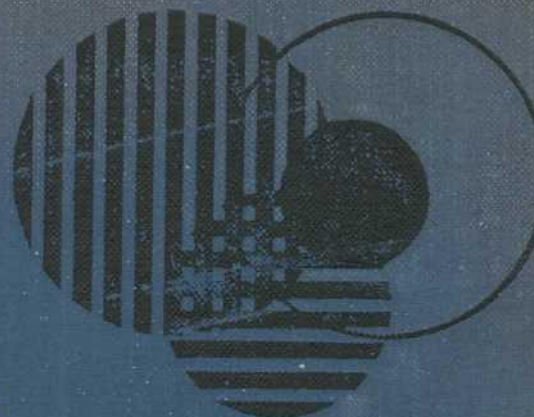


Л. ЛАСТЕД
ВВЕДЕНИЕ
В ПРОБЛЕМУ
ПРИНЯТИЯ
РЕШЕНИЙ
В МЕДИЦИНЕ



612.014

126

INTRODUCTION
TO MEDICAL DECISION MAKING

LEE B. LUSTED, M. D.
Professor and Chairman
Department of Radiology
Stritch School of Medicine
Loyola University Medical Center
Hines, Illinois

CHARLES C. THOMAS PUBLISHER
SPRINGFIELD, ILLINOIS, USA

Л. Ластед

ВВЕДЕНИЕ В ПРОБЛЕМУ
ПРИНЯТИЯ РЕШЕНИЙ
В МЕДИЦИНЕ

Перевод с английского
И. М. БЫХОВСКОЙ

Под редакцией
ПРОФ. М. Л. БЫХОВСКОГО

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1971

12.014

128 УДК 21.531

103083

БИБЛИОТЕКА
Государственного
медицинского института
гор. Аялжап

В книге с позиций кибернетики рассмотрены три важных этапа деятельности врача: 1) принятие решения в процессе диагностики, 2) выбор наиболее подходящего метода лечения и 3) выбор наиболее подходящего метода оценки данных, полученных в результате исследования больного. Центральное место отводится проблеме эффективного взаимодействия врача и вычислительной машины. Этот вопрос, представляющий большой интерес, рассматривается впервые и иллюстрируется обширным фактическим материалом. Заключительные главы посвящены оценке будущего развития вычислительных методов в приложении к медицине, особенно к проблеме «врач — машина»; подробно описана вычислительная система для диагностики заболеваний щитовидной железы. В приложениях приведены соответствующие программы для цифровой вычислительной машины.

Книга предназначена для врачей, биологов, специалистов по вычислительной технике, работающих в области биологической и медицинской кибернетики.

ADTI
AXB-RESURS MARKAZI
INV № 7160

Редакция литературы по новой технике

Инд. 2-10-2; 3-3-14
144-71

ПРЕДИСЛОВИЕ К РУССКОМУ ИЗДАНИЮ

На протяжении последнего десятилетия мы являемся свидетелями все более глубокого проникновения математических методов и электронных вычислительных машин в биологию и медицину.

Известно, что появление электронных вычислительных машин привело к необычайно быстрому развитию вычислительных методов. Многие задачи естествознания, при решении которых математические методы раньше не применялись вследствие огромных вычислительных трудностей, ныне стали доступными для математического анализа и точного решения. С другой стороны, появление вычислительных машин и усилия исследователей в направлении использования их в различных областях науки и практики привели к созданию новых математических методов, существенно расширивших области применения математики.

Этот процесс представляет собой в сущности то, что мы называем в последнее время математизацией наук. Особенно важное значение процесс математизации имеет для биологии и медицины, которые до последнего времени были в большой степени «качественными» науками. Использование количественных методов важно здесь не только с точки зрения возможности получения более точных решений; оно является (и это не менее существенно) одним из средств теоретического развития этих наук.

Большое внимание уделяют исследователи использованию в биологии и медицине методов кибернетики. Это естественно, так как всякий живой организм фактически представляет собой весьма совершенную саморегулирующуюся систему, т. е. относится к системам такого типа, для исследования которых и разрабатывались методы теории автоматического управления. Конечно, живой организм несоизмеримо сложнее любой технической системы, и поэтому только при очень большом упрощении можно для некоторых живых систем непосредственно использовать развитые ранее методы этой теории. В подавляющем же большинстве случаев мы сталкиваемся с необходимостью значительного усовершенствования этих методов. Другими словами, эта область представляет собой необычайно благодатное поле деятельности для специалиста по автоматическому управлению.

