

Оуэнс К. Д., Сокс Г. К. мл.

Принятие решений в медицине: вероятностное медицинское обоснование

Owens K. D., Sox H. C. Jr.

Medical decision making: probabilistic medical reasoning

Edward Shortliffe/Leslie Perreault, Medical Informatics: Computer Applications in Health Care. Addison-Wesley Publishing Company. Addison-Wesley Publ.Co. 1990, Chpt. 3, P. 70-116

2725 Sand Hill Road, Menlo Park, CA 94025

Ключевые слова Принятие решений о лечении

Анализ полезности

Темы Системы информационного обеспечения
здравоохранения

Управление медицинской помощью

Прочитав эту главу [Данная работа является третьей главой книги “Медицинская информатика. Компьютерные приложения для медицины” под редакцией Э. Шортлифа и Л. Перре (выходные данные приведены в предыдущей сноске). Ссылки на другие главы относятся к указанной книге. Во всех ссылках внутри главы удален префикс третьей главы. – прим. переводчика.], вы должны будете уметь отвечать на следующие вопросы:

Каким образом для анализа результатов тестов и принятия медицинских решений, включающих неопределенность, используется понятие вероятности?

Как можно охарактеризовать способность теста разграничивать болезнь и здоровье?

Какая информация требуется для правильной интерпретации результатов теста?

Что такое принятие решений при помощи математического ожидания? Каким образом эта методика помогает понимать конкретные медицинские проблемы?

Что такое анализ чувствительности? Каким образом он используется для проверки устойчивости решения и определения важных для решения переменных?

3.1. Природа клинических решений: неопределенность и процесс диагностирования

Поскольку клинические данные несовершенны, а результат лечения заранее не определен, медицинские работники часто встают перед трудным выбором. В этой главе мы введем понятие вероятностного медицинского обоснования, одного из подходов, помогающего поставщикам медицинских услуг справляться с неопределенностью, присущей в медицине многим решениям. Для принятия медицинских решений можно поменять широкий набор методов, и не следует думать, что наш подход всегда необходим или даже приемлем. Однако на протяжении этой главы мы будем приводить простые клинические примеры для иллюстрации широкого класса проблем, в которых вероятностное медицинское обоснование дает понимание.

Как мы видели в предыдущей главе, медицинская практика есть принятие решений. В этой главе обсуждается сам процесс принятия этих решений. Таким образом, главы 2 и 3 в совокупности дают подготовительные знания для оставшихся глав, в которых мы будем обсуждать использование компьютеров для принятия медицинских решений. Особое внимание будет уделено системе взаимоотношений между требуемой информацией и разработкой и внедрением системы.

Материал в этой главе представлен в контексте решений, принимаемых отдельным практикующим врачом. Вводимые понятия, однако, могут применяться более широко. Чувствительность и специфичность – важные параметры лабораторных систем, сигнализирующих о ненормальных результатах исследований (глава 9), а также систем курации пациентов (глава 12) и систем библиографических правок (глава 14). Понимание вероятности и того, как ее корректировать после поступления новой информации, будет основой изучения систем принятия клинических решений (глава 15). Значимость вероятности для принятия медицинских решений – идея не новая: “...Хорошая медицина состоит не в неразборчивом применении к пациенту лабораторных обследований, а в четком понимании вероятностей и возможностей того, что именно данное обследование может дать нужную для этого случая информацию”. (Reabody, 1992, стр. 325)

Пример 1. Вы – директор банка крови в большом городе. Все потенциальные доноры проходят проверку на антитела к вирусу иммунодефицита (ВИЧ). ВИЧ вызывает синдром приобретенного иммунодефицита (СПИД). Будучи инфицирован, человек выделяет антитела, которые можно обнаружить в крови. Поскольку СПИД передается через кровь, среди доноров ни в коем случае не должно оказаться ВИЧ-инфицированных. Тест на антитела дает положительный результат в 98% случаев, когда антитела присутствуют, и отрицательный – в 99% случаев, когда их в крови нет [Чувствительность и специфичность этого теста близки к значениям для ранних тестов на антитела к ВИЧ (Weiss et al., 1985). Сейчас доступны тесты с гораздо лучшими параметрами.]

Если результат теста положителен, значит ли это, что в крови донора есть антитела к ВИЧ? Если результат отрицателен, можно ли быть уверенным, что антител в крови действительно нет? На интуитивном уровне эти вопросы не кажутся сложными. Кажется, что если результат положителен, то скорее всего образец сданной крови содержит антитела. Мы будем потрясены, узнав, что если реально инфицирован только один донор из тысячи, тест чаще дает ошибочный результат, чем правильный. Фактически, из ста доноров с положительными результатами теста, инфицированы меньше десяти, то есть на каждый правильный результат приходится десять неправильных! Как понимать такое соотношение? прежде чем ответить на этот вопрос, рассмотрим еще один пример.

Пример 2. Мистеру Джонсу 59 лет. У него болезнь коронарной артерии (сужение или блокировка кровеносных сосудов, снабжающих сердечную мышцу). Когда кровь не достигает сердечной мышцы, она не получает достаточного количества кислорода (гипоксия), и пациент часто испытывает боль в грудной клетке (стенокардия). Мистеру Джонсу дважды делали операцию по шунтированию коронарной артерии трансплантантом, при которой на старый сосуд пересаживается новый, взятый обычно из ноги, таким образом, что кровь обходит заблокированную область. К сожалению, у него снова появляются боли в грудной клетке. Несмотря на медикаментозное лечение, они становятся все сильнее. Если в сердечную мышцу не будет поступать кислород, результатом может стать инфаркт миокарда, от которого часть мышцы погибнет.

Идти ли мистеру Джонсу на третью операцию? Лечение не помогает, и без хирургического вмешательства у него есть большой риск получить инфаркт, возможно, смертельный. С другой стороны, операция тоже опасна, и дело не только в том, что процент смертности при третьей операции выше, чем при двух первых. Шанс, что хирургическое вмешательство облегчит боли, тоже ниже, чем при первой операции. в этом примере любой вариант содержит значительную неопределенность. Более того, риск весьма серьезен – неверное решение основательно повысит вероятность того, что мистер Джонс умрет. Решение будет нелегким даже для опытного врача.

Эти примеры иллюстрируют ситуации, когда интуиция вводит в заблуждение или просто неверна. Врач, который недостаточно критически результаты теста в примере 1, ошибочно скажет людям, что они заражены СПИДом – ошибка с глубоким эмоциональным и социальным последствием. В примере 2 умение врача принимать решения повлияет на состояние здоровья и продолжительность жизни пациента. Подобные ситуации в медицине встречаются постоянно. В этой главе мы задались целью показать, как решению таких проблем могут помочь теория вероятностей и анализ решений.

Принятие решений – дело, которым профессиональному врачу приходится

заниматься постоянно. Некоторые решения получают на основе дедуктивного обоснования или физиологических принципов. Однако многие решения принимаются на основе знаний, полученных из коллективного опыта: чтобы оценить проблему, врач часто должен полагаться на эмпирическое знание связей между симптомами и диагнозами. Решение, основанное на таких нечетких ассоциациях, будет в какой-то мере неопределенным. В следующем разделе мы рассмотрим решения, принятые при условии неопределенности, и представим обзор диагностического процесса. По словам Lloyd H. Smith, Jr., “Медицинские решения, принимаемые с помощью вероятностных методов, необходимы, но и опасны. Даже самый проникательный врач может случайно ошибиться” (Smith, 1985, стр. 3).

3.1.1. Принятие решений при условии неопределенности

Пример 3. Мистер Сантос, 33-х лет, уже один раз перенесший тромб в вене левой ноги, жалуется на то, что нога снова, уже в течение пяти дней, болит и опухла. Врачебный осмотр показал, что нога болезненна и раздута до середины икры – признак [Тут употребляется слово *sign*, означающее симптом – результат медицинского осмотра, например, температура 101,2 градуса по Фаренгейту. *Symptom* же – субъективное ощущение пациента, например жар или озноб. Если ощущения пациента могут быть описаны врачом, разница между этими двумя терминами размывается.], указывающий на вероятность тромбоза глубоких вен. Была проведена венография – исследование, при котором в вену вводится рентгеноконтрастный краситель, который виден на рентгеновском снимке. Вены в ненормальном состоянии, но рентгенолог не может сказать, есть ли новый тромб.

Лечиться ли мистеру Сантосу от тромбоза? Главная диагностическая проблема – опасность повторения тромба в ноге. Тромб может сместиться, и тогда кровь перенесет его к легким, где он может вызвать угрожающую жизни закупорку сосудов – так называемую легочную эмболию. Но у пациентов с распухшей ногой причиной этого лишь в половине случаев является тромб; есть множество других причин, по которым может распухнуть нога. Имея дело с опухшей ногой, врач не может быть уверен, что вызвано это тромбом. Таким образом, полученные сведения содержат значительную неопределенность. Более того, в этом примере даже самый определяющий тест даст сомнительный результат, так как более ранний тромбоз в любом случае вызвал ненормальное состояние вен. Лечение тромбоза состоит в приеме антикоагулянтов (лекарств, препятствующих образованию тромба), которые создают для пациента риск чрезмерного кровотечения. Поэтому врач не хочет назначать их, пока не уверен в существовании тромба.

Этот пример проиллюстрировал важную проблему: клинические данные несовершенны. Степень их несовершенства может быть различна, но клинические данные, включая результаты диагностических исследований, рассказ пациента, результаты врачебного осмотра, неопределенными.

3.1.2. Вероятность: альтернативный способ выражения неопределенности

Язык, используемый практикующими врачами для описания состояния пациента, не определен, что еще больше усложняет проблему неопределенности для принятия медицинских решений. Такие слова, как “вероятно” и “весьма похоже”, используемые врачами для описания степени их уверенности в данном диагнозе, имеют для разных людей абсолютно разные значения (см. рис. 1). Такая широко распространенная несогласованность значений описательных выражений создает обширные возможности для взаимного непонимания.

Проблема выражения уровней вероятности существует не только в медицине. Как она решается в других областях? Рассмотрим, к примеру, скачки. Когда опытный игрок решает, на какую лошадь поставить, его вряд ли удовлетворит информация, что у данной лошади “есть шансы” победить. Ему требуется знание шансов.

Шансы – просто альтернативный способ выразить вероятность. Использование вероятности или шансов для описания неопределенности позволяет избежать разноречивых толкований, присущих обиходным описательным выражениям.

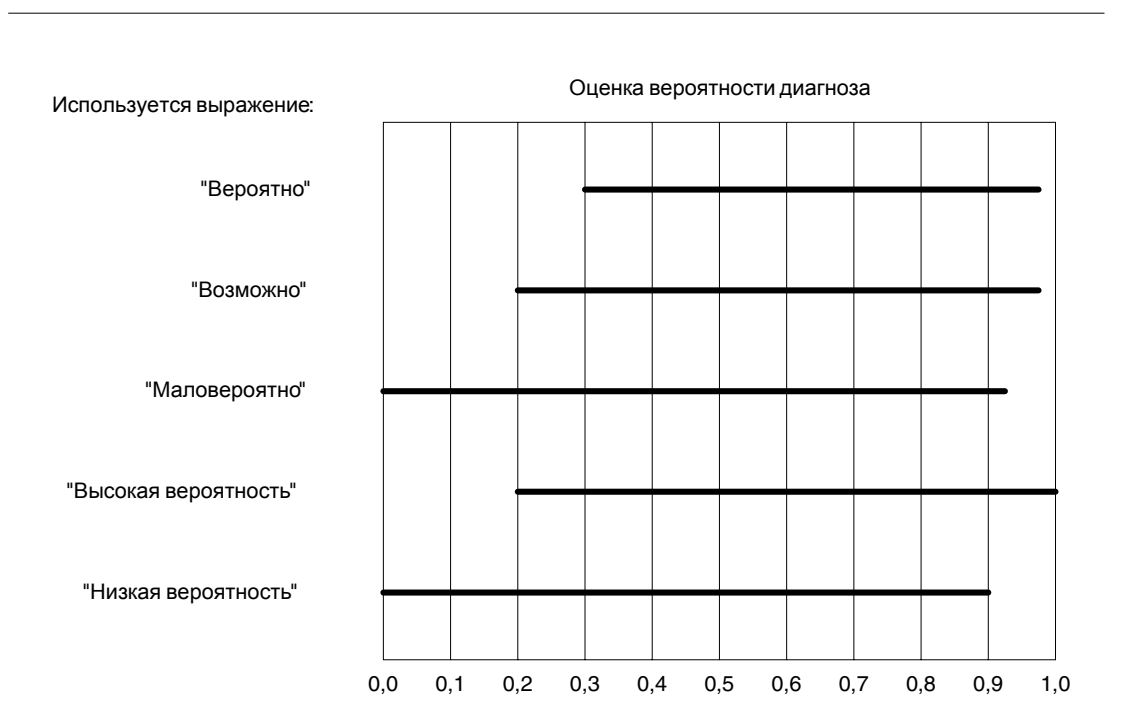


Рис. 1. Вероятность и выражения для нее. Разные врачи вкладывают в одни и те же слова разные значения. График показывает широкий разброс значений, даваемых терминам медицинскими работниками, которых попросили интерпретировать их.

ИСТОЧНИК: перепечатано с разрешения авторов из статьи Bryant G. D., Norman G. R. Expressions of probability: Words and Numbers. The New England Journal of Medicine, 302; 411, 1980).

3.1.3. Рассмотрение диагностического процесса

В главе 2 мы описали гипотетико-дедуктивный подход, диагностическую стратегию, основанную на последовательных итерациях образования гипотез, сборе данных и их интерпретации. Мы обсудили, как наблюдения могут вызвать гипотезу, и как новая информация впоследствии могут снизить или повысить уверенность в этой гипотезе. Здесь мы снова кратко рассмотрим этот процесс на примере конкретного случая. Для обсуждения разобьем процесс диагностики на три этапа.

Первый этап включает в себя первоначальное суждение о том, насколько вероятно заболевание пациента. После осмотра и разговора с пациентом врач интуитивно развивает мнение о вероятности болезни. Это суждение может базироваться на предыдущем опыте или знании медицинской литературы. Мнение врача о вероятности болезни обычно не выражается явно и может быть усовершенствовано с помощью явной оценки вероятности. Оценка вероятности, сделанная до получения дальнейшей информации, называется априорной вероятностью или вероятностью до теста.

Пример 4. Мистер Рубенштейн, 60 лет, жалуется своему врачу на давящие боли в грудной клетке, появляющиеся при быстрой ходьбе. После его рассказа и медицинского осмотра врач считает, что с довольно большой вероятностью у мистера Рубенштейна большое сердце, и назначает ЭКГ при физической нагрузке (мистер Рубенштейн делает зарядку). Только при физической нагрузке, когда требуется больше кислорода и, соответственно, сердце должно работать интенсивнее, становятся видны многие состояния сердца. Результаты мистера Рубенштейна показывают ненормальные изменения ЭКГ во время физической нагрузки – симптом болезни сердца.

Каким образом врач может описать этот случай в числах? Прежде всего он должен

спросить пациента о виде, продолжительности и силе боли. Традиционно далее ему следует опираясь на интуитивные представления об этиологии (причине) боли решить, что делать дальше. Наш подход состоит в том, чтобы выразить первоначальную интуицию врача в явной форме, оценив априорную вероятность болезни. В приведенном примере врач мог, опираясь на то, что узнал из разговора с пациентом, оценить вероятность болезни сердца в 50% (0,5). В разделе 2 мы более аккуратно разоведем методы оценки априорной вероятности.

Следующий этап диагностического процесса состоит в получении дополнительной информации, обычно – результатов диагностических исследований. Врач из примера 4 назначил тест, позволяющий уменьшить неопределенность диагноза болезни сердца. Положительные результаты теста поддержали диагноз (это уменьшение неопределенности показано на рисунке 2а). Хотя в примере 4 при диагностике болезни сердца врач выбрал ЭКГ при физической нагрузке, это лишь одно из многих доступных исследований, и врачу хотелось бы знать, какое из них назначить первым. Одни исследования уменьшают неопределенность больше, чем другие (см. рис. 2b). Чем больше исследование уменьшает неопределенность, тем оно полезнее. В разделе 3 мы разработаем способы измерения, насколько тест уменьшает неопределенность, и используем понятия чувствительности и специфичности, введенные в главе 2.

