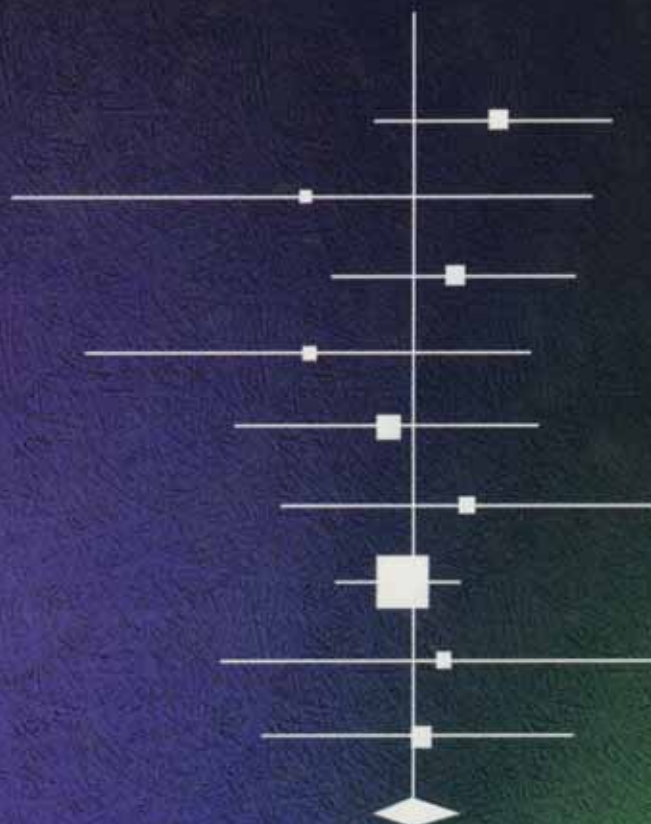


Р. ФЛЕТЧЕР, С. ФЛЕТЧЕР, Э. ВАГНЕР

Клиническая Эпидемиология

ОСНОВЫ ДОКАЗАТЕЛЬНОЙ МЕДИЦИНЫ



Издательство Медиа Сфера

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

к.м.н. С.Е. Башинского и к.м.н. С.Ю. Варшавского

Медиа Сфера

Москва 1998

УДК 616
ББК 519
Ф71

ОБЯЗАТЕЛЬНЫЙ
ЭКЗЕМПЛЯР

Перевод с английского:
проф. А.Д. Деев, проф. Г.С. Жуковский, к.м.н. С.А.Шальнова
Редакторы перевода:
д.б.н. Е.М.Аксель, проф. Я.В.Двойрин,
д.б.н. 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 №063375
Подписано в печать 14.04.98. Формат 60 × 90/16
Объем 22,0 бум. л. Тираж 2000. Зак. № 174
Отпечатано с диапозитивов издательства "Медиа Сфера"
в типографии ООО "Информполиграф"
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.

Книга написана одними из основателей новой дисциплины – клинической эпидемиологии. Рассматриваются основные положения этой науки и ее применение в медицинской практике: диагностике, лечении, профилактике. Книга богато иллюстрирована клиническими примерами, что облегчает ее понимание. В книгу включен список основных терминов клинической эпидемиологии с их подробным объяснением. Представляет интерес для практикующих врачей, исследователей, студентов и преподавателей медицинских институтов, организаторов здравоохранения.

527809
Книга содержит показания к применению медицинских препаратов, их дозировки, схемы введения и указания на побочные эффекты. Учитывая, что со временем эта информация может претерпевать изменения, 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 (руссск.)

© 1996 Williams & Wilkins
© Перевод на русский язык и оформление, "Медиа Сфера", 1997.

СОДЕРЖАНИЕ

	Предисловие	VI
	Предисловие к русскому изданию	VIII
	Благодарности	IX
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	230
Глава 10	Изучение случаев	257
Глава 11	Выявление причины	282
Глава 12	Подводя итоги	306
Приложение		
	Словарь терминов	333

ПРЕДИСЛОВИЕ

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 вся 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 наилучшей. 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. И пациенты и врачи сталкиваются с одними и теми же вопросами о диагнозе, прогнозе и лечении, и ставят прежние цели: облегчить страдания, восстановить функции и предупредить преждевременную смерть. Для ответов на эти вопросы мы полагаемся на основные типы 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. Мы не изменили построение книги в целом, но обновили 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 сделать изложение чуть более подробным в надежде на то, что теперь читатели ожидают от книги большего, чем в период становления этой области знаний.

Это издание — начальный курс и для клиницистов, и для начинающих исследователей, которые хотят понять, каким образом добываются и оцениваются медицинские данные. Книга должна быть полезной читателю независимо от уровня его клинической подготовки: будь то студент-медик или практикующий врач.

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 пневмонию 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 (эта 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 больше пользы, чем вреда. 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. Эта книга отражает и то, чему мы научились у них, и наш собственный опыт. Наши учителя – Archie Cochrane, John Cassel, Alvan Feinstein, David Sackett, Kerr White и другие – были пионерами в этой области.

Мы многому научились и у молодых врачей, которые верили в 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. Их число слишком велико и перечислить здесь каждого не представляется возможным.

Роберт и Сюзан Флетчеры особенно благодарны редакционному коллективу журнала *Annals of Internal Medicine*, где они были главными редакторами в течение четырех лет. Статьи в рецензируемых клинических журналах – письменное свидетельство развития клинической эпидемиологии. Мы многому научились при редактировании рукописей. Начало работы над третьим изданием книги пришлось на разгар нашего переезда из Филадельфии в Бостон. Мы благодарны докторам Steven Jones и Wendy Levinson – нашим хозяевам во время творческого отпуска в Орегоне. Tom Inui пригласил нас в наш новый академический дом – Гарвардскую школу медицины, и предоставил все возможности и ресурсы, необходимые для завершения работы над книгой.

Эдвард Вагнер благодарен за предоставленную возможность применить принципы клинической эпидемиологии при разработке и внедрении программы Group Health Cooperative of Puget Sound. Я благодарен за поддержку моим 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 коллег, которые помогли нам улучшить понимание клинической эпидемиологии и выразить его более ясно.

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 необходимостью 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 признание 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 сопровожда-

ется 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? Каковы его патогенетические механизмы?
Стоимость	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: 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 *p24*. 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 <i>CD4+</i>	Оппортунистические инфекции
	Дидезоксинозин	Антигенемия <i>p24</i>	Качество жизни
	Дидезоксицитидин	Вирусемия	Смерть
Связь известна или предполагается?			

Количественный подход

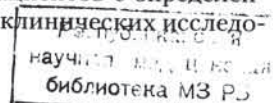
Клиническая наука особенно убедительна, когда может обеспечить количественный подход при измерениях. Отчасти это обусловлено тем, что количественные результаты дают более достоверные доказательства, позволяют оценить ошибку и облегчают обмен информацией как между врачами, так и между врачом и больным. 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 исследо-

527809



ваниях). 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, одностороннее) 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 истории 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 приемом 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 из-за вмешивающихся факторов: 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 пожелавших 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, пожелавших 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 внутриа-териальной 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 приходится 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.

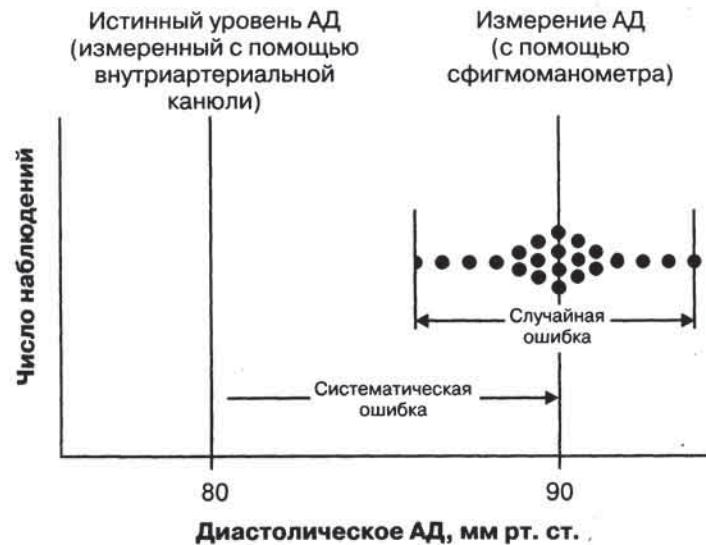


Рис. 1.2. Соотношение между систематической и случайной ошибкой: измерение артериального давления путем введения внутриаартериальной канюли и с помощью сфигмоманометра.

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 безупречно 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.

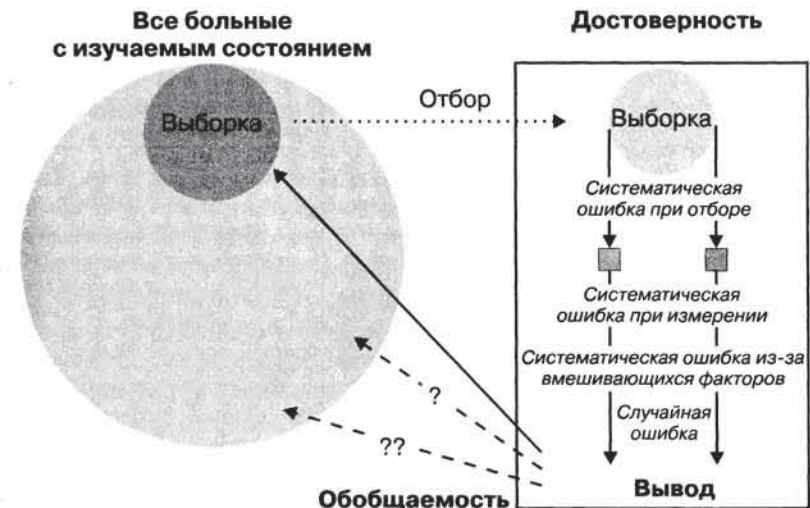


Рис. 1.3. Достоверность и обобщаемость.

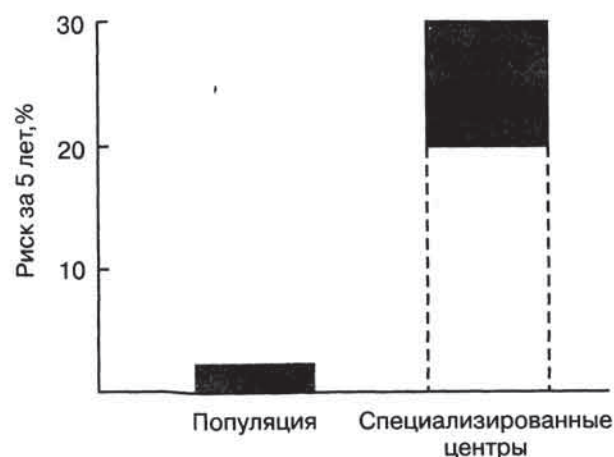


Рис. 1.4. Систематическая ошибка при отборе. Диапазон значений риска разрыва аневризмы брюшной аорты (закрашенная область) за 5 лет наблюдения больных с аневризмой (диаметром менее 5 см) в общей популяции и в специализированном центре [6].

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 многоцентрового исследования мо-

жет 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, и это приводит к тому, что 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 последовательно-сти 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 когортные исследования (cohort studies), 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

- White K.L. Healing the schism. Epidemiology, medicine and public health. New York: Springer-Verlag, 1991.
- Fletcher R.H. Clinical medicine meets modern epidemiology and both profit. *Ann Epidemiol* 1992;2:325–333.
- Murphy E.A. The logic of medicine. Baltimore: Johns Hopkins University Press, 1976.
- Sackett D.L. Bias in analytic research. *J Chronic Dis* 1979; 32:51–63.
- Jha PKS, 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.
- Ballard D.J. Abdominal aortic aneurysm [Letter]. *New Engl J Med* 1993; 329:1275.
- Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. *New Engl J Med* 1989;321:129–135.
- 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.

Sirithamorn 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.

Gehlbach 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. Clinimetric. St-Hyacinthe, Que., Canada: Edisen, 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 & 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 and 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 and 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 and 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 and 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.

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 зоб 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 врачам 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 эпиглоттита, вызванного *Haemophilus influenzae*, нередко приводящего к смерти. 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. Если патология не выявлена, то заключение в истории болезни звучит как "в пределах нормы", "без особенностей", "без существенных изменений". Если обнаруженные изменения расценены как патологические, то устанавливается предположительный или окончательный диагноз и назначается лечение.

Оценивать полученные клинические данные лишь как норму или патологию было бы слишком примитивно и даже ошибочно. Такой упрощенный подход оправдывается тем, что во многих случаях нет практической необходимости в детальном рассмотрении фактических данных. Бертран Рассел сказал: "Чтобы тебя хорошо понимали, приходится быть не совсем точным, а чтобы быть абсолютно точным, приходится быть непонятым". Врачи обычно выбирают крайность, предпочитая быть понятными для других и для самих себя, даже ценой известной потери точности. Другая причина стремления к упрощению данных состоит в том, что работа врача всегда заканчивается принятием решения – проводить обследование или выжидать, назначить лечение или просто успокоить пациента. Вследствие этого, использование некой упрощенной классификации типа "присутствует/отсутствует", становится необходимостью.

Из табл. 2.1, видно, как из сложных клинических данных выводятся относительно простые формулировки заключений о патологии. В левой части таблицы перечислены патологические состояния, в правой – некоторые данные, на основании которых принято решение о наличии этой патологии. Конкретные выводы, основанные на этих результатах, никак нельзя считать бесспорными. Например, среднее арифметическое значение из четырех измерений диастолического АД составляет 94 мм рт.ст. Некоторые могут возразить, что этот уровень недостаточен высок для диагноза артериальной гипертензии, к тому же есть свои отрица-

Таблица 2.1
Обобщение клинических данных: перечень патологических состояний
пациента и данные, на которых основаны эти заключения

Патология	Фактические данные
1. Артериальная гипертензия	Величина артериального давления, мм рт.ст.: 170/102, 150/86, 166/92, 172/96
2. Сахарный диабет	Тест на толерантность к глюкозе: Время, ч 0 0,5 1 2 Глюкоза плазмы, мг/100 мл 110 190 170 140
3. Почечная недостаточность	8) Биохимические показатели сыворотки: Креатинин 2,7 мг/100 мл Азот мочевины 40 мг/100 мл Гидрокарбонаты 18 мэкв/л

тельные 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, незави-

симо 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/100 мл, но обычно регистрируется как 193 мг/100 мл. *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 (assigasy).

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 концентрации 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. Информацию 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 такие характеристики, как тупая, пульсирующая, жгучая, жалающая 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 данные, полученные в ходе измерений, позволяют предсказать результат, который предполагалось получить (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 биологическими – 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 показаний (неточен). Однократное измерение, сделанное на

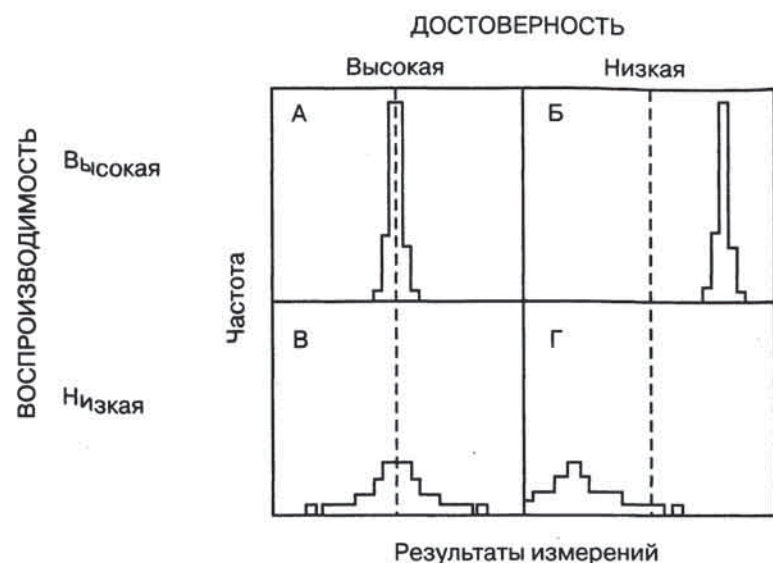


Рис. 2.1. Достоверность и воспроизводимость. А. Высокая достоверность и высокая воспроизводимость. Б. Низкая достоверность и высокая воспроизводимость. В. Высокая достоверность и низкая воспроизводимость. Г. Низкая достоверность и низкая воспроизводимость. Прерывистые линии показывают истинные значения.

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, 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
Метод измерения	
Инструмент	Средства измерения
Наблюдатель	Лица, производящие измерения
Биологические причины	
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 ошиб-

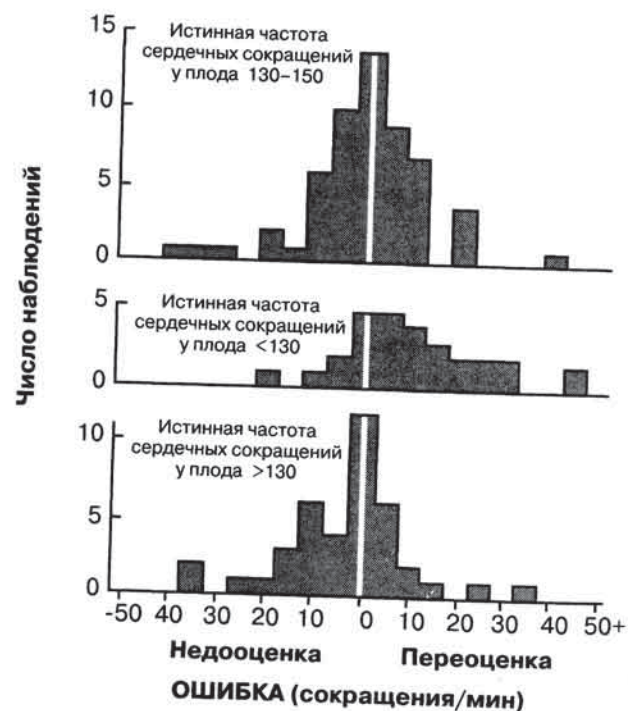


Рис. 2.2. Вариация результатов, связанная с наблюдателем. В зависимости от истинной ЧСС у плода, установленной с помощью электронного монитора, оценка ЧСС при аускультации оказывается в пределах нормы, заниженной или завышенной.

ке. 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, определяя 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 мониторных записей ЭКГ, сделанных в течение 24 ч, отличить снижение частоты желудочковых экстрасистол, обусловленное лечением, от биологических или спонтанных вариаций, снижение частоты должно превышать 83%".

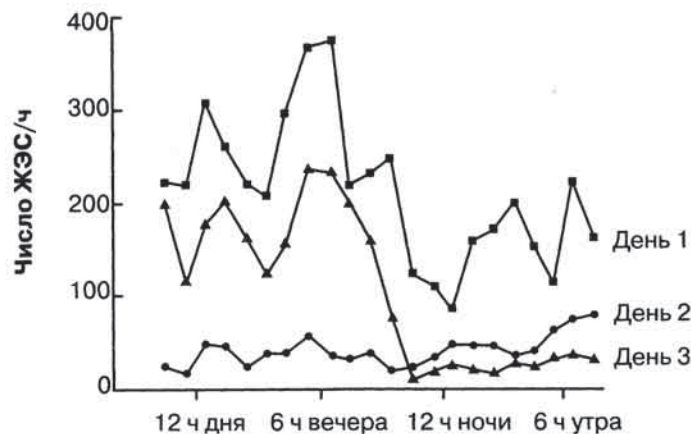


Рис. 2.3. Биологическая вариация. Число желудочковых экстрасистол (ЖЭС) у нелеченого больного при холтеровском мониторировании в течение трех последовательных суток.

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 воз-

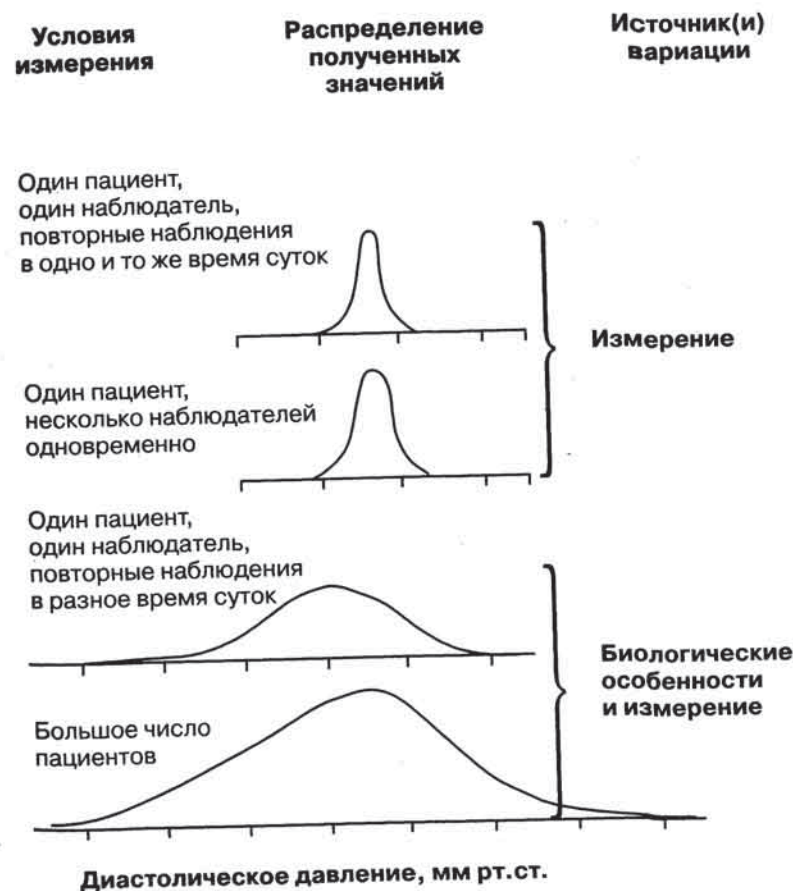


Рис. 2.4. Источники вариации. Измерение диастолического (фаза V) артериального давления [Fletcher R.H., Fletcher S.W., Boe J. et al. The blood pressure in a population (Special issue). Acta med Scand 1957; 321:5 – 313].

44 Клиническая эпидемиология

буждении, во 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

Другой аспект особенности 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 гранул, биопсия может дать отрицательный результат только в силу случайной вариации. Еще один пример случайной вариации был представлен ранее (см. рис. 2.3, стр. 44): 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
Центральные тенденции			
Средняя величина	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 пригодна для описания ненормальных (негауссовых) распределений)
Процентиля, децили, квантили и т.д.	Доля наблюдений, находящихся в определенной части распределения	Характеризует "необычность" значения независимо от типа распределения	Неудобна для статистических расчетов

$$\sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}$$

где X – каждое единичное наблюдение; $\bar{X}$ – среднее из всех наблюдений; N – число наблюдений.

Описание распределений

Для описания распределений случайных величин используются два основных свойства распределения: центральные тенденции (середина 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.

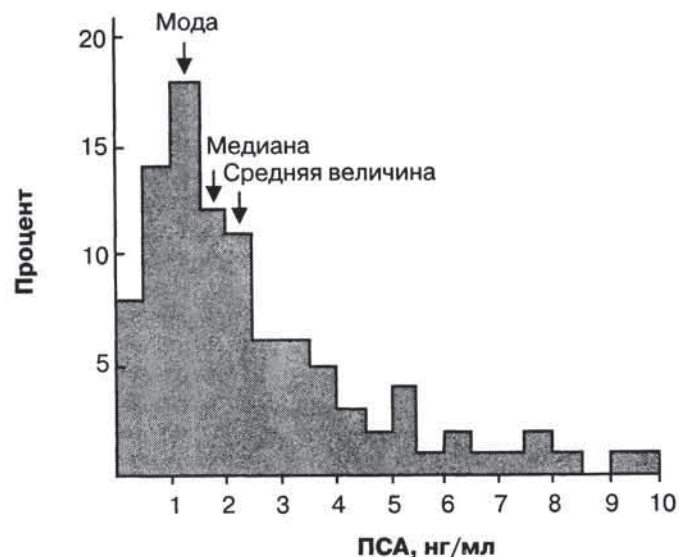


Рис. 2.5. Оценки центральной тенденции и рассеяния. Распределение уровней простат-специфического антигена (ПСА) у предположительно здоровых мужчин [Kane R.A., Littrup P.J., Babaian K. e.a. Prostate-specific antigen levels in 1695 men without evidence of prostate cancer. Cancer 1992;69:1201–1207].

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 сло-

вами, 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.

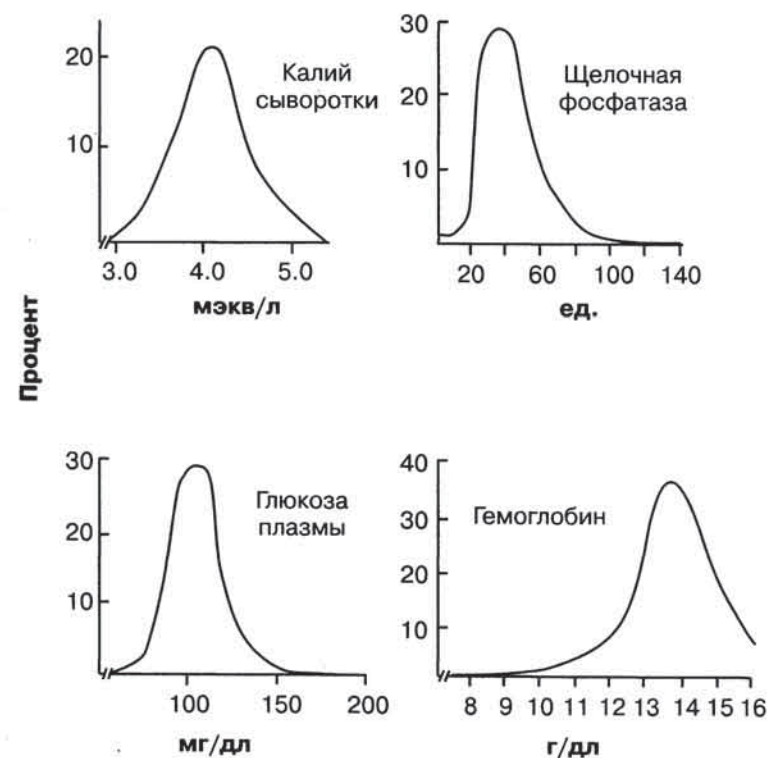


Рис. 2.6. Фактические распределения в клинической медицине [Martin H.F., Gudzinowicz B.J., Fanger H. Normal values in clinical chemistry. New York: Marcel Dekker, 1975].

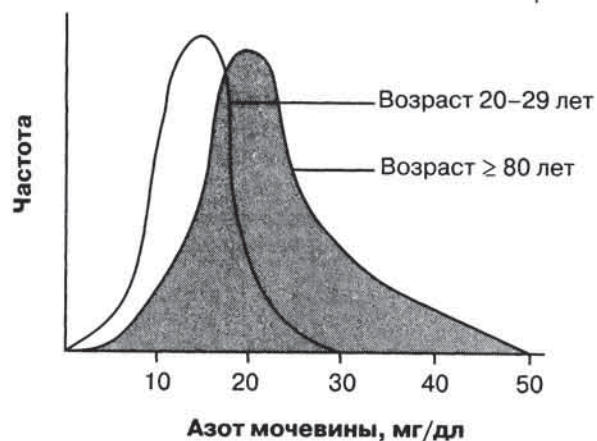


Рис. 2.7. Распределения переменных в клинической медицине в зависимости от возраста: азот мочевины у людей в возрасте 20 – 29 лет и старше 80 лет [Martin H.F., Gudzinowicz B.J., Fanger H. Normal values in clinical chemistry. New York: Marcel Dekker, 1975].

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 площадь под кривой, ограниченная средней величиной ± 2 стандарт-

ных 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

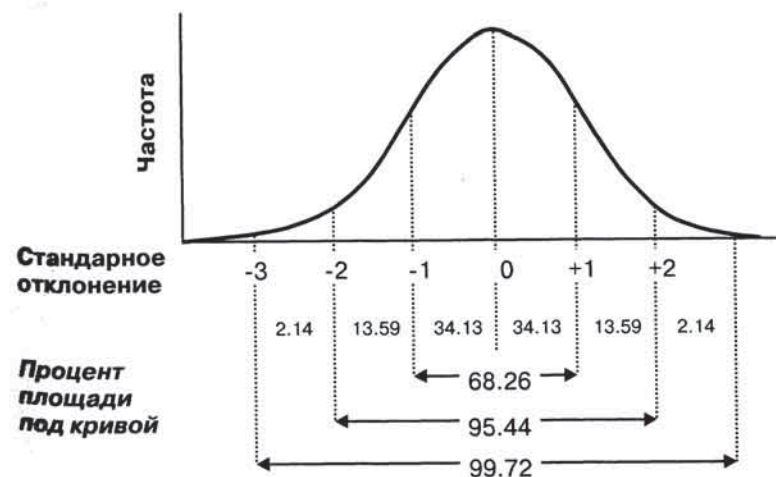


Рис. 2.8. Нормальное (гауссово) распределение.

нам 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 мутанты; 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 высокие уровни фенилаланина, например, из-за нарушений метабо-

лизма 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 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 некоего 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 договориться (в некотором смысле произвольно), что все значения признака, выходящие за два стандартных отклонения от средней вели-

чины, 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 только специфическим 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 процентиля.

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 повы-

шению уровня гемоглобина в среднем на 2,30 г/дл, но не ослабило 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 уровень состояния здоровья у обследуемого выше среднего, 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 нормальщиими.

Патология — это болезнь

Более строгий подход к разграничению нормы и патологии состоит в том, чтобы называть патологическими те признаки, которые обычно обусловлены болезнью, инвалидностью или смертью, 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 предположе-



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].

ние, что 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 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 лече-

нию: 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, меньше отклоняющиеся от нормы. 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 показателя 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 значения, на самом деле характерны более низкие значения. Эти лица были отобраны только в силу случайной вариации, из-за кото-

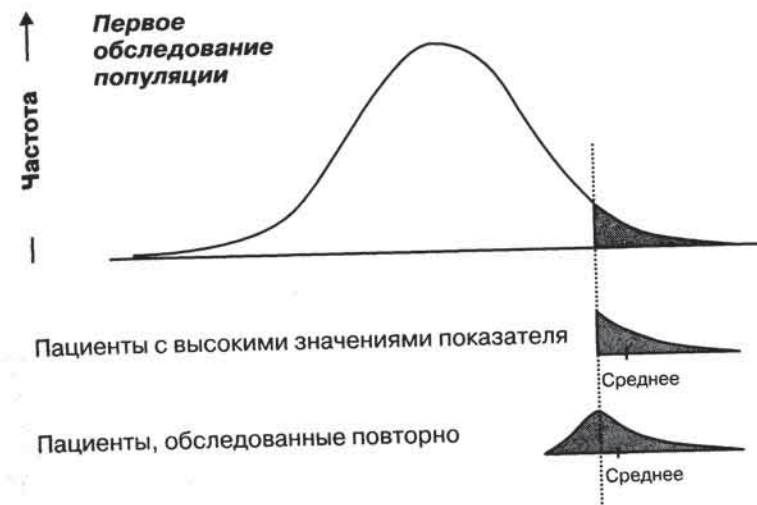


Рис. 2.11. Смещение к среднему.

рой 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 в сторону средней величины для всей подгруппы лиц, имевших при первом измерении значение признака выше точки разделения.

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 основания. Например, показано, что среди всех лиц, у которых при

скрининге выявляется патологический уровень тироксина (T_4) 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

4. 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 (T_4): 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.

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 похудания — в качестве признаков рака легкого у курящих.

Упрощение данных

В главе 2 отмечалось, что клинические показатели, включая данные диагностических тестов, бывают разных типов: качественные, порядковые и количественные. Независимо от типа 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 положительный результат при наличии заболевания или отрица-

тельный в его отсутствие. 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 достаточно трудоемкие, дорогие или сопряженные с риском исследования – 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 условиях. Врачи исходят из того, что в интересах больного

применение более подробного обследования оправданно только в случае, если предварительные тесты дали положительный результат. Врачи 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 простат-специфического 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 заболеваний

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 интересующим нас 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 $\frac{2}{3}$ обследованных, а это лишь немного 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, под-

дающихся объективной оценке, например, стенозом коронарных артерий, выявляемым при ангиографии, при сцинтиграфии 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 дольше применяется, либо, исходя из согласованного мнения экспертов. Так можно прийти к парадоксу: 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 действительно 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	$-PV = \frac{d}{c+d}$
		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.

ца, 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 ли 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

		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	

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

$$LR+ = \frac{\frac{27}{37}}{\frac{10}{35+77}} = 2.3 \quad LR- = \frac{\frac{10}{35+77}}{\frac{77}{77+35}} = 0.39$$

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].

доля 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 позволят исключить некоторые, 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 положительного результата при отсутствии 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, 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 в 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),

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: U.S. Government Printing Office, 1960]



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

то вряд ли кто-то из больных будет 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 в 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 достижении компромисса 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 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 диагностически менее эффективных тестов характеристическая кривая располагается ближе к диагонали, соединяющей левый нижний и правый верхний углы. Диагональ отражает соотношение частот истинно и ложноположительных результатов для абсолютно 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 удается работать с такими тестами, так что приходится применять различные способы нахождения компромисса между чувствительностью и специфич-

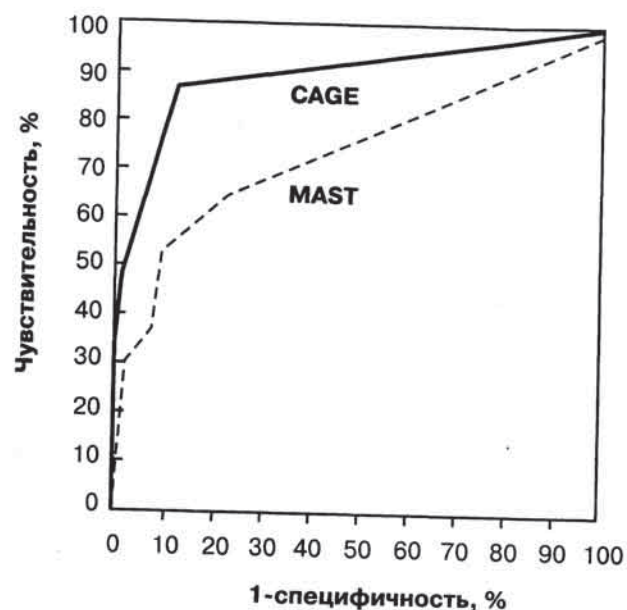


Рис. 3.5. Характеристические кривые по результатам применения опросников CAGE и MAST у пожилых людей, страдающих и не страдающих алкоголизмом [Jones T.V. et al. J Gen Intern Med 1993;8:674-678].

ностью. 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 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 вовсе не недобросовестностью авторов, выполнивших начальные исследования, и не последующим несправедливым скептицизмом медицинской общественности. Скорее,

это обусловлено ограниченными возможностями методов, с помощью которых в первую очередь 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, как 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 на 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.

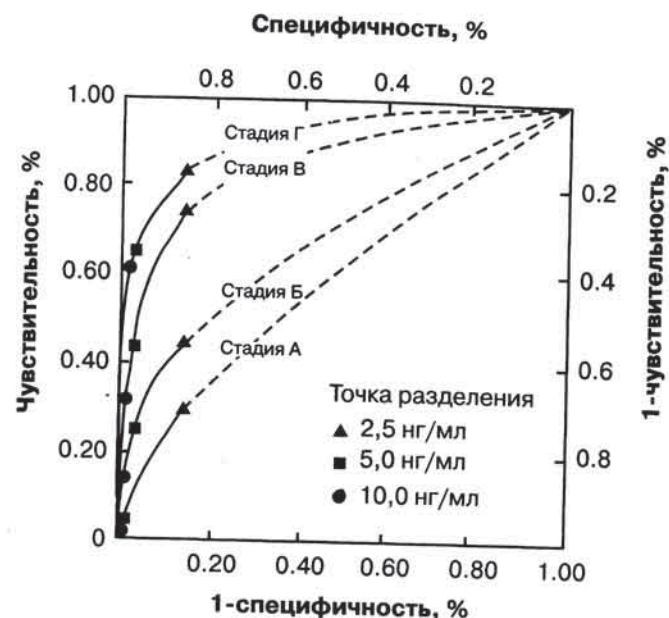


Рис. 3.6. Характеристическая кривая теста КЗА для диагностики рака толстой кишки на разных стадиях заболевания. Чувствительность и специфичность теста меняются в зависимости от стадии заболевания [Fletcher R.H. Ann Intern Med 1986; 104:66–73].

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 тестирование 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 стандартным 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 отношений 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 искажаться, даже если в исследовании отсутствует систематическая ошибка. Наблюдаемые конкретные значения соот-

ветствуют 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.

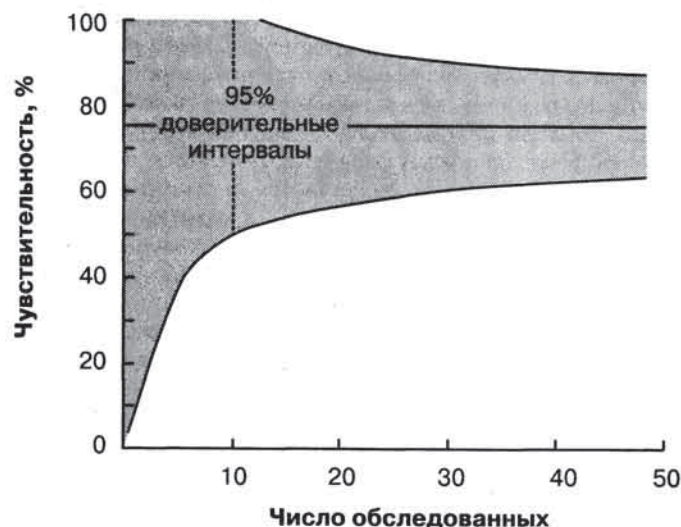


Рис. 3.7. Точность оценки чувствительности. Значения 95% доверительного интервала для чувствительности, равной 75%, в зависимости от размера выборки.

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 (положитель-

ные 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 положителен (отрицателен)"? 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, 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 неми-

нуемо 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 усиливается.

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%, тест считается отличным. Другими словами, на практике чувствительность и специфичность редко варьируют более чем в двухкратных 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.

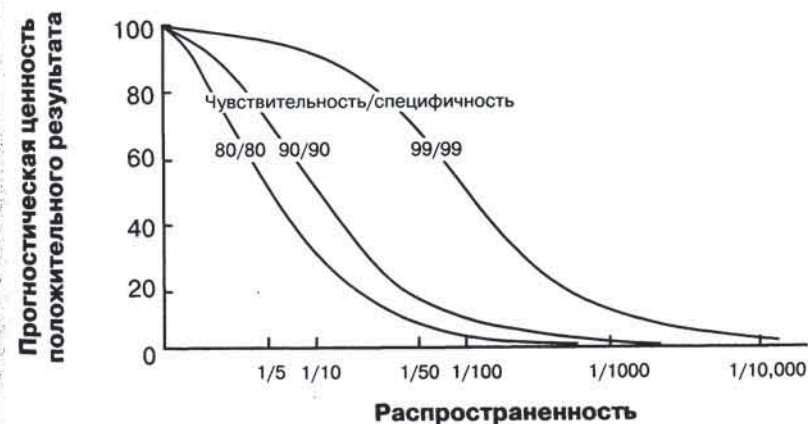


Рис. 3.8. Прогностическая ценность положительного результата теста в зависимости от чувствительности, специфичности и распространенности заболевания.

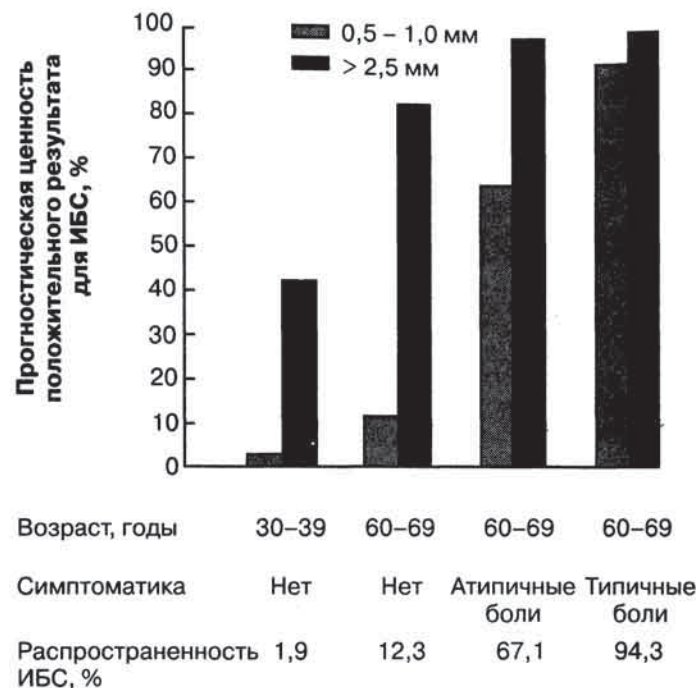


Рис. 3.9. Влияние распространенности заболевания на прогностическую ценность положительного результата теста. Вероятность наличия ишемической болезни сердца (ИБС) у мужчин в зависимости от возраста, симптомов и снижения сегмента ST на ЭКГ при пробе с физической нагрузкой [Diamond G.A., Forrester J.S.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

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, признаков 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 соблазнительно 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 подтвердит 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. 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 вероятности 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 избирательных 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 "Сиэтл Суперсоникс" 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 или 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:

Претестовые шансы $\times$ отношение правдоподобия = 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 (распространенность), 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 использования конкретного результата 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. Сравнивая 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)		↑ Диагноз подтвержден
1,1 – 2,0	3 (11,1)		
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)	↑ Диагноз исключен
9,1 – 10,0	–	20 (21,5)	
10,1 – 11,0	–	11 (11,8)	
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 отношению 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 заболевания 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 результат, поскольку в случае отрицательного результата диагностический поиск прекращается.

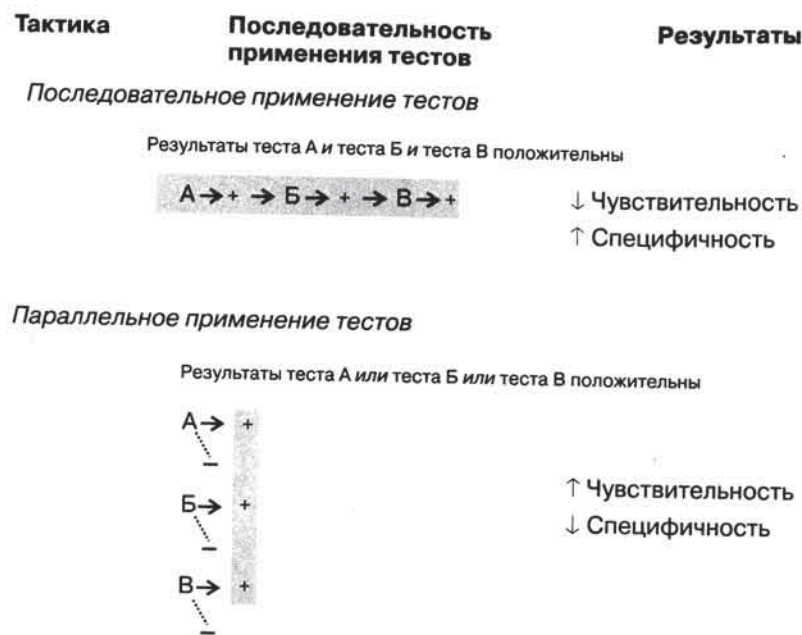


Рис. 3.10. Последовательное и параллельное применение нескольких диагностических тестов.

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 (параллельное и последовательное применение тестов) 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 определению прогностической ценности теста. Однако для большинства

Таблица 3.4
Эффект последовательности применения тестов: "А после Б" и "Б после А" *

Распространенность заболевания									
Число пациентов, прошедших обследование		1000							
Число больных		200 (распространенность 20%)							
Чувствительность и специфичность тестов									
Тест	Чувствительность		Специфичность						
А	80		90						
Б	90		80						
Последовательность применения тестов									
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

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 пациентов 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.

Претестовая вероятность

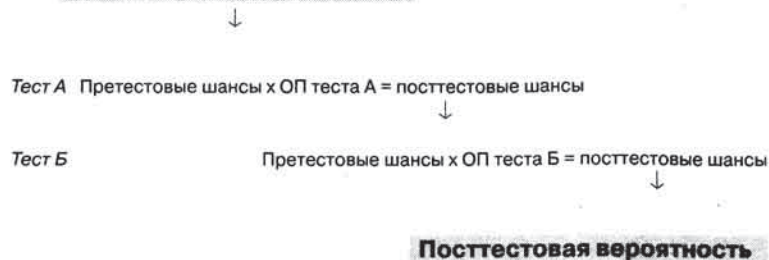


Рис. 3.11. Использование отношений правдоподобия (ОП) при последовательном применении тестов

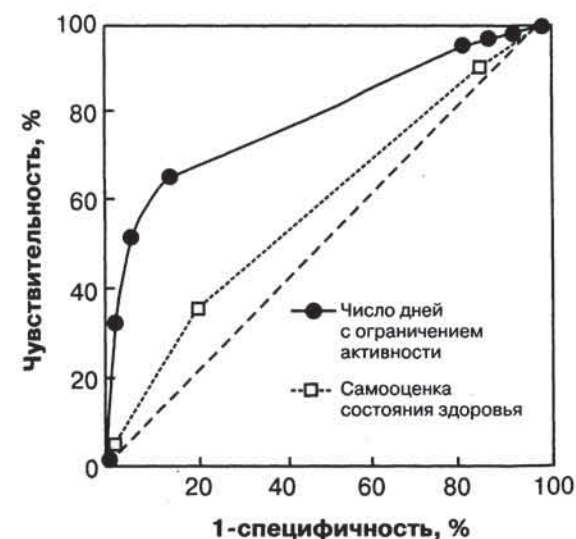


Рис. 3.12. Характеристики двух опросников по оценке состояния здоровья. Способность различать пациентов с серьезными заболеваниями и без них у пожилых пациентов [Wagner E.H. et al. J Am Geriatr Soc 1993;41:241–248].

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 состо-

яния здоровья 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.

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 простат-специфического 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 "клинической оценки" – 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 вероятности или частоты 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 разошлись. 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 ве-

роятность 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

состояние 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.

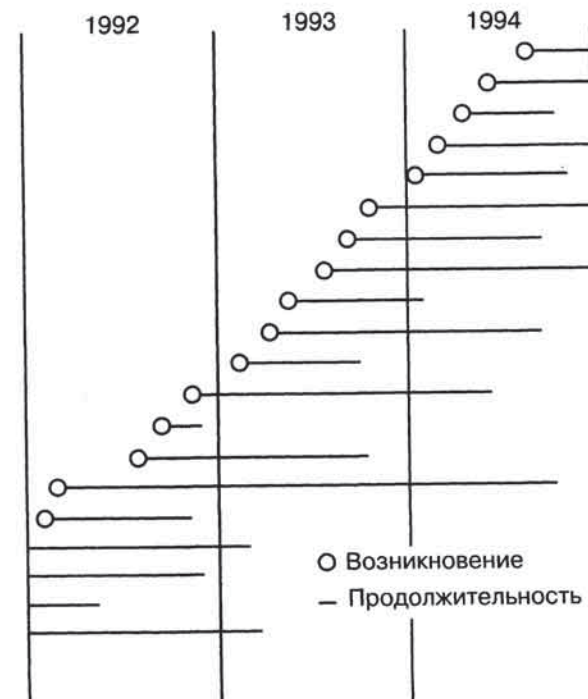


Рис. 4.1. Возникновение заболевания в обследуемой группе риска из 100 человек в период с 1992 по 1994 г.

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 (распространенности)

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 результа-



Рис. 4.2. Методика исследования распространенности.