Всякая система управления осуществляет некоторую последовательность двух основных процессов: определяет состояние системы и вырабатывает для данного состояния оптимальное воздействие (управление). С другой стороны, центральная проблема медици-

ны — проблема лечения — также включает в себя два аналогичных основных процесса: определение состояния больного и назначение наиболее эффективного воздействия на больной организм (план лечения). Естественно, что именно эти две задачи и составляют главное содержание того нового направления, которое сейчас называют медицинской кибернетикой.

В свою очередь медицинская кибернетика развивается в настоящее время в виде двух ветвей, тесно соприкасающихся между собой. Первая ветвь — мы ее условно назовем «внешней кибернетикой» живого — имеет своей целью разработку теории и методов внешнего управления живым организмом. Например, как мы упоминали выше, задачу диагностики и лечебной тактики.

Вторая ветвь — назовем ее «внутренней кибернетикой» живого — ставит своей целью изучение внутренних систем саморегуляции и управления организма. Большое значение имеет здесь использование электронных вычислительных машин для моделирования процессов, протекающих в живом организме.

Касаясь проблемы диагностики, следует заметить, что в настоящее время значительно возросли требования к точности диагноза, к его достоверности. Удовлетворить их старыми, классическими способами становится все труднее. Разработка новых методов исследования больного, конечно, очень важна, но не решает проблемы, так как остается открытым вопрос, как из имеющейся информации о состоянии больного сделать верный диагностический вывод. В сердечной хирургии, например, требования к точности диагноза особенно высоки. Известно, что летальность среди больных, которые были направлены на операцию с неверным диагнозом, в несколько раз выше, чем среди больных с правильным диагнозом. Вообще нужно сказать, что последнее десятилетие ознаменовалось значительными успехами в хирургии: усовершенствовалась техника операций, расширились границы вмешательства и т. д. Тем не менее, например, число радикально оперированных раковых больных до сегодняшнего времени составляет лишь 25—30% числа первично диагностируемых случаев. Причина этих низких показателей состоит не только в запоздалом обращении больных в лечебные учреждения, но в не меньшей степени в ошибках и промедлении диагностики заболевания, допускаемых врачами. Полученные в настоящее время результаты позволяют прийти к заключению, что использование электронных вычислительных машин приведет к ускорению процесса диагностики и повышению точности предоперационного диагноза.

Следует отметить, что с проблемой управления мы встречаемся уже на этапе диагностики. Действительно, при организации диагностического процесса не следует требовать, чтобы у больного были проведены все или даже главные методы исследования, применяемые для диагностики данного класса заболеваний. Может оказаться, что некоторые из них сами по себе не безразличны, а иногда и опас-

ны для организма больного. Поэтому диагностический процесс, реализуемый с помощью электронной вычислительной машины, следует строить как некую динамическую процедуру, которая начинается с проведения простых исследований. Затем на основании полученных данных проводится диагностическая оценка. Если оказывается, что этих данных достаточно, то система ставит окончательный диагноз. Если же этих данных недостаточно для постановки диагноза с заданной степенью точности, то система указывает, какое следующее исследование (из группы более тяжелых) нужно провести для того, чтобы в данной конкретной ситуации получить максимальную информацию, и т. д. Таким образом, система осуществляет не только постановку диагноза, но и управление сбором информации, т. е. собственно процессом диагностики.

Во всех упомянутых основных задачах — сборе информации о больном, диагностике, выборе тактики лечения — мы встречаемся с одной общей проблемой — проблемой принятия решения. Именно этому важному вопросу и посвящена предлагаемая читателю книга американского ученого Л. Б. Ластеда, известного своими работами в области медицинской кибернетики. Среди имеющейся в настоящее время уже довольно многочисленной литературы по применению математических методов и электронных вычислительных машин в медицине данная книга занимает особое место. Может быть, именно потому, что книга написана высококвалифицированным специалистом в области медицинской радиологии, хорошим врачом, и потому, что предназначалась она для врачей, в ней есть нечто такое, чего не найдешь ни в какой другой книге по медицинской кибернетике. Это «нечто» — раскрытие самой сущности проблемы решения в медицинском аспекте, в таком виде, как она возникает перед врачом, когда он встречается с больным и на его плечи ложится вся ответственность за состояние больного и его лечение. Именно поэтому содержание книги глубоко интересует не только врачей, и даже не столько врачей, сколько специалистов по автоматическому управлению, вычислительной технике, математиков, работающих в области медицинской кибернетики.

С другой стороны, здесь в значительно меньшей степени представлены математические результаты и технические решения, достигнутые в этой области, и особенно в работах советских ученых. Это, конечно, является недостатком книги.

При изложении проблемы решения автор рассматривает три важные области деятельности врача: принятие решения в процессе диагностики; выбор наиболее подходящего метода лечения для данного больного; оценка данных, полученных в результате проведенных исследований (в частности, в области рентгенологии) с целью максимально возможного исключения субъективных ошибок. Рассмотрен и ряд практических систем, в том числе системы для диагностики заболеваний щитовидной железы, первичных опухолей

костей, врожденных пороков сердца. При этом постоянное внимание уделяется проблеме ошибок врача при оценке результатов обследования больного. Кроме того, диагностическим ошибкам посвящена целая глава. Очень интересен пример эффективного взаимодействия врача и вычислительной машины в виде предлагаемой системы вероятностной обработки медицинской информации (гл. 5).

В целом книга, несомненно, будет встречена с интересом советским читателем.

При переводе мы сочли уместным сделать некоторые (впрочем, совсем небольшие) сокращения и исправить некоторые опечатки в формулах и тексте.

М. Быховский

ПРЕДИСЛОВИЕ

Начинайте с убеждением, что абсолютная истина трудно достижима в вопросах, относящихся к нашим собратьям, здоровым или больным, что ошибки в наблюдении не исключаются даже при самых лучших способностях и знаниях, что неправильные суждения неизбежны в практике врачебного искусства, которое в большой степени состоит в балансировании вероятностями.

Сэр Вильям Ослер¹⁾

Медицинская диагностика как логический процесс ставит две основные проблемы. Первая состоит в том, что мы не знаем всех заболеваний, которые дают одни и те же или сходные признаки и симптомы. Если бы мы их знали, то можно было бы проследить весь процесс от причины заболевания до его проявления, выраженного в симптомах и признаках. Эта неопределенность связей между признаками, симптомами и заболеваниями придает медицинской диагностике вероятностный характер. Продолжающиеся исследования основных биологических процессов и все более глубокое понимание признаков и симптомов заболеваний помогут решить эту первую проблему. Но по мере накопления знаний, связанных с решением первой проблемы, становится все более серьезной вторая проблема. Уже сейчас мы имеем дело с таким количеством заболеваний и сопутствующих признаков, симптомов и лабораторных данных, что врач испытывает затруднения при необходимости вспомнить всю нужную диагностическую информацию. Многие специалисты считают, что ошибки в дифференциальной диагностике происходят чаще из-за ошибок пропуска, чем по другим причинам.

Одним из результатов ускорения медицинских исследований за последние пятнадцать лет является большой объем новой информации. В настоящее время мы являемся свидетелями борьбы за такую организацию медицинских знаний, чтобы они стали более полезными для практической деятельности врачей. Под организацией медицинских знаний мы понимаем отбор, систематизацию и представление необходимых врачу данных в таком виде, чтобы он смог принять более авторитетное решение о возможной последовательности своих действий.

Очень много усилий уже затрачено на разработку методов, которые сделают медицинскую литературу более доступной и удобной для пользования. Примечательным примером этого является система MEDLARS Национальной медицинской библиотеки, основанная

¹⁾ Teacher and Student, in: Aequanimitas with Other Addresses, 2nd ed., Philadelphia, Blakiston, 1930, p. 38.

на использовании электронной вычислительной машины. Эта разработка заслуживает похвалы за ту активную деятельность, которую она стимулировала в области библиотечного дела. Возможно, все же наряду с машинной обработкой медицинской литературы следует решить проблему отбора, систематизации и представления врачу данных, в которых он заинтересован. Кто должен делать это?

Начиная отбор медицинской информации, нужно ясно понять, что лечащие врачи и теоретики должны постоянно прилагать усилия для определения того, что является существенным для теории и что является существенным для клиники. Решение этих вопросов требует опыта, квалифицированного суждения и, конечно, обработки литературных материалов. Машинный поиск литературы и скопированные машиной библиографические списки, по-видимому, могут сделать их решение посильным для врача. Для каждой области медицины нужно иметь возможность отобрать наиболее важные результаты. Врач не в состоянии беспредельно добавлять новые сведения к своему «умственному багажу» и в то же время надлежащим образом использовать эти сведения. Методы систематизации и представления информации лечащим врачам также очень важны и требуют рассмотрения многих возможных носителей информации — от справочников с таблицами вероятностей типа «заболевания — симптомы» и до систем вычислительных машин.

Конечно, хотелось бы иметь большую информацию о течении заболеваний и относительно больного организма в целом, но, к сожалению, врач слишком часто оказывается в положении, когда ему нужна была бы более полная информация о данном пациенте, о возможных диагнозах и об эффективности различных планов лечения. В такой ситуации врач как человек, принимающий решение, неизбежно сталкивается с неопределенностью. Он должен принимать решения о диагнозе и лечении. Чтобы сделать это, он часто вынужден пользоваться информацией, лишь косвенно связанной с тем, что он в действительности хочет знать. Мы приходим к заключению, что предоставление все большего и большего количества информации врачу, принимающему решение, само по себе еще не дает ответа на поставленную задачу и что врачу необходимы некоторые дополнительные средства, помогающие ему справиться с неопределенностью. Поиск методов и технических устройств для реализации этих средств приводит к исследованию вопроса о том, каким образом люди обрабатывают информацию и принимают решения; это весьма глубокий вопрос, и ответ на него имеет значение для многих областей человеческой деятельности. В этой книге особое внимание обращено на те исследования, которые связаны с указанными проблемами. Эти работы имеют своим источником такие области науки, как психология, психофизика и статистика; они играют существенную роль при разработке методов, которые помогут врачу принимать решение относительно диагноза и лечения.

Уместно будет задать вопрос: почему не существует хорошо определенной технологии постановки диагноза в медицине или в областях военного дела, коммерции, права или государственного управления? Эдвардс [19] предлагает возможный ответ, заключающийся в том, что наука до недавнего времени находилась под влиянием ложного представления, что неопределенность не может быть измерена количественно. Однако, как отмечает далее Эдвардс, тот недавно обнаруженный факт, что неопределенность может быть измерена некоторым способом, зависящим от человеческого суждения, позволил приступить к разработке технологии диагностики. Подход, лежащий в основе этой технологии и базирующийся на персоналистском взгляде на вероятность, стали называть байесовским подходом. В этой книге исследование медицинской диагностики проводится именно с такой точки зрения.

Разработка систем, предназначенных для диагностики и выбора действия, интересна и важна применительно ко многим областям, так как такие системы позволяют сфокусировать внимание ученых для планомерного исследования вопроса о том, как принимать решения в реальном мире. *Диагноз и выбор действия* — это термины, используемые в исследованиях процесса принятия решения в различных областях деятельности человека, и они эквивалентны терминам *диагностика и лечение* в медицине. Процессы принятия решения о диагнозе и о выборе лечения должны рассматриваться совместно, так как диагноз сам по себе не очень нужен. Диагностика — это также и принятие решения о выборе действия, и есть смысл проводить диагностические процедуры только в том случае, если, по крайней мере при каких-то определенных обстоятельствах, врач предполагает предпринять некоторые действия, и ему необходим диагноз для того, чтобы решить, что именно делать. Другой аргумент в пользу изучения связи между лечением и диагностикой заключается в том, что реакцию пациента на лечение можно рассматривать как диагностический тест, результат которого может изменить мнение врача о диагнозе, о дальнейшем лечении или о том и о другом.

Для выбора действия требуется проведение большого количества исследований. К счастью, исследования в области выбора лечения могут быть в значительной степени отделены от исследований в области диагностики. Поскольку до сих пор проведено очень мало исследований по выбору лечения, мы ниже укажем на некоторые проблемные вопросы и продемонстрируем тесную связь между выходными данными байесовской диагностической системы и методом выбора действия, основанным на принципе максимизации ожидаемой величины.

Все исследования, которые мы отобрали для изложения в этой книге, имеют своей главной целью более эффективное использование медицинских данных в процессе принятия решения врачом.

Для некоторых исследователей, чьи работы мы цитируем, эта цель, может быть, не является основной, но, я надеюсь, они согласятся, что широкая интерпретация их работ допускает такую точку зрения. Излагаемые исследования сгруппированы в три общих предметных раздела: 1) вероятностный вывод, 2) диагностический процесс и деревья решений и 3) процессы решения и ошибки наблюдателя. Границы этих трех областей пересекаются, и, поскольку они представляются тесно связанными, мы излагаем их под общим заглавием процедуры принятия решения в диагностике и лечении.

Указанные направления исследований сами по себе не являются новыми. Действительно, теорема Байеса, которую мы обсуждаем в главе, посвященной вероятностному выводу, известна с 1763 г. Тем не менее исследователи вновь заинтересовались некоторыми из этих известных областей, так как наличие электронных вычислительных машин и автоматических методов обработки информации вдохновило ученых на поиски ответов на некоторые старые и трудные вопросы.

В этой книге постоянно присутствуют два основных лица — пациент и врач. Хотя излагаемые методология и технология были разработаны для помощи врачу в процессе принятия решения о диагнозе и лечении, читатель не должен забывать, что именно забота о пациенте является стимулом для любой такой работы. Итак, не забывайте пациента!

Основная идея книги заключается в следующем: эффективное принятие медицинского решения требует, чтобы мы регулярно, через определенные интервалы времени пересматривали и перестраивали медицинскую информацию. Однако получение все большего количества информации само по себе не является достаточным для того, чтобы снять неопределенность, в условиях которой часто должен действовать врач, когда он сталкивается с конкретным пациентом. Поэтому необходимо искать методы и развивать технологию, которые помогут врачу использовать любые доступные данные наиболее эффективным образом для принятия решения о диагнозе и составления плана лечения.

Цель нашей книги — показать, что начала методологии и технологии диагностики и лечения уже заложены экспериментальными работами; эта книга также выдвигает для будущих исследований некоторые проблемные вопросы, ответы на которые смогли бы привести к практическому симбиозу человека и машины в процессе принятия решения при диагностике и выборе лечения.

Многие данные, подтверждающие сказанное, могут оказаться новыми для читателей-медиков. До сих пор они не были собраны в одном месте, и некоторые из относящихся к делу материалов могут показаться несвязанными. Имея это в виду, мы систематизировали основные положения так, как мы их себе представляем, и попробовали указать некоторые из связанных с ними вопросов:

1. Врач располагает все увеличивающимся объемом медицинской информации, и тем не менее при диагностике и выборе лечения он постоянно встречается с ситуациями, в которых эта информация оказывается неполной и только косвенно связанной с тем, что ему в действительности нужно знать о пациенте.

2. В таких ситуациях врач вынужден принимать решения о диагнозе и лечении в условиях неопределенности.

3. Предоставление все большего количества информации врачу не является само по себе средством для уменьшения неопределенности, хотя дополнительная информация всегда полезна. Если это так, то *каким же образом можно помочь врачу справиться с неопределенностью?*

Для получения ответов на эти вопросы проводятся исследования по двум направлениям: а) проекты, предусматривающие разработку технических средств и устройств, основанных на использовании вычислительных машин в помощь врачу при принятии решения о диагнозе и выборе лечения, и б) проекты, которые предусматривают изучение самого врача в роли человека, принимающего решение.

4. Исследовательские проекты, связанные с применением вычислительных машин в помощь врачу при диагностике и выборе лечения в свою очередь характеризуются двумя способами использования этих машин: а) анализ медицинских данных с целью выявления связи между признаками, симптомами и диагнозами и б) выполнение необходимых вычислений для обработки информации в системе человек — машина.

4. 1. С помощью вычислительных машин были проанализированы медицинские данные с точки зрения их диагностической ценности, т. е. была сделана попытка определить, какие признаки и симптомы имеют наибольший вес для определенных диагнозов. Были испробованы теорема Байеса, отношение правдоподобия и дискриминантные функции. Из всех этих методов наиболее широко использовали теорему Байеса. Попытки сгруппировать связанные признаки, симптомы и заболевания основывались на использовании известных методов группирования, машинных процедур сравнения, и численной таксонометрии; каждый из этих методов позволяет достичь какого-то успеха, тем не менее все они нуждаются в дальнейшей проверке. Если суть медицинского суждения заключается в процессе сравнения, то группирование сходных признаков и симптомов может уменьшить число отдельных элементов информации, которые должны рассматриваться врачом в процессе сравнения. Важное значение имеет разработка методов уменьшения необходимого количества признаков (симптомов).

4. 2. Человек является консервативным устройством обработки информации. Он не способен выделить из обрабатываемых данных всю ту степень определенности, которая в них содержится в скрытой форме.

Практическую систему обработки информации, которая подходящим образом ассоциируется с человеческим консерватизмом, можно было бы построить таким образом, чтобы вычислительные машины собирали и обрабатывали оценки правдоподобия, придаваемые человеком каждому наблюдаемому симптому и признаку, если у пациента имеется любое из заболеваний, которое приводит к появлению этих