Третьим этапом будет корректировка первоначальной оценки вероятности с учетом новой информации. Врач из примера 4 должен задать себе вопрос: “Какую вероятность болезни дал ненормальный результат теста?” Он хочет узнать **вероятность** после теста, или **апостериорную вероятность** болезни (см. рис. 2а). В разделе 4 мы еще раз рассмотрим теорему Байеса, введенную в главе 2, и обсудим ее использование для вычисления апостериорной вероятности болезни. Как мы уже заметили, для вычисления апостериорной вероятности болезни нужно знать вероятность априорную, а также чувствительность и специфичность теста [Заметим, что априорная и апостериорная вероятности соотносятся с понятиями распространенности предсказываемого значения. Именно последние термины мы использовали в главе 2, так как обсуждали использование исследований для групп пациентов. Для группы априорная вероятность болезни – это просто ее распространенность среди данной группы населения].

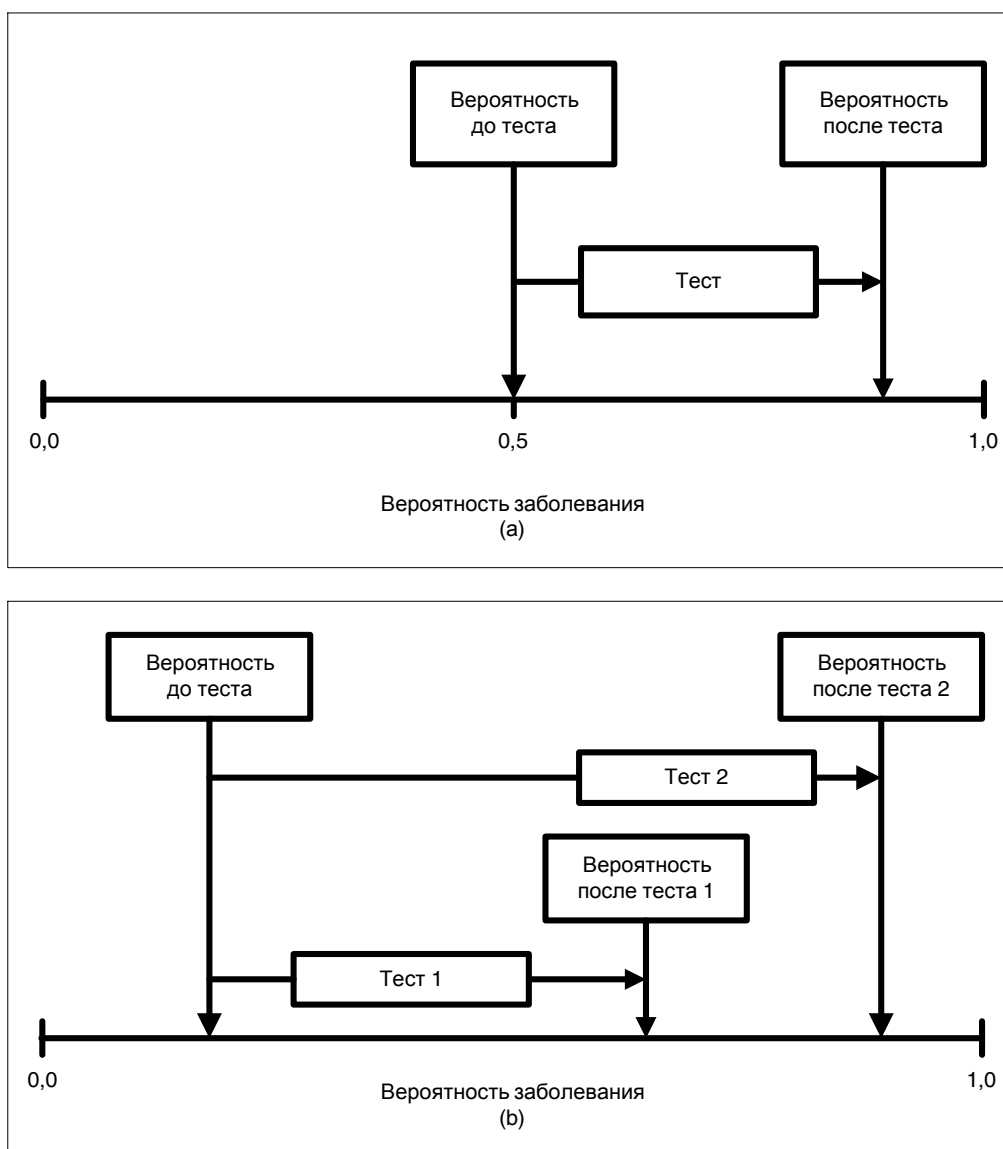


Рис 2. (а) Положительный результат теста повышает вероятность диагноза. (б) Тест 2 уменьшает неопределенность в отношении наличия болезни (увеличивает вероятность болезни) больше, чем тест 1

3.2. Оценка вероятности: методы оценки априорной вероятности

В этом разделе мы разработаем методы, которые врач может использовать для выработки суждения о вероятности болезни до назначения исследований.

Вероятность – предпочитаемое нами средство выражения неопределенности. В нашем контексте вероятность p выражает мнение врача о степени возможности события числом от нуля до единицы. Событие, которое произойдет наверняка, имеет вероятность, равную 1; событие, которое наверняка не произойдет, имеет вероятность 0 [Мы используем байесовскую интерпретацию вероятности; существуют и другие статистические интерпретации этого понятия].

Вероятность события A обозначается $p[A]$. Сумма вероятностей всех взаимоисключающих исходов случайного события должна быть равна единице.

Например, при бросании монеты

$$p[\text{“орел”}] + p[\text{“решка”}] = 1,0$$

Вероятность того, что произойдут оба события A и B обозначается $p[A \text{ and } B]$ или $p[A, B]$.

Говорят, что события A и B **независимы**, если вероятность одного из них не зависит от того, происходит ли второе. Вероятность двух независимых событий A и B может быть выражена через вероятности каждого из них по отдельности:

$$p[A, B] = p[A] \times p[B].$$

Например, если монету бросают последовательно два раза, то вероятность того, что оба раза выпадет “орел”, равна $0,5 \times 0,5 = 0,25$ (вне зависимости от исхода первого броска вероятность выпадения “орла” все равно равна 0,5).

Вероятность события A при условии того, что событие B происходит, называется **условной вероятностью** события A **при условии события B**. Эта вероятность обозначается через $p[A|B]$, что читается “вероятность A при условии B”. Таким образом, апостериорная вероятность – это условная вероятность, условием для которой являются результаты теста или полученные сведения. Например, если тромб обнаруживается у 30% пациентов, у которых распухла нога, то говорят, что вероятность тромба при условии распухшей ноги равна 0,3:

$$p[\text{тромб} | \text{распухшая нога}] = 0,3$$

Перед тем, как упомянута распухшая нога, априорная вероятность тромба – это просто распространенность тромбов в ноге среди группы населения, к которой принадлежит пациент, число много меньше, чем 0,3.

Теперь, решив использовать вероятность для выражения неопределенности, рассмотрим, как можно ее оценить. Это можно сделать либо субъективными, либо объективными методами; каждый из этих подходов имеет свои достоинства и ограничения.

3.2.1. Субъективная оценка вероятности

Большую часть оценок практикующие врачи делают на основе личного опыта, то есть сравнивая имеющуюся проблему со встречавшимися ему ранее или задавая себе вопрос: “Как часто встречается диагноз у виденных мной пациентов?”

Чтобы произвести такую субъективную оценку вероятности, люди полагаются на несколько дискретных, часто бессознательных ментальных процессов. Физиологи познания изучали и описывали такие процессы [Tversky and Kahneman, 1974], называемые **эвристическими**. Точнее, эвристический познавательный процесс – это ментальный процесс, с помощью которого мы узнаем, вспоминаем или обрабатываем информацию; при этом мы можем полагаться на практику (а не на теорию) и не особенно гнаться за строгостью рассуждений. Знание эвристических процессов полезно, так как помогает понять фундамент интуитивной оценки вероятности. И наивные, и умудренные опытом люди, включая врачей и ученых, занимающихся статистикой, принимая решения, злоупотребляют эвристикой и, таким образом, допускают систематические, часто серьезные ошибки в оценке вероятности. Так же, как в ясный день легко можно недооценить расстояние, так в обманчивых клинических ситуациях можно ошибиться в оценке вероятности. [Tversky and Kahneman, 1974]

Три эвристических процесса можно обозначить как важные для оценки вероятности: эвристику репрезентативности, эвристику наличия и эвристику начального значения и корректировки.

Эвристика репрезентативности. Физиологи показали, что один из путей, с помощью которых люди оценивают вероятность – это вопрос “Какова вероятность того, что событие A принадлежит классу B?” Например, “какова вероятность того, что пациент с распухшей ногой принадлежит к классу пациентов с тромбозом?” Отвечая на подобные вопросы, мы часто полагаемся на эвристику

репрезентативности, в которой о вероятности судят по уровню того, насколько А характерно для В, или похоже на В. Врач будет судить о вероятности тромбоза по тому, насколько пациент с распухшей ногой ассоциируется для него с пациентом с тромбозом. Если у пациента есть все классические симптомы, ассоциирующиеся с тромбозом, врач решает, что, скорее всего, у пациента и на самом деле тромбоз. Трудности с использованием такой эвристики появляются, если болезнь очень редкая, то есть очень низка априорная вероятность, она же распространенность, если опыт врача нетипичен и дает неверное представление, или если нетипичен клинический профиль пациента.

Эвристика наличия. На оценку вероятности события влияет, насколько легко вспоминаются подобные события. события, вспоминающиеся легче других, кажутся более вероятными; это эвристика **наличия**, часто ведущая к ошибке. Запоминаются случаи драматические, нетипичные, вызвавшие много эмоций, а поэтому легко переоценить их вероятность. Врач, который вел пациента, умершего от тромбоза, мгновенно представит себе именно тромбоз причиной распухшей ноги. Другие случаи, когда у пациента распухла нога, вспоминаются не сразу, и врач, возможно, переоценит вероятность тромбоза у пациента с распухшей ногой.

Эвристика начального значения и корректировки. Это еще одна широко распространенная эвристика. Врач делает начальную оценку вероятности, а затем корректирует ее в соответствии с полученной информацией. Например, врач из примера 4 сначала оценивает вероятность болезни сердца в 0,5. Если же он затем узнает, что все братья пациента умерли от болезней сердца, ему следует повысить это значение, поскольку из литературы известно, что семейная предрасположенность увеличивает вероятность диагноза. Обычной ошибкой здесь будет недостаточный сдвиг первоначальной оценки в свете новой информации. вместо того, чтобы повысить оценку априорной вероятности, скажем, до 0,8, врач может откорректировать ее всего до 0,6.

Эвристики часто вводят ошибки в рассуждения об априорной вероятности. Ошибки в оценке вероятности априорной неизбежно отражаются на оценке вероятности апостериорной. Поэтому значения эвристики важны для принятия медицинских решений. Избавиться от некоторых из этих трудностей врач может, если будет использовать для оценки вероятности опубликованные результаты исследований.

3.2.2. Объективная оценка вероятности

Опубликованные результаты исследований могут послужить проводником для более объективных оценок вероятности. Для оценки вероятности болезни могут быть использованы распространенность болезни среди населения, или среди данной группы населения, а также *правила клинических прогнозов*.

Как говорилось в главе 2, **распространенность** – это частота события среди населения; это полезная отправная точка для оценки вероятности. Например, если требуется оценить вероятность рака простаты у мужчины 50 лет, полезной отправной точкой будет распространенность этой болезни у мужчин такого возраста (от 5 до 14%). Из этих значений можно исходить, добавляя или уменьшая вероятность по мере поступления новых результатов. Оценки вероятности диагноза для определенной группы населения можно часто встретить в медицинской литературе.

Симптомы, такие как пальпируемые узелки простаты, могут стать основанием для помещения пациента в **клиническую подгруппу**, для которой известна вероятность диагноза. Среди пациентов, направленных к урологу с узелками простаты, распространенность рака составляет примерно 50%. Неудобство такого подхода состоит в сложности помещения пациента в нужную клинически определенную подгруппу, особенно если плохо определены критерии классификации пациентов. Новой тенденцией стало развитие “путеводителей”, известных как правила клинических прогнозов, с целью помочь врачам относить пациентов к четко определенным подгруппам, для которых известны вероятности диагнозов.

Правила клинических прогнозов были разработаны на основе систематического

изучения пациентов с данной диагностической проблемой; они определяют, каким образом для оценки вероятности может быть использована комбинация имеющихся факторов. Определены симптомы и признаки, дающие независимый вклад в вероятность данного диагноза. Они выражены в численных значениях “веса”, основанных на статистическом анализе вклада признаков в вероятность. Результатом является список симптомов и признаков для конкретного пациента, каждый – с соответствующим численным вкладом в общее значение. Общее значение помещает пациента в подгруппу с известной вероятностью диагноза.

Пример 5. У 65-летней мисс Мартин, недавно пережившей сердечный приступ, наблюдается ненормальный ритм сердца (аритмия). Она находится в плохом состоянии и собирается лечь на операцию.

Какова вероятность того, что у мисс Мартин возникнет осложнение на сердце? Для оценки такого риска было разработано правило клинических прогнозов [Goldman et al., 1977]. Таблица 1 показывает список клинических данных и соответствующих диагностических весов. Веса всех признаков, имеющихся у пациента, суммируются, и по суммарному значению пациент попадает в группу с определенной вероятностью осложнений на сердце, как показано в таблице 2. Мисс Мартин получает 20 баллов, что, как может оценить врач, составляет 11% вероятности возникновения серьезных осложнений на сердце.

Таблица 1. Диагностические веса для оценки риска осложнений на сердце при операции на сердце

Клинический признак	Диагностический вес
Свидетельства о застойных сердечных явлениях	+11
Документированный сердечный приступ в последнее время	+10
Аритмия на последней ЭКГ – больше пяти преждевременных вентрикулярных сокращений	+7
Возраст более 70 лет	+7
Срочная операция	+4
Желудочная или торакальная операция	+3
Свидетельства о вальвулярном стенозе аорты	+3
Плохое медицинское состояние	+3

ИСТОЧНИК: Goldman, L., et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. Перепечатано с разрешения автора из The New England Journal of Medicine, 297:848, 1977.

Таблица 2. Правила клинических прогнозов для диагностических весов таблицы 1

Суммарный вес	Распространенность осложнений на сердце со смертельным исходом, %	Распространенность осложнений на сердце, угрожающих жизни, %
0 - 5	0,2	0,7
6 - 12	2,0	5,0
13 - 25	2,0	11,0
26 и более	56,0	22,0

ИСТОЧНИК: Goldman, L., et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. Перепечатано с разрешения автора из The New England Journal of Medicine, 297:848, 1977.

Объективные оценки априорной вероятности чреватой ошибкой по причине смещенности оценок исследований, на которых оценка основана. Например, опубликованная распространенность может не относиться напрямую к конкретному пациенту. Проиллюстрировать это может такой факт. Большинство врачей считало, что если у пациента обнаружено незначительное присутствие крови в моче (микрогематурия), то ему надо проходить дополнительные анализы, так как, согласно ранним работам, у большей части таких пациентов был обнаружен рак или другие серьезные болезни. Анализы предполагали некоторый риск, дискомфорт и расходы пациента. Однако практика назначения их каждому пациенту с микрогематурией была широко распространена в течение нескольких лет. Более поздние исследования показали, что у пациентов без других симптомов, кроме незначительного присутствия крови в моче, вероятность серьезного диагноза в действительности очень низка [Mohr et al., 1986]. До опубликования второй работы много людей подвергалось ненужным анализам, затратив на это нервы и деньги.

Чем объясняется разница между двумя оценками распространенности болезни? Первая работа, показавшая распространенность болезней у пациентов с микрогематурией, была основана на статистике для пациентов, направленных к специалистам-урологам. Обычные практикующий врач первой помощи направляет к специалисту тех пациентов, у которых уже подозревает болезнь. К специалисту попадает мало пациентов, данные которых подразумевают низкую вероятность болезни. Поэтому распространенность болезни среди пациентов, попавших в сферу деятельности специалиста, часто бывает гораздо выше, чем в первичной медицинской практике. Поэтому работы, основанные на статистике для больных, попавших к врачам-специалистам, почти всегда завышают вероятность диагноза. Это пример необъективности выбора. Необъективность выбора – явление обычное, так как большинство опубликованных работ приводят значения именно для пациентов, направленных к специалистам. По этой причине, прежде чем использовать опубликованные результаты для оценки априорной вероятности в других клинических условиях, возможно, понадобится скорректировать их.

Теперь мы можем использовать наработанную технику для иллюстрации того, как врач из примера 4 может оценить априорную вероятность болезни сердца у своего пациента, мистера Рубенштейна, жалующегося на давящую боль в грудной клетке. Начнем с использования доступных объективных данных. Отправной точкой может служить распространенность болезни сердца у шестидесятилетних мужчин. Однако в данном случае мы можем получить более точную оценку, поместив пациента в клиническую подгруппу, для которой известна распространенность болезни. Распространенность по клинической подгруппе, такой как мужчины с типичными для болезни сердца симптомами, предскажет априорную вероятность точнее, чем распространенность по группе, разнородной по отношению к симптомам, такой как население в целом. Примем, что исследования показали распространенность болезни сердца среди мужчин с типичными для болезни сердца симптомами, равную 0,9. Эта распространенность полезна как начальная оценка, которую можно корректировать по мере поступления информации о конкретном пациенте. Хотя распространенность болезни сердца среди мужчин с типичными симптомами и очень высока, некоторые пациенты с подобной историей ей все же не больны.

Далее врач может использовать субъективные методы дальнейшей корректировки этой оценки, основываясь на специфической для данного пациента информации. Например, он может повысить вероятность с 0,9 до 0,95 или выше, узнав о семейной предрасположенности к болезни сердца. Однако следует быть осторожным, чтобы избежать ошибок, которые могут появиться при эвристической субъективной оценке вероятности. В частности, следует знать об ошибочной тенденции оставаться слишком близко к первоначальной оценке при ее

регулировании по добавочной информации. Комбинируя объективные и субъективные методы оценки априорной вероятности, врач может достигнуть приемлемого значения оценки.

В этом разделе мы рассмотрели субъективные и объективные методы определения априорной вероятности и научились корректировать ее значение после оценки вероятности болезни в специфичной подгруппе населения, к которой принадлежит пациент. Следующий шаг диагностического процесса – сбор дальнейшей информации, обычно в виде формальных диагностических исследований (лабораторные анализы, рентгеноскопия и тому подобное). Чтобы помочь лучше понять этот этап, в следующих двух разделах мы обсудим, как измерять аккуратность теста и использовать вероятность для интерпретации его результатов.

3.3. Измерение оперативных характеристик диагностического процесса

Первая проблема, встающая при интерпретации результата теста – определить, нормален он или нет. В этом разделе мы представим спорные вопросы, которые необходимо решить для этого.

3.3.1. Классификация результатов теста как ненормальных

Большинство биологических измерений в популяции здоровых людей содержат переменные, принимающие разные значения для разных людей. Распределение значений часто аппроксимируется кривой нормального (гауссовского, или колоколообразного) распределения (рис. 3). 95% населения попадает между двумя стандартными значениями отклонения от среднего; около 2,5% окажутся с каждой стороны от этих значений. Таким же образом можно аппроксимировать нормальным распределением значения для больных людей. Два эти распределения обычно накладываются друг на друга (рис. 3).

Каким образом классифицировать результат теста как ненормальный?

Большинство клинических лабораторий сообщают **верхний предел нормы**, который обычно определяется как два стандартных отклонения от среднего. Другими словами, результат теста классифицируется как ненормальный (или положительный), если он превышает среднее более, чем на два стандартных отклонения, и как нормальный (или отрицательный), если он меньше этого предельного разделяющего значения. К примеру, если средняя концентрация холестерина в крови равна 220 мг/декалитр, клиническая лаборатория может избрать в качестве верхнего предела нормы 280 мг/декалитр, так как это количество больше среднего на два стандартных отклонения. Заметим, что разделяющее значение, основанное на произвольном статистическом критерии может не иметь биологического значения.

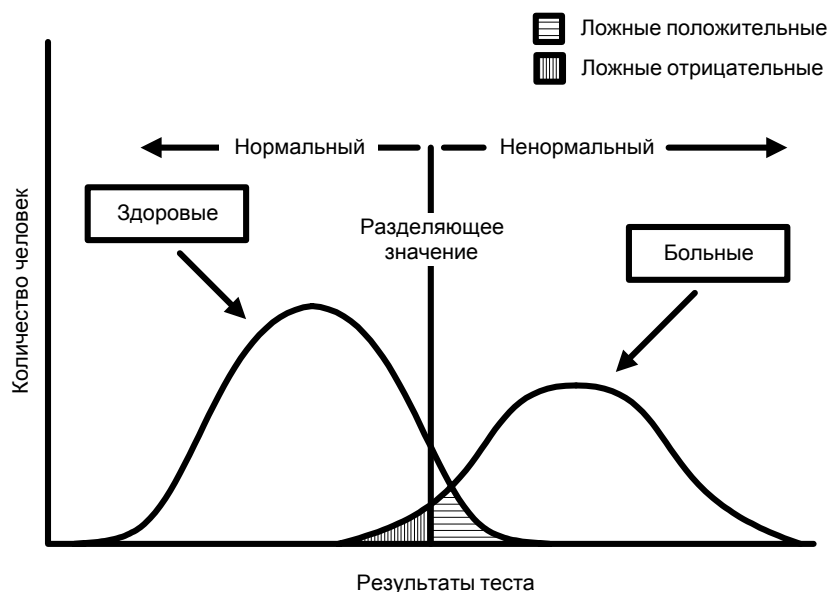


Рис. 3. Распределение результатов теста среди больных и здоровых. Сдвиг разделяющего значения по длительному интервалу будет менять отношение между количеством ложных положительных и ложных отрицательных результатов.

В идеальном тесте нет значений, на которых распределения для здоровых и больных перекрываются. Такой тест, если установлено подходящее разделяющее значение, дает нормальный результат для всех здоровых людей и ненормальный – для всех больных. Такому стандарту соответствует мало тестов. Если результат теста определяется как нормальный по статистическому критерию, у 2,5% здоровых людей результат будет ненормальный. Если распределения результатов теста для здоровых и больных перекрываются, некоторая часть больных пациентов будет иметь нормальные результаты теста (см. рис. 3). Следует хорошо знать термины, используемые для обозначения этих групп.

Истинные положительные (TP) результаты – положительные результаты у людей действительно больных (тест правильно классифицирует их как больных).

Истинные отрицательные (TN) результаты – отрицательные результаты у людей действительно не больных данной болезнью (тест правильно классифицирует их как здоровых).

Ложные положительные (FP) результаты – положительные результаты у людей, данной болезнью не больных (тест неправильно классифицирует их как больных).

Ложные отрицательные (FN) результаты – отрицательные результаты у людей больных (тест неправильно классифицирует их как здоровых).

Рисунок 3 показывает, что сдвиг разделяющего значения (на рисунке – вертикальная линия) между условно нормальными и ненормальными результатами теста изменит относительное соотношение между этими группами. Если разделяющее значение отодвигается дальше от среднего нормального значения, возрастает число ложных отрицательных и уменьшается число ложных положительных результатов.

Однажды установив разделяющее значение, мы можем удобным образом подвести итог характеристики теста – его способности отделять больных от небольных – в **таблице случайности** 2×2 , как показано в таблице 3. Таблица резюмирует количество пациентов в каждой группе – TP, FP, TN и FN. Заметим, что сумма по первой колонке есть общее количество больных пациентов TP+FN. Сумма по второй колонке – общее число пациентов, не больных данной болезнью,

TN+FP. Сумма по первой строке, TP+FP – количество пациентов с положительным результатом теста, по второй – с отрицательным (FN+TN).

Совершенный тест вообще не дает неправильных результатов – FN и FP. Однако ошибочные результаты теста все же возможны, и таблицу случайности можно использовать для определения меры этих ошибок.

Таблица 3. Таблица случайности 2 × 2 для результатов теста

Результаты	Болезнь		Всего
	есть	нет	
Положительные	TP	FP	TP+FP
Отрицательные	FN	TN	FN+TN
	TP+FN	TN+FP	

TP – истинные положительные, TN – истинные отрицательные, FP – ложные положительные, FN – ложные отрицательные результаты теста.

3.3.2. Меры функционирования теста

Меры функционирования теста бывают двух видов: *меры соответствия* и *меры несоответствия*. В таблице 3 появляются два вида *соответствующих* результатов теста: TP и TN. Относительная частота этих результатов формирует основу мер соответствия. Они аналогичны понятиям чувствительности и специфичности теста, введенным в главе 2. Мы определим каждую меру в терминах как таблиц 2 × 2, так и условной вероятности.

Доля истинных положительных результатов (true-positive rate – TPR), или **чувствительность** – это вероятность того, что у больного пациента будет положительный результат, то есть вероятность положительного результата **при условии** болезни:

p [положительный результат | болезнь]

Другим способом можно представить себе TPR как дробь: отношение количества больных пациентов с положительным результатом теста к общему количеству больных пациентов.

$$TPR = \frac{\text{число больных с положительным результатом теста}}{\text{общее число больных}}$$

Из нашей таблицы 2 × 2 (таблица 3) мы можем определить числитель и знаменатель этой дроби. В числителе, по определению, стоит TP, а в знаменателе – сумма по первой колонке, то есть TP+FN. Таким образом,

$$TPR = \frac{TP}{TP + FN}$$

Доля истинных отрицательных результатов (true negative rate – TNR), или **специфичность** – это вероятность того, что у не больного данной болезнью пациента результат теста будет отрицательным:

p [отрицательный результат | болезни нет].

В виде дроби TNR представляется как отношение количества небольших пациентов с отрицательными результатами теста к общему числу больных:

$TNR = \frac{\text{число небольных с отрицательным результатом теста}}{\text{общее число не больных}}$

Из таблицы 3:

$$TNR = \frac{TN}{TN + FP}$$

Аналогично определяются меры несоответствия – доли ложных положительных и ложных отрицательных результатов. Доля ложных отрицательных результатов (false negative rate – FNR) – это вероятность того, что у больного пациента результат теста будет отрицательным, то есть отношение числа больных с отрицательным результатом к общему числу больных, равное

$$FNR = \frac{FN}{FN + TP}$$

Доля ложных положительных результатов (FPR) – это вероятность положительного результата теста у небольного пациента, то есть отношение числа здоровых с положительным результатом к общему числу здоровых, равное

$$FPR = \frac{FP}{FP + TN}$$

Пример 6. Вернемся к проверке донорской крови на антитела к ВИЧ. В современной практике для этой цели используется так называемый твердофазный иммунодефицитный анализ (ТИФА). В таблице 4 представлены гипотетические результаты теста ТИФА, взятого у 400 пациентов [Предполагается, что у нас есть совершенный (отличный от ТИФА) метод определения присутствия или отсутствия антител. В разделе 3.3.4 мы обсудим понятие “золотого стандарта” теста]. Требуется измерить его характеристику.

Вычислим TPR (чувствительность) и TNR (специфичность) теста ТИФА. По определению,

$$TPR = \frac{TP}{TP + FN} = \frac{98}{98 + 2} = 0,98$$

Таким образом, вероятность, что если в крови пациента есть антитела к ВИЧ, то вероятность положительного результата теста для него равна 0,98. Если тест проводился у 100 пациентов, у которых действительно есть антитела, нужно ожидать, что у двух из них результат будет неправильный, отрицательный, то есть доля ложных отрицательных результатов равна 2%. (Уясните для себя, что сумма TPR и FNR по определению равна единице: $TPR + FNR = 1$).

Таблица 4. Таблица случайности для теста ТИФА на антитела к ВИЧ

Результат теста ТИФА	Антитела		Всего
	есть	нет	
положительный	98	3	101
отрицательный	2	297	299
	100	300	400

Таким же образом получаем:

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{297}{297 + 3} = 0,99.$$

Вероятность, что у пациента без антител результат теста будет отрицательным, равна 0,99. Таким образом, из 100 пациентов, не зараженных СПИДом, ТИФА, скорее всего, даст отрицательный результат у 99 человек, и у одного – положительный. (Уясните для себя, что сумма TNR и FPR также по определению равна единице: $\text{TNR} + \text{FPR} = 1$).

3.3.3. Смысл чувствительности и специфичности: как выбрать тест

Возможно, вам уже ясно, что вычисленные значения чувствительности и специфичности теста с непрерывным распределением зависят от выбора конкретного разделяющего значения между “нормальными” и “ненормальными” результатами. Заметьте, что по рисунку 3 рост разделяющего значения (его сдвиг вправо) заметно снизит число ложных положительных и увеличит число ложных отрицательных результатов. Другими словами, тест станет более специфичным и менее чувствительным. Таким же образом более низкое разделяющее значение повысит число ложных положительных результатов и уменьшит – ложных отрицательных, увеличив тем самым чувствительность и уменьшив специфичность. В любом случае в выборе разделяющего значения содержится философский выбор: что лучше допускать – ложные отрицательные результаты (пропущенные случаи ненормальности) или ложные положительные результаты (то есть неадекватно классифицировать здоровых людей как больных)? Выбор разделяющего значения зависит от вида болезни и цели тестирования. Если диагноз очень серьезен, и если доступно спасительное для жизни лечение, следует стараться свести к минимуму число ложных отрицательных результатов. С другой стороны, если болезнь не слишком серьезна, а лечение опасно, разделяющее значение следует установить так, чтобы минимизировать количество ложных положительных результатов.

Мы подчеркиваем, что чувствительность и специфичность – характеристики не теста как такового, а теста и критерия, по которому его результаты объявляются ненормальными. Сдвиг разделяющего значения на рисунке 3 не влияет на сам тест (способ его проведения и результаты пациентов), однако он меняет чувствительность и специфичность. Таким образом, наилучший способ характеристики теста – описать значения чувствительности и специфичности, которые он может давать в зависимости от разделяющего значения. Типичный способ показать эти значения – сравнить при изменении разделяющего значения чувствительность теста с величиной, равной единице минус специфичность (то есть долю истинных положительных результатов с долей ложных положительных результатов); эти две величины являются альтернативными друг другу (рис. 4). Получающаяся известная как характеристическая кривая (ROC), первоначально была описана исследователями, разрабатывавшими методы детектирования электромагнитных сигналов, а затем применена в области психологии [Swets, 1973; Peterson and Birdcall, 1953]. Произвольно выбранная на кривой ROC точка сообщает о чувствительности и специфичности теста для данного порога “нормальности”.

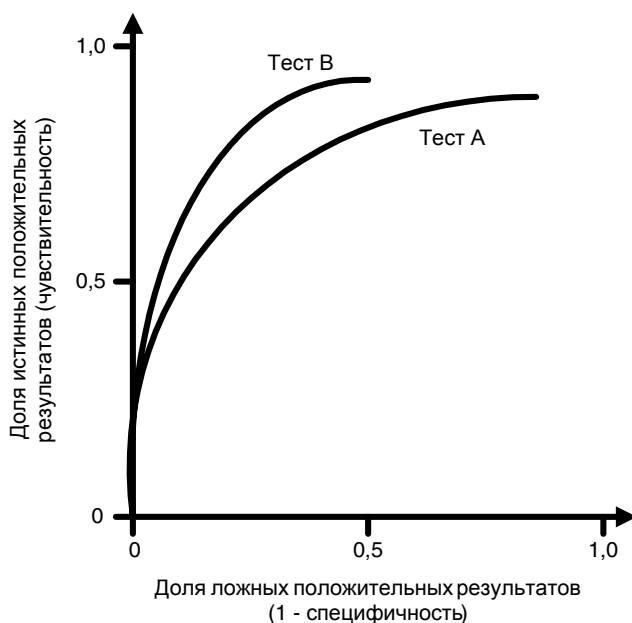


Рис. 4. Характеристическая кривая для двух гипотетических тестов. Тест В лучше теста А (то есть имеет большую мощность распознавания), поскольку его кривая выше (например, доли ложных положительных результатов для теста В ниже, чем для теста А при любом значении доли истинных положительных результатов). Однако, лучший тест не всегда предпочтителен в клинической практике (см. ниже).

Подобные кривые можно нарисовать для **любого** теста, имеющего целью связать полученные клинические данные с конкретными болезнями или группами болезней. Например, рентгенолог, исследующий 100 снимков грудной клетки и определяющий по ним, есть ли у пациентов пневмония, в большинстве случаев некоторую часть снимков классифицирует неправильно [Komaroff, 1979]. Могут быть подсчитаны чувствительность и специфичность его интерпретации. Если он смотрит на те же 100 снимков второй раз, он может классифицировать их уже по-другому; соответственно, другими будут чувствительность и специфичность. Если он много раз по-разному классифицирует снимки, то по полученному множеству значений чувствительности и специфичности можно построить по точкам кривую ROC, отвечающую его прочтению снимков.

Предположим, что в дополнение к имеющимся, введен новый тест на присутствие болезни. Например, пусть стала доступна новая рентгенографическая процедура, определяющая наличие или отсутствие пневмонии. Новый тест может быть оценен с точки зрения чувствительности и специфичности, для него можно построить кривую ROC. Как показано на рис. 4, второй тест лучше первого (то есть имеет большую мощность распознавания), если его кривая ROC выше. Другими словами, тест В лучше теста А, если его специфичность ($1 - FPR$) больше специфичность теста А при любом значении чувствительности (а чувствительность выше чувствительности теста А при любом значении специфичности).

Кривая ROC важна для понимания источников выбора теста и интерпретации данных. Однако врачу не всегда следует выбирать тест по лучшей кривой ROC. Важны и такие данные, как стоимость проведения теста, риск и дискомфорт для пациента, скорость получения результатов. Когда нужно выбирать среди нескольких доступных тестов, следует избирать тест с наибольшими чувствительностью и специфичностью *при прочих равных*, то есть с учетом таких факторов, как стоимость и риск для пациента. Чем выше чувствительность и

специфичность теста, тем больше тест уменьшит неопределенность, касающуюся вероятности диагноза.

3.3.4. Разработка исследования характеристики теста

В пункте 3.2 мы обсудили характеристики теста – способности теста отличать больных от небольных. Классифицируя результаты теста как TP, TN, FP, FN, мы предполагали, что можем узнать с уверенностью, болен пациент или нет. Таким образом, обоснованность результатов любого теста надо измерять по сравнению с некоторым “золотым стандартом” – тестом, отражающим действительное состояние пациента, таким как биопсия пораженной ткани или хирургическая операция. Тест “золотого стандарта” – это процедура, используемая для однозначного определения наличия или отсутствия болезни. Оцениваемый тест называют **индексным**. Тест “золотого стандарта” обычно дороже, опаснее или труднее в исполнении, чем индексный (иначе вообще не было бы смысла оценивать тест, который по всем параметрам хуже уже имеющегося).

Характеристика индексного теста измеряется на небольшой выбранной для исследования группы людей. Однако нас больше интересует ее значение в действительной практике, то есть для более широкой группы пациентов. Поскольку в двух группах людей тест может работать по-разному, мы различаем исследуемую группу, охватывающую пациентов, на которых была измерена опубликованная характеристика теста (обычно это подмножество группы относящейся к клинической практике), и группу, относящуюся к клинической практике, состоящую из типичных пациентов, которые подвергаются тесту.

3.3.5. Необъективность при измерении характеристики теста

Ранее мы упомянули проблему необъективности выбора. Опубликованные оценки распространенности болезни (полученные по контрольной группе) могут отличаться от распространенности в группе, относящейся к клинической практике, поскольку больных пациентов чаще включают в исследования, чем здоровых. Подобным образом и опубликованные значения чувствительности и специфичности взяты по исследуемой группе, которая может отличаться от основной группы по среднему уровню здоровья и по распространенности болезни. Такое различие может повлиять на характеристику теста настолько, что опубликованные значения во многих случаях не будут подходить для клинической практики.

Пример 7. В начале 70-х годов был разработан анализ крови на канцерозембриональный антиген (К-антиген), применявшийся для определения рака толстой кишки. Сообщения ранних исследований свидетельствовали о высоких чувствительности и специфичности этого анализа. Позднейшие работы, однако, доказали, что К-антиген полностью непригоден для диагностики рака толстой кишки. Анализы обследования использовались на неотобранном контингенте, и различия между характеристикой теста для исследуемой и основной групп были частично ответственны за первоначальные ошибки в вычислении значений TPR и TNR для К-антигенного анализа [Ransahoff and Feinstein, 1978].

Ситуация, подобная опыту с К-антигеном, повторилась уже со множеством тестов. Ранние измерения характеристики теста бывают излишне оптимистичны, а последующие результаты разочаровывают. Проблемы возникают, когда TPR и TNR, вычисленные для исследуемой группы, не соответствуют своим настоящим значениям. Такое несоответствие обычно является результатом необъективности в разработке начальных исследований, среди видов необъективности можно выделить необъективность спектра, необъективность назначения и необъективность интерпретации.

Необъективность спектра появляется, когда исследуемая группа состоит только из лиц с четко выраженной болезнью (самых что ни на есть больных) и здоровых добровольцев, как это часто бывает, когда тест разрабатывается впервые. Продвинутой стадии болезни легче диагностировать, чем раннюю стадию. Например, легче определить рак, когда он распространился по всему телу (пошли метастазы), чем когда он локализован в небольшой части толстой кишки. По сравнению с исследуемой группой, в клинической практике будут чаще

встречаться ранние стадии болезни, которые с большей вероятностью не будут определены индексным тестом (возрастет FN). Таким образом, результаты по исследуемой группе дадут слишком низкое значение FNR, влекущее за собой слишком большое значение TPR ($TPR = 1 - FNR$). В добавление к этому, здоровые добровольцы с меньшей вероятностью, чем настоящие пациенты, больны другими болезнями, которые могут вызвать ложный положительный результат [Добровольцы обычно очень здоровы, в то время как средний пациент часто болен не одной определенной болезнью, а сразу несколькими. Эти другие болезни могут вызвать ложный положительный результат. Например, пациенты с доброкачественным (а не злокачественным) увеличением предстательной железы скорее, чем здоровые добровольцы, будут иметь ложный положительный уровень энзима крови, называемого фосфатазой. У людей с раком простаты уровень энзима часто повышен, и это повышение используется для подтверждения диагноза.]; контрольная группа будет иметь неоправданно низкую FPR, и поэтому специфичность $TNR = 1 - FPR$ будет переоценена. Неаккуратность ранних оценок TPR и TNR в тесте на К-антиген была отчасти обусловлена необъективностью спектра.

Необъективность назначения появляется, когда результаты индексного теста влияют на назначение теста “золотого стандарта”. В клинической практике пациенты с отрицательным результатом индексного теста с меньшей вероятностью будут подвержены тесту “золотого стандарта”, чем пациенты с положительными результатами. Другими словами, исследуемая группа, включающая людей с положительными результатами индексного теста, имеет больший процент больных людей, чем группа, относящаяся к клинической практике. Получается, что пациенты с результатами из TN и FN в контрольной группе представлены недостаточно. Результат – завышение TPR и завышение TNR.

Необъективность интерпретации возникает, когда интерпретация индексного теста влияет на трактовку теста “золотого стандарта” или наоборот. Такая необъективность влечет за собой искусственную согласованность результатов двух тестов и неоправданно увеличивает меры согласованности (чувствительность и специфичность) по контрольной группе. (Помните, что основа мер согласованности – относительная частота TP и TN).

Чтобы противостоять необъективности этих трех видов, возможно, окажется нужным корректировать значения TPR и TNR для новой группы населения. Результат любой необъективности – завышение, а не завышение TPR. Поэтому если подозревается необъективность, для применения к новой группе населения TPR (чувствительность) следует уменьшить.

Корректировка TNR (специфичности) зависит от вида необъективности. Необъективности спектра и интерпретации повышают TNR, и при их наличии специфичность следует взять ниже, чем значение на контрольной группе. С другой стороны, необъективность назначения занижает специфичность, и поэтому при подозрении на нее, специфичность следует повысить.

Этот пункт касался второго этапа диагностического процесса – извлечения дополнительной информации из диагностических тестов. Мы научились находить такие характеристики теста, как чувствительность (TPR) и специфичность (TNR). Две эти величины определяют вероятность того, что результат теста отразит истинное состояние пациента. Но для группы, относящейся к клинической практике, эта вероятность – еще не ответ на вопрос “какова вероятность болезни при условии положительного результата теста?” Чтобы ответить на него, мы рассмотрим методы вычисления апостериорной вероятности болезни.

3.4. Апостериорная вероятность: теорема Байеса и предсказывающие значения

Третий этап диагностического процесса (см. рис. 2а) – корректировка оценки вероятности с учетом новой информации, полученной из диагностических тестов,

то есть вычисление апостериорной вероятности.

3.4.1. Предсказывающее значение теста

В главе 2 мы дали формулу для вычисления предсказывающего значения теста (PV+) через чувствительность, специфичность и распространенность болезни:

PV+ =

$$\frac{\text{чувствительность} \times \text{распространенность}}{\text{чувств.} \times \text{распростр.} + (1 - \text{специфичность})(1 - \text{распростр.})}$$

Для лучшего понимания сути этой формулы, напомним, что положительное предсказывающее значение теста – это вероятность того, что пациент с положительным результатом теста действительно болен. Таким образом, PV+ также может быть вычислен непосредственно по таблице случайности 2 × 2 (см. таблицу 3) как отношение числа больных пациентов с положительными результатами к числу всех пациентов с положительными результатами:

$$PV+ = \frac{TP}{TP + FP}$$

Отрицательное предсказывающее значение (PV-) – это вероятность того, что пациент с отрицательным результатом теста действительно здоров, то есть отношение числа здоровых пациентов с отрицательными результатами к числу всех пациентов с отрицательными результатами. В терминах таблицы 3:

$$PV- = \frac{TN}{TN + FN}$$

Пример 8. Вычислим предсказывающие значения для теста ТИФА из примера 6 (см. таблицу 4):

$$PV+ = \frac{98}{98 + 3} = 0,97$$

$$PV- = \frac{297}{297 + 2} = 0,99$$

Вероятность присутствия антител у пациента с положительным результатом индексного теста ТИФА равна в данном исследовании 0,97; примерно у 97 из 100 таких пациентов антитела действительно есть. Вероятность того, что в крови пациента с отрицательным результатом теста ТИФА антител нет, равна примерно 0,99.

Следует еще раз подчеркнуть разницу между предсказывающими значениями и чувствительностью и специфичностью, так как и те, и другие вычисляются по таблице 2 × 2, и их часто путают. Чувствительность и специфичность дают вероятность данного результата теста у пациента в соответствующем состоянии (здоров/болен). Предсказывающие значения же дают вероятность правильного определения состояния пациента по известному результату теста.

Положительное предсказывающее значение, вычисленное из таблицы 4, равно 0,97, то есть мы ожидаем, что 97 из 100 пациентов с положительным результатом индексного теста действительно заражены. Однако, как мы видели, в примере 1 антитела присутствуют в крови менее, чем у одной десятой пациентов с положительными результатами. Чем объясняется такое расхождение значений? Чувствительность и специфичность в этих двух примерах одинаковы. Расхождение же обязано своим присутствием чрезвычайно важной характеристике предсказывающего значения, которую часто забывают учитывать: ведь предсказывающее значение зависит еще и от распространенности болезни среди данной группы населения. Величина предсказывающего значения не может быть

перенесена на новую группу населения, поскольку в двух группах может различаться распространенность болезни.

Различие предсказывающих значений теста ТИФА в примере 1 и в примере 6 проистекает от различия значений распространенности диагноза в группах, взятых для этих примеров. В примере 1 распространенность антител была взята равной 0,001, а в примере 6 – равной 0,25 (распространенность можно найти из таблицы 4 как $TP+TN$, деленное на общее число пациентов).

Теперь, наконец, мы можем понять неожиданный результат первого примера: тест, имеющий высокие значения чувствительности и специфичности, может дать низкое предсказывающее значение, если низка распространенность болезни среди данной группы населения.

3.4.2. Теорема Байеса

Как подчеркивалось ранее в этой главе, понятия предсказывающего значения и апостериорной вероятности идентичны. Распространенность болезни среди группы пациентов играла у нас роль априорной вероятности болезни. Однако стоит врачу начать собирать информацию о пациенте, он начинает пересматривать свою оценку вероятности диагноза. Роль априорной вероятности для следующего проводимого им теста исполняет уже не распространенность диагноза, а пересмотренная оценка. Поэтому нам необходимы более общие формулы для $PV+$ и $PV-$ (апостериорной вероятности), то есть основанные на априорной вероятности, извлеченной с использованием всей полученной информации о пациенте. Нужная нам математическая формула носит название теоремы Байеса.

Теорема Байеса дает количественный способ вычисления апостериорной вероятности с использованием априорной вероятности, а также чувствительности и специфичности теста. Теорема эта следует из определения условной вероятности и свойств вероятности (доказательство теоремы Байеса дано в приложении в конце главы).

Напомним, что условная вероятность – это вероятность события A при условии, что произойдет событие B (см. раздел 2). Мы же, вообще говоря, хотим знать вероятность того, что пациент болен (событие A) при условии, что результат теста положительный (событие B). Если обозначить присутствие болезни как D , отсутствие – как $-D$, нужный нам результат теста как R , то вероятность болезни будет обозначаться $p[D]$, а вероятность болезни при условии положительного результата теста – как $p[DIR]$.

Теорема Байеса гласит, что:

$$p[DIR] = \frac{p[D] \times p[R|D]}{p[D] \times p[R|D] + p[-D] \times p[R|-D]}$$

Мы можем переформулировать это общее уравнение для положительного результата теста (+), заменив $p[DIR]$ на $p[D|+]$, $p[R|D]$ – на $p[+|D]$, $p[R|-D]$ – на $p[+|-D]$, и $p[-D]$ – на $(1 - p[D])$. Вспомнив (из пункта 3), что $p[+|D] = TPR$, а $p[+|-D] = FPR$, получаем:

$$p[D|+] = \frac{p[D] \times TPR}{p[D] \times TPR + (1 - p[D]) \times FPR}$$

Вспомнив, что $p[D|+]$ и $PV+$ – одно и то же, и заменив $p[D]$ на распространенность, получаем, что отсюда следует формула для $PV+$ из пункта 4.1.

Аналогично выводим теорему Байеса для отрицательного результата теста:

$$p[D|-] = \frac{p[D] \times FNR}{p[D] \times FNR + (1 - p[D]) \times TNR}$$

Пример 9. Теперь мы можем сосчитать вероятность болезни сердца у мистера Рубенштейна из примера 4 после получения положительных результатов ЭКГ при физической нагрузке. В конце раздела 2 мы, основываясь на распространенности болезни сердца среди пациентов с типичными симптомами и ее распространенности среди пациентов с семейной предрасположенностью, оценим априорную вероятность болезни в 0,95. Допустим, что TPR и FPR для ЭКГ при физической нагрузке равны 0,65 и 0,20 соответственно. Применяя формулу Байеса для положительных результатов теста, получаем:

$$p[D|+] = \frac{0,95 \times 0,65}{0,95 \times 0,65 + 0,05 \times 0,20} = 0,98$$

Получается, что положительный результат теста повышает вероятность болезни с 0,95 до 0,98 – изменение довольно скромное. Причина этого – высокая априорная вероятность (0,95) и высокое значение FPR (0,20). Повторив вычисления для априорной вероятности, равной 0,75, получаем апостериорную вероятность 0,91; предположив, что FPR теста было равно 0,05 вместо 0,20, а априорная вероятность – снова 0,95, получаем $p[D|+] = 0,996$.

3.4.3. Теорема Байеса для отношений шансов

Хотя формула теоремы Байеса и проста, она неудобна для устного вычисления. Можно вывести более удобную форму теоремы Байеса, выразив вероятность через шансы и используя другие меры характеристики теста.

Вероятность и шансы соотносятся формулами:

$$\text{шансы} = p / (1-p)$$

$$p = \text{шансы} / (1+\text{шансы})$$

Другими словами, если вероятность дождя 0,75, то шансы равны 3:1.

Меры характеристики теста, которые мы обсудили раньше, можно скомбинировать и получить другой показатель – **коэффициент правдоподобия** (Likelihood ratio – LR), равный **отношению вероятности результата для больного к вероятности результата для здорового**.

Коэффициент правдоподобия может быть использован для характеристики полученных клинических данных (таких, как распухшая нога) или результатов теста. Мы опишем два коэффициента правдоподобия, один для положительных результатов теста (или наличия признака), другой – для отрицательных результатов (или отсутствия признака); они обозначаются соответственно LR+ и LR-.

LR+ – это отношение вероятности положительного результата для больного к вероятности положительного результата для здорового:

$$LR+ = \frac{TPR}{FPR}$$

Если тест хорошо различает болезнь, TPR будет высоким, а FPR – низким, и LR+ будет много больше единицы. Коэффициент правдоподобия, равный единице, означает, что вероятность результата для больного и здорового одинакова, и тест не имеет смысла.

Подобным же образом, LR- равен отношению вероятности отрицательного результата для больного к вероятности отрицательного результата для здорового:

$$LR- = \frac{FNR}{TNR}$$

У хорошего теста FNR низко, а TNR – высоко, то есть LR- много меньше единицы.

Между шансами до и после теста существует простое соотношение:

Шансы после теста = (шансы до теста) × LR,

то есть:

$$\frac{p[D|R]}{p[-D|R]} = \frac{p[D]}{p[-D]} \times \frac{p[R|D]}{p[R|-D]}$$

Это теорема Байеса для отношений шансов [Некоторые авторы называют ее формой “шансы-правдоподобие” теоремы Байеса]. Она может быть напрямую получена из теоремы Байеса и определения условной вероятности, введенных нами ранее. Таким образом, для получения шансов после теста надо просто перемножить шансы перед тестом и соответствующий коэффициент правдоподобия.

Пример 10. Мы можем сосчитать апостериорную вероятность при положительном результате ЭКГ при физической нагрузке у 60-летнего мужчины, априорная вероятность у которого была равна 0,75. Шансы до теста были

$$\frac{p}{1-p} = \frac{0,75}{1-0,75} = 3$$

(то есть 3:1).

Коэффициент правдоподобия теста равен

$$LR+ = \frac{TPR}{FPR} = \frac{0,65}{0,20} = 3,25$$

Шансы после теста, по теореме Байеса для отношений шансов, равны:

шансы после теста = $3 \times 3,25 = 9,75 : 1$.

Выразим, наконец, вероятность через шансы: $p = (\text{шансы}) / (1 + \text{шансы}) =$
 $\frac{9,75}{1 + 9,75} = 0,91$

Как и ожидалось, результат согласуется с полученным ранее (см. обсуждение примера 9).

Теорема Байеса для отношений шансов дает возможность проводить вычисления быстро, так что можно вычислить вероятность, например, прямо у постели больного. Доступны вероятностные значения многих тестов [Sox et al., 1988]. Зная шансы до теста, шансы после него можно сосчитать в один прием.

3.4.4. Смысл теоремы Байеса

В этом пункте мы изучим смысл теоремы Байеса для интерпретации теста. Эти идеи необычайно важны, и при этом до сих пор часто бывают поняты неправильно.

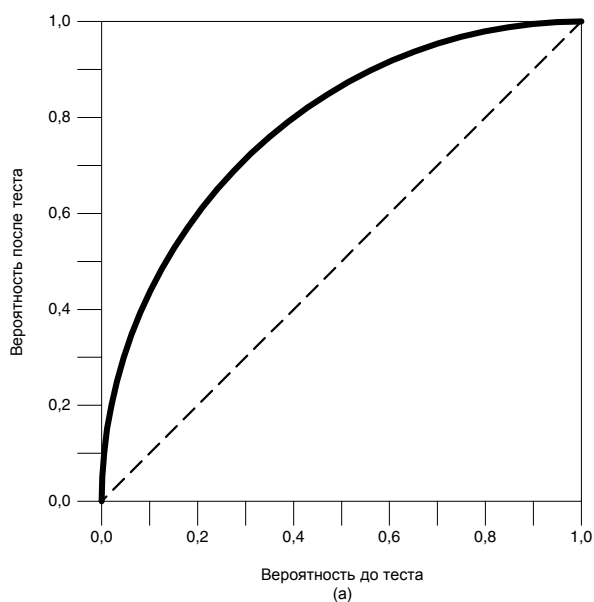
Рисунок 5 иллюстрирует одну из наиболее существенных концепций этой главы: апостериорная вероятность болезни увеличивается по мере роста априорной вероятности. Рисунок 5a основан на вычислениях вероятности после теста при его положительном результате для всех возможных значений вероятности до теста, рисунок 5b построен аналогично для отрицательного результата теста.

Линии, проведенные под углом 45° на обоих рисунках обозначают бесполезный тест, то есть тест, вероятности до и после которого равны (коэффициент правдоподобия равен единице). Кривая с рисунка 5a показывает зависимость между вероятностями до и после теста с чувствительностью и специфичностью, равными 0,9. Заметим, что при низких априорных вероятностях положительный результат теста значительно увеличивает вероятность болезни. Если же априорная вероятность сама по себе велика, то положительный результат теста лишь слегка увеличивает ее.

Рисунок 5b показывает отношение между вероятностями до и после теста,

результаты которого отрицательны. При высокой априорной вероятности апостериорная вероятность при таком результате гораздо меньше исходной, однако отрицательный результат мало влияет на вероятность, если она и сначала была низка.

Это подчеркивает еще одну ключевую идею этой главы: интерпретация результата теста зависит от априорной вероятности болезни до него. Если вероятность до теста низка, то эффект его положительного результата велик, а отрицательного – мал; если априорная вероятность велика, то положительный результат теста мало, а отрицательный – сильно влияет на апостериорную вероятность. Другими словами, если до тестирования врач почти уверен в диагнозе (то есть априорная вероятность близка к нулю или к единице), то подтверждающий результат теста слабо влияет на апостериорную вероятность (см. пример 9). Если вероятность до теста средняя, или результат теста опровергает сложившееся клиническое впечатление, то под влиянием теста вероятность изменится сильно.



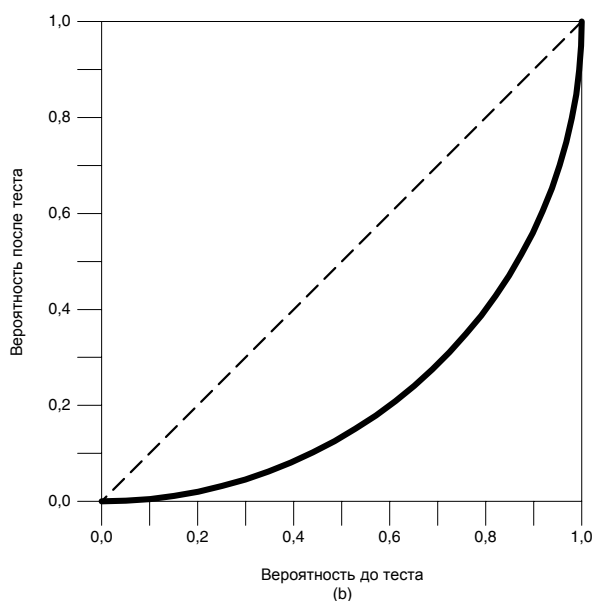


Рис. 5. Связь между вероятностью болезни до и после теста. Пунктирная линия соответствует тесту, не влияющему на вероятность болезни. Чувствительность и специфичность теста были приняты равными 0,9 для обоих примеров. (А) По теореме Байеса была вычислена вероятность после теста для всех значений вероятности до него при условии положительного результата теста. (В) То же для отрицательного результата теста.

ИСТОЧНИК: Sox, Jr., H. C. Probability theory in the use of diagnostic tests: Application to critical study of the literature. In Sox, Jr., H. C. (ed), *Common Diagnostic Tests: Use and Interpretation*. Philadelphia: American College of Physicians, 1987, p. 1-17. Публикуется с разрешения Американского врачебного колледжа.

Заметим из рисунка 5а, что если вероятность до теста очень низка, положительный его результат может увеличить вероятность лишь до среднего уровня. Предположим, что рисунок 5а представляет отношения между вероятностями до и после ЭКГ при физической нагрузке. Если врач верит, что априорная вероятность заболевания коронарной артерии равна 0,1, то вероятность после теста будет около 0,5. Хотя вероятность изменится сильно, вероятность после теста примерно средняя, что оставляет значительную неопределенность диагноза.

Таким образом, если вероятность до теста низка, положительный результат теста вряд ли увеличит вероятность болезни до уровня, достаточного, чтобы врач принял окончательное решение. Таким же образом, если вероятность до теста очень высока, вряд ли отрицательный результат теста изменит апостериорную вероятность достаточно для того, чтобы исключить диагноз.

Рисунок 6 иллюстрирует еще одну важную идею: специфичность теста главным образом влияет на интерпретацию положительного результата, чувствительность – главным образом на интерпретацию отрицательного результата. На обоих рисунках (а) и (б) верхнее семейство кривых соотносится с положительным, а нижнее – с отрицательными результатами теста. Рисунок 6а показывает вероятность после теста для тестов с меняющейся специфичностью (TNR). Заметим, что изменение специфичности сильно влияет на верхнее семейство кривых (положительные результаты). Это значит, что рост специфичности теста заметно влияет на апостериорную вероятность, если результат теста положительный, но сравнительно слабо, если результат теста отрицательный.

Таким образом, при попытке подтверждения диагноза [Подтвердить диагноз значит в медицине удостовериться, что пациент болен этой болезнью, опровергнуть – удостовериться в обратном. Врач, сильно подозревающий, что у пациента бактериальная инфекция, назначает посев, дабы подтвердить диагноз; другой врач почти уверен, что у пациента обычная ангина, и не назначает посева, чтобы опровергнуть стрептококковую инфекцию миндалин.] следует выбирать тест с высокой специфичностью. Рисунок 6b отвечает изменениям чувствительности. Она, как можно заметить, является причиной больших изменений в нижнем семействе кривых (отрицательных результатов теста), но маленьких – в верхнем семействе. Получается, что для опровержения диагноза следует выбирать тест с высокой чувствительностью.

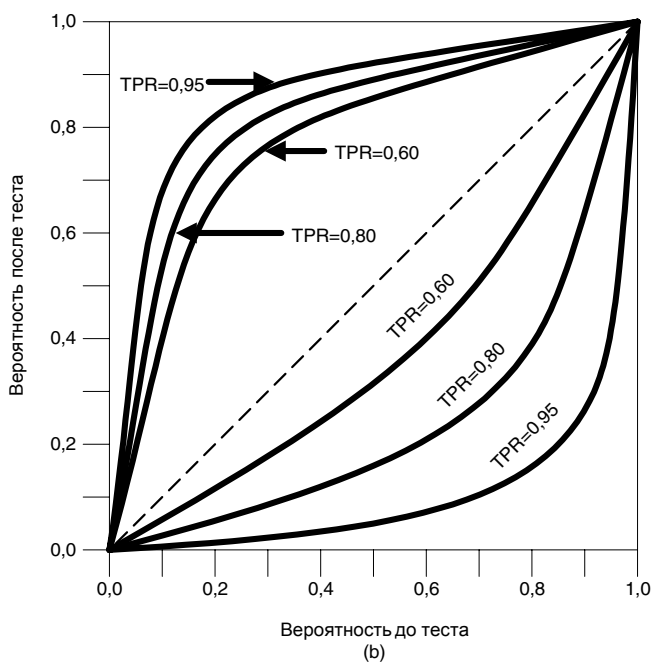
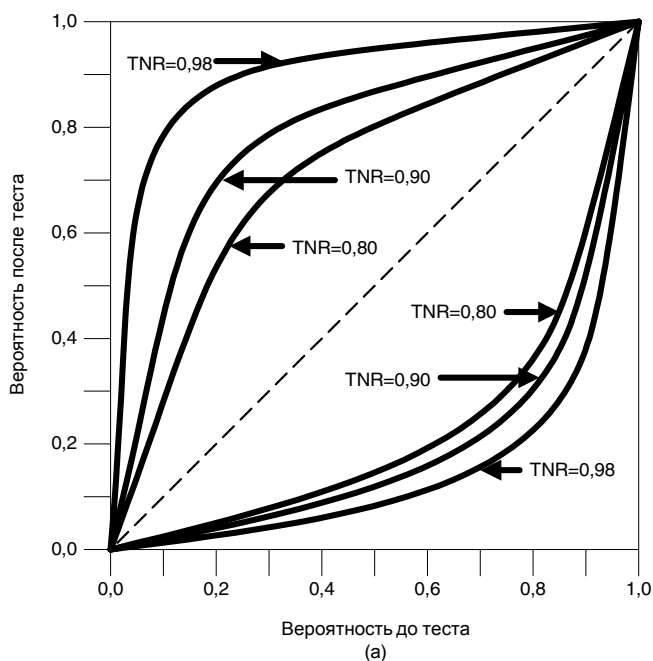


Рис. 6. Влияние чувствительности и специфичности на вероятность после теста. Рисунки (а) и (b) подобны рисункам 5а и 5b, кроме как в том, что вычисления были повторены для нескольких значений чувствительности (TPR, доли истинных положительных результатов) и специфичности (TNR, доли истинных отрицательных результатов) теста. В (а) чувствительность теста была положена равной 0,90, а значения специфичности менялись. В (b) менялась чувствительность, а специфичность принята равной 0,90. На обоих графиках верхнее семейство кривых отвечает положительным, а нижнее – отрицательным результатам теста.

ИСТОЧНИК: Sox, Jr., H. C. Probability theory in the use of diagnostic tests: Application to critical study of the literature. In Sox, Jr., H. C. (ed), *Common Diagnostic Tests: Use and Interpretation*. Philadelphia: American College of Physicians, 1987, p. 1-17. Публикуется с разрешения Американского врачебного колледжа.

3.4.5. Предосторожности при применении теоремы Байеса

Теорема Байеса дает мощный метод вычисления апостериорной вероятности. Однако следует знать возможные ошибки при применении этого метода. Распространены четыре проблемы: неаккуратная оценка априорной вероятности, применение неверных параметров, характеризующих тест, неверное предположение об условной независимости и взаимном исключении.

Теорема Байеса обеспечивает средства для корректировки оценки вероятности (полученной до теста) с целью учета новой информации. Однако точность вычисления апостериорной вероятности ограничена точностью оцененной априорной вероятности. Точность же априорной оценки достигается с помощью правильного использования опубликованных значений распространенности, эвристических методов и правил клинических прогнозов, и обычно может быть сделана приемлемой. В анализе решений, как мы увидим, часто достаточно знать *интервал* априорной вероятности. Тем не менее, если ненадежна оценка априорной вероятности, теорема Байеса вряд ли будет иметь ценность.

Вторая потенциальная ошибка – использование при вычислениях чувствительности и специфичности без учета возможной необъективности источников (см. пункт 3.5).

Третья потенциальная проблема появляется, когда теорема Байеса используется для интерпретации последовательности тестов. Если пациент подвергается последовательно двум тестам, можно использовать вероятность после первого из них, вычисленную по теореме Байеса, в качестве априорной для второго теста. По ней с помощью теоремы Байеса можно вычислить вероятность после второго теста. Такой подход, однако, обоснован, только если тесты *условно независимы*. Тесты на один и тот же диагноз называются **условно независимыми**, если вероятность конкретного результата второго теста не зависит от результата первого *при условии* данного состояния болезни/здоровья. В терминах условной вероятности для случая, когда болезнь присутствует, это выражается так:

$$p[\text{второй тест положителен} \mid \text{первый тест положителен и болезнь присутствует}] = \\ p[\text{второй тест положителен} \mid \text{первый тест отрицателен и болезнь присутствует}] = \\ p[\text{второй тест положителен} \mid \text{болезнь присутствует}].$$

При применении теоремы Байеса последовательно в ситуации, когда условная независимость нарушена, точности вычислений не будет.

Четвертая распространенная проблема возникает при предположении, что всякий ненормальный результат теста есть следствие одного и только одного диагноза. В байесовском подходе, как мы его описали, в общем принимается, что болезни, подлежащие рассмотрению, **взаимно исключают** друг друга. Если это не так, применять формулу Байеса надо с большой осторожностью.

Мы показали, как вычислять апостериорную вероятность. В следующем разделе мы вернемся к проблеме, как принимать решения, когда результат действий врача

(например, лечения) не определен.

3.5. Принятие решений методом математического ожидания

Проблемы принятия решений в медицине часто не могут быть решены с помощью обоснований из области патофизиологии. Например, когда исход лечения (скажем, хирургической операции) непредсказуем, врачу нужен метод для выбора лечения. Для разрешения таких сложных проблем можно использовать идеи, развитые в предыдущих разделах. Мы обсудим два метода: дерево решений, метод представления и сравнения ожидаемых исходов для каждого возможного варианта действий, и пороговую вероятность, метод решения того, может ли новая информация изменить определяющее решение. Эти методы помогают внести ясность в проблему принятия решений и таким образом выбрать вариант, который с наибольшей вероятностью поможет пациенту.

3.5.1. Сравнение неопределенных перспектив

Подобно большинству биологических событий, исход болезни пациента непредсказуем. Каким образом врач может определить, какая линия действий имеет наибольшие шансы на успех?

Пример 11. Существуют два метода лечения смертельной болезни. Продолжительность жизни пациента после каждого из них непредсказуема, как показывают частотное распределение на рис. 7 и таблица 5. Каждое лечение связано с неопределенностью: вне зависимости от того, какое лечение получает пациент, он умирает не позже конца четвертого года, и нет способа узнать, какой год будет для пациента последним. Как видно из рисунка 7, пациент доживет до начала четвертого года скорее при лечении В, но может случиться и так, что он умрет в первый год при лечении В или доживет до четвертого при лечении А.

Какой из методов лечения предпочесть? Этот пример демонстрирует важный факт: выбор метода лечения – это выбор между рисками. Как мы обычно делаем выбор между рисками? Чаще всего мы полагаемся на кривую распределения или на “шестое чувство”. Как следует выбирать между рисками? Мы предложим метод выбора, называемый методом математического ожидания: каждому риску ставится в соответствие число, и выбор сводится к сравнению этих чисел [Более общий термин для метода математического ожидания – метод ожидаемой полезности. Однако понятие полезности находится вне наших интересов в этой главе, и поэтому мы используем термин “математическое ожидание”]. Оба лечения А и В из примера 11 рискованы по отношению к продолжительности жизни после лечения. Мы хотим ввести меру (численную) каждого лечения, которая резюмировала бы исходы таким образом, что мы могли бы решить, какое лечение следует предпочесть.

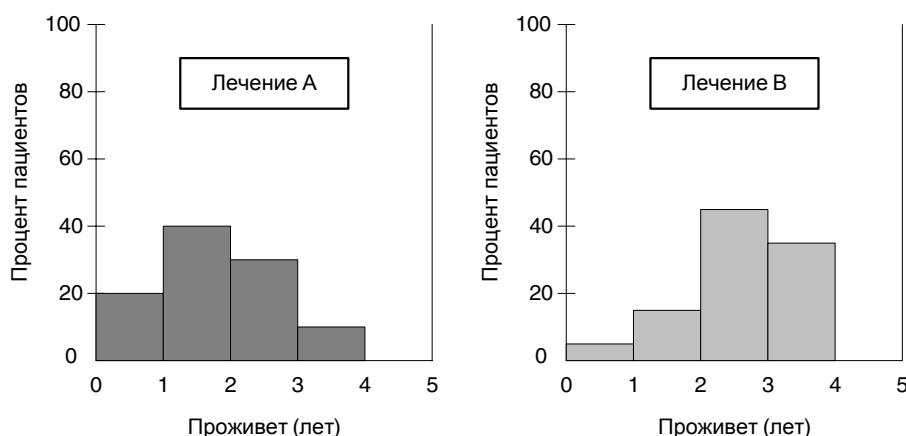


Рис. 7. Продолжительность жизни при разных методах лечения смертельной болезни. Доступны два метода; результат каждого непредсказуем.

Таблица 5. Распределение вероятностей для двух методов лечения (см. рис. 7)

Годы после начала лечения	Вероятность смерти	
	Метод А	Метод В
1	0,20	0,05
2	0,40	0,15
3	0,30	0,45
4	0,10	0,35

Идеальный критерий выбора – число, которое отражает предпочтительность исходов риска. Мера предпочтения называется полезностью и является желанным свойством для принятия решений: следует предпочесть риск с большей полезностью. Мы лишь бегло коснемся этого понятия, подробно эта тема и детали анализа решений изложены в других учебниках (см. список рекомендованной литературы в конце этой главы) [Этот метод, прежде чем быть примененным в медицине, уже использовался во многих областях]. Мы в качестве критерия выбора будем использовать среднюю продолжительность жизни, но следует помнить, что это лишь упрощенная специально для примера модель. Далее мы рассмотрим другие факторы, такие как качество жизни.

Поскольку мы не уверены в продолжительности жизни каждого пациента, мы характеризуем метод лечения через среднюю продолжительность жизни, полученную путем рассмотрения большого числа пациентов, прошедших лечение. Первый наш шаг в вычислении средней продолжительности жизни для метода лечения – разделение пациентов, подвергнувшихся лечению, на группы с одинаковой продолжительностью жизни. Затем надо умножить время жизни каждой группы на ее долю среди всех пациентов, и, наконец, сложить полученные произведения для всех значений продолжительности жизни [В этом простом примере мы принимаем, что смерть всегда наступает в конце года.].

Произведем эти вычисления для примера 11. Средняя продолжительность жизни для метода А равна

$$(0,2 \times 1,0) + (0,4 \times 2,0) + (0,3 \times 3,0) + (0,1 \times 4,0) = 2,3 \text{ года.}$$

Средняя продолжительность жизни для метода В равна

$$(0,05 \times 1,0) + (0,15 \times 2,0) + (0,45 \times 3,0) + (0,35 \times 4,0) = 3,1 \text{ года.}$$

Продолжительность жизни при лечении случайна. Лечение А – это риск, характеризующийся средней продолжительностью жизни, равной 2,3 года; лечение В характеризуется средней продолжительностью жизни 3,1 года. Если критерием выбора является длина жизни, следует выбрать лечение В.

3.5.2. Представление выбора через дерево решений

На рисунке 8 в виде диаграммы изображен выбор между лечениями А и В. Случайные события могут быть представлены в виде узла шансов. По соглашению, узел шансов изображается как окружность, из которой исходят несколько линий. Каждая линия соответствует одному из возможных исходов; с ней связана вероятность этого исхода. Для одного пациента возможен только один исход. Кстати говоря, некоторые врачи под этим предлогом отказываются от использования вероятностей: “как можно полагаться на данные по популяции, если каждый пациент индивидуален?” Фактически же часто для того, чтобы сформулировать мнение о том, что может случиться с одним пациентом, мы обязаны использовать частоту исходов, полученную по опыту многих пациентов. Мы можем скорректировать такую частоту исходов по специфике конкретного пациента и оценить таким образом вероятность каждого исхода из узла шансов.

Узел шансов может показать больше, чем просто случайное событие. Исход случайного события, для конкретного случая неизвестный, может быть представлен **математическим ожиданием** узла шансов. Понятие математического ожидания очень важно и легко для понимания. Мы можем вычислить среднюю продолжительность жизни, ожидаемую у пациентов, подвергшихся лечению, узел шансов которого изображен на рисунке 8. Это средняя длина жизни называется ожидаемой продолжительностью жизни, или, в более общих терминах, математическим ожиданием (для) узла шансов. Процесс вычисления мы уже описали: мы перемножаем продолжительность жизни, связанную с данным исходом, и его вероятность, а затем складываем получившиеся числа по всем исходам. Таким образом, если нескольким сотням пациентов было назначено лечение А или лечение В, мы получим среднюю продолжительность жизни, равную 2,3 года для лечения А и 3,1 года для лечения В.

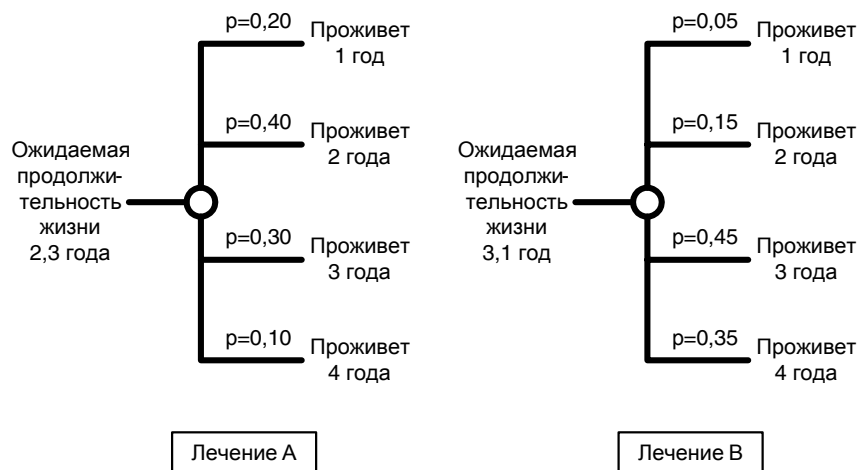


Рис. 8. Представление продолжительности жизни при двух методах лечения с рисунка 7. в виде узла шансов. Вероятность того, что пациент проживет данное число лет, умножается на это количество лет, затем эти произведения складываются и образуют ожидаемую продолжительность жизни.

Мы описали основу принятия решений с использованием математического ожидания. Термин математическое ожидание используется для характеристики случайного события, например, исхода лечения. Если исход лечения измеряется в единицах продолжительности жизни, степени здоровья [Если в качестве меры предпочтения используется полезность, процесс называется принятием решения на основе максимальной полезности.] или в долларах, то лечение, соответственно, будут характеризовать такие величины, как ожидаемая продолжительность жизни, ожидаемая степень здоровья, или ожидаемая денежная стоимость лечения для пациента.

Используя принятие решений с помощью математического ожидания при выборе между методами с неопределенными исходами, действуют по следующему алгоритму:

Вычисляют математическое ожидание для каждого возможного варианта, а затем выбирают вариант с самым высоким математическим ожиданием.

3.5.3. Анализ решений

Внесем ясность в понятие “принятие решений методом математического ожидания”, рассмотрев пример. Прежде всего заметим. Что анализ решений состоит из четырех этапов:

1. Создание дерева решений (это самый сложный этап, так как он требует формулировки проблемы выбора, назначения вероятностей и измерения исходов).
2. Вычисление математического ожидания для каждого возможного варианта выбора.
3. Выбор варианта с самым высоким математическим ожиданием.
4. Анализ чувствительности с целью проверки полученного решения.

Многие работники здравоохранения, впервые узнав о технике анализ решений, оставляют эту информацию без внимания, поскольку видят возможность ошибки в оценке как вероятностей, так и полезности (по дереву решений). Они говорят, что эта техника потворствует решениям, принятым на основе маленькой разницы между ожидаемыми значениями, которые являются в лучшем случае лишь оценками. Осознав это, профессиональные аналитики разработали защиту от такого мнения – анализ чувствительности. Мы осудим этот важный четвертый этап анализа решений в пункте 5.4.

Первый шаг в анализе решений – создание дерева решений, представляющего нашу проблему. Рассмотрим такой пример.

Пример 12. Пациенту, мистеру Дэнби, 66 лет, он страдает настолько жестоким артритом обоих коленных суставов, что может передвигаться по дому лишь с помощью двух палок, а во всех других случаях вынужден ездить на инвалидной коляске. Другая его проблема – эмфизема, болезнь, при которой легкие теряют способность обмена кислородом и диоксидом углерода между кровью и воздухом, что в свою очередь влечет за собой одышку (затрудненное дыхание). Передвигаясь на инвалидной коляске, он может дышать спокойно, но ходьба с палками заставляет его тяжело дышать и плохо себя чувствовать. Несколько лет назад он обдумывал перспективу операции по протезированию коленного сустава, но отказался от нее – во многом по совету своего терапевта, предупредившего его о серьезном риске не выжить после этой операции из-за больных легких. Однако недавно жена мистера Дэнби перенесла инсульт, и с тех пор частично парализована; в теперешнем состоянии пациент не может оказать ей должного ухода. Он сообщил своему врачу, что решил согласиться на операцию.

Терапевт мистера Дэнби знаком с методами анализа решений. Он осознает, что проблема напичкана неопределенностями. Под сомнением способность мистера Дэнби выжить при операции, а сама операция не всегда восстанавливает способность двигаться до нужного пациенту уровня. Мало того, существует риск, хотя и небольшой, что с протезом (искусственным коленом) будет занесена инфекция, и тогда мистеру Дэнби предстоит еще одна опасная операция по удалению протеза. С удаленным протезом мистер Дэнби

вообще не сможет ходить, даже на костылях. Возможные исходы протезирования включают смерть при первой операции и смерть при второй операции, если будет занесена инфекция (мы предположим, что в последнем случае, если смерть произойдет, то сразу после операции). Возможные функциональные исходы включают полное восстановление двигательных функций или их улучшение, и то, что способность двигаться останется на прежнем низком уровне. Идти мистеру Дэнби на операцию или оставаться в прежнем состоянии?

Используя принятые в анализе решений обозначения, врач составляет дерево решений, показанное на рис. 9. Квадрат используется для обозначения **узла решений**, а линии, исходящие из него – для представления возможных действий.

Прежде всего, согласно методам принятия решений с использованием математического ожидания, терапевт должен задать вероятность каждой ветви каждого узла. Для этого он спрашивает нескольких хирургов-ортопедов об их оценке шансов на полное восстановление двигательных функций (и получает ответ, что вероятность полного восстановления – 0,60) и шансов занесения инфекции в протезе сустава (вероятность инфекции – 0,05). Субъективными методами оценки вероятности он получает, что пациент умрет во время или сразу после операции с вероятностью 0,05.

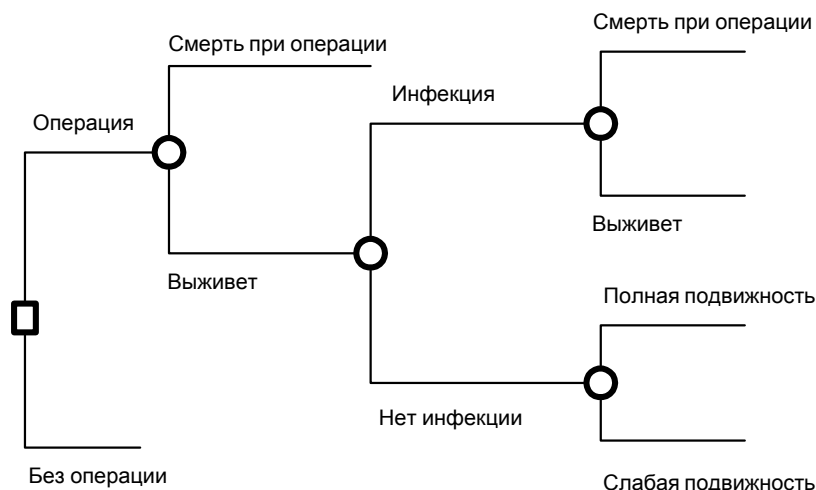


Рис. 9. Дерево решений для операции по протезированию коленного сустава. Квадрат обозначает узел решения (идти ли на операцию), круги – узлы шансов

Таблица 6. Возможные исходы в примере 12

Годы жизни	Функциональный статус	Годы полного функционирования, соответствующие исходу
10	Полная подвижность (удачная операция)	10
10	Слабая подвижность (неудачная операция или status quo)	6
10	Прикованность к инвалидной коляске	3
0	Смерть	0

Далее он должен приписать каждому исходу свое значение. Для этого он, прежде

всего, перечисляет возможные исходы. Как видно из таблицы 6, исходы различаются по двум параметрам: продолжительность жизни и качество жизни (функциональный статус). Для точной характеристики каждого исхода врачу следует разработать меру для подсчета каждого из этих параметров. Здесь недостаточно использовать только продолжительность жизни, так как пять лет жизни при хорошем здоровье – это лучше, чем пять же лет при плохом состоянии здоровья. Терапевт учесть этот фактор, выразив оба измерения через одно: продолжительность жизни при хорошем здоровье. Полученная мера называется **годы жизни с учетом качества (QALY)** [годы жизни с учетом качества – мера полезности, часто используемая в медицинском анализе решений и анализе решений в здравоохранении (см., например, обсуждение анализа “затраты-эффективность” в главе 19)]. Врач может сопоставить годы жизни при слабом здоровье и годы жизни при хорошем здоровье, попросив мистера Дэнби обозначить самый короткий срок жизни при хорошем здоровье, на который он согласился бы в обмен на полный ожидаемый срок жизни (10 лет) в теперешнем состоянии. Итак, врач спрашивает мистера Дэнби: “Многие говорят, что променяли бы долгую жизнь с серьезным физическим недостатком на жизнь короткую, но здоровую. Как вам кажется, сколько лет жизни при полной способности двигаться были бы для вас эквивалентны десяти годам жизни в данном состоянии?” Этот вопрос он повторяет для каждого возможного исхода. Ответы пациента изображены в третьей колонке таблицы 6. Пациент считает, что десять лет ограниченной способности двигаться эквивалентны шести годам нормальной подвижности, в то время как десять лет на инвалидной коляске соответствуют только трем годам полной подвижности. На рисунке 10 изображено окончательное дерево решений – дополненное оценками вероятности и значениями полезности для каждого исхода [при более тонком анализе решений врачу следовало бы скорректировать значения полезности для случаев, требующих операции, чтобы учесть боль и неудобства, связанные с операцией и послеоперационной реабилитацией.].

Вторая задача, стоящая перед врачом – вычисление математического ожидания, выражаемого в “годах здоровья” при условии операции и без нее. Врач вычисляет математическое ожидание для каждого узла шансов, двигаясь справа (от “листьев” дерева) налево (к “корню” дерева). Рассмотрим, например, математическое ожидание для узла шансов, соответствующего исходу операции по извлечению зараженного протеза (узел А на рис. 6). Вычисление состоит из трех этапов:

Вычисление математического ожидания QALY для смерти при операции по удалению зараженного протеза – умножение вероятности смерти при операции (0,05) на QALY исхода (смерти) – 0 лет:
 $0,05 \times 0 = 0$ (QALY).

Вычисление математического ожидания QALY для выживания при этой операции – умножение вероятности выживания (0,95) на число лет жизни с учетом качества, соответствующее исходу (инвалидной коляске) – 3 года: $0,95 \times 3 = 2,85$ (QALY).

Сложение значений, полученных на шаге 1 (0 QALY) и на шаге 2 (2,85 QALY) для получения математического ожидания развития инфекции в протезе: $0 + 2,85 = 2,85$ (QALY).

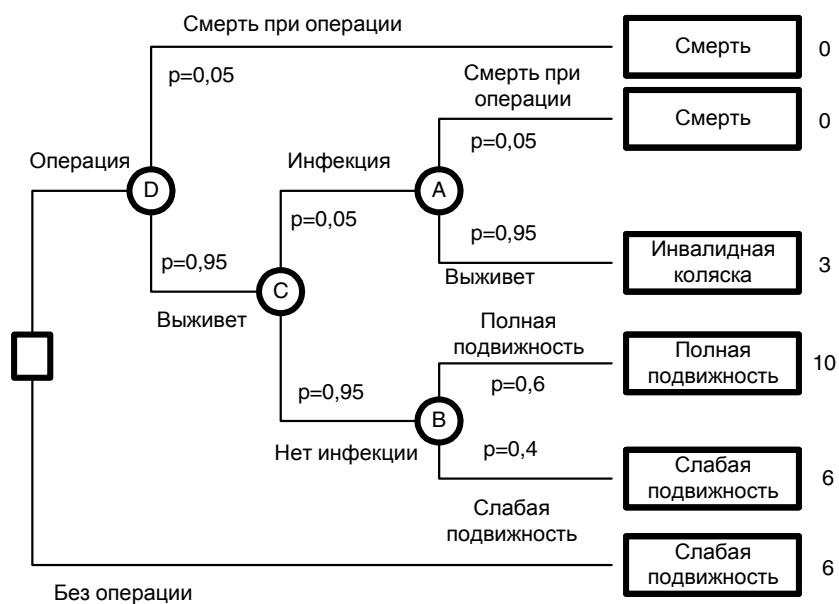


Рис. 10. Дерево решений для операции по протезированию коленного сустава. Каждой ветви каждого узла шансов приписана вероятность. Каждому “листу” (концу каждой ветви) дерева приписана оценка этого исхода, сделанная пациентом и выражающаяся в годах полной подвижности

Таким же образом вычисляется математическое ожидание QALY для узла B: $(0,6 \times 10) + (0,4 \times 6) = 8,4$ QALY. Для получения значения математического ожидания QALY для случая выживания после первой операции, мы действуем следующим образом:

Умножаем уже полученное значение математического ожидания для случая инфекции (2,85 QALY) на вероятность заражения протеза (0,05): $2,85 \times 0,05 = 0,143$ QALY.

Умножаем уже полученное значение математического ожидания для случая. Когда в протез не попадает инфекция (8,4 QALY) на вероятность такого исхода (0,95): $8,4 \times 0,95 = 7,98$ QALY

Сложить значения, полученные на этапе 1 (0,143 QALY) и на этапе 2 (7,98 QALY): $0,143 + 7,98 = 8,123$ QALY.

Этот процесс, называемый **усреднением по узлам шансов**, с тем же успехом работает и для узла D. Мы движемся к корню дерева, пока не получаем значение математического ожидания QALY для операции. В нашем примере, если мистер Дэнби идет на операцию, то ожидание продолжительности жизни, выраженное в годах полной подвижности, будет равно 7,7. Что значит это число? Оно не означает, что при условии операции мистеру Дэнби гарантированы 7,7 лет подвижной жизни. При одном взгляде на дерево решений видно, что некоторые пациенты при такой операции умирают, некоторые – получают инфекцию, некоторым операция не приносит улучшения. Таким образом, для конкретного пациента нет никаких гарантий. Однако если операции подвергаются, скажем, 100 одинаковых пациентов, то **среднее** число лет полной подвижности будет 7,7. Рассмотрев тот вариант, при котором мистер Дэнби не идет на операцию, мы поймем, что для него значит этот результат (7,7 лет).

Без операции средняя продолжительность жизни, выраженная в годах нормальной подвижности, будет, по оценке мистера Дэнби, равна 6,0 лет на каждые 10 лет слабой подвижности. Некоторые пациенты проживут дольше, а некоторые –

меньше 10 лет (при слабой подвижности), но средняя продолжительность жизни, выраженная в годах нормальной подвижности, будет равна 6. Поскольку 6 меньше 7,7, **в среднем** операция обеспечивает лучший исход для пациента, и поэтому терапевт рекомендует ее.

Из этого примера видна ключевая идея для понимания принятия решений с помощью математического ожидания: анализ решений не может гарантировать хорошего исхода. Но он может помочь врачу определить лечение, которое в среднем по большому числу подобных пациентов даст наилучший результат. Анализ решений подгоняется под конкретного пациента, так как и функции полезности, и оценки вероятности корректируются индивидуально. Тем не менее, результаты анализа представляют исходы, которые могли бы произойти **в среднем** в группе подобных пациентов. При условии, что исход для конкретного пациента непредсказуем, лучшим выбором для него будет вариант, дающий наилучший результат в среднем у подобных пациентов.

3.5.4. Анализ чувствительности

Анализ чувствительности – это проверка обоснованности заключений анализа по широкому интервалу предположений о вероятностях и значениях полезности. Вероятность исхода в узле шансов может быть оценена наилучшим возможным образом, но часто велик интервал “приемлемых” вероятностей, которые врач может использовать почти с равными на то основаниями. Анализ чувствительности используется для ответа на вопрос: меняется ли решение о предпочтительном выборе, когда оценки вероятностей и исходов меняются в соответствующем интервале?

Вопрос о протезировании коленного сустава из примера 12 показывает силу анализа чувствительности. Если заключение анализа (лучше делать операцию, чем не делать) остается тем же, несмотря на широкий разброс принимаемых значений вероятностей и мер исходов, то эта рекомендация заслуживает доверия. Рисунки 11 и 12 показывают ожидаемую продолжительность жизни с учетом качества при условии операции и без нее при меняющихся принятых значениях вероятности смерти при операции и вероятности восстановления полной подвижности.

Каждая точка на этих линиях представляет собой математическое ожидание лет жизни с учетом качества, вычисленное по дереву с рис. 9 для соответствующего значения параметра. Из рис. 11 видно, что операция предпочтительнее, чем отсутствие операции при широком интервале значений смертности при операции – ожидаемая продолжительность жизни с учетом качества становится ниже, чем без операции, лишь когда смертность превышает 25%. Рисунок 12 показывает влияние вероятности восстановления полной подвижности: следует предпочесть операцию уже когда эта вероятность превышает 20%, а в нашем примере это число гораздо выше (60% по оценке консультировавших хирургов-ортопедов). Поэтому терапевт может смело продолжать рекомендовать операцию. Мистер Дэнби не может быть уверен в благополучном ее исходе, но он имеет веские основания считать, что с большей вероятностью будет хорошо себя чувствовать, сделав операцию, чем отказавшись от нее.

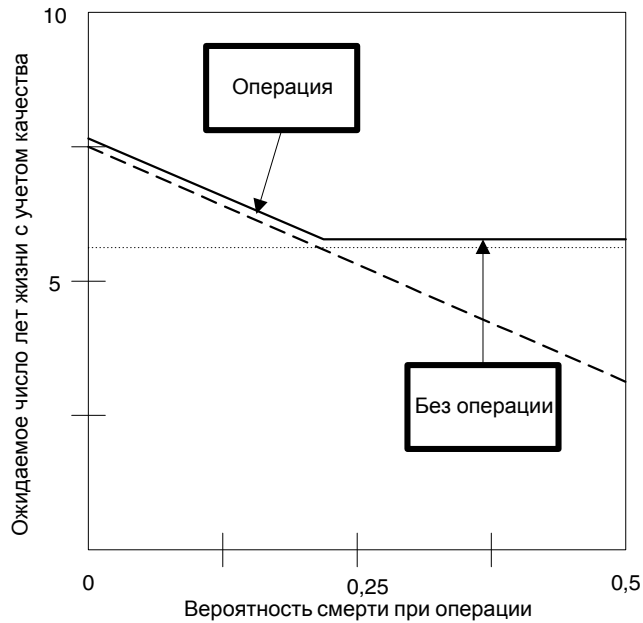


Рис. 11. Анализ чувствительности влияния смертности при операции на ожидаемую длину здоровой жизни (пример 12). По мере роста вероятности смерти меняется отношение значений при условии операции и без нее. Точка пересечения двух линий представляет вероятность, при которой становится предпочтительным не делать операцию. Сплошная линия показывает предпочтительный выбор в зависимости от вероятности.

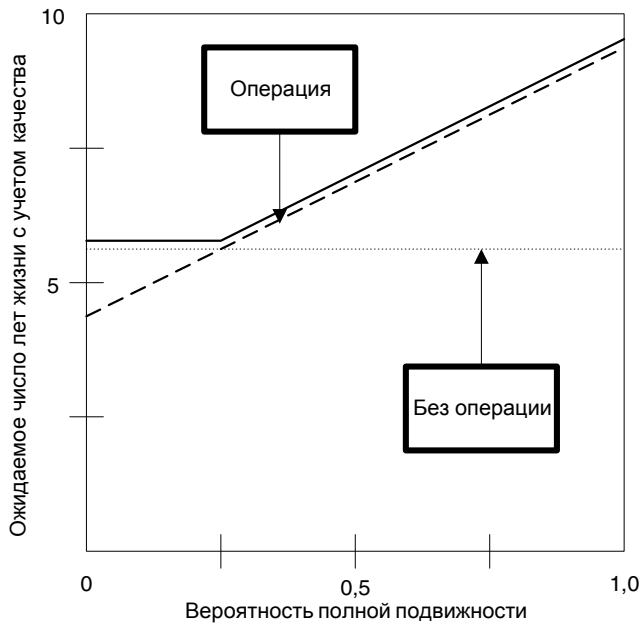


Рис. 12. Анализ чувствительности влияния вероятности успешной операции на ожидаемую длину здоровой жизни (пример 12). По мере роста вероятности удачной операции меняется отношение значений при условии операции и без нее. Точка пересечения двух линий представляет вероятность, при которой становится выгодным делать операцию.

Сплошная линия показывает предпочтительный выбор в зависимости от вероятности.

Другой способ утверждения заключения с помощью анализа чувствительности – обозначить интервал вероятностей, при котором заключение действительно. Точка пересечения двух линий на рисунке 11 – это вероятность смерти при операции, при которой ожидаемые значения продолжительности жизни с учетом качества при условии операции и без нее становятся равны. Если эти ожидаемые значения – основа выбора, то при вероятности смерти, равной 25% врачу (и пациенту) должно быть безразлично, делать операцию или нет [Смертность 20% при операции может показаться слишком большой, однако это значение корректно, если в качестве основы выбора используется мера QALY. При более сложном анализе можно использовать функцию полезности, отражающую уровень дискомфорта пациента, рискующего жизнью. Функции полезности обсуждаются в дополнительной литературе (см. список в конце главы)]. Когда вероятность ниже этого значения, следует выбрать операцию, когда выше – отказаться от нее. Вероятность, при которой меняется предпочтительный выбор (точка пересечения линий вариантов выбора), называется пороговой вероятностью. В следующем пункте мы обсудим это важное для принятия решений понятие.

3.6. Лечить, обследовать или ничего не делать?

Врач, столкнувшийся с диагностической проблемой и оценивший в числах один из симптомов пациента, должен выбрать между следующими действиями:

1. Ничего больше не делать (не назначать дополнительных тестов и не лечить пациента).
2. Перед выбором, лечить ли пациента, затребовать дополнительную информацию (обследовать).
3. Лечить, не ожидая дальнейшей информации.

Этот выбор упрощается, если врач знает истинное состояние пациента: обследование тогда не нужно, и врачу надо только выбрать между возможными терапевтическими действиями (см. пример 12). Однако знание истинного состояния пациента обычно требует дорогостоящих, требующих времени и часто опасных и неприятных пациенту диагностических процедур. Поэтому практикующие врачи часто хотят решить, лечить ли пациента, если даже не абсолютно уверены в его истинном состоянии. В этом случае они рискуют, так как могут по ошибке оставить без лечения действительно больного пациента, а могут, назначив лечение пациенту, у которого рассматриваемой болезни нет, заставить его страдать от побочных эффектов этого лечения.

Решить, что делать – лечить, обследовать или бездействовать – бывает трудно, но выше уже описаны все принципы, используемые при решении проблем такого рода. Решение состоит из трех этапов:

1. Определить пороговую вероятность лечения болезни.
2. Определить априорную вероятность диагноза.
3. Решить, может ли результат теста повлиять на ваше решение о лечении.

Пороговая вероятность лечения болезни – это вероятность болезни, при которой все равно, лечить болезнь или не лечить. Если вероятность выше пороговой, лечить следует, если ниже, то нет (рис. 13).

Проблему, лечить ли, когда диагноз не определен, можно решить с помощью дерева решений, такого, как показано на рис. 14. Это дерево можно использовать для установления пороговой вероятности лечения, оставив вероятность болезни неизвестной и установив математическое ожидание QALY для операции равным математическому ожиданию QALY для терапевтических (то есть нехирургических, таких как прием лекарств и физиотерапия) методов лечения. По этим условиям

надо найти вероятность болезни (в этом примере хирургическое вмешательство соответствует ветви дерева “лечить” с рис. 14, а нехирургическое лечение – ветви “не лечить”). Поскольку при найденной вероятности выбор метода лечения безразличен, это и будет пороговая вероятность лечения. Такое ее определение составляет первый шаг решения проблемы. На практике пороговую вероятность лечения часто устанавливают скорее интуитивно, чем аналитически.

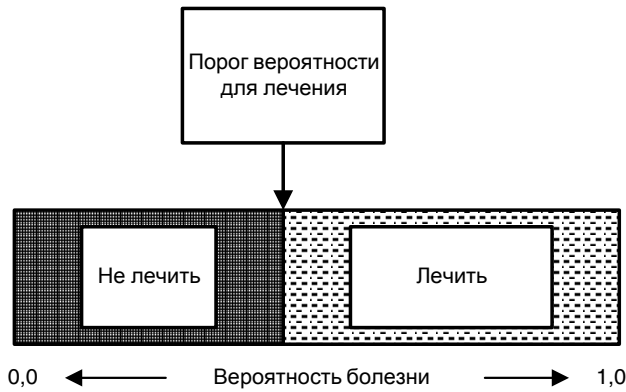


Рис. 13. Пороговая вероятность лечения (схема)

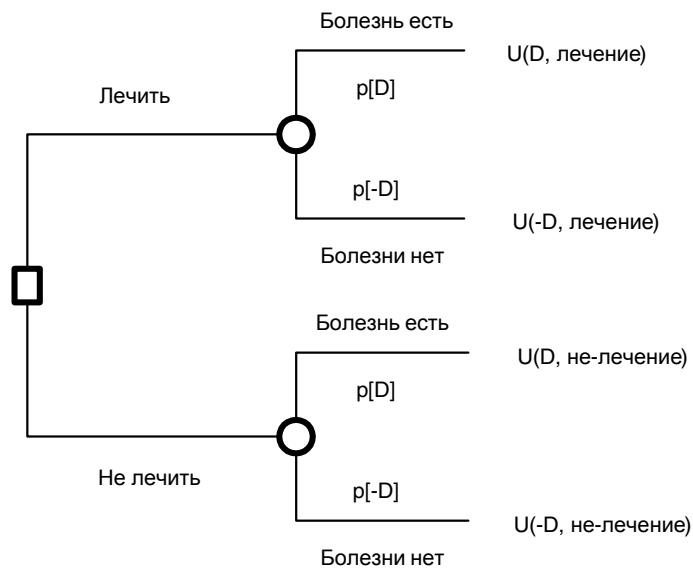


Рис. 14. Дерево решений для вычисления пороговой вероятности лечения болезни. Установив полезность для вариантов лечить и не лечить равной, можно вычислить вероятность, при которой выбор врачу безразличен. Помните, что $p[-D] = 1 - p[D]$.

Поскольку априорная вероятность известна, известны и действия при отсутствии дальнейшей информации о пациенте. Если априорная вероятность выше

пороговой, то лечить следует, если ниже – то не следует. Таков второй шаг решения проблемы.

Один из ведущих принципов принятия медицинских решений гласит: не надо назначать обследование, если оно не может изменить действия по отношению к пациенту. В нашей терминологии этот принцип формулируется так: назначать обследование следует только если его результат может повлечь изменение статуса вероятности относительно пороговой вероятности. Другими словами, если вероятность до теста превышала пороговую, то отрицательный результат теста должен снижать вероятность болезни до уровня ниже пороговой; и наоборот, если вероятность была ниже пороговой, то положительный результат теста должен давать апостериорную вероятность, превышающую пороговую. В каждом из этих случаев результат теста может изменить решение о том, лечить ли пациента.

Чтобы установить, может ли тест изменить действия по отношению к пациенту, просто применяется теорема Байеса. По ней вычисляется апостериорная вероятность для результата теста (положительного или отрицательного), который мог бы сдвинуть вероятность болезни относительно пороговой вероятности лечения. Если априорная вероятность болезни больше пороговой, то вероятность болезни вычисляется для отрицательного результата теста, если меньше – для положительного. Эти вычисления составляют третий шаг.

Пример 13. Врач, специалист по болезням легких, подозревает, что у пациента легочная эмболия (тромб, застрявший в сосудах легких). Обычный метод – сделать радиоизотопное сканирование легких, тест, при котором в вену вводятся крохотные радиоактивные частички, которые проходят по тонким сосудам легких. Сканирующее устройство улавливает радиацию частиц. Частицы не могут пройти в часть легких, которую снабжает заблокированный тромбом сосуд. К сожалению, есть и другие причины белых пятен в картине сканирования, так что врач не может быть уверен, что белое пятно означает именно тромб. Поэтому, если сканирование показало белое пятно, врач должен провести еще один тест для подтверждения диагноза. Этот тест – ангиограмма легких, при которой в артерии легких вводится радиоактивная краска, и делается рентгеновский снимок. Эта процедура требует некоторого риска и дискомфорта для пациента и стоит значительную сумму. Если результат сканирования отрицательный, ни проводить дополнительные тесты не нужно.

Чтобы установить правильную стратегию, врач предпринимает следующие шаги:

1. Определяет пороговую вероятность лечения легочной эмболии.
2. Оценивает априорную вероятность легочной эмболии.
3. Решает, может ли повлиять на его решение результат теста.

Прежде всего, предположим, что пороговая вероятность для этого пациента равна 0,10. Что это значит? Если врач не может получить дополнительную информацию, то при априорной вероятности выше 0,10 (то есть если врач верит, что шанс легочной эмболии у пациента больше одной десятой) ему надо начать лечить эмболию, а если априорная вероятность ниже 0,10, от лечения следует воздержаться. Это означает, что если вероятность эмболии равна пороговой, то врач будет лечить от нее девять пациентов без эмболии на одного настоящего больного. Такой подход может быть обоснован, потому что больному эмболией (которая может быть смертельной) лечение помогает очень хорошо, в то время как лечение пациента без эмболии от нее сравнительно безопасно. Эти рассуждения составляют первый шаг.

Врач оценил априорную вероятность легочной эмболии в 0,05. Поскольку это значение меньше порогового, то следует ничего не делать, если только положительный результат сканирования легких не может увеличить вероятность легочной эмболии до 0,10 или выше. Это рассуждение составляет второй шаг.

Чтобы определить, может ли результат теста повлиять на решение о лечении, врач должен решить, увеличит ли положительный результат сканирования легких вероятность легочной эмболии до уровня выше 0,10 (пороговой вероятности). Врач просматривает литературу и находит в ней данные, что доля истинных положительных результатов при сканировании легких равна 0,75, а доля истинных отрицательных результатов – 0,25.

Отрицательный результат сканирования уведет вероятность дальше от пороговой и не поменяет решение. Положительный же результат сдвинет вероятность диагноза в сторону пороговой и, если вероятность превысит пороговую, может изменить решение о действиях по отношению к пациенту. Итак, врач вычисляет вероятность диагноза после теста (сканирования легких), если его результат положителен. По теореме Байеса:

$$p[D|+] = \frac{p[D] \times \text{TPR}}{p[D] \times \text{TPR} + (1 - p[D]) \times \text{FPR}} = \frac{0,05 \times 0,75}{0,05 \times 0,75 + 0,95 \times 0,25} = 0,14.$$

Поскольку полученная вероятность легочной эмболии выше пороговой, положительный результат сканирования изменит решение врача, и поэтому ему следует назначить сканирование. Этот результат составляет третий шаг.

Этот пример особенно полезен по двум причинам. Во-первых, он демонстрирует один из методов принятия решений. Во-вторых, он показывает, как в принятии медицинских решений используются понятия, введенные в этой главе.

3.7. Роль вероятности и анализа решений в медицине

Прочитав эту главу, вы возможно, хотели бы знать, каким же образом теория вероятностей и анализ решений могут быть плавным образом введены в медицинскую практику. Понимание вероятности и параметров, характеризующих тест, предотвратит некоторое число несчастных случаев. В примере 1 мы обсуждали гипотетический тест, который на вид казался точным способом проверки донорской крови на СПИД. Качественный анализ, однако, показал, что из-за низкой распространенности антител к ВИЧ среди населения, результаты теста чаще были ошибочными, чем правильными [Мы подчеркиваем, что сейчас доступны тесты на ВИЧ с гораздо лучшими чувствительностью и специфичностью, чем в нашем примере. На практике кровь, давшая положительную реакцию, подвергается второму тесту с высокой специфичностью; использование последовательности тестов заметно ослабляет проблему ложных положительных результатов [Burke et al., 1988].].

Необходимость грамотной интерпретации тестов возникает очень часто. В середине 80-х годов федеральное правительство ввело практику проверки гражданских служащих, занимающих высокие должности, на наркотики. Эта практика достигла самых высоких уровней: в 1986 году сам президент Рейган предложил свою мочу для анализа на наркотики. Если у теста для проверки служащих чувствительность и специфичность были равны 0,95, а в действительности наркотики употребляло 10% служащих, то одна треть положительных результатов была бы ложной! (Случись такое в Белом Доме, положительный результат президента с вероятностью 33% был бы ложным). Понимание такого рода спорных вопросов должно вызвать у людей большой интерес; работникам здравоохранения надо быть готовыми к ответам на вопросы своих пациентов.

Хотя нам и следует аккуратно интерпретировать любой вид результатов теста, анализ решений играет в медицине более выборочную роль. Не все решения, принимаемые врачом, требуют анализа решений. Некоторые решения основаны на физиологических принципах и дедуктивном основании. Другие решения содержат мало неопределенности. Тем не менее, приходится принимать много решений, основанных на несовершенных данных, и решений, исход которых нельзя узнать в момент их принятия. Анализ решений обеспечивает технику управления такими ситуациями.

Полный анализ решений часто производит на работников здравоохранения впечатление подхода слишком громоздкого, чтобы им можно было пользоваться в трудных, но требующих быстрого решения ситуациях. Это опасение вполне разумно. Некоторые решения, конечно, дают основания для полного их анализа, но многие его не требуют. Полный анализ решений, включающий поиск нужных вероятностей в литературе и оценку значений полезности для данного пациента,

может занять много времени.

Для принятия многих решений прояснить вопрос может уже дерево, четко изображающее возможные исходы. Когда ограничено время, полезным может быть даже сделанный на скорую руку грубый анализ. Используя оценки вероятностей, данные врачами-экспертами, и угадывая, какими могут быть значения полезности для данного пациента, врач может быстро провести анализ и узнать, какие вероятности и полезности особенно сильно влияют на решение. Потратив сколько-то времени в библиотеке или у себя дома, врач может получить более точные оценки этих важных вероятностей и полезностей. Просчитав некогда одну проблему решения, врач затем может использовать ее дерево решений по мере надобности, меняя переменные для конкретного случая. Анализы решений можно найти в таких журналах, как *Medical Decision Making*; следует только скорректировать их на специфику конкретного пациента. Впервые проведя количественный анализ, врач часто будет обнаруживать, что некоторые из переменных дерева лишь незначительно влияют на решение. Поэтому в дальнейшем, анализируя эту проблему, он не будет уделять этим переменным много внимания.

Иногда работники здравоохранения неохотно используют анализ решений, так как его результат может зависеть от вероятностей, которые надо оценивать – например, априорной вероятности. Человек вдумчивый будет озабочен тем, что оценка может быть ошибочной, поскольку в медицинской литературе часто трудно достать требуемые данные. Мы, однако, утверждаем, что неопределенность клинических данных является проблемой для любого метода принятия решений, а анализ решений может по крайней мере явно оценить влияние этой неопределенности. Средство для этого – анализ чувствительности: мы можем проверить каждую переменную, и узнать, близко ли ее значение к критическому для выбора. Таким образом, можно, например определить, влияет ли на окончательное решение изменение значения априорной вероятности от 0,6 до 0,8. При таких действиях часто обнаруживается, что необходимо оценивать лишь интервал значений данной переменной, а не точное ее значение. В некоторых случаях надо знать только выше оно или ниже порогового значения переменной. Таким образом, с помощью анализа чувствительности можно решить, должна ли нас беспокоить неопределенность, касающаяся данной переменной.

В этой главе не упоминались компьютеры. Однако их использование может заметно упростить некоторые аспекты анализа решений (см. главу 15). *MEDLINE* и другие системы поиска библиографии (см. главу 14) упрощают получение характеристик тестов и опубликованных оценок распространенностей болезней. На данных, собранных в информационной системе больницы, могут быть использованы компьютерные программы для статистического анализа. Программное обеспечение для анализа решений, доступное для персональных компьютеров, может помочь врачам строить деревья решений, вычислять математическое ожидание и проводить анализ чувствительности. Со временем станут широко доступны системы клинических консультаций, которые могут оценивать данные пациента и выдавать рекомендации по диагностике и лечению.

Принятие решений в медицине часто включает в себя для врача – неопределенность, а для пациента – риск. Многие профессионалы здравоохранения приветствовали бы возможность уменьшить неопределенность в решениях. Существуют важные и трудные медицинские проблемы, для которых эту возможность предоставляет анализ решений.

Приложение. Вывод теоремы Байеса

Теорема Байеса доказывается следующим образом. Мы обозначаем условную вероятность болезни D при условии результата R теста через $p[D|R]$, априорную вероятность D – через $p[D]$. По определению условной вероятности

$$p[D|R] = \frac{p[R, D]}{p[R]} \quad (1)$$

Вероятность результата R – это сумма вероятностей для больных и небольных пациентов:

$$p[R] = p[R,D] + p[R,-D].$$

Подставляя это равенство в (1), получаем:

$$p[D|R] = \frac{p[R,D]}{p[R,D] + p[R,-D]} \quad (2)$$

Опять воспользовавшись определением условной вероятности, получаем:

$$p[R|D] = \frac{p[R,D]}{p[D]}, \text{ а } p[R|-D] = \frac{p[R,-D]}{p[-D]}.$$

Эти выражения могут быть переписаны в виде:

$$p[R,D] = p[D] \times p[R|D] \quad (3)$$

$$p[R,-D] = p[-D] \times p[R|-D] \quad (4)$$

Подставляя (3) и (4) в (2), получаем формулу Байеса:

$$p[D|R] = \frac{p[D] \times p[R|D]}{p[D] \times p[R|D] + p[-D] \times p[R|-D]}.$$

Литература

Goldman, L., et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. *New England Journal of Medicine*, 297:845, 1977.

Статья описывает разработку правила клинических прогнозов для оценки вероятностей угрожающих жизни осложнений на сердце среди пациентов, подвергающихся операциям не на сердце.

Raiffa, H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison-Westley, 1970.

Книга дает развернутое, не относящееся к медицине введение в анализ решений, теорию полезности и дерево решений.

Sox, Jr., H. C. Probability theory in the use of diagnostic tests. *Annals of Internal Medicine*, 104:60, 1986.

Статья написана для практикующих врачей и содержит обзор понятий, относящихся к вероятности и интерпретации тестов.

Sox, Jr., H. C., Blatt, M.A., Higgins, M. S., and Marton, K. I. *Medical Decision Making*. Boston, MA: Butterworths, 1988.

В этом вводном учебнике более подробно, чем в этой главе, рассмотрены те же вопросы, а также многие другие темы. Приложение содержит вероятностные отношения для ста распространенных диагностических тестов.

Tversky, A. And Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124, 1974.

Статья содержит ясное и интересное обсуждение экспериментальных результатов правильного и неправильного использования эвристики в ситуациях неопределенности.

Weinstein, M. C. And Fineberg, H. V. *Clinical Decision Analysis*. Philadelphia, PA: W. B. Saunders, 1980.

Учебник дает клиническую перспективу анализа решений. Он более детально раскрывает все вопросы, освещенные в этой главе, и охватывает оценку полезности (в том числе – годы жизни с учетом качества) и анализ эффективности стоимости.

Упражнения

1. Вычислите следующие вероятности для пациента, собирающегося на операцию по шунтированию коронарной артерии:

- а) Единственные возможные взаимоисключающие исходы – смерть, облегчение состояния или продолжение симптомов. Вероятность смерти равна 0,02, вероятность облегчения симптомов – 0,80. Какова вероятность того, что у пациента сохранятся симптомы?
- б) Два возможных осложнения при операции – инсульт и инфаркт, их вероятности равны 0,02 и 0,05 соответственно. Пациент спрашивает, каков шанс, что у него будут оба осложнения. Предположив, что осложнения условно независимы, вычислите эту вероятность.
- в) Пациенту хочется знать вероятность инсульта при условии, что в качестве осложнения у него будет и инфаркт. Предположив, что оба осложнения имеет один из пятисот пациентов, вероятность инфаркта равна 0,05, и эти события независимы, вычислите ответ.

2. Таблица 7 показывает результат гипотетического исследования по тесту ТИФА на антитела к ВИЧ (см. пример 1).

- а) Вычислите чувствительность, специфичность, распространенность болезни, положительные и отрицательные предсказывающие значения.
- б) Используя вычисленные в (А) TPR и TNR, заполните таблицу 8. Вычислите распространенность болезни и положительные и отрицательные предсказывающие значения.

3. Вас попросили интерпретировать положительный результат теста ТИФА на антитела к ВИЧ у донора-добровольца без симптомов СПИДа. Взяв его историю, вы узнали, что он употребляет наркотики (внутривенно). Вы знаете, что средняя распространенность антител к вирусу ВИЧ в вашем городе равна одной пятисотой, но среди таких наркоманов она в 20 раз выше.

- а) Оцените априорную вероятность присутствия антител в крови этого человека.
- б) Пациент рассказал вам, что два человека, с которыми он пользовался одними и теми же иглами, впоследствии умерли от СПИДа. Какой эвристический процесс вы примените для субъективной корректировки оценки из (а)?
- в) Используйте значения чувствительности и специфичности, найденные в упражнении (2а) и вычислите с их помощью вероятность того, что у пациента есть антитела к ВИЧ, после теста для его положительного и отрицательного результата. Априорную вероятность примите равной 0,10.
- г) Следует ли вам изменить TPR и TNR теста для того, чтобы положительный результат теста увеличивал вероятность болезни?

4. У вашего пациента рак. Лечить его можно с помощью химиотерапии, а можно – хирургическим путем. Если пациент выбирает операцию, то с вероятностью 2% он от нее умрет (продолжительность жизни равна нулю), с вероятностью 50% излечится (ожидаемая продолжительность жизни 15 лет), и с вероятностью 48% не вылечится (ожидаемая продолжительность жизни 1 год). Если же пациент выбирает химиотерапию, то он умрет с вероятностью 5% (ожидаемая продолжительность жизни 0 лет), вылечится – с вероятностью 65% (ожидаемая продолжительность жизни 15 лет), и с вероятностью 30% рак будет задержан, но не остановлен (ожидаемая продолжительность жизни 2 года). Создайте дерево решений и найдите математическое ожидание для каждой альтернативы в терминах ожидаемой продолжительности жизни.

5. Вы подозреваете, что у вашего пациента с больным горлом бактериальная инфекция, требующая лечения антибиотиками (в отличие от вирусной инфекции, которая вообще не лечится). Пороговую вероятность лечения вы приняли равной 0,4, а на основе осмотра оценили вероятность бактериальной инфекции в 0,8. Доступен тест с TPR = 0,75, TNR = 0,85, который показывает присутствие или

отсутствие бактериальной инфекции. Назначать ли его? Объясните ваше решение.

6. Перечислите три рода необъективности, которая может повлиять на измерение характеристики теста. Объясните, что представляет из себя каждый род, и установите, в какую сторону надо корректировать вероятность болезни после теста для компенсации необъективности каждого рода.

7. Каким образом компьютерная система может облегчить задачу комплексного анализа решений? Посмотрите на названия глав второй части этой книги. Какую роль каждая система играет в процессе принятия медицинского решения?

8. Какой самый важный вопрос вы рассматриваете, когда ищете в литературе вероятности для пациента, подобного вашему? Как следует корректировать вероятности в свете ответа на этот вопрос?

9. Как вы думаете, почему врачи иногда назначают тесты, результат которых не может повлиять на их действия по отношению к пациенту. Кажутся ли вам найденные причины имеющими основания? Имеют ли они основания только в определенных случаях? Объясните ответы.

10. Постройте дерево решений для какого-нибудь важного выбора, который вы должны сделать (относящегося к вашей жизни). Что новое вы поняли, когда строили его?