения прогностической ценности диагностического теста.

Однако для каждого пациента в отдельности факторы риска обычно не считаются такими надежными прогностическими показателями развития заболевания, как ранние клинические признаки [3]:

"Развитие серьезного заболевания точнее всего можно предсказать на основании начальных симптомов. Обнаруженное сегодня снижение легочной вентиляции позволяет с уверенностью предсказать его дальнейшее снижение в будущем, а выявленное повышение артериального давления дает повод утверждать, что в дальнейшем оно будет еще выше. Ранние клинические проявления стенокардии позволяют прогнозировать развитие в будущем летального исхода лучше, чем все традиционные факторы риска".

Знание факторов риска может использоваться наиболее эффективно, если этот риск высок и если, кроме того, имеются клинические признаки, увеличивающие вероятность заболевания. Например, возраст и пол – относительно значимые факторы риска развития ИБС, но распространенность заболевания среди пожилых мужчин, т.е. в группе наиболее высокого риска, составляет лишь 12%. Если же при оценке распространенности учитывается и клиническая картина – наличие и характер болей в грудной клетке, а также результаты ЭКГ-пробы с физической нагрузкой, то вероятность ИБС возрастает до 99% [4].

Информация об отсутствии фактора риска может быть использована для исключения диагноза, особенно если этот фактор определяет развитие заболевания. Так, при дифференциальной диагностике опухоли плевры у больного, работавшего на асбестовом производстве, разумно предположить мезотелиому, однако возникновение ее маловероятно, если больной никогда не имел контакта с асбестом.

Знание факторов риска используется также для формирования групп повышенного риска с целью повышения эффективности программ скрининга.

Профилактика

Если фактор риска является к тому же причиной болезни, то его устранение можно использовать в качестве профилактической меры, независимо от того, известен или нет патогенез заболевания. Иллюстрацией служат некоторые классические примеры из истории эпидемиологии. Так, еще до того как был идентифицирован возбудитель, Сноу заметил, что холерой чаще заболевают люди, употребляющие воду из определенных источников, и предотвратил эпидемию, перекрыв эти источники. В наше время, прежде чем был открыт вирус иммунодефицита человека (ВИЧ), удалось выявить, что наличие многих партнеров у мужчин-гомосексуалистов служит фактором риска заболевания СПИДом [5]. Концепции причины болезни и взаимоотношение причины и профилактики обсуждаются в главе 11.

Исследования риска

Лучшим способом выяснить, действительно ли воздействие потенциального фактора риска увеличивает вероятность заболевания, был бы эксперимент. Для этого лиц, не имеющих в данный момент изучаемой болезни, нужно объединить в группы с одинаковой предрасположенностью к этой болезни. Затем одну группу пришлось бы подвергнуть воздействию определенного фактора риска, а другая служила бы контролем. Во всех других отношениях две группы должны быть в одинаковых условиях. Различие в заболеваемости между этими двумя группами можно было бы отнести на счет фактора риска.

К сожалению, воздействие большинства факторов риска на человека нельзя изучать с помощью *экспериментальных исследований*, когда исследователь определяет, кого подвергать воздействию. Назовем некоторые вопросы, касающиеся оценки риска, которые занимают ученых-медиков сегодня. Насколько повышен риск сердечно-сосудистых заболеваний у людей с пониженной физической активностью, которые во всех других отношениях не отличаются от активных людей? Вызывают ли радиотелефоны развитие опухолей мозга? Увеличивает ли прием алкоголя риск развития рака молочной железы? Для того чтобы ответить на подобные вопросы, необходим эксперимент, провести который, однако, невозможно. Во-первых, он продолжался бы десятилетия. Во-вторых, неэтично подвергать группу людей воздействию возможного фактора риска. Наконец, большинство людей не захотят, чтобы их диета и поведение в течение длительного времени определялись посторонними лицами. Следовательно, для изучения риска необходимы более щадящие методы.

Клинические исследования, в которых исследователь собирает данные путем простого наблюдения событий в их естественном течении, не вмешиваясь в них активно, называются *обсервационными исследованиями* (observational studies). Таково большинство исследований риска; сюда относятся *когортные исследования* (cohort studies), о которых речь пойдет далее, и *исследования случай–контроль* (case control studies), рассматриваемые в главе 10.

Когорты

Термин "*когорта*" (cohort) означает группу лиц, изначально объединенных каким-либо общим признаком и наблюдаемых в течение определенного периода времени, чтобы проследить, что с ними произойдет в дальнейшем. В табл. 5.1 перечислены некоторые примеры использования когорт для клинических исследований.

Таблица 5.1
Когорты и их назначение

Признак, по которому составлена когорта	Используется для оценки	Пример
Возраст	возраста	Ожидаемая продолжительность жизни для лиц в возрасте 70 лет (независимо от даты рождения)
Дата рождения	календарного времени	Заболеваемость туберкулезом лиц 1910 года рождения
Воздействие	фактора риска	Рак легкого у курильщиц
Заболевание	прогноза	Выживаемость больных раком молочной железы
Профилактическое лечение	профилактики	Снижение заболеваемости пневмонией после применения пневмококковой вакцины
Лечебное вмешательство	лечения	Повышение выживаемости больных лимфогранулематозом после комбинированной химиотерапии

Если нужно получить достоверную информацию о риске, то, независимо от способа формирования когорты, наблюдение за ней должно проводиться с соблюдением следующих двух условий.

Во-первых, когорты должны наблюдаться в течение срока, превышающего продолжительность естественного течения изучаемого заболевания. Это необходимо для того чтобы дать риску возможность проявиться. Если мы хотим выяснить, приводит ли облучение области шеи в детстве к возникновению рака щитовидной

железы, то пятилетнее проспективное наблюдение вряд ли позволит проверить гипотезу о связи между раком щитовидной железы и облучением, поскольку обычно период между этим воздействием и проявлением заболевания гораздо продолжительнее.

Во-вторых, все лица, включенные в когорту, должны проследиваться в течение всего периода наблюдения. Поскольку сам факт выбывания людей из исследования и причины такого выбывания влияют на результат, информация, полученная на неполной когорте, может исказить истинное положение вещей.

Когортные исследования

При проведении когортного исследования формируется группа (когорта), в которую включаются лица, у которых отсутствует изучаемое заболевание, но у которых оно может появиться в дальнейшем (рис. 5.2). (Например, в исследовании факторов риска развития рака эндометрия все включенные в когорту женщины не должны иметь болезней матки.) Лица, включенные в когорту, классифицируются по характеристикам (потенциальным факторам риска), которые могут влиять на исход. Затем эту когорту наблюдают в течение определенного периода, чтобы установить, у кого из ее участников возникнет изучаемый исход. При этом есть возможность выявить, какие из изучаемых факторов риска связаны с последующими исходами. Когортные исследования называют также *продольными* (longitudinal) – подчеркивается, что пациенты прослеживаются во времени, или *проспективными* (prospective) – имеется в виду, что группа сформирована в настоящее время и прослежена в будущем, или *исследованиями заболеваемости* (incidence) – обращено внимание на то, что основным способом оценки является регистрация новых случаев заболевания в течение определенного срока.

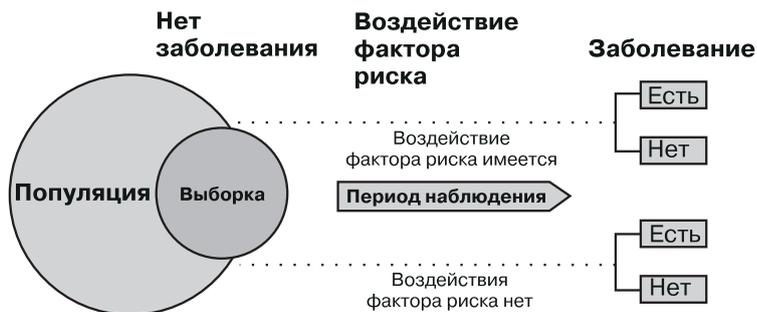


Рис. 5.2. Методика когортного исследования риска.

Опишем в качестве примера классическое когортное исследование, которое внесло важнейший вклад в понимание этиологии сердечно-сосудистых заболеваний.

Пример. Фремингемское исследование [6] было начато в 1949 г. для определения факторов, связанных с повышенным риском развития ИБС. Репрезентативная выборка общей численностью 5209 мужчин и женщин в возрасте 30–59 лет была отобрана примерно из 10 000 лиц этого возраста, проживавших во Фремингеме, – небольшом городке вблизи Бостона (США). Из исключенных в исследование 5127 не страдали ИБС на момент первого обследования и, следовательно, обладали только риском развития ИБС. Эти люди затем обследовались каждые два года для выявления признаков ИБС. Исследование продолжалось 30 лет и показало, что риск развития ИБС связан с повышенным артериальным давлением, высоким уровнем холестерина, курением, нарушением толерантности к глюкозе и гипертрофией левого желудочка. Были выявлены существенные различия риска развития ИБС между теми, кто не имел ни одного из этих факторов, и теми, у кого присутствовали все факторы.

Исторические когортные исследования

Существует два способа проведения когортных исследований (рис. 5.3). Когорта может быть сформирована в настоящее время и прослежена в будущем (*проспективное когортное исследование*, concurrent cohort study) или выделена по архивным документам и прослежена до настоящего времени (*ретроспективное когортное исследование*, historical cohort study).

Рассматриваемые ниже достоинства и недостатки примерно в равной степени относятся к параллельным и к историческим когортным исследованиям. Однако эти два вида проспективных исследований различаются по качеству получаемых данных. В про-



Рис. 5.3. Историческое и параллельное когортные исследования.

спективных исследованиях сбор данных производится непосредственно для целей исследования с полным пониманием поставленных задач. Это позволяет избежать систематических ошибок, которые могут снизить точность данных. Сбор данных для ретроспективных когортных исследований проводится по архивным данным (обычно по историям болезни), где они фиксировались независимо от целей исследования. Такие данные не всегда отвечают требованиям, предъявляемым к строгим научным исследованиям.

Достоинства и недостатки когортных исследований

Некоторые достоинства и недостатки когортных исследований по изучению факторов риска обобщены в табл. 5.2. Когортные исследования риска – наилучшая замена истинного эксперимента в ситуации, когда эксперимент невозможен. Они основаны на той же логике, что и клинические испытания, и позволяют выявить воздействие возможного фактора риска, избегая систематических ошибок, которые непременно возникают, если оценка воздействия производится уже после того, как стал известен исход.

Главный их недостаток заключается в том, что если исход редкий (что обычно и бывает на самом деле), то в группу приходится включать большое число людей. Чтобы получить результаты, группа должна оставаться под наблюдением продолжительное время. Например, фремингемское исследование ИБС – одного из наиболее частых хронических заболеваний в Америке – было в свое время самым крупным исследованием такого рода. Однако пришлось наблюдать более 5000 лиц в течение несколь-

Таблица 5.2
Достоинства и недостатки когортных исследований

Достоинства	Недостатки
Единственный способ непосредственной оценки заболеваемости (т.е. абсолютного риска)	Число включенных в исследование лиц должно быть значительно больше, чем число пациентов с изучаемым заболеванием; поэтому метод непригоден для редких заболеваний
Отвечают на клинический вопрос: заболеют ли люди, если они подверглись воздействию фактора риска?	Высокая стоимость из-за того, что приходится исследовать большое число людей в течение продолжительного времени
При оценке эффекта воздействия фактора риска позволяют избежать систематической ошибки, которая обычно возникает, если исход известен заранее	Результаты долгое время остаются неизвестными
Позволяет оценить связь между воздействием фактора риска и несколькими заболеваниями	Позволяет оценить связь между заболеванием и воздействием относительно небольшого числа факторов (тех, что были определены в начале исследования)

ких лет, прежде чем удалось опубликовать первые предварительные результаты. В течение первых 8 лет различные проявления ИБС были зарегистрированы только у 5% обследуемых!

Другой недостаток когортных исследований обусловлен тем, что лица, включенные в группу, обычно живут свободно, а не под контролем исследователей. Для поддержания контакта с ними требуются значительные усилия и средства. Поэтому когортные исследования стоят дорого, иногда миллионы долларов.

Поскольку затраты времени и денег ограничивают использование когортных исследований, они применяются для получения ответов только на самые важные вопросы. Это послужило стимулом к поиску менее трудоемких и дорогих, но достаточно надежных методов оценки риска. (Наиболее распространенный из них – исследование случай–контроль, обсуждается в главе 10.)

В научном аспекте основной недостаток обсервационных исследований, включая когортные, состоит в том, что они в большей степени, чем экспериментальные, подвержены систематическим ошибкам. Лица, испытывающие влияние фактора риска, при естественном течении событий по многим показателям отличаются от группы сравнения, на которую этот фактор не воздействовал. Если эти различия имеют отношение к рассматриваемому заболеванию, то они могут повлиять на любые связи, существующие между предполагаемым фактором риска и заболеванием.

Отсюда главное требование к обсервационным исследованиям: чтобы устранить "мешающие" различия между основной и контрольной группами, следует максимально приблизить исследование к условиям эксперимента. "Мешающими" считаются такие различия, которые скрывают от исследователя истинные причинно-следственные связи. Один из способов устранения таких различий иллюстрируется ниже.

Пример. Хотя серповидно-клеточная анемия (HbAS) обычно считается доброкачественным состоянием, некоторые исследования показали, что она обуславливает нарушения роста и интеллектуального развития. Поэтому было предпринято исследование, чтобы выяснить, действительно ли у детей с HbAS дефекты роста и развития наблюдаются чаще, чем у детей с нормальным гемоглобином (HbAA). По всем другим признакам две группы были идентичны [7]. Учитывалось, что с ростом, развитием и с наличием HbAS одновременно связан ряд других факторов: раса, пол, дата рождения, масса тела, гестационный возраст при рождении, оценка по шкале Апгар на 5-й минуте жизни, а также социально-экономический статус. Если бы все эти факторы не были приняты во внимание, то некоторые из них могли бы сместить результаты исследования так, что эффект HbAS не удалось бы отделить от эффектов других фак-

торов. Авторы исключили влияние других факторов, применив метод подбора соответствующих пар. Для каждого ребенка с НвАС они подобрали ребенка с НвАА, который не отличался от первого по семи другим факторам. Было прослежено всего 100 новорожденных (50 с НвАС и 50 с НвАА) до 3–5-летнего возраста. Никаких различий в росте и развитии детей обнаружено не было.

Основные источники систематических ошибок в обсервационных исследованиях и методы их устранения описываются в главе 6.

Сравнение рисков

По сути дела оценка риска совпадает с оценкой заболеваемости, которая определена в главе 4 как число новых случаев заболевания, возникших в популяции в течение заданного срока. Однако часто требуется сравнить заболеваемость в двух или более группах, отличающихся по степени воздействия предполагаемого фактора риска. Существует несколько способов оценки взаимосвязи между воздействием фактора риска и заболеванием. Эти способы обычно применяются для сравнения рисков и называются *оценками эффекта* (measures of effect). Они соответствуют разным концепциям риска и применяются для разных целей. Различают четыре вида оценки эффекта (табл. 5.3, 5.4).

Добавочный риск

Прежде всего можно спросить: "Какой добавочный риск развития заболевания связан с воздействием фактора риска помимо того, который существует для лиц, не подвергавшихся воздействию

Таблица 5.3
Оценки эффекта

Термин	Вопрос	Определение*
Добавочный риск (Разница рисков) (AR)	Какова заболеваемость, обусловленная воздействием фактора риска?	$AR = I_E - I_{\bar{E}}$
Относительный риск (Отношение рисков) (RR)	Во сколько раз заболеваемость лиц, подвергавшихся воздействию, выше, чем не подвергавшихся?	$RR = \frac{I_E}{I_{\bar{E}}}$
Добавочный популяционный риск (AR _p)	Какова заболеваемость в популяции, связанная с распространенностью фактора риска?	$AR_p = AR \times P$
Добавочная доля популяционного риска (AF _p)	Какая доля случаев заболевания в популяции обусловлена воздействием фактора риска?	$AF_p = \frac{AR_p}{I_T}$

* Где I_E – заболеваемость среди лиц, подвергавшихся воздействию фактора риска; $I_{\bar{E}}$ – заболеваемость среди лиц, не подвергавшихся этому воздействию; P – распространенность фактора риска; I_T – общая заболеваемость в популяции.

данного фактора? *Добавочный* (атрибутивный, attributable) риск определяется как заболеваемость у лиц, подвергавшихся воздействию фактора риска, минус заболеваемость у лиц, не подвергавшихся этому воздействию. Если допустить, что исходная заболеваемость связана с другими причинами, то добавочный риск – это дополнительные случаи развития заболевания, обусловленные воздействием фактора риска. Заметим, что при этом способе сравнения фактор риска рассматривается как причина болезни, а не просто ее маркер. Учитывая способ вычисления добавочного риска, его также называют *разницей рисков* (risk difference).

Таблица 5.4
Расчет оценки эффекта: курение и смертность от рака легкого*

Простые риски

Смертность от рака легкого среди курящих –	0,96/1000/год
Смертность от рака легкого среди некурящих –	0,07/1000/год
Распространенность курения –	56%
Общая смертность от рака легкого –	0,56/1000/год

Сравнительные риски

Добавочный риск = $0,96/1000/\text{год} - 0,07/1000/\text{год} = 0,89/1000/\text{год}$
 Относительный риск = $0,96/1000/\text{год} : 0,07/1000/\text{год} = 13,7$
 Добавочный популяционный риск = $0,89/1000/\text{год} \times 0,56 = 0,50/1000/\text{год}$
 Добавочная доля популяционного риска = $0,50/1000/\text{год} : 0,56/1000/\text{год} = 0,89$

* [Doll R., Hill A.B. Br Med J 1964;1:1399–1410].

Относительный риск

С другой стороны, можно спросить: "Во сколько раз выше заболеваемость среди лиц, подвергавшихся, чем среди не подвергавшихся воздействию фактора риска?" Для ответа на этот вопрос вводится понятие *относительного риска* (relative risk), или *отношения рисков* (risk ratio), т.е. отношения заболеваемости среди лиц, подвергавшихся и не подвергавшихся воздействию факторов риска. Относительный риск не несет информации о величине абсолютного риска (заболеваемости). Даже при высоких значениях относительного риска абсолютный риск может быть совсем небольшим, если заболевание редкое. Относительный риск показывает силу связи между воздействием и заболеванием. Таким образом, относительный риск – это мера влияния фактора риска, которая важна при изучении этиологии болезни.

Интерпретация оценки индивидуального риска

Понятия относительного и добавочного риска отличаются по клиническому смыслу. Применение того или иного показателя риска определяется постановкой вопроса.

Пример. Принято считать, что факторы риска сердечно-сосудистых заболеваний менее значимы для пожилых, чем для лиц среднего возраста. Это предположение было проверено путем сравнения относительных и добавочных рисков для хорошо известных факторов риска развития сердечно-сосудистых заболеваний в различных возрастных группах [8]. В качестве примера возьмем риск развития инсульта при курении (табл. 5.5). Относительный риск убывает от 4,0 для лиц в возрасте 45–49 лет до 1,4 для лиц в возрасте 65–69 лет. Однако добавочный риск с возрастом слегка увеличивается в основном за счет того, что у пожилых инсульт встречается чаще, независимо от курения. Таким образом, хотя причинная связь между курением и инсультом с возрастом становится слабее, у курящего пожилого человека повышается риск развития инсульта в такой же и даже несколько в большей мере, чем у молодого.

Поскольку добавочный риск отражает дополнительную вероятность заболевания, его применение в качестве индивидуального показателя риска в большинстве клинических ситуаций более информативно по сравнению с относительным риском. С другой стороны, относительный риск лучше показывает силу причинной связи.

Таблица 5.5

Сравнение относительного и добавочного рисков при изучении связи между курением, инсультом и возрастом*

Возраст	Заболеваемость (на 1000)		Относительный риск	Добавочный риск
	некурящие	курящие		
45–49	7,4	29,7	4,0	22,3
50–54	17,2	37,0	2,2	19,8
55–59	27,9	64,7	2,3	36,7
60–64	47,4	76,9	1,6	29,5
65–69	80,2	110,4	1,4	30,2

* [Psaty B.M. et al. J Clin Epidemiol 1990;43:961–970, с изменениями].

Популяционный риск

Мы можем посмотреть на риск и с другой стороны, задав вопрос: каков вклад фактора риска в общую заболеваемость группы людей, а не отдельных индивидуумов? Такого рода информация позволяет определить, какие факторы риска действительно важны, а какие не имеют особого значения для здоровья общества. Это помогает административным органам системы здравоохранения определять приоритеты при распределении ресурсов. Относительно слабый фактор риска (с низким относительным риском), но с высокой распространенностью в данной популяции может создать более значительную заболеваемость, чем сильный, но редкий фактор риска.

Для того чтобы оценить риск в популяции, необходимо знать, с какой частотой члены рассматриваемой популяции подвергаются воздействию фактора риска. *Популяционный добавочный* (атрибутивный) *риск* (population attributable risk) рассчитывается как произведение добавочного риска на распространенность фактора риска в популяции. Этот показатель отражает дополнительную заболеваемость в популяции, связанную с фактором риска. Кроме того, можно определить долю заболеваемости в популяции, связанную с данным фактором риска, т.е. добавочную долю популяционного риска. Она рассчитывается путем деления добавочного популяционного риска на общую заболеваемость в популяции.

Влияние распространенности фактора риска на соотношение между индивидуальным и популяционным риском иллюстрирует рис. 5.4, где показан добавочный риск смерти в зависимости от уровня диастолического артериального давления (рис. 5.4,А). Риск увеличивается при возрастании артериального давления. Однако очень высокое артериальное давление наблюдается лишь у немногих людей (рис. 5.4,Б). Если определить артериальную гипертензию как состояние, при котором диастолическое артериальное давление превышает 90 мм рт.ст., то окажется, что артериальная гипертензия – распространенное явление. Однако лишь у незначительного числа людей давление является чрезвычайно высоким (115 мм рт.ст. и более). В результате максимальный процент дополнительных смертей в популяции (58,4%) обусловлен мягкой артериальной гипертензией – 90–105 мм рт.ст. (рис. 5.4,В). Парадоксально, но получается, что врачи могли бы сохранить больше жизней посредством эффективного лечения больных с мягкой артериальной гипертензией, чем с высокой. Этот факт настолько противоречит интуитивному клиническому мышлению, что был назван парадоксом профилактики [9].

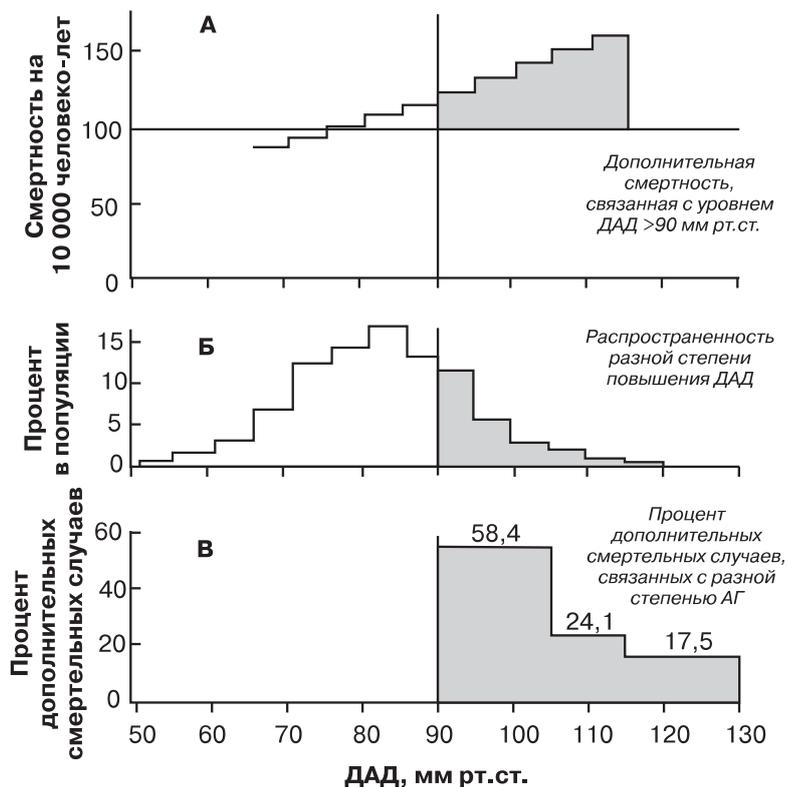


Рис. 5.4. Артериальная гипертензия: соотношение между добавочным риском, распространённостью фактора риска и популяционным риском [The Hypertension Detection and Follow-up Cooperative Group. Ann N Y Acad Sci 1978;304:254–266, с изменениями].

ДАД – диастолическое артериальное давление.

Оценки популяционного риска встречаются в клинической литературе реже, чем оценки индивидуального риска, а именно, добавочного и относительного рисков. Однако надо учитывать, что в клинической практике врача его больные точно так же составляют популяцию, как и население целого региона – для сотрудника системы здравоохранения. Кроме того, при лечении конкретных пациентов важно иметь в виду влияние распространенности фактора риска на риск для всего общества. Если больной не может рассказать историю своей болезни или же ему трудно связать ее с определенным воздействием, то вероятность того или иного заболевания оценивается на основании данных о рас-

пространности факторов риска. Например, когда речь идет о возможных причинах цирроза печени у больного в Северной Америке, логичнее предположить, что цирроз обусловлен злоупотреблением алкоголя, а не шистосомозом, поскольку обитатели этого материка редко имеют контакт с *Schistosoma mansoni*. Разумеется, совершенно другая ситуация в дельте Нила, где эти трематоды широко распространены, и люди, в основном мусульмане, редко употребляют алкоголь.

Резюме

Факторы риска связаны с повышением вероятности развития заболевания. Независимо от того, является ли некий фактор риска причиной заболевания, его присутствие позволяет прогнозировать вероятность возникновения болезни.

Воздействие большинства предполагаемых факторов риска нельзя изучать в эксперименте, поэтому обычно приходится проводить наблюдения, отслеживая воздействие факторов и оценивая частоту заболевания. Один из методов такого наблюдения состоит в отборе когорты, некоторые представители которой подвергались воздействию фактора риска. После определенного периода наблюдения сопоставляют частоту развития заболевания у лиц, подвергавшихся и не подвергавшихся воздействию этого фактора. Хотя когортные исследования риска с научной точки зрения предпочтительны, этот подход не всегда можно применить на практике, поскольку он требует значительных затрат времени, сил и средств.

Результаты сравнения заболеваемости в группах, отличающихся по воздействию фактора риска, можно представить с помощью нескольких показателей. Добавочный риск – это дополнительная заболеваемость, связанная с воздействием фактора риска. Относительный риск показывает, во сколько раз вероятность заболевания выше у лиц подвергавшихся, чем у не подвергавшихся воздействию фактора риска. Воздействие фактора риска на группу лиц зависит не только от силы фактора риска, но и от его распространенности.

Литература

1. Weiss N.S., Liff J.M. Accounting for the multicausal nature of disease in the design and analysis of epidemiologic studies. *Am J Epidemiol* 1983;117:14–18.
2. Prabhat K.S.J., Beral V., Peto J., Hack S., Hermon C., Deacon J. Antibodies to human papillomavirus and to other genital infectious agents and invasive cervical cancer risk. *Lancet* 1993;341:1116–1118.
3. Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985;14:32–38.

4. Diamond G.A., Forrester J.S. Analysis of probability as an aid in the clinical diagnosis of coronaryartery disease. *N Engl J Med* 1979;300:1350–1358.
5. Jaffe H.W. et al. National case-control study of Kaposi's sarcoma and *Pneumocystis carinii* pneumonia in homosexual men. Part I, Epidemiologic results. *Ann Intern Med* 1983;99:145–151.
6. Dawber T.R. The Framingham Study. The epidemiology of atherosclerotic disease. Cambridge, MA: Harvard University Press, 1980.
7. Kramer M.S., Rooks Y., Pearson H.A. Growth and development in children with sickle-cell trait. *N Engl J Med* 1978;299:686–689.
8. Psaty B.M. et al. Risk ratios and risk differences in estimating the effect of risk factors for cardiovascular disease in the elderly. *J Clin Epidemiol* 1990;43:961–970.
9. Hofman A., Vandenbroucke J.P. Geoffrey Rose's big idea. *Br Med J* 1992;305:1519–1520.

Рекомендуемая литература

- Detsky A.S., O'Rourke K., Corey P.N., Johnston N., Fenton S., Jeejeebhoy K.N. The hazards of using active clinic patients as a source of subjects for clinical studies. *J Gen Intern Med* 1988;3:260–266.
- Feinstein A.R. Scientific standards in epidemiologic studies of the menace of daily life. *Science* 1988;242:1257–1263. Response by Savitz D.A., Greenland S., Stolley P.D., Kelsey J.L. Scientific standards of criticism: a reaction to "Scientific standards in epidemiologic studies of the menace of daily life" by Feinstein. *Epidemiology* 1990;1:78–83.
- Malenka D.J., Baron J.A., Johansen S., Wahrenberger J.W., Ross J.M. The framing effect of relative and absolute risk. *J Gen Intern Med* 1993; 8:543–548.
- Morganstern H., Kleinbaum D.G., Kupper L.L. Measures of disease incidence used in epidemiologic research. *Int J Epidemiol* 1980;9:97–104.
- Naylor C.D., Chen E., Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916–921.

6 ПРОГНОЗ

У заболевшего сразу возникает множество вопросов: "Опасна ли болезнь? Есть ли риск умереть? Будет ли больно? Удастся ли сохранить обычный образ жизни? Возможно ли выздоровление?" Большинство пациентов и их родственников хотят знать, что их ожидает, даже если болезнь неизлечима.

Прогноз – это предсказание будущего течения болезни. В данной главе мы рассмотрим, как можно описать течение болезни, какие при этом возможны систематические ошибки и как их избегать. Мы хотим, чтобы читатели разобрались в трудной, но важной проблеме – как точнее предсказать будущее пациента. Цель состоит в том, чтобы прогноз не оказался, с одной стороны, слишком неопределенным, а с другой – уверенным, но ошибочным.

Говоря о прогнозе, врачи и пациенты имеют в виду несколько аспектов. Во-первых, они хотят знать обычное течение заболевания. Например, молодого пациента, страдающего постгерпетической невралгией, можно заверить, что боль продлится меньше месяца. Во-вторых, желательно, по возможности, иметь прогноз для конкретного случая. Ведь даже при смертельной ВИЧ-инфекции инфицированные могут жить от нескольких месяцев до 10 лет и более, так что больного интересует, сколько он сможет прожить. В-третьих, пациенты особенно хотят знать, как заболевание повлияет на их жизнь, причем их волнует вопрос не только о риске смерти, но и смогут ли они работать, ходить, говорить, изменятся ли их взаимоотношения с семьей и друзьями, предостоят ли им боль и страдания.

Прогностические исследования

Перечисленные вопросы рассматриваются в прогностических исследованиях аналогично тому, как когортные исследования решают проблемы риска. Для этого формируется группа пациентов, имеющих некоторый общий признак (для прогностических исследований – это заболевание или состояние), затем проводится наблюдение в течение определенного срока с оценкой клинических исходов. Часто выявляются состояния, связанные с конкретным исходом, т.е. *прогностические факторы* (prognostic factors).

Клиническое и естественное течение болезни

Прогноз заболевания может быть сделан либо для клинического, либо для естественного течения болезни. Под *клиническим течением* (clinical course) подразумевается развитие болезни при оказании пациенту медицинской помощи, т.е. когда на ход событий влияет лечение. Обычно пациенты обращаются за медицинской помощью через некоторое время после начала болезни, при появлении таких симптомов, как боль, недомогание, внешние дефекты или необычное поведение. В качестве примеров упомянем сахарный диабет, рак легкого, бешенство. После установления диагноза обычно назначают соответствующее лечение.

В отсутствие медицинского вмешательства прогноз относится к *естественному течению* (natural history) болезни, т.е. к тому, что произойдет с пациентом, если он не будет лечиться. Даже в странах с развитой системой медицинской помощи многие больные не получают лечения. Заболевания могут оставаться нераспознанными из-за того, что протекают бессимптомно или рассматриваются как обычные недомогания. Примерами служат легкая степень депрессии, анемия или некоторые злокачественные новообразования, развивающиеся медленно и незаметно (некоторые виды опухолей щитовидной или предстательной железы).

Точка отсчета

Наблюдение когорты в прогностических исследованиях производится начиная с некоторого момента, называемого *точкой отсчета* (zero time). Эта точка должна быть одинаковой для всех пациентов и должна соответствовать четко определяемому моменту в течении заболевания (например, появлению симптомов, установлению диагноза или началу лечения). Термином "*исходная когорта*" (inception cohort) обозначают группу, сформированную из пациентов в начале (в момент первого проявления) заболевания.

Если наблюдение разных пациентов в когорте начинается в разные моменты течения заболевания, то последующее описание теряет точность. Сопоставление сроков таких событий, как выздоровление, рецидив или смерть будет затруднительным или ошибочным.

Допустим, что нужно описать клиническое течение заболевания у больных раком легкого. Наберем когорту пациентов с этим заболеванием и будем наблюдать их вплоть до развития осложнений или смерти. Но каких больных следует включать? Если точкой отсчета для одних пациентов станет обнаружение бо-

лезни при скрининге, для других – появление симптомов, для третьих – госпитализация или начало лечения, то получаемый при исследовании прогноз будет зависеть от сочетания нескольких точек отсчета. Более того, если точно не описать, в какой момент заболевания пациенты включались в когорту, то будет невозможно интерпретировать или использовать полученный прогноз.

Описание исходов заболевания

Описание прогноза должно содержать полный спектр проявлений состояний, важных для пациента. Это не только болезнь и смерть, но и боль, страдание, неспособность обслуживать себя или совершать обычные действия. (Пять основных клинических исходов перечислены ранее – см. табл. 1.2 на стр. 16.)

В стремлении к наукообразию врачи иногда неоправданно преувеличивают значение отдельных исходов. Клинические эффекты, которые не могут прямо осознаваться пациентами (например, уменьшение размера опухоли, нормализация биохимических и серологических показателей крови), не являются целью лечения. Эти биологические феномены целесообразно рассматривать в качестве клинически значимых исходов только в том случае, если известно, что между ними существует причинно-следственная связь. Так, гиперкальциемия является важным клиническим исходом в течении гиперпаратиреоза, только если она вызывает такие симптомы, как сонливость или жажда, либо если есть основания полагать, что она приведет к осложнениям в виде заболеваний почек и костей. Если же исход нельзя связать с каким-либо признаком, который ощущается пациентом, то эта информация не должна использоваться для оценки лечения, хотя она может быть весьма важной для оценки этиологии и патогенеза заболевания.

Оценка качества жизни, связанного со здоровьем

Люди все лучше осознают, что "здоровье" – это не просто отсутствие таких событий, как болезнь или тем более смерть. Клинические вмешательства должны положительно влиять на жизнь и существование пациента, что определяется понятиями "качество жизни", "состояние здоровья", или "функциональный статус". Для оценки качества жизни больных были разработаны опросники. Иногда в результате опросов усиливаются аргументы в пользу некоторых клинических вмешательств. Например, было показано [1], что при хронической почечной недостаточности лечение эритропоэтином не только повышает у пациентов гематокрит, но и улучшает качество жизни. С другой стороны, иногда при оценке

качества жизни обнаруживаются сложные соотношения положительных и отрицательных феноменов. Анализ результатов лечения зидовудином пациентов со слабовыраженными симптомами ВИЧ-инфекции показал, что хотя препарат задерживает развитие СПИДа в среднем на 0,9 мес, положительный эффект сводится на нет побочными действиями препарата. Так, пациенты, получавшие зидовудин, жили в среднем 14,5 мес без прогрессирования заболевания или развития серьезных побочных эффектов от действия лекарства, в то время как пациенты, не получавшие препарат, жили 14,7 мес [2]. Таким образом, достижение в виде небольшой задержки развития СПИДа стало сомнительным, когда в качестве исхода стали оценивать качество жизни.

Прогностические факторы

Хотя большинство пациентов интересуются тем, как вообще протекает их заболевание, каждого особенно волнует его индивидуальный прогноз. *Прогностические факторы* помогают идентифицировать группы пациентов с одной болезнью, но с разным прогнозом.

Различия между прогностическими факторами и факторами риска

Исследования факторов риска обычно проводятся на здоровых людях, тогда как прогностические факторы (состояния, связанные с конкретными исходами заболевания), по определению, изучаются у больных. Существуют и другие важные различия, о которых речь пойдет далее.

Различия факторов

Факторы риска не обязательно совпадают с прогностическими факторами, причем для одного и того же заболевания эти факторы могут быть совершенно разными. Например, низкое артериальное давление снижает риск возникновения острого инфаркта миокарда, но служит неблагоприятным прогностическим фактором при развившемся остром инфаркте (рис. 6.1). Аналогичным образом применение эстрогенов во время менопаузы увеличивает риск заболевания раком матки, но возникающие при этом формы рака выявляются на более ранних стадиях и их прогноз, по-видимому, более благоприятен, чем в среднестатистическом случае.

Некоторые факторы одинаково влияют как на риск, так и на прогноз. Например, риск возникновения острого инфаркта миокарда и риск смерти от него увеличиваются с возрастом.

Различия исходов

Риск и прогноз описывают различные явления. Для риска учитываемым событием является возникновение заболевания. С точки зрения прогноза рассматривается множество исходов, включая смерть, осложнения, инвалидность, страдание.

Различия показателей частоты

Факторы риска обычно предсказывают события малой вероятности (например, заболеваемость для многих болезней составляет от 1 на 100 до 1 на 10 000 случаев в год), поэтому соотношения между воздействием фактора риска и заболеванием обычно остаются неясными даже для опытных клиницистов, пока они не возьмут данные тщательно проведенных исследований, включающих большое число пациентов в течение длительного периода времени. Прогноз оперирует относительно частыми событиями. Обычно врач может самостоятельно дать точные оценки прогноза на основе личного опыта. Например, известно, что больные раком легкого или поджелудочной железы редко живут более 5 лет, тогда как большинство пациентов с хроническим лимфолейкозом живут значительно дольше.

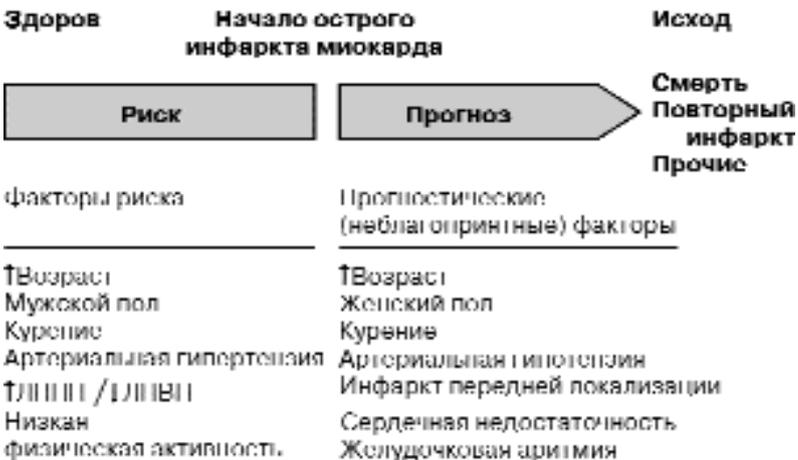


Рис. 6.1. Различия между прогностическими факторами и факторами риска для острого инфаркта миокарда.

ЛПНП – липопротеиды низкой плотности

ЛПВП – липопротеиды высокой плотности.

Множественные прогностические факторы и правила прогнозирования

Сочетание факторов обычно дает более точный прогноз, чем каждый из факторов, взятый по отдельности. Правила клинического прогнозирования позволяют оценивать вероятность исхода на основании набора характеристик пациента.

Пример. Если у ВИЧ-инфицированного пациента развился СПИД, то прогноз неблагоприятен, время выживания короткое. Однако прежде чем применение антивирусных средств и терапии, направленной против сопутствующих инфекций, стало стандартным лечением, выяснилось, что некоторые больные СПИДом живут значительно дольше, чем другие. Было проведено исследование для определения того, какие характеристики пациентов влияют на выживаемость [3]. С выживаемостью оказалось связано несколько физиологических характеристик. Используя сочетания этих факторов, исследователи разработали многоступенчатую прогностическую систему, оценивая в 1 балл присутствие каждого из 7 факторов:

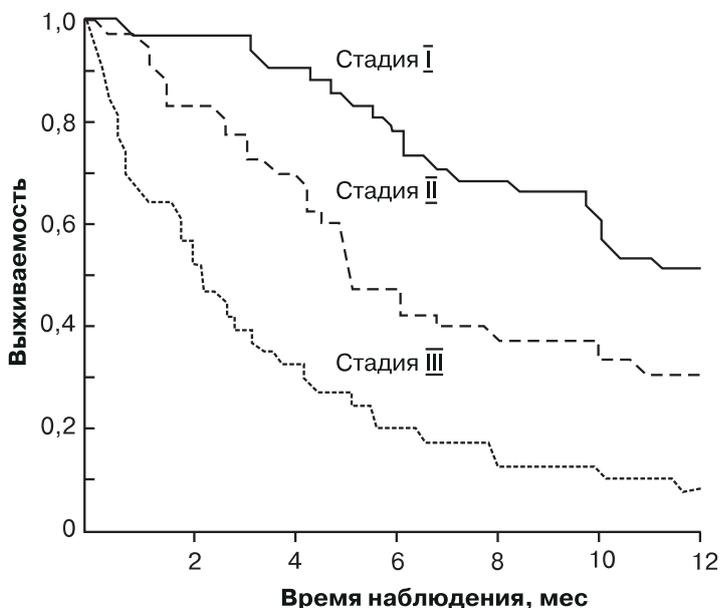


Рис. 6.2. Выживаемость больных СПИДом в зависимости от прогностической стадии. Медиана выживаемости (мес): стадия I — 11,6; стадия II — 5,1; стадия III — 2,1 [Justice A.C., Feinstein A.R., Wells C.K. A new prognostic staging system for the acquired immunodeficiency syndrome. N Engl J Med 1989;320:1388—1393, с изменениями].

сильная диарея или содержание сывороточного альбумина менее 2,0 г/дл, любые неврологические расстройства, насыщение крови кислородом 50 мм рт.ст. и менее, гематокрит менее 30%, число лимфоцитов менее 150 мл⁻¹, число лейкоцитов менее 2500 мл⁻¹, число тромбоцитов менее 140 000 мл⁻¹. Были выделены три прогностические стадии в зависимости от суммы набранных баллов: I стадия – 0 баллов, II стадия – 1 балл, III стадия – 2 балла и более. На рис. 6.2 представлена выживаемость больных СПИДом в трех прогностических стадиях. Учитывая все прогностические факторы, авторы отмечают, что предсказанная средняя выживаемость варьирует от 11,6 мес для пациентов I стадии до 2,1 мес для пациентов III стадии.

Описание прогноза

Прогноз как частота

В качестве характеристики течения болезни удобно использовать один показатель – частоту, т.е. долю лиц, у которых отмечается изучаемое событие. Показатели, которые обычно служат для этих целей, приведены в табл. 6.1. Они основаны на одной и той же концепции заболеваемости и учитывают события, возникающие в когорте с течением времени.

При оценке частоты необходимо четко определить точку отсчета, специфические клинические характеристики пациентов, виды исходов, продолжительность наблюдения. Срок наблюдения должен быть достаточно большим, чтобы произошли все возможные события, иначе полученный показатель будет ниже истинной частоты.

Таблица 6.1

Показатели, используемые для описания прогноза

Показатель	Определение*
Пятилетняя выживаемость	Число больных, проживших 5 лет после некоторого момента течения болезни, %
Летальность	Число умерших от данного заболевания, %
Смертность от заболевания	Число умерших от данной болезни, на 10 000 (или 100 000) населения
Эффективность лечения	Число больных, у которых наблюдалось улучшение состояния в ответ на лечение, %
Ремиссия	Число больных, у которых исчезли все проявления болезни, %
Рецидив	Число больных, у которых заболевание появилось вновь после выздоровления, %

* Срок наблюдения либо указывается, либо имеется в виду, что он достаточно продолжителен для проявления всех исходов.

Компромисс: простота против более подробной информации

Достоинство выражения прогноза в виде частот заключается в простоте этих показателей. Их легко удержать в памяти и использовать в сообщениях. Недостаток заключается в том, что показатели относительно малоинформативны и за одинаковыми обобщенными цифрами могут скрываться значительные различия прогноза.

На рис. 6.3 представлена пятилетняя выживаемость пациентов с четырьмя разными состояниями, при каждом из которых через 5 лет выживает примерно 10%. Однако такие сходные суммарные показатели пятилетней выживаемости скрывают различия, чрезвычайно важные для пациентов. Ранняя выживаемость паци-

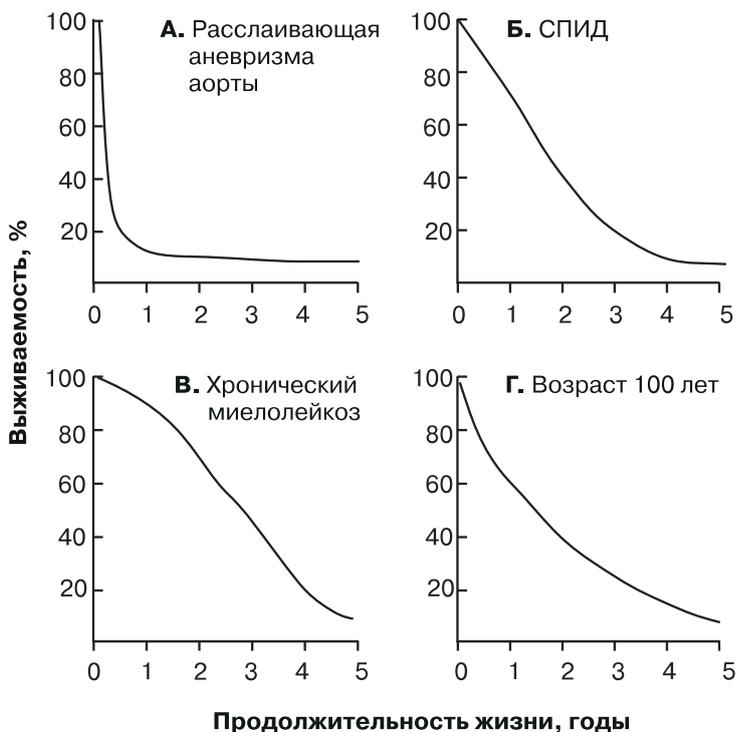


Рис. 6.3. Ограниченность показателей пятилетней выживаемости: четыре состояния с одинаковой пятилетней выживаемостью, равной 10% [Anagnostopoulos C.D. et al. Aortic dissections and dissecting aneurysms. *Am J Cardiology* 1972;30:263—273; Saah J.A., Hoover D.R. et al. Factors influencing survival after AIDS: report from the Multicenter AIDS Cohort Study (MACS). *J Acquir Immune Defic Syndr* 1994;7:287—295; Kardinal C.G. et al. Chronic granulocytic leukemia. Review of 536 cases. *Arch Intern Med* 1976;136:306—313; American College of Life Insurance. 1979 life insurance fact book. Washington, DC:ACLI 1979].

ентов с расслаивающей аневризмой аорты очень низка, однако если они выжили в течение первых нескольких месяцев, то в дальнейшем риск смерти уже не зависит от наличия аневризмы (рис. 6.3,А). С другой стороны, ВИЧ-положительные пациенты, у которых развился СПИД, умрут в течение 5 лет (рис. 6.3,Б). Хронический миелолейкоз относительно мало влияет на выживаемость в первые годы после установления диагноза (рис. 6.3,В). В последующие годы летальность увеличивается, так что практически все пациенты умирают через 5 лет после установления диагноза. Только в возрасте 100 лет люди в общей популяции имеют такую же пятилетнюю выживаемость, как пациенты с указанными тремя заболеваниями (рис. 6.3,Г).

Анализ дожития

Занимаясь прогнозированием, мы хотели бы знать среднюю вероятность изучаемого исхода у пациентов с данным заболеванием для любого момента времени. Если прогноз выражен в виде обобщенного показателя частоты, он не содержит такой информации. Однако существуют методы представления информации о среднем времени наступления исхода.

Дожитие когорты

Непосредственный способ определения выживаемости состоит в том, чтобы набрать когорту пациентов в некоторый момент течения заболевания (например, появление симптомов, установление диагноза, начало лечения) и наблюдать их, пока у всех не появится изучаемый исход. Для небольших когорт результаты наблюдения можно представить так, как это сделано на рис.6.4,А. График дожития состоит из ступеней, соответствующих смерти

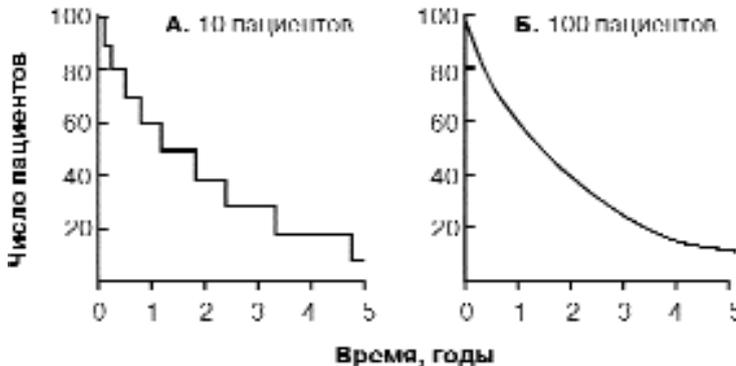


Рис. 6.4. Выживаемость в двух когортах, малой и большой численности, члены которых прослежены в течение всего периода наблюдения.

каждого из 10 пациентов когорты. Если число пациентов увеличить (рис.6.4,Б), то размер ступеней будет уменьшаться. При очень большом числе пациентов график приблизится к гладкой кривой. Эта информация затем может использоваться для ежегодного или даже еженедельного прогноза для пациентов с таким же заболеванием.

К сожалению, получить подобную информацию на практике сложно. Некоторые из пациентов выйдут из исследования до окончания периода наблюдения – из-за другого заболевания, перемены места жительства или вследствие нежелания участвовать. Этим пациентам придется исключить из когорты, несмотря на усилия, затраченные на сбор данных о них до момента выбывания. Кроме того, необходимо ждать, пока все члены когорты достигнут того срока, для которого требуется вычислить выживаемость. Поскольку обычно пациенты включаются в исследование на протяжении некоторого периода времени, то для любой календарной даты срок наблюдения будет относительно большим для пациентов, включенных в исследование первыми, и коротким – для вступивших в исследование позднее. Прежде чем станет доступной информация о выживаемости за каждый последующий год, последний включенный пациент должен его прожить.

Кривые дожития

Для эффективного использования всех данных о каждом пациенте в когорте применяют метод оценки выживаемости, называемый анализом дожития (обычно метод называют по имени его авторов анализом Каплана–Мейера). Цель анализа дожития не всегда состоит в том, чтобы отразить только выживание или смерть пациентов. Метод позволяет описать любой исход, дихотомический по своему характеру и имеющий место лишь однажды в процессе наблюдения, например время до развития инфаркта миокарда или до рецидива злокачественного новообразования. Если метод применен не собственно для анализа выживания, а для оценки иных событий, иногда используется термин "анализ времени наступления события".

На рис.6.5 показана типичная кривая дожития. Для того чтобы дать представление о вкладе случайности в наблюдаемые показатели, отмечают число пациентов, остающихся под наблюдением к началу каждого периода.

Вероятность дожить до некоторого момента оценивается как суммарная вероятность дожития в предшествующие временные интервалы. Эти интервалы могут быть заданы сколь угодно малыми: при анализе Каплана–Мейера интервалы соответствуют

промежутку времени между наступлением каждого нового события (например, смерти) и предшествующим событием. На протяжении большинства интервалов не умирает никто, следовательно, вероятность дожития равна единице. Когда умирает один или более больных, вероятность дожития рассчитывается как отношение числа выживших к числу наблюдавшихся в начале интервала. Пациенты, которые уже умерли, выбыли или не наблюдались необходимое время, подвергаются цензурированию (censoring) и их не учитывают при оценке дожития для данного временного интервала. Цензурированные пациенты считаются выбывшими, т.е. в дальнейшем они не учитываются в знаменателе¹. Вероятность дожития не изменяется в течение тех интервалов, где ни один пациент не умер, поэтому на практике она рассчитывается только для интервалов, в которых есть случаи смерти. Хотя вероятность, относящаяся к каждому интервалу, не очень точна из-за малого числа происшедших в нем событий, общая вероятность дожития до определенного момента времени (представляющая

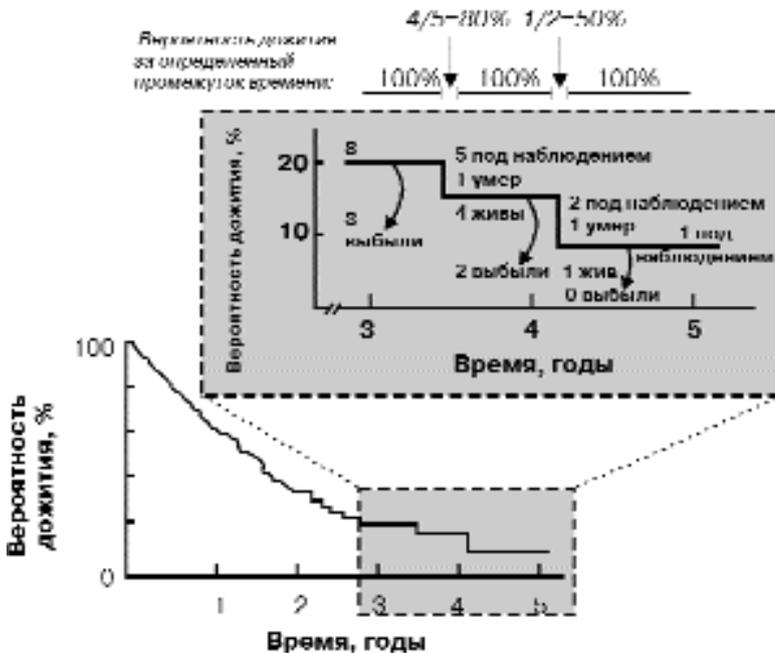


Рис. 6.5. Типичная кривая дожития с детализацией одного фрагмента.

¹ Процесс исключения выбывших пациентов из анализа данных называют *цензурированием*. – Примеч. ред.

собой произведение всех предшествующих вероятностей) является весьма точной.

Часть кривой дожития (см. рис. 6.5 между 3-м и 5-м годом от точки отсчета) представлена более подробно, чтобы проиллюстрировать данные, используемые при оценке дожития: число наблюдающихся больных, число выбывших больных, а также число больных, у которых развился рассматриваемый исход для каждого момента времени.

Интерпретация кривых дожития

При интерпретации кривых дожития надо иметь в виду следующее. Во-первых, на вертикальной оси откладываются показатели оцениваемой вероятности дожития для членов гипотетической когорты, а не процент выживающих в реальной когорте. Во-вторых, точки на кривой дожития лучше всего (для имеющейся совокупности данных) оценивают вероятность дожития членов когорты. Однако точность этих оценок зависит, как всегда при выборочных исследованиях, от числа наблюдений, на которых основана оценка. В левой части кривой оценки должны быть более надежны, поскольку в этот начальный период число находящихся под наблюдением пациентов бывает больше. В правой части кривой число пациентов, по которым оценивается дожитие, часто становится весьма незначительным из-за смертей, выбывания или позднего включения в исследование. Следовательно, к концу периода наблюдения оценки дожития становятся неточными и сильно зависят от того, что происходит с оставшимся относительно малым числом пациентов. Например, как видно из рис. 6.5, вероятность прожить 5 лет составляет 8%. Если в этот период один из оставшихся пациентов умрет, то вероятность дожития упадет до нуля. Естественно, такому результату нельзя полностью доверять. Итак, показатели дожития в правой части кривых следует интерпретировать с осторожностью.

Наконец, форма некоторых кривых дожития, особенно таких, где изучаемое событие наступает у большинства пациентов, создает впечатление, что события случаются чаще в начале, чем в более поздний период, когда кривая достигает плато, и кажется, что риск событий значительно снизился. Однако это впечатление обманчиво. Со временем показатели выживаемости применяются к уменьшающемуся числу пациентов, так что кривая становится более пологой, даже если частота исходов не изменяется.

В медицинской литературе встречаются разные типы кривых дожития (рис. 6.6). Часто на вертикальной оси откладывается доля лиц с наблюдаемым исходом, а не доля лиц, у которых он не наступил; в этом случае кривая поднимается вправо вверх.

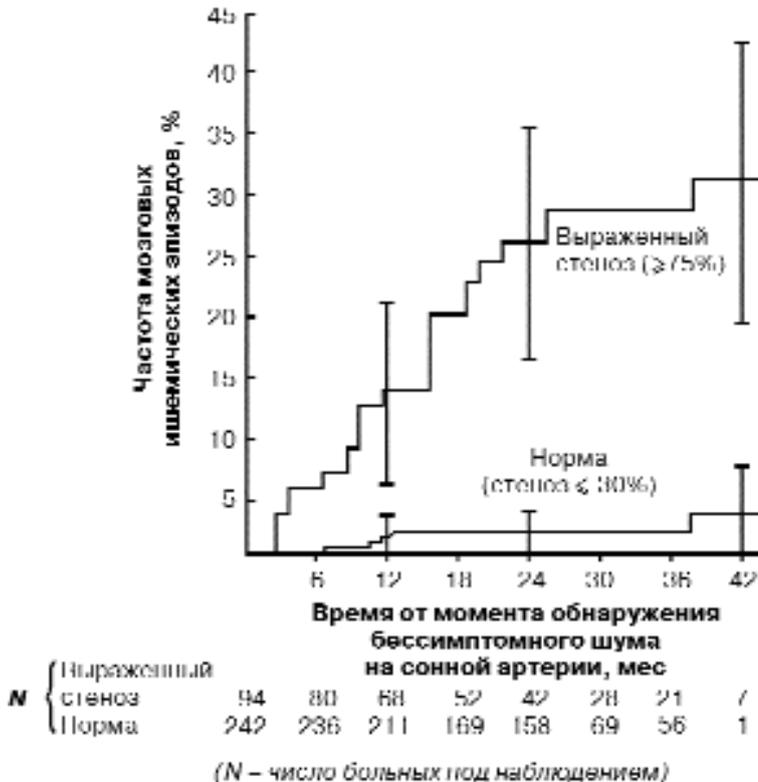


Рис. 6.6. Кривые дожития в двух когортах. Показаны число пациентов, находящихся под наблюдением, и 95% доверительные интервалы для изучаемых показателей. Кривые отражают суммарную вероятность мозговых ишемических эпизодов начиная с момента установления диагноза в зависимости от степени стеноза сонной артерии [Chambers B.R., Norris J.W. Outcome in patients with asymptomatic neck bruits. N Engl J Med 1986;315:860 – 865].

Объем информации, представляемой на кривой, можно увеличить. Число пациентов, находящихся под наблюдением в различные моменты времени, может быть показано под горизонтальной осью; точность оценок дожития, убывающая со временем из-за того, что все меньшее число пациентов остаются под наблюдением, может быть показана в виде доверительных интервалов (см. главу 9). Наконец, кривые дожития для пациентов с разными характеристиками (например, с разными прогностическими факторами или разным лечением) могут быть помещены для сравнения на одном рисунке. Иногда на кривых дожития отмечают момент, когда пациент выбыл из-под наблюдения.

Кривые дожития можно построить для сочетаний прогностических факторов. Для этого пациентов группируют по наличию или отсутствию оцениваемых прогностических факторов, как упоминалось ранее в этой главе. Статистическая методика, известная как *регрессионная модель пропорционального риска Кокса* (Cox proportional hazards regression model), может применяться для оценки роли сочетания факторов, наилучшим образом прогнозирующих события в группах изучаемых пациентов, а также независимого эффекта отдельных факторов (см. главу 9).

Систематические ошибки в когортных исследованиях

Возможность систематической ошибки существует при любых наблюдениях. Систематическая ошибка в когортном исследовании, независимо от того, изучается риск или прогноз, может создать видимость различий, когда в действительности их нет, либо скрыть различия, которые на самом деле существуют.

Обнаружить систематическую ошибку легче, если знать, где наиболее вероятно ее возникновение в ходе исследования. Во-первых, важно определить, может ли систематическая ошибка вообще присутствовать при данных условиях исследования. Во-вторых, нужно установить, имеется ли она в данном исследовании. В-третьих, следует решить, вызовет ли эта ошибка клинически значимое искажение результатов. Если результаты не будут искажены слишком сильно, это позволит избежать неверных выводов и опасных последствий. Некоторые характерные источники систематических ошибок в когортных исследованиях показаны на рис. 6.7 и описаны далее.

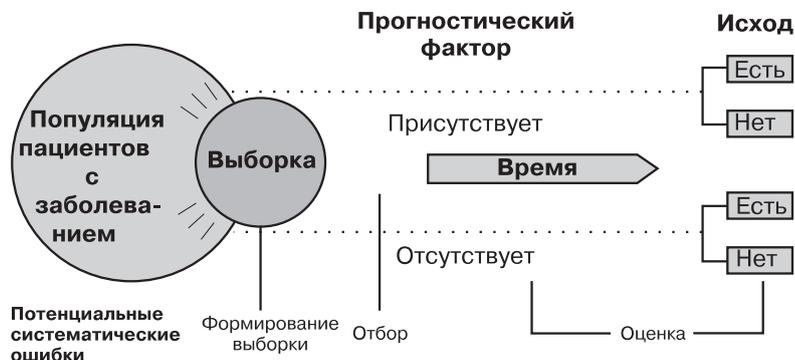


Рис. 6.7. Возможные источники возникновения систематических ошибок в когортных исследованиях.

Систематическая ошибка, возникающая при формировании выборки

Один из видов систематической ошибки возникает в процессе формирования когорты, когда группы пациентов, включенных в исследование, отличаются друг от друга по иным характеристикам, нежели изучаемые факторы. В этом случае результат может определяться именно этими не изучаемыми в исследовании характеристиками.

Систематическая ошибка, возникающая при формировании выборки в прогностических исследованиях, может быть обусловлена одним или несколькими различиями между когортами, включая тяжесть заболевания, наличие сопутствующих заболеваний, стадию развития болезни, предшествующее лечение. Следующий пример иллюстрирует, как эта ошибка была выявлена в исследовании по оценке прогностической значимости сывороточного карциноэмбрионального антигена (КЭА) у больных раком толстой кишки.

Пример. Повышенные уровни КЭА обнаружены для нескольких типов рака, включая рак толстой кишки. Было проведено исследование, для того чтобы определить, возможно ли предсказание по дооперационным уровням КЭА рецидива заболевания после удаления опухоли [4]. Уровни КЭА коррелировали со стадией заболевания, обычно оцениваемой по классификации Dukes: А – опухоль ограничена стенкой кишки, В – опухоль прорастает стенку кишки, но не поражает лимфатические узлы, С – вовлечены регионарные лимфатические узлы, D – опухоль дает отдаленные метастазы. Средние уровни КЭА изменялись в зависимости от стадии заболевания следующим образом: 4 нг – стадия А, 9 нг – стадия В, 32 нг – стадия С и 251 нг – стадия D. Как стадия процесса, так и уровни КЭА хорошо прогнозировали рецидив заболевания. Однако является ли уровень КЭА самостоятельным прогностическим признаком, независимым от стадии заболевания, или вероятность рецидива зависит только от стадии по Dukes? Для ответа на этот вопрос была исследована связь между дооперационными уровнями КЭА и рецидивами в пределах каждой стадии по Dukes. На рис. 6.8 показано, что для стадии В уровни КЭА независимо предсказывают рецидив. Аналогичные результаты получены для стадии С. Следовательно, связь между уровнями КЭА и вероятностью рецидива рака толстой кишки нельзя свести к систематической ошибке при формировании выборки, по крайней мере для стадий В и С, т.е. уровень КЭА является важным независимым прогностическим фактором.

Когорты дожития

Истинные когортные исследования следует отличать от исследований *когорт дожития* (survival cohort), в которые включа-

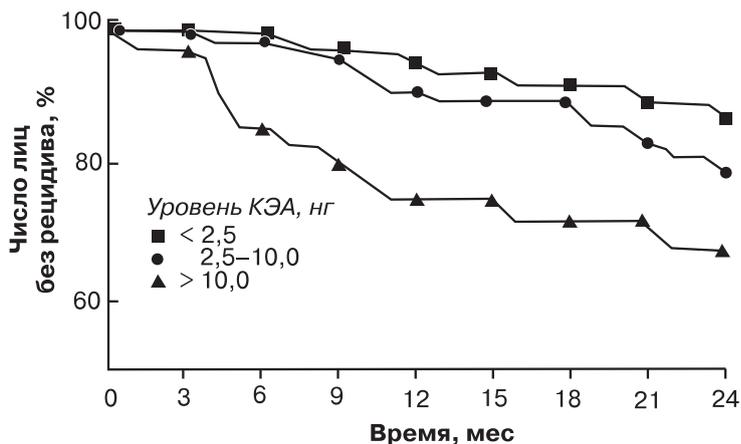


Рис. 6.8. Безрецидивная выживаемость больных раком толстой кишки (стадия В по Дюке) в зависимости от уровня КЭА [Wolmark N. Ann Surg 1984;199:375 – 382].

ются пациенты, имеющие заболевание и доступные для наблюдения, – например, пациенты специализированной клиники. Другой термин для таких групп – *“когорты доступных пациентов”* (available patients cohort). Публикации о когортах дожития вводят в заблуждение, если их представляют как истинные когорты. В когорту дожития больные включаются на разных этапах течения заболевания, а не в его начале, как при истинном когортном исследовании. Клиническое течение затем описывается путем ретроспективного анализа состояния пациентов от начала заболевания до настоящего времени (рис. 6.9).

Данные, полученные на когортах дожития, часто представляют так, как если бы они описывали течение заболевания с самого его начала. Однако эти результаты могут быть подвержены систематической ошибке, поскольку когорта состоит из пациентов, которые включились в наблюдение уже спустя некоторое время после начала болезни. При смертельных заболеваниях когорта дожития состоит из пациентов, которые дожили до начала исследования и, следовательно, доступны для последующего наблюдения. При болезнях, протекающих с ремиссиями, в когорту попадают те больные, у которых ремиссии отсутствуют. В итоге, исследования когорт дожития описывают прошлую историю распространенных заболеваний и не дают нам представлений о том, как будет развиваться болезнь через некоторое время от ее начала. Таким образом, когорта дожития есть особый случай систематической ошибки при формировании выборки.

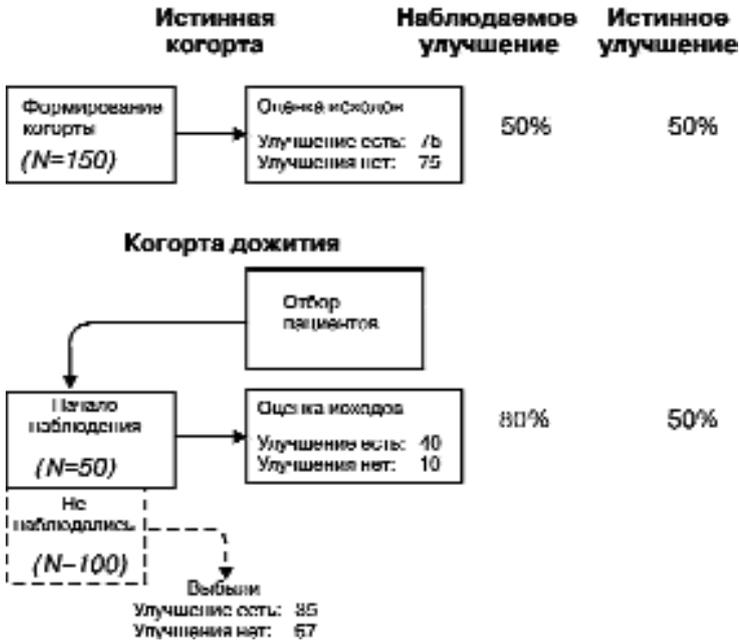


Рис. 6.9. Сравнение истинных когорт и когорт дожития. В когортах дожития некоторые пациенты, имеющие заболевание, выпадают из-под наблюдения.

Публикации о когортах дожития относительно часты в медицинской литературе, в частности, в виде описаний серий случаев (см. главу 10). Такие исследования могут давать важные результаты, особенно при публикации первых сведений о новых болезнях, но это только предварительные, а не окончательные наблюдения.

Пример. Возникло подозрение, что силиконовые протезы молочной железы могут быть причиной аутоиммунных ревматоидных реакций. Было обследовано 156 женщин с силиконовыми протезами молочных желез и с жалобами, характерными для ревматических болезней [5]. Исследуемая группа состояла из пациенток, обратившихся к трем ревматологам, которые были известны как специалисты по диффузным заболеваниям соединительной ткани у женщин с силиконовыми протезами. Результаты серологических тестов у этих больных сравнивались с таковыми у пациенток без протезов, но с фибромиалгией, а также у пациенток с протезами, но без признаков ревматического заболевания. Анализ данных, полученных у пациенток с протезами и имеющих жалобы, характерные для ревма-

тического заболевания, показал, что у большинства из них отсутствуют критерии диагностики ревматоидного артрита, а иммунологические показатели находятся в пределах нормы. Однако у 14 пациентов выявлены склеродермия и аномальные серологические показатели, чего не наблюдалось в группах сравнения.

Из-за возможных систематических ошибок при формировании групп авторы были осторожны в выводах: "Предположения, высказанные в этой и других аналогичных работах, следует проверить в крупномасштабных популяционных исследованиях". Публикация результатов первого же большого исследования не подтвердила выдвинутую гипотезу [6].

Систематическая ошибка вследствие миграции

Эта форма систематической ошибки может возникнуть, если пациенты одной из групп покидают ее, выбывая из исследования или переходя в другую группу. Если такие изменения значительны, они могут повлиять на достоверность выводов.

Почти во всех исследованиях некоторые члены исходной группы выбывают с течением времени. Если выбывание происходит случайным образом, так что характеристики лиц, выбывших из сравниваемых групп, в среднем одинаковы, то систематическая ошибка не возникнет, независимо от того, велико ли и одинаково ли число лиц, выбывших из каждой группы. Но обычно характеристики выбывших для разных групп различны. Причины выбывания – смерть, выздоровление, осложнения терапии и т. д., часто связаны с прогнозом и могут повлиять на характеристики одной группы сильнее, чем другой. В результате группы, сопоставимые вначале, через некоторое время перестают удовлетворять этому требованию.

По мере увеличения доли членов когорты, выбывающих из исследования, возрастает и вероятность систематической ошибки. Величину этой ошибки нетрудно оценить. Для этого нужно знать число лиц в когорте, число выбывших из-под наблюдения и наблюдаемую частоту исходов.

Пример. Thompson et al. [7] описали отдаленные результаты гастроэктомии. Когорта из 123 пациентов с патологическим ожирением обследовалась через 19 – 47 мес после операции. За положительный результат принималась потеря более 30% избыточной массы тела.

Удалось проследить только 103 (84%) пациентов. Из них доля успешно леченных составила 60 из 103 (58%). Для того чтобы определить диапазон, в котором может находиться истинный показатель эффективности операции, авторы провели анализ "наилучший вариант – наихудший вариант" (best case / worst case analysis). Частота успешных операций подсчитывалась при допущении, что все вы-

бывшие из-под наблюдения имели, в одном варианте, положительный результат (наилучший вариант) или, в другом варианте – отрицательный результат (наихудший вариант). Из общей когорты (123 пациента) 103 пациента завершили исследование и 20 выбыли. Наблюдаемая доля успешных результатов составила 60 на 103, или 58%. В наилучшем варианте все 20 выбывших больных считаются успешно леченными, и тогда показатель эффективности будет (60+20) на 123, или 65%. В наихудшем варианте все 20 больных считаются безуспешно леченными, и тогда этот показатель равен 60 на 123, или 49%. Таким образом, истинный показатель эффективности операции должен находиться между 49 и 65% и, по-видимому, он близок к 58%, так как маловероятно, что все выбывшие оказались либо успешно, либо безуспешно леченными.

В процессе наблюдения пациенты иногда переходят из одной группы в другую в пределах когорты. При этом основания, по которым они были исходно отнесены к той или иной группе, уже не имеют значения. Если обмен больными между группами значителен, то наблюдаемое различие риска между группами может оказаться меньше того, которое получилось бы при неизменном составе групп. Систематическая ошибка вследствие миграции, обусловленная взаимными переходами пациентов, чаще составляет проблему в исследованиях риска, чем в прогностических исследованиях, поскольку исследования риска обычно продолжаются в течение многих лет. С другой стороны, факт перехода больных из одной группы в другую может быть учтен при анализе результатов исследования.

Пример. Исследование связи между образом жизни и смертностью было организовано следующим образом: 10 269 выпускников Гарвардского университета были обследованы для оценки физической активности, курения, массы тела и артериального давления [8]. Обследование проводилось дважды – в 1966 и 1977 г. Затем изучалась смертность за 9 лет – с 1977 по 1985 г. Авторы понимали, что исходные характеристики могут изменяться, затушевывая связь между образом жизни и смертностью. Для того чтобы разрешить эту проблему, исследователи распределили обследуемых на 4 категории: лица, сохраняющие образ жизни с высоким риском; лица, перешедшие по образу жизни с низкого уровня риска на высокий; лица, перешедшие с высокого уровня риска на низкий; лица, сохраняющие низкий уровень риска. После поправки на другие факторы риска оказалось, что у тех выпускников, которые увеличили физическую активность до умеренного уровня, бросили курить, нормализовали массу тела и/или кровяное давление, смертность была ниже, чем у тех, кто сохранял или приобрел образ жизни с высоким риском, но тем не менее все же выше, чем у лиц, никогда не имевших факторов риска.

Систематическая ошибка, возникающая при оценке исходов

Эта ошибка может возникнуть, если шансы обнаружения исхода у пациентов в одной группе выше, чем в другой. Разумеется, некоторые исходы, такие как смерть, сердечно-сосудистые катастрофы, тяжелые формы злокачественных новообразований настолько очевидны, что вряд ли будут пропущены. Однако для менее четко определяемых исходов (специфические причины смерти, субклинические проявления заболевания, побочные эффекты или инвалидизация) систематическая ошибка может возникнуть просто из-за различий в методах их выявления и классификации.

Эту ошибку можно минимизировать с помощью трех приемов: организовать исследование так, чтобы наблюдатель не знал, к какой группе принадлежит каждый пациент; однозначно определить критерии диагностики изучаемых исходов (и следовать им); прилагать одинаковые усилия к выявлению исходов во всех группах.

Пример. Изучались исходы у пациентов с клинически бессимптомным шумом на сонной артерии [9]. Всего в течение 4 лет наблюдалось 500 таких пациентов, которые были классифицированы по степени исходного стеноза сонной артерии при доплеровском исследовании. В качестве исходов рассматривались изменения степени стеноза и частота мозговых ишемических эпизодов.

Для того чтобы избежать систематической ошибки, авторы определяли степень стеноза сонной артерии, используя однозначные критерии оценки доплерограмм, причем эту оценку производили специалисты, не знавшие ни результатов аускультации, ни данных предыдущего доплеровского исследования. Клинические и доплерографические оценки повторялись каждые 6 мес; у всех неявившихся на обследование пациентов выясняли по телефону, не было ли у них мозговых ишемических эпизодов.

Исследование показало, что у пациентов со стенозом сонной артерии более 75% частота ишемических эпизодов за 3 года превышала 20%, что более чем в 4 раза чаще, чем у пациентов со стенозом менее 30% (см. рис. 6.6).

Как бороться с систематической ошибкой, возникающей при отборе

В идеале, для определения связи фактора с прогнозом лучше всего было бы сравнивать две абсолютно идентичные когорты, различающиеся только тем, что в одной из них изучаемый фактор имеется, а в другой – нет, т.е. все остальное, за исключением этого фактора, у этих когорт одинаково. Но в реальных когортных исследованиях "все остальное" обычно одинаковым не бывает.

Как разрешить эту проблему? Существует несколько приемов, которые позволяют нивелировать различия между группами сравнения либо на стадии планирования исследования, либо при анализе его результатов (табл. 6.2). Читатель должен скептически относиться к любому исследованию, в котором не применялся один или несколько таких приемов. Основной вопрос, на который должно ответить исследование, заключается в том, связаны ли различия в прогнозе для разных групп только с исследуемым фактором или они зависят от каких-либо других факторов.

Таблица 6.2

Методы устранения систематических ошибок, возникающих при отборе

Метод	Описание	Этап исследования	
		планирование	анализ
Рандомизация	Распределить пациентов по группам таким образом, чтобы каждый пациент имел равные шансы попасть в ту или иную группу	+	
Введение ограничений	Ограничить диапазон характеристик пациентов, включаемых в исследование	+	
Подбор соответствующих пар	Каждому пациенту в одной группе подобрать одного или более пациентов с такими же характеристиками (кроме изучаемой) для группы сравнения	+	
Стратификация	Сравнить влияние изучаемого фактора на частоту исходов внутри подгрупп, имеющих одинаковый исходный риск		+
Стандартизация			
простая	Математически скорректировать исходное значение какой-либо одной характеристики таким образом, чтобы уравнивать подгруппы по исходному риску		+
множественная	Скорректировать различия по многим влияющим на исход факторам, применяя методы математического моделирования		+
Анализ			
"наилучший вариант–наихудший вариант"	Описать результаты, которые могут получиться, если исходить из предположения о максимальной выраженности систематической ошибки при отборе		+

Рандомизация

Рандомизация – метод случайного отнесения больного к той или иной исследуемой группе. Единственный способ уравнивать все посторонние факторы ("все остальное") состоит в том, чтобы распределить пациентов по группам случайным образом так, чтобы каждый пациент имел равный шанс попасть в группу воздействия и в группу без него. Рандомизация уравнивает вероятность воздействия на пациента не только тех факторов, которые, как мы предполагаем, могут влиять на прогноз, но и тех, о существовании которых мы даже не знаем. Таким образом, рандомизация защищает нас от неверных заключений относительно прогностических факторов. Однако рандомизация обычно неприменима в исследованиях прогноза. Конкретные ситуации, в которых можно применять рандомизацию (обычно – при исследовании того, как лечение влияет на прогноз), будут обсуждаться в главе 7.

Введение ограничений

Для того чтобы уравнивать вероятность воздействия важных посторонних факторов, можно ограничить контингент пациентов, включенных в исследование, только теми, кто обладает узким диапазоном характеристик. Например, влияние возраста на прогноз после острого инфаркта миокарда может быть исследовано на мужчинах с неосложненным инфарктом передней локализации. Однако необходимо иметь в виду, что хотя ограниченное включение способствует однородности групп, это достигается ценой отказа от обобщаемости результатов. В процессе исключения потенциальных участников могут быть сформированы нетипичные когорты, и результаты таких исследований нельзя будет переносить на большинство пациентов с данным состоянием.

Подбор соответствующих пар

Испытуемые могут подбираться таким образом, чтобы для каждого пациента из одной группы существовал один или несколько пациентов в группе сравнения с такими же характеристиками, кроме изучаемого фактора. Часто пациентов подбирают по возрасту и полу, поскольку эти факторы сильно связаны с прогнозом для многих заболеваний. Однако такой подбор можно рекомендовать и для многих других факторов, таких как стадия или степень тяжести заболевания, скорость прогрессирования и предшествующее лечение. Пример применения метода подбора пар в когорте пациентов с серповидно-клеточной анемией был рассмотрен при обсуждении наблюдательных исследований в главе 5.

Хотя подбор пар используется часто и может быть весьма полезен, он позволяет нивелировать систематическую ошибку, обусловленную только теми факторами, которые учитываются при подборе пар. Кроме того, как правило, невозможно подобрать пары, схожие более чем по нескольким признакам, поскольку на практике трудно найти пациентов, удовлетворяющих всем критериям отбора. К тому же, если характеристики, использованные при подборе пар, относительно грубы, между отобранными группами остаются различия. Например, если исследование риска рождения ребенка с синдромом Дауна провести путем подбора пар по возрасту матерей в 10-летнем диапазоне, то частота рождения больного ребенка в зависимости от возраста матери будет различаться почти в 10 раз между группами, большинство в одной из которых составляют 30-летние, а в другой – 39-летние женщины. И еще – поскольку вводимое ограничение или подбор пар производится по некоторому признаку, его влияние на исходы оценить не удастся.

Стратификация (stratification)

После того как данные собраны, производится их анализ и результаты представляются по подгруппам пациентов со сходными характеристиками (*стратификация*).

Пример. Предположим, что нужно сравнить послеоперационную летальность при операции аортокоронарного шунтирования в больницах А и Б. Всего в больнице А отмечено 48 смертей на 1200 операций (4%), а в больнице Б – 64 смерти на 2400 операций (2,6%). Согласно этим огрубленным показателям, больница Б лучше, чем А. Так ли это в действительности? Не исключено, что пациенты двух больниц исходно несопоставимы по прогнозу. На основании возраста, функции миокарда, степени стеноза и других характеристик можно подразделить пациентов на подгруппы с разным дооперационным риском (табл. 6.3), а затем сравнить послеоперационную летальность в пределах каждой подгруппы. После того как пациенты были разделены по дооперационному риску, послеоперационная летальность в каждой подгруппе риска оказалась абсолютно одинаковой для обеих больниц: 6,0, 4,0 и 0,67% для пациентов с высоким, средним и низким риском соответственно. Очевидный источник ошибочного первоначального вывода кроется в том, что для оценки используется только грубый показатель летальности, тогда как на самом деле характеристики пациентов в двух больницах существенным образом различались: высокий риск имели 42% пациентов в больнице А против 17% пациентов в больнице Б.

Стратификация – один из наиболее часто применяемых и действенных приемов борьбы с систематической ошибкой.

Таблица 6.3

Пример стратификации: гипотетические показатели летальности после аортокоронарного шунтирования в двух больницах после распределения больных на подгруппы в зависимости от дооперационного риска

Дооперационный риск	Больница А			Больница Б		
	число больных	число летальных исходов	летальность, %	число больных	число летальных исходов	летальность, %
Высокий	500	30	6	400	24	6
Средний	400	16	4	800	32	4
Низкий	300	2	0,67	1200	8	0,67
Всего	1200	48	4	2400	64	2,6

Стандартизация

От систематической ошибки можно избавиться при сравнении двух показателей, если скорректировать их таким образом, чтобы веса́ факторов, влияющих на исход, были равны. Этот процесс называется *стандартизацией*, или *коррекцией* (standartization, or adjustment). Суть его заключается в том, что мы рассчитываем частоту исходов для каждой подгруппы, а затем формируем две гипотетические сравниваемые группы, в которых каждая соответствующая подгруппа со своим специфическим показателем частоты исхода представлена равным числом пациентов. В предыдущем примере показатель летальности для пациентов с высоким риском, равный 6%, получил удельный вес 500/1200 в больнице А и значительно меньший удельный вес, 400/2400, в больнице Б и т.д.; в итоге грубый показатель летальности для больницы А = $(500/1200 \times 0,06) + (400/1200 \times 0,04) + (300/1200 \times 0,0067) = 0,04$, а для больницы Б = $(400/2400 \times 0,06) + (800/2400 \times 0,04) + (1200/2400 \times 0,0067) = 0,026$.

Если ввести для подгрупп равные веса́, например, $1/3$ (цифры могут основываться на данных для той или иной больницы или любой другой референсной популяции), тогда стандартизованный показатель летальности в больнице А составит $(1/3 \times 0,06) + (1/3 \times 0,04) + (1/3 \times 0,0067) = 0,036$ и то же самое – в больнице Б. Благодаря заданному равному весу всех подгрупп в обеих группах полностью устранен кажущийся избыточный послеоперационный риск в больнице А.

Различия между грубым показателем послеоперационной летальности в двух больницах возникают в результате систематической ошибки, обусловленной различиями дооперационного риска у пациентов. Но нас интересуют только различия, связанные с а-

мой больницей и квалификацией хирургов, а не с пациентами как таковыми. Различия в грубых показателях летальности вызваны различиями пациентов, в то время как стандартизованные показатели летальности уравнивают веса дооперационного риска у пациентов обеих больниц. Стандартизация чаще применяется в исследованиях риска (в которых показатели частоты обычно стандартизируются по возрасту, полу и/или расе), чем в прогностических исследованиях. В отличие от метода стратификации (часто используемой в прогностических исследованиях), стандартизация нивелирует эффект посторонних факторов. При стратификации эффект фактора может быть выявлен, даже если именно он используется в качестве признака, по которому производится разделение. Таким образом, благодаря стандартизации мы показали, что в больницах А и Б пациенты имеют одинаковый прогноз. При помощи стратификации был получен тот же результат.

Множественная стандартизация

В большинстве клинических ситуаций одновременно действуют многие факторы. Между ними существуют сложные взаимные влияния. Эти переменные могут быть связаны как друг с другом, так и с изучаемым исходом, один фактор может изменять эффект другого, а общий эффект двух или нескольких факторов иногда больше, чем простая сумма их индивидуальных эффектов.

Многофакторный анализ (multivariable analysis) – это метод одновременного рассмотрения воздействий многих переменных (см. главу 9). Он используется для того чтобы одновременно корректировать эффекты многих переменных для выявления независимого действия одного фактора. Кроме того, метод позволяет выделить из большого числа переменных малое их подмножество, которое вносит независимый и существенный вклад в исход, и упорядочить переменные по силе их влияния на исход. *Анализ пропорционального риска Кокса* (Cox's proportional hazard analysis) – это вид многофакторного анализа, используемый в случае, когда исходом служит время наступления события (как в анализе выживаемости).

Многофакторный анализ – единственный реальный способ обработки многих данных (переменных), когда на этапе анализа исследования приходится иметь дело сразу со многими переменными. (Рандомизация также контролирует множество переменных, но на этапах планирования и проведения исследования.) Более простые методы, такие как стратификация или подбор соответствующих пар, позволяют одновременно рассматривать только несколько переменных, причем в ущерб статистической мощности.

Анализ чувствительности

В отсутствие данных о важных прогностических факторах их потенциальный эффект можно оценить, предполагая различную степень диспропорции в распределении этих факторов между сравниваемыми группами и оценивая, как это повлияет на результат. Такой процесс носит общее название "*анализ чувствительности*" (sensitivity analysis). Анализ "наилучший вариант – наихудший вариант", о котором уже шла речь в этой главе, представляет собой специальный вид анализа чувствительности, когда результаты сравниваются, исходя из предположения наименьшей и наибольшей диспропорции в распределении переменной, используемой для прогнозирования¹.

Предположение худшего – это более строгий тест для проверки влияния фактора на результаты исследования. Менее жесткий подход состоит в предположении того, что фактор распределен между группами маловероятным образом.

Пример. Исследование методов лечения при легких формах сахарного диабета показало, что у пациентов, которым назначали препарат сульфонилмочевины толбутамид, был более высокий риск смерти от сердечно-сосудистых заболеваний, чем у тех, которым назначали инсулин или только диету. Результат подвергся критике, поскольку авторы не собирали и не учитывали при анализе данные о курении – известном факторе, связанном с сердечно-сосудистой смертностью. Было высказано предположение, что если курящие распределились по группам неодинаково, а именно так, что среди получавших толбутамид их оказалось больше, то различия следовало бы связать с курением, а не с назначением толбутамида. Однако Cornfield [10] отметил, что даже если в группе леченных толбутамидом число курящих было бы на 20% больше, чем в контрольной группе (маловероятная ситуация: 1 шанс из 50 000), то повышенный риск смерти в группе больных, леченных толбутамидом, все равно бы сохранился. Следовательно, наблюдаемые различия между группами нельзя объяснить за счет систематической ошибки, обусловленной диспропорцией в распределении курящих.

Общая стратегия

За исключением рандомизации, все методы устранения влияния различий между группами имеют следующий недостаток: они эффективны только в отношении тех факторов, которые выделены для рассмотрения, но не затрагивают неизвестные на момент исследования или известные, но не принятые во внимание прогностические факторы.

¹ Анализ чувствительности может также использоваться для оценки потенциального влияния неточности данных при анализе принятия решений (см. главу 4).

Обычно исследователи не ограничиваются каким-либо одним методом устранения систематической ошибки, а используют сразу несколько. Так, проводя исследование с целью выяснения, снижается ли при желудочковых экстрасистолах выживаемость после перенесенного острого инфаркта миокарда, можно поступить следующим образом: а) не включать в исследование пациентов слишком молодого или старческого возраста, а также необычные случаи инфаркта (например, вследствие септической аневризмы); б) произвести подбор пар по возрасту – фактору, сильно связанному с прогнозом, но постороннему по отношению к изучаемому фактору; в) анализировать результаты раздельно по подгруппам с разной степенью тяжести заболевания (например, наличие или отсутствие сердечной недостаточности либо других заболеваний, таких как хронические обструктивные заболевания легких); г) с помощью многофакторного анализа скорректировать грубый результат по эффектам всех вместе взятых переменных, которые могут быть связаны с прогнозом, кроме аритмии.

Обобщаемость результатов и систематическая ошибка, возникающая при отборе

Прогностические исследования, проводимые на базе специализированных центров, могут давать ошибочную картину применительно к больным, которые не подвергались специальному отбору. Это утверждение справедливо даже для аккуратно выполненных исследований, в которых тщательно контролируются систематические ошибки и полученный прогноз справедлив для конкретной выборки больных. Однако из-за особенностей использованной выборки может оказаться, что на большинство других больных с данным состоянием или на вашего конкретного больного распространить результаты нельзя.

Иногда в рандомизированных контролируемых испытаниях пациенты контрольной группы изучаются для более глубокого понимания обычного клинического течения заболевания. Но такие пациенты не обязательно отражают большинство популяции, поскольку у включенных в исследование добровольцев обычно течение заболевания более благоприятно, чем у больных, которые не согласились участвовать в испытании. Например, в большом канадском исследовании по выявлению рака молочной железы у 40-летних женщин показано, что 90% женщин контрольной группы с инвазивным раком были живы спустя 7 лет и смертность от рака молочной железы была ниже, чем среди всей популяции канадских женщин [11].

Резюме

Прогноз – это описание течения заболевания с момента его начала. По сравнению с факторами риска, прогностические факторы имеют относительно высокую распространенность и могут оцениваться даже на основании индивидуального клинического опыта. Вместе с тем случаи, наблюдаемые в медицинских центрах и описываемые в медицинской литературе, часто представляют собой смещенную выборку из всех случаев соответствующего заболевания, в результате наблюдается склонность к завышению степени их тяжести.

Прогноз лучше всего описывается вероятностью возникновения исхода в любой момент развития заболевания. В принципе, это может быть сделано при наблюдении когорты больных до тех пор, пока изучаемый исход не наступит у всех, у кого он может произойти. Однако поскольку такой подход неэффективен, используется другой метод, называемый анализом дожития, или анализом времени наступления событий. Наступление событий во времени оценивается путем накопления исходов для всех наблюдаемых пациентов в течение всех предшествующих интервалов времени.

Как и любое наблюдение за когортами, исследования по сравнению прогноза между разными группами пациентов могут быть подвержены систематической ошибке из-за способа формирования когорты, непостоянного состава групп в процессе наблюдения или неодинакового подхода к регистрации исходов. Методы коррекции возможного систематических ошибок обеспечивают правильное (несмещенное) сравнение. К таким методам относятся введение ограничений, подбор пар, стратификация, стандартизация, многофакторный анализ и анализ чувствительности. Один или несколько таких методов должны обязательно использоваться в любом сравнительном исследовании.

Литература

1. Laupacis A. Changes in quality of life and functional capacity in hemodialysis patients treated with recombinant human erythropoietin. The Canadian Erythropoietin Study Group. *Semin Nephrol* 1990;2(Suppl 1):11–19.
2. Gelber R.D. et al. Quality-of-life evaluation in a clinical trial of zidovudine therapy in patients with mildly symptomatic HIV infection. *Ann Intern Med* 1992;116:961–966.
3. Justice A.C., Feinstein A.R., Wells C.K. A new prognostic staging system for the acquired immunodeficiency syndrome. *N Eng J Med* 1989;320:1388–1393.
4. Wolmark N. et al. The prognostic significance of preoperative carcinoembryonic antigen levels in colorectal cancer. Results from the NSABP clinical trials. *Ann Surg* 1984; 199:375–382.
5. Bridges A.J., Conley C., Wang G., Burns D.E., Vasey F.B. A clinical and immunologic evaluation of women with silicone breast implants and symptoms of rheumatic disease. *Ann Intern Med* 1993; 118:929–936.

6. Gabriel S.E., O'Fallon W.M., Kurland L.T., Beard C.M., Woods J.E., Melton L.J. III. Risk of connective-tissue diseases and other disorders after breast implantation. *N Engl J Med* 1994;330:1697–1702.
7. Thompson K.S., Fletcher S.W., O'Malley M.S., Buckwalter J.A. Long-term outcomes of morbidly obese patients treated with gastrogastrostomy. *J Gen Intern Med* 1986;1:85–99.
8. Paffenbarger R.S., Hyde R.T., Wing A.L., Lee I.M., Jung D.L., Kampert J.B. The association of changes in physical-activity level and other lifestyle characteristics with mortality among men. *N Engl J Med* 1993;328:538–545.
9. Chambers B.R., Norris J.W. Outcome in patients with asymptomatic neck bruits. *N Engl J Med* 1986; 315:860–865.
10. Cornfield J. The University Group Diabetes Program. A further statistical analysis of the mortality findings. *JAMA* 1971;217:1676–1687.
11. Miller A.B., Baines C.J., Teress Te., Wall C. Canadian National Breast Screening Study. 1. Breast cancer detection and death rates among women aged 40 to 49 years. *Can Med Assoc J* 1992;147:1459–1476.

Рекомендуемая литература

- Colton T. *Statistics in medicine*. Boston: Little, Brown, 1975, pp. 237–250.
- Concato J., Feinstein A.R., Holford T.R. The risk of determining risk with multivariable models. *Ann Intern Med* 1993; 118:201–210.
- Feinstein A.R., Sosin D.M., Wells C.K. The Will Rogers phenomenon—stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* 1985;312:1604–1608.
- Guyatt G.H., Feeny D.H., Patrick D.L. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622–629.
- Horwitz R.I. The experimental paradigm and observational studies of cause-effect relationships in clinical medicine. *J Chron Dis* 1987;40:91–99.
- Laupacis A., Wells G., Richardson S., Tugwell P. Users' guides to the medical literature. V. How to use an article about prognosis. *JAMA* 1994;272:234–237.
- Peto R. et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. Analysis and examples. *Br J Cancer* 1977;35:1–39.
- Wasson J.H., Sox H.C., Neff R.K., Goldman L. Clinical prediction rules: applications and methodological standards. *N Engl J Med* 1985;313:793–799.
- Weiss N.S. *Clinical epidemiology: the study of the outcomes of illness*. New York: Oxford University Press, 1986.

7 ЛЕЧЕНИЕ

Как только больному поставлен диагноз и предсказано ожидаемое течение заболевания, сразу возникают вопросы: "Что можно предпринять? Существует ли лечение, которое улучшит исход заболевания?" В этой главе будет рассмотрен вопрос о том, как определить эффективность назначенного лечения.

Предположения и доказательства

При разработке новых методов лечения требуются, с одной стороны, перспективные идеи, а с другой стороны – способы их надежной проверки.

Предположения

Гипотезы о возможных способах лечения возникают в результате любой медицинской деятельности. Некоторые из них основываются на знаниях механизма заболевания. Препараты против устойчивых к антибиотикам бактерий разработаны на основе представлений о механизме резистентности. Гормональные препараты получены благодаря знаниям о структуре естественных гормонов. Эффективный метод лечения сердечной недостаточности путем снижения постнагрузки предложен после того, как патофизиологические исследования выявили важность этого параметра в развитии данной патологии.

Другие гипотезы о способах лечения возникли благодаря наблюдениям проницательных врачей. Приведем два примера таких наблюдений: у пациентов с болезнью Паркинсона, получавших амантадин для профилактики гриппа, отмечено улучшение неврологического статуса; при назначении колхицина по поводу подагры наблюдалось уменьшение частоты приступов семейной формы желтушного лептоспироза. Эффективность этих методов лечения не была предсказана на основе знаний о природе болезней, а механизмы действия этих препаратов до сих пор изучены не полностью. Аналогичным образом во всем мире применяются народные средства, польза которых подкреплена вековым опытом, а не научными исследованиями.

Гипотезы могут возникать в результате проб и ошибок. Неко-

торые противоопухолевые препараты были обнаружены путем проверки активности огромного числа веществ.

Идеи относительно методов лечения и профилактики возникают также в результате эпидемиологических исследований. Было замечено, что в странах Африки, где пища богата клетчаткой, распространенность заболеваний кишечника ниже по сравнению с развитыми странами, население которых потребляет мало клетчатки. Это наблюдение привело к попыткам профилактики заболеваний кишечника (синдрома раздраженной кишки, дивертикулита, аппендицита, рака толстой кишки) с помощью соответствующей диеты. Сравнительные исследования распространенности заболеваний в разных странах выявили роль красного вина в профилактике сердечно-сосудистых заболеваний и фтора – в предотвращении кариеса.

Проверка предположений

Некоторые методы лечения действуют настолько быстро и эффективно, что их ценность очевидна даже без формальной проверки. Врачи не имеют ни малейших сомнений относительно эффективности пенициллина при пневмонии, хирургической операции при аппендиците или колхицина при подагре. Здесь вполне достаточно клинического опыта.

Однако обычно эффект лечения не столь впечатляющий. В таких случаях предположения об эффективности лечения необходимо подвергнуть проверке посредством клинических исследований, поскольку выявлению истинного эффекта лечения может мешать множество факторов – случайные совпадения, неправомерные сравнения, спонтанные изменения течения болезни, принятое желаемого за действительное.

Иногда знание механизмов заболевания, основанное на изучении лабораторных моделей или физиологических исследованиях у людей, достигает столь высокого уровня, что возникает искушение предсказать эффект лечения, не проводя формальной проверки. Однако, полагаясь только на современные представления о механизмах развития заболевания, без проверки изучаемого метода на людях, можно получить неприятные “сюрпризы”, поскольку понимание патогенеза далеко не полно.

Пример. Клинические проявления инсульта часто бывают обусловлены инфарктом мозга в бассейне стенозированного сегмента внутренней сонной артерии. Теоретически предотвратить инсульт у этих больных можно, обеспечив обходной кровоток пораженного сегмента для нормального кровоснабжения области возможного инфаркта. Создание анастомоза между поверхностной височной арте-

рией и внутренней сонной артерией в дистальной по отношению к участку стеноза области технически осуществимо. Поскольку возможность такого вмешательства казалась очевидной (на основе физиологических представлений, а также в связи с уже доказанным успехом аналогичной процедуры – аортокоронарного шунтирования), операцию стали широко использовать.

Группа по исследованию экстракраниально-интракраниального шунтирования [1] провела рандомизированное контролируемое испытание эффективности этой операции. Пациенты с ишемией мозга и стенозом внутренней сонной артерии были распределены случайным образом на две группы – хирургического вмешательства и консервативного лечения. Операция была технически успешной – по ее окончании функционировали 96% анастомозов. Однако она оказалась неэффективной: летальность и заболеваемость инсультом через 5 лет были практически одинаковыми среди оперированных и среди больных, леченных консервативно, причем у оперированных продолжительность жизни была меньше.

Это исследование показывает, что лечение, основанное на знаниях патогенеза заболевания и кажущееся целесообразным с научной точки зрения, может быть неэффективным, что выясняется в итоге тщательной проверки на пациентах. Конечно, далеко не все идеи оказываются столь бесплодными – эффективность каротидной эндартерэктомии, предложенной вместо описанного вмешательства, полностью подтвердилась [2].

Следовательно, гипотезы о методах лечения почти всегда необходимо проверять клиническими исследованиями, в которых сопоставляются данные о течении заболевания у леченых и нелеченых пациентов. Как заметил Л. Опи, "лечение назначается не потому, что оно должно подействовать, а потому, что оно действует" [3].

Исследования эффективности лечения

Под лечением обычно понимаются меры, назначаемые врачом пациенту с определенным заболеванием: проведение хирургической операции, прием лекарств, соблюдение диеты или занятия физическими упражнениями. Однако существует множество других способов улучшения здоровья. Сюда относятся меры по предотвращению заболевания у конкретного пациента (консультирование и раннее выявление патологии с последующим лечением обсуждаются в главе 8), профилактика заболеваний среди населения, изменение организации и финансирования системы здравоохранения. Независимо от того, какие действия предпринимаются для лечения, методы проверки их эффективности не меняются.

Оценка эффективности лечения может производиться методом наблюдения (обсервационные исследования) или путем проведения эксперимента (клинического испытания).



Рис. 7.1. Структура клинического испытания.

Обсервационные исследования (observational studies) эффективности лечения – это особый вид исследований прогноза, где в качестве изучаемого прогностического фактора выступает терапевтическое вмешательство. Все, что говорилось выше о когортных исследованиях (см. главы 5 и 6), применимо и к обсервационным исследованиям эффективности лечения. Главное преимущество этих исследований – относительная простота проведения. Основной недостаток обусловлен тем, что в исследуемых группах, помимо различий в самом лечении, могут существовать и другие систематические различия, которые приводят к ошибочным выводам.

Клинические испытания (clinical trials) – специальный вид когортных исследований, условия проведения которых (отбор групп вмешательства, характер вмешательства, организация наблюдения и оценка исходов) обеспечивают устранение систематических ошибок. По степени контролируемости и управляемости клинические испытания превосходят обычные когортные исследования. В сущности, это эксперимент, аналогичный тем, которые проводятся в лаборатории. Проводя клиническое испытание, исследователь изучает влияние только одного отдельно взятого фактора: другие факторы, влияющие на исход, должны быть, насколько это возможно, одинаковыми в сравниваемых группах.

Рандомизированные контролируемые испытания служат стандартом качества научных исследований эффективности лечения. Сначала будет рассмотрен подробно именно этот способ получения информации, а затем – альтернативные методы поиска ответа на тот же вопрос.

Рандомизированные контролируемые испытания

Схематично ход клинического испытания представлен на рис. 7.1. Для исследования сначала отбираются пациенты из боль-

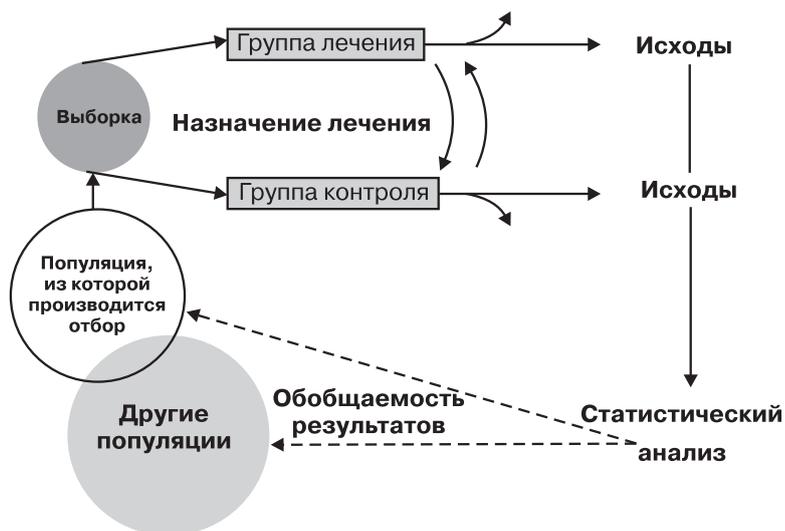


Рис. 7.2. Систематические ошибки в клинических испытаниях.

шого числа людей с изучаемым состоянием. Затем этих пациентов разделяют случайным образом на две группы, сопоставимые по основным прогностическим признакам. Одна группа, называемая экспериментальной, или группой лечения, подвергается вмешательству, которое, как ожидается, будет полезным. Другая группа, называемая контрольной, или группой сравнения, находится точно в таких же условиях, как первая, за исключением того, что составляющие ее пациенты не подвергаются изучаемому вмешательству. Затем исследователи наблюдают клиническое течение заболевания в обеих группах и любые отличия в исходах приписывают изучаемому вмешательству.

Основная причина подобного построения клинических испытаний состоит в стремлении избежать систематических ошибок при сравнении эффективности двух или большего числа вмешательств. Достоверность клинических испытаний зависит от того, насколько в сравниваемых группах удалось обеспечить одинаковое распределение всех факторов, определяющих прогноз, кроме изучаемого вмешательства.

Далее в этой главе будет описана методика проведения и интерпретации клинических испытаний на примере схемы, изображенной на рис.7.2.

Формирование выборки

Характеристики пациентов, включаемых в испытание, определяются в какой степени полученные выводы могут быть перенесены на других больных. Среди многих причин, по которым пациенты с изучаемым заболеванием не включаются в исследование, основными являются следующие три: больные не отвечают установленным критериям включения, отказываются участвовать в нем или не соблюдают условий его проведения.

Во-первых, критерии включения формулируются для ограничения разнородности пациентов. Основные критерии, ограничивающие включение, таковы: нетипичный характер заболевания, наличие других болезней, необычно плохой прогноз (который может стать причиной выбывания пациента во время исследования) или высокая вероятность несоблюдения предписанного лечения. Естественно, в исследование не включаются также больные, у которых имеются противопоказания к одному из видов исследуемого лечения. Такое ограничение разнородности повышает достоверность исследования – сокращается возможность исходов, не связанных с самим лечением. Кроме того, это позволяет точно оценить обобщаемость, поскольку известно, по отношению к каким больным применимы результаты. Однако исключение части пациентов приводит к снижению обобщаемости результатов, так как в клинической практике могут достаточно часто встречаться больные, подобные тем, которые не были включены в исследование.

Во-вторых, пациенты могут отказаться от участия в исследовании. Некоторые не хотят лечиться предлагаемым методом либо возражают против того, что выбор лечения определяется случайным образом, а не их лечащим врачом. Пациенты, отказавшиеся участвовать в исследовании, обычно отличаются от тех, кто дает свое согласие, по социально-экономическому статусу, тяжести заболевания и другим признакам, связанным или не связанным со здоровьем.

В-третьих, исключаются пациенты, которые на ранних стадиях испытания показали неспособность строго следовать предписанной методике лечения. Это позволяет избежать напрасной траты усилий и снижения достоверности исследования вследствие выбывания пациентов или их перехода из группы в группу.

В результате указанных причин пациенты в клинических испытаниях обычно представляют собой высокоселективную выборку, смещенную относительно общей совокупности больных с изучаемым состоянием (рис. 7.3). Поэтому переносить полученные в клинических испытаниях результаты в обычную клиническую практику нужно с осторожностью.

Вмешательство

Исследуемое вмешательство характеризуется тремя свойствами: практической применимостью, сложностью и эффективностью.

Во-первых, применимо ли предлагаемое вмешательство в обычной клинической практике? Пытаясь стандартизовать лечение таким образом, чтобы оно могло быть легко описано и воспроизведено в других лечебных учреждениях, некоторые исследователи прекращают дальнейшую работу, убедившись, что их результаты не найдут применения в повседневной практике.

Во-вторых, в научных исследованиях применяют однокомпонентные высокоспецифичные виды вмешательств, которые могут быть точно описаны и воспроизведены. Что касается практикующих врачей, то им постоянно приходится делать выбор из альтернативных видов лечения, включающих многие элементы. Многокомпонентные вмешательства поддаются точной оценке лишь в той мере, в какой они воспроизводимы в других условиях.

Пример. Известно, что по различным причинам пожилые люди часто падают. Rubenstein et al. [4] изучали эффективность программы профилактики падений у пожилых. Обслуживаемые местной поликлиникой пожилые люди, у которых отмечались случаи падения, были случайным образом распределены на две группы – в одной группе применялась специальная программа профилактики (экспериментальная группа), а в другой – обычный уход (контрольная группа). Программа включала в себя подробное обследование, лабораторные тесты и оценку окружающей обстановки пациента; кроме того, пациенты получали рекомендации от врача общей практики. В течение последующих 2 лет в экспериментальной группе по сравнению с контрольной уменьшилось число падений, было на 26% меньше госпитализаций и на 52% меньше число дней, проведенных в больнице.

В-третьих, достаточно ли изучаемое вмешательство отличается от альтернативных способов лечения, чтобы можно было ожидать существенного влияния на исход? Некоторые заболевания могут быть излечены путем устранения единственной причины (например, у больных с гипертиреозом). Однако большинство заболеваний обусловлено сочетанием многих факторов, так что трудно ожидать значительного терапевтического эффекта от вмешательства, изменяющего, причем незначительно, лишь один из факторов. Поэтому вряд ли стоит удивляться, если в исследовании вмешательства может быть сделан вывод о его неэффективности.

Процент больных



Рис. 7.3. Отбор пациентов для клинических испытаний. Исследование эффективности программы по снижению риска поражения нижних конечностей у пациентов с сахарным диабетом [Litzelman D.K., Slemenda C.W. et al. Ann Intern Med 1993;119:36–41].

Группы сравнения

О ценности изучаемого метода лечения можно судить только путем сравнения его результатов с эффектом других лечебных мероприятий. Для оценки результата исследуемого лечения следует провести сравнение между группами, получающими различные виды лечения. Вопрос заключается не в том, использовать ли метод сравнения вообще, а в том, как обеспечить правомерность сравнений.

Отсутствие вмешательства

Можно сравнивать эффект исследуемого вмешательства с отсутствием вмешательства. При этом одна группа больных получает лечение, а другая не получает его и не находится под медицинским наблюдением. Такое сравнение позволит оценить общий эффект медицинской помощи как связанный с исследуемым вмешательством, так и не связанный с ним.

Наблюдение

Можно сравнивать эффект исследуемого лечения с обычным наблюдением без какого-либо вмешательства. Дело в том, что все пациенты, участвующие в исследовании, неизбежно ощущают к себе особое внимание. Осознавая, что они являются предметом специального интереса, люди невольно меняют свое поведение, независимо от характера получаемого ими лечения. Этот феномен называется *эффектом Готорна* (Hawthorne effect). Причины его не ясны. Возможно, пациенты стремятся порадовать своих докторов, или им хочется внести вклад в получение "хороших" результатов. Сравнение результата лечения с итогами простого наблюдения позволяет отделить вклад вмешательства от эффекта Готорна.

Лечение плацебо

Можно сравнивать исследуемое лечение с назначением плацебо. Плацебо – это лекарственная форма, не отличимая от исследуемого препарата по внешнему виду, цвету, вкусу и запаху, но не оказывающая специфического действия (например, таблетки глюкозы или инъекции физиологического раствора). Показано, что плацебо, которое больной уверенно принимает за лекарство, уменьшает послеоперационную боль, тошноту и зуд приблизительно у $\frac{1}{3}$ пациентов. Этот феномен называется *эффектом плацебо* (placebo effect).

Пример. Пациенты с тяжелым хроническим зудом участвовали в испытании лекарственных препаратов, купирующих зуд [5]. В течение каждой из трех недель в случайной последовательности 48 пациентов получали либо ципрогептадина хлорид, либо тримепразина тартрат, либо плацебо. Кроме того, в график приема препаратов и плацебо был введен (также в случайном порядке) однонедельный перерыв. Результаты оценивались в отсутствие сведений о принимаемом препарате и представлялись в баллах выраженности зуда: чем выше балл, тем сильнее зуд. Полученные средние значения были следующими: ципрогептадина хлорид – 28, тримепразина тартрат – 35, плацебо – 30, отсутствие лечения – 50. Таким образом, оба активных препарата, равно как и плацебо, были одинаково эффективны и дали гораздо лучшие результаты, чем отсутствие лечения.

Эффект плацебо имеет разный смысл для исследователей и для клиницистов. Исследователи больше, чем врачи-практики заинтересованы в установлении специфических эффектов, имеющих отношение к современным теориям о причинах заболевания, и рассматривают эффект плацебо как базальный уровень для измерения специфических лечебных эффектов. Врачи должны воспринимать эффект плацебо в качестве лечебного и стараться максимально использовать его, так же как и любой другой способ помощи пациентам.

Многие клинические вмешательства имеют как специфические, так и неспецифические эффекты (рис.7.4). Для врачей и пациентов важен полный эффект вмешательства, определяемый как отличие в состоянии пациента в результате лечения по сравнению с тем, что с ним произойдет при естественном течении заболевания, т.е. без лечения. Однако важно знать, какая часть лечебного эффекта имеет специфический характер, а какая – неспецифический, чтобы избежать применения опасного, неприятного или дорогостоящего вмешательства, лишь малая доля эффекта которого обусловлена специфическим действием.

Обычное лечение

Можно сравнивать исследуемое лечение с обычным лечением. Это целесообразно и приемлемо с точки зрения этики только в том случае, если известно, что обычное лечение эффективно.

Таким образом, суммарный лечебный эффект исследуемого препарата складывается из четырех компонентов, что продемонстрировано на рис. 7.5.

Назначение лечения

Для исследования специфического лечебного действия вмешательства целесообразнее всего распределять пациентов по груп-

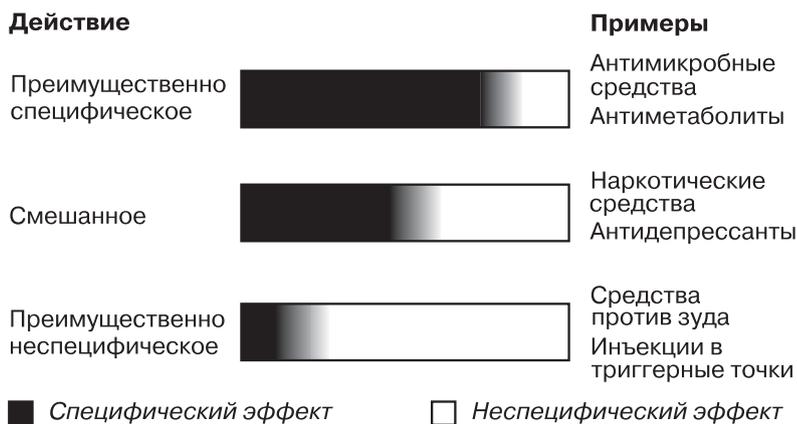


Рис. 7.4. Соотношение между специфическим и неспецифическим эффектами лекарственных препаратов [Fletcher R.H. Fam Med REV 1983; 1:40–48].

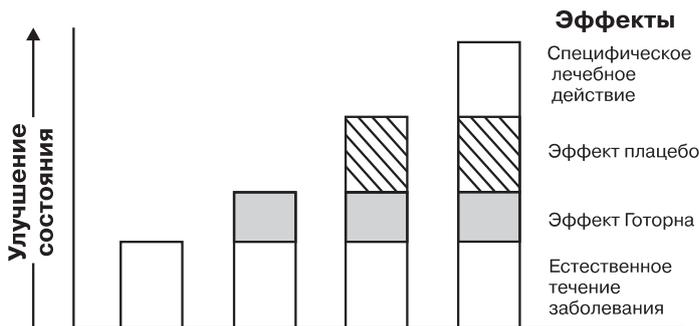


Рис. 7.5. Вклад различных причин в суммарный лечебный эффект.

пam случайным образом, т.е. путем *рандомизации* (randomization). При этом назначение исследуемого или контрольного лечения производится при помощи процедуры, аналогичной подбрасыванию монеты, обеспечивающей равные (или по крайней мере известные) шансы попасть в ту или другую группу.

Рандомизация – оптимальный метод выбора лечения, так как позволяет избежать систематической ошибки при разделении пациентов на группы. Это означает, что пациенты одной группы будут в среднем обладать теми же характеристиками, что и пациенты другой. Достичь такого распределения в изучаемых группах всех факторов, связанных с прогнозом, независимо от того, известны они до исследования или нет, можно только посредством рандомизации.

В конечном счете при участии в испытании большого числа пациентов рандомизация обычно "работает" так, как описано выше. Однако случайное распределение не гарантирует, что все группы действительно будут одинаковыми. Хотя процесс случайного выбора лечения свободен от систематической ошибки, результат все же может оказаться смещенным. Различие между группами может возникнуть просто случайно, особенно если число включенных в исследование пациентов невелико. Для того чтобы оценить, не случилось ли такое "невезение", авторы рандомизированных контролируемых испытаний часто представляют таблицу, в которой сравнивается частота различных характеристик (особенно тех, которые заведомо связаны с исходом) в экспериментальной и контрольной группах. Всегда необходимо убедиться в том, что важные характеристики пациентов распределились практически одинаково между группами сравнения. Если же имеются некоторые различия, то их можно оценить и попытаться нивелировать при анализе (см. главу 6).

По мнению некоторых исследователей, для снижения вероятности "невезения", следует до рандомизации убедиться в том, что по крайней мере некоторые характеристики, о которых известно, что они сильно влияют на исход, встречаются в экспериментальной и контрольной группах с одинаковой частотой. Для этого предлагается сначала распределить пациентов по подгруппам с одинаковым прогнозом и только затем рандомизировать их отдельно в каждой подгруппе – такой процесс называется *стратифицированной рандомизацией* (stratified randomization). При этом подходе группы обязательно должны получиться сравнимыми по характеристикам, использованным при формировании подгрупп. Другие исследователи возражают, утверждая, что различия, случайно возникающие в результате "невезения", как правило, незначительны, и с ними можно справиться математическими методами уже после того, как данные собраны.

Различия, возникающие после рандомизации

Не все пациенты участвуют в клиническом испытании в соответствии с его первоначальным планом. Например, может выяс-

ниться, что у некоторых больных нет изучаемого заболевания, как считалось при включении их в испытание; другие прекращают контакт с исследователем, не принимают назначенных препаратов, выбывают из-за побочных эффектов или других заболеваний, по тем или иным причинам получают лечение, предписанное другой группой, или не применяемое в данном исследовании. В итоге сопоставимость групп, достигнутая в результате рандомизации, нарушается к тому времени, когда необходимо анализировать исходы.

Отсутствие изучаемого заболевания

В клиническом испытании, да и на практике иногда приходится назначать лечение прежде, чем удастся полностью удостовериться в том, что у пациента именно то заболевание, по поводу которого показано данное вмешательство.

Пример. Для того чтобы выяснить, повышают ли моноклональные антитела к эндотоксину выживаемость при сепсисе, 543 пациента с сепсисом и подозреваемой грамотрицательной инфекцией были случайным образом разделены на две группы – больные одной группы получали антиэндотоксин, больные другой группы – плацебо [6]. В подгруппе пациентов, у которых действительно была грамотрицательная бактериемия, летальность снизилась с 49 до 30%, что существенно больше случайной разницы. Однако лишь у 200 пациентов (37%) имелась грамотрицательная бактериемия, подтвержденная результатами посева крови. Следовательно, причины, объясняющие эффективность препарата у остальных 63% больных, отсутствовали. Для всех пациентов с сепсисом (как имевших, так и не имевших бактериемию), летальность составляла 43% в группе получавших плацебо и 39% в группе получавших антиэндотоксин. Это различие незначительно и может быть обусловлено случайностью. Таким образом, исследование показало, что препарат эффективен в отношении грамотрицательных бактерий, а не сепсиса. Важны оба результата: первый – для ученых, заинтересованных в выяснении биологического действия антиэндотоксина на бактерии, второй – для клиницистов, которые должны знать клинический эффект применения препарата у больных сепсисом, учитывая, что решение приходится принимать прежде, чем станет известно, действительно ли у пациента имеется бактериемия.

Если предположение о наличии у пациента изучаемого заболевания не подтверждается, то за такую ошибку приходится расплачиваться. Обследование дополнительного числа пациентов, которые не получают пользы от специфического лечения, уменьшает эффективность испытания – чтобы выявить эффект, нужно включить в исследование больше пациентов. Иначе говоря, из-за

того, что пациенты, у которых может проявиться специфический эффект, смешаны с теми, у которых его быть не может, эффективность лечения оказывается меньше, чем в исследовании, в которое включены только пациенты с наличием заболевания. Это уменьшает шансы обнаружения эффекта при заданном числе пациентов в испытании (см. главу 9). Однако важное достоинство исследований подобного типа состоит в том, что они предоставляют информацию о последствиях принятия решения так, как это видится с точки зрения клинициста (см. далее в этой главе раздел "Испытания с анализом данных в зависимости от назначенного или полученного лечения").

Соблюдение предписаний врача

Говоря о соблюдении пациентом предписаний врача, мы имеем в виду то, насколько точно больной, участвующий в клиническом испытании, выполняет требования, предъявляемые к нему врачом-исследователем. Пациент может сознательно идти на нарушение предписаний, однако чаще такое несоблюдение требований связано с тем, что больной не понял, какие препараты, в какой форме и дозе следует принимать, или препарат закончился, или у него нет денег на приобретение лекарств. Нарушение предписаний врача может привести к тому, что лечение, показавшее свою эффективность в специальных условиях, окажется неэффективным на практике.

Соблюдение предписаний особенно важно при амбулаторном наблюдении. В клинике многие факторы ограничивают поведение больного и делают его более дисциплинированным. В стационаре у пациентов создается впечатление, что их состояние намного тяжелее, чем на самом деле, и они этим обеспокоены. Они оказались в непривычной обстановке, зависят от квалификации и внимания персонала во всем, даже в сохранении самой жизни. Кроме того, медицинский персонал больниц действует по схеме, гарантирующей получение пациентом назначенного лечения. Таким образом, клинический опыт и данные медицинской литературы, отражающие условия стационара, могут недооценивать фактор соблюдения больным предписаний врача вне больницы, где лечится большинство больных и где гораздо труднее добиться выполнения назначений.

В клиническом испытании сравнение эффекта лечения у пациентов, соблюдающих и не соблюдающих назначения, может привести к ошибочным выводам.

Пример. В крупном исследовании [7] влияния нескольких гиполипидемических препаратов на развитие стенокардии 1103 мужчи-

нам был назначен клофибрат и 2789 мужчинам – плацебо. Пятилетняя летальность составила 20,0% в группе принимавших клофибрат и 20,9% в группе принимавших плацебо, что указывает на неэффективность препарата.

В то же время было обнаружено, что не все пациенты принимали лекарство. Был ли клофибрат эффективен у тех, кто действительно принимал препарат? Ответ оказался утвердительным. В экспериментальной группе пятилетняя летальность среди пациентов, действительно принимавших клофибрат, составила 15,0% по сравнению с 26,4% у пациентов, не соблюдавших схему лечения ($p < 10^{-5}$). Однако летальность среди больных, которые строго выполняли назначения и принимали плацебо, также снизилась. Здесь пятилетняя летальность у действительно принимавших препарат (плацебо) составила 15,0% против 28,3% среди тех, кто не принимал его ($p < 10^{-15}$). Таким образом, связь между приемом препарата и прогнозом не была полностью обусловлена специфической активностью препарата. Авторы [7] предостерегают от оценки эффектов лечения в подгруппах, сформированных по степени соблюдения пациентами протокола исследования после рандомизации.

Сопутствующие вмешательства

После рандомизации пациенты могут подвергаться воздействию множества факторов, помимо изучаемого вмешательства. Если эти факторы действуют в разных группах по-разному и связаны с исходами, то они создают систематические различия между сравниваемыми группами.

Пример. Лечение больных СПИДом связано с эмоциями, в частности потому, что болезнь поражает молодых людей и неизбежно заканчивается смертью в течение нескольких лет после появления симптомов. Изучению эффективности методов лечения препятствует нарушение обычных процедур рандомизированных испытаний, поскольку пациенты пытаются увеличить свои шансы на выживание. В ходе рандомизированных испытаний пациенты иногда обмениваются препаратами (исследователи называют обмен схемами лечения среди пациентов "контаминацией") или получают через "лекарственные клубы" препараты, не используемые в испытании. О таких поступках они не сообщают исследователям, так что эти факторы невозможно учесть при подведении итогов испытания. Вследствие этого исследование оказывается смещенным в сторону отсутствия эффекта, поскольку различия между результатами лечения в опытной и контрольной группах уменьшаются.

Сравнение пациентов, реагирующих и не реагирующих на лечение

Иногда в клинических испытаниях, особенно касающихся лечения злокачественных новообразований, сравниваются исходы у

пациентов, у которых в процессе лечения улучшилось состояние (отреагировавших на лечение), с исходами у тех, у кого оно не улучшилось (не отреагировавших на лечение). Предполагается, что это позволяет судить об эффективности лечения.

Такой подход научно не обоснован и часто ошибочен, поскольку наличие или отсутствие эффекта может быть обусловлено многими факторами, влияющими на окончательный исход: стадией заболевания, скоростью прогрессирования, соблюдением схемы лечения, дозой, побочными явлениями и, наконец, наличием других заболеваний. Если на самом деле ни у кого из пациентов состояние не улучшилось в связи с лечением и дальнейшее течение заболевания определялось какими-то другими причинами, то отреагировавшими на лечение будут названы те, у кого почему-либо наступило улучшение, а не отреагировавшими – те, у кого течение заболевания оказалось неблагоприятным. Разумеется, у отреагировавших исходы будут лучше, независимо от того, получили они исследуемое лечение или нет.

Слепой метод

Если участники испытания знают, кто какой вид лечения получает, их поведение может измениться определенным образом (т.е. стать причиной систематической ошибки). Один из способов минимизации этого эффекта состоит в использовании *слепого метода* (blinding). Исследование организуется таким образом, чтобы его участники не знали, кто получает изучаемый препарат, а кто – препарат сравнения, и это не влияло бы на их действия, снижая тем самым достоверность исследования. Более точно этот метод следовало бы называть "маскированием", но термин "слепой метод" давно стал общеупотребительным¹.

Слепой метод может быть реализован в клиническом испытании на четырех уровнях. Во-первых, те, кто распределяет пациентов по группам вмешательства, не должны знать, какое лечение будет назначено каждому последующему больному, чтобы это не нарушало включения пациентов в исследование в порядке их поступления. Во-вторых, пациенты не должны знать, какое именно лечение они получают; тогда менее вероятны несоблюдение схемы лечения или субъективность при описании своего состояния. В-третьих, проводящие наблюдение врачи не должны знать, какое лечение назначено пациенту, тогда не возникнут невольные различия в ведении больных. Наконец, если исследователи, оценивающие исходы, не будут знать о различиях методов лечения между группами, их мнение будет беспристрастным.

¹ В исследованиях, посвященных лечению людей с нарушением зрения, этот метод по этическим соображениям по-прежнему называется маскированием. – *Примеч. ред.*

Иногда используются термины "простой слепой метод" (не информирован пациент) или "двойной слепой метод" (не информированы и пациент, и исследователь), но они передают смысл недостаточно ясно. Лучше просто подробно описать методику проведения исследования. Испытание, в котором не пытались назначать лечение вслепую, называется открытым.

Слепой метод обычно применяется в исследованиях с плацебо (в основном это исследования эффективности лекарственных средств). Однако при исследовании многих важных клинических проблем, таких как оценка эффективности хирургических вмешательств, радиотерапии, диеты или организации медицинской помощи, слепой метод невозможен ни для пациентов, ни для лечащих врачей.

Следует отметить, что применение слепого метода чаще декларируется, чем осуществляется в действительности. Физиологические эффекты, например урежение пульса при приеме бета-блокаторов или угнетение кроветворной активности костного мозга при химиотерапии злокачественных новообразований, четко связаны с приемом конкретных препаратов. Определенные симптомы могут указывать на прием активного препарата.

Пример. В исследовании [8] первичной профилактики сердечно-сосудистых заболеваний было использовано практически идеальное плацебо. Одни участники получали холестирамин (экспериментальная группа), другие – порошок такого же вида, вкуса и запаха – плацебо (контрольная группа). Однако побочные эффекты встречались существенно чаще в экспериментальной группе. В конце первого года испытания у пациентов экспериментальной группы чаще, чем в контрольной, отмечались запоры (39% против 10%), изжога (27% против 10%), отрыжка и метеоризм (27% против 16%), а также тошнота (16% против 8%). Появление этих новых симптомов могло подсказать пациентам, какое лекарство они получают.

Имеются объективные свидетельства того, что пациенты и врачи в некоторых слепых испытаниях могут догадаться, кто какое лекарство получает.

Пример. Было предпринято двойное слепое рандомизированное испытание, чтобы выяснить, предотвращает ли прием пропранолола повторный инфаркт миокарда [9]. После окончания исследования, но до раскрытия кодов, пациентов и медицинский персонал попросили отгадать, какой препарат принимал пациент. Среди пациентов 79,9% правильно отгадали пропранолол и 57,2% – плацебо. Врачи и медицинские сестры показали такой же результат. Клинический персонал мог догадываться по частоте пульса; на чем основывали свою догадку пациенты – неизвестно.

Оценка исходов

Если исход испытания оценивается однозначно (например, выжил или умер пациент), то вероятность ошибочного определения исхода невелика. С другой стороны, если исходы определяются на основании мнения одного из участников, то возможность систематической ошибки значительно возрастает. Несмотря на то что факт смерти очевиден, причина ее часто неясна. Большинство людей умирают вследствие комплекса причин или из-за неясных причин. Это позволяет выдвигать различные предположения о причинах смерти. Вывод может зависеть от знания предшествовавших событий, включая характер лечения. Возможность систематической ошибки еще больше увеличивается при оценке таких симптомов, как боль, тошнота или депрессия. Систематической ошибки при оценке исходов можно избежать посредством применения одинакового для всех пациентов подхода к выявлению исходов, использования четких критериев их оценки, а также с помощью слепого метода.

Иногда, стремясь быстро провести исследование, немедленно получить и опубликовать его результаты, используют косвенные, легко оцениваемые исходы (surrogate outcomes) вместо настоящих клинических исходов (true outcomes). Например, в клинических исследованиях лечения ВИЧ-инфекции довольно часто за основной исход принимали результаты биологических тестов, показывающих выраженность инфекции (число CD4-лимфоцитов или присутствие антигена p24), а не клиническое прогрессирование заболевания (развитие сопутствующих инфекций или смерть). Однако число CD4-лимфоцитов оказалось несовершенным маркером клинического эффекта лечения. Как отмечалось в главе 1, замещение клинических исходов биологическими (косвенными) в исследованиях, результаты которых должны использоваться при лечении, считается оправданным, только если известно, что косвенный исход сильно связан с клиническим исходом.

Существует несколько способов оценки относительной эффективности двух методов лечения (табл. 7.1). В качестве наиболее подходящего клинического показателя предлагается *число больных, которых необходимо лечить* (ЧБНЛ, number needed to treat – NNT) для предотвращения одного неблагоприятного исхода [10]. Этот показатель является величиной, обратной снижению абсолютного риска.

На восприятие оценки эффективности лечения как пациентами, так и врачами влияет способ представления результатов. Результаты, выраженные через относительный риск, кажутся значительнее по величине, чем те же результаты, выраженные через до-

бавочный риск, а наименьшими кажутся эффекты, выраженные с помощью ЧБНЛ [11,12]. Кроме того, пациенты, которым сообщили их вероятность выжить, воспринимают эту цифру как более высокий шанс выживания, чем те, которые узнали другой показатель – вероятность умереть [13]. Таким образом, чтобы представить себе, каков эффект лечения, и передать это представление другим, необходимо выражать результаты испытания различными способами. Что такое "правильная" статистика – это спорный вопрос.

Испытания с анализом данных в зависимости от назначенного или полученного лечения

Результаты контролируемых рандомизированных испытаний могут анализироваться и представляться двумя способами: либо на основании факта назначения того или иного лечения при рандомизации, либо на основании лечения, полученного пациентом фактически (рис.7.6). Правильное представление результатов зависит от постановки вопроса.

Если вопрос состоит в том, какая тактика лечения более предпочтительна для принятия клинического решения, то следует применить анализ в зависимости от лечения, назначенного при рандомизации, независимо от того, все ли пациенты на самом деле получали это лечение. Этот подход называется *анализом в зависимости от назначенного лечения* (intention to treat analysis) [14]. Достоинства такого подхода в том, что поставленный вопрос соответствует тому, который обычно интересует клинициста при назначении лечения, и что сравниваемые пациенты действительно распределяются на группы случайным образом. Его недостаток обусловлен тем, что если многие пациенты не получали предпри-

Таблица 7.1
Способы оценки терапевтических эффектов*

Обобщенная оценка**	Определение
Снижение относительного риска	$\frac{\text{Частота событий в контрольной группе} - \text{частота событий в экспериментальной группе}}{\text{частота событий в контрольной группе}}$
Снижение абсолютного риска	Частота событий в контрольной группе – частота событий в экспериментальной группе
ЧБНЛ	$\frac{1}{\text{Частота случаев в контрольной группе} - \text{частота случаев в экспериментальной группе}}$

* [Laupacis A. et al. New Engl J Med 1988;318:1728 –1733].

** Для непрерывных данных, когда имеются результаты измерений до и после лечения, аналогичные оценки могут делаться либо путем сравнения средних величин для экспериментальной и контрольной групп после лечения, либо путем оценки различий между значениями до и после лечения.

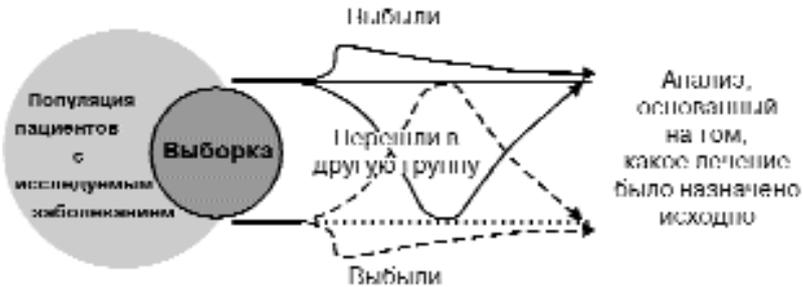
Анализ данных в зависимости от назначенного лечения**Анализ данных в зависимости от полученного лечения**

Рис. 7.6. Испытания с анализом данных в зависимости от назначенного или полученного лечения.

санного лечения, то различия между экспериментальной и контрольной группами стираются, и вероятность отрицательного результата исследования увеличивается. В таком случае отсутствие различия между группами может быть трактовано по-разному: то ли экспериментальное вмешательство на самом деле неэффективно, то ли оно просто не применялось.

Другое дело, если нас интересует действительно ли экспериментальное лечение лучше, чем контрольное. В этом случае для ответа больше подходит анализ в зависимости от полученного лечения, т.е. оценка эффекта того лечения, которое каждый больной получал в действительности и вне зависимости от того, какое лечение ему было назначено при рандомизации. При этом выясняется механизм изучаемого эффекта. Недостаток этого подхода в том, что если большинство пациентов не получили того лечения, кото-

рое им было предписано, испытание перестает быть рандомизированным и становится обычным когортным исследованием. Это значит, что все различия между группами, исключая способ лечения, должны быть каким-то образом нивелированы (путем введения ограничений, подбора пар, разделения на подгруппы или стандартизации) для достижения полной сопоставимости, как это делается при неэкспериментальных исследованиях.

Эффективность лечения

Результаты испытаний оцениваются в двух аспектах. Во-первых, может ли лечение быть эффективным при идеальных условиях? Во-вторых, является ли оно эффективным при реальных обстоятельствах?¹ (рис.7.7).

Вопрос о том, эффективно ли лечение в принципе, относится к проблеме идеальной эффективности. Идеально эффективное лечение обеспечивает желаемый эффект у тех больных, которые его получают. Идеальная эффективность определяется путем ограничения участвующих в исследовании пациентов лишь теми, кто будет полностью соблюдать медицинские предписания.

С другой стороны, лечение является реально эффективным, если приносит пользу в реальной ситуации. Для определения реальной эффективности пациентам предлагают схему лечения, допуская, что они будут либо следовать ей, либо нарушать ее, как это обычно происходит на практике. Лишь в немногих клинических испытаниях ставят вопрос о реальной эффективности. Отчасти это обусловлено риском, что результаты окажутся неубедительными. Обнаруженная в исследовании неэффективность лечения может быть следствием того, что оно в принципе неэффективно, что пациенты не соблюдали схему лечения или и тем и другим вместе взятым.

Применение результатов испытаний к конкретному больному

Клинические испытания включают в себя данные наблюдений за многими отличающимися друг от друга больными и дают усредненное описание того, что с ними происходит. Как получить более точные оценки для конкретного больного? Существует два пути: изучение подгрупп и изучение отдельных больных с помощью строгих научных методов, подобных тем, которые используются в рандомизированных испытаниях.

¹ В английском языке для обозначения этих двух концепций применяются соответственно термины "efficacy" и "effectiveness". Учитывая отсутствие общепринятого перевода этих терминов на русский язык, мы используем для их обозначения соответственно понятия "идеальной" и "реальной" эффективности. —Примеч. ред.

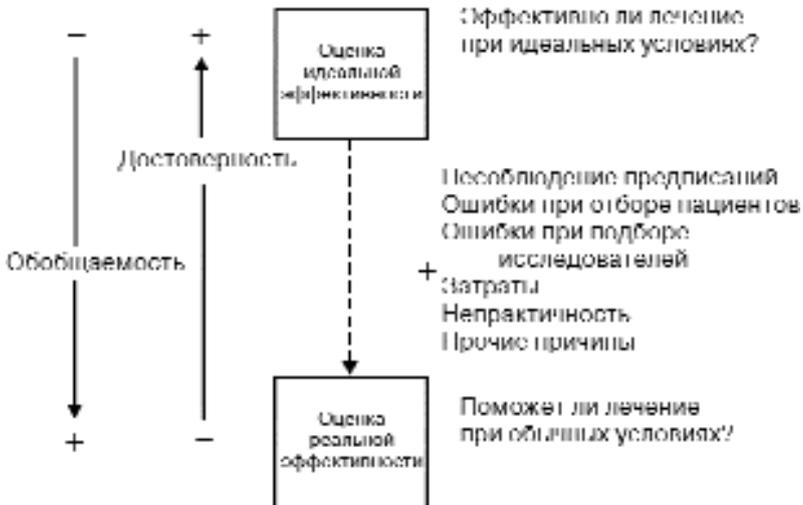


Рис. 7.7. Идеальная и реальная эффективность лечения.

Подгруппы

Главный результат клинических испытаний представляет собой описание наиболее важных исходов для каждой из основных групп воздействия. Но весьма заманчиво также провести более детальный анализ результатов, не ограничиваясь рамками общих выводов. Для этого рассматриваются подгруппы пациентов с особыми характеристиками или исходами. Однако в таких случаях возникает риск ошибочных заключений, что следует учитывать при анализе данных в подгруппах. (Некоторые понятия, на которых основано изложение в данном разделе, обсуждаются в главе 9.)

Одна из опасностей при изучении подгрупп состоит в повышенном шансе обнаружения в подгруппе эффектов, не существующих на самом деле. Этот эффект возникает из-за того, что множественные сравнения повышают шансы ложноположительных результатов в большей степени, чем это оценивается по значению p для каждого сравнения в отдельности (см. главу 9). Некоторые общие соображения, полезные при решении вопроса о реальности обнаруженных различий между подгруппами, перечислены в табл. 7.2.

Вторая опасность состоит в возможности ложноотрицательного вывода. Изучение подгрупп в клиническом испытании (пациентов с определенными характеристиками или особыми исходами) резко уменьшает объем данных, так что часто невозможно сделать

обоснованные заключения. Тем не менее искушение заняться анализом в подгруппах присутствует, и некоторая предварительная информация для такого анализа может быть собрана.

Пример. Исследование здоровья врачей [15] – рандомизированное контролируемое испытание было проведено для выяснения того, снижает ли ежедневный прием аспирина смертность от сердечно-сосудистых заболеваний у практически здоровых врачей-мужчин. Исследование было остановлено преждевременно, поскольку смертность среди врачей оказалась гораздо ниже, чем предполагалось исходно. Другая причина прекращения исследования – меньшее число случаев инфаркта миокарда в группе принимавших аспирин по сравнению с контролем. Авторы полагали, что хотя им и не удалось получить ответа на основной вопрос исследования, влияние ежедневного приема аспирина на частоту возникновения инфаркта миокарда было подлинным, поскольку оно биологически объяснимо, обнаружено также в других исследованиях и, наконец, поскольку шансы ложноположительного вывода были оценены как очень низкие (1 на 10 000). С другой стороны, авторы наблюдали небольшое увеличение риска развития инсульта в экспериментальной группе. Но они не могли быть уверены в реальности этого эффекта, так как испытуемых с подобным исходом оказалось слишком мало. Таким образом, в исследовании, которое не могло ответить на главный поставленный вопрос, авторы оценили свои наблюдения в подгруппах как достоверные, опираясь на общую информацию, имеющуюся по данному вопросу.

Таблица 7.2 Принципы оценки реальности наблюдаемых различий эффектов в подгруппах*

Вопросы, касающиеся данного исследования

- Является ли величина наблюдаемого различия клинически важной?
- Насколько вероятно случайное возникновение эффекта, учитывая:
 - число исследованных подгрупп?
 - величину p ?
- Была ли гипотеза о наличии эффекта:
 - высказана еще до его выявления?
 - выдвинута для объяснения уже выявленного эффекта?
 - одной из малого числа гипотез?

Вопросы, касающиеся других исследований

- Было ли различие обнаружено путем сопоставления данных одного исследования или при сравнении данных нескольких исследований?
- Наблюдался ли подобный эффект в других исследованиях?
- Есть ли косвенные свидетельства в пользу существования эффекта?

* [Oxman A.D., Guyatt G.H. Ann Intern Med 1992; 116:78–84].

Оценка реальной эффективности лечения для конкретных больных

Лечение, эффективное в среднем для группы пациентов, может оказаться безуспешным у конкретных больных. Результаты достоверного клинического испытания служат достаточным основанием, чтобы начать лечение конкретного больного, но для продолжения лечения более важен опыт наблюдения за этим больным. Следовательно, прежде чем приступить к выполнению программы лечения, полезно задать себе следующие вопросы:

- Известно ли, что это лечение эффективно в идеальных условиях для любых больных?
- Известно ли, что в среднем это лечение реально эффективно для аналогичных пациентов?
- Оправдывает ли польза от лечения неудобства и риск, связанные с вмешательством?
- Эффективно ли лечение для моего пациента?

Задавая эти вопросы, а не просто руководствуясь результатами испытаний, можно обезопасить себя от неудачного выбора лечения или упрямого неприятия плохих результатов.

Испытания на единственном больном

Тщательные клинические исследования, проводимые с должным вниманием к систематическим и случайным ошибкам, могут быть выполнены за определенное время с каждым больным в отдельности [16]. Метод, называемый *испытанием на единственном больном*, представляет собой усовершенствованный вариант более общего неформализованного процесса проб и ошибок, который так обычен для клинической практики. Пациенту последовательно назначается то или иное лечение (например, активный препарат или плацебо) в случайном порядке, на короткий период – 1–2 нед. Ни пациент, ни врач не знают, какое лекарство назначено. Исходы (например, просто субъективное отношение пациента к лечению или выраженность симптомов) оцениваются после каждого периода и подвергаются статистическому анализу.

Этот метод целесообразен, когда течение заболевания непредсказуемо, реакция на лечение проявляется быстро и нет наложения фармакологических эффектов после смены препаратов. Примерами заболеваний, для которых может применяться метод, служат мигрень, бронхоспазм, фибромиозит, функциональные заболевания кишечника.

Испытания на единственном больном помогают принимать клинические решения, хотя и для относительно небольшой доли пациентов. Они также полезны при рассмотрении интересных клинических гипотез с целью отбора некоторых из них для последующей оценки в полноценном рандомизированном клиническом испытании, включающем большое число больных.

Альтернативы рандомизированным испытаниям

Рандомизированные контролируемые испытания с использованием слепого метода служат "золотым стандартом" для сравнения эффектов лечения. Их результатам следует всегда отдавать предпочтение перед любой другой имеющейся информацией об эффектах лечения. Однако опираться на результаты клинических испытаний не всегда возможно.

Ограничения рандомизированных испытаний

Клинические испытания имеют ряд ограничений. Может не оказаться достаточного числа пациентов с изучаемым заболеванием, в одном месте и в один период времени, чтобы провести весомое с научной точки зрения испытание. Клинические испытания требуют больших средств – иногда более 50–100 млн долларов. К тому же проходят годы, прежде чем будут получены результаты, что может быть неприемлемо для таких серьезных, общественно значимых заболеваний, как СПИД.

Метод лечения может быть настолько общепринятым, несмотря на отсутствие обоснованных доказательств его пользы, что трудно убедить врачей и пациентов в необходимости его испытания. Можно, конечно, настаивать на том, что если эффект лечения неизвестен, то единственный этический подход состоит в проведении исследования (и что неэтично продолжать применять лечение, польза которого не доказана). Однако для восприятия этих доводов требуется определенный уровень аналитических способностей, нетипичный для пациентов и практикующих врачей. Некоторые врачи выступают за "рандомизацию с первого пациента", иными словами за то, чтобы начинать испытания тотчас же после появления нового метода лечения. Другие считают, что лучше проводить тщательные клинические испытания несколько позднее, после уточнения методики применения нового средства, чтобы оценивать эффективность оптимального режима вмешательства. Во всяком случае, все соглашаются, что если откладывать контролируемое испытание слишком долго, может быть упущена сама возможность его проведения.

По этим причинам для многих клинических решений отсут-

ствуют рекомендации, основанные на результатах клинических испытаний. Но это не снимает необходимости принимать такие решения. Каковы альтернативы рандомизированным испытаниям и насколько эти альтернативы заслуживают доверия?

Преимущества и недостатки альтернативных методов

Альтернативные методы оценки эффективности обычно основаны на использовании больших баз данных – клинических, финансовых, административных. Часто используются данные, собиравшиеся для ответа на какой-то другой вопрос. Задачи и методика исследования могут быть определены уже после того, как собраны данные, так что ресурсы затрачиваются в основном на анализ данных. Этот процесс называется *вторичным анализом данных* (secondary data analysis), поскольку ответ на вопрос исследования не был первичной целью сбора информации.

Использование таких данных для исследования имеет ряд преимуществ, все они – практические. Во-первых, если база данных содержит результаты наблюдения за большим количеством пациентов, ответ на изучаемый вопрос с высокой вероятностью не будет следствием случайности. Даже анализ в подгруппах (например, принимающие эстрогены женщины пожилого возраста или молодые мужчины с впервые перенесенным инфарктом миокарда передней локализации) можно проводить с высоким уровнем статистической значимости. В большинстве клинических испытаний подобное обилие пациентов не планируется из-за высоких затрат, лучшие из клинических испытаний обычно обеспечивают ответ только на главный вопрос исследования применительно ко всем участвующим пациентам, но редко позволяют ответить на тот же вопрос в отношении подгрупп.

Во-вторых, упомянутые базы данных содержат сведения, более приближенные к обычной клинической практике, чем данные клинических испытаний. Они отражают опыт работы учреждений системы здравоохранения или даже целых регионов и стран, а не наблюдения за селективной группой экспериментальных пациентов. Следовательно, обобщаемость результатов выше.

В-третьих, использование уже собранных данных дешевле, чем сбор новых при клиническом испытании. В рандомизированных испытаниях на поиск, обследование, включение и наблюдение одного пациента затрачиваются тысячи долларов, тогда как анализ имеющихся данных относительно недорог.

Наконец, работа с существующими данными позволяет получить ответ на важный вопрос за довольно короткий срок. Клинические испытания от включения первого пациента до конца про-

спективного наблюдения часто делятся годами. Иногда клиницисты нуждаются в быстром ответе, каким бы несовершенным он ни был, поскольку ответственные решения о выборе способа лечения приходится принимать ежедневно.

Этим практическим преимуществам противостоят недостатки. Сбор данных обычно проводится не так тщательно, как в хорошо организованных клинических испытаниях. Например, диагноз "артериальная гипертензия" означает то, что под этим подразумевает установивший его врач, тогда как в испытании будут указаны уровень артериального давления, метод и частота измерения, а также может быть сделана поправка на возраст. В базе данных могут отсутствовать некоторые важные для исследования показатели, поскольку они не имели значения для первоначальной цели сбора данных. И, конечно, существует проблема проведения сравнений, свободных от систематических ошибок.

Компромисс между скоростью и простотой, с одной стороны, и достоверностью, с другой, будет рассмотрен для каждого альтернативного метода исследования в следующем разделе.

Исследования с непараллельным контролем

Пациенты контрольной группы могут отличаться от пациентов экспериментальной группы по времени и месту их включения в исследование. Например, сравнивается прогноз для пациентов, недавно леченных современными препаратами, и для пациентов, лечившихся в прошлом, когда таких препаратов еще не было. Или проводится сравнение результатов хирургической операции в двух больницах, где применяются разные методики вмешательства. Такой подход удобен, но проблема состоит в том, что время и место почти всегда тесно связаны с прогнозом. Клинические испытания с попыткой точного сравнения групп пациентов, получающих лечение в разные временные периоды или в разных учреждениях, ставят очень трудную задачу.

Иногда результаты современного лечения сравниваются с данными наблюдения аналогичных пациентов в прошлом. Это называется *историческим*, или *непараллельным контролем* (historical or nonconcurrent control). Хотя работа может быть тщательно выполнена, такая методика исследования имеет множество недостатков. Методы диагностики со временем изменяются, одновременно меняется и средний прогноз. Так, новые диагностические технологии создавали впечатление, что с течением времени прогноз при лечении рака легкого улучшается, хотя это было не так [17]. При современных возможностях выявления метастазов больные были отнесены к группе с более поздней стадией заболе-

вания, чем подобные больные раньше; и именно эта "миграция по стадиям" обеспечила более благоприятный прогноз для каждой стадии, чем это сообщалось ранее. Поддерживающее лечение (например, антибиотики, пищевые добавки или профилактика пептической язвы) тоже со временем совершенствуется, способствуя общему улучшению прогноза, которое не обязательно обусловлено появившимся новым специфическим лечением.

Пример. Sacks et al. [18] опубликовали обзор клинических испытаний по шести видам лечения, чтобы выяснить, отличаются ли результаты испытаний с параллельным контролем от результатов, которые получены в исследованиях тех же видов лечения с историческим контролем. Было рассмотрено 50 рандомизированных испытаний и 56 исследований с историческим контролем. В 79% исследований с историческим контролем и только в 20% рандомизированных испытаний с параллельным контролем экспериментальное лечение оказалось эффективно. Различия между результатами двух видов исследований возникали преимущественно из-за того, что в испытаниях с историческим контролем пациенты контрольной группы имели более тяжелое течение заболевания. Стандартизация по прогностическим факторам не изменила результат; по-видимому, различия были обусловлены общим улучшением лечения либо отбором пациентов с меньшей тяжестью заболевания.

Таким образом, если брать в качестве стандарта достоверности рандомизированные контролируемые испытания, то результаты исследований с историческим контролем оказываются смещенными в пользу экспериментального лечения, причем это смещение не может быть преодолено поправкой на известные прогностические факторы.

При использовании исторического контроля сравнение тем надежнее, чем короче период между формированием экспериментальной и контрольной групп и чем меньше изменились за этот период другие аспекты медицинской помощи. Так, в некоторых онкологических центрах эффективность методов химиотерапии исследуется при сравнении нового метода с последним использовавшимся, например, в прошлом году. Наилучший способ избежать систематической ошибки – использование *параллельного контроля* (concurrent control) (т.е. пациентов, наблюдающихся в тот же период времени).

Опыт другого учреждения, с иными методами лечения, может служить стандартом для сравнения. Однако более предпочтительно отбирать пациентов для экспериментальной и контрольной групп в одном и том же учреждении, поскольку множество факторов (контингент пациентов, организация лечебного процесса, ква-

лификация персонала и т.д.) часто приводит к значительным отличиям прогноза в разных учреждениях, независимо от изучаемого вмешательства.

Пример. Внутрибольничная летальность в больницах центральной Пенсильвании, проводивших операции аортокоронарного шунтирования, различалась почти в 3 раза (рис.7.8) [19]. Степень тяжести заболевания и, следовательно, прогноз пациентов в этих больницах также различались. После того как были рассчитаны с учетом прогностических факторов ожидаемые частоты смертельных случаев, в одной больнице летальность оказалась ниже ожидаемой, в другой совпала, а в третьей была выше ожидаемой. Любое заслуживающее доверия сравнение результатов вмешательства в этих больницах должно принимать во внимание не только различия в степени тяжести пациентов, но и квалификацию хирургов и персонала.

Неконтролируемые исследования

Неконтролируемые исследования описывают течение заболевания в одной группе пациентов, подвергающихся изучаемому

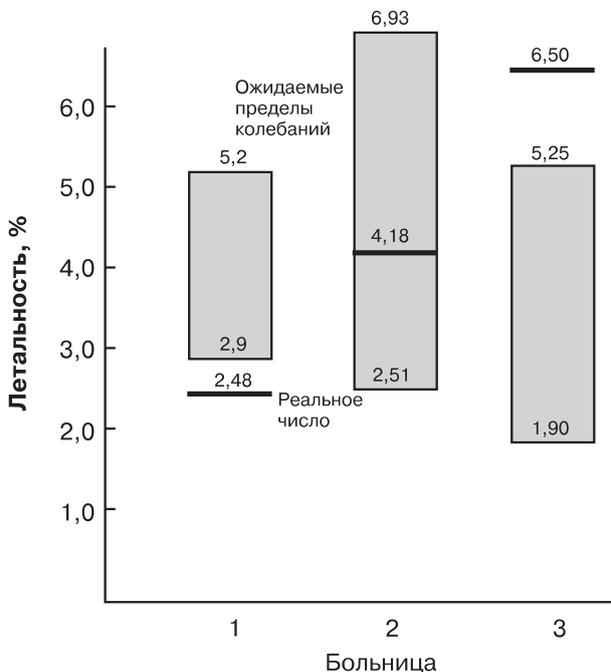


Рис. 7.8. Наблюдаемая и ожидаемая (с учетом состава пациентов) летальность при операции аортокоронарного шунтирования в трех больницах [Topol E.J., Califf R.M. Ann Intern Med 1994;120:65–70].

по 1964 г. [20]. Хотя эффективное лечение не применялось (поскольку в те годы его просто не было), болезнь протекала с чередованием периодов обострений и продолжительных ремиссий. Разумеется, периоды обострений вызывают как у больного, так и у врача тревогу и желание что-либо предпринять. Если лечение начать на пике обострения, за ним последует период улучшения. В отсутствие другого сравнения (иного, чем активность болезни в предыдущий период) улучшение можно было бы отнести за счет лечения.

Неспецифические эффекты

В неконтролируемых испытаниях невозможно отделить эффекты Готорна и плацебо от воздействия лечения как такового. Но если имеются контрольные пациенты, получающие такой же уход, как те, кому дают исследуемое лечение, и если они получают плацебо, то при сравнении эти эффекты устраниваются.

Смещение к среднему

Лечение часто назначается из-за крайних или необычных проявлений заболевания, например очень высокого артериального давления или лихорадки. В подобной ситуации последовательно проводимые измерения выявят улучшение по чисто статистическим причинам. Как обсуждалось в главе 2, пациенты, отобранные по экстремально высоким значениям показателя, должны в среднем иметь более низкие значения при последующих оценках. Если таких пациентов подвергают вмешательству после того, как у них впервые обнаружено необычное проявление, а затем проводят оценку эффекта вмешательства, то улучшения следует ожидать, даже если лечение неэффективно.

Спонтанное улучшение

Некоторые заболевания протекают со спонтанным улучшением. В таких случаях лечение может совпадать по времени с улучшением, но не быть его причиной. Например, пациенты часто обращаются по поводу таких острых саморазрешающихся заболеваний, как респираторная инфекция или желудочно-кишечное расстройство на пике проявления симптомов. После обращения к врачу больные начинают выздоравливать вследствие естественно-го течения заболевания, независимо от назначенного лечения.

Нерандомизированные исследования

При включении в клиническое испытание пациенты могут быть разделены на экспериментальную и контрольную группы по

решению лечащего врача. Такое исследование имеет все преимущества и недостатки когортных исследований.

Преимущество когортных исследований обусловлено тем, что решения о выборе способа лечения все равно должны приниматься, невзирая на имеющийся уровень знаний. В отсутствие четкого мнения о том, какой метод лечения предпочтителен, часто применяются разные методы. Поэтому в неэкспериментальных исследованиях большое число больных получает разное лечение со свойственными каждому методу эффектами. Если опыт подобного наблюдения собрать и должным образом проанализировать, он может помочь в принятии клинических решений.

К сожалению, часто трудно быть уверенным в том, что проводимые в обсервационном исследовании сравнения не подвержены систематическим ошибкам. Решения о способе лечения принимаются на основании множества факторов – степени тяжести заболевания, наличия сопутствующих заболеваний, местных традиций, согласия больного и т.д. Пациенты, получающие различные виды лечения, скорее всего отличаются и по другим признакам. Следовательно, возможность определения изолированного эффекта лечения, свободного от влияния других факторов, в таком исследовании весьма сомнительна.

Фазы испытания лекарственных средств

Испытание лекарственных средств обычно проходит через три последовательные фазы [21]. В фазу I устанавливается диапазон доз, которые легко переносятся больными и безопасны (по крайней мере в отношении частых и тяжелых побочных реакций). В исследовании фазы I включается очень малое число пациентов (около 10), контрольная группа отсутствует. Фаза II обеспечивает предварительную информацию об идеальной эффективности препарата и о соотношении между дозой и эффективностью. В исследуемые группы включается небольшое число пациентов, чтобы выявить только самые значительные эффекты; испытание может проводиться слепым методом и иметь контрольную группу. Фаза III окончательно выясняет идеальную эффективность и обычные побочные реакции; в испытание включается достаточное число пациентов (от десятков до тысяч), чтобы обнаружить клинически важные эффекты лечения; результаты обычно публикуются в медицинских журналах и используются регулирующими инстанциями при лицензировании лекарственных средств.

Испытания фазы III недостаточно масштабны, чтобы обнаружить различия в частоте (и даже существовании) необычных побочных реакций. (См. обсуждение статистической чувствительно-

сти в главе 9.) Для этого необходимо проследить очень большое число пациентов уже после появления препарата в продаже – процесс, называемый постмаркетинговым исследованием, или испытанием фазы IV.

Резюме

Ценные идеи относительно возможных средств лечения следует подвергать строгой научной проверке, прежде чем использовать для принятия клинических решений. Наилучшей проверкой служат рандомизированные контролируемые испытания – особый тип когортных исследований. В этих исследованиях вмешательства распределяются случайным образом, поэтому они свободны от систематических ошибок. В клинических испытаниях пациенты обычно представляют собой тщательно отобранную группу, что снижает обобщаемость результатов. Пациенты случайным образом распределяются на получающих либо экспериментальное вмешательство, либо некоторое сравнительное воздействие: обычное лечение, плацебо или простое наблюдение. Сразу после рандомизации (перед вмешательством) сравниваемые группы имеют в среднем одинаковый прогноз, однако не обусловленные лечением различия между группами могут возникнуть позднее. Они связаны с отказом от предписанного лечения, выбыванием из исследования, приемом лечения, предназначенного для другой группы, различиями в тактике ведения больных и в сопутствующей терапии. Использование слепого метода для всех участников испытания помогает минимизировать систематические ошибки при рандомизации больных, их ведении и оценке исходов, но успешное применение этого метода возможно не всегда.

Существует два способа анализа данных в рандомизированном испытании. Первый способ – анализ в зависимости от назначенного лечения, т.е. в соответствии с группами, сформированными при рандомизации; результат служит критерием для принятия клинических решений. Второй способ – анализ в зависимости от фактически полученного лечения; результат позволяет судить о биологических механизмах действия вмешательства, но не имеет прямого влияния на клиническое решение; недостаток этого способа состоит в том, что если большинство пациентов не получили изначально предписанного им лечения, испытание перестает быть рандомизированным. Для того чтобы получить информацию, применимую к конкретному пациенту в большей степени, чем это обеспечивает обобщенный результат испытания, клиницисты могут опираться на результаты наблюдений за подгруппами пациен-

тов (хотя это сопряжено с дополнительным риском ошибочных выводов) или проводить испытания на своих собственных пациентах.

При решении многих клинических вопросов невозможно или непрактично опираться на результаты рандомизированных клинических испытаний. В качестве компромисса используются результаты сравнения с пациентами, получавшими другое лечение в прошлом, с этими же больными, получавшими ранее другое лечение, или данные нерандомизированных исследований. Такие компромиссы уменьшают достоверность данных.

Литература

1. EC/IC Bypass Study Group. Failure of extracranial-intracranial arterial bypass to reduce the risk of ischemic stroke. *N Engl J Med* 1985;313:1191–1200.
2. Mayberg M.R., Wilson E., Yatsu F., Weiss D.G., Messina L., Hershey L.A., Coiling C., Eskridge J., Deykin D., Winn H.R. Carotid endarterectomy and prevention of cerebral ischemia in symptomatic carotid stenosis. *JAMA* 1991;266:3289–3294.
3. Opie on the heart [Editorial]. *Lancet* 1980; 1:692.
4. Rubenstein L.Z., Robbins A.S., Josephson K.R., Schulman B.L., Osterweil D. The value of assessing falls in an elderly population. A randomized controlled trial. *Ann Intern Med* 1990; 113:308–316.
5. Fischer R.W. Comparison of antipruritic drugs administered orally. *JAMA* 1968;203:418–419.
6. Ziegler E.J., et al. Treatment of gram negative bacteremia and septic shock with HA-IA human monoclonal antibody against endotoxin. *New Engl J Med* 1991; 324:429–436.
7. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1980;303:1038–1041.
8. Lipid Research Clinics Program. The Lipid Research Clinics coronary primary prevention trial results. 1. Reduction in incidence of coronary heart disease. *JAMA* 1984;251:351–364.
9. Byington R.P., et al. Assessment of double-blindness at the conclusion of the beta-blocker heart attack trial. *JAMA* 1985;253:1733–1736.
10. Laupacis A., Sackett D.L., Roberts R.S. An assessment of clinically useful measures of the consequences of treatment. *New Engl J Med* 1988;318:1728–1733.
11. Naylor C.D., Chen E., Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann Intern Med* 1992;117:916–921.
12. Malenka D.J., Baron J.A., Johansen S., Wahrenberger J.W., Ross J. The framing effect of relative and absolute risk. *J Gen Intern Med* 1993;543–48.
13. McNeil B.J., Pauker S.G., Sox H.C. Jr., Tversky A. On the elicitation of preferences for alternative therapies. *New Engl J Med* 1982;306:1259.
14. Sackett D.L., Gent M. Controversy in counting and attributing events in clinical trials. *N Engl J Med* 1979;301:1410–1412.
15. Steering Committee of the Physicians' Health Study Research Group. Final report of the aspirin component of the ongoing Physicians' Health Study. *New Engl J Med* 1989;321:129–135.
16. Guyatt G., Sackett D., Taylor D.W., Chong J., Roberts R., Pugsley S. Determining optimal therapy—randomized trials in individual patients. *N Engl J Med* 1986;314:889–892.
17. Feinstein A.R., Sosin D.M., Wells C.K. The Will Rogers phenomenon. Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *New Engl J Med* 1985;312:1604–2608.
18. Sacks H., Chalmers T.C., Smith H. Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982;72:233–240.
19. Topol E.J., Califf R.M. Scorecard cardiovascular medicine. Its impact and future direction. *Ann Intern Med* 1994;120:65–70.

20. Ropes M. Systemic lupus erythematosus. Cambridge, MA: Harvard University Press, 1976.
21. Spilker B. Guide to clinical interpretation of data. New York: Raven Press, 1986.

Рекомендуемая литература

- Chalmers T.C., Smith H. Jr, Blackburn B., Silverman B., Schroeder B., Reitman D., Ambroz A. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981;2:31–49. Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ont. How to read clinical journals. V: To distinguish useful from useless or even harmful therapy. *Can Med Assoc J* 1981;124:1156–1162.
- DerSimonian R., Charette L.J., McPeck B., Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982;306:1332–1337.
- Feinstein A.R. An additional basic science for clinical medicine. II: The limitations of randomized trials. *Ann Intern Med* 1983;99:544–550.
- Friedman L.M., Furberg C.D., De Mets D.L. *Fundamentals of clinical trials*, 2nd ed. Littleton, MA: John Wright PSG, 1985.
- Guyatt G.H., Sackett D.L., Cook D.J. How to read clinical journals. II: How to use an article about therapy or prevention. A: Are the results of the study valid? *JAMA* 1993;270:2598–2601.
- Guyatt G.H., Sackett D.L., Cook D.J. How to read clinical journals. II: How to use an article about therapy or prevention. B: What were the results and will they help me in caring for my patients? *JAMA* 1994;271:59–63.
- Guyatt G., Sackett D., Taylor D.W., Chong J., Roberts R., Pugsley S. Determining optimal therapy-randomized trials in individual patients. *N Engl J Med* 1986;314:889–892.
- Hellman S., Hellman D.S. Of mice and men. Problems of the randomized clinical trial. *New Engl J Med* 1991;324:1585–1589.
- Laupakis A., Wells G., Richardson S., Tugwell P. Users' guide to the medical literature. V: How to use an article about prognosis. 1994; 272:234–237.
- Lavori P.W., Louis T.A., Bailar J.C. III, Polansky M. Design for experiments-parallel comparisons of treatment. *N Engl J Med* 1983;309:1291–1298.
- Meinert C.L. *Clinical trials: design, conduct and analysis*. New York: Oxford University Press, 1986.
- Mosteller F., Gilbert J.P., McPeck B. Reporting standards and research strategies for controlled trials. *Control Clin Trials* 1980; 1:37–58.
- Oxman A.D., Guyatt G.H. A consumer's guide to subgroup analysis. *Ann Intern Med* 1992;116:78–84.
- Peto R., Pike M.C., Armitage P., Breslow N.E., Cox D.R., Howard S.V., Mantel N., McPherson K., Peto J., Smith P.G. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, part 1. *Br J Cancer* 1976;34:585–612.
- Peto R., Pike M.C., Armitage P., Breslow N.E., Cox D.R., Howard S.V., Mantel N., McPherson K., Peto J., Smith P.G. Design and analysis of randomized clinical trials requiring prolonged observation of each patient, part 2. *Br J Cancer* 1977;35:1–39.
- Pocock S.J. *Clinical trials: a practical approach*. New York: John Wiley & Sons, 1983.
- Yusuf S., Collins R., Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; 3:409–420.

8

ПРОФИЛАКТИКА

Веди себя благоразумно – среди тысяч людей только один умирает естественной смертью, остальные погибают вследствие безрассудной манеры бытия.

Маймонид (средневековый философ, 1135–1204)

Большинство врачей посвящают себя медицине потому, что стремятся излечивать больных. Однако люди хотят вообще никогда не болеть, а уж если это случилось, то чтобы болезнь была обнаружена как можно раньше и с ней было бы покончено прежде, чем она причинит им вред. Поэтому людей, у которых нет характерных жалоб, обследуют для выявления и коррекции факторов риска. Это позволяет предотвратить или распознать заболевание на ранней стадии, когда своевременное вмешательство сохраняет здоровье. Такие *периодические обследования* относятся к мерам по *охране здоровья населения*.

Охрана здоровья занимает важное место в клинической практике [1]. Эта деятельность может составлять часть медицинского обслуживания, например, когда врач измеряет артериальное давление у пациента, жалующегося на боль в горле. Иногда люди посещают врача специально для периодического обследования.

Врач должен понимать концептуальную основу и содержание периодических обследований. Он должен быть готов ответить на вопросы пациента: "Почему в этом году мне снова нужно делать мазок Папаниколау?" или "Моему соседу делают рентгеноскопию грудной клетки каждый год, почему Вы не направляете меня на это исследование?"

В данной главе рассматривается индивидуальная профилактическая работа врачей с пациентами. Однако эффективна профилактика и на популяционном уровне. Обязательные вакцинации учащихся, запрет на курение в общественных местах, законодательные ограничения продажи огнестрельного оружия – вот примеры профилактики на государственном уровне. Для решения таких проблем, как предупреждение огнестрельных ранений, наиболее эффективна общественная профилактика, для других, таких как профилактика рака толстой кишки, – обследование в услови-

ях стационара. Для третьих необходимо сочетание усилий врачей с широкой профилактической деятельностью на уровне государства (например, профилактика курения складывается как из индивидуальной работы врачей, помогающих конкретным больным бросить курить, так и из общественных мер – образование, законы, налоги имеют целью предупреждение курения среди подростков).

Большинство научных подходов к профилактике в клинической медицине, в частности принципы, лежащие в основе выбора диагностических тестов, прогнозирования заболеваний и оценки эффективности вмешательств, уже рассматривались выше. Данная глава посвящена развитию этих принципов и стратегии их применения для профилактики.

Уровни профилактики

Словарь Вебстера [2] определяет профилактику как "действие, направленное на предупреждение события". Исходя из такого определения, можно утверждать, что почти вся деятельность в медицине может быть описана как профилактика. В конечном счете усилия врачей направлены на предотвращение преждевременной смерти, болезни, инвалидизации, дискомфорта, неудовлетворенности и лишений (см. главу 1). Тем не менее в практической медицине, как будет показано далее, дается более строгое определение профилактики. Хотя в настоящее время профилактические мероприятия применяются намного шире, чем ранее, клиницисты по-прежнему тратят основную часть своего времени на диагностику и лечение, а не на предупреждение заболеваний.

В зависимости от того, на каком этапе осуществляется вмешательство, возможны три типа, или уровня, профилактики (рис. 8.1).

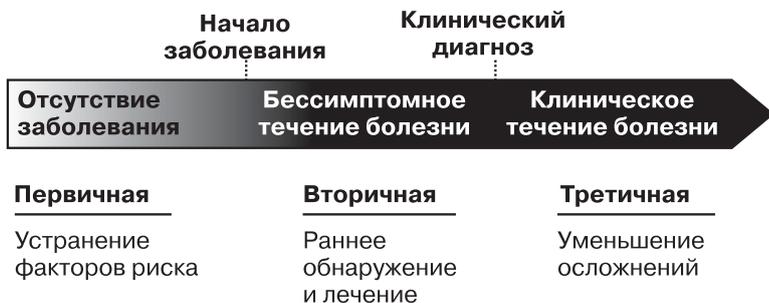


Рис. 8.1. Уровни профилактики.

Первичная профилактика All Caps

Первичная профилактика предупреждает само возникновение болезни, устраняя ее причины. Назначение фолиевой кислоты для предупреждения пороков развития нервной трубки, вакцинация против многих инфекционных заболеваний или обучение пациентов здоровому образу жизни (например, помощь в прекращении курения, рекомендации по потреблению пищи с низким содержанием холестерина и высоким содержанием клетчатки, поддержание необходимого уровня физической активности, ведение безопасной сексуальной жизни) – все это примеры первичной профилактики.

Первичная профилактика часто осуществляется на государственном уровне, вне системы здравоохранения, как это отмечалось выше. Примерами служат хлорирование и фторирование воды, принятие законов о применении ремней безопасности в автомобилях и шлемов для мотоциклистов. Некоторые меры первичной профилактики применяются на определенных производствах (использование наушников или противопылевых масок), в школах (вакцинация), в специализированных медицинских учреждениях (тестирование донорской крови на вирус гепатита В и ВИЧ).

Вторичная профилактика

Меры *вторичной профилактики* позволяют обнаруживать болезнь на ранних стадиях, когда она протекает бессимптомно и своевременное лечение может остановить ее развитие. Примерами служат кольпоцитодиагностика по Папаниколау, маммография и анализ кала на скрытую кровь. Большинство мероприятий вторичной профилактики осуществляется в медицинских учреждениях и в этой деятельности участвуют все врачи, особенно работающие со взрослыми. Существуют общественные программы, например скрининг на глаукому в больших магазинах.

Скрининг

Скрининг (screening) – это идентификация не распознанного ранее у пациента заболевания или фактора риска путем опроса (например, относительно курения), физикального исследования (например, пальпация предстательной железы), лабораторного исследования (например, определение фенилаланина в сыворотке) или с помощью других процедур (например, сигмоскопии), которые могут быть выполнены относительно быстро. Среди лиц, считающих себя здоровыми, с помощью скрининга выделяют тех, у кого имеется заболевание или факторы риска. Скрининг входит в комплекс многих мероприятий по первичной профилактике и

всех мероприятий по вторичной профилактике. Скрининговые тесты не предназначены для диагностики. Если врач не собирается проводить дальнейшее обследование и лечение лиц с отклонениями от нормы, то в таких тестах вообще нет смысла.

Третичная профилактика

Третичная профилактика относится к действиям, направленным на предотвращение ухудшения или осложнений заболевания после того, как болезнь проявилась. Пример: использование бета-блокаторов для уменьшения риска смерти у пациентов, перенесших инфаркт миокарда. Третичная профилактика совпадает с лечением, однако при достаточно высоком уровне организации она способна решать еще ряд проблем, связанных с болезнью. Например, для больных сахарным диабетом третичная профилактика выходит за рамки простого контроля уровня глюкозы в крови: требуются регулярное обследование офтальмологом для раннего обнаружения диабетической ретинопатии, обучение уходу за стопами, поиск и устранение других факторов риска сердечно-сосудистых заболеваний, регулярный контроль за содержанием белка в моче для своевременного применения ингибиторов ангиотензинпревращающего фермента с целью предупреждения почечной недостаточности.

Третичная профилактика особенно важна при уходе за пациентами со смертельными заболеваниями. Задача здесь состоит не в предотвращении смерти, а в максимальном продлении периода с приемлемым качеством жизни. Например, сейчас не существует специфического лечения больных с боковым амиотрофическим склерозом, который заканчивается параличом дыхательных и глотательных мышц. Однако тщательный медицинский уход включает в себя наложение гастростомы, что предупреждает голодание и обезвоживание организма, трахеостомы для облегчения дыхания; по желанию пациента возможно подключение портативного дыхательного аппарата для отдыха дыхательных мышц. Без подобного активного подхода у больного может возникнуть острая дыхательная недостаточность как результат самого заболевания, обезвоживания и пневмонии. В этом случае пациент, его семья и врач столкнутся с необходимостью госпитализации больного в палату интенсивной терапии в надежде недопустить дальнейшего ухудшения состояния и восстановить приемлемое качество жизни еще на некоторый период. Третичная профилактика может предотвратить подобное течение событий.

Осуществление программ третичной профилактики невозможно вне системы здравоохранения; в них участвуют не только врачи, но и многие другие медицинские работники.

Цель периодических обследований

При рассмотрении вопроса о рутинном обследовании лиц без характерных симптомов следует сначала определить, какие медицинские проблемы или заболевания предполагается предотвратить. Это утверждение кажется банальным. Однако на самом деле многие профилактические процедуры, особенно скрининг, осуществляются без ясного представления о том, что же на самом деле хотят найти. Например, при рутинном обследовании часто назначается общий анализ мочи. Но ведь общий анализ мочи может использоваться для диагностики самых разных заболеваний, включая сахарный диабет, бессимптомные инфекции мочевыводящего тракта или мочекаменную болезнь. Поэтому прежде чем назначать этот анализ, необходимо решить, есть ли смысл его делать, а если есть, то на какое заболевание проводится скрининг.

Для решения вопроса о том, какие состояния требуют скрининга, важны три критерия (табл. 8.1): 1) тяжесть заболевания; 2) достоинства метода, применяемого для скрининга; 3) эффективность мер первичной профилактики (например, консультирования по поводу безопасной половой жизни) или эффективность

Таблица 8.1
Критерии необходимости включения изучаемого состояния в периодические обследования

1. Насколько велика тяжесть заболевания, с точки зрения вызываемых им		
Смерти	Дискомфорта	
Болезни	Неудовлетворенности	
Инвалидизации	Лишений	
2. Насколько эффективен тест, используемый для скрининга, с точки зрения его		
Чувствительности	Стоимости	Эффекта ярлыка*
Специфичности	Безопасности	
Простоты	Приемлемости	
3. а) Какова эффективность вмешательства в качестве меры первичной профилактики		
или		
б) Какова эффективность лечения при вторичной профилактике (если заболевание обнаружено), с точки зрения		
Идеальной эффективности		
Соблюдения пациентом назначений		
Большей эффективности раннего лечения, по сравнению с поздним.		

* См. соответствующий раздел в этой главе.

лечения при вторичной профилактике, если заболевание будет обнаружено при скрининге (например, лечения рака предстательной железы).

Тяжесть заболевания

Оправдан ли скрининг серьезностью изучаемого состояния с точки зрения смертности, заболеваемости и характера течения? Целесообразен скрининг только на состояния, угрожающие жизни или здоровью (см. главу 1, табл. 1.2). Серьезность заболевания определяется прежде всего его прогнозом или порождаемым риском (см. главы 5 и 6). Например, последствия бессимптомной бактериурии для здоровья неясны (кроме как при беременности или предстоящей урологической операции). Неизвестно, может ли она стать причиной почечной недостаточности или артериальной гипертензии. И тем не менее бактериологический анализ мочи часто проводится при периодических обследованиях.

Некоторые заболевания, причиняющие больному тяжкие мучения, встречаются достаточно редко (в определенных возрастных группах), чтобы стать поводом для скрининга. Два примера – рак молочной железы и рак толстой кишки. Хотя обе формы рака не исключены и у более молодых людей, в основном они встречаются у лиц старше 50 лет. У 20-летних женщин заболеваемость раком молочной железы составляет 1 на 100 000 ($\frac{1}{5}$ от заболеваемости у 70-летних мужчин) [3]. Таким образом, скрининг на рак молочной железы оправдан для женщин старше 50 лет, но не для 20-летних женщин (или 70-летних мужчин). При проведении скрининга для выявления очень редких заболеваний число людей, получивших от него пользу, весьма мало, а пострадавших из-за эффекта ярлыка – велико (см. далее).

Особенно сложна ситуация, когда сведения об эффективности раннего лечения отсутствуют. Как поступить? Например, имеются данные, что при синдроме Барретта (замещение плоского эпителия цилиндрическим в дистальном отделе пищевода) в 30–40 раз повышен риск развития рака пищевода [4]. Однако неизвестно, какова эффективность раннего лечения рака, выявленного при скрининге таких больных с помощью периодического эндоскопического исследования.

Простого ответа в такой ситуации не существует. Врачам следует помнить, что скрининг бесполезен, если отсутствует ранняя эффективная терапия. Нужно тщательно взвесить все данные по поводу лечения конкретного пациента. Если эти данные свидетельствуют о неэффективности лечения, то скрининг скорее повредит, чем поможет больному.

Требования, предъявляемые к тестам

Для любых скрининговых тестов, будь то опрос, физикальное или лабораторное исследование, применимы следующие критерии оценки качества.

Чувствительность и специфичность

При выборе теста для поиска болезни среди лиц с отсутствием симптомов следует учитывать очень низкую распространенность заболевания даже среди групп высокого риска, отобранных по возрасту, полу и другим характеристикам. Следовательно, хороший скрининговый тест должен обладать высокой чувствительностью – чтобы не пропускать те немногие случаи болезни, которые присутствуют, а также иметь высокую специфичность – чтобы снизить число лиц с ложноположительными результатами.

Чувствительность и специфичность для скринингового теста устанавливаются так же, как и для диагностического, за исключением того, что "золотым стандартом" служит не другой тест, а результаты проспективного наблюдения. Например, чувствительность скринингового теста на скрытую кровь для выявления рака толстой кишки определяется отношением числа случаев рака толстой кишки, диагностируемых во время скрининга, к сумме этого числа и числа случаев рака, впоследствии обнаруженных в течение года среди лиц с отрицательными результатами теста (исходя из предположения, что эти случаи рака присутствовали во время скрининга, но были пропущены, т.е. результаты теста были ложноотрицательными) [5]. Подобный способ определения чувствительности и специфичности для скринингового теста иногда называют *методом выявляемости* (detection method).

Этот метод хорошо работает при определении чувствительности многих скрининговых тестов, но для оценки скрининга на некоторые виды новообразований его использование затруднено. Применение данного метода возможно при следующих условиях. Во-первых, необходимо, чтобы был известен период, в течение которого должно проводиться наблюдение. Как правило, этот период неизвестен и его продолжительность устанавливается наугад. Во-вторых, необходимо, чтобы у всех лиц с положительными результатами скринингового теста обязательно развилось бы заболевание, если они не будут лечиться. Этот второй аспект составляет проблему при скрининге на рак предстательной железы. Поскольку гистологически выявляемый рак предстательной железы очень распространен (у 25 % мужчин в возрасте 50 лет и практически у всех мужчин в возрасте 90 лет), скрининговые тесты обнаружат этот вид рака у многих мужчин, но у большинства из них ни-

когда не будет клинических проявлений злокачественного роста. Таким образом, если методом выявляемости оценить чувствительность такого теста, как определение простатспецифического антигена (ПСА), то тест выглядит весьма ценным, поскольку числитель содержит все обнаруженные случаи, а не только те, которые в последующем проявят тенденцию к злокачественному росту.

Справиться с этими проблемами позволяет *метод заболеваемости* (incidence method), с помощью которого рассчитывают чувствительность, используя величину заболеваемости среди не подвергавшихся скринингу, а также частоту возникновения рака за определенный период времени среди подвергавшихся скринингу. Такой подход основан на том, что чувствительность теста должна влиять на частоту выявления заболевания за определенный отрезок времени, а не на заболеваемость. Для рака предстательной железы метод заболеваемости определяет чувствительность теста как единица минус отношение частоты выявления рака предстательной железы за определенный период времени в группе мужчин, периодически подвергающихся скринингу, к заболеваемости раком этой же локализации в группе лиц, не подвергавшихся скринингу (контрольная группа). Метод заболеваемости при вычислении чувствительности диагностического теста обходит проблему учета "доброкачественного" рака предстательной железы, но может снизить чувствительность, поскольку исключает случаи рака с длительным периодом развития. Следовательно, истинная чувствительность теста скорее всего находится между оценками, полученными методами выявляемости и заболеваемости.

Из-за низкой распространенности большинства заболеваний прогностическая ценность положительного результата для основной массы скрининговых тестов (даже для тестов с высокой специфичностью) низка. Врачи, осуществляющие профилактические мероприятия с помощью скрининговых тестов, должны смириться с тем, что им приходится затрачивать массу усилий впустую, проводя исследования у лиц, которые никогда не заболеют. Однако ошибку можно минимизировать, если применять скрининг только среди лиц с более высокой вероятностью заболевания.

Пример. Заболеваемость раком молочной железы увеличивается с возрастом: от 1 на 100 000/год в возрасте 20 лет до 1 на 200/год для лиц старше 70 лет. Следовательно, уплотнение в молочной железе, обнаруженное во время скрининга у молодой женщины, будет доброкачественным с большей вероятностью, чем у пожилой. В большом исследовании по скринингу на рак молочной железы [6] результаты биопсии объемного образования молочной железы заметно варьировали в разных возрастных группах: отношение числа случаев доброкачественных опухолей к числу злокачественных соста-

вило более чем 16 к 1 у женщин в возрасте до 40 лет, но было менее чем 3 к 1 у женщин старше 70 лет (рис. 8.2). Показатели чувствительности и специфичности для клинического обследования молочной железы и для маммографии тоже оказались выше в группе старшего возраста.

Выявляемость заболеваний при скрининге снижается, если он применяется повторно (рис. 8.3). В самом деле, при первом скрининге (*скрининг распространенности*, prevalence screen) выявляются случаи заболевания, длительность которого у разных больных различна. При повторном скрининге большинство вновь выявленных случаев будут составлять заболевания, начавшиеся между первым и вторым скринингами и лишь небольшое число случаев составят те, что были пропущены во время первого скрининга. Следовательно, второй и последующие скрининги – это *скрининги заболеваемости* (incidence screen). На рис. 8.3 видно, как уменьшается число случаев заболевания в группе при повторных обследованиях после скрининга распространенности. Это означает, что предсказательная ценность положительного результата теста будет убывать при последующих обследованиях.

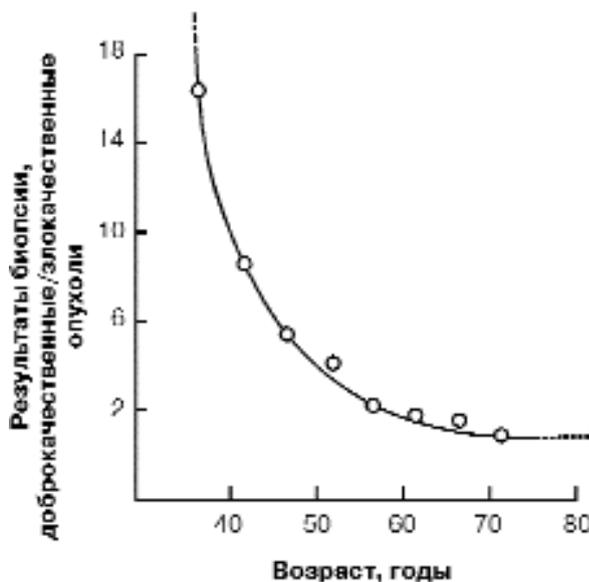


Рис. 8.2. Эффективность скрининга в зависимости от возраста пациентов. Отношение числа доброкачественных биопсий к числу злокачественных по результатам биопсии при скрининге женщин на рак молочной железы [Baker L.H. CA 1982;32: 195 – 231].

ний, требующихся для пациентов с положительным результатом теста. Таким образом, на стоимость теста влияют чувствительность, специфичность и предсказательная ценность положительного результата. Стоимость также зависит от того, требуется ли специальный визит к врачу. Скрининговые тесты, проводимые во время посещения пациентом врача по другим поводам (например, измерение артериального давления), значительно дешевле, чем тесты, требующие специальных визитов, освобождения от работы и дополнительных затрат на транспорт.

Безопасность

Для диагностических тестов, применяемых у больных с конкретными жалобами, некоторая степень риска разумна и этична. Если пациент обращается за помощью и симптомы неясны, врач не может медлить и делает все возможное для установления диагноза. Совсем другое дело, когда риску подвергают людей, считающих себя здоровыми. В таких случаях к безопасности процедуры предъявляются особые требования. Отчасти это обусловлено низкими шансами обнаружения заболевания у здоровых людей. Хотя колоноскопия не считается опасной процедурой при обследовании пациентов с патологией желудочно-кишечного тракта, она недостаточно безопасна в качестве метода скрининга, особенно с учетом потенциальной возможности прободения кишечника. Если частота прободения составляет 0,2% и колоноскопия используется для скрининга рака толстой кишки у женщин в возрасте 50–55 лет, то число случаев прободения составит около двух на каждый случай обнаружения рака; для женщин старше 70 лет отношение обратное, поскольку в этом возрасте колоректальный рак встречается гораздо чаще [7].

Приемлемость для больного и врача

Значение приемлемости теста иллюстрируется опытом обследований, направленных на выявление ранних форм рака шейки матки и толстой кишки. Женщины с наиболее высоким риском рака шейки матки реже всего проходят кольпоцитодиагностику по Папаниколау. Та же проблема характерна и для рака толстой кишки: жители Северной Америки, у которых отсутствуют симптомы заболевания, не желают периодически подвергаться обследованию, что, впрочем, вовсе неудивительно.

Из данных, приведенных в табл. 8.2, видна готовность разных групп людей к скринингу на рак толстой кишки. Лица, посещающие врача по собственной инициативе, проявляют высокую степень сотрудничества: они готовы собрать кал, сделать гваяковую

Таблица 8.2

Готовность пациентов к участию в скрининге: частота участия пациентов в проведении гваяковой пробы в разных группах*

Группа	Число участников, приславших пробу,%
Программа скрининга на рак толстой кишки	85
Программа скрининга на рак молочной железы	70
Лица, находящиеся под наблюдением страховой медицинской организации в возрасте 50 – 74 лет	27
Лица, находящиеся под наблюдением страховой медицинской организации в возрасте 50 – 74 лет, получившие набор для гваяковой пробы, инструкции и напоминания по почте и по телефону	48

* [Myers R.E. et al. Med Care 1991;29:1039 – 1050; Fletcher S.W., Dauphinee W.D. Clin Invest Med 1981;4:23 –31].

пробу (мазок) и отослать материал своему врачу для клинического анализа. Пациенты, которые не обращаются сами к врачу, менее склонны к сотрудничеству. Наименьшая готовность к участию в скрининге отмечается у пожилых людей, при том что риск развития рака толстой кишки в этой возрастной группе самый высокий. За счет дополнительных усилий можно привлечь большее число людей, хотя вряд ли это позволит охватить и половину популяции.

Приемлемость теста для врачей – критерий, на который, как правило, обращают внимание только те, кто осуществляет обследование. После одного большого и успешно проведенного исследования целесообразности скрининга врачи отказались от сигмоскопии, поскольку гастроэнтерологи, проводившие процедуру, нашли, что она слишком сложна, занимает много времени и не оправдывает полученных результатов [8]. При этом только 38% пациентов откликнулись на предложение пройти тест.

Эффект ярлыка

Эффект ярлыка (labeling) – это психологическое воздействие результатов теста или диагноза на пациента. Исследования показывают, насколько важна иногда бывает при тестировании роль психологического фактора.

Эффект ярлыка может либо помочь, либо повредить пациенту. Если пациент узнает, что результаты теста у него нормальные, то эффект положительный. Большинству врачей приходилось слышать реакцию: "Отлично, значит, я смогу поработать еще год". Получение "carte blanche" по поводу состояния здоровья способ-

ствует тому, что человек с охотой выполняет свои обычные обязанности.

С другой стороны, если человеку сказать, что результат теста сомнительный, это может оказать отрицательное психологическое воздействие. Исследование группы женщин с ложноположительными результатами маммографии (женщины с сомнительным результатом маммографии, у которых при последующем обследовании рак не был обнаружен) показало, что спустя несколько месяцев 47% пациенток испытывали тревогу в связи с результатами маммографии и 41% боялись заболеть раком молочной железы, причем 17% сообщили, что это состояние мешало им выполнять ежедневные обязанности [9].

Роль эффекта ярлыка вызывает особую озабоченность в связи с прогрессом в области генетического скрининга. Идентификация гена хореи Гентингтона позволяет родственникам больного пройти тестирование на носительство этого гена. Такой тест может помочь людям, которые хотели бы знать, могут ли они вступать в брак и иметь детей. Куда сложнее оказываются ситуации (более частые), когда наличие гена обуславливает риск, но не обязательно развитие заболевания. Например, известно несколько генов, связанных с развитием рака толстой кишки и молочной железы. Далеко не у всех носителей данных генов возникнет рак, а многие из тех, кто не имеет их, все-таки заболеют. Поскольку эти события ожидаются только в будущем, лицам, которые узнали о наличии у них одного из упомянутых генов, предстоит долго жить в ожидании страшной болезни.

Этический аспект эффекта ярлыка вызывает особую тревогу, когда это касается лиц с ложноположительными результатами тестов. В подобных ситуациях проведение скрининга может вызвать у пациента ощущение уязвимости, т.е. причинит больше вреда, чем пользы.

Риск ложноположительного результата

Преыдушие рассуждения относятся в отдельности к каждому тесту, который может выполнять врач при периодическом обследовании больных. Однако большинство врачей используют при этом не один и не два теста. Одно из исследований показало, что практикующие врачи были убеждены, что при периодическом обследовании следует применять 57 различных тестов [10]. Современные технологии и, вероятно, опасение судебного преследования побуждают врачей к тому, чтобы "учесть любые варианты". Автоматизированные системы анализа крови позволяют назначать до нескольких десятков легко выполнимых тестов.

Таблица 8.3

Соотношение между числом назначаемых тестов и долей здоровых людей хотя бы с одним выходящим за пределы нормы результатом теста*

Число тестов	Число лиц хотя бы с одним выходящим за пределы нормы результатом, %
1	5
5	23
20	64
100	99,4

*[Sackett D.L. Clin Invest Med 1978;1:37 – 43].

Если результат скринингового теста выражен количественно (а чаще всего это именно так) и если за нормальный результат принимаются все значения показателя, укладывающиеся в 95% доверительный интервал (как это и бывает обычно), то чем больше тестов назначит врач, тем выше риск ложноположительного результата. Как видно из данных табл. 8.3, при достаточно большом количестве тестов "отклонения от нормы" будут обнаружены практически у всех здоровых людей.

Эффективность лечения

К вмешательствам по первичной профилактике относятся вакцинация (например, введение противостолбнячной сыворотки для профилактики столбняка), назначение лекарственных препаратов (например, аспирин для предупреждения инфаркта миокарда), консультирование по коррекции образа жизни (например, помощь в отказе от курения или рекомендации по употреблению пищи с низким содержанием холестерина). Каково бы ни было воздействие, оно должно быть эффективным как в идеальной, так и в реальной ситуации. Эффективность обычно лучше документируется в отношении лекарственных средств, нежели для советов по коррекции образа жизни. Федеральные законы США требуют наличия строгих научных данных об эффективности препаратов, разрешенных к использованию. Это не предусмотрено для методов психологической помощи по коррекции образа жизни, но клиницисты должны располагать строгими научными данными, прежде чем включать эти методы в качестве обычных профилактических мер. Образ жизни – один из наиболее важных факторов здоровья в современном обществе. Эффективное консультирование способствует сохранению здоровья, неэффективное – всего лишь трата времени и средств, к тому же оно может повредить.

Пример. В рамках программы борьбы с курением оценивались две стратегии психологической помощи – групповые занятия продолжительностью 1 ч, индивидуальные беседы по 10 – 20 мин, то и другое еженедельно в течение 8 нед; обе формы консультирования сочетались с заместительной никотиновой терапией [11,12]. По сравнению со случайным образом отобранной контрольной группой, психологическая помощь была эффективна: $\frac{1}{3}$ пациентов прекратили курить за 6 мес наблюдения. Однако среди пациентов, консультировавшихся индивидуально, от курения отказались менее 20%. Кроме того, авторы обнаружили, что для тех, кто время от времени продолжал курить в течение первых 2 нед программы, высока вероятность неудачи лечения через 6 мес. Эти наблюдения показывают, что консультирование лучше проводить "перед лицом товарищей по несчастью". Такого рода исследования, тщательно оценивающие результаты психологической помощи, выявляют, какой подход наиболее эффективен.

Меры вторичной профилактики обычно те же, что и при лечении. Аналогично мерам первичной профилактики они должны быть эффективны. Если не существует эффективного раннего лечения, то скрининг в отношении такого заболевания не оправдан, независимо от того, насколько легко оно обнаруживается. В этом случае раннее выявление не помогает пациенту, а всего лишь увеличивает время, в течение которого он знает о своей болезни.

Другой критерий, важный для оценки мер вторичной профилактики, определяет, улучшится ли исход, если болезнь обнаружена при скрининге на бессимптомной стадии, а не позднее, когда появились симптомы и пациент сам обратился к врачу. Если в обеих ситуациях исход одинаков, то скрининг не нужен.

Пример. Исследовалось влияние скрининга на исходы рака легкого [13]. Каждые 4 мес проводился скрининг у курящих мужчин, включая рентгеноскопию грудной клетки и цитологическое исследование мокроты. Выявленные и соответственно сразу получившие лечение больные не отличались по исходам от больных в группе, где скрининг не проводился: смертность от рака легкого составила (на 1000 человеко-лет наблюдения) 3,2 в группе, прошедшей скрининг, против 3,0 в группе, где скрининг не проводился. Таким образом, раннее обнаружение и лечение помогли больным раком легкого не больше, чем лечение, начатое при появлении симптомов.

Систематические ошибки All Caps

Как отмечалось в главе 7, наилучший метод оценки эффективности лечения – рандомизированное контролируемое испытание. Это справедливо для всех видов вмешательств, но особенно – для раннего лечения после скрининга. Для определения эффективности профилактических мероприятий обычно требуются го-

ды наблюдений за большим числом людей. Например, раннее лечение после скрининга на рак толстой кишки может уменьшить летальность при этой локализации рака приблизительно на $\frac{1}{3}$, но чтобы выявить этот эффект, потребовалось 13-летнее проспективное наблюдение [5]. Простого "клинического впечатления" о роли скрининга здесь недостаточно.

Тщательные исследования необходимы также из-за систематических ошибок. Ниже описано три специфических вида систематических ошибок в исследованиях эффективности программ скрининга.

Систематическая ошибка, обусловленная задержкой диагностики

Задержкой диагностики (lead time) называется период между моментом обнаружения заболевания при скрининге и моментом, когда диагноз был бы поставлен, если бы пациент обратился к врачу по поводу появившихся симптомов (рис. 8.4). Величина задержки диагностики для каждого заболевания зависит как от биологической скорости его прогрессирования, так и от способности теста к раннему обнаружению болезни. Если задержка диагностики очень мала, что соответствует современной ситуации для рака легкого, то лечение больных, выявленных при скрининге, не будет более эффективным, чем лечение, начатое после появления симптомов. С другой стороны, если задержка диагностики велика, как это бывает при раке шейки матки (прогрессирование карциномы *in situ* до клинического проявления инвазивного рака происходит в среднем в течение 30 лет), то лечение больных, выявленных при скрининге, может быть очень эффективным.

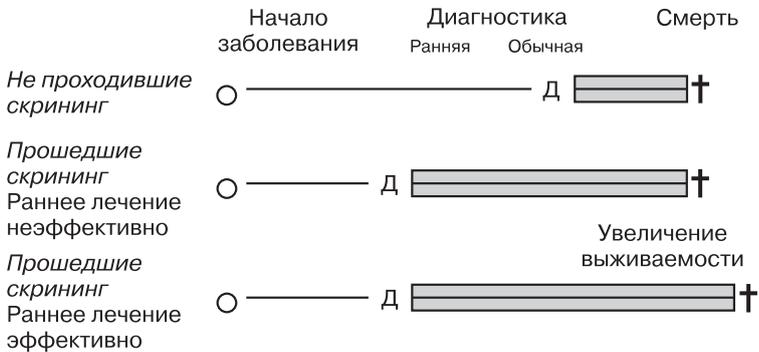


Рис. 8.4. Влияние сроков диагностики на выживаемость после скрининга: заштрихованные участки соответствуют продолжительности жизни после установления диагноза (Д).

Каким образом задержка диагностики может привести к появлению систематической ошибки в исследовании, посвященном эффективности раннего лечения?

Благодаря скринингу, болезнь обнаруживается раньше, чем это могло бы произойти после появления симптомов (см. рис. 8.4). В результате при смертельном заболевании средняя продолжительность выживания от момента постановки диагноза окажется в среднем больше у пациентов, выявленных при скрининге, чем у выявленных при начавшихся симптомах, даже если раннее лечение не более эффективно, чем лечение в стадии клинических проявлений. Такая ситуация создает ложное впечатление, что скрининг помогает пациентам прожить дольше, тогда как в действительности увеличилась не продолжительность жизни, а продолжительность болезни.

Для того чтобы избежать *систематической ошибки, обусловленной задержкой диагностики* (lead time bias), следует изучать как группу, прошедшую скрининг, так и контрольную группу и сравнивать возрастные показатели смертности, а не выживаемость от момента установления диагноза. Эффективность ранней диагностики и лечения рака толстой кишки можно считать доказанной, поскольку исследования показали, что среди подвергающихся скринингу летальность ниже, чем в сравниваемой контрольной группе [5].

Систематическая ошибка, обусловленная быстрым прогрессирующим болезнью

Систематическая ошибка, обусловленная быстрым прогрессирующим болезнью (length time bias) (рис. 8.5, 8.6) может возникнуть в скрининговых исследованиях вследствие того, что доля диагностированных при скрининге медленно прогрессирующих заболеваний больше, чем их доля в обычной клинической практике. За счет включения в исследование большего числа случаев с медленно растущим раком будет казаться, что скрининг и раннее лечение более эффективны, чем обычное лечение.

Систематическая ошибка, обусловленная быстрым прогрессирующим болезнью, возникает следующим образом. Скрининг эффективнее при медленном прогрессирующем болезнью. Однако скорость развития злокачественных новообразований варьирует в широких пределах. Некоторые формы прогрессируют медленно, другие – очень быстро. Скрининговые тесты будут в основном обнаруживать медленно растущие опухоли, поскольку в этих случаях появлению симптомов предшествует более длительный период. При быстром развитии опухоли, наоборот, больше вероят-

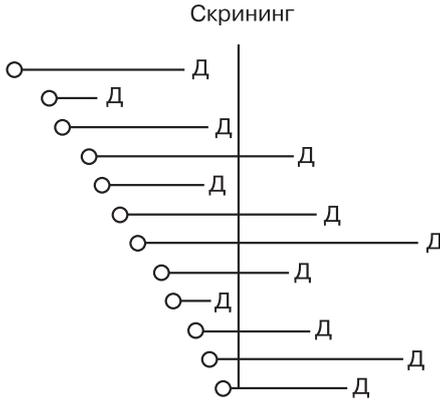


Рис. 8.5. Систематическая ошибка, обусловленная быстрым прогрессированием болезни. Быстро прогрессирующие болезни, при которых от возникновения (О) до появления симптомов и установления диагноза (Д) проходит мало времени, с меньшей вероятностью будут обнаружены при скрининге.

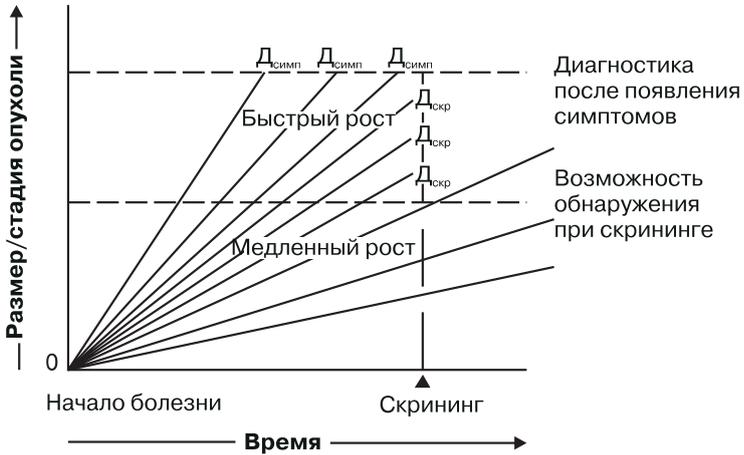


Рис. 8.6. Систематическая ошибка, обусловленная быстрым прогрессированием болезни. Быстро прогрессирующие злокачественные новообразования обратят на себя внимание еще до проведения скрининга, тогда как при более медленном прогрессировании есть время для обнаружения опухоли. $Д_{симп}$ – диагноз поставлен после появления симптомов, $Д_{скр}$ – диагноз поставлен в результате скрининга.

ность, что появятся симптомы и будет установлен диагноз в интервале между обследованиями. Следовательно, при скрининге существует тенденция к обнаружению опухолей, исходно имеющих более благоприятный прогноз. Вследствие этого летальность при злокачественных новообразованиях, выявляемых при скри-

нингe, может быть ниже, чем летальность в обычной практике, однако это не связано с эффективностью скрининга.

Систематическая ошибка, обусловленная несоблюдением назначений

Третий основной тип систематических ошибок в исследованиях эффективности профилактического лечения – это *систематическая ошибка, обусловленная несоблюдением назначений* (compliance bias). Она зависит от тщательности выполнения пациентами врачебных предписаний. У пациентов, точно соблюдающих схему лечения, наблюдается тенденция к лучшему прогнозу независимо от проведения скрининга. Если исходы в группе участвующих в программе скрининга добровольцев сравниваются с исходами в другой группе лиц, то более благоприятные результаты у добровольцев могут быть обусловлены не самим лечением, а другими факторами, связанными с выполнением назначений.

Пример. В исследовании эффективности программы сохранения здоровья экспериментальная группа пациентов приглашалась, а группа сравнения не приглашалась на ежегодные обследования [14]. Однако через какое-то время некоторые пациенты контрольной группы обратились с просьбой о проведении им периодических обследований. При этом среди пациентов контрольной группы, прошедших обследования по доброй воле, летальность была более низкой, чем у пациентов, которых специально приглашали на скрининг (рис. 8.7). Это обусловлено тем, что приглашаемая группа включала в себя не только пациентов, готовых к сотрудничеству, но также и тех, кто участвовал неохотно.

Для предотвращения систематических ошибок, обусловленных скоростью развития болезни или степенью соблюдения врачебных предписаний, следует проводить рандомизированные контролируемые испытания, поскольку они учитывают все исходы в группах независимо от метода диагностики или степени готовности пациентов к сотрудничеству. Популяции пациентов при случайном распределении на экспериментальную и контрольную группы должны в среднем содержать примерно одинаковое число больных с медленным и быстрым прогрессированием опухоли и иметь сравнимую степень готовности к сотрудничеству. Затем эти группы можно будет проследить во времени, оценивая показатели смертности, а не выживаемости, чтобы избежать систематической ошибки вследствие задержки диагностики.

Поскольку рандомизированные контролируемые испытания сложны, продолжительны и дороги, исследователи иногда пытаются использовать для изучения методов профилактики и эффек-

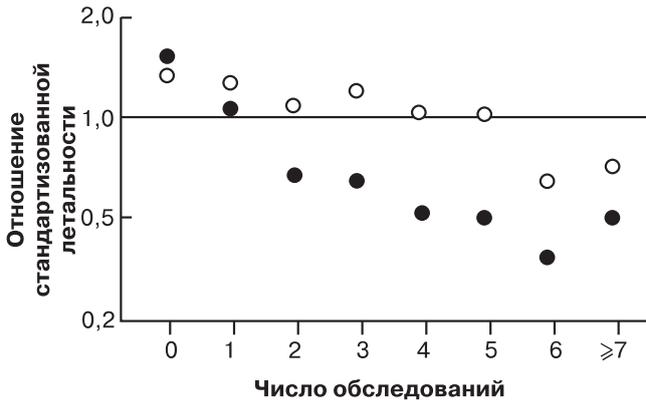


Рис. 8.7. Эффект готовности к сотрудничеству у пациентов в программах скрининга. Контрольная группа пациентов (●) имеет более низкое отношение стандартизованной летальности (отношение наблюдаемого числа смертей к ожидаемому, стандартизованное по возрасту), чем экспериментальная группа, прошедшая скрининг (○); однако контрольная группа состоит только из лиц, которые сами просили их обследовать, в то время как экспериментальная группа включает в себя всех лиц, которым скрининг был назначен [Friedman G.D. et al. J Chron Dis 1986;39:453 – 463].

тивности лечения после скрининга другие виды исследований – когортные (см. главу 5) или исследования случай–контроль (см. главу 10).

Пример. Для того чтобы установить, снижается ли благодаря периодическому скринингу летальность при раке толстой кишки, локализованного в пределах досягаемости сигмоскопии, J.V. Selby et al. [15] сопоставили частоту скрининговой сигмоскопии за предыдущие 10 лет у пациентов, умерших от рака толстой кишки, и среди выживших, подобранных с учетом возраста и пола. Чтобы устранить систематические ошибки, обусловленные задержкой диагностики и скоростью развития заболевания, исследовали влияние скрининга только у тех пациентов, о которых было достоверно известно, умерли они от рака толстой кишки (экспериментальная группа) или нет (контрольная группа). Для устранения систематической ошибки, обусловленной степенью готовности к сотрудничеству, авторы стандартизовали результаты по числу периодических обследований, в которых участвовал каждый пациент. Результаты стандартизовались также с учетом других факторов, которые могли бы способствовать как более частому скринингу, так и повышению вероятности развития рака толстой кишки. Оказалось, что пациенты, которые умерли от рака с локализацией в прямой кишке или дистальном отделе ободочной кишки, реже подвергались обследованию с помощью сигмоскопии за предыдущие 10 лет (8,8%), чем пациенты контрольной группы (24,2%); раннее лечение, начатое вслед за сигмоскопией, предотвратило почти 60% случаев смерти от рака дисталь-

ного отдела толстой кишки. Поскольку при локализации опухоли выше уровня досягаемости сигмоскопии летальность в группе больных, прошедших скрининг, была такой же, как в контроле, авторы сочли это еще одним доказательством эффективности скрининга.

Часто для оценки эффективности скрининговых программ применяется метод изучения серий случаев, когда группа участников скрининга прослеживается во времени. Однако этот метод в данном случае непригоден, поскольку получаемые результаты подвержены всем видам систематических ошибок, обсуждавшихся ранее (подробнее о методе изучения серий случаев см. в главе 10).

Вред и польза профилактики

Пропаганда здорового образа жизни и профилактика заболеваний приобретают все большую популярность. Разумеется, максимальное сохранение здоровья людей – достойная цель, но, как показано в этой главе, за ней стоят сложные теоретические построения. Следует учитывать, что мероприятия по поддержанию здоровья способны нанести вред, хотя и непреднамеренный. Профилактические меры стоят денег, отнимают у пациента время и причиняют неудобства. В худшем варианте у некоторых пациентов могут возникнуть серьезные осложнения, связанные либо с самим скрининговым тестом, либо с побочными эффектами последующих тестов и лечения – особенно у пациентов с ложноположительным результатом. Кроме того, ложноположительные результаты могут оказать отрицательное психологическое воздействие. Таким образом, врачи должны располагать надежными данными о пользе и вреде профилактических мероприятий. Одних добрых намерений здесь недостаточно.

Прежде чем проводить профилактическое мероприятие, особенно если оценка его специалистами неоднозначна, врач должен обсудить с пациентом все доводы "за" (вероятность известных и предполагаемых благоприятных эффектов для здоровья) и "против" (вероятность отрицательных последствий).

Пример. Хотя скрининг с использованием физикального исследования и маммографии по поводу рака молочной железы повсеместно рекомендован для женщин старшего возраста, существуют разногласия относительно его применения у женщин в возрасте 40 – 49 лет. Рандомизированные контролируемые испытания показали, что скрининг в этой возрастной группе малоэффективен, хотя профилактический эффект, равный примерно 15%, сохраняется в течение многих лет. Мнения разных групп экспертов неоднозначны. При обсуждении этой проблемы с пациенткой следует объяснить как пользу, так и вред скрининга (рис. 8.8). Такой подход не только наиболее честен по отношению к больной, но и помогает ей понять



Рис. 8.8. Оценка пользы и вреда скрининга. Результаты 10-летнего ежегодного проведения маммографии 1000 женщинам (исходный возраст 40 лет).

ситуацию, принять осознанное решение. Анализ эффективности затрат позволяет формализовать подход к внедрению профилактических программ администраторами системы здравоохранения [16].

Современные рекомендации

Благодаря научному прогрессу в области профилактической медицины, современные рекомендации о здоровом образе жизни уже не те, что были в прошлом. Некоторые специалисты советуют отказаться от рутинных ежегодных обследований, предпочитая избирательный подход, при котором применяемые тесты зависят

от возраста, пола и клинических характеристик пациента. Тем самым увеличиваются вероятность наличия заболевания и предсказательная ценность положительного результата теста. Существует также тенденция к использованию меньшего числа тестов, чем раньше (это уменьшает число ложноположительных результатов). Уделяется специальное внимание вопросу о том, поиск каких именно заболеваний следует проводить. Все более осознается необходимость четкого определения критериев, которым должны отвечать тесты, пригодные для проведения периодических обследований.

Резюме

Болезнь лучше всего предотвратить методами первичной профилактики. Эффективность подобных воздействий нужно оценивать столь же тщательно, как и всех других видов медицинских вмешательств.

Отрицательные последствия болезни можно предупредить путем проведения скрининга, позволяющего выявить заболевание на бессимптомной стадии, когда лечение наиболее эффективно (вторичная профилактика). Основания для проведения скрининга таковы: заболевание является достаточно тяжелым; существует хороший скрининговый тест; лечение в начальной стадии более эффективно, чем применяемое после появления симптомов. Скрининговые тесты должны быть достаточно чувствительны, чтобы обнаружить большинство случаев заболевания, достаточно специфичны, чтобы не давать слишком много ложноположительных результатов, недороги, безопасны и приемлемы с точки зрения как пациентов, так и врачей.

Исследования вторичной профилактики и эффективности раннего лечения подвержены систематическим ошибкам трех видов: из-за впечатления, что больные живут дольше, создаваемого более ранним обнаружением заболевания; из-за тенденции скрининга выявлять непропорционально большое число медленно прогрессирующих случаев вследствие более высокой их распространенности; из-за относительно большего числа случаев благоприятного прогноза, обусловленных участием в скрининге пациентов с высокой степенью сотрудничества. Выявленный положительный эффект вмешательства может быть связан с этими ошибками, а не с самими профилактическими мероприятиями.

В результате лишь ограниченное число мер первичной и вторичной профилактики может быть рекомендовано для включения в программы по здравоохранению. Профилактические программы должны проводиться с учетом пола, возраста и клинического состояния пациентов.

Литература

1. Schappert S.M. National ambulatory medical care survey: 1991 summary. Advance data from vital and health statistics. No. 230. Hyattsville, MD: National Center for Health Statistics, 1993.
2. Webster's ninth new collegiate dictionary. Springfield, MA: Merriam-Webster, 1991.
3. Ries L.A.G., Miller B.A., Hankey B.F., Kosary C.L., Hargis A., Edwards B.K., eds. SEER cancer statistics review, 1973–1991: tables and graphs. NIH Publication No. 94–789. Bethesda, MD: National Cancer Institute, 1994.
4. Cameron A.J., Ott B.J., Payne W.S. The incidence of adenocarcinoma in columnar-lined (Barrett's) esophagus. *N Engl J Med* 1985;313:857–859.
5. Mandel J.S., et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993;328:1365–1371.
6. Baker L.H. Breast Cancer Detection Demonstration Project: five-year summary report. *CA* 1982;2:195–31.
7. Eddy D.M. Screening for colorectal cancer. *Ann Intern Med* 1990; 113:373–384.
8. Dales L.G., Friedman G.D., Collen M.F. Evaluating periodic multiphasic health check-ups: a controlled trial. *J Chron Dis* 1979;32:385–404.
9. Lerman C., Trock B., Rimer B.K., Boyce A., Jepson C., Engstrom P.F. Psychological and behavioral implications of abnormal mammograms. *Ann Intern Med* 1991;114:657–661.
10. Romm F.J., Fletcher S.W., Hulka B.S. The periodic health examination: comparison of recommendations and internists' performance. *South Med J* 1981; 74:265–271.
11. Fiore M.C., Kenford S.L., Jorenby D.E., Wetter D.W., Smith S.S., Baker T.B. The nicotine patch: clinical effectiveness with different counseling treatments. *Chest* 1994;105:524–533.
12. Kenford S.L., Fiore M.C., Jorenby D.E., Smith S.S., Wetter D., Baker T.B. Predicting smoking cessation: who will quit with and without the nicotine patch. *JAMA* 1994; 271:589–594.
13. Fontana R.S., Sanderson D.R., Woolner L.B., Taylor W.F., Miller W., Muhm J.R. Lung cancer screening: the Mayo program. *J Occup Med* 1986;28:746–750.
14. Friedman G.D., Collen M.F., Fireman B.H. Multiphasic health checkup evaluation: a 16-year follow-up. *J Chron Dis* 1986;39:453–463.
15. Selby J.V., Friedman G.D., Quesenberry C.P., Weiss N.S. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992;326:653–657.
16. Eddy D.M., Hasselblad V., McGivney W., Hendee W. The value of mammography screening in women under age 50 years. *JAMA* 1988;259:1512–1519.

Рекомендуемая литература

- Eddy D.M. (ed). Common screening tests. Philadelphia, American College of Physicians, 1991.
- Goldbloom R.B., Lawrence R.S., eds. Preventing disease: beyond the rhetoric. New York: Springer-Verlag, 1990.
- Guyatt G.H., Sackett D.L., Cook D.J. Users' guides to the medical literature. II: How to use an article about therapy or prevention. A: Are the results of the study valid? *JAMA* 1993;270:2598–2601.
- Guyatt G.H., Sackett D.L., Cook D.J. Users' guides to the medical literature. II: How to use an article about therapy or prevention. B: What were the results and will they help me in caring for my patients? *JAMA* 1994;271:59–63.
- Hayward R.S., Steinberg E.P., Ford D.E., Roizen M.F., Roach K.W. Preventive care guidelines: 1991. *Ann Intern Med* 1991;114:758–783.
- Miller A.B., ed. Screening for cancer. Orlando, FL: Academic Press, 1985.
- Miller A.B., Chamberlain J., Day N.E., Hakama M., Prorok P.C., eds. Cancer screening. Cambridge, England: International Union Against Cancer, 1991.
- Russell L.B. Educated guesses: making policy about medical screening tests. Berkeley: University of California Press, 1994.
- U.S. Preventive Services Task Force. Guide to preventive services: an assessment of the effectiveness of 169 interventions. Baltimore: Williams & Wilkins, 1989.

9

СЛУЧАЙНОСТЬ

Любая попытка врача получить необходимую ему информацию на основе клинического опыта (будь то специальное исследование или повседневная практика) осложняется двумя явлениями: систематической ошибкой и случайностью.

Как отмечалось в главе 1, систематическая ошибка (называемая также смещением) – это результат какого-либо процесса, обуславливающего систематические отличия полученных при наблюдении данных от истинных значений. В клинических исследованиях приходится прилагать значительные усилия, чтобы по возможности избежать систематической ошибки, а если она неизбежна, то контролировать и оценивать ее влияние на результаты.

С другой стороны, случайная ошибка присуща любому наблюдению. Ее можно минимизировать, но никогда нельзя избежать полностью. Случайные вариации параметров могут быть заложены в самом процессе измерения или в природе изучаемого биологического феномена (см. главу 2). Эта ошибка называется случайной, поскольку в среднем вероятность отклонения как в одну, так и в другую сторону от истинной величины одинакова.

Большинство из нас при интерпретации данных склонны переоценивать значение случайной ошибки и недооценивать значение систематической ошибки. Часто можно услышать: "Если $p < 0,001$, то небольшая систематическая ошибка не причинит значительного вреда". Однако если при сборе данных имела место неизвестная систематическая ошибка, то никакая статистическая обработка положения не исправит. Один исследователь, придерживающийся, очевидно, крайней точки зрения, заметил: "Результаты хорошо спланированного и тщательно выполненного исследования, как правило, понятны и без формального анализа, а при существенных недостатках в планировании или выполнении исследования не поможет никакая статистика [1]".

В этой главе случайность обсуждается в контексте контролируемых клинических испытаний, так как это более простой путь изложения материала. Однако применение обсуждаемых понятий не ограничивается сравнением способов лечения в клинических исследованиях. Статистические методы используются всегда, ког-

да выводы, касающиеся популяции, делаются на основе информации, полученной на выборке.

Случайная ошибка

В клиническом исследовании различия, наблюдаемые между экспериментальной и контрольной группами, не отражают истинных различий, поскольку в обеих сравниваемых группах происходят случайные отклонения. Статистические критерии помогают оценить степень приближения наблюдаемого различия между группами к истинному. Но почему бы не измерить явление точно, покончив с неопределенностью? Дело в том, что исследование обычно выполняется на выборке пациентов, а не на всех, у кого имеется изучаемое состояние. В итоге всегда возникает вероятность того, что выборка больных в исследовании, даже отобранная без систематической ошибки, может не соответствовать всей популяции таких больных.

Для оценки роли случайности в клинических наблюдениях используются два общих подхода. Первый, называемый *проверкой гипотез*, ставит вопрос о том, присутствует ли обнаруженный эффект (различие) в действительности. Для ответа на этот вопрос применяются статистические методы проверки гипотезы об отсутствии различия, или "нулевой гипотезы". Этот традиционный способ оценки роли случайности ассоциируется с известным показателем "*p*". Метод пользуется популярностью с тех пор, как статистические методы были разработаны в начале XX века. При другом подходе, называемом *методом доверительных интервалов*, с помощью статистических методов определяется диапазон значений, который с определенной вероятностью включает в себя истинную величину. Этот подход стал широко применяться недавно и сейчас предпочитается большинством журналов по причинам, рассмотренным ниже. Начнем с описания традиционного подхода.

Проверка гипотез

В обычной ситуации, когда главные выводы исследования выражены в дихотомической форме (например, лечение рассматривается либо как успешное, либо нет), результаты статистической проверки тоже имеют дихотомический характер (результат либо статистически значим – т.е. неслучаен, либо статистически незначим). Существуют четыре варианта соотношений между выводами исследования и действительностью (рис. 9.1).

Из четырех вариантов заключений статистического теста возможны два правильных: а) эффекты разных методов лечения дей-

		Истинное различие	
		Присутствует	Отсутствует
Результат статистического теста	Различие значимо	Верно	Ошибка I рода (α)
	Различие незначимо	Ошибка II рода (β)	Верно

Рис.9.1. Соотношение между заключением статистического теста и истинным различием двух групп больных, леченных разными методами. ("Отсутствует" – это фактически упрощение, которое означает, что истинное различие не превышает определенную величину.)

ствительно различны и именно таков вывод исследования; б) методы лечения не различаются по своему эффекту, что и соответствует выводу исследования.

Два других варианта ошибочны. Это бывает в случае, если изучаемые методы лечения в действительности равноэффективны, но сделан вывод, что один из методов лечения лучше. Такого рода ошибка, приводящая к ложноположительному заключению об эффективности изучаемого метода лечения, называется *альфа-ошибкой* (α -error), или *ошибкой I рода* (Type I error). Альфа-ошибка – это вероятность того, что одно лечение будет считаться лучше другого, когда в действительности это различие отсутствует. С другой стороны, лечение может быть эффективным, хотя исследование привело к выводу об отсутствии эффекта. Такое ложноотрицательное заключение называется *бета-ошибкой* (β -error), или *ошибкой II рода* (Type II error). Бета-ошибка – это вероятность того, что оба лечения будут считаться равноэффективными, когда в действительности одно лечение лучше, чем другое. Вывод об "отсутствии различия" в упрощенном виде указывает на то, что истинное различие не превышает определенную величину, поскольку невозможно точно установить, что между двумя методами лечения вовсе нет различий.

На рис. 9.1 представлена четырехпольная таблица, аналогичная той, в которой результаты диагностических тестов сравниваются с истинным диагнозом (см. главу 3). Здесь "тест" – это заключение клинического испытания, основанное на статистической оценке результатов исследования, проведенного на выборке.

Истинные различия между двумя методами лечения устанавливаются путем наблюдения всех пациентов с изучаемой болезнью или большого числа выборок, состоящих из таких больных, если бы это можно было сделать на практике. Альфа-ошибка аналогична ложноположительным, а бета-ошибка – ложноотрицательным результатам теста. В отсутствие систематической ошибки неопределенность статистического заключения обусловлена случайной вариацией.

Случайная вариация присуща любым наблюдениям, поэтому ставить вопрос о том, случайны или нет результаты исследования, было бы чрезмерным упрощением. Скорее нужно искать ответ на вопрос, какова вероятность того, что результаты, полученные при конкретных условиях исследования, объясняются случайной вариацией. Вероятность ошибки, обусловленной случайной вариацией, оценивается методами статистики. Эти методы позволяют рассчитать вероятность того, что конкретные результаты получены исключительно за счет случайности.

Статистика – это специальная область математики со своей, неизвестной многим врачам, терминологией: "нулевая гипотеза", "дисперсия", "регрессия", "мощность", "моделирование". Для неспециалистов в этой области статистическая оценка – лишь средство подведения итога исследования, оценки влияния случайной вариации.

В следующих двух разделах обсуждаются альфа- и бета-ошибки соответственно. Мы постараемся продемонстрировать, как метод проверки гипотез используется для оценки вероятности этих ошибок в медицине. Однако мы не будем пытаться оперировать строгими количественными категориями – для этого читателю следует ознакомиться с существующими пособиями по биостатистике (см. "Рекомендуемая литература" в конце этой главы).

Вывод о том, что лечение эффективно

В современной медицинской литературе статистические показатели в основном рассчитываются для определения вероятности альфа-ошибки, выражаемой в виде всем знакомой величины p . Это количественная оценка того, с какой вероятностью выявленные в конкретном исследовании различия между лечебными вмешательствами обусловлены случайностью, исходя из предположения, что в действительности различий между сравниваемыми группами нет. Смысл величины p можно проиллюстрировать следующим образом. Допустим, что различие между двумя способами лечения в действительности отсутствует, но в клиническом испытании выявлено, что одно лечение более эффективно, чем дру-

гое. Если провести множество таких исследований, то какая доля испытаний покажет такую же или большую эффективность одного лечения по сравнению с другим?

Такую величину p мы будем называть p_{ω} , чтобы отличать ее от оценки другого рода случайной ошибки, обозначенной нами как p_{β} . Встречающийся в научной литературе показатель p соответствует, как правило, тому, что мы называем p_{α} .

Ошибка, оцениваемая при помощи p_{α} , существует всегда, когда делается заключение о более высокой эффективности одного метода лечения по сравнению с другим. Вывод о том, что между способами вмешательства нет различий, поскольку p_{α} превышает некоторый уровень – некорректен; в этой ситуации следует применять показатель p_{β} (вероятность бета-ошибки).

Дихотомическая и точная оценка величины p

Стало привычным придавать особое значение тем случаям, когда $p < 0,05$, поскольку риск ошибочного заключения менее одного шанса из 20 считается очень низким. Действительно, трудно себе представить, что событие с вероятностью $< 0,05$ могло возникнуть только по чистой случайности, так как имеется один шанс из 20, что оно могло произойти случайно.

О различиях, при которых $p_{\alpha} < 0,05$ говорят как о статистически значимых. Однако важно помнить, что точка разделения устанавливается на уровне 0,05 произвольно. Специалист может принять более высокие или более низкие оценки в зависимости от последствий ложноположительного заключения в конкретной ситуации.

Для примирения различных точек зрения по поводу того, какая величина p достаточна, а какая – нет, некоторые исследователи приводят точные значения вероятностей p_{α} (например, 0,03; 0,07; 0,11 и т.д.) вместо разделения их на две категории $< 0,05$ или $> 0,05$. Тогда уже сам читатель должен судить, что считать статистически значимым, а что – нет. Однако если величина p больше, чем 1 из 5, то обычно пишут просто $p > 0,20$, так как практически все согласятся, что вероятность альфа-ошибки более чем 1 из 5 неприемлемо высока. Аналогичным образом при очень малых значениях p (таких как $p < 0,001$) вероятность того, что наблюдаемое различие объясняется случайностью чрезвычайно низка, поэтому дальнейшее уточнение шанса этого случайного события не прибавит информации.

Статистическая и клиническая значимость

Статистически значимое различие, независимо от того, насколько мала величина p , не означает, что оно клинически значи-

мо. Если величина $p < 0,0001$ получена в хорошо спланированном исследовании, то она с высокой степенью достоверности свидетельствует о том, что различие действительно существует. Но на основании величины p нельзя судить ни о степени этого различия, ни о его клинической важности. Высокую статистическую значимость могут иметь совсем небольшие различия – при условии, что исследовано достаточно большое число больных.

Пример. В начале 90-х годов горячо обсуждался вопрос о том, какой тромболитический препарат эффективнее при остром инфаркте миокарда – стрептокиназа или тканевый активатор плазминогена (tPA). Крупномасштабные исследования выявили различия в частоте реперфузии, но не летальности. В большом рандомизированном контролируемом испытании GUSTO [2], в которое был включен 41 021 пациент из 15 стран, tPA применялся в более интенсивном режиме, чем в предшествующих исследованиях. Через 30 дней после инфаркта миокарда среди получавших tPA наблюдалась более низкая летальность (6,3%), чем среди получавших стрептокиназу (7,2 или 7,4% в зависимости от способа введения гепарина, который назначали в качестве дополнительной терапии), причем вероятность того, что это различие случайно, была мала ($p < 0,001$). Тем не менее различие в эффективности препаратов было невелико; для предотвращения одного случая смерти около 100 больных пришлось бы лечить препаратом tPA, а не стрептокиназой. Препараты tPA гораздо дороже стрептокиназы, так что предотвращение этой одной смерти стоило бы почти 250 000 долларов [3]. К тому же tPA повышает риск развития геморрагического инсульта. Поэтому некоторые оспаривают целесообразность применения tPA при таком небольшом преимуществе в его эффективности, т.е. сомневаются в "клинической значимости" различий летальности при использовании tPA и стрептокиназы.

С другой стороны, менее впечатляющие значения p могут быть получены в исследованиях со значительным эффектом лечения, если они выполнены на малом числе больных (см. следующий раздел).

Статистические критерии

Обычные статистические критерии, знакомые многим читателям, используются для оценки вероятности альфа-ошибки. Статистический критерий применяется по отношению к данным для получения статистики критерия (т.е. статистики, определяющей критическую область критерия для проверки гипотез), которая в свою очередь используется для оценки вероятности ошибки (рис. 9.2). Критерий обеспечивает проверку нулевой гипотезы, т.е. предположения об отсутствии реального различия в исходах меж-

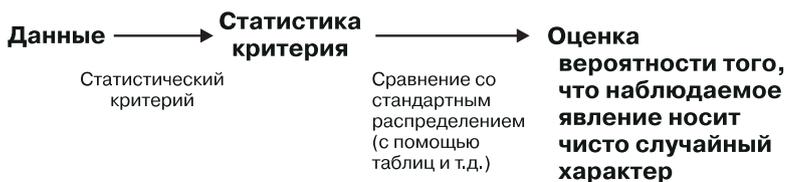


Рис. 9.2. Статистическая оценка.

Таблица 9.1
Некоторые статистические методы, широко используемые в клинических исследованиях

Метод	Область применения
<i>Проверка статистической значимости различий</i>	
Хи-квадрат (χ^2)	Между относительными частотами двух или большего числа событий (когда имеется большое число наблюдений)
Точный критерий Фишера	Между относительными частотами двух событий (когда имеется малое число наблюдений)
Критерий Манна–Уитни (U)	Между двумя медианами
Критерий Стьюдента (t)	Между двумя средними
Критерий Фишера (F)	Между двумя и более средними
<i>Описание степени связи</i>	
Коэффициент регрессии	Между независимой (прогностический фактор) и зависимой (исход) переменными
Критерий Пирсона (r)	Между двумя переменными
<i>Моделирование влияния множества переменных</i>	
Логистическая регрессия	При дихотомических исходах
Модель пропорционального риска Кокса	Если исходом считается время наступления события

ду двумя группами больных, леченных разными методами. Это допущение – чисто математическое, оно не означает, что предположение об отсутствии различий составляет рабочую гипотезу исследования. Проверка заканчивается либо отклонением нулевой гипотезы (заключением о наличии различия), либо ее принятием (заключением об отсутствии различия).

Некоторые общепринятые статистические методы представлены в табл. 9.1. Достоверность каждого критерия зависит от оп-

ределенных допущений относительно данных. Если данные не удовлетворяют этим допущениям, то вычисленная величина p_α может ввести в заблуждение. Описание этих статистических критериев (которые также называются статистическими тестами) можно найти в любом пособии по биостатистике.

Пример. Назначение критерия χ^2 для качественных данных гораздо легче понять, чем большинство других, поэтому он может быть использован для иллюстрации применения статистической оценки. Рассмотрим данные рандомизированного испытания двух схем введения гепарина: по номограмме, определяющей дозу препарата в зависимости от массы тела, и по стандартной схеме лечения [4]. В качестве исхода был выбран такой показатель, как превышение терапевтического уровня активированного частичного тромбопластинового времени (АЧТВ) в пределах 24 ч от начала лечения антикоагулянтом. В группе больных, которым вводили гепарин по номограмме, указанный результат был у 60 из 62 (97%) больных, в группе больных, получавших гепарин по стандартной схеме, такой результат был у 37 из 48 (77%) больных (табл. 9.2).

Какова вероятность того, что в повторных исследованиях с той же численностью наблюдений разница приведенных выше показателей будет такой же или большей, если в действительности сравниваемые методы лечения не различаются по эффективности? Это зависит от того, насколько наблюдаемые результаты отличаются от тех, которых были бы получены в случае одинаковой эффективности двух методов лечения. Если методы лечения не различаются по эффективности, то ожидаемое число успешно леченных больных (табл. 9.3) в каждой из двух сравниваемых групп можно получить с помощью расчетного показателя частоты успешного лечения среди общего числа больных (88%).

Критерий χ^2 дает количественную оценку расхождения между наблюдаемым и ожидаемым числом больных и представляет собой сумму четырех выражений, рассчитываемых по приведенным выше таблицам:

$$\frac{(\text{Наблюдаемое число больных} - \text{Ожидаемое число больных})^2}{\text{Ожидаемое число больных}}$$

Величина критерия χ^2 определяется тем, насколько все наблюдаемые числа отличаются от ожидаемых при условии отсутствия различий в эффективности сравниваемых схем лечения. Поскольку разности возводятся в квадрат, то не важно, выше или ниже наблюдаемые значения по сравнению с ожидаемыми. Деля квадрат разности на ожидаемое значение, получаем величину, стандартизованную по ожидаемому числу больных.

Таблица 9.2
Наблюдаемые частоты

	АЧТВ превышает терапевтический уровень		
	да	нет	всего
Определение дозы по номограмме	60	2	62
Стандартная схема лечения	37	11	48
Всего	97	13	110

Таблица 9.3
Ожидаемые частоты (округленные до ближайшего целого)

	АЧТВ превышает терапевтический уровень		
	да	нет	всего
Определение дозы по номограмме	55	7	62
Стандартная схема лечения	42	6	48
Всего	97	13	110

Для представленных данных χ^2 равен:

$$\frac{(60 - 55)^2}{55} + \frac{(2 - 7)^2}{7} + \frac{(37 - 42)^2}{42} + \frac{(11 - 6)^2}{6} = 8,79.$$

Полученное значение χ^2 затем сравнивается с табличными величинами этого критерия, соответствующим определенным вероятностям (таблицы имеются в книгах и компьютерных программах), чтобы найти вероятность такого же или большего значения χ^2 . Очевидно, чем больше значение χ^2 , тем меньше вероятность того, что наблюдаемое различие имеет случайный характер. В приведенном случае $p=0,004$, т.е. такова вероятность ложноположительного заключения о различии эффективности двух методов лечения.

Вывод о том, что лечение неэффективно

В некоторых исследованиях делается вывод о том, что один метод лечения не более эффективен, чем другой. Примером могут служить исследования, показывающие, что прием антиоксидантов не предотвращает развитие рака и что введение антител к эн-

дотоксину не улучшает прогноз у большинства больных при септическом шоке.

Возникает вопрос: могли ли подобные результаты получиться чисто случайно, например, вследствие того, что эти исследования были выполнены в нетипичных условиях? Какова вероятность ложноотрицательного результата (бета-ошибки, или ошибки II рода)? Риск ложноотрицательного результата особенно высок в исследованиях с относительно небольшим числом больных.

Бета-ошибке уделяется меньше внимания, чем альфа-ошибке, по нескольким причинам. Во-первых, ее труднее рассчитать. Во-вторых, большинство из нас просто предпочитают положительный результат. Отрицательные результаты воспринимаются как нежелательные, авторы менее охотно их публикуют в журналах. Если же эти исследования все-таки публикуются, то авторы предпочитают делать акцент на подгруппах больных, для которых найдено различие в лечебном эффекте, даже если это различие не достигает уровня статистической значимости. Иногда авторы стараются объяснить отсутствие различия между эффектами какими-то особыми причинами, но только не случайностью. Однако главный вопрос, который нужно поставить, если результаты указывают на отсутствие различий в эффективности методов – какова вероятность бета-ошибки.

Вероятность того, что в исследовании будет найдено статистически значимое различие, когда это различие действительно существует, называется *статистической мощностью исследования*.

$$\text{Статистическая мощность} = 1 - p_{\beta}.$$

Мощность и p_{β} – дополняющие друг друга способы выражения одного и того же понятия. Мощность исследования аналогична чувствительности диагностического теста.

Об исследовании говорят как о статистически мощном, если оно имеет высокую вероятность выявления различий между методами лечения, когда эти различия действительно существуют.

Сколько больных достаточно для испытания?

Допустим, Вы читаете сообщение о клиническом испытании, в котором многообещающий новый метод лечения сравнивается со старым. Вы знаете, что источником наблюдаемых различий может стать случайная вариация, и хотите знать, достаточным ли было число больных (*размер выборки, sample size*) в этом исследовании, чтобы исключить случайность в качестве объяснения полученного результата. Сколько больных необходимо включить в

исследование, чтобы обеспечить адекватное сравнение эффективности двух методов лечения? Ответ зависит от четырех характеристик исследования: величины различия в частоте исходов между группами, p_a , p_b и природы данных. Эти характеристики должен учитывать и исследователь, планирующий эксперимент, и читатель, решающий, следует ли доверять публикации.

Величина эффекта

Размер выборки зависит от того, какова ожидаемая величина различий, которые предстоит выявить. В принципе можно искать различия любой величины и, конечно, исследователь надеется, что сможет обнаружить даже самые небольшие различия. Однако при прочих равных условиях для выявления малых различий требуется большее число пациентов. Поэтому лучше ставить вопрос таким образом: какое число больных достаточно, чтобы выявить наименьший клинически значимый эффект? Если нас интересуют только очень большие различия между экспериментальной и контрольной группами больных (т.е. очень сильный лечебный эффект), то допустимо меньшее число пациентов.

Альфа-ошибка

Размер выборки зависит также от риска альфа-ошибки (вывода об эффективности лечения, которое на самом деле неэффективно). Приемлемая величина такого риска выбирается произвольно – от 1 до 0. Если исследователь готов к последствиям высокой вероятности ложного вывода об эффективности метода, то он может взять небольшое число пациентов. Если же он стремится сделать риск ошибочного вывода достаточно малым, то требуется увеличить число больных. Как обсуждалось выше, обычно p_a устанавливается на уровне 0,05 (1 из 20), а иногда 0,01 (1 из 100).

Бета-ошибка

Другой фактор, определяющий размер выборки, – это выбранный риск бета-ошибки, который тоже произволен. Вероятность бета-ошибки часто устанавливается на уровне 0,20, т.е. допускается 20% вероятность не выявить существующие в действительности различия. общепринятые допустимые величины бета-ошибок гораздо больше, чем альфа-ошибок, т.е. мы относимся более требовательно к утверждениям об эффективности лечения. Если говорят, что лечение эффективно, оно должно быть эффективным в действительности.

Природа данных

Статистическая мощность исследования определяется еще и природой данных. Когда исходы выражены качественными признаками и описываются частотой событий, статистическая мощность исследования зависит от этой частоты. Чем больше число событий, тем выше статистическая мощность исследования для данного числа испытуемых. R. Peto et al. [5] установили, что в клинических испытаниях с оценкой времени до наступления смерти (или до наступления других событий – рецидива, метастазов, первых эпизодов тромбоза, инсульта, выздоровления либо смерти от конкретной причины) возможность выявления различий в эффективности двух методов лечения зависит в большей степени от числа умерших больных (или имевших соответствующий исход), чем от числа больных, включенных в исследование. Исследование 100 больных, 50 из которых умерли, имеет примерно такую же чувствительность (мощность), что и исследование 1000 больных, из которых умерли тоже 50 пациентов.

Если исход выражается непрерывной количественной переменной (например, артериальное давление или уровень холестерина в сыворотке), то мощность исследования определяется степенью различий пациентов внутри каждой группы. Чем больше различия между пациентами по изучаемым характеристикам, тем меньше уверенности в том, что наблюдаемая разница (или ее отсутствие) между группами обусловлена истинными различиями в эффективности методов лечения. Другими словами, чем больше различия между пациентами внутри групп, тем ниже статистическая мощность исследования.

При планировании исследования автор выбирает такие величины клинической значимости лечебного эффекта, α - и β -ошибок, которые считает приемлемыми. Он может спланировать исследование таким образом, чтобы сделать его мощность максимальной для данного размера выборки, например путем отбора больных с высокой вероятностью развития исходов или с одинаковыми характеристиками (разумеется, в пределах поставленной задачи). Однако, получив данные и имея конкретную научную задачу, исследователь уже не может повлиять на статистическую мощность исследования, поскольку она определяется характеристиками полученных данных.

Взаимосвязь характеристик исследования

Обсуждавшиеся выше взаимоотношения четырех характеристик исследования представлены в табл. 9.4.

Эти взаимоотношения носят характер взаимного compro-

мисса. В принципе для любого числа включенных в исследование пациентов существует определенный баланс между α - и β -ошибками: при прочих равных условиях, чем больше допускаемая величина ошибки одного рода, тем меньше должен быть риск ошибки другого рода. При этом по сути своей ни одна из них не "хуже" другой. Последствия принятия ошибочной информации за истинную зависят от клинической ситуации. Если имеется острая необходимость в более эффективном методе лечения (например, болезнь очень опасна и нет эффективного альтернативного метода лечения) и предлагаемое лечение не опасно, то разумнее предпочесть относительно высокий риск вывода о том, что вмешательство эффективно, когда в действительности это не так (большая α -ошибка), минимизируя вероятность отвергнуть эффективный метод (β -ошибка мала). С другой стороны, если болезнь менее серьезна и существуют альтернативные методы лечения либо новый метод лечения более дорог или опасен, следует минимизировать риск применения нового вмешательства, которое может быть неэффективным (α -ошибка мала), даже за счет относительно высокой вероятности упустить действительно эффективное лечение (большая β -ошибка). Конечно, можно уменьшить обе ошибки – если число исследуемых больных велико, частота исходов высока, изучаемый показатель внутри групп варьирует мало, а предполагаемый лечебный эффект значителен.

Соотношение между величиной лечебного эффекта и необходимым числом пациентов для общепринятых уровней p_α и p_β иллюстрируют два приведенных ниже примера. Один из них представляет ситуацию, когда достаточно относительно малого числа больных, а другой – когда требуется очень большое их число.

Таблица 9.4
Определение объема выборки

	Воля исследователя	Природа данных
N пропорционально	$\frac{1}{\Delta \cdot P_\alpha \cdot P_\beta}$	и $\frac{V}{1} \text{ или } \frac{1}{P}$

где N – число пациентов; Δ – величина различия в исходах между группами; P_α – вероятность α -ошибки, т.е. ложноположительного результата; P_β – вероятность β -ошибки, т.е. ложноотрицательного результата; V – вариабельность наблюдений (для количественных данных); P – доля больных с изучаемым исходом (для качественных данных)

Пример. Согласно наблюдениям серий случаев, нестероидный противовоспалительный препарат Сулиндак эффективен при полипах толстой кишки. Это предположение было проверено в рандомизированном испытании [6] на 22 больных с семейным аденоматозным полипозом, 11 из которых получали сулиндак, а другие 11 – плацебо. Через 9 мес у получавших сулиндак среднее число полипов было на 44% меньше, чем у получавших плацебо; различие статистически значимое ($p=0,014$). Поскольку лечебный эффект значителен, а на каждого пациента приходилось большое количество полипов (у некоторых более 100), для доказательства того, что лечебный эффект неслучаен, достаточно небольшого числа больных.

Пример. Исследование GUSTO [2], о котором упоминалось выше, было спланировано таким образом, чтобы при включении 41 000 пациентов оно с вероятностью 90% обеспечивало бы обнаружение снижения летальности в экспериментальной группе на 15% или частоты летальных исходов на 1% по сравнению с контрольной группой, в зависимости от того, какой из этих показателей будет больше. При этом допустимый уровень $p_{\alpha}=0,05$, а предполагаемая летальность в контрольной группе не ниже 8%. Здесь необходим большой объем выборки, так как доля больных с неблагоприятным исходом (смерть) относительно мала, величина лечебного эффекта невелика (15%) и авторы хотели иметь относительно высокую вероятность обнаружить эффект терапии, если он присутствует (90%).

Для получения ответов на большинство возникающих в наше время вопросов относительно эффективности того или иного вмешательства требуется изучение результатов лечения очень большого числа больных. Вместе с тем эффективность таких действительных вмешательств, как введение инсулина при диабетическом кетоацидозе или хирургической операции при аппендиците, можно установить при анализе данных небольшого числа больных. Однако подобные методы лечения появляются редко и многие из них уже хорошо изучены. Теперь нам приходится рассматривать патологию с хроническим течением и с множественными взаимодействиями этиологическими факторами; эффективность предлагаемых новых методов лечения таких заболеваний, в общем, невелика. В подобной ситуации необходимо обращать особое внимание на то, достаточна ли численность больных в клиническом испытании для того, чтобы отличить истинный лечебный эффект от случайного результата.

Клиницисты должны иметь возможность оценить статистическую мощь опубликованных исследований. С этой целью мы помещаем здесь рис. 9.3, который показывает соотношение между размером выборки и разницей в эффективности лечения для заранее известной частоты исходов. Очевидно, что исследова-

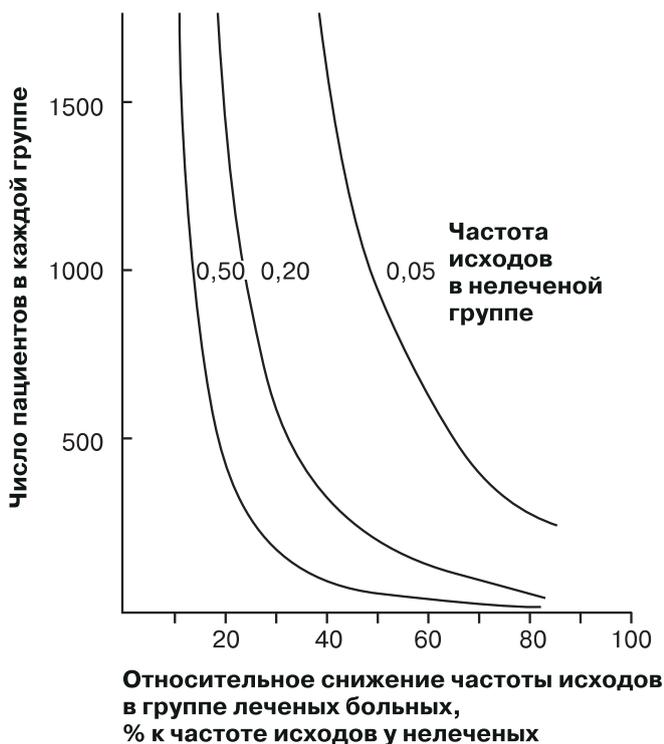


Рис. 9.3. Расчет числа пациентов, которое должно быть включено в каждую из двух одинаковых по численности и получающих разное лечение групп, при $p=0,05$ и заданных соотношениях разницы в частоте исходов между леченой и нелеченой группами (p) в зависимости от частоты исходов в нелеченой группе. (Рассчитано по формуле, представленной в: Weiss N.S. "Clinical epidemiology. The study of the outcome of illness". New York: Oxford University Press, 1986.)

ния, включающие менее 100 больных, имеют незначительные шансы выявить статистически значимые различия в частоте исходов даже при большом лечебном эффекте изучаемого метода. Вообще, трудно обнаружить эффект вмешательства, если он меньше 25%. На практике статистическую мощность исследования можно оценивать с помощью доступных формул, таблиц, номограмм или компьютерных программ.

Точечные оценки и доверительные интервалы

Величина эффекта (например, лечебный эффект в клиническом испытании или относительный риск в когортном исследовании), наблюдаемая в конкретном исследовании, называется *то-*

ческой оценкой эффекта. Эта оценка лучше всего отражает истинную эффективность вмешательства и, кроме того, является обобщающим статистическим показателем, на который обычно делается основной акцент в научных публикациях.

Однако истинная величина лечебного эффекта обязательно отличается от наблюдаемой в исследовании. В каждом конкретном исследовании результат будет либо выше, либо ниже истинной величины из-за случайной вариации. Следовательно, нужна обобщающая оценка степени вариации, которую можно ожидать вследствие случайности.

Статистическая точность оценки наблюдаемой величины эффекта выражается *доверительным интервалом* (confidence interval), обычно 95%, охватывающим область по обе стороны от точечной оценки. Доверительный интервал величины эффекта интерпретируется следующим образом: если в исследовании отсутствовала систематическая ошибка, то этот интервал с вероятностью 95% включает в себя истинную величину лечебного эффекта. Чем уже доверительный интервал, тем ближе точечная оценка эффекта к истинной величине. Истинная величина с наибольшей вероятностью находится вблизи точечной оценки, менее вероятно – у границ доверительного интервала и с еще меньшей вероятностью (5 из 100) может выходить за пределы интервала. Статистическая точность возрастает с увеличением статистической мощности исследования.

Доверительные интервалы дают информацию, аналогичную статистической значимости. Если точка, соответствующая полному отсутствию лечебного эффекта (например, относительный риск, равный 1, или различие в эффективности лечения, равное 0), лежит за пределами 95% доверительного интервала наблюдаемого эффекта, то результаты скорее всего статистически значимы при уровне значимости 0,05. Если эта точка находится внутри доверительного интервала, то результаты статистически незначимы.

Доверительные интервалы имеют ряд преимуществ. Они позволяют сделать акцент на величине эффекта. Доверительные интервалы позволяют читателям увидеть диапазон правдоподобных оценок и с учетом этого решить, согласуется ли величина эффекта, которую они считают клинически значимой, с результатами исследования [7]. Кроме того, доверительные интервалы дают представление о статистической мощности. Если размах доверительного интервала относительно велик и величина, соответствующая отсутствию эффекта, лежит на периферии интервала, то для установления статистически значимого различия следует увеличить мощность исследования.

Пример. На рис. 9.4 показаны точечные оценки и их доверительные интервалы для относительного риска развития трех заболеваний: рака эндометрия, рака молочной железы и перелома шейки бедра при приеме эстрогенов. (Обратите внимание, что риск представлен на логарифмической шкале, а это создает впечатление более узких, чем на самом деле, доверительных интервалов для более высоких рисков.) Наблюдаемый относительный риск развития рака эндометрия (после 8 и более лет лечения эстрогенами) составляет 8,22, но это не точно соответствует истинной величине, которая вполне может быть более высокой – 10,61 или более низкой – 6,25. В любом случае маловероятно, что она окажется равной 1,0 (отсутствие риска). Что касается рака молочной железы, то, согласно результатам исследования, прием эстрогенов вряд ли может быть фактором риска развития этого заболевания; точечная оценка относительного риска близка к 1,0. Данные также указывают на то, что либо вредное воздействие эстрогенов незначительно, либо их профилактическая эффективность мала. И наконец, прием эстрогенов

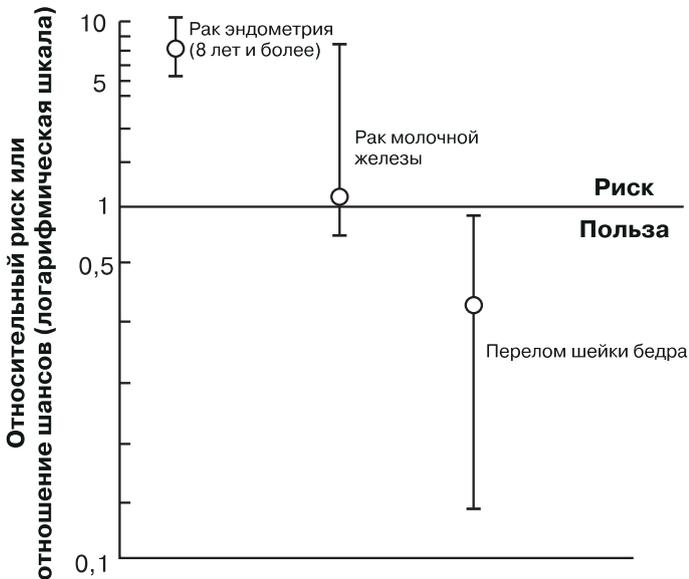


Рис. 9.4. Точечные оценки (O) и доверительные интервалы (I): риск и польза применения препаратов эстрогенов женщинами в постменопаузе [Grady D., Rubin S.M., Petitti D.B., Fox G.S., Black D., Ettinger B., Ernster V.L., Cummings S.R. Hormonal therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 1992; 117: 1016 – 1037; Colditz G.A., Stampfer M.J., Willett W.C., Hennekens C.H., Rosner B., Speizer F.E. Prospective study of estrogen replacement therapy and risk of breast cancer in postmenopausal women. *JAMA* 1990; 264: 2648 – 2653; and Paganini-Hill, Ross R.K., Gerkins V.R., Henderson B.E., Arthur M., Mack N.M. Menopausal estrogen therapy and hip fractures. *Ann Intern Med* 1981; 95: 28 – 31].

вполне вероятно может предупреждать переломы шейки бедра. Верхняя граница доверительного интервала находится ниже 1,0, т.е. профилактический эффект статистически значим при $p = 0,05$.

Точечные оценки и доверительные интервалы используются для характеристики статистической точности любого показателя частоты (заболеваемости и распространенности), сравнения показателей частоты (относительных и добавочных рисков) и других обобщающих статистических показателей. Например, отдельные исследования показали, что 34% взрослых американцев пользовались нетрадиционными методами лечения (95% доверительный интервал от 31 до 37% [8]), что при интенсивной терапии больных инсулинзависимым сахарным диабетом риск развития ретинопатии снижается на 76% (95% доверительный интервал от 62 до 85%) по сравнению с обычным лечением [9], что чувствительность клинического обследования для выявления спленомегалии составляет 27% (95% доверительный интервал от 19 до 36%) [10].

Определение доверительных интервалов стало обычным способом представления главных результатов клинических исследований, поскольку такой подход обладает многими преимуществами перед методом проверки гипотез (оценкой p). Однако оценка p все еще применяется – в силу традиции и из-за удобства использования в исследованиях, где представлено много данных и для всех трудно привести доверительные интервалы.

Статистическая мощность до и после выполнения исследования

Вычисление статистической мощности, основанное на методе проверки гипотез, производят до выполнения исследования, чтобы определить, достаточное ли число больных будет включено для выявления клинически значимого лечебного эффекта (если он существует). Однако после того, как исследование выполнено, вычисление статистической мощности не имеет смысла [11] – величина лечебного эффекта, частота исходов и вариабельность характеристик больных теперь уже известны. Поэтому для исследователей, публикующих результаты клинических испытаний, и для читателей, пытающихся понять их значение, применение доверительных интервалов более уместно. На стадии планирования исследования статистическая мощность определяется для произвольно выбранной величины лечебного эффекта. После завершения исследования важна величина лечебного эффекта и то, с какой статистической точностью эта наблюдаемая величина позволяет оценить величину истинную.

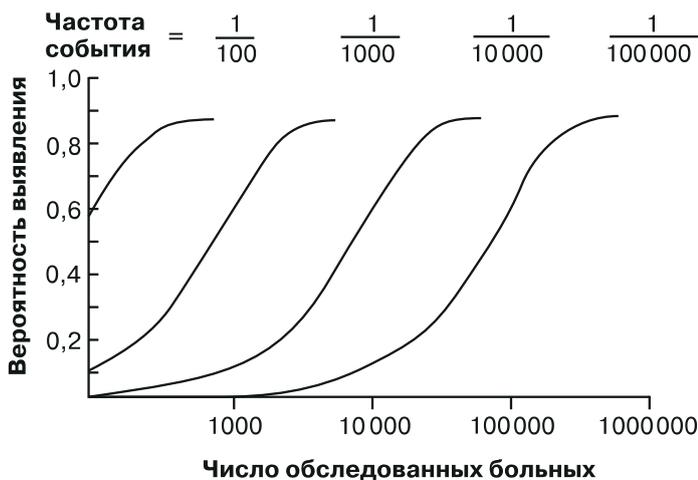


Рис. 9.5. Вероятность выявления события в зависимости от его частоты и от числа наблюдаемых пациентов [Guess H.A., Rudnick S.A. Use of cost effectiveness analysis in planning cancer chemoprophylaxis trials. Control Clin Trials 1983; 4: 89 – 100].

Выявление редких событий

Иногда важно выявить относительно редкие события (например, 1 на 1000), особенно если речь идет о серьезных заболеваниях, например медикаментозной апластической анемии или угрожающей жизни аритмии. В подобной ситуации нужно наблюдать очень большое число пациентов, чтобы выявить хотя бы один случай. И все равно это число будет куда меньше, чем необходимо для получения сколько-нибудь надежной оценки частоты явления.

На рис. 9.5 показана вероятность выявления события как функция от числа обследованных больных. Эмпирическое правило для расчета таково: чтобы иметь высокую вероятность выявления события, встречающегося с частотой $1/n$, необходимо обследовать $3 \times n$ пациентов [12]. Например, чтобы выявить событие, встречающееся с частотой $1/1000$, необходимо обследовать 3000 больных.

Множественные сравнения

Статистические выводы исследований имеют ауру особой доказательности, особенно для неспециалистов. Однако, как справедливо замечают скептики, возможна "ложь с помощью статистики", пусть даже неумышленная. Более того, это возможно даже когда

исследование хорошо спланировано, математические методы корректны, а намерения авторов безупречны.

Статистические выводы могут вводить в заблуждение, так как точность статистических критериев зависит от количества рассматриваемых в работе вопросов и от того, когда эти вопросы поставлены. Если в большом массиве данных проводятся сравнения между многими переменными, то величина p для каждого индивидуального сравнения недооценивает вероятность случайного результата этого сравнения. Как это ни удивительно, но интерпретация величины p для статистического критерия зависит от контекста.

Рассмотрим следующий пример. Допустим, что было выполнено масштабное исследование с большим количеством подгрупп больных и различными исходами. В частности, это может быть клиническое испытание эффективности лечения ишемической болезни сердца, в которое включено несколько разных групп больных (с поражением одной, двух или трех коронарных артерий; с нарушением и без нарушения функции левого желудочка; с нарушениями ритма сердца и без них; с различными сочетаниями этих состояний) и рассматривается несколько исходов (смерть, инфаркт миокарда, стенокардия). Предположим также, что в действительности не существует связи между разными видами лечения и исходами в любой из подгрупп. И наконец, предположим, что эффекты лечения оцениваются раздельно для каждой подгруппы и для каждого исхода. Это процесс, который включает в себя очень много сравнений. Как указывалось ранее в этой главе, при уровне значимости 0,05 одно из 20 таких сравнений по случайным причинам даст статистически значимый результат. В общем случае, если производится 20 сравнений, то в среднем одно покажет статистически значимый результат; если 100 сравнений, то примерно 5 и т.д. Таким образом, при большом числе сравнений некоторые из них просто вследствие случайности превысят уровень статистической значимости различий, даже если на самом деле истинных связей между переменными не существует. Чем больше проводится сравнений, тем выше вероятность выявления статистически значимых различий.

Этот феномен называется проблемой множественных сравнений. Вследствие этой проблемы, доказательность данных клинических исследований зависит от того, планировалось ли до начала исследования проведение этих сравнений.

К сожалению, из публикуемых результатов исследования не всегда можно узнать, сколько в действительности было сделано сравнений. Часто важные находки отбираются из большого числа несущественных. Выборочное принятие решения о том, что важно

и что неважно в отношении большого объема данных, может привести к значительному искажению действительности.

Как учесть статистический эффект множественных сравнений при интерпретации результатов исследования? Хотя с этой целью предлагались методы стандартизации p , все же лучше всего, понимая проблему, с осторожностью относиться к положительным заключениям исследований с множественными сравнениями. Один специалист по статистике выразил это следующим образом [13]:

"Если Вы анализируете данные достаточно глубоко и делаете это достаточно часто, Вы наверняка найдете что-нибудь необычное. Многие из этих странных находок связаны со случайностью. Я не имею ввиду, что манипуляции с данными – предосудительное занятие, просто находки, которые не были исходно сформулированы в качестве главных задач испытания, следует воспринимать с предельной осторожностью. Когда-нибудь теоретическая статистика научит нас, что делать с подобными результатами. А сейчас, я думаю, самая лучшая позиция – это осторожность, соединенная со стремлением подтвердить или отвергнуть находки в дальнейших исследованиях".

Об оценке достоверности статистически значимых различий в подгруппах сообщалось в главе 7.

Описание взаимосвязи

Статистические методы используются также для определения степени связи между переменными, например между массой тела и артериальным давлением. Хорошо известны такие показатели связи, как коэффициент корреляции Пирсона (r) для количественных данных и коэффициент ранговой корреляции Спирмена для порядковых данных. Каждый из этих показателей определяет количественную связь между двумя переменными и имеет соответствующий статистический критерий, который позволяет оценить, случайна ли наблюдаемая связь.

Многофакторные методы

Большинство клинических исходов представляет собой итог сложного взаимодействия многих переменных. Например, ИБС развивается как комплексный результат взаимодействия нарушений липидного обмена, артериальной гипертензии, курения, наследственных факторов, сахарного диабета, гиподинамии и индивидуальных особенностей. Целесообразно вначале попытаться понять эти взаимосвязи путем изучения относительно простых сочетаний параметров, таких, которые представлены в четырехпольных таблицах (с одной зависимой переменной) или в многополь-

ных таблицах сопряженности (анализ, оценивающий, как изменяется эффект одной переменной в присутствии или в отсутствие одной либо большего числа других переменных); такой способ представления данных облегчает их восприятие. Однако при большом числе переменных этот метод обычно неприемлем, поскольку из-за отсутствия достаточного числа больных с каждым сочетанием факторов невозможно получить стабильные оценки показателей. Например, в исследование включено 120 больных, по 60 в каждой из групп с разным лечением. Даже если в расчет принимается только одна дополнительная дихотомическая переменная, в каждой подгруппе в лучшем случае окажется лишь по 15 больных, а при неравномерном распределении некоторые подгруппы будут содержать еще меньшее число больных.

В подобной ситуации в дополнение к построению таблиц сопряженности требуется метод оценки одновременного влияния нескольких переменных. Это *многофакторное моделирование* (multivariable modeling), которое обеспечивает математическое выражение сочетанного влияния многих переменных. Метод называется многофакторным, так как он оценивает эффект многих переменных. Термин "моделирование" означает, что это математическая конструкция, которая рассчитывается из полученных данных, но основана она на упрощающих допущениях о характеристиках этих данных (например, все переменные имеют нормальное распределение и одинаковую дисперсию).

Математические модели могут использоваться в исследованиях причинно-следственных связей, когда нужно определить вклад одной переменной, независимый от влияния нескольких других (посторонних) переменных. Включение в прогностическую модель одновременно нескольких переменных обеспечивает более точные предсказания, чем это позволяют отдельные переменные.

Основная структура многофакторной модели такова:

Переменная, характеризующая исход = константа + ($\beta_1 \times$ переменная1) + ($\beta_2 \times$ переменная2) + ...

где β_1 и β_2 ... – коэффициенты, которые определяются на основании имеющихся данных; переменная 1, переменная 2, ... – прогностические переменные, связанные с исходом. Самые точные значения коэффициентов определяются математически, что в значительной мере зависит от вычислительной возможности современных компьютеров.

Моделирование включает в себя несколько этапов.

- Выявление и измерение всех переменных, которые могут быть связаны с изучаемым исходом.
- Уменьшение количества рассматриваемых в модели перемен-

ных до приемлемого числа (обычно не более нескольких). Часто это делается путем отбора переменных, которые при одновременном их влиянии связаны с исходом наиболее сильно. Если статистический критерий используется на этой стадии, то обычно ошибка совершается при включении в модель переменных, у которых связь с изучаемым исходом имеет уровень значимости $p < 0,10$. При отборе должны учитываться также сведения о биологическом значении переменной.

- Если несколько переменных сильно связаны друг с другом, то следует включать в модель лишь одну из них, поскольку все эти переменные несут практически одинаковую информацию.
- Включение в модель остальных переменных в определенной последовательности, в соответствии с изучаемым вопросом. Например, если при анализе причин некоторые переменные должны быть фиксированными, они включаются в модель первыми, далее следует переменная, представляющая наибольший интерес. Тогда модель позволит определить независимый вклад главной переменной. С другой стороны, если исследователь желает сделать прогноз, основанный на нескольких переменных, то они могут быть включены в порядке убывания силы их связи с исходом, что определяется самой моделью.

Моделирование – постоянный атрибут современной медицинской литературы, оно присутствует примерно в 18% статей в основных медицинских журналах [14] и почти во всех крупных исследованиях причинно-следственных связей. Широко используются такие виды моделей, как логистическая регрессия (для дихотомических исходов, имеющих место в исследованиях случай–контроль) и модель пропорционального риска Кокса (в исследованиях, где исходом является время наступления события).

Многофакторное моделирование – неотъемлемая часть многих клинических исследований. Другого метода для одновременного учета или изучения многих переменных не существует. Однако эти методы имеют оборотную сторону. Модель – это "черный ящик", в который трудно заглянуть, чтобы понять, что в нем происходит. Достоверность моделей базируется на допущениях, которые могут не подтверждаться. На моделях плохо распознаются модификации эффекта (различия эффектов в разных подгруппах больных). Переменная, сильно связанная с исходом, может не появиться в модели из-за того, что она встречается редко; в результате модель оказывается нечувствительной к эффекту этой переменной по причине недостаточной информации.

Наконец, на результаты моделирования сильно влияет случайная вариабельность характеристик больных в исследуемых вы-

борках. Показано, например, что в разных случайных выборках из одного и того же массива данных модель часто идентифицировала разные прогностические факторы и распределяла их в разном порядке [15]. Для того чтобы избежать этого, следует, пользуясь эмпирическим правилом, включать в модель лишь те прогностические факторы, с которыми связано не менее 10 случаев исхода.

По этим причинам модели сами по себе не могут служить стандартом достоверности; их достоверность должна специально проверяться. Обычно для этого выясняют, предсказывает ли модель то, что выявляется в другой, независимой выборке больных (см. главу 12). Результаты первой модели считаются гипотезой, которая должна быть проверена на новых данных. Если результаты первой модели обусловлены главным образом случайностью, то маловероятно, чтобы та же случайность имела место в выборке, используемой для проверки модели. В пользу достоверности модели свидетельствует также ее соответствие биологическим представлениям и согласованность ее результатов с результатами анализа данных, выполненного более простым и ясным способом.

Резюме

Основу клинической информации составляют наблюдения на выборках больных. Однако даже выборки, отобранные без систематической ошибки, могут давать искаженное представление о событиях в общей популяции из-за случайной вариации показателей у включенных в выборку пациентов.

Есть два общих подхода для выявления роли случайности в клинических исследованиях – проверка гипотез и метод доверительных интервалов. При проверке гипотез с помощью статистических критериев производится оценка предположения, что наблюдаемый результат имел случайный характер. Когда в испытании сравниваются два метода лечения, возможны два рода ошибок: ошибочный вывод о предпочтительности одного метода перед другим, хотя на самом деле они не различаются по эффективности, и ошибочный вывод об отсутствии различий в эффективности при том, что в действительности один метод лучше другого. Вероятность одной и другой ошибки обозначается p_α и p_β соответственно.

Мощность статистического критерия $(1-p_\beta)$ – это вероятность выявления статистически значимого различия, когда оно действительно существует. Статистическая мощность зависит от числа больных в исследовании, величины лечебного эффекта, p_α и частоты исходов, либо индивидуальной варибельности реакции больных на воздействие. При прочих равных условиях статисти-

ческую мощность можно повысить, увеличив число больных в исследовании, но это не всегда осуществимо.

Метод доверительных интервалов заключается в использовании полученных данных для определения диапазона, который с заданной вероятностью включает в себя истинную величину эффекта. Используются точечные оценки и доверительные интервалы. Этот подход имеет много преимуществ перед методом проверки гипотез. Он позволяет сделать акцент на величине эффекта, а не на величине p – уровне статистической значимости; показывает диапазон правдоподобных величин эффекта, так что врач может соотнести представленные таким образом данные с клиническим значением эффекта; обеспечивает информацию о статистической мощности исследования.

Риск ложноположительного результата возрастает при множественных сравнениях. Риск ложноотрицательного результата увеличивается, если статистическая мощность исследования невелика (обычно из-за небольшого числа включенных больных или малой распространенности изучаемого исхода).

В большинстве клинических исследований рассматривается влияние многих взаимодействующих переменных. Путем многофакторного моделирования можно оперировать со всеми переменными одновременно. Это позволяет контролировать посторонние переменные в исследовании причин и обеспечивает более точное прогнозирование, чем это возможно при оценке влияния только одной переменной. Однако результаты, полученные на этих моделях, следует интерпретировать с осторожностью, поскольку не вполне понятно, как эти модели работают, насколько они чувствительны к случайной вариации показателей и допущения, на которых эти модели основаны, могут быть не верны.

Литература

1. Johnson A.F. Beneath the technological fix. Outliers and probability statements. *J Chron Dis* 1985;38:957–961.
2. The GUSTO Investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *New Engl J Med* 1993;329:673–682.
3. Farkouh M.E., Lang J.D., Sackett D.L. Thrombolytic agents: the science of the art of choosing the better treatment. *Ann Intern Med* 1994;120:886–888.
4. Raschke R.A., Reilly B.M., Guidry J.R., Fontana J.R., Srinivas S. The weight-based heparin dosing nomogram compared with a "standard care" nomogram. A randomized controlled trial. *Ann Intern Med* 1993;119:874–881.
5. Peto R., Pike M.C., Armitage P., Breslow N.E., Cox D.R., Howard S.V., Mantel N., McPherson K., Peto J., Smith P.G. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. 1. Introduction and design. *Br J Cancer* 1976;34:585–612.
6. Giardiello F.M., Hamilton S.R., Krush A.J., Piantadosi S., Hyland L.M., Celano P., Booker S.V., Robinson C.R., Offerhaus G.J.A. Treatment of colonic and rectal adenomas with sulindac in familial adenomatous polyposis. *New Engl J Med* 1993;328:1313–1316.

7. Braitman L.E. Confidence intervals assess both clinical significance and statistical significance. *Ann Intern Med* 1991;114:515–517.
8. Eisenberg D.M., Kessler R.C., Foster C., Norlock F.E., Calkins D.R., Delbanco T.L. Unconventional medicine in the United States. Prevalence, costs and patterns of use. *New Engl J Med* 1993;328:246–252.
9. Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *New Engl J Med* 1993;329:977–986.
10. Grover S.A., Barkun A.N., Sackett D.L. Does the patient have splenomegaly? *JAMA* 1993;270:2218–2221.
11. Goodman S.N., Berlin J.A. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121:200–206.
12. Sackett D.L., Haynes R.B., Gent M., Taylor D.W. Compliance. In: Inman WHW, ed. *Monitoring for drug safety*, Lancaster, UK: MTP Press, 1980.
13. Armitage P.I. Importance of prognostic factors in the analysis of data from clinical trials. *Control Clin Trials* 1981;1:347–353.
14. Concato J, Feinstein A.R., Holford T.R. The risk of determining risk with multivariable models. *Ann Intern Med* 1993;118:201–210.
15. Diamond G.A. Future imperfect: the limitations of clinical prediction models and the limits of clinical prediction. *J Am Coll Cardiol* 1989;14:12A–22A.

Рекомендуемая литература

- Altman D.G., Gore S.M., Gardner M.J., Pocock S.J. Statistical guidelines for contributors to medical journals. *Br Med J* 1983;286:1489–1493.
- Bailer J.C. III, Mosteller F, eds. *Medical uses of statistics*. Waltham, MA: New England Journal of Medicine Books, 1986.
- Cupples L.A., Heeren T., Schatzkin A., Colton T. Multiple testing of hypotheses in comparing two groups. *Ann Intern Med* 1984;100:122–129.
- Detsky A.S., Sackett D.L. When was a "negative" clinical trial big enough? How many patients you need depends on what you found. *Arch Intern Med* 1985;145:709–712.
- Carpenter L.M. Is the study worth doing? *Lancet* 1993;343:221–223.
- Datta M. You cannot exclude the explanation you haven't considered. *Lancet* 1993;342:345–347.
- Glynn J.R. A question of attribution. *Lancet* 1993;342:530–532.
- Grisso J.A. Making comparisons. *Lancet* 1993;342:157–160.
- Jolley T. The glitter of the t table. *Lancet* 1993;342:27–29.
- Leon D. Failed or misleading adjustment for confounding. *Lancet* 1993;342:479–481.
- Mertens T.E. Estimating the effects of misclassification. *Lancet* 1993;342:418–421.
- Sitthi-Amorn C., Poshachinda V. Bias. *Lancet* 1993;343:286–288.
- Victoria C.G. What's the denominator? *Lancet* 1993;342:97–99.
- Friedman G.D. *Primer of epidemiology*, 4th ed. New York: McGraw-Hill, 1994.
- Gardner M.J., Altman D.G. *Statistics with confidence—confidence intervals and statistical guidelines*. London: BMJ Books, 1989.
- Goodman S.N., Berlin J.A. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121:200–206.
- Hanley J.A., Lipman-Hand A. If nothing goes wrong is everything right? Interpreting zero numerators. *JAMA* 1983;249:1743.
- Hennekens C.H., Buring J.E. *Epidemiology in medicine*. Boston: Little, Brown & Co. 1987.
- Ingelfinger J.A., Mosteller F, Thibodeau L.A., Ware J.H. *Biostatistics in clinical medicine*. New York: Macmillan 1983.
- Moses L.E. *Statistical concepts fundamental to investigations*. *N Engl J Med* 1985;312:890–897.
- Riegelman R.K., Hirsch R.P. *Studying and study and testing a test*. 2nd ed. Boston: Little, Brown & Co. 1989.
- Rothman K.J. A show of confidence. *N Engl J Med* 1978;299:1362–1363.
- Young M.J., Bresnitz E.A., Strom B.L. Sample size nomograms for interpreting negative clinical studies. *Ann Intern Med* 1983;99:248–251.

10

ИЗУЧЕНИЕ СЛУЧАЕВ

Любой случай поучителен, но урок не всегда усваивается.

Сэр Уильям Ослер

Медицинские знания создаются прежде всего в результате тщательного изучения пациентов. Привычный для всех образ – усталый, сосредоточенный врач, склонившийся над постелью лихорадящего ребенка. Основу современного медицинского образования составляет описание и обсуждение клинических случаев. Большинство клинических и патологоанатомических конференций и "круглых столов" начинается представлением интересного случая с последующим обсуждением на его примере общих принципов и связей. Медицинская литература тоже уделяет много внимания исследованиям случаев, будь то простое изложение материала о нескольких больных (*описание случаев, case report*), количественный анализ группы из большого числа больных (*исследование серий случаев, case series*) или сравнение двух групп, в одну из которых отобраны пациенты с определенной патологией, а в другую – лица без нее (*исследование случай – контроль, case control study*).

Описание случаев

Описание случая – подробное изложение данных, полученных путем наблюдения одного или нескольких случаев. Это хороший метод привлечения внимания медиков к новым или малоизвестным болезням, проявлениям или сочетаниям болезней. Примерно 20 – 30% оригинальных статей в ведущих медицинских журналах представляют собой исследования групп, состоящих не более чем из 10 больных.

Зачем нужны сообщения о случаях

Описания случаев преследуют несколько целей. Прежде всего это единственный способ сообщения о редком клиническом событии, богатый источник гипотез о патогенезе, риске, прогнозе и лечении. Описания случаев редко могут использоваться для проверки таких гипотез, у них другая задача: поставить проблему пе-

ред медицинским сообществом и побудить к проведению более убедительных исследований. Благодаря описаниям случаев были впервые выявлены врожденные уродства вследствие приема талидомида во время беременности, алкогольный синдром плода, токсический шок, лаймская болезнь и хантавирусная инфекция.

Сообщения о случаях служат также для выяснения механизмов заболевания и лечения, поскольку подробно описывают методически сложные клинические и лабораторные исследования, проводимые при изучении одного или нескольких больных. Это ситуации, когда сложность, высокая стоимость и часто экспериментальный характер исследования ограничивают возможность его проведения лишь небольшим числом больных. Подобные работы внесли существенный вклад в понимание генетических, метаболических и физиологических основ заболеваний человека. Они являются мостом между лабораторным экспериментом и клиническим исследованием и занимают важное место в прогрессе медицины.

Приведем пример, показывающий, как сообщение о единичном случае может дать важную информацию о механизме заболевания.

Пример. Предполагалось, что галотан, средство для ингаляционного наркоза, может вызывать гепатит. Однако представлению о "галотановом гепатите" противоречила низкая частота гепатита после применения галотана и возможная роль многих других причин для развития гепатита после хирургического вмешательства.

Прояснить проблему помог опыт наблюдения единичного случая [1]. У анестезиолога был обнаружен рецидивирующий гепатит, приведший к циррозу печени. Обострения гепатита регулярно повторялись через несколько часов после возвращения анестезиолога к работе. Экспериментальное воздействие малой дозы галотана немедленно приводило у него к рецидиву гепатита, что было документировано данными клинического наблюдения, результатами биохимических тестов и гистологическим исследованием биоптата печени.

Благодаря этому необычному случаю стало ясно, что галотан действительно может вызывать гепатит. Однако описание случая не давало информации о том, насколько часто возникает такая реакция на галотан. Последующие исследования, показавшие, что это отнюдь не редкий побочный эффект, способствовали замене галотана на менее гепатотоксичные средства для наркоза.

Другое направление использования сообщений о случаях – описание необычных проявлений болезни. Иногда это может принимать форму медицинской версии сборника "Хочешь верь, хочешь нет", неформального собрания удивительных медицинских

случаев, привлекающих интерес своей невероятностью. Чем значительнее повреждение, чем невероятнее размеры инородного тела, тем скорее сообщение о случае будет опубликовано. Описания редких случаев, представляющих собой отклонения от обычного течения заболеваний, могут волновать воображение, но, как правило, клинически менее важны, чем другие типы исследований.

В то же время некоторые описания необычных случаев оказываются результатом более свежего, более пронизательного взгляда на проблему и первым представлением того, что впоследствии окажется полезной находкой. Задача читателя заключается в том, чтобы суметь отличить новый взгляд на проблему от обычного отклонения от нормы. Правил здесь не существует. Когда ничто другое не помогает, приходится полагаться только на здравый смысл и хорошо развитый скептицизм.

Предвзятое описание

Описания случаев в небольших специально подобранных группах часто бывают предвзятыми. Поскольку журналы более охотно принимают к публикации работы с описанием случаев успешного лечения, авторы стремятся представлять именно такие данные, опуская свой негативный опыт. Вероятно, при изучении сообщений, описывающих случаи, правильнее всего воспринимать их как стимул к поиску дальнейших доказательств представленного феномена в медицинской литературе или среди своих больных.

Пример. В сообщении [2] описана больная 23 лет, у которой появились сильные боли в животе на фоне лечения эналаприлом по поводу гипертонической болезни. Повышенный уровень липазы в сыворотке указывал на наличие панкреатита. Вскоре после отмены эналаприла симптомы исчезли и уровень липазы нормализовался. Исследователи нашли только один аналогичный опубликованный случай и начали интенсивный поиск опубликованных и неопубликованных данных. Было выявлено еще 60 случаев, большинство из них – неопубликованные отчеты фирм, производящих препарат. Это увеличило уверенность в наличии причинно-следственной связи между приемом эналаприла и развитием панкреатита.

За редкими исключениями описание случаев не служит основанием для изменения клинической практики, так как оно не позволяет оценить ни частоту возникновения таких случаев, ни роль систематических и случайных ошибок.

Совпадение редких событий

Сообщения о случаях часто описывают совпадение редких событий, особенно если наблюдаемое совпадение наводит на

мысль об интересном биологическом механизме. Однако даже редкие события иногда совпадают лишь по чистой случайности; факт их совпадения еще не означает, что они биологически связаны. Как указывает J. Mulvihill [3]: "Когда речь идет о большой популяции, то вопрос не в том, возможны ли редкие события, а в том, происходят ли они чаще, чем это ожидается в силу случайности".

Данные, представленные в табл. 10.1, иллюстрируют, насколько часто могли бы совпадать по чистой случайности два относительно редких события – хроническая почечная недостаточность и прием нестероидных противовоспалительных средств. В отсутствие биологической связи между этими двумя событиями (а как будет видно из дальнейших рассуждений, такая связь вполне возможна) вероятность их совпадения равна произведению частот каждого из них. Только в США ежегодно должно наблюдаться 100 таких случаев – более чем достаточно для публикации нескольких сообщений.

Существуют определенные причины, вследствие которых такие редкие сочетания встречаются в медицинских центрах и служат предметом публикаций гораздо чаще, чем они встречаются в общей популяции. Во-первых, пациент, страдающий сразу двумя серьезными заболеваниями, с большей вероятностью будет госпитализирован (хотя бы вследствие тяжести его состояния), чем пациент с одним заболеванием. Во-вторых, показано, что две болез-

Таблица 10.1

**Сочетание двух редких событий:
оценка частоты и числа совпадений случаев применения нестероидных
противовоспалительных средств и хронической почечной
недостаточности на основании предположения об отсутствии
биологической связи между этими двумя состояниями***

Частота отдельных событий	
Распространенность применения лекарственного средства (гипотетическая)	1/100 человек
Частота хронической почечной недостаточности	40/1 000 000 / год
Частота сочетания двух событий	$1/100 \times 40/1\ 000\ 000$ / год $= 4/10\ 000\ 000$ / год
Население США	250 000 000
Число совпадений двух событий в США	$4/10\ 000\ 000$ / год \times 250 000 000 $= 100$ / год

*[Hiatt K.A., Friedman G.D. Characteristics of patients referred for treatment of end-stage renal disease in a defined population. Am J Public Health 1982;72:829 – 833].

ни, не связанные между собой в общей популяции, в условиях стационара могут встречаться в сочетании, если больные с двумя разными болезнями поступают в больницу с разной частотой [4]. Кроме того, пациент, страдающий сразу двумя заболеваниями, интересен в научном отношении, так что такой случай с более высокой вероятностью будет описан, а соответствующая статья принята к публикации.

Следовательно, нужно скептически относиться к выводам о взаимосвязи болезней, которые основаны только на описании случаев. Это всего лишь гипотезы, которые сначала должны быть проверены более строгими методами.

Исследования серий случаев

Исследование серий случаев – это изучение группы численностью 10 пациентов и более с определенным заболеванием. Большее число случаев позволяет оценить роль случайности, величину p и других статистических показателей, которые часто фигурируют в исследованиях серий случаев, в отличие от описаний отдельных случаев. Серия случаев – самый распространенный способ описания клинической картины заболевания; он достаточно информативен и удобен для выполнения этой задачи, но имеет серьезные ограничения.

Недостаток исследований серий случаев заключается в отсутствии групп сравнения. Иногда это оказывается не принципиальным, о чем свидетельствует следующий пример.

Пример. В период с июня 1981 г. по февраль 1983 г., спустя несколько лет после того как впервые был описан СПИД и определены его признаки, ученые из Центра по контролю за заболеваниями собрали информацию о проживающих в США 1000 больных, у которых симптомы соответствовали критериям диагностики этого заболевания. При этом были описаны демографические и поведенческие характеристики больных, а также осложнения. Пневмоцистная пневмония была выявлена у 50% больных, саркома Капоши – у 28%, а оба эти заболевания – у 8% больных; у 14% выявлены инфекции, отличные от пневмоцистной пневмонии; 94% больных отнесены к одной или нескольким из следующих групп: мужчины гомо- или бисексуалисты, лица, применяющие наркотики внутривенно, уроженцы Гаити и больные гемофилией [5].

В этом исследовании отсутствовала контрольная группа, составленная из лиц без СПИДа. Кроме того, некоторые критерии диагностики СПИДа были описаны позднее. Тем не менее, поскольку наблюдавшиеся сопутствующие заболевания достаточно

редки у лиц без СПИДа и характеристики групп риска весьма выразительны, это сообщение прояснило представление о СПИДе и положило начало этапу более детального изучения его проявлений и факторов риска.

С другой стороны, если два признака часто встречаются вместе, то отсутствие группы сравнения может привести к ошибочному выводу.

Пример. Многие врачи объясняют боли в пояснично-крестцовой области выпячиванием одного или нескольких межпозвоночных дисков. В нескольких исследованиях серий случаев с помощью магнитно-резонансной томографии изучались анатомические особенности пояснично-крестцового отдела позвоночника при таких болях. У большинства больных в межпозвоночных дисках имелись отклонения от нормы, что подтверждало представление о роли патологии дисков в возникновении боли в пояснично-крестцовой области. Однако, как отмечалось в главе 3, такая же распространенность патологии дисков была установлена при магнитно-резонансной томографии у лиц, симптомы заболевания у которых отсутствовали. Это ставит под сомнение причинно-следственную связь между болями в пояснично-крестцовой области и выпячиванием дисков, выявляемым при магнитно-резонансной томографии [6].

Другой недостаток исследований серий случаев состоит в том, что они рассматривают клинические проявления и лечение болезни в группе, сформированной в определенный момент времени, т.е. в когорте дожития (см. главу 6). Поэтому их нужно отличать от когортных исследований или испытаний методов лечения, в которых исходная когорта пациентов с изучаемым заболеванием прослеживается в течение определенного периода для регистрации исходов. Исследования серий случаев часто являются ретроспективными и это ограничивает их ценность для анализа прогноза или изучения причинно-следственных связей.

Исследования случай – контроль

Для того чтобы выяснить, действительно ли некоторый признак или возможная причина чаще встречаются у лиц с изучаемым заболеванием, необходимо соблюдать ряд условий. Во-первых, и это самое главное, кроме группы с изучаемым заболеванием должна присутствовать контрольная группа без заболевания. Во-вторых, число наблюдений должно быть достаточно большим, чтобы свести к минимуму вероятность случайного результата. В-третьих, для обеспечения надежного сравнения группы должны быть сходными по всем признакам, за исключением наличия изучаемого заболевания. Наконец, чтобы доказать, что фактор риска

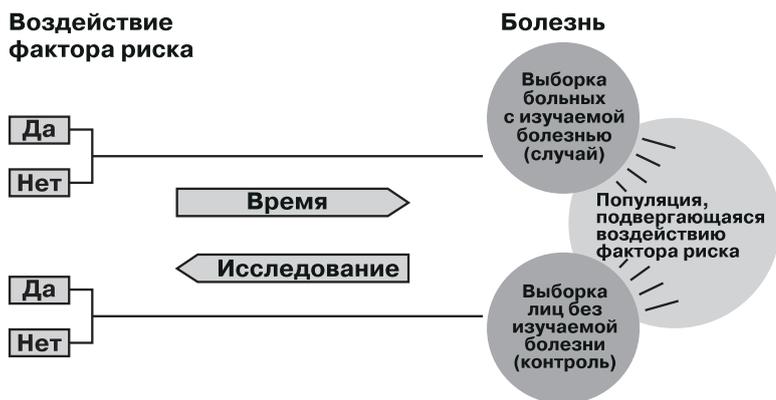


Рис. 10.1. Методика исследования случай – контроль.

не зависит от других факторов и потому может считаться независимой причиной, при анализе данных нужно контролировать все другие существенные различия между группами.

Всего этого не могут обеспечить описания случаев и исследования серий случаев. Этим условиям во многих ситуациях не удовлетворяют и когортные исследования, поскольку не позволяют накопить достаточное число наблюдений для устранения случайности. Подобным требованиям отвечают исследования случай – контроль, в которых сравнивается распространенность предполагаемого фактора риска (обычно называемого воздействием) в экспериментальной и контрольной группах.

Методика исследования

Принципиальный план исследования случай–контроль представлен на рис. 10.1. Сначала производится отбор группы пациентов с изучаемым заболеванием и сходной по остальным признакам группы лиц без этого заболевания. Затем исследователи ретроспективно определяют частоту воздействия изучаемого фактора в обеих группах. Полученные данные позволяют рассчитать относительный риск развития заболевания в связи с изучаемым фактором.

Пример. Увеличивает ли прием нестероидных противовоспалительных препаратов риск развития заболевания почек? Такой вопрос был поставлен в исследовании случай – контроль [7]. Как оно выполнялось?

Во-первых, необходимо было определить, что понимается под заболеванием почек и собрать достаточную по численности группу лиц с данной патологией. По понятным причинам поиск пациентов проводился в специализированных клиниках, где сосредоточены подобные больные.

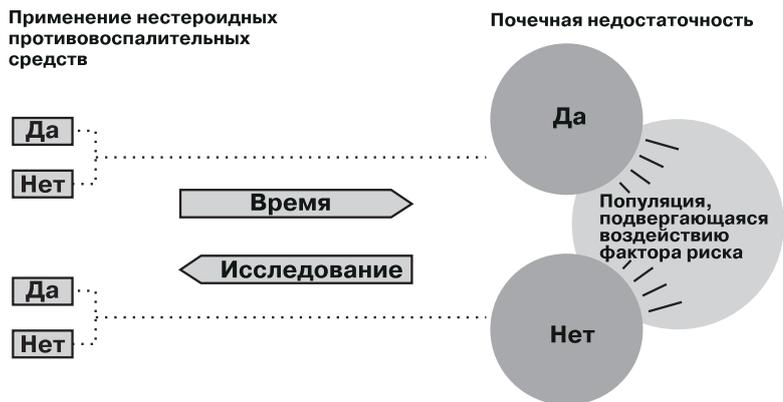
Разумеется, отобранные случаи включали только больных, у которых диагноз был установлен в процессе медицинского обследования. Вероятность включения бессимптомных больных с умеренной почечной недостаточностью в основную группу была невелика.

После того как случаи были отобраны и диагноз верифицирован, производился отбор группы контроля. Этот отбор происходил на основе сформулированной цели исследования. Она такова: установить, чаще ли больные с почечной недостаточностью получали в прошлом нестероидные противовоспалительные препараты, чем пациенты сходной группы без заболевания почек.

Исследователи установили, что относительный риск применения нестероидных противовоспалительных препаратов для развития почечной недостаточности составляет 2,1, и что повышенный риск наблюдается у мужчин пожилого возраста. Расчет произведен на основании данных о частоте использования этих препаратов в основной и контрольной группах.

Что означает "сходная группа"? В толковании этого понятия имеются противоречия. В когортном исследовании риска развития заболевания почек при лечении нестероидными противовоспалительными средствами понятие "сходная" означает, что пациенты контрольной группы должны быть представителями той же когорты, из которой набрана основная группа, и к началу исследования не должны иметь заболевания почек. Например, это могут быть лица, проживающие в одном и том же регионе или находящиеся под наблюдением одного и того же медицинского учреждения. Но существует ли естественная когорта, из которой может быть набрана группа пациентов, получающих помощь в определенной специализированной клинике? Система направления больных в специализированную клинику организована таким образом, что поступающие туда больные обычно проживают в разных населенных пунктах, лечатся у разных врачей и до заболевания не принадлежали к одной и той же популяции. Поэтому контрольной группы лиц без заболевания почек не существовало, ее требовалось создать. Это было сделано путем случайного отбора среди лиц, проживавших в районе обслуживания каждой клиники. Можно было надеяться, что сформированная таким способом контрольная группа позволит оценить вероятную распространенность использования нестероидных противовоспалительных средств среди больных с почечной недостаточностью в отсутствие связи между заболеванием почек и приемом этих препаратов [8].

Исследование случай–контроль



Когортное исследование

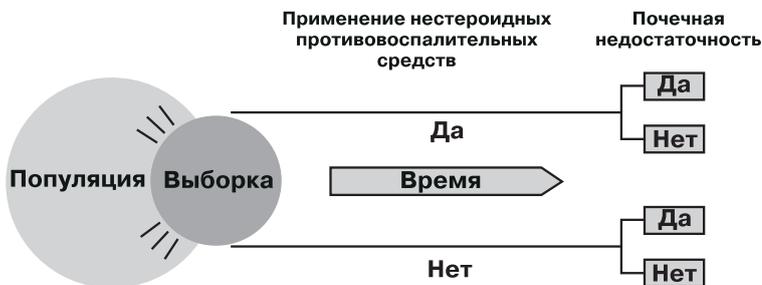


Рис. 10.2. Сопоставление когортного исследования и исследования случай — контроль.

После того, как основная и контрольная группы были отобраны, приступили к следующему этапу, который состоял в оценке воздействия изучаемого фактора риска. Необходимо было собрать анамнез у каждого участника основной и контрольной групп в отношении применявшихся лекарственных средств. В отличие от когортных исследований, где использование препаратов можно проследить во времени, факт приема лекарства в исследовании случай – контроль восстанавливается по памяти.

Таким образом, в исследовании случай – контроль используется информация о событиях, имевших место в прошлом, и здесь заложена возможность систематической ошибки. Всегда существует тенденция толковать прошлое в свете изучаемого события. Это особенно естественно для случаев, когда в настоящем имеется столь серьезная патология, как почечная недостаточность. Для того чтобы избежать систематической ошибки, исследователи стараются использовать объективные методы: слепой метод отбора пациентов, слепой метод анализа случаев, определение критериев достоверности факта приема нестероидных противовоспалительных средств в прошлом.

Отличия когортных исследований от исследований случай – контроль

Как когортное исследование, так и исследование случай – контроль – это наблюдательные исследования факторов риска. Иногда их путают между собой. Отличительная особенность методики исследования случай – контроль состоит в том, что к моменту начала исследования все изучаемые исходы уже наступили. В когортном исследовании в начале наблюдения, когда оцениваются факторы риска, участники еще не имеют изучаемого заболевания. На рис. 10.2 обобщены различия между методами когортного исследования и исследования случай – контроль. Поскольку наличие связи во времени между предполагаемой причиной и результатом служит важным критерием для оценки причинно-следственных отношений (см. главу 11), когортные исследования обеспечивают более строгую основу такой оценки.

В табл. 10.2 обобщены основные характеристики и показаны различия методов когортного исследования, исследования случай – контроль и одномоментного исследования. Как будет видно из дальнейшего изложения материала, именно вследствие этих различий исследования случай – контроль особенно подвержены систематическим ошибкам.

Отношение шансов

Как определить, повышен ли риск в действительности? На рис. 10.3 показан пример расчета риска для исследований – когортного и случай–контроль. В когортном исследовании популяция, подвергающаяся воздействию фактора риска, разделяется на две группы: одна из них (A+B) в начале исследования подвергалась воздействию нестероидных противовоспалительных препаратов, другая (C+D) – не подвергалась. С течением времени слу-

Таблица 10.2
Основные характеристики когортного исследования, исследования случай—контроль и одномоментного исследования

Когортное исследование	Исследование случай—контроль	Одномоментное исследование
Начинается с определения популяции, подвергающейся воздействию фактора риска	Популяция, подвергающаяся воздействию фактора риска, не обязательно определена	Начинается с определения популяции
Случаи не отбираются, а устанавливаются в процессе непрерывного наблюдения (предпочтительно все случаи)	Случаи отбираются исследователем из имеющейся совокупности больных	Случаи не отбираются, а устанавливаются при одномоментном обследовании населения
Контрольная группа (без изучаемого заболевания) не отбирается, а формируется естественным образом	Контрольная группа отбирается исследователем таким образом, чтобы она была сходной с экспериментальной группой	Контрольная группа включает в себя лиц, у которых не выявлено заболевания при одномоментном обследовании
Воздействие оценивается до развития заболевания	Факт воздействия оценивается и восстанавливается по памяти после развития заболевания	Факт воздействия оценивается и восстанавливается по памяти после развития заболевания
Риск или заболеваемость, а также относительный риск измеряются непосредственно	Риск или заболеваемость нельзя измерить непосредственно; относительный риск воздействия можно оценить по отношению шансов	Риск или заболеваемость нельзя измерить непосредственно; относительный риск воздействия можно оценить по отношению шансов

чаи заболевания почек возникают как в группе, получавшей нестероидные противовоспалительные препараты (А), так и в группе их не получавшей (С). Это дает нам числитель и знаменатель для расчета заболеваемости почечной недостаточностью в когорте, подвергавшейся воздействию $[A/(A+B)]$, и в когорте, не подвергавшейся воздействию $[C/(C+D)]$. Кроме того, можно рассчитать относительный риск.

$$\text{Относительный риск} = \frac{\text{заболеваемость в когорте, подвергавшейся воздействию}}{\text{заболеваемость в когорте, не подвергавшейся воздействию}} = \frac{A/(A+B)}{C/(C+D)}$$

С другой стороны, исследование случай–контроль начинается с отбора основной группы с заболеванием почек (А+С) и группы контроля без заболеваний почек (В+D). Определить показатели частоты этого заболевания невозможно, так как эти группы формируются не естественным образом, а создаются согласно критериям отбора, установленным исследователем. Поэтому способ вычисления заболеваемости среди тех, кто подвергся и кто не подвергся воздействию нестероидных противовоспалительных средств, не существует. Следовательно, нельзя рассчитать относительный риск путем деления заболеваемости среди подвергшихся на заболеваемость среди не подвергавшихся воздействию. Однако известны относительные частоты воздействия нестероидных противовоспалительных средств в основной и контрольной группах.

Показано, что сравнение частот воздействия в основной и контрольной группах дает нам показатель риска, который и по своей сути, и математически равнозначен относительному риску. Этот показатель, называемый *отношением шансов* (odds ratio), определяется как шанс наличия воздействия в основной группе, деленный на шанс наличия воздействия в группе контроля:

$$\frac{[A/(A+C) / C/(A+C)]}{[B/(B+D) / D/(B+D)]}$$

Упрощенно:

$$\frac{A/C}{B/D} \text{ или } \frac{AD}{BC}$$

Итак, отношение шансов можно вычислить, перемножив по обеим диагоналям данные в четырехпольной таблице, а затем разделить полученные произведения одно на другое (см. рис. 10.3).

Обратим внимание на то, что если частота воздействия выше в основной группе, то отношение шансов будет больше единицы,

	Основная группа	Контрольная группа	
Было воздействие	A	B	A + B
Не было воздействия	C	D	C + D
	A + C	B + D	

Когортное исследование

Относительный риск =

$$\frac{A/(A+B)}{C/(C+D)}$$

Исследование случай–контроль

Отношение шансов =

$$\frac{A/(A+C)}{C/(A+C)} = \frac{A/C}{B/D} = \frac{AD}{BC}$$

Рис. 10.3. Вычисление относительного риска в когортном исследовании и отношения шансов в исследовании случай–контроль.

что указывает на повышенный риск. Следовательно, чем сильнее связь между воздействующим фактором и заболеванием, тем выше отношение шансов. И наоборот, если частота воздействия в основной группе ниже, отношение шансов будет меньше единицы, что можно расценивать как защитную роль воздействующего фактора. Таким образом, по своему смыслу отношение шансов аналогично относительному риску, полученному в когортных исследованиях. Сходство информации, выраженной через отношение шансов и через относительный риск, приводит к тому, что некоторые исследователи представляют отношение шансов как "относительный риск".

Отношение шансов примерно равняется относительному риску только когда заболеваемость низкая, поскольку требуются некоторые допущения в вычислениях. Насколько низкой должна быть заболеваемость? Ответ отчасти зависит от величины относительного риска [9]. В принципе, искажение относительного риска становится достаточно большим, когда частота заболевания в группе без воздействия превышает 1/100. К счастью, большинство заболеваний, особенно те, которые изучаются в исследованиях случай–контроль, наблюдаются гораздо реже.

Преимущества исследований случай – контроль

Исследование случай–контроль стало распространенным и важным методом изучения этиологии и клинического течения заболеваний. Его преимущества заключаются прежде всего в том, что исследователи могут выявлять случаи и делать сравнения, независимо от распространенности заболевания. Когортные исследования для такой цели не слишком эффективны. Например, чтобы получить информацию о риске применения нестероидных противовоспалительных препаратов у 100 пациентов с хронической почечной недостаточностью, придется проследить когорту из 1 000 000 человек, принимающих эти препараты в течение 2,5 лет (см. табл. 10.1). Понятно, что из-за стоимости и организационных трудностей такое исследование вряд ли возможно. В то же время относительно несложно и недорого набрать в клиниках и других лечебных учреждениях 100 или более случаев почечной недостаточности, найти сходные группы без заболевания и сравнить их по частоте применения нестероидных противовоспалительных препаратов в прошлом. При таком подходе можно опросить несколько сотен испытуемых в течение нескольких недель или месяцев, затратив гораздо меньше средств, чем при когортном исследовании.

Реальное преимущество исследований случай–контроль при установлении роли причинных или прогностических факторов состоит в отсутствии необходимости ожидать ответа длительное время. Многие заболевания имеют продолжительный латентный период – промежуток времени от воздействия фактора риска до проявления патологического эффекта. Так, канцерогенность некоторых химических веществ проявляется только через 10–20 лет после воздействия этих факторов. От исследователей и медиков потребовалось бы слишком много терпения, чтобы выяснить, подтвердится ли подозреваемый риск.

Исследования случай–контроль занимают все более значимое место в медицинской литературе благодаря тому, что позволяют решать важные вопросы быстро и эффективно. Для тех, кто хочет выяснить причинно-следственную связь с помощью относительно надежного метода, исследование случай–контроль – единственно возможный подход к изучению ряда болезней. Доля исследований случай–контроль среди всех научных публикаций постоянно растет и составляет большую часть эпидемиологических статей. Скорость выполнения и дешевизна оправдывают их популярность в той мере, в какой результаты этих исследований достоверны. Однако здесь существует проблема, поскольку результаты исследований случай–контроль особенно подвержены

систематическим ошибкам. Эти ошибки обсуждаются в следующем разделе.

Как избежать систематических ошибок в исследованиях случай – контроль

Во многих исследованиях случай–контроль авторы сами формируют группы сравнения, вместо того чтобы, как это бывает в когортных и одномоментных исследованиях, позволить это сделать природе. Этот элемент искусственности при формировании групп составляет неизбежное зло, поскольку достоверность исследований случай–контроль зависит от сопоставимости основной и контрольной групп.

Основная и контрольная группы сопоставимы, если контрольную группу можно рассматривать как основную при условии развития у ее участников изучаемого состояния. Другими словами, чтобы быть сопоставимыми, члены основной и контрольной групп должны принадлежать к одной и той же общей популяции. Второе, более спорное положение: чтобы быть сопоставимыми, пациенты основной и контрольной групп должны иметь равную возможность подвергаться воздействию изучаемого фактора [10]. Например, вероятность применения нестероидных противовоспалительных препаратов (см. ранее), по-видимому, выше у тех, кто регулярно лечится и, очевидно, еще больше у больных с сочетанием симптомов. Обязательно ли пациенты в обеих группах, основной и контрольной, должны иметь одинаковые симптомы и получать одинаковую медицинскую помощь? Мнения различны, но следует учитывать, что оценка риска будет невозможна, если пациенты из разных групп будут лечиться у врачей, придерживающихся разной тактики назначения нестероидных противовоспалительных препаратов.

Таким образом, для обеспечения сопоставимости основной и контрольной групп требуется учитывать все обстоятельства, при которых больной подвергался воздействию изучаемого фактора.

Формирование основной группы

Раньше основные группы в исследованиях случай–контроль, как правило, набирались из больных, наблюдавшихся в клиниках или других лечебных учреждениях. Создание регистров заболеваний, таких как Система наблюдения за онкологическими больными Национального института рака США, и компьютерных баз диагностических данных сделали реальным отбор представительной выборки из всех случаев, возникших в определенной популяции.

Случаи, выделенные из общей популяции, должны быть более типичными и включать более широкий спектр проявлений заболевания.

Больные, которые отбираются в основную группу для исследования случай–контроль, должны по возможности иметь недавно возникшее заболевание (для оценки заболеваемости), а не уже давно существующее (для оценки распространенности). Причины этого обсуждались в главе 4. Распространенность заболевания на некоторый момент времени является функцией двух переменных – заболеваемости и продолжительности болезни. Продолжительность в свою очередь определяется скоростью, с которой пациенты лишаются статуса больного (в связи с выздоровлением или смертью) или сохраняют его из-за медленного течения заболевания либо паллиативного характера лечения. Вследствие этих причин, факторы риска для распространенности заболевания могут одновременно быть факторами риска как для заболеваемости, так и для продолжительности болезни или для того и другого вместе; их относительные вклады невозможно различить. Воздействие, приводящее к развитию смертельной формы заболевания уменьшает его распространенность и приводит к снижению относительного риска, оцениваемого на основе распространенности.

Если отношения шансов, полученные для болезней с малой и большой продолжительностью, одинаковы, то читатель может быть хотя бы отчасти уверен, что результаты исследования случай–контроль не подверглись влиянию систематической ошибки вследствие отбора.

Формирование контрольной группы

Главный источник систематических ошибок в исследованиях случай–контроль в том, что контрольная группа представляет собой не естественно сформировавшуюся группу, а специально создается исследователем. Какой контроль считается сопоставимым с основной группой?

Существует несколько способов правильного формирования контрольной группы. Лучший способ минимизировать связанную с отбором систематическую ошибку – формировать обе группы, основную и контрольную, из одной и той же популяции. Если основная группа охватывает все случаи или представляет собой несмещенную выборку из всех случаев, возникающих в популяции (накапливаемых в когортном исследовании или выявленных в одномоментном исследовании), то группа контроля должна представлять собой случайную выборку из всех других лиц той же популяции. Такая тактика называется *популяционным* (population-

based), или *гнездовым* (nested) исследованием случай–контроль. Контрольная группа должна формироваться по тем же критериям включения/исключения, что и основная группа, и должна быть отобрана из популяции или когорты примерно в то же время, в которое отбирается экспериментальная группа.

Пример. Предупреждает ли регулярная интенсивная физическая активность риск остановки сердца у лиц без явного заболевания сердечно-сосудистой системы? Наличие информационной системы неотложной медицинской помощи облегчило выполнение популяционного исследования случай–контроль для ответа на этот вопрос [11]. Больные были отобраны из 1250 жителей Сиэтла и пригородной зоны графства Кинг, штат Вашингтон, США, которые в течение определенного периода времени перенесли вне стационара остановку сердца. Данные собирались из отчетов персонала скорой помощи.

Контрольная группа была сформирована по случайным выбранным телефонным номерам в той же местности (большинство жителей имели домашние телефоны). Обе группы, основная и контрольная, должны были удовлетворять единым критериям включения: возраст 25–75 лет; отсутствие клинически распознаваемого заболевания сердца; отсутствие предшествующего заболевания, ограничивающего физическую активность; наличие супруга/супруги, которые могли сообщить информацию о привычной физической нагрузке, т.е. об изучаемом воздействии. Участники контрольной группы соответствовали участникам основной группы по возрасту, полу, семейному положению и месту жительства (город или пригород). Супруги участников обеих групп опрашивались о способе проведения досуга. Критерии включения обеспечивали принадлежность участников обеих групп к одной и той же популяции и равную степень физической активности.

Результаты, основанные на изучении 163 лиц, включенных в основную и контрольную группы, подтвердили находки предшествующих исследований. Риск остановки сердца у лиц с интенсивной физической активностью во время досуга оказался на 65–75% ниже, чем у лиц, ведущих малоподвижный образ жизни.

Хотя отбор основной и контрольной групп из определенной популяции или когорты предпочтителен, часто более реальным оказывается отбор обеих групп в клиниках или других медицинских учреждениях. Однако такой путь опаснее, поскольку контингент больных в лечебных учреждениях, как правило, представляет собой смещенную выборку из популяции, на которую должны распространяться результаты исследования.

Второй метод формирования сопоставимой контрольной группы включает в себя подходы, иллюстрированные примерами в этой главе и изложенные в главе 6: наложение ограничений, подбор пар, стратификация и стандартизация. Наиболее сложна процедура подбора пар, которая обсуждается далее.

Метод подбора пар при формировании основной и контрольной групп заключается в том, что каждому испытуемому, отнесенному к основной группе, соответствует один или несколько испытуемых с такими же характеристиками в контрольной группе. Пары обычно подбираются по возрасту, полу и месту жительства, так как эти характеристики часто связаны с заболеванием. Однако если известны другие важные факторы, то при подборе пар не следует ограничиваться демографическими данными. Подбор пар увеличивает полезную информацию, поскольку уменьшает различия между основной и контрольной группами по другим (помимо изучаемого) факторам, влияющим на развитие заболевания. Это повышает чувствительность метода для выявления причинно-следственных связей. Но процедура подбора пар сопряжена с риском. Если исследователь подбирает пару по фактору, связанному с изучаемым воздействием, то повышается шанс, что участники такой пары будут иметь одинаковый анамнез в отношении этого воздействия. Например, если пары подбирали по наличию симптомов артрита, по поводу которого обычно назначают нестероидные противовоспалительные препараты, то возрастает вероятность, что такие пары будут иметь одинаковый анамнез в отношении применения этих лекарственных средств. Данный процесс, называемый *переуравновешиванием* (overmatching), вызывает смещение отношения шансов по направлению к единице и ослабляет способность исследования выявлять статистически значимое увеличение или уменьшение отношения шансов.

Третий подход состоит в формировании нескольких контрольных групп. Вследствие трудностей подбора истинно сопоставимых контрольных групп, при использовании одной контрольной группы может возникнуть систематическая ошибка при оценке отношения шансов. Для того чтобы этого не произошло, можно отобрать несколько контрольных групп из различных источников¹. При отборе основной группы в клинике имеется возможность сформировать одну контрольную группу из других больных той же клиники, а вторую – из лиц, живущих по соседству с пациентами, включенными в основную группу. Если при использовании разных контрольных групп получатся одинаковые отношения шансов, это свидетельствует об отсутствии систематической ошибки, так как маловероятно, чтобы смещение в

¹ Формирование нескольких контрольных групп для одной основной группы отличается от подбора нескольких контрольных испытуемых для одного испытуемого основной группы. Последнее делается для повышения статистической мощности (или точности оценки относительного риска). В принципе, подбор нескольких контрольных испытуемых к каждому испытуемому основной группы приводит к небольшому, но полезному повышению статистической мощности, однако увеличение этого числа свыше трех-четырех не дает особого эффекта.

одинаковом направлении и в одинаковой степени повлияло на результаты, полученные с участием отличающихся друг от друга контрольных групп. Если же оценки относительных рисков окажутся разными, то это признак того, что при подборе одной или нескольких групп не была учтена систематическая ошибка; при этом существует возможность разобраться, в чем она состояла.

Пример. В исследовании случай–контроль, проводившемся для выявления возможной связи между приемом эстрогенов и развитием рака эндометрия, основная группа набиралась в пределах одной университетской клиники. Были созданы две контрольные группы: одна – среди гинекологических больных той же клиники, вторая – из случайной выборки женщин, живущих в районе, обслуживаемом клиникой.

Сопутствующие заболевания, например артериальная гипертензия, сахарный диабет и желчнокаменная болезнь, встречались намного чаще в основной группе и в больничной контрольной группе, что отражает, очевидно, поводы для госпитализации. Несмотря на эти различия, в обеих контрольных группах длительный прием эстрогенов отмечался гораздо реже, чем в основной группе, и для них были получены очень близкие отношения шансов (4,1 и 3,6).

Авторы [12] заключили, что такое постоянство результатов при двух совершенно разных группах сравнения свидетельствует об отсутствии систематической ошибки и о достаточной точности результатов.

Возможные варианты подбора основной и контрольной групп обобщены на рис. 10.4. Если вся основная группа набрана в определенной популяции (или в представительной выборке из всех случаев), то лица контрольной группы должны быть набраны из этой же популяции. Такова оптимальная ситуация. Если основная группа является смещенной выборкой из всех случаев, как это и бывает для большинства больничных выборок, то контрольная группа должна формироваться с таким же смещением.

Оценка воздействия

Даже если удалось избежать систематической ошибки при создании основной и контрольной групп, перед исследователем встает задача оценки достоверности воздействия после того, как заболевание или исход уже произошли. Исследования случай–контроль подвержены систематическим ошибкам при оценке достоверности воздействия. Это связано с тем, что воздействие оценивается после начала изучаемого заболевания или возникновения исхода. Различают три вида таких систематических ошибок:

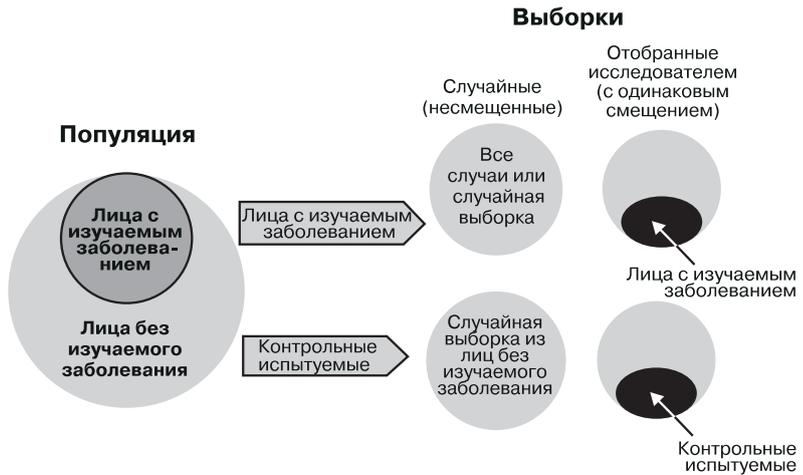


Рис. 10.4. Две тактики создания основной и контрольной групп из общей популяции; несмещенные выборки и выборки с одинаковым смещением.

1. Влияние исхода на воздействие.
2. Влияние исхода на воспоминания о воздействии.
3. Влияние исхода на оценку или регистрацию воздействия.

Систематическая ошибка первого вида особенно важна в ситуации, когда в качестве воздействия изучается лечение, поскольку решение о его назначении может быть продиктовано ранними проявлениями болезни. Это систематическая ошибка, вызванная наличием показаний к лечению. Обратимся к примеру. Исследование случай–контроль было проведено для того чтобы установить, предупреждает ли прием бета-блокаторов развитие инфаркта миокарда у пациентов, лечившихся по поводу артериальной гипертензии [13]. Поскольку стенокардия – главное показание к назначению бета-блокаторов, авторы исключили всех пациентов с указаниями на стенокардию в анамнезе и с другими проявлениями ИБС. Результаты показали, что в группе больных с артериальной гипертензией, леченных бета-блокаторами, риск развития не смертельного инфаркта миокарда значительно снижился даже после исключения всех пациентов со стенокардией или другими проявлениями ИБС.

Систематическая ошибка второго вида возникает из-за того, что воспоминания больных о воздействии изучаемого фактора отличаются от воспоминаний лиц контрольной группы. При всеобщей доступности сведений о возможном риске в связи с экологическими факторами или применением лекарств вполне естественно, что заболевшие люди лучше вспоминают факты, предшествовавшие развитию болезни, чем лица без соответствующего заболевания. Эту ошибку следует учитывать, когда оценка воздействия основана на сборе анамнеза.

Существуют две меры защиты от такой систематической ошибки. Во-первых, нужно иметь альтернативный источник той же информации – документы (история болезни или другие письменные данные) или результаты опроса родственников либо других осведомленных лиц. Во-вторых, цель исследования должна быть скрыта от обследуемых лиц. Разумеется, неэтично вообще не информировать их об общей сути исследования. Однако подробный рассказ о конкретных гипотезах может исказить получаемые результаты и тем самым привести к другому нарушению этики – вовлечению людей в бесполезный исследовательский проект.

Третья проблема – влияние исхода на оценку или регистрацию воздействия должна быть знакома всем, кто изучал приемы диагностики. Если врач, который обследует поступившего в стационар пациента с заболеванием почек, знает о возможной связи между применением нестероидных противовоспалительных препаратов и развитием почечной недостаточности, то можно ожидать, что он специально расспросит пациента об использовании анальгетиков и подробно зафиксирует эту информацию. Исследователь, знающий цель исследования и исходы у обследуемых лиц, может проявить предвзятость при сборе анамнеза. Меры предупреждения этого вида систематической ошибки те же, что и упомянутые выше: множественные источники информации и слепой метод сбора данных, т.е. тот, кто проводит сбор данных, не должен быть осведомлен о гипотезе, проверяемой в исследовании.

Научные стандарты для исследований случай – контроль

Для того чтобы судить о достоверности исследования случай–контроль, предлагается сначала рассмотреть, каким образом тот же вопрос решался бы в рандомизированном контролируемом испытании [14]. Конечно, в действительности выполнить исследование таким способом нельзя, но рандомизированное контролируемое испытание послужит научным стандартом при обсуждении последствий компромиссов, присущих исследованию случай–контроль.

В клиническое испытание обычно включают только тех больных, которые выражают согласие на участие в исследовании. Аналогично, для исследования случай–контроль нужно подбирать тех лиц, которые в принципе могли подвергнуться воздействию. Например, в исследование нестероидных противовоспалительных препаратов в качестве возможной причины развития почечной недостаточности следует включать лиц обоего пола, не имеющих противопоказаний к назначению этих препаратов (таких как язвенная болезнь, например). Так же как в клиническом испытании, все лица из обеих групп должны быть подвергнуты одинаковому обследованию для выявления болезни почек. Эти и другие параллели между клиническими испытаниями и исследованиями случай–контроль могут быть использованы, чтобы разобраться, какие ошибки могли произойти, насколько они серьезны и как их устранить.

Для установления критериев достоверности исследований случай–контроль предпринимались определенные усилия [15]. Однако, чтобы применять эти критерии, требуется глубокое понимание множества факторов, потенциально влияющих на воздействие и заболевание, а также на выявление того и другого в реальных клинических ситуациях.

Исследования случай – контроль при анализе системы здравоохранения

Исследования случай–контроль используются прежде всего для проверки гипотез об этиологии заболевания. В последнее время эта методика стала применяться для изучения вопросов, связанных с обеспечением и качеством медицинской помощи.

Пример. Можно ли предупредить детский церебральный паралич и внутриутробную смерть плода? Британские авторы [16] в исследовании случай–контроль сравнивали данные об антенатальном лечении 141 младенца с церебральным параличом и 62 младенцев, умерших в интранатальном или неонатальном периоде (основная группа). На каждый случай болезни или смерти были подобраны два здоровых младенца, родившихся в то же время и в том же месте (контрольная группа). Недостаточные меры, направленные на коррекцию патологического состояния плода, отмечены чаще в основной, чем в контрольной группе, но этим удавалось объяснить лишь очень небольшой процент случаев детского церебрального паралича.

Поскольку наиболее серьезные последствия плохого качества медицинской помощи относительно редки, исследование случай–контроль можно рекомендовать как эффективную тактику

для изучения связи между отклонением от официальных рекомендаций и неблагоприятными последствиями этого.

Резюме

Прогресс медицины в значительной степени зависит от тщательного исследования случаев заболевания. Описание случаев – это исследование нескольких случаев (10 или менее). Они обеспечивают описание необычных проявлений болезни, помогают изучать патогенез и разрабатывать гипотезы об этиологии и способах лечения. Однако описания случаев особенно подвержены случайным и систематическим ошибкам. Исследование серий случаев – это исследование групп из большего числа больных. Недостатки заключаются в отсутствии контрольной группы, с которой можно было бы сравнивать результаты, а также в том, что случаи изучаются в разные моменты времени и на разных стадиях течения заболевания.

В исследованиях случай–контроль группа лиц с изучаемым заболеванием (основная группа) сравнивается со сходной группой лиц, не имеющих данного заболевания (контрольная группа). Главное преимущество такой методики заключается в возможности набирать основную группу в специализированных клиниках или из регистров заболеваний, а не искать или ожидать появления случаев в определенной популяции риска.

Поэтому исследования случай–контроль стоят гораздо дешевле и выполняются быстрее, чем когортные исследования. Кроме того, они являются единственной возможной тактикой изучения факторов риска редких заболеваний. Хотя частоту новых случаев или относительный риск невозможно рассчитать прямо, их можно оценить по отношению шансов. Все недостатки исследований случай–контроль обусловлены их подверженностью систематическим ошибкам. Эта проблема связана прежде всего с двумя характеристиками исследования случай–контроль. Во-первых, группы сравнения обычно формируются исследователем, а не образуются естественным путем; во-вторых, воздействие оценивается после того, как заболевание уже случилось.

Каково место исследований случай–контроль в клинической эпидемиологии с учетом их подверженности систематическим ошибкам? Некоторые считают такие исследования ненаучными, нелогичными и даже порочными. Другие рассматривают их как важный первый этап в изучении многих медицинских проблем. Почти все согласны с тем, что когортные исследования обеспечивают более строгие, более достоверные доказательства и что в ситуациях, при которых они осуществимы, когортные исследования

являются методом выбора. Тем не менее при соответствующем внимании к возможным источникам систематических ошибок, исследования случай–контроль могут служить надежным и эффективным подходом для поиска ответов на многие вопросы как клинические, так и касающиеся организации здравоохранения.

Литература

1. Klatskin G., Kimberg D.V. Recurrent hepatitis attributable to halothane sensitization in an anesthetist. *N Engl J Med* 1969;280:515–522.
2. Roush M.K., McNutt R.A., Gray T.F. The adverse effect dilemma: quest for accessible information. *Ann Intern Med* 1991;114:298–299.
3. Mulvihill J.J. Clinical ecogenetics. Cancer in families. *N Engl J Med* 1985;312:1569–1570.
4. Berkson J. Limitations of the application of fourfold table analysis to hospital data. *Biomed Bull* 1946;2:47–53.
5. Jaffe H.W., Bregman D.J., Selik R.M. Acquired immune deficiency syndrome in the United States: the first 1,000 cases. *J Infect Dis* 1983;148:339–345.
6. Jensen M.C., Brant-Zawadzki M.N., Obuchowski N., Modic M.T., Malkasian D., Ross J.S. Magnetic resonance imaging of the lumbar spine in people without back pain. *N Engl J Med* 1994;331:69–73.
7. Sandler D.P., Burr F.R., Weinberg C.R. Non-aspirin nonsteroidal antiinflammatory drugs and risk of chronic renal disease. *Ann Intern Med* 1991; 115:165–172.
8. Schlesselman J.J. Case control studies: design, conduct, analysis. New York: Oxford University Press, 1982.
9. Feinstein A.R. The bias caused by high values of incidence for P1 in the odds ratio assumption that 1-P1 1. *J Chron Dis* 1986;39:485–487.
10. Wacholder S., McLaughlin J.K., Silverman D.T., Mandel J.S. Selection of controls in case-control studies. *Am J Epidemiol* 1992;135:1019–1028.
11. Siscovick D.S., Weiss N.S., Hallstrom A.P., Inui T.S., Peterson D.R. Physical activity and primary cardiac arrest. *JAMA* 1982;248:3113–3117.
12. Hulka B.S., Fowler W.C. Jr., Kaufman D.G., Greenberg B.G., Hogue C.J.R., Berger G.S., Pulliam C.C. Estrogen and endometrial cancer: cases and two control groups from North Carolina. *Am J Obstet Gynecol* 1980;137:92–101.
13. Psaty B.M., Koepsell T.D., LoGerfo J.P., Wagner E.H., Inui T.S. Beta-blockers and primary prevention of coronary heart disease in patients with high blood pressure. *JAMA* 1989; 261:2087–2094.
14. Feinstein A.R., Horwitz R.L. Double standards, scientific methods and epidemiologic research. *N Engl J Med* 1982;307:1611–1617.
15. Horwitz R.L., Feinstein A.R. Methodologic standards and contradictory results in case control research. *Am J Med* 1979;66:556–564.
16. Gaffney G., Sellers S., Flavell V., Squier M., Johnson A. Case-control study of intrapartum care, cerebral palsy, and perinatal death. *Br Med J* 1994;308:743–750.

Рекомендуемая литература

- Feinstein A.R. Clinical biostatistics XX: the epidemiologic trohoc, the ablative risk ratio, and "retrospective research". *Clin Pharmacol Ther* 1973; 14:291–307.
- Feinstein A.R., Horwitz R.I., Spitzer W.O., Battista R.N. Coffee and pancreatic cancer: the problems of etiologic science and epidemiological case-control research. *JAMA* 1981;246:957–961.
- Hayden G.F., Kramer M.S., Horwitz R.I. The case-control study. A practical review for the clinician. *JAMA* 1982; 247:326–331.
- Ibrahim M.A., Spitzer W.O. The case-control study: consensus and controversy. New York: Pergamon Press, 1979.
- Rothman K.J. Modern epidemiology. Boston: Little, Brown & Co., 1986.

11 ПРИЧИНА

Пример. Несколько лет тому назад студентам-медикам предложили прокомментировать результаты исследования, выявившего корреляцию между состоянием новорожденных и привычкой акушеров к курению. Состояние новорожденных оценивали по шкале Апгар; высокая оценка (9 – 10 баллов) указывала на то, что младенец здоров, низкая оценка означала возможность патологии и необходимость тщательного наблюдения. Результаты показали, что если акушеры курят (разумеется, не в родильном блоке!), то состояние новорожденных по шкале Апгар хуже. Когда студентов спросили, какая ошибка была допущена в этом исследовании, то после многих предположений кто-то, наконец, сказал, что вывод просто лишен смысла.

После этого организаторы эксперимента признались, что в предложенном студентам примере воздействие и заболевание были специально изменены. На самом деле цифры были взяты из исследования, которое опубликовал в 1843 г. Оливер Уэнделл Холмс (в то время профессор анатомии и физиологии, а позднее декан медицинского факультета Гарвардского университета). В этом исследовании сопоставлялись привычка акушеров мыть руки и развитие сепсиса у родильниц. Холмс [1] пришел к заключению: "Болезнь, известная как послеродовая лихорадка, является заразной, поскольку часто врачи и средний медицинский персонал переносят ее от одной больной к другой".

Отклик одного из врачей середины XIX века на утверждение О. Холмса о том, что немытые руки служат причиной послеродовой лихорадки, удивительным образом напоминает вышеупомянутую реакцию студентов-медиков: наблюдение лишено смысла. "Я скорее буду приписывать их (случаи заболевания послеродовым сепсисом) несчастному случаю или воле Провидения – по крайней мере, в этом есть хоть какой-то смысл, нежели объяснять их заражением, из чего невозможно сформировать никакой ясной концепции, по крайней мере, относительно этой болезни", – писал доктор Чарльз Д. Мейгс, профессор акушерства, женских и детских болезней медицинского колледжа им. Т. Джефферсона [1].

Итак, Холмс и Мейгс разошлись во мнениях о причине. Холмс, исходя из своих данных, был уверен: распространение послеродового сепсиса обусловлено тем, что акушеры не мыли руки

между приемами родов. Однако он не мог объяснить, каким образом немытые руки могут стать причиной болезни, поскольку возбудитель еще не был известен. Поэтому Мейгс не поверил, что причина послеродового сепсиса раскрыта (и, наверное, продолжал не утруждать себя мытьем рук).

Клиницисты часто встречаются с информацией о возможных причинно-следственных связях. В этой книге очень часто обсуждался вопрос о методах установления причинно-следственных связей, но до сих пор не было уделено специального внимания понятию причины.

В этой главе будет рассмотрено понятие причины в клинической медицине. Затем мы покажем, какого рода доказательства дают основания предполагать, что связь между явлениями, или корреляция, обусловлена причинно-следственными отношениями. Кроме того, кратко коснемся одного из типов исследований, который еще не описывался в нашей книге: это исследования, в которых известно воздействие возможной причины на группу в целом, а не на отдельных членов этой группы.

Понятие причины

Словарь Вебстера [2] определяет причину, как "то, что вызывает эффект или результат". В медицинских учебниках причина обычно обсуждается в связи с такими понятиями, как "этиология", "патогенез" или "факторы риска".

Для практикующих врачей причина важна прежде всего потому, что она определяет подходы к решению трех клинических задач: профилактике, диагностике и лечению. Клинический пример в начале этой главы демонстрирует, как знание причинно-следственных связей может помогать в деле профилактики. Когда врач периодически проверяет артериальное давление, он исходит из доказанности того, что повышенное артериальное давление – причина заболеваемости и смертности, и что нормализация артериального давления предупреждает развитие инсульта и инфаркта миокарда. Диагностика, особенно при инфекционном заболевании, часто включает в себя поиск причинного воздействия. Информация о причине нужна также в ситуации, когда по наличию факторов риска выделяют группы с высокой распространенностью заболевания (см. главу 3). В принципе, представление о причинной связи лежит в основе любого лечебного воздействия в клинической медицине. Разве мы назначили бы при пневмококковой пневмонии пенициллин, если бы не предполагали, что это приведет к выздоровлению? Или зачем бы мы советовали проводить химиотерапию больному с метастазами, если бы не считали,

что подавление метаболизма раковых клеток обеспечит регрессию метастазов, и, в результате, увеличение продолжительности и качества жизни?

По большому счету врачи больше заинтересованы в выяснении устранимых или обратимых причин болезней, чем таких, с которыми ничего нельзя поделать. С другой стороны, исследователи изучают и такие причинные факторы, в отношении которых нет эффективного лечения или профилактики, в надежде разработать эти меры в будущем.

Причины единичные и множественные

В 1882 г., через 40 лет после полемики между О. Холмсом и Ч. Мейгсом, Роберт Кох разработал правила для определения роли инфекционного агента как причины болезни. Основой его подхода стало предположение о том, что каждая болезнь имеет одну причину, а конкретная причина вызывает одну болезнь. Сформулированные им постулаты таковы:

1. Микроорганизм должен быть обнаружен в каждом случае заболевания.
2. Микроорганизм должен быть выделен и выращен в виде чистой культуры.
3. Микроорганизм должен вызывать специфическое заболевание при введении животному.
4. Микроорганизм должен быть затем выделен у животного и идентифицирован.

Интересно, что при установлении причины Кох не учитывал эффект лечения. Наверное он бы добавил эффект лечения к своим критериям, если бы жил на сто лет позже, когда эффективное лечение стало более частым явлением.

Подход Коха способствовал существенному развитию понятия причины в медицине. До этого считалось, что одну и ту же болезнь могут вызывать разные бактерии. Постулаты Коха помогли внести порядок в существовавший хаос. Они полезны и сегодня. Постулат о том, что конкретный микроорганизм является причиной заболевания, послужил основой для недавних открытий. В 1977 г. была обнаружена грамотрицательная бактерия, вызывающая болезнь легионеров, а в 1980 г. было установлено, что СПИД обусловлен вирусом иммунодефицита человека (ВИЧ).

Однако для установления причины большинства болезней недостаточно одних только постулатов Коха. Иногда слишком строгое следование правилам Коха вело медицину в тупик. Конечно, было бы удобно, если бы всегда существовало соотношение "одна причина – одна болезнь". Однако курение способствует раз-

витию рака легких, хронических обструктивных заболеваний легких, язвенной болезни, рака желчного пузыря, ишемической болезни сердца. Ишемическая болезнь сердца в свою очередь имеет множественные причины, включая курение, артериальную гипертензию, гиперхолестеринемию и наследственные факторы, а иногда возникает и в отсутствие этих известных факторов риска.

Обычно заболевание вызывается комплексом совместно действующих факторов, сочетанием причин [3]. Такое сочетание хорошо известно, например, для ишемической болезни сердца, но это же справедливо и в отношении инфекционных болезней, для развития которых присутствие микроорганизма необходимо, но не всегда достаточно, чтобы вызвать болезнь. Так, чтобы заболеть СПИДом, необходимо инфицирование ВИЧ, но при инфицировании не всегда развивается заболевание. Например, контакт с ВИЧ при порезе скальпелем редко приводит к развитию СПИДа (3–4 случая на 1000), так как данный вирус гораздо менее контагиозен, чем, например, вирус гепатита В.

Временные соотношения причины и следствия

Специалисты в области биологии и медицины, изучающие причину, обычно ищут патогенетический механизм или общий конечный путь развития заболевания. Например, серповидно-клеточная анемия обусловлена генетической аномалией молекулы гемоглобина, при которой вследствие полимеризации гемоглобина S происходит серповидная деформация эритроцитов. Выяснение патогенеза заболеваний сыграло очень важную роль в прогрессе медицины в этом столетии.

Однако развитие болезни определяется и менее специфическими, более отдаленными причинами или факторами риска, такими как поведение людей или воздействие окружающей среды. Эти факторы бывают даже более важными причинами болезни, чем патогенетические механизмы. Например, значительная доля случаев смерти от сердечно-сосудистых заболеваний и рака в США в конечном счете сводится к поведенческим и экологическим факторам. Распространение СПИДа связано прежде всего с половым поведением и применением наркотиков, а случаи смерти от насилия и несчастных случаев – с социальными условиями, доступностью оружия, злоупотреблением алкоголем, несоблюдением требований по использованию ремней безопасности.

Если рассматривать причину исключительно на клеточном и субклеточном уровне, то это ограничивает возможность лечебного вмешательства. Даже когда патогенетический механизм неясен, знание факторов риска может обеспечить эффективное лечение и

профилактику. Следовательно, О. Холмс был прав, утверждая, что акушеры должны мыть руки, хотя он не имел и понятия о бактериях.

Как для распространения, так и для контроля многих заболеваний важны патогенетические механизмы и неспецифические факторы риска. Иногда происходит сложное взаимодействие множества различных причин.

Пример. Постулаты Коха впервые были использованы для проверки гипотезы о том, что туберкулез вызывается попаданием бактерии *Mycobacterium tuberculosis* в чувствительный к ней организм. Общий конечный путь развития туберкулеза – внедрение бактерий в ткани организма хозяина. С точки зрения патогенеза, бороться с болезнью можно с помощью антибиотиков или вакцин, эффективных против бактерий. Благодаря биомедицинским исследованиям удалось получить оба средства.

Однако развитие туберкулеза – весьма сложный процесс. Здесь важны такие причины, как чувствительность организма и интенсивность воздействия возбудителя (рис. 11.1). Эти причины определяют, произойдет ли внедрение бактерии в ткани организма хозяина.

Некоторые врачи сомневаются в том, что чувствительность организма хозяина и интенсивность воздействия возбудителя можно считать причинами заболевания туберкулезом, но это очень существенные компоненты этиологии. В самом деле, улучшение социальных и экономических условий, влияющих на восприимчивость организма человека (уменьшение скученности про-

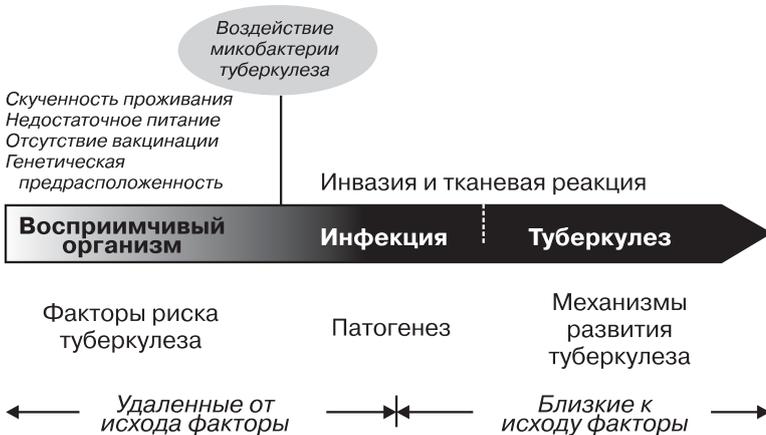


Рис. 11.1. Причины заболевания туберкулезом.

живания и улучшение питания), сыграло более важную роль в снижении заболеваемости туберкулезом в развитых странах, чем методы лечения, разработанные на основе представлений о патогенезе. Так, смертность от туберкулеза в Англии резко снизилась задолго до того, как началось применение антибиотиков (рис. 11.2, А). Вакцина же была разработана еще позже.

Начиная с 1985 г. число случаев заболевания туберкулезом в США стало возрастать [4] (рис. 11.2, Б). Почему это произошло? Во-первых, из общего числа дополнительно заболевших 60% составляют эмигранты. Во-вторых, важным фактором стала ВИЧ-инфекция, распространение которой способствует увеличению числа лиц с ослабленной иммунной системой, восприимчивых к туберкулезу. К тому же эти восприимчивые лица имеют больше шансов, чем представители общей популяции, заразиться туберкулезом, поскольку распространенность обоих заболеваний – СПИДа и туберкулеза, выше в группах населения с худшими экономическими условиями. Наконец, изменился и сам возбудитель, в результате эволюции которого возникли штаммы, устойчивые ко многим лекарствам, причем это усугубляется рядом обстоятельств. Генетические изменения микобактерии более вероятны при несоблюдении схемы лечения [5], которое чаще наблюдается у наркоманов, составляющих важную группу риска для ВИЧ-инфекции. Кроме того, изменениям генома бактерии способствует высокая интенсивность ее размножения в организме хозяина, страдающего иммунодефицитом. Итак, при поисках причины может обнаружиться невероятно сложное взаимодействие разных факторов – среды, поведения, молекулярно-биологических механизмов.

Другим примером необходимости двух подходов – патогенетического и эпидемиологического, служит недавнее снижение смертности от ишемической болезни сердца в США.

Пример. В течение последних двух десятилетий смертность от ишемической болезни сердца снизилась более чем на $\frac{1}{3}$. Это снижение совпало с уменьшением воздействия на все население нескольких факторов риска сердечно-сосудистых заболеваний: больше больных артериальной гипертензией стали получать эффективное лечение, мужчины среднего возраста стали меньше курить, снизилось потребление жиров и холестерина с пищей. Эти достижения, по крайней мере отчасти, были следствием эпидемиологических и биомедицинских исследований и способствовали сохранению десятков тысяч жизней ежегодно. Вряд ли можно было бы добиться такого успеха без понимания как непосредственных механизмов, так и более отдаленных причин сердечно-сосудистых заболеваний [6].

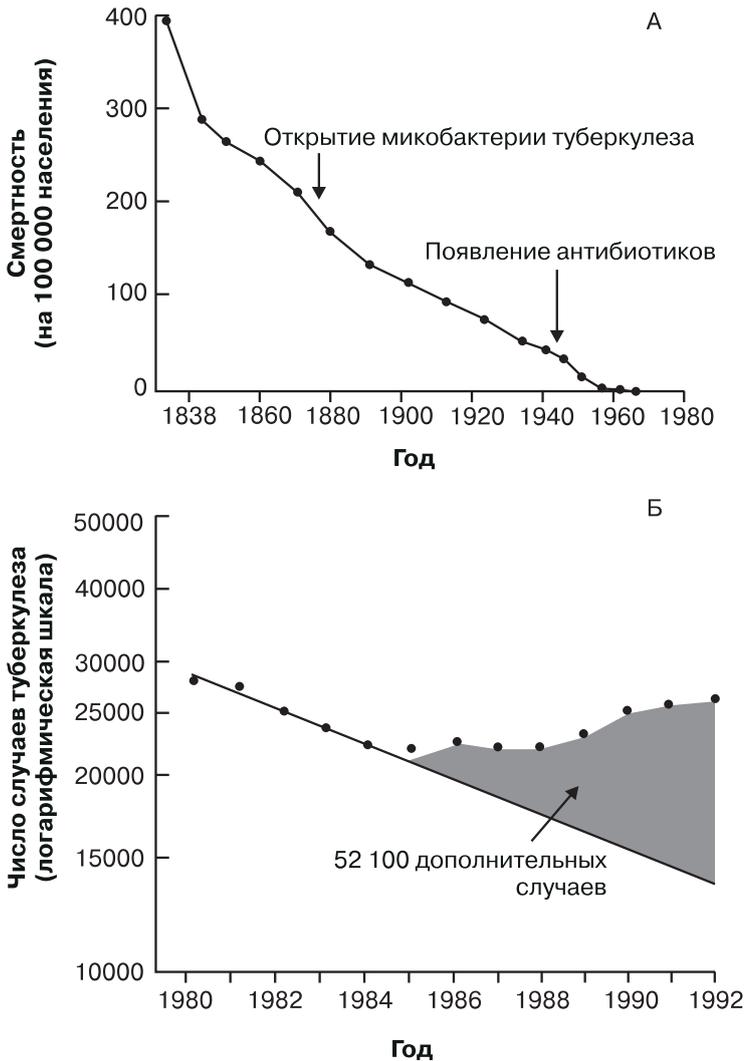


Рис. 11.2. Динамика смертности и заболеваемости туберкулезом.

А – снижение смертности от туберкулеза органов дыхания в Англии и Уэльсе в течение последних 150 лет [McKeown T. The role of medicine: dream, mirage or nemesis. London; Nuffield Provincial Hospital Trust, 1976.]; Б – повышение заболеваемости туберкулезом в США в 1985 – 1992 гг. Разница между ожидаемым и наблюдаемым числом случаев. Пунктирная линия – наблюдаемые случаи; сплошная линия – ожидаемые случаи [Cantwell M.F. et al. Epidemiology of tuberculosis in the United States. 1985 through 1992. JAMA 1994; 272:535–539].

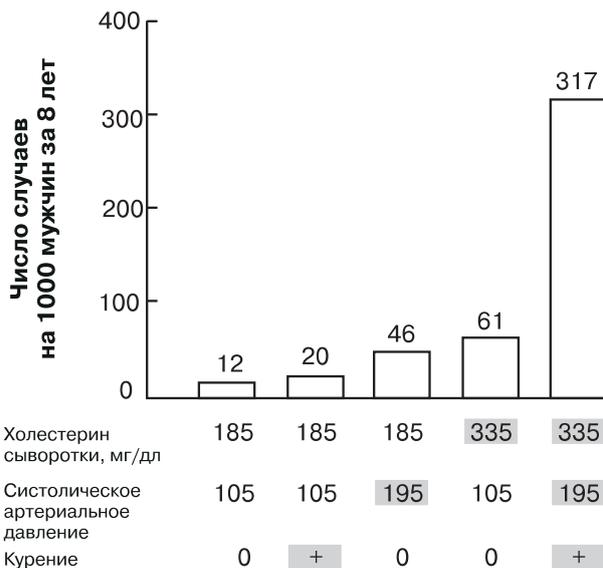


Рис. 11.3. Взаимодействие множественных причин заболевания. Риск развития сердечно-сосудистых заболеваний у мужчин в зависимости от каждого фактора риска в отдельности и от их сочетаний. Величины, выходящие за пределы нормы, находятся в темных клетках [Kannel W.B. Preventive cardiology. Postgrad Med 1977;61:71 – 85].

Взаимодействие многих причин

При одновременном воздействии нескольких причин результирующий риск может быть больше или меньше, чем ожидается при простом суммировании влияния отдельных причин. Если общий результат больше, чем сумма влияний отдельных причин, клиницисты называют это явление синергизмом, а если меньше – то антагонизмом¹.

Пример. На рис. 11.3 показана вероятность развития сердечно-сосудистого заболевания в течение 8-летнего периода у мужчин в возрасте 40 лет. Мужчины, которые не курили, имели низкий уровень холестерина крови и низкие показатели систолического артериального давления, относились к группе низкого риска развития заболевания (12/1000). Риск увеличивался до 20–61/1000, когда эти факторы риска присутствовали по отдельности. Но когда все три фактора действовали одновременно, абсолютный риск развития

¹ О *статистическом взаимодействии* (statistical interaction) говорят, когда комбинации переменных в математической модели дают некую дополнительную информацию по сравнению с той, которая была получена в результате оценки эффектов всех прогностических переменных по отдельности. Это понятие близко к явлениям биологического синергизма и антагонизма, но представляет собой математическое построение, а не природный феномен.

сердечно-сосудистого заболевания (317/1000) почти в 3 раза превышал сумму рисков, связанных с отдельными факторами [7].

Выяснение причины затрудняется, когда действует не один, а много факторов. Если имеется много причинных факторов и они взаимодействуют между собой, то можно существенным образом повлиять на состояние здоровья пациента, изменив только одну или несколько причин. В только что приведенном примере у больных, которые прекратят курение и будут соблюдать назначения врача по нормализации артериального давления, риск развития сердечно-сосудистых заболеваний может заметно уменьшиться, даже если другие факторы останутся без изменений.

Модификация эффекта

Особый тип взаимодействия – *модификация эффекта* (effect modification). Это означает, что сила корреляции между двумя переменными зависит от уровня некоторой третьей переменной, называемой *модификатором эффекта* (effect modifier).

Пример. Поскольку существовали противоречивые данные об эффективности тиазидных диуретиков в профилактике ишемической болезни сердца, было проведено исследование, целью которого являлось определение соотношения между дозой тиазидного диуретика и риском внезапной смерти, а также влияние на эффект тиазидов дополнительного назначения калийсберегающих диуретиков. Обобщенные результаты представлены на рис. 11.4. Эффект тиазидных диуретиков определяется их дозой: прием низкой дозы (25 мг) предотвращает, тогда как прием высокой дозы (100 мг) увеличивает шансы внезапной смерти. Дополнительное назначение калийсберегающих диуретиков в различных дозах модифицирует эффект тиазидов, усиливая их защитное влияние [8].

Установление причин

В клинической медицине невозможно доказать причинно-следственный характер связи, полностью исключив все сомнения. Можно лишь с помощью эмпирических доказательств повысить уверенность до уровня, при котором причина может считаться доказанной с точки зрения поставленных задач и целей. И, наоборот, данные, отрицающие некоторую причину, следует накапливать до тех пор, пока не будет исключена вероятность существования причинно-следственной связи. Доказательства причинно-следственной связи необходимо проверять самыми разными способами. Поэтому для получения убедительных свидетельств за или против причины приходится выполнить несколько разных исследований.

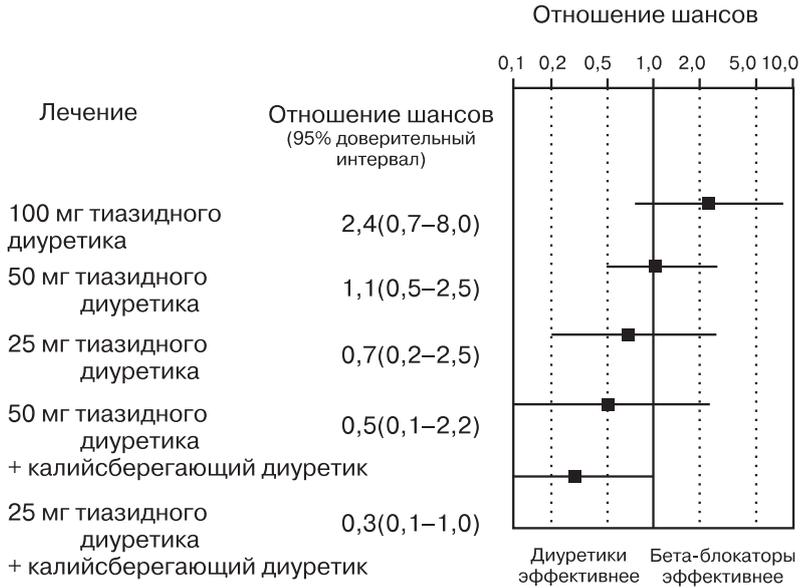


Рис. 11.4. Пример модификации эффекта: риск внезапной смерти при лечении тиазидными диуретиками по сравнению с бета-блокаторами изменяется под влиянием калийсберегающих диуретиков. Отношение шансов увеличивается с повышением дозы тиазидных диуретиков, и это дает основание предположить, что безопаснее использовать бета-блокаторы. Однако при дополнительном назначении калийсберегающих диуретиков риск внезапной смерти при использовании тиазидных диуретиков становится меньше, чем при лечении бета-блокаторами [Siscovick D.S. et al. Diuretic therapy for hypertension and the risk of primary cardiac arrest. N Engl J Med 1994; 330: 1852 – 1857].

Связь и причина

Два фактора могут рассматриваться в качестве причины и следствия, если между ними имеется корреляция. Однако корреляция не всегда бывает обусловлена причинно-следственной связью. На рис. 11.5 представлены другие виды связей, которые нужно исключать. Во-первых, следует выяснить, реально ли существует связь между предполагаемой причиной и следствием или же это просто артефакт вследствие систематической ошибки либо случайности. Корреляции, которые на самом деле не существуют, чаще всего обусловлены случайностью или систематическими ошибками при отборе и измерении. Если вероятность этих ошибок мала, то связь действительно имеется. Однако прежде чем решить, что связь является причинно-следственной, необходимо выяснить, косвенная она (т.е. опосредованна неким другим, вме-

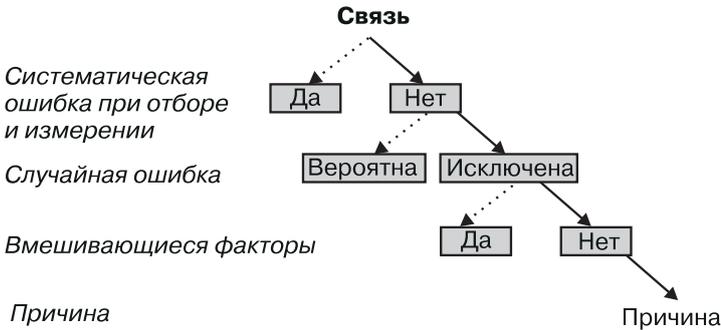
Объяснение**Находка**

Рис. 11.5. Связь и причина.

шивающимся фактором) или прямая. Если вмешивающиеся факторы не обнаружены, то наличие причинной связи вполне возможно.

При этом не исключается, что в будущем может быть найден другой фактор, который окажется более непосредственной причиной. Например, в нескольких исследованиях отмечено, что женщины хуже, чем мужчины, переносят операцию аортокоронарного шунтирования; было сделано предположение, что послеоперационный прогноз связан с полом. В дальнейшем выяснилось, что не женский пол, а меньшая площадь поверхности тела – поскольку она коррелирует с диаметром коронарных артерий – является важной переменной, которая связана с развитием сердечной недостаточности и смертью [9]. Таким образом, иногда факторы, прежде считавшиеся причиной, при появлении новых сведений оказываются связанными с заболеванием лишь опосредованно.

Методы изучения причин

При изучении причинно-следственных связей следует придавать большое значение методике исследования.

Надежное доказательство причинно-следственной связи возможно только путем тщательного проведения рандомизированного контролируемого испытания с соблюдением следующих условий: 1) наличие достаточного числа больных; 2) применение тройного слепого метода; 3) тщательная стандартизация методов сбора

и анализа данных. Метод рандомизированного контролируемого испытания позволяет избежать различий между сопоставляемыми группами как по известным факторам, которые можно учесть и другими способами, так и по вмешивающимся факторам.

Рандомизированные контролируемые испытания обычно используются для доказательства причинных связей при оценке методов лечения и профилактики. Однако, как указывалось в главе 6, с помощью таких испытаний редко удается изучать причины заболеваний. Для этой цели служат обсервационные исследования.

В принципе, чем сильнее отличается методика исследования от рандомизированного испытания, тем меньше это исследование защищено от возможных систематических ошибок, и тем слабее доказательство причинно-следственной связи. С этой точки зрения более предпочтительно правильно выполненное когортное исследование, поскольку оно позволяет минимизировать систематические ошибки, возникающие вследствие отбора, измерений и влияния вмешивающихся факторов. Одномоментные исследования уязвимы, так как не обеспечивают непосредственных данных о последовательности событий. Истинные исследования распространенности – одномоментные исследования целой популяции, защищая от систематической ошибки, возникающей при отборе, но подвержены систематическим ошибкам, обусловленным измерением и влиянием вмешивающихся факторов. Исследования случай–контроль также подвержены систематической ошибке вследствие отбора. Наиболее методически слабы исследования серий случаев, поскольку в них нет ни заранее определенной популяции, ни группы сравнения.

Эта иерархия доказательности исследований, основанная на степени их подверженности систематической ошибке, дает только приблизительный ориентир для выбора. В процессе выполнения каждого конкретного исследования его достоверность можно существенно повысить или понизить независимо от типа исследования.

Популяционные исследования

До сих пор обсуждались методы доказательства причинно-следственных связей в ситуациях, когда воздействие и само заболевание известны для каждого пациента, включенного в исследование. При эпидемиологическом обследовании больших популяций факт воздействия известен только для групп, а не для конкретных членов этих популяций.

Исследования, в которых фактор риска характеризуется средним воздействием на группу, называются исследованиями *совокупного риска* (aggregate risk studies). Другое их название – *эколо-*

гические исследования (ecological studies), поскольку их задача – оценить влияние на людей факторов окружающей среды в целом.

Пример. Какие факторы связаны со смертностью от болезней сердца в развитых странах? St. Leger et al. [10] собрали данные о смертности от ишемической болезни сердца (ИБС) в 18 развитых странах, чтобы изучить роль экономических факторов, медицинского обслуживания и особенностей питания. При этом неожиданно выявилась отрицательная связь между смертностью от ИБС и потреблением вина (рис. 11.6).

На основании исследования была выдвинута гипотеза о том, что алкоголь защищает от развития ИБС. Позднее проведенные исследования на отдельных пациентах показали, что потребление алкоголя увеличивает содержание липопротеидов высокой плотности в сыворотке, предохраняющих от развития сердечно-сосудистых заболеваний.

Результаты исследований совокупного риска редко считаются окончательными. Главная проблема – это потенциальная система-

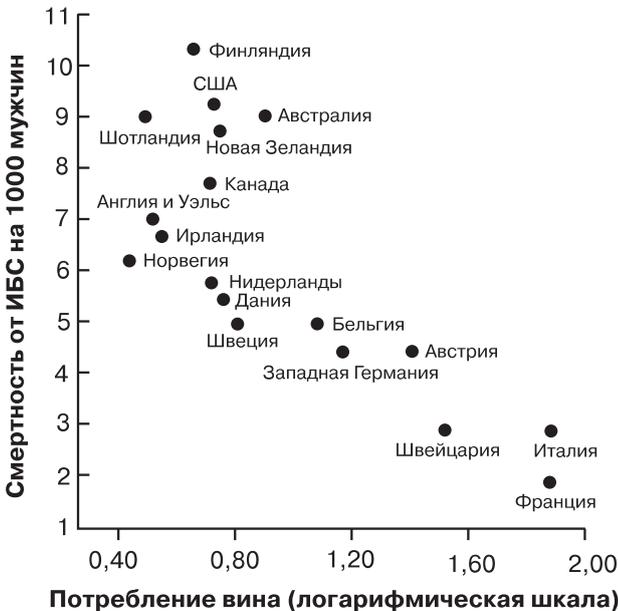


Рис. 11.6. Пример исследования совокупного риска: связь между потреблением вина и смертностью от болезней сердца в экономически развитых странах [St. Leger A.S., Cochrane A.L., Moore F. Factors associated with cardiac mortality in developed countries with particular reference to the consumption of wine. Lancet 1979; 1:1017–1020].

тическая ошибка, называемая *экологической ошибкой* (ecological fallacy): заболевшие лица в общей группе воздействия на самом деле могли и не испытывать влияния факторов риска. Кроме того, воздействие – не всегда единственная характеристика, которая отличает участников группы воздействия от лиц группы, не подвергавшейся воздействию, т.е. не исключено влияние каких-то дополнительных факторов. Таким образом, исследования совокупного риска наиболее полезны для выдвижения гипотез, которые должны затем быть проверены в более строгом исследовании.

Исследования серий случаев во времени

Полученное в исследовании совокупного риска доказательство того, что фактор действительно связан с эффектом, можно усилить, если проводить наблюдения более чем в двух временных точках (до и после воздействия) и в нескольких местах. При исследованиях серий случаев во времени эффект оценивают в разные моменты времени – до и после воздействия предполагаемой причины, а затем определяют, изменяется ли эффект так, как это ожидалось. Если после изменения предполагаемой причины следует изменение предполагаемого результата, то маловероятно, что связь между ними ложная.

Пример. Было показано, что риск развития диареи, вызываемой *Clostridium difficile*, и псевдомембранозного энтероколита связан с приемом антибиотиков, в частности клиндамицина, ампициллина и цефалоспоринов. Повышение назначаемых доз клиндамицина привело в 1990 г. к эпидемии диареи в клинике (рис. 11.7) [11]. Меры по санитарному просвещению, контролю инфекции и гигиене помещений не смогли остановить эпидемию. Заболевания прекратились после того как больные перестали принимать клиндамицин. Для дальнейшего уточнения этой связи авторы провели исследование случай–контроль, подтвердившее результаты исследования серий случаев во времени.

Исследования множественных серий случаев во времени можно проводить, изучая влияние возможной причины в нескольких группах в различные моменты времени. При этом проверяется, вызывает ли предполагаемая причина изменения в разных группах. Регулярное возникновение эффекта вслед за воздействием, происходившим в разные моменты времени и в разных местах, более надежно доказывает причинную роль изучаемого воздействия, чем то же явление, выявленное однократно. В самом деле, трудно представить себе, чтобы какие-то посторонние факторы одинаково влияли на эффект изучаемого воздействия в разное время и в разных местах.

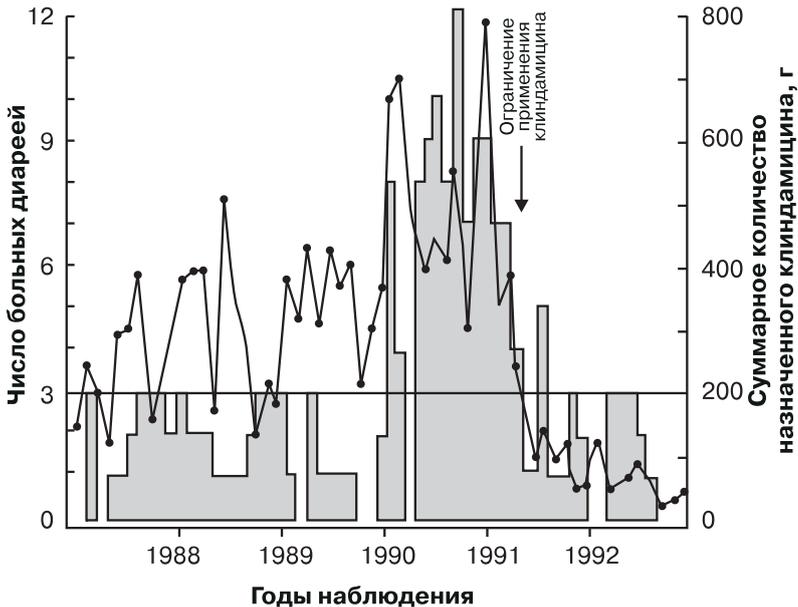


Рис. 11.7. Исследование серий случаев во времени для выявления связи между применением клиндамицина и развитием диарей, вызываемой *Clostridium difficile* [Pear S.M., Williamson T.H., Bettin K.M. et al. Decrease in nosocomial *Clostridium difficile*-associated diarrhea by restricting clindamycin use. Ann Intern Med 1994; 120:272 – 277].

Пример. Из-за отсутствия рандомизированных контролируемых испытаний эффективность скрининга по выявлению рака шейки матки пришлось оценивать с помощью обсервационных исследований. Некоторые из наиболее убедительных доказательств эффективности скрининга были получены путем исследований множественных серий случаев во времени [12]. Данные были получены из программ скрининга, начатых в различных провинциях Канады в разное время в течение 10-летнего периода в 60–70-е годы. Снижение смертности систематически следовало за введением таких программ, независимо от времени и места их внедрения. На основании этих данных было сделано заключение, что "скрининг способствует снижению смертности от рака шейки матки".

Аргументы за и против

В 1965 г. английский статистик сэр Остин Брэдфорд Хилл [13] предложил набор признаков, на основании которых можно решить вопрос о том, является ли связь между болезнью и некото-

рым фактором среды причинно-следственной или только корреляционной. Эти критерии широко используются, иногда с небольшими изменениями (табл. 11.1). Мы кратко прокомментируем отдельные критерии. Не все они имеют одинаковое значение для решения вопроса о причине.

Временное соотношение между причиной и эффектом

Разумеется, причина должна предшествовать следствию. Этот принцип кажется очевидным, но его недооценивают при интерпретации большинства одномоментных исследований, в которых оценку предполагаемой причины и следствия производят в один и тот же момент. Иногда делают допущение, что одно явление предшествует другому, тогда как на самом деле это не установлено. В ряде случаев бывает трудно определить последовательность событий.

Пример. Давно замечено, что при избыточной массе тела риск смерти, особенно от сердечно-сосудистых заболеваний, выше, чем при нормальной массе. Исходя из этого, разумно допустить, что снижение массы тела окажет профилактическое действие у больных с ожирением. Однако в нескольких когортных исследованиях обнаружено увеличение смертности среди лиц, терявших массу тела, даже если до этого они не болели. Эти тревожные наблюдения

Таблица 11.1
Доказательства наличия причинно-следственной связи*

Критерии	Комментарии
Последовательность во времени	Причина предшествует эффекту
Сила связи	Большой относительный риск
Зависимость эффекта от дозы	При усилении воздействия заболеваемость повышается
Обратимость	При ослаблении воздействия заболеваемость снижается
Устойчивость	Эффект наблюдается разными исследователями, независимо от места, условий и времени
Биологическое правдоподобие	Эффект согласуется с современными научными представлениями
Специфичность	Одна причина приводит к одному эффекту
Аналогия	Причинно-следственная связь уже установлена для сходного воздействия или болезни

* [Bradford-Hill A.B. The environment and disease: association and causation. Proc R Soc Med 1965; 58:295–300, с изменениями].

можно объяснить, предположив, что снижение массы тела было доклиническим проявлением смертельного заболевания [14]. В этом случае смертельное заболевание предшествует и вызывает снижение массы тела, а не наоборот. (Для того чтобы исключить эту возможность, нужно знать, было ли похудание преднамеренным.)

Причина должна обязательно предшествовать клиническому эффекту, поэтому нарушение такой последовательности служит сильным аргументом против возможной причинно-следственной связи. Тем не менее само по себе наличие соответствующей последовательности событий во времени – слабое доказательство.

Сила связи

Сильная связь между предполагаемой причиной и эффектом, которая выражается высоким относительным или абсолютным риском, свидетельствует о существовании причинной связи более надежно, чем слабая связь. Как было показано во многих проспективных исследованиях, заболеваемость раком легких среди курящих в 4 – 16 раз выше, чем среди некурящих, тогда как относительный риск развития рака почки, что следует из тех же исследований, невелик (1,1 – 1,6) [15]. Следовательно, связь между курением и развитием рака легкого гораздо сильнее, чем между курением и развитием рака почки.

Аналогичным образом относительный риск выявления вирусного гепатита В при гепатоцеллюлярной карциноме, равный почти 300, оставляет мало сомнений в том, что вирус – причина рака печени [16]. Иногда высокий относительный риск является следствием систематической ошибки. Однако нераспознанная систематическая ошибка чаще приводит к уменьшению, а не к увеличению относительного риска.

Зависимость эффекта от дозы

Зависимость эффекта от дозы означает, что количественные изменения предполагаемой причины сопровождаются количественными изменениями результата. Наличие такой зависимости служит аргументом в пользу существования причинно-следственной связи. На рис. 11.8 показана четкая зависимость эффекта от дозы при соотношении уровня смертности от рака легких (эффект) с числом выкуриваемых сигарет (доза).

Хотя зависимость эффекта от дозы убедительно свидетельствует в пользу причинно-следственной связи, особенно в сочетании с высокими значениями относительного или абсолютного риска, все это не исключает влияния каких-то вмешивающихся факторов.

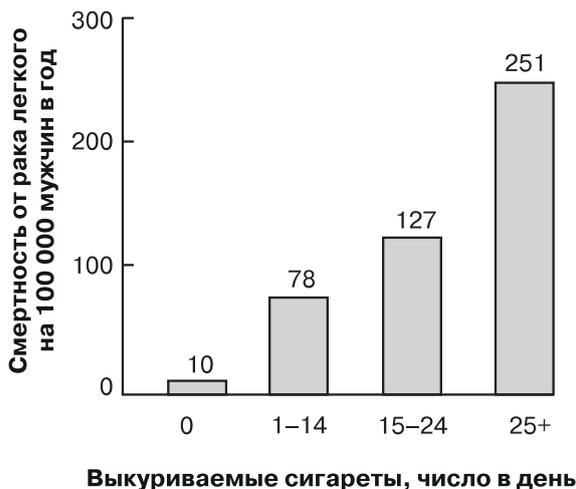


Рис. 11.8. Пример зависимости эффекта от дозы: смертность от рака легких у врачей-мужчин в зависимости от числа выкуриваемых в день сигарет [Doll R., Petro R. Mortality in relation to smoking: 20 years' observations on male British doctors. Br Med J 1976; 2:1525–1536].

Пример. Представители табачной промышленности утверждают, что сильная связь между курением и заболеваемостью раком легкого, а также зависимость эффекта от дозы обусловлены влиянием неких вмешивающихся факторов. Их аргумент таков: некоторая неизвестная переменная (вмешивающийся фактор) служит причиной никотиновой зависимости и одновременно увеличивает риск заболевания раком легкого. Чем в большей степени присутствует этот фактор, тем чаще курение сочетается с заболеванием раком легкого, отсюда – зависимость эффекта от дозы. Теоретически такой аргумент может объяснять наблюдаемую связь между курением и развитием рака легкого, однако природа вмешивающегося фактора совершенно непонятна. Без рандомизированного контролируемого испытания (в котором лица с вмешивающимся фактором будут одинаково распределены между группами курящих и некурящих) доводы о влиянии вмешивающегося фактора трудно полностью опровергнуть.

Обратимость связи

Вероятность того, что фактор является причиной заболевания, возрастает, если при устранении этого фактора риск развития заболевания уменьшается, т.е. связь между предполагаемой причиной и следствием обратима. На рис. 11.9 показано, что у отказавшихся от курения уменьшается вероятность заболеть раком легкого.

Обратимость связи – сильное, но не бесспорное доказательство ее причинно-следственного характера, поскольку и обратимость может объясняться влиянием вмешивающихся факторов. Например, можно предположить (хотя это маловероятно), что люди, желающие бросить курить, имеют меньший уровень некоего неидентифицированного фактора, чем те, кто продолжают курение (см. рис.11.9).

Устойчивость эффекта

Когда несколько исследований, выполненных в разное время, в разных условиях и с различными больными приводят к одному и тому же заключению, доказательство причинного характера связи усиливается. Эффективность скрининга на колоректальный рак была доказана после того, как профилактическое значение скрининга подтвердилось в рандомизированном контролируемом испытании по применению анализа кала на скрытую кровь [17], и в исследовании случай–контроль с проведением сигмоскопии [18]. Особенно убедительным подтверждением причинной связи служит получение одинакового результата разными методами, поскольку в исследованиях с одинаковой методикой возможна одна и та же ошибка.

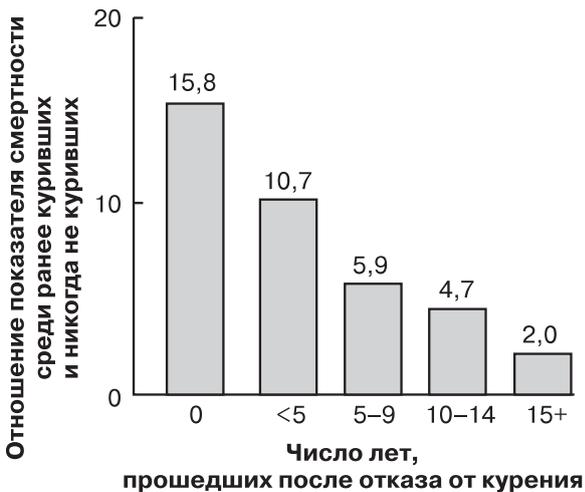


Рис. 11.9. Обратимость связи: снижение смертности от рака легких среди отказавшихся от курения. Исключены лица, которые бросили курить после того как заболели раком [Doll R., Petro R. Mortality in relation to smoking: 20 years' observations on male British doctors. Br Med J 1976;2:1525 – 1536].

Однако разные исследования часто дают противоречивые результаты. Отсутствие постоянства эффекта необязательно перечеркивает результаты конкретного исследования. Одно хорошо выполненное исследование может перевесить несколько плохих.

Биологическое правдоподобие

При оценке причинно-следственных отношений важно учитывать, насколько совместима выявляемая связь с современными знаниями о патогенезе. Если у нас нет никаких объяснений механизма возникновения такой связи, то мы сомневаемся в ее реальности, и часто этот скептицизм справедлив. Например, в начале 80-х годов для лечения рака был предложен препарат Амигдалин. Однако ученые не могли представить себе биологический механизм противоопухолевого действия экстракта абрикосовых косточек, химическое строение которого весьма отлично от строения известных соединений с противоопухолевой активностью. Чтобы разрешить эту проблему, было проведено рандомизированное контролируемое испытание, в котором было показано, что данное вещество неэффективно против изученных форм рака [19].

Важно помнить, что представление о биологическом правдоподобии зависит от уровня знаний в данное время. В прошлом веке, во времена Мейгса, инфекционная природа заболеваний казалась нереальной. В наше время нетрудно согласиться с наблюдениями Холмса, поскольку известен биологически правдоподобный механизм возникновения послеродового сепсиса вследствие стрептококковой инфекции. С другой стороны, по-прежнему неясен механизм анестезирующего эффекта иглоукалывания. Многие ученые на Западе считают биологически неправдоподобным, что втыкание в тело иглолок и их вращение должно снимать боль, а потому не верят в эффективность акупунктуры.

Итак, биологическое правдоподобие результата усиливает доводы о причинно-следственном характере связи. Если же оно отсутствует, нужно пытаться найти другой способ доказательства причинности связи. Если это другое доказательство окажется достаточно сильным, то отсутствие биологического правдоподобия укажет скорее на ограниченность современных знаний, чем на то, что изучаемый фактор не является причиной.

Специфичность

Специфичность (одна причина – один эффект) чаще имеет место при острых инфекционных болезнях (полиомиелит, столбняк) и врожденных нарушениях обмена веществ (подагра, охроноз, семейная гиперхолестеринемия). Как уже отмечалось, при

хронических болезнях часто существует множество причин одного и того же заболевания. И наоборот, одна и та же причина приводит к возникновению многих заболеваний. Развитию рака легкого способствуют курение, асбестоз, ионизирующее излучение. Вследствие курения развиваются не только рак легкого, но также бронхит, язвенная болезнь, пародонтоз. Выявление специфического действия – сильный аргумент в пользу причинно-следственной связи. Однако отсутствие специфичности – слабый аргумент против нее.

Аналогия

Доказательство причинно-следственной связи усиливается при наличии аналогичных примеров для точно установленных причин. Если известно, что медленная вирусная инфекция может вызвать хроническое дегенеративное заболевание центральной нервной системы (подострый склерозирующий панэнцефалит), то по аналогии легче допустить, что другой вирус может вызвать дегенерацию иммунной системы (синдром приобретенного иммунодефицита). Но вообще-то аналогия – слабый аргумент в пользу причины.

Взвешивание доказательств

В этой главе в основном обсуждалось, на что именно надо обращать внимание в отдельных исследованиях, когда рассматривается возможность причинной связи. Однако, решая вопрос о причине, следует учитывать все доказательства, полученные во всех доступных исследованиях. После всестороннего изучения этих доказательств вывод о причинном характере связи может быть усилен или признан ошибочным. Такой подход требует детального рассмотрения результатов различных исследований, особенно если они противоречивы. В подобных случаях клиницисты должны взвешивать доказательства.

На рис. 11.10 представлены различные аргументы за и против причинного характера связи в зависимости от метода исследования и признаков, которые усиливают или ослабляют степень доказанности причины. Показана относительная сила ряда аргументов для принятия или отклонения гипотезы о причине. Так, выявление сильной связи и зависимости эффекта от дозы в тщательно выполненном когортном исследовании служит сильным доказательством в пользу причинной связи. В то же время, если в одномоментном исследовании не удастся выявить эффект воздействия причин, это является слабым доказательством отсутствия причинно-следственной связи.

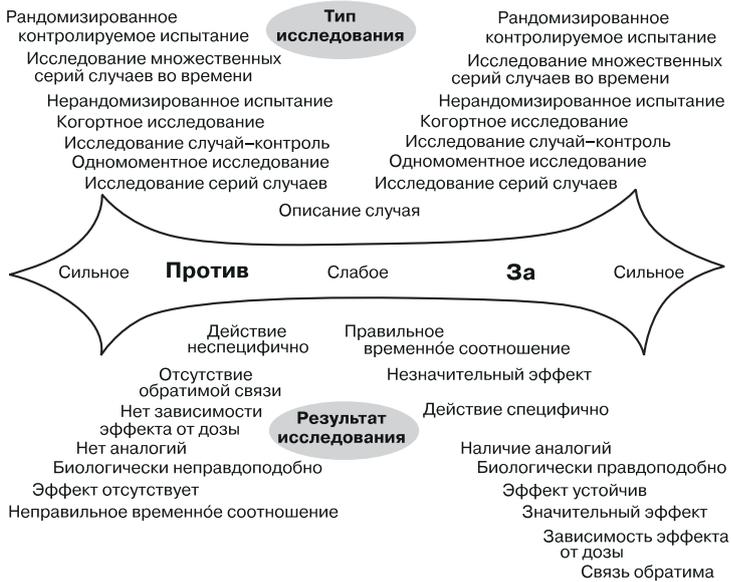


Рис. 11.10. "Шкала доказательности" исследований. Обратите внимание на то, что сила, с которой каждый из методов доказывает наличие причинного характера связи, равна силе, с которой он доказывает отсутствие таковой. Это справедливо для всех методов, кроме описаний случаев и исследований множественных серий случаев во времени. Противоположные результаты исследований не совпадают по силе аргументации за или против причинности связи.

Резюме

Причинно-следственные связи составляют основу диагностической, профилактической и лечебной деятельности в клинической медицине.

Обычно заболевания имеют много причин, хотя иногда может преобладать одна причина. Часто несколько причин взаимодействуют таким образом, что риск развития заболевания становится больше, чем можно было бы ожидать при простом суммировании эффектов каждой отдельной причины. В других случаях сила причинно-следственной связи между двумя переменными изменяется, благодаря присутствию третьей переменной, модифицирующей эффект.

Причинами заболевания могут быть непосредственные этиологические факторы или более отдаленные – генетические, экологические, поведенческие. Медицинские вмешательства, направ-

ленные на предупреждение или обратное развитие заболевания, возможны на любой стадии, как на уровне отдаленных, так и непосредственных патогенетических причин.

Причинно-следственную связь обычно устанавливают путем постепенного накопления результатов разных исследований. Сила доказательств определяется в первую очередь типом исследования. Для установления причины проводятся наблюдательные исследования, поскольку использовать рандомизированные контролируемые испытания с этой целью редко бывает возможно. Некоторые популяционные исследования (исследования единичных и множественных серий случаев во времени) позволяют предположить наличие причинной связи, когда воздействие на группы сопровождается определенным эффектом.

К признакам, усиливающим доказательство существования причинно-следственной связи, относятся определенная временная последовательность событий, сильная связь между предполагаемой причиной и эффектом, зависимость эффекта от дозы, уменьшение риска при устранении предполагаемой причины и постоянство результатов в нескольких исследованиях. Дополнительные признаки наличия причинной связи – биологическое правдоподобие и аналогия с уже известными фактами.

Литература

1. Holmes O.W. On the contagiousness of puerperal fever. *Med Classics* 1936; 1:207–268 [Originally published 1843.]
2. Webster's ninth new collegiate dictionary. Springfield, MA: Merriam-Webster 1991.
3. MacMahon B., Pugh T.F. *Epidemiology: principles and methods*. Boston: Little, Brown & Co 1970.
4. Cantwell M.F., Snider D.E., Cauthen G.M., Onorato I.M. Epidemiology of tuberculosis in the United States, 1985 through 1992. *JAMA* 1994;272:535–539.
5. Weis S.E., et al. The effect of directly observed therapy on the rates of drug resistance and relapse in tuberculosis. *N Engl J Med* 1994;330:1179–1184.
6. Goldman L. Analyzing the decline in the CAD death rate. *Hosp Pract* 1988;23(2):109–117.
7. Kannel W.B. Preventive cardiology. *Postgrad Med* 1977;61:74–85.
8. Siscovick D.S., et al. Diuretic therapy for hypertension and the risk of primary cardiac arrest. *N Engl J Med* 1994;330:1852–1857.
9. O'Conner G.T., Morton J.R., Diehl M.J., Olmstead E.M., Coffin L.H., Levy D.G. Differences between men and women in hospital mortality associated with coronary artery bypass graft surgery. *Circulation* 1993;88(1):2104–2110.
10. St. Leger A.S., Cochrane A.L., Moore F. Factors associated with cardiac mortality in developed countries with particular reference to the consumption of wine. *Lancet* 1979;1:1017–1020.
11. Pear S.M., Williamson T.H., Bettin K.M., Gerding D.N., Galgian J.N. Decrease in nosocomial *Clostridium difficile*-associated diarrhea by restricting clindamycin use. *Ann Intern Med* 1994;120:272–277.
12. Cervical Cancer Screening Programs. 1. Epidemiology and natural history of carcinoma of the cervix. *Can Med Assoc J* 1976; 114:1003–1033.
13. Bradford-Hill A.B. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
14. Kuller L., Wing R. Weight loss and mortality. *Ann Intern Med* 1993;119:630–632.

15. Centers for Disease Control. Highlights of the surgeon general's report on smoking and health. *MMWR* 1979;28:1–11.
16. Beasley R.P., Lin C.C., Hwang L.Y., Chien C.S. Hepatocellular carcinoma and hepatitis B virus. *Lancet* 1981;2:1129–1133.
17. Mandel J.S., et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993;328:1365–1371.
18. Selby J.V., Friedman G.D., Quesenberry C.P., Weiss N.S. A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992;326:653–657.
19. Moertel C.G., et al. A clinical trial of amygdalin (laetrile) in the treatment of human cancer. *N Engl J Med* 1982;306:201–206.

Рекомендуемая литература

Buck C. Popper's philosophy for epidemiologists. *Int J Epidemiol* 1975;4:159–168.

Chalmers A.F. *What Is this Thing Called Science?* 2nd ed. New York: University of Queensland Press 1982.

Department of Clinical Epidemiology and Biostatistics, McMaster University Health Sciences Centre. How to read clinical journals. IV: To determine etiology or causation. *Can Med Assoc J* 1981;124:985–990.

Rothman K.J. (ed). *Causal inference*. Chestnut Hill, M.A.: Epidemiology Resources 1988.

12

ПОДВОДЯ ИТОГИ

"Где то знание, которое затерялось в информации?"
Т.С.Элиот, "Судьба"

Польза от клинического исследования определяется, с одной стороны, его научной обоснованностью, признаваемой вдумчивыми беспристрастными учеными, а с другой стороны, связью с проблемами, стоящими перед врачами и их пациентами. Врачи, клинические решения которых основываются на знаниях, получаемых из медицинской литературы, и исследователи, создающие эти знания, должны понимать, в чем состоят сильные и слабые стороны научного исследования.

Для того чтобы судить о научной обоснованности читатель должен проявить активность. Он должен составить себе четкое представление о том, что именно он хочет извлечь из медицинской литературы, затем выяснить, существует ли такая информация и соответствует ли она стандартам научной обоснованности. При пассивном чтении, без опоры на научные принципы, читатель не заметит недостатков в опубликованном материале и легко может впасть в заблуждение.

В этой главе описано, каким образом занятый своей повседневной работой врач может на практике применить рассмотренные в этой книге методологические основы *доказательной медицины* (evidence-based medicine)¹. Во-первых, мы обсудим, как выявлять научные статьи, отвечающие на конкретный клинический вопрос, и как сократить их число до приемлемого уровня, не жертвуя важной информацией. Во-вторых, мы обобщим основные правила оценки качества статьи, возвращаясь к понятиям, уже обсуждавшимся в книге. В-третьих, мы рассмотрим, как проводить сравнительную оценку статей по определенной проблеме, чтобы выбрать наиболее достоверные. Именно на этих достоверных результатах врачи должны основывать свои клинические решения, пока в их распоряжение не поступят новые, более совершенные данные. На протяжении всей главы будет показано, как эти шаги –

¹ Понятие "Evidence-based medicine" может переводиться на русский язык по-разному, в частности, как аналитическая, научно обоснованная или доказательная медицина. –Примеч. ред.

выявление статей, оценка исследований и синтез информации соотносятся со стратегической задачей врача: в течение всей своей практической деятельности быть в курсе современной медицинской литературы.

Какова бы ни была сила наиболее надежного из имеющихся фактов, врач должен брать его за основу для действий, будучи в то же время готовым к переоценке фактов [1]. Наука, как социальный институт, опирается на готовность ученых рассматривать честные обоснованные аргументы за и против бытующей точки зрения, принимать новые доказательства и изменять свое мнение.

Как читать медицинскую литературу

Врачи изучают медицинскую литературу по-разному в зависимости от поставленных целей. Это может быть просмотр публикаций в поисках важных данных, чтение интересных статей для поддержания необходимого профессионального уровня, поиск ответов на конкретные клинические вопросы, систематическое изучение литературы по определенной клинической проблеме для выработки или изменения стандартов качества. К сожалению, врачи редко находят время для исчерпывающего изучения литературы. Однако если врач владеет основными принципами поиска информации, то даже при беглом просмотре ему легче определить заслуживающие внимания статьи и отобрать важные для себя результаты.

Какие статьи важны для принятия клинических решений

Не все статьи одинаково важны для принятия клинического решения. Вдумчивые клиницисты должны находить и оценивать в океане информации самое существенное.

На рис.12.1 схематически показано, как отличать наиболее важные для принятия клинического решения статьи, отсеивая второстепенные. Во многих статьях – обзорных, учебных, редакционных – описано то, что принято считать истиной, при этом сами статьи не являются публикациями результатов оригинальных исследований, выполненных для установления этой истины. Такие статьи – удобный источник обобщенной информации, но они только интерпретируют информацию, не внося в нее независимого вклада. Кроме того, их обычно пишут авторы с установившимся мнением, поэтому не исключена известная предвзятость.

Пример. Насколько объективно обзорные статьи и учебники отражают имеющиеся научные данные по определенному клиническому вопросу? На основании результатов *мета-анализа* (meta-

Вклад в решение клинических проблем

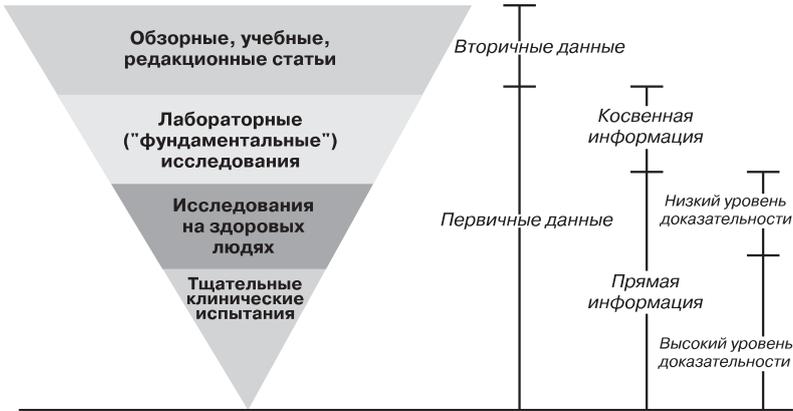


Рис. 12.1. Характеристика литературы по научной проблеме: относительная ценность различных типов статей применительно к поиску ответа на клинический вопрос.

analysis) (см. далее в этой главе) рандомизированных клинических испытаний [2] авторы оценили эффективность различных вмешательств для снижения заболеваемости и смертности от инфаркта миокарда. Оценки сопоставляли с рекомендациями экспертов, опубликованными в тот же период в обзорных статьях и учебниках. Оказалось, что точка зрения экспертов на несколько лет запаздывает по сравнению с результатом мета-анализа и часто не согласуется с ним. Так, к 1980 г. были опубликованы данные 12 рандомизированных контролируемых испытаний, посвященных оценке эффективности профилактического применения лидокаина при лечении острого инфаркта миокарда. Буквально все они показали, что применение лидокаина не дает никаких преимуществ, а иногда даже ухудшает результат по сравнению с плацебо. Вместе с тем большинство обзорных статей и учебников, опубликованных в 80-е годы, продолжали рекомендовать лидокаин в качестве стандартного метода лечения.

В других статьях описаны оригинальные исследования, выполненные в лабораторных условиях для выяснения биологической сущности заболевания. Эти работы служат богатым источником гипотез. Однако результаты таких исследований нельзя непосредственно распространять на человека – сложный организм в сложном физическом и социальном окружении – поскольку вли-

яние всех этих факторов и особенностей было преднамеренно исключено при эксперименте.

Исследования на людях (клинические исследования) нацелены на разработку рекомендаций для принятия клинических решений. Они выполняются с разной степенью научной строгости. Большинство из них не отвечают даже низким стандартам качества. Например, недавний обзор методов клинических исследований, опубликованных в трех хирургических журналах, выявил, что более чем в 80% исследований не только не было рандомизированной контрольной группы, но и вообще отсутствовала группа сравнения [3].

В этой книге мы доказывали, что ценность клинического исследования зависит от строгости используемых в нем методов (достоверность) и возможности приложения его результатов к другой клинической ситуации (обобщаемость). Следовательно, небольшое количество хороших публикаций гораздо ценнее, чем множество слабых или не соответствующих реальной практике. Таким образом, общий вывод из медицинских публикаций часто зависит от интерпретации относительно небольшого числа статей. Составление обзора литературы подразумевает тщательный отбор публикаций, оценку их сильных и слабых сторон и обобщение различных результатов.

Как искать полезные статьи

При систематическом изучении литературы первая задача состоит в том, чтобы найти полезные статьи. Это особенно трудно, когда объем литературы таков, что приходится просматривать множество, часто тысячи заголовков, прежде чем удастся найти несколько полезных публикаций. При этом необходимо уменьшить объем литературы до приемлемого количества без потери важной информации – задача ответственная и требующая много времени. Опишем план действий для проведения исчерпывающего обзора литературы.

Первый этап состоит в разработке критериев сортировки заголовков – для отбора статей, относящихся к теме поиска, и исключения гораздо большего числа ненужных статей. Критерии должны стать чувствительным (скрининговым) тестом для выявления необходимых статей, т.е. нужно, чтобы как можно меньше полезных статей было упущено. Вначале можно пожертвовать специфичностью ради обеспечения требуемой чувствительности, сознавая, что затем придется повозиться с "ложноположительными" статьями, вылавливая те материалы, которые соответствуют окончательным критериям отбора. Часто полезный скрининго-

вый алгоритм представляет собой сочетание ключевых слов, которые должны присутствовать в заглавии, например "саркоидоз легочный, кортикостероиды" или "рак поджелудочной железы, диагноз".

На втором этапе поиска критерии отбора применяются к списку названий журналов, который ведется *Национальной медицинской библиотекой США* (National Library of Medicine), для чего используется база данных MEDLINE. Несмотря на наличие соответствующих печатных изданий, многие клиницисты сейчас имеют доступ к электронным версиям этой базы данных на компакт-дисках (CD-ROM) или через глобальные компьютерные сети. Поскольку при компьютерном поиске обычно пропускаются некоторые важные статьи, необходимо параллельно использовать другие источники, например новые обзорные статьи, другие публикации по той же теме, учебники, советы специалистов. В результате поиска будет сформирован обширный список заголовков, часть которых отражает нужные статьи, но многие в действительности лишние.

Третий этап поиска состоит в идентификации с помощью специфических критериев тех публикаций, которые непосредственно относятся к изучаемому вопросу. Часто применяются три критерия отбора публикаций:

- Обозначен ли в заголовке статьи конкретный клинический вопрос, изначально составляющий предмет поиска?
- Представляет ли статья оригинальное исследование или вторичную информацию (точку зрения)?
- Основано ли исследование на относительно надежных методах?

Многие статьи, не относящиеся к теме поиска, удается исключить путем изучения полного заголовка. Однако часто этой информации недостаточно и тогда очень помогает реферат. При внимательном чтении реферата можно увидеть, использовались ли в исследовании метода лечения контрольная группа и случайное распределение пациентов по группам, проводилось ли исследование прогноза на исходной когорте, достаточна ли статистическая мощность или точность и т.д.

Структурированные рефераты [4], публикация которых взята за правило многими ведущими медицинскими журналами, облегчают выявление потенциально полезных статей. В структурированном реферате по определенному шаблону приведены основные характеристики исследования (цель и метод исследования, клиническая база, характеристики пациентов и вмешательств,

оцениваемые исходы, результаты и выводы), которые требуются, чтобы отличить достоверные и информативные публикации от многих других, неоригинальных или недостаточно строго выполненных.

На последнем этапе этой работы просматривают тексты отобранных статей, чтобы отсеять те из них, которые не соответствуют окончательным критериям. К этому моменту число статей уже должно уменьшиться настолько, чтобы такая задача стала выполнимой. На рис. 12.2 обобщены эти три этапа и проиллюстрирован процесс поиска литературы по конкретному вопросу – исходы протезирования коленного сустава [5].

Если времени для полного и основательного поиска недостаточно, либо читатель хочет лишь бегло просмотреть литературу, чтобы быть в курсе последних достижений, то начальные этапы процесса сокращаются. Иногда изучаются только журналы, публикующие оригинальные исследования, выполненные на высоком методическом уровне. Однако это малочувствительная тактика: чтобы найти 80% наиболее надежных статей по конкретному вопросу, придется изучить по крайней мере 11 лучших журналов [6].

Другой подход заключается в том, чтобы скрининг проводился специалистами – экспертами в области клинической медицины и методологии клинических исследований, которые должны изучать все публикуемые в мире статьи, применяя для отбора строгие критерии. Журнал "ACP Journal Club"¹ публикует структурированные рефераты надежных в научном отношении оригинальных клинических исследований; отбор статей производится по строгим критериям, публикуемым в каждом выпуске журнала. Результаты отбора впечатляют: в 1993 г. из более 6 млн статей, опубликованных в клинических журналах, лишь около 350 (приемлемое для изучения число) соответствовали этим критериям.

Третий подход обеспечивает база данных, создаваемая организацией "*Кохрейн Коллаборейшн*" (Cochrane Collaboration). Экспертные группы во всем мире объединенными усилиями отбирают лучшие исследования по клиническим вмешательствам, обобщают их в стандартной форме, распространяют информацию с помощью электронных систем, при обнаружении ошибок вносят коррективы в базу данных и поддерживают ее на современном уровне. Результаты этой работы, по крайней мере по некоторым клиническим вопросам, должны стать доступными в середине 90-х годов. (Уже стали доступны. – *Примеч. ред.*)

¹ С 1996 г. большая часть материалов этого журнала переводится на русский язык и публикуется в "Международном журнале медицинской практики". – *Примеч. ред.*

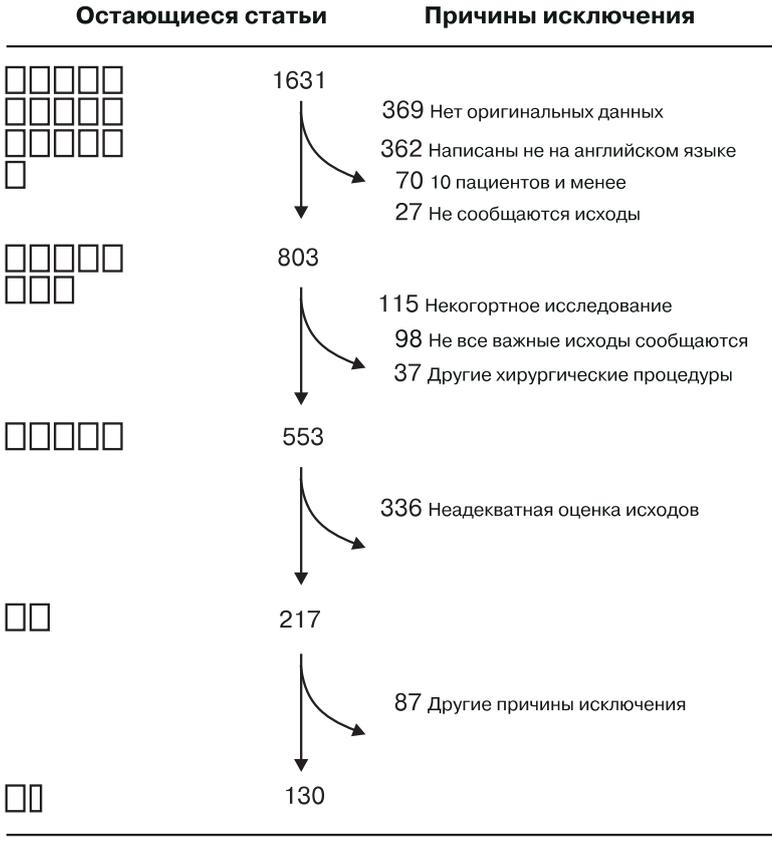


Рис. 12.2. Поиск литературы: идентификация небольшого числа наиболее важных статей из массива медицинской литературы [Callahan C.M., Drake B.G., Heck D.A., Dittus R.S. Patient outcome following tricompartmental total knee replacement. JAMA 1994;271:1349–1357].

Оценка статей

После выявления статей, представляющих интерес, следующая задача состоит в том, чтобы оценить качество представленных в них данных. При этом используются научные принципы, описанные в этой книге.

В приложении к данной главе описаны фундаментальные критерии оценки научной надежности различных типов исследования. Мы сформулировали ряд вопросов, которые должны ставить как исследователи, так и читатели при знакомстве с резуль-

татами исследований. Вопросы касаются цели исследования, обобщаемости и клинической применимости результатов, а также двух факторов, влияющих на достоверность исследования: систематической и случайной ошибки. Мы приводим основные правила оценки диагностических и прогностических исследований, исследований частоты заболеваний, лечения и установления причин. В приложении перечислены важнейшие вопросы, ответив на которые, можно сделать вывод о том, заслуживает ли статья включения в обзор литературы, а также оценить степень ее доказательности при обобщении результатов разных статей.

Соответствует ли методика исследования заявленной цели?

Нельзя говорить о "хорошей" или "плохой" методике исследования вообще, абстрагировавшись от вопроса, на который это исследование должно ответить. Многие клинически ориентированные схемы оценки методики рассматривают наблюдательные исследования, например одномоментные как наименее доказательные. Это утверждение верно, если вопрос касается профилактики или лечения, но несправедливо, если речь идет об исследовании диагностических тестов. В табл.12.1 приведены клинические вопросы и методы исследований, лучше всего им соответствующие. Этими данными можно руководствоваться при выборе метода, однако они не исключают применения других творческих научно обоснованных подходов наряду с теми, которые названы в таблице. Например, самое надежное из имеющихся доказательств

Таблица 12.1
Оптимальные методы исследований применительно к разным клиническим вопросам

Вопрос	Метод исследования
Диагностика	Одномоментное исследование
Распространенность	Одномоментное исследование
Частота возникновения новых случаев (заболеваний/исходов)	Когортное исследование
Риск	Когортное исследование Исследование случай—контроль
Прогнозирование	Когортное исследование
Лечение	Клиническое испытание
Профилактика	Клиническое испытание
Причина	Когортное исследование Исследование случай—контроль

того, что периодический скрининг с применением сигмоскопии может способствовать уменьшению смертности от колоректального рака, получено в строгом исследовании случай–контроль [7]. Этот метод исследования служит единственно возможной формой доказательства на сегодняшний день, поскольку для проверки эффективности сигмоскопии в рандомизированном исследовании потребовались бы большое число больных и длительный период наблюдения.

Выдвижение и проверка гипотез

Выводы конкретных исследований в зависимости от применяемого метода обладают разной убедительностью (см. главу 11, рис. 11.10, "шкала доказательности"). На одном конце шкалы находятся потенциально полезные сообщения, которые только предполагают наличие связи (без проверки выдвинутых предположений). Большинство описаний случаев выполняют именно эту роль. Выводы исследований носят предварительный характер, многие из них впоследствии не подтверждаются. На другом конце шкалы находятся исследования, подвергающие гипотезы строгой проверке – например крупные рандомизированные контролируемые испытания. Выводы этих исследований более категоричны. Большинство исследований располагается между двумя крайними точками этой шкалы.

Сформулированные до начала исследования гипотезы очень важны – они препятствуют проникновению в литературу ложноположительных результатов. Допустим, исследуется большое число переменных, ни одна из которых в действительности не связана с другими. Как отмечалось в главе 9, при изучении большого числа связей между переменными имеется вероятность того, что некоторые из связей по чисто случайным причинам достигнут уровня, позволяющего считать их реально существующими. На принятом уровне статистической значимости ($p < 0,05$) примерно 1 из 20 таких сравнений значимо уже по определению. Конечно, установленные связи "реальны" только для конкретной выборки и не обязательно проявятся в популяции, поскольку изучаемая выборка может быть не похожа на другие выборки из рассматриваемой популяции.

Теперь предположим, что одной из таких связей уделено особое внимание, потому что она хорошо согласуется с существующими биомедицинскими теориями. Предположим также, что в окончательном отчете исследования о проверке других связей не сообщается. В этом случае данная единственная связь при ее рассмотрении вне контекста исследования может показаться очень

Таблица 12.2

Характеристики исследований, позволяющие определить, способны ли они проверить или лишь выдвигать гипотезы

Характеристика	Выдвижение гипотез	Проверка гипотез
Методика исследования	Нестрогая	Строгая
Гипотеза	Отсутствует или формулируется после сбора и анализа данных	Сформулирована до начала исследования
Число сравнений	Большое	Небольшое
Значение p	Велико	Мало
Результаты подтверждены на другой выборке	Нет	Да

важной. Такое явление – предвзятый выбор какой-то интересной случайно обнаруженной связи – нередко встречается в публикациях.

Существует несколько способов оценки того, в какой степени исследование является проверяющим или только выдвигающим гипотезу (табл.12.2).

Строгость методики исследования нельзя рассматривать в отрыве от других факторов. Предварительное выдвижение гипотезы и ограничение числа изучаемых сравнений уменьшают вероятность появления связей, лишь кажущихся достоверными. В результате применения распространенной тактики анализа данных, при которой сравниваются данные в различных подгруппах большого популяционного исследования, могут выявляться случайные или ложные связи. Если подтверждаются априорные гипотезы (выдвинутые до начала исследования), то доверие к полученным данным возрастает. Другой путь повышения надежности результатов – ограничение числа сравнений, что уменьшает вероятность ложноположительных результатов в исследовании. Кроме того, можно установить очень малое значение p , чтобы исключить роль случайности при объяснении конкретных находок.

Еще один способ застраховаться от ошибочного признания реальности ложных или случайных связей заключается в выдвижении гипотезы на основании обработки одного массива данных с последующей проверкой ее на другом массиве (рис.12.3). Доступность больших массивов данных и наличие статистических компьютерных программ позволяют проводить анализ с включением многих переменных, рассматриваемых в моделях по отдельности или совместно. Анализ многих переменных должен рассматри-



Рис. 12.3. Формирование и проверка гипотезы на разных массивах данных.

ваться как метод выдвижения гипотез, так как исследователи заранее редко оговаривают ожидаемый результат, что снижает его ценность. Если массив данных достаточно велик, его можно случайным способом разделить на две части. Одна часть может быть использована для построения, а вторая – для проверки модели. Можно проверять гипотезу на другом массиве данных.

Пример. Исследователи разработали индекс, включающий семь прогностических признаков, для предсказания раннего рецидивирования астмы после выписки из отделения неотложной помощи [8]. Индекс был разработан на группе из 205 больных, наблюдавшихся в медицинском центре, и для этой группы имел чувствительность 95% и специфичность 97%. Результаты были настолько впечатляющими, что индекс начали внедрять в клиническую практику повсеместно.

Позднее две другие группы исследователей независимо друг от друга оценили результат применения индекса на других клинических базах [9, 10] и получили разочарывающие данные. Чувствительность и специфичность составили соответственно 40 и 71% в одном исследовании и 18,1 и 82,4% – в другом.

Этот пример показывает, что опасно слишком доверять связи, выявленной по результатам анализа одного массива данных, без проверки на другой независимой группе больных. Возможно несколько причин обнаруженных расхождений результатов: систематические отличия больных в первом исследовании, разный способ применения индекса, обусловленные случайностью нетипичные результаты в исходном исследовании.

Независимо от того, какая тактика используется для того чтобы придать исследованию проверочный характер, обязанность автора – четко пояснить, какому участку шкалы между крайними позициями (выдвижение гипотез – проверка гипотез) принадлежит его статья и почему. Задача читателя – найти эту информа-

цию или воссоздать ее, если она не очевидна. Однако не следует отвергать исследования, которые только выдвигают гипотезы – они тоже важны, хотя и не дают однозначных ответов.

Объединение результатов многих исследований

Современное состояние знаний по интересующей нас проблеме обычно определяется по результатам всех относящихся к делу исследований, а не по данным какой-либо одной публикации. До недавнего времени обычный путь решения этого вопроса заключался в формировании некоего суждения без применения формальных правил оценки вклада индивидуальных исследований. Оценки такого типа часто принимают форму традиционной (повествовательной) обзорной статьи, подготовленной специалистом в соответствующей области или группой экспертов, представляющих разные взгляды на проблему (например, согласительной конференцией Национального института здоровья США).

Сейчас при обобщении публикаций применяются формализованные методы. Их преимущество состоит в том, что благодаря приданию относительного веса рассматриваемым при обзоре исследованиям формируются согласованные общие выводы. Кроме того, научно обоснованные методические принципы используются как на этапах предварительного определения критериев оценки, сбора информации (в данном случае – результатов отдельных исследований) и анализа результатов, так и в процессе подведения итогов в соответствии с этими критериями и собранными данными.

Процесс обобщения исследований по определенному вопросу с использованием формализованных методов называется мета-анализом, буквально – анализом анализов, или синтезом информации. Такой подход особенно полезен, когда рассматривается один конкретный вопрос и существует по крайней мере несколько относительно надежных исследований с разными выводами. Особенно широко эти методы стали применяться в последние годы. Среди публикаций 1990 – 1994 гг. база данных MEDLINE содержала почти 2000 статей под предметной рубрикой "мета-анализ". (В настоящее время их число приближается к 5000. – *Примеч. ред.*)

Существует три основных этапа выполнения мета-анализа.

Первый – выявление из всех имеющихся по теме статей наиболее надежных, как это описывалось в начале главы.

Второй – оценка соответствия каждой статьи заранее установленным критериям. В некоторых мета-анализах при такой

оценке выводится общий показатель качества исследований; в других случаях качественные характеристики исследований, такие как метод исследования, количество пациентов, источник выборки, способ сбора данных – оцениваются отдельно.

Третий этап – объединение количественных результатов многих исследований; фактически при этом формируется одно большое исследование с более высокой статистической мощностью, чем отдельные исследования. Каждое отдельное исследование "взвешивается" по размеру выборки, т.е. исследования с большим числом наблюдений получают больший вес, чем исследования с меньшим числом больных.

Во многих мета-анализах при взвешивании учитывается качество каждого исследования. Затем с помощью статистических методов, обычно разновидности регрессионного анализа, выводится объединенный результат, например относительный риск или процентное снижение смертности. Результат мета-анализа представляется в виде графика, на котором изображены как результаты отдельных исследований, так и обобщенный показатель эффекта.

Несмотря на то что мета-анализ стал стандартом оценки других подходов к обобщению литературы, сохраняются противоречия относительно многих его элементов, в частности определения показателей качества индивидуальных исследований и их включения в итоговую оценку. Нужно иметь в виду, что показатели качества, хотя и объединяют разные с методологической точки зрения характеристики, помогают объяснить различия между исследованиями.

Пример. Хотя вакцина БЦЖ используется для иммунизации против туберкулеза на протяжении более 50 лет во многих странах, ее эффективность неоднозначна; в частности, в нескольких крупных клинических испытаниях получены противоречивые результаты. В одном из ранних мета-анализов [11], сравнившим методы и результаты этих исследований, было показано, что выявление туберкулеза в группе, вакцинированной БЦЖ, и в контрольной группе не было подвержено систематической ошибке только в трех исследованиях, показавших эффективность вакцинации 75% и более.

В более позднем мета-анализе по тому же вопросу [12] были объединены результаты 7 рандомизированных и 6 нерандомизированных испытаний, а также 10 исследований случай–контроль, опубликованных до 1994 г. включительно (рис.12.4). Мета-анализ показал, что независимо от применявшегося метода исследования риск заболевания туберкулезом у вакцинированных уменьшается примерно в два раза по сравнению с невакцинированными. Для

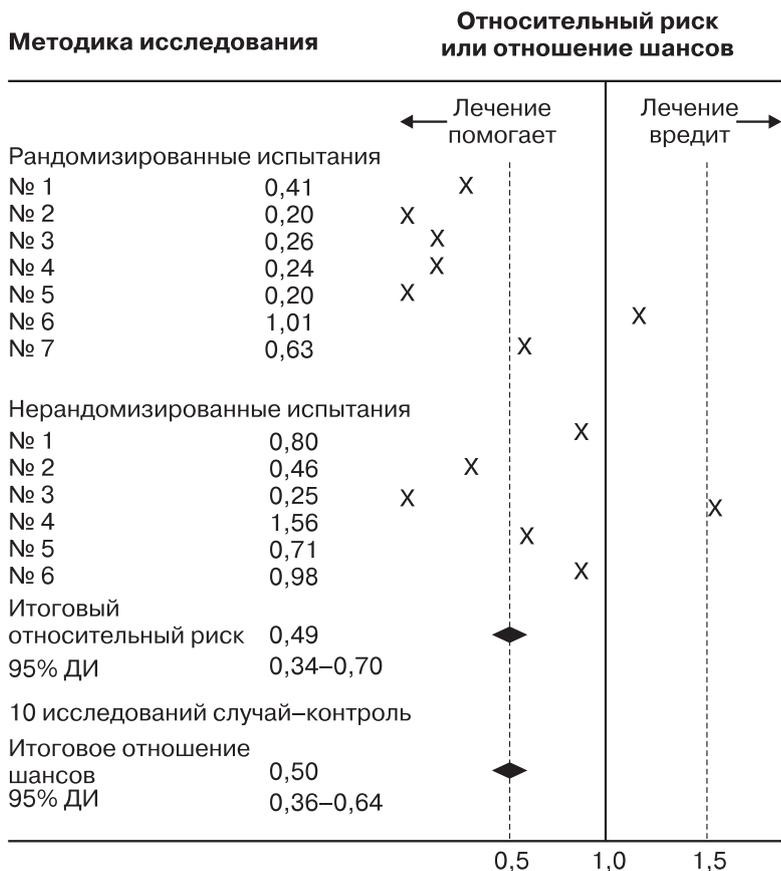


Рис. 12.4. Результаты мета-анализа эффективности вакцинации БЦЖ для профилактики туберкулеза. ДИ – доверительный интервал [Colditz G.A., Brewer T.F., Berkey C.S., Wilson M.E. et al. Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. JAMA 1994; 271:698–702].

объяснения различий наблюдаемой эффективности между разными исследованиями были разработаны обобщенные показатели качества для каждого из трех использованных методов. С помощью регрессионного анализа установлено, что более высокий показатель качества исследования соответствует данным о более высокой эффективности вакцинации. Таким образом, мета-анализ четко установил эффективность вакцинации БЦЖ.

Однако зачастую не удастся выявить связь между обобщенным показателем качества и результатами исследования. В этом

случае при мета-анализе нужно обратить внимание на конкретные методические особенности исследований, чтобы понять, почему их выводы противоречивы [13].

Объединение данных

Нередко результаты отдельных исследований бывают неопределенными: из-за малого числа больных или исходов не обеспечивается достаточная статистическая мощность или точность. В таких исследованиях оценки показателей нестабильны, а сравнения отягощены неприемлемо высоким риском пропустить эффект (бета-ошибка).

Выходом из такой ситуации является объединение данных нескольких относительно небольших исследований, посвященных одному и тому же вопросу. В результате такого объединения формируется один массив данных, который условно можно рассматривать как полученный в результате одного большого исследования. Это допустимо, только если отдельные исследования настолько сходны между собой (по характеристикам пациентов, вмешательствам, показателям исходов), что их можно трактовать как составные части единого исследования. Объединение данных преследует цель собрать достаточное количество наблюдений, чтобы получить точную общую оценку эффекта, а не объяснять различия выводов разных исследований. Преимущество объединения данных состоит в том, что оно позволяет обеспечить достаточную статистическую мощность для выявления значимых различий, если они существуют. Объединение данных особенно полезно, когда изучается редкое заболевание и/или исход. При таких обстоятельствах объединение данных – единственный способ достижения требуемой статистической мощности.

Пример. Имеется много сообщений о развитии язвенной болезни при приеме кортикостероидов. Тем не менее в обсервационных исследованиях трудно установить, действительно ли данные препараты вызывают язву, поскольку многие ситуации, при которых назначаются кортикостероиды (например, при стрессе или в сочетании с раздражающими желудок веществами), могут самостоятельно вести к развитию язвенной болезни. Возможно также, что у людей, получающих кортикостероиды, язва обнаруживается вследствие более тщательного обследования, тогда как у других больных она остается нераспознанной.

Лучший способ определения причины и следствия – проведение рандомизированных контролируемых испытаний. Известно много рандомизированных испытаний, в которых изучалось применение кортикостероидов для лечения различных заболеваний, при этом одним из регистрируемых побочных явлений была язвенная

болезнь. Ни одно из этих исследований не было достаточно большим, чтобы самостоятельно проверить гипотезу о том, что данные препараты вызывают язвенную болезнь. Но все вместе эти исследования позволяют изучить частоту относительно редкого события.

В обзоре [14], объединяющем результаты 71 рандомизированного контролируемого испытания кортикостероидов, в качестве одного из исходов рассматривалось развитие язвенной болезни. В среднем на одно исследование приходилось примерно 86 больных и 1 случай язвенной болезни; случаи язвы (хотя бы один) были выявлены лишь в 31 испытании. Авторы мета-анализа объединили результаты 71 испытания для повышения статистической мощности. В объединенном исследовании оказалось 6111 больных, из них около 80 больных с язвой. Частота язвенной болезни составила 1,8% в группе получавших кортикостероиды и 0,8% – в контрольной группе (относительный риск 2,3; 95% доверительный интервал от 1,4 до 3,7). Сходный результат был получен и при анализе данных с учетом других факторов риска (разные дозы, пути введения кортикостероидов и продолжительность лечения), а также особенностей проявления и диагностики язвы (подозрение на язву, наличие кровотечения, диагностика с помощью специальных методов).

Таким образом, объединение результатов многих исследований, каждое из которых было достаточно строгим с методической точки зрения, но включало слишком малое число наблюдений, чтобы ответить на поставленный вопрос, обеспечило необходимую статистическую мощность для выявления повышенного риска заболевания язвенной болезнью.

Сторонники метода объединения данных указывают, что этот метод позволяет по-новому увидеть сильные и слабые стороны разных исследований.

Например, на одном и том же рисунке можно представить число строгих исследований, точечные оценки эффективности и их доверительные интервалы для каждого исследования, соотношение между величиной эффекта и точностью оценки и, наконец, точечную оценку эффективности и доверительный интервал для объединенного результата (см. рис. 12.5).

Противники объединения данных возражают, что принципы отбора больных, методы вмешательств и оценки исходов в разных исследованиях настолько различаются, что объединять эти данные нельзя. "Разъединители" в отличие от "объединителей" не удовлетворены получаемым результатом. К тому же объединение данных обеспечивает только повышение статистической мощности; оно не ликвидирует систематические ошибки, которым подвержены отдельные методы исследований, и не создает условий для взаимного устранения этих ошибок в процессе объединения

данных. В любом случае мета-анализ только дополняет, но не заменяет того знания, которое достигается путем тщательного изучения каждого из лучших исследований, посвященных определенному клиническому вопросу.

Публикация результатов и систематическая ошибка

Клиницист, как и любой другой человек, предпочитает хорошие новости, поэтому заголовки журнальных статей пестрят такими терминами, как "эффективность", "прогнозирование", "корреляция". Ведь гораздо менее интересно обсуждать то, что не получилось; часто негативный результат считается просто неудачным. Друзья могут не без ехидства посоветовать таким авторам направить публикацию в "Журнал отрицательных результатов".

Предпочтение, отдаваемое положительным результатам, приводит к предвзятости при отборе материалов для публикации в медицинских журналах.

Пример. Английским комитетом по этике исследований была изучена судьба 285 статей [15]. В 54% исследований содержались статистически значимые результаты, в 16% – статистически незначимые тенденции, в 30% – нулевой результат. Среди исследований со статистически значимыми результатами 85% были опубликованы или представлены к публикации, тогда как среди исследований с отрицательным результатом – только 56% (отношение шансов 4,54; 95% доверительный интервал 2,4–8,6). Исследования с нулевым результатом не были хуже по качеству и не чаще отвергались редакциями журналов.

Массив опубликованных в конечном итоге данных представляет собой смещенную выборку результатов всех проводимых исследований. В результате методы выявления причин, диагностические тесты и лечение представляются более эффективными, чем в действительности. Например, мета-анализ относительной эффективности лечения больных раком яичника одним или несколькими препаратами выявил значительно лучшую выживаемость при комплексной терапии. Однако, когда автор мета-анализа [16] добавил неопубликованные данные, различие исчезло. Нет оснований утверждать, что смещение оценки преднамеренно. Каждый старается опубликовать свою "лучшую" работу, но публикация – не случайный процесс. Существуют факторы, способствующие тому, что предпочтение отдается положительным результатам – независимо от соотношения количества этих результатов среди всех выполненных научных проектов. Читатели должны знать об этой систематической ошибке, чтобы реалистически от-

носиться к новым многообещающим данным, которые появляются в медицинских журналах.

Один из путей избежать указанной систематической ошибки – больше доверять крупным исследованиям, чем небольшим. Основная масса исследований, выполнение которых потребовало значительных усилий и материальных затрат, будет опубликована, независимо от того, какие результаты получены – положительные или отрицательные. С другой стороны, маленькие, относительно недорогие исследования, скорее отсеиваются в процессе отбора для публикации.

Различия результатов исследований

До сих пор в основном обсуждалось, каким образом исследования могут привести к разным выводам из-за использования разных методов, более адекватных в одних и менее адекватных в других случаях. Однако существует и альтернативное объяснение: научные вопросы, будучи на первый взгляд сходными, в действительности могут существенно отличаться. В этом случае будут верны результаты не одного, а обоих исследований. Это может быть связано с тем, что наблюдаемые различия обусловлены не методами исследования, а особенностями биологии человека.

Пример. Несколько авторов [17, 18] провели мета-анализ исследований, оценивающих эффективность медикаментозного лечения артериальной гипертензии. Ранние исследования продемонстрировали существенное и статистически значимое снижение частоты инсультов, но менее выраженное, и часто недостоверное снижение частоты ИБС (рис. 12.5). Доверительный интервал для относительно высокого риска ИБС, вычисленного при объединении всех исследований, включает в себя единицу.

Более поздние исследования, которые были проведены на пациентах пожилого возраста или по крайней мере включали их, показали такую же степень снижения риска инсульта, но более выраженное и значимое снижение риска развития ИБС. Хотя большая эффективность лечения в более поздних испытаниях позволяет предполагать, что медикаментозное лечение при артериальной гипертензии у пожилых пациентов эффективнее, чем это считалось раньше, нужно иметь в виду более низкие дозы использованных в этих испытаниях средств – диуретиков и бета-блокаторов, что тоже могло способствовать более эффективному снижению риска ИБС.

Исследования причинно-следственных связей, где, казалось бы, поставлены одни и те же вопросы, могут в действительности рассматривать по крайней мере четыре разных аспекта: особенности пациентов, вмешательства, наблюдения и критерии

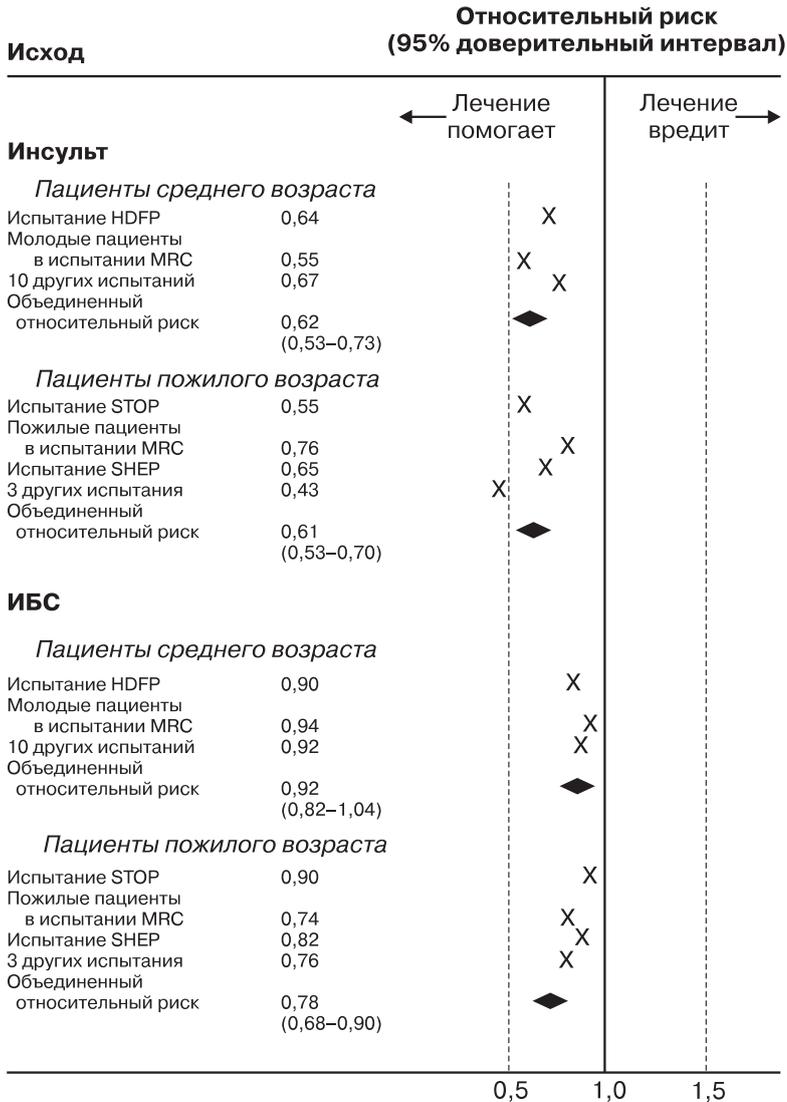


Рис. 12.5. Результаты мета-анализа влияния лечения артериальной гипертензии на риск развития инсульта и ИБС [Cutler J.A., Pstay B.M., MacMahon S., Furberg C.D. Public health issues in hypertension control. What has been learned from clinical trials? In: Laragh J.H., Brenner B.M., eds. Hypertension: physiology, diagnosis and management. 2nd ed. New York: Raven Press 1995].

эффективности. Различия по любому из этих аспектов достаточно, чтобы в исследованиях получились разные результаты.

Другие источники информации

До сих пор мы рассматривали в качестве главного источника информации журнальные статьи с сообщениями об оригинальных исследованиях и основанный на них мета-анализ. Что можно сказать о других источниках информации?

Учебники удобны и им можно доверять в том, что касается твердо установленных фактов. Их недостатки заключаются в запаздывании сведений, а также в отражении взглядов только авторов учебника, с ограниченными ссылками на другие точки зрения. Источниками информации могут служить коллеги, особенно специалисты в определенной области, но их мнения играют только консультативную роль, поскольку могут быть субъективны или подвержены финансовым интересам. Например, гастроэнтерологи доверяют эндоскопии больше, чем рентгенологическому исследованию с введением контрастного вещества, а хирург предпочитает медикаментозному лечению хирургическое вмешательство.

В последнее время постоянно увеличивается количество баз данных с полным и современным отражением информации; они широко доступны по телефону, факсу, на дискетах, компакт-дисках (CD-ROM), по электронной почте, в сети Интернет. Приведем примеры: круглосуточная телефонная связь с Центром контроля и профилактики заболеваний США обеспечивает информацию о профилактике заболеваний для отправляющихся в любую точку земного шара; база данных "Toxline" содержит информацию об отравлениях; база данных PDQ предоставляет рекомендации по современным химиотерапевтическим средствам для лечения злокачественных новообразований; целый набор баз данных содержит информацию о лекарственных препаратах, их токсичности и дозах с корректировкой для случаев почечной недостаточности. Эти базы данных содержат информацию о фактах, которые важны для практической медицины, но требуются нечасто или трудно запоминаются. Клиницисты должны знать о существовании этой информации и иметь к ней доступ. Кроме того, пользоваться базами данных нужно с учетом уроков этой книги: найденные факты надежны настолько, насколько хороши методы, которыми они получены. Многие базы данных, такие как клинические рекомендации Агентства по организации здравоохранения и научных исследований, Профилактической службы особого назначения США и Американской Коллегии врачей, создаются с помощью лучших методов и обеспечивают надежную информацию.

Некоторые базы данных несут отпечаток конфликта интересов отдельных лиц или промышленных кругов и должны восприниматься критически.

Клинические рекомендации

На протяжении всей книги мы доказывали, что клинические исследования составляют основу для внедрения новых методов в медицинскую практику и для принятия клинических решений. Такой подход в отличие от сообщений о случаях из жизни и личного опыта врача получил название доказательной медицины (evidence-based medicine)[19]. Важный элемент доказательной медицины – преобразование результатов исследований в ясные, четкие рекомендации для врачей. Практические рекомендации представляют собой методично разработанные установки, нацеленные на помощь врачам при решении конкретных клинических проблем [19]. Их подготовка и использование стали обычными во многих медицинских учреждениях. Достоверность лучших из рекомендаций обеспечивается путем включения в группы по их разработке специалистов по всем аспектам проблемы (от высококвалифицированных исследователей до клиницистов, экономистов и пациентов), для учета всех важных сторон проблемы и достижения баланса интересов. Лучшие из рекомендаций основаны на научных фактах, а не только на мнении экспертов, и поэтому при их создании часто используются формализованные методы изучения литературы и ее обобщения, как описывалось в этой главе [20].

Рекомендации предназначены для того чтобы давать советы, а не навязывать клинические решения. На практике встречаются достаточно веские причины, чтобы не следовать рекомендациям при лечении некоторых больных.

Влияют ли рекомендации на поведение врачей? В мета-анализе [21] было собрано и обобщено 59 публикаций, посвященных влиянию подробных рекомендаций, основанных на результатах строго организованных исследований (рандомизированные испытания, нерандомизированные сравнительные испытания, исследования серий случаев во времени). Показано, что более чем в 90% случаев клинические рекомендации приводили к существенным изменениям клинической практики, а 9 из 11 исследований продемонстрировали, что это ведет и к улучшению результатов лечения.

Литература

1. Tukey J.W. Conclusions and decisions. *Technometrics* 1960;2:423–433.
2. Antman E.M., Lau J., Kupelnick B., Mosteller F., Chalmers T.C. A comparison of results of meta-analyses of randomized trials and recommendations of clinical experts: treatment of

- myocardial infarction. *JAMA* 1992;268:240–248.
3. Soloman M.J., McLeod R.S. Clinical studies in surgical journals—have we improved? *Dis Colon Rectum* 1993;36:43–48.
 4. Haynes R.B., Mulrow C.D., Huth E.J., Altman D.G., Gardner M.J. More informative abstracts revisited. *Ann Intern Med* 1990;113:69–75.
 5. Callahan C.M., Drake B.G., Heck D.A., Dittus R.S. Patient outcome following tricompartmental total knee replacement. A meta-analysis. *JAMA* 1994;271:1349–1357.
 6. Haynes R.B. Where's the meat in clinical journals? [Editorial]. *ACP J Club* 1993; Nov-Dec:A22-A23 [Originally published in *Ann Intern Med* 1993; 114(Suppl 1)].
 7. Selby J.V., Friedman G.D., Quesenberry C.P. Jr., Weiss N.S. A case control study of screening sigmoidoscopy and mortality from colorectal cancer. *N Engl J Med* 1992;326:653–657.
 8. Fischl M.A., Pitchenik A., Gardner L.B. An index predicting relapse and need for hospitalization in patients with acute bronchial asthma. *N Engl J Med* 1981;305:783–789.
 9. Rose C.C., Murphy J.G., Schwartz J.S. Performance of an index predicting the response of patients with acute bronchial asthma to intensive emergency department treatment. *N Engl J Med* 1984;310:573–577.
 10. Centor R.M., Yarbrough B., Wood J.P. Inability to predict relapse in acute asthma. *N Engl J Med* 1984;310:577–580.
 11. Clemens J.D., Chuong J.H., Feinstein A.R. The BCG controversy: a methodological and statistical reappraisal. *JAMA* 1983;249:2362.
 12. Colditz G.A., Brewer T.F., Berkey C.S., Wilson M.E., Burdick E., Fineberg H.V., Mosteller F. Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. *JAMA* 1994;271:698–702.
 13. Greenland S. A critical look at some popular meta-analytic methods [Invited Commentary]. *Am J Epidemiol* 1994;140:290–296.
 14. Messer J., Reitman D., Sacks H.S., Smith H. Jr., Chalmers T.C. Association of adrenocorticosteroid therapy and peptic-ulcer disease. *N Engl J Med* 1983;309:21–24.
 15. Easterbrook P.J., Berlin J.A., Gopalan R., Matthews D.R. Publication bias in clinical research. *Lancet* 1991;337:867–872.
 16. Simes R.J. Confronting publication bias: a cohort design for meta-analysis. *Stat Med* 1987;6:11–29.
 17. Cutler J.A., Psaty B.M., MacMahon S., Furberg C.D. Public health issues in hypertension control: what has been learned from clinical trials? In: Laragh J.H., Brenner B.M., eds. *Hypertension: physiology, diagnosis, and management*, 2nd ed. New York: Raven Press, 1995.
 18. Hebert P.R., Fiebach N.H., Eberlein K.A., Taylor J.O., Hennekens C.H. The community based randomized trials of pharmacological treatment of mild-to-moderate hypertension. *Am J Epidemiol* 1988;127:581–589.
 19. Evidence based medicine working group. Evidence based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;268:2420–2425.
 20. Hayward R.S.A., Wilson M.C., Tunis S.R., Bass E.B., Rubin H.R., Haynes R.B. More informative abstracts of articles describing clinical practice guidelines. *Ann Intern Med* 1993;118:731–737.
 21. Grimshaw J.M., Russell I.T. Effect of clinical guidelines on medical practice. A systematic review of rigorous evaluations. *Lancet* 1993;342:1317–1321.

Рекомендуемая литература

- Epidemiology Work Group of the Interagency Regulatory Liaison Group. Guidelines for documentation of epidemiologic studies. *Am J Epidemiol* 1981;114:609–713.
- Goldschmidt P.G. Information synthesis: a practical guide. *Health Services Res* 1986;21:215–237.
- Haynes R.B., McKibbin K.A., Fitzgerald D., Guyatt G.H., Walker C.J., Sackett D.L. How to keep up with the medical literature. I: Why try to keep up and how to get started. *Ann Intern Med* 1986;105:149–153.
- Haynes R.B., McKibbin K.A., Fitzgerald D., Guyatt G.H., Walker C.J., Sackett D.L. How to keep up with the medical literature. II: Deciding which journals to read regularly. *Ann Intern Med* 1986;105:309–312.
- Haynes R.B., McKibbin K.A., Fitzgerald D., Guyatt G.H., Walker C.J., Sackett D.L. How to keep

up with the medical literature.III: Expanding the number of journals you read regularly. Ann Intern Med 1986;105:474–478.

Haynes R.B., McKibbin K.A., Fitzgerald D., Guyatt G.H., Walker C.J., Sackett D.L. How to keep up with the medical literature. IV: Using the literature to solve clinical problems. Ann Intern Med 1986;105:636–640.

Haynes R.B., McKibbin K.A., Walker C.J., Mousseau J., Baker L.M., Fitzgerald D., Guyatt G., Norman G.R. Computer searching of the medical literature: an evaluation of MEDLINE searching systems. Ann Intern Med 1985;103:812–816.

Irwig L., Tosteson A.N., Gastonis C., Lau J., Colditz G., Chalmers T.C., Mostellar F. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994;120:667–676.

L'Abbe K.A., Detsky A.S., O'Rourke K. Meta-analysis in clinical research. Ann Intern Med 1987;107:224–233.

Light R.J., Pillmer D.B. Summing up. The science of reviewing research. Cambridge, MA: Harvard University Press, 1984.

Mulrow C.D. State of the science [The Medical Review Article]. Ann Intern Med 1987;106:485–488.

Thompson S.G., Pocock S.J. Can meta-analyses be trusted? Lancet 1991;338:1127–1130.

Основные рекомендации для определения достоверности клинических исследований

Все исследования

1. Какой *клинический вопрос* поставлен в исследовании?
 - Методика исследования должна соответствовать этому клиническому вопросу (см. табл.12.1).
2. Какие *пациенты, переменные* и *исходы* изучались?
 - Эти категории определяют обобщаемость результатов.
3. Насколько вероятно наличие систематической ошибки?
 - Систематические различия между сравниваемыми группами (например, по характеристикам больных, вмешательствам или факторам риска; по типам исходов; по методам измерений) снижают достоверность исследований.
4. Насколько значителен эффект?
 - Клинические решения зависят от величины эффекта, а не просто от его наличия.
5. Какова вероятность того, что результаты обусловлены случайностью?
 - Врач должен знать, в каком диапазоне (доверительный интервал) с заданной вероятностью находятся истинные значения или (что менее полезно) насколько вероятно, что наблюдаемый эффект имел место вследствие случайности (значение p для "положительных" результатов и статистическая мощность для "отрицательных" результатов).

Исследования методов диагностики

1. Имеется ли *четкое описание* теста (включая точку разделения нормы и патологии)?

- Если результат теста может быть представлен диапазоном значений, то заключение зависит от выбора точки разделения.
2. Установлено ли для всех больных истинное наличие или отсутствие заболевания (с помощью метода, являющегося “золотым стандартом”)?
 - Установить все важные характеристики теста возможно, только если имеются данные для всех клеток четырехпольной таблицы.
 3. Соответствуют ли *характеристики группы*, на которой изучался тест, характеристикам тех пациентов, у которых тест будет использоваться на практике?
 - Чувствительность теста часто зависит от степени тяжести болезни, а специфичность – от доли обследованных лиц без заболевания.
 4. Исключена ли систематическая ошибка при оценке теста и заболевания?
 - Систематическая ошибка может возникнуть, если оценка результатов теста проводится при наличии сведений о болезни и наоборот.
 5. Приведены ли данные об *эффективности теста* – чувствительность, специфичность или отношение правдоподобия?
 - Эта информация нужна, чтобы решить, использовать ли тест в клинической практике.
 6. Для тестов с диапазоном значений: как влияет на эффективность теста выбор точки разделения?
 - Информация, которую обеспечивает тест, зависит от степени отклонения от нормы.
 7. Если сообщается прогностическая ценность результатов теста, соотносится ли она с распространенностью заболевания на практике?
 - Прогностическая ценность результата теста зависит от распространенности заболевания (а также от чувствительности и специфичности теста). Если группы лиц с заболеванием и без него набираются отдельно, без учета реальной распространенности, то получаемая прогностическая ценность не имеет клинического смысла.

Одномоментные исследования

1. Каковы критерии выявления *случая*?
 - Распространенность зависит от того, какой смысл вкладывается в понятие случая.
2. В какой *популяции* выявлены случаи?
 - Распространенность зависит от характеристик группы лиц, в которой она оценивается.

3. Изучается ли распространенность в несмещенной выборке из популяции?
 - Распространенность для выборки отражает распространенность в популяции с точностью, соответствующей репрезентативности выборки.

Когортные исследования

1. Все ли члены когорты:
 - а) включены в начале исследования (*исходная когорта*)?
 - В противном случае не будут учтены пациенты, у которых заболевание протекает слишком легко или слишком тяжело;
 - б) имеют *риск* развития исхода?
 - Нет смысла описывать развитие исходов у лиц, которые уже заболели или не могут заболеть;
 - в) находятся в одинаковой *точке отсчета (по времени)* в течении заболевания?
 - Прогноз зависит от того, с какого момента в процессе заболевания начинается отсчет исходов.
2. Все ли члены когорты *завершили* исследование?
 - Выбывание членов когорты из исследования может привести к появлению систематической ошибки, если у выбывших в среднем болезнь протекает тяжелее или, наоборот, легче, чем у завершивших исследование.
3. У всех ли членов когорты выявление исходов проводилось с *одинаковой тщательностью*?
 - В противном случае различия в частоте исходов могут не отражать истинные различия, а быть обусловлены систематической ошибкой при измерении.
4. Исключена ли систематическая ошибка при *сравнении*? (Были ли члены сопоставляемых когорт одинаковыми по всем признакам, кроме изучаемого?)
 - Для того чтобы приписать исход изучаемому фактору, другие влияющие на исход факторы должны быть представлены в сравниваемых группах в равной мере.

Рандомизированные испытания

1. Соблюдены ли основные *требования к когортным исследованиям*?
 - Клинические испытания – разновидность когортных исследований.
2. *Случайным ли образом* больные разделены на опытную и контрольную группы?
 - Это единственный способ сделать сравнение методов лечения полностью свободным от систематической ошибки.

3. Проводилось ли исследование слепым методом, т.е. знали ли больные, клинический персонал и исследователи к какой группе принадлежал каждый испытуемый?
 - Неосведомленность участников испытания позволяет получить несмещенную оценку.
4. Обеспечивались ли в обеих группах *одинаковые методы всех вмешательств*, кроме изучаемого?
 - Различия условий для двух групп могут нарушить сопоставимость, которая была достигнута путем рандомизации.
5. Какой подход анализа данных применялся: *исходя из допущения, что все больные получили лечение, предписанное при рандомизации, или исходя из фактически полученного лечения*?
 - Если не все больные получали предписанное лечение, то существует два вида анализа данных, с разной задачей и разной научной силой. Анализ с учетом назначенного вмешательства важен для принятия клинических решений и проводится на группах, сформированных случайным образом. Анализ с учетом фактически полученного вмешательства позволяет объяснить эффект вмешательства и является когортным исследованием.

Исследование случай–контроль

1. Производилось ли включение пациентов в *начале* заболевания?
 - Факторы риска для распространенных заболеваний могут быть связаны с началом или с длительностью заболевания.
2. *Одинаковы ли* основная и контрольная группы по всем признакам, кроме изучаемого фактора?
 - Для обоснованной оценки относительного риска необходима сопоставимость сравниваемых групп.
3. Применялись ли в основной и контрольной группах *одинаковые и не подверженные систематической ошибке методы выявления воздействия* изучаемого фактора?
 - Систематическая ошибка при оценке воздействия может увеличить или уменьшить величину относительного риска.

Мета-анализ

1. *Все ли исследования* (опубликованные и неопубликованные) по изучаемому вопросу выявлены?
 - Задача состоит в обобщении результатов всех завершенных исследований, а не смещенной выборки из них.
2. Включает ли мета-анализ *только строгие научно обоснованные* исследования (с низкой вероятностью систематической ошибки)?

– Задача состоит в обобщении наиболее надежных результатов.

3. Если в мета-анализе оценивается суммарный эффект:
 - а) *однородны ли* включенные в мета-анализ исследования (с точки зрения характеристик больных, типов вмешательств и исходов)?
 - Неправомочно выводить единую общую меру эффекта на основе разных по сути исследований.
 - б) *взвешены ли* включаемые исследования по размеру выборки (числу наблюдений)?
 - Более крупные (более точные) исследования получают больший вес, чем небольшие (менее точные) исследования.
4. *Учитывается ли качество* включаемого исследования при оценке его результата?
 - Результаты хорошо спланированных и проведенных исследований заслуживают большего доверия.

СЛОВАРЬ ТЕРМИНОВ

- **Альфа-ошибка (α -error),** или **ошибка I рода (Type I error)** – вероятность ошибочного отклонения нулевой гипотезы (см. *Проверка гипотез*). В клинических исследованиях это вероятность того, что одно лечение будет считаться эффективнее другого, когда в действительности это различие отсутствует. Используя юридическую терминологию, можно сказать, что альфа-ошибка – это вероятность наказания невиновного.
См. стр. 231, 239.
- **Анализ в зависимости от назначенного лечения (intention to treat analysis)** – способ анализа данных в контролируемом испытании. Анализ проводится в зависимости от того, к какой группе – экспериментальной или контрольной – был отнесен пациент при рандомизации, независимо от того, получил он в действительности исследуемое лечение или нет.
См. стр. 188, 202.
- **Анализ дожития (survival analysis, time-to-event analysis)** – статистический метод анализа времени, прошедшего от заданного начального момента до момента наступления определенного исхода (обычно – смерти). Метод позволяет описать любой исход, дихотомический по своему характеру и имеющий место лишь однажды в процессе наблюдения, например время до развития инфаркта миокарда или до рецидива злокачественного новообразования. Преимущество метода заключается в том, что в расчетах используются сведения обо всех пациентах, в том числе еще не умерших к моменту проведения анализа или утерянных в процессе наблюдения, что позволяет оценить среднее время выживания более точно. К способам анализа времени наступления события относятся анализ Каплана–Мейера (Kaplan–Meier analysis), регрессионная модель пропорционального риска Кокса (Cox proportional hazards regression model), лог-ранк тест (log-rank test) и др.
См. стр. 149, 150, 154.

- **Анализ принятия решений (decision analysis)** – статистический подход к принятию решений, осуществляемый путем сопоставления альтернатив с учетом затрат и последствий для больного в заданных условиях.
См. стр. 28, 115, 118–119.
- **Анализ чувствительности (sensitivity analysis)** – метод, используемый в мета-анализе, фармакоэкономике и анализе принятия решений для оценки влияния различных параметров на конечный результат. Анализ *наилучший вариант–наихудший вариант* (best case / worst case analysis) представляет собой специальный вид анализа чувствительности.
См. стр. 117, 166, 168.
- Апостериорная вероятность – см. *Прогностическая ценность*.
- Априорная вероятность – см. *Распространенность*.
- **Бета-ошибка (β -error), или ошибка II рода (Type II error)** – вероятность ошибочного принятия нулевой гипотезы (см. *Проверка гипотез*). В клинических исследованиях это вероятность того, что оба лечения будут считаться одинаково эффективными, когда в действительности одно лечение лучше, чем другое. Используя юридическую терминологию, можно сказать, что бета-ошибка – это вероятность отпустить виновного на свободу (например за недостаточностью улик).
См. стр. 231–233, 238–239, 241, 317.
- **Величина p (p value)** – вероятность того, что полученный результат абсолютно случаен. Величина p может изменяться от 1 (результат точно случаен) до 0 (результат точно не случаен). Величина p меньшая или равная заданному уровню альфа-ошибки (например 0,05) свидетельствует о статистической значимости полученного различия. В то же время она не несет информации о клинической значимости результата.
См. стр. 232–233, 246.
- **Величина p_α** – вероятность альфа-ошибки. Обычно, когда пишут p , имеют в виду именно p_α .
См. стр. 233.
- **Величина p_β** – вероятность бета-ошибки.
См. стр. 233, 238–240

- **Вероятностная выборка (probability sample)** – такая выборка, вероятность попадания в которую для каждого индивидуума определена заранее.
См. стр. 109.
- Вероятность, посттестовая – см. *Прогностическая ценность.*
- Вероятность, претестовая – см. *Распространенность.*
- **Вмешивающийся фактор (confounding factor)** – фактор, связанный с известным исследуемым фактором и некоторым образом влияющий на результат (исход). Например известно, что мужчины чаще, чем женщины болеют ИБС. Однако это может быть связано не с исследуемым фактором (пол), а с тем, допустим, что мужчины чаще курят, больше подвергаются стрессам, потребляют больше пищи, богатой холестерином. Влияние вмешивающихся факторов приводит к возникновению систематической ошибки (confounding bias).
См. стр. 20–21, 289–290, 295–297.
- **Вторичная профилактика (secondary prevention)** – это вмешательства, направленные на то, чтобы замедлить или остановить развитие заболевания у больных.
См. стр. 207, 227.
- **Выборка (sample)** – это часть популяции, полученная путем отбора. Клинические исследования обычно выполняются на выборках. Оценку характеристик популяции по практическим причинам приходится осуществлять путем оценки этих характеристик по выборке.
См. стр. 17, 109.
- Выборка, вероятностная – см. *Вероятностная выборка.*
- Выборка, случайная – см. *Случайная выборка.*
- Выборка, смещенная – см. *Смещенная выборка.*
- Выборки, размер – см. *Размер выборки.*
- Группа, контрольная – см. *Контрольная группа.*
- Группа, экспериментальная – см. *Экспериментальная группа.*

- Данные, количественные – см. *Количественные данные*.
- Данные, дискретные – см. *Дискретные данные*.
- Данные, дихотомические – см. *Дихотомические данные*.
- Данные, непрерывные – см. *Непрерывные данные*.
- Данные, порядковые – см. *Порядковые данные*.
- **Дискретные данные (discrete data)** – это количественные данные, которые выражаются целыми числами. Например: число беременностей; число рождений живых младенцев; число судорожных припадков у больного в течение месяца.
См. стр. 35.
- **Дихотомические данные (dichotomous data)** – признаки, которые могут иметь только два значения (присутствует–отсутствует, да–нет, жив–умер).
См. стр. 34.
- **Добавочная доля популяционного риска (population attributable risk fraction)** – доля заболеваемости (смертности, инвалидности) в популяции, связанная с данным фактором риска; рассчитывается путем деления добавочного популяционного риска на общую заболеваемость (смертность, инвалидность) в популяции.
См. стр. 134–135.
- **Добавочный, или атрибутивный риск (attributable risk)** – определяется как заболеваемость для лиц, подвергавшихся воздействию фактора риска, минус заболеваемость для лиц, не подвергавшихся этому воздействию. Добавочный риск – это дополнительные случаи развития заболевания, обусловленные воздействием фактора риска. Учитывая способ вычисления добавочного риска, его также называют разницей рисков (risk difference), или непосредственным риском.
См. стр. 134, 136.
- **Доверительный интервал (confidence interval)** – статистический показатель, позволяющий оценить в каких пределах может находиться истинное значение параметра в популяции; диапазон колебаний истинных значений. Величины, полученные

в исследованиях на выборке больных, отличаются от истинных величин в популяции вследствие влияния случайности. Так, 95% доверительный интервал означает, что истинное значение величины с вероятностью в 95% лежит в его пределах. Доверительные интервалы помогают сориентироваться, соответствует ли данный диапазон значений представлениям читателя о клинической значимости эффекта и каких результатов можно ожидать, применив описанную методику на сходной группе больных. Величина доверительного интервала характеризует степень доказательности данных, в то время как величина *p* указывает на вероятность ошибочного отклонения нулевой гипотезы (см. *Величина p*).
См. стр. 243, 318, 320.

- **Достоверность (validity)** – характеристика, показывающая в какой мере результат измерения соответствует истинной величине. Достоверность исследования определяется тем, в какой мере полученные результаты справедливы в отношении данной выборки (internal validity). Это внутренняя характеристика, она касается именно данной группы больных и не обязательно распространяется на другие группы.
См. стр. 24, 35, 40, 58, 173–175, 185, 196–197, 203, 234–235, 249, 269, 275, 251, 290, 306, 310, 323.
- **Исследование до–после (before-after study)** – неконтролируемое клиническое исследование, описывает течение заболевания в одной группе пациентов, подвергающейся изучаемому вмешательству. Подход основан на предположении, что любое улучшение, наблюдаемое после лечения, обусловлено именно лечением. Это предположение может оказаться ложным, что делает указанный метод весьма уязвимым.
См. стр. 199.
- Исследование распространенности (prevalence study) – см. *Одномоментное исследование*.
- **Исследование серий случаев (case series)** – описательное исследование, представляющее собой количественный анализ группы больных.
См. стр. 157, 225, 242, 255, 259, 261, 277.
- **Исследование случай–контроль (case control study)** – ретроспективное исследование, в котором по архивным данным,

воспоминаниям или суждениям пациентов производится сравнение двух групп, в одну из которых отобраны пациенты с определенной патологией, а в другую – лица без нее. При этом сначала производится отбор группы пациентов с изучаемым заболеванием и сходной по остальным признакам группы лиц без этого заболевания. Затем ретроспективно оценивается частота воздействия возможного фактора риска в обеих группах. Полученные данные позволяют рассчитать относительный риск развития заболевания в связи с изучаемым фактором.

См. стр. 260.

- **Исследование совокупного риска (aggregate risk study), или экологическое исследование (ecological study)** – исследование, в котором фактор риска характеризуется средним воздействием на группу; его задача – оценка влияния на людей факторов окружающей среды в целом.
См. стр. 290.
- **Исторический, или непараллельный контроль (historical or nonconcurrent control)** – способ клинического исследования, при котором результаты современного лечения сравниваются с данными наблюдения аналогичных пациентов в прошлом. Противоположность параллельному контролю (concurrent control), при котором контрольная группа набирается одновременно и по тем же правилам, что и экспериментальная группа
См. стр. 196.
- **Историческое когортное исследование (historical cohort study)** – когорта выделена по архивным документам и прослежена до настоящего времени.
См. стр. 131–132.
- **Качественные данные (nominal data)** – это такие признаки, которые нельзя выразить количественно. Например: диагноз, группа крови, страна проживания. Частным случаем качественных данных являются *дихотомические данные*.
См. стр. 34.
- **Клиническая эпидемиология (clinical epidemiology)** – наука, разрабатывающая методы клинических исследований, которые дают возможность делать справедливые заключения, контролируя влияние систематических и случайных ошибок.
См. стр. 12.

- **Клинические испытания (clinical trials)** – специальный вид когортных исследований для оценки разных вмешательств, условия проведения которых (отбор группы вмешательства, характер вмешательства, организация наблюдения и оценка исходов) направлены на устранение влияния систематических ошибок на получаемые результаты.
См. стр. 173.
- **Когорта (cohort)** – группа лиц, изначально объединенных каким-либо общим признаком (например: здоровые лица или больные на определенной стадии заболевания) и наблюдаемых в течение определенного периода времени, чтобы проследить, что с ними произойдет в дальнейшем.
См. стр. 129.
- **Когорта дожития (survival cohort)** – когорта, в которую включаются пациенты, имеющие заболевание и доступные для наблюдения – например пациенты специализированной клиники. Другой термин для таких групп – *доступные когорты* (available patients cohort).
См. стр. 156.
- **Когортное исследование (cohort study)** – исследование, в котором определенная когорта пациентов прослеживается в течение некоторого периода времени. Когортные исследования называют также *продольными*, или *лонгитудинальными* (longitudinal study) – подчеркивается, что пациенты прослеживаются во времени, *проспективными* (prospective study) – имеется в виду, что группа сформирована в настоящее время и прослежена в будущем, или *исследованиями заболеваемости* (incidence study) – обращается внимание на то, что основным способом оценки является регистрация новых случаев заболевания в течение определенного срока.
См. стр. 130.
- **Количественные данные (numerical, or interval data)** – это такие величины, которым присущ естественный порядок расположения с равными интервалами между последовательными значениями, независимо от их места на шкале. Примеры: масса тела, концентрация глюкозы в крови.
См. стр. 34.
- Контроль, параллельный – см. *Параллельный контроль*.

- Контроль, исторический – см. *Исторический контроль*.
- Контроль, непараллельный – см. *Исторический контроль*.
- **Контрольная группа (control group)**, или **группа сравнения** – группа испытуемых, получающих обычное лечение, или не получающих лечения, или получающих плацебо. Результаты измерений в контрольной группе сравниваются с результатами измерений в экспериментальной группе для оценки эффекта исследуемого метода лечения.
См. стр. 174.
- **Мета-анализ (meta-analysis)** – количественный анализ объединенных результатов нескольких клинических испытаний одного и того же вмешательства. Такой подход обеспечивает большую статистическую мощь, чем в каждом отдельном испытании за счет увеличения размера выборки. Используется для обобщенного представления результатов многих испытаний и для увеличения доказательности результатов испытаний.
См. стр. 305, 314–319.
- **Многофакторное моделирование (multivariable modeling)** – метод, обеспечивающий математическое выражение сочетанного влияния многих переменных. Метод называется многофакторным, так как он оценивает эффект многих переменных. Термин "моделирование" означает, что это математическая конструкция, которая рассчитывается из полученных данных, но основана она на упрощающих допущениях о характеристиках этих данных (например, что все переменные имеют нормальное распределение и одинаковую дисперсию).
См. стр. 250.
- **Многофакторный анализ (multivariable analysis)** – это совокупность статистических методов, которые одновременно рассматривают влияние многих переменных на какой-либо один фактор. Если после устранения влияния этих переменных действие фактора сохраняется, его воздействие считается независимым. Кроме того эти методы применяются для выделения из большого числа признаков малого подмножества, которое вносит независимый и существенный вклад в конечный результат (исход), что используется для ранжирования переменных по силе их воздействия на исход и для других целей.
См. стр. 165, 167–168, 249.

- **Модификация эффекта (effect modification)** – особый тип взаимодействия факторов, при котором сила корреляции между двумя переменными зависит от уровня некоторой третьей переменной, называемой *модификатором эффекта* (effect modifier).
См. стр. 287.
- **Мощность, статистическая** – см. *Статистическая мощность*.
- **Непрерывные данные (continuous data)** – это количественные данные, которые могут принимать любое значение на непрерывной шкале. Примеры: масса тела, артериальное давление, парциальное давление кислорода в артериальной крови.
См. стр. 35.
- **Обобщаемость (external validity, or generalizability)** – внешняя характеристика, которая определяется тем, в какой мере результаты данного исследования применимы к другим группам больных.
См. стр. 24, 109, 162, 167, 175, 195, 202, 306, 310.
- **Обсервационное исследование (observational studies)** – исследование без преднамеренного вмешательства, в противоположность экспериментальному исследованию.
См. стр. 129.
- **Одномоментное исследование (cross-sectional study, prevalence study, survey)** (иногда также называется **поперечным исследованием, в противоположность продольным, или лонгитудинальным, исследованиям**) – вариант описательного исследования, проводимого в определенный момент времени с целью оценки распространенности заболевания или исхода, изучения течения заболевания и т.п.
См. стр. 103, 111, 113–114, 118, 171, 290.
- **Описание случая (case report, case study)** – простое изложение материала об одном или нескольких больных.
См. стр. 255.
- **Открытое исследование (open, open-label, unmasked study)** – исследование, в котором не применялся слепой метод.
См. стр. 185.

- **Относительный риск (relative risk)**, или **отношение рисков (risk ratio)** – отношение заболеваемости среди лиц, подвергавшихся и не подвергавшихся воздействию факторов риска. Относительный риск не несет информации о величине абсолютного риска (заболеваемости). Даже при высоких значениях относительного риска абсолютный риск может быть совсем небольшим, если заболевание редкое. Относительный риск показывает силу связи между воздействием и заболеванием.
См. стр. 135.
- **Отношение правдоподобия (likelihood ratio)** – это отношение вероятности получить положительный результат диагностического теста у больных к вероятности получить положительный результат теста у здоровых лиц. Отношение правдоподобия для положительного результата теста – это чувствительность деленная на 1 минус специфичность. Таким образом, отношение правдоподобия отражает одновременно и чувствительность, и специфичность теста. Если отношение правдоподобия положительного результата теста равно 1, то это значит, что вероятность положительного результата теста у больного такая же, как вероятность положительного результата теста у здорового. Если отношение правдоподобия положительного результата теста равно 3,5, то это значит, что вероятность положительного результата теста у больного в 3,5 раза выше, чем вероятность положительного результата теста у здорового.
См. стр. 86.
- **Отношение шансов (odds ratio)** – определяется как отношение шансов события в одной группе к шансам события в другой группе, или как отношение шансов того, что событие произойдет к шансам того, что событие не произойдет. В исследованиях случай–контроль отношение шансов используется для оценки относительного риска.
См. стр. 264–266.
- Ошибка, систематическая – см. *Систематическая ошибка.*
- Ошибка, случайная – см. *Случайная ошибка.*
- Ошибка, экологическая – см. *Экологическая ошибка.*
- **Параллельный контроль (concurrent control)** – способ клинического исследования, при котором контрольная группа на-

бирается одновременно и по тем же правилам, что и экспериментальная группа. Этот метод позволяет избежать некоторых видов систематических ошибок, которые неизбежны при исследованиях с историческим контролем.

См. стр. 197.

- **Первичная профилактика (primary prevention)** – вмешательства, направленные на предупреждение возникновения болезни, обычно за счет устранения ее причин.

См. стр. 207.

- **Плацебо (placebo)** – лекарственная форма, неотличимая от исследуемого препарата по внешнему виду, цвету, вкусу и запаху, но не оказывающая специфического действия (например таблетки глюкозы или инъекции изотонического раствора хлорида натрия), или иное безразличное вмешательство, используемое в медицинских исследованиях для имитации лечения с целью устранения систематической ошибки, связанной с *плацебо-эффектом*.

См. стр. 178.

- **Плацебо-эффект (placebo effect)** – изменение состояния пациента (отмечаемое самим пациентом или лечащим врачом), связанное с фактом лечения, а не с биологическим действием препарата.

См. стр. 178.

- **Популяционный добавочный (атрибутивный) риск (population attributable risk)** – рассчитывается как произведение добавочного риска на распространенность фактора риска в популяции. Этот показатель отражает дополнительную заболеваемость в популяции, связанную с фактором риска.

См. стр. 134–137.

- **Популяция (population)** – это совокупность индивидуумов, из которой отбирается выборка и на которую могут быть распространены результаты, полученные для этой выборки. Популяция может представлять собой все население (обычно таковы популяции в эпидемиологических исследованиях причин заболеваний), или же состоять из пациентов, госпитализированных в определенную клинику, или из пациентов с определенным заболеванием (что чаще имеет место в клинических исследованиях). Таким образом, можно говорить об общей популя-

ции, госпитальной популяции или популяции пациентов с конкретным заболеванием. Эпидемиологическое определение популяции отличается от биологического (экологического).

См. стр. 17.

- **Порядковые данные (ordinal data)** – это такие показатели, которые могут быть расположены в естественном порядке (ранжированы), например от малого до большого или от хорошего до плохого, но размер интервала между такими категориями не может быть выражен количественно (например: стадии болезни; оценки – "высокий, средний, низкий" или "отсутствует, слабый, умеренный, тяжелый").

См. стр. 34.

- Посттестовая вероятность – см. *Прогностическая ценность*.

- Претестовая вероятность – см. *Распространенность*.

- **Проверка гипотез (hypothesis testing) о равенстве средних** – метод статистического доказательства гипотезы о равенстве средних величин. Нулевая гипотеза (H_0) заключается в том, что различий между двумя выборками, оцененным по их средним значениям, нет. Противоположная гипотеза (H_1) утверждает обратное. Принятие нулевой гипотезы означает, что обнаруженная разница случайна, отклонение нулевой гипотезы – что разница неслучайна. Вероятность ошибочного отклонения нулевой гипотезы называется *альфа-ошибкой*, уровень статистической значимости обнаруженных различий обозначается как величина p , вероятность ошибочного принятия нулевой гипотезы называется *бета-ошибкой*.

См. стр. 230, 252, 311.

- **Прогностическая ценность (predictive value)** – вероятность наличия заболевания при известном результате диагностического теста. Прогностическая ценность называется также *априорной* (или *посттестовой*) *вероятностью*, поскольку это вероятность наличия (или отсутствия) болезни после того, как стали известны результаты теста.

См. стр. 76–77, 88, 94.

- **Прогностическая ценность отрицательного результата (negative predictive value)** – вероятность отсутствия заболевания при отрицательном ("нормальном") результате теста.

См. стр. 77–78.

- **Прогностическая ценность положительного результата (positive predictive value)** – вероятность заболевания при положительном ("ненормальном") результате теста.
См. стр. 77–78.
- Профилактика, вторичная – см. *Вторичная профилактика.*
- Профилактика, первичная – см. *Первичная профилактика.*
- Профилактика, третичная – см. *Третичная профилактика.*
- p – см. *Величина p .*
- p_α – см. *Величина p_α .*
- p_β – см. *Величина p_β .*
- **Размер выборки (sample size)** – число больных, которое необходимо включить в исследование для исключения случайности в качестве объяснения полученного результата. Размер выборки зависит от четырех характеристик исследования: величины различия в частоте исходов между группами, p_α , p_β и природы данных. Размер выборки следует учитывать исследователю, планирующему эксперимент, и читателю, решающему, следует ли доверять опубликованным результатам.
См. стр. 238.
- **Рандомизация (randomization)** – процедура, обеспечивающая случайное распределение больных в экспериментальную и контрольную группы. Случайным распределением достигается отсутствие различий между двумя группами и, таким образом, снижается вероятность систематической ошибки в клинических исследованиях вследствие различий групп по каким-либо признакам.
См. стр. 162, 180.
- **Распространенность (prevalence)** – это частота некоторого состояния в группе. Рассчитывается как отношение числа лиц, у которых на момент обследования наблюдается изучаемое состояние (болезнь или исход), к числу всех лиц в группе (популяции, обследованным больным, работникам предприятия). Распространенность указывает также на априорную, или претестовую вероятность заболевания или исхода.
См. стр. 78, 80, 81, 84, 86, 100, 103, 110, 114.
- Риск, добавочный (атрибутивный) – см. *Добавочный риск.*

- Риск, относительный – см. *Относительный риск*.
- Риск, популяционный добавочный (атрибутивный) – см. *Популяционный добавочный риск*.
- **Систематическая ошибка, смещение (systematic error, bias)** – это неслучайное, однонаправленное отклонение результатов от истинных значений. Систематическая ошибка может возникнуть вследствие отбора, т.е. при создании выборки (sampling, или assembling bias), вследствие измерений (measurement bias), при воздействии вмешивающихся факторов (confounding bias) и во многих других случаях. О систематической ошибке говорят также имея в виду предвзятость при публикации положительных результатов исследований и отклонении отрицательных – publication bias. Для борьбы с систематическими ошибками и получения достоверных данных используются организационные методы (например: рандомизация, слепой метод и т.п.), а также внесение поправок, учитывающих величину смещения.
См. стр. 14, 18, 75, 111, 123, 154, 219, 269, 319.
- **Скрининг (screening)** – массовое обследование лиц, не считающих себя больными, для выявления скрыто протекающих заболеваний или других состояний (факторов риска будущих заболеваний). Обычно проводится с использованием дешевых, простых, неинвазивных диагностических процедур, имеющих высокую чувствительность.
См. стр. 64, 51, 127, 143, 207–218, 297, 308.
- **Слепой, или маскированный, метод (blinding, or masking)** – процедура, обеспечивающая отсутствие информации о том, к какой группе – экспериментальной или контрольной – отнесен каждый испытуемый. При простом слепом методе информация отсутствует только у пациента, при двойном слепом – у пациента и исследователя, при тройном слепом – у пациента, исследователя и лиц, проводящих статистическую обработку результатов исследования. Применяется для устранения систематической ошибки в клинических исследованиях.
См. стр. 185, 289.
- **Случайная выборка (random sample)** – такая выборка, вероятность попадания в которую для каждого индивидуума в популяции одинакова.
См. стр. 109.

- **Случайная изменчивость**, или **случайная вариация (random variation)** – отклонение результата отдельного наблюдения (измерения) от его истинного значения, обусловленное исключительно случайностью.
См. стр. 39.
- Смещение – см. *Систематическая ошибка.*
- **Смещение к среднему (regression to the mean)** – закономерность в биомедицинских исследованиях, выражающаяся в том, что величины, отклоняющиеся от среднего значения, при последующих измерениях оказываются ближе к среднему.
См. стр. 56.
- **Смещенная выборка (biased sample)** – это такая выборка, которая систематическим образом отличается от популяции, представляющей предмет исследования, или от популяции, по отношению к которой должны применяться результаты исследования. Например лица, добровольно участвующие в программах профилактики заболеваний, отличаются от прочих людей в популяции и представляют собою смещенную выборку из популяции.
См. стр. 26–27.
- **Специфичность диагностического теста (specificity)** – вероятность отрицательного результата диагностического теста в отсутствие болезни.
См. стр. 66–74.
- **Статистика (statistics)** – это специальная область математики со своей, незнакомой многим врачам терминологией: нулевая гипотеза, дисперсия, регрессия, мощность, моделирование и т.п.
См. стр. 232.
- **Статистическая мощность исследования (statistical power)** – вероятность того, что в исследовании будет найдено статистически достоверное различие, когда это различие действительно существует. Статистическая мощность равняется $1 - \text{минус } p_{\beta}$ (см. *Величина p_{β}*). Мощность исследования аналогична чувствительности диагностического теста.
См. стр. 238, 246, 315, 317.

- **Точка разделения (cut-off point, cutpoint)** – величина, используемая для разделения ряда величин на две части. В диагностическом процессе точка разделения отделяет "нормальные" показатели от "ненормальных". Результаты теста, лежащие в области "нормы", называются отрицательным результатом, а лежащие в области "патологии" – положительным результатом.
См. стр. 58, 69.
- **Точность теста (test accuracy)** – доля правильных результатов теста (истинно положительных и истинно отрицательных) в общем количестве полученных результатов.
См. стр. 79, 96.
- **Третичная профилактика (tertiary prevention)** – мероприятия, направленные на предотвращение ухудшения течения или осложнений заболевания после того, как болезнь проявилась.
См. стр. 208.
- **Фактор риска (risk factor)** – особенность организма или внешнее воздействие, приводящие к увеличению риска возникновения заболевания или иному неблагоприятному исходу.
См. стр. 121.
- **Характеристическая кривая (receiver operating characteristic, or ROC curve)** – график, характеризующий диагностическую точность теста. На оси ординат откладывается чувствительность (доля истинно положительных результатов), на оси абсцисс – 1 минус специфичность (доля ложноположительных результатов). Чем дальше кривая отстоит от диагонали, тем выше точность теста.
См. стр. 70, 95.
- **Центральная тенденция (central tendency)** – статистический показатель, наиболее полно отражающий величины, наблюдаемые в выборке или популяции. К таким показателям относятся: мода (наиболее часто встречающаяся величина), медиана (величина, занимающая середину в ряду ранжированных величин) и среднее значение.
См. стр. 45.
- **Частота новых случаев (incidence)** – это отношение числа лиц, у которых в течение определенного времени развилось

изучаемое состояние, ко всем обследованным в группе, где исходно этого состояния никто не имел. Частота новых случаев болезни в популяции, где это заболевание исходно отсутствовало, называется заболеваемостью. Частота новых случаев исходов при определенном заболевании указывает, соответственно, на инвалидность или летальность при этом заболевании. Для оценки частоты новых случаев отбирают группу лиц без заболевания (или без изучаемого исхода) и периодически обследуют ее, подсчитывая число новых случаев, появившихся за определенный период.

См. стр. 100.

- **Частотное распределение (frequency distribution)** – величины, расположенные в порядке возрастания и частота, с которой каждая величина встречается в популяции или выборке. Обычно представляется в виде графика, на котором по горизонтальной оси откладываются наблюдаемые величины, а по вертикальной – частоты, с которой эти величины наблюдаются.
См. стр. 44, 46, 48.
- **Число больных, которых необходимо лечить, ЧБНЛ (number needed to treat, NNT)** – способ оценки относительной эффективности двух методов лечения. Показывает, какое количество больных необходимо подвергнуть лечению исследуемым методом для предотвращения одного случая изучаемого исхода. Например, если для предотвращения одного случая инфаркта миокарда надо лечить антигипертензивным средством 33 пациентов с артериальной гипертензией в течение 5 лет, то ЧБНЛ равно 33 за 5 лет. Этот показатель является величиной, обратной снижению абсолютного риска.
См. стр. 187.
- **Чувствительность диагностического теста (sensitivity)** – вероятность положительного результата диагностического теста при наличии болезни.
См. стр. 66.
- Шансов, отношение – см. *Отношение шансов.*
- **Шансы (odds)** – 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. Если вероятность того, что событие

произойдет, обозначить P , то шансы этого события будут равны $P/(1-P)$. Например, если вероятность выздоровления 0,3, то шансы выздороветь равны $0,3/(1-0,3) = 0,43$. Для некоторых расчетов шансы удобнее использовать, чем вероятности.

См. стр. 85.

- **Экологическая ошибка (ecological fallacy)** – систематическая ошибка, связанная с тем, что заболевшие лица в группе обследования на самом деле могли не испытывать влияния факторов риска. Возникает, когда группу выделяют по общим косвенным признакам (проживание на одной территории, принадлежность к одной профессии и т.д.), а не по наличию фактора риска у отдельных субъектов (например по потреблению алкоголя).

См. стр. 292.

- **Экспериментальная группа (experimental group)** – это группа, подвергающаяся вмешательству (лечению) в ходе исследования. Иначе называется *группой лечения*, или *группой вмешательства* (treatment group, or intervention group).
- **Экспериментальное исследование (experimental study)** – сравнительное исследование, специально спланированное и посвященное изучению влияния по крайней мере одного вмешательства. Сравнения могут производиться между двумя и более группами или внутри одной группы до и после назначения вмешательства.
- **Эффект ярлыка (labeling)** – психологическое воздействие результатов теста или диагноза на пациента – участника скрининга.

См. стр. 209, 216–217.

Клиническая эпидемиология. Основы доказательной медицины.

Р. Флетчер, С. Флетчер, Э. Вагнер.

Перевод с английского под общей редакцией

канд. мед. наук С.Е. Бащинского

и канд. мед. наук С.Ю. Варшавского

Переводчики:

проф. А.Д. Деев, проф. Г.С. Жуковский, канд. мед. наук С.А.Шальнова

Редакторы перевода:

доктор биол. наук Е.М.Аксель, проф. Я.В.Двойрин,

доктор биол. наук М.А.Каменская, проф. Р.Г.Оганов, М.Н.Соловова

Научный консультант:

проф. В.В.Власов

Оригинал-макет и верстка:

О.П.Пономаренко

Векторная графика:

С.М.Камышов

Дизайн обложки:

О.Б.Захарова

Корректоры:

С.М.Казинцева, Т.В.Полухина, Т.Р.Тверитнева, В.И.Федорова

Оригинал-макет подготовлен в издательстве "Медиа Сфера"

Лицензия ЛП №064159

Подписано в печать 14.08.98. Формат 60 × 90/16

Объем 22,0 усл. печ. л. Тираж 10 000. Зак. №

Отпечатано с диапозитивов издательства "Медиа Сфера"

в Gummerus Printing, Finland, FIN-40351 Jyväskylä, P.O. Box 444.

Издательство "Медиа Сфера". Москва, Дмитровское ш. 46 - 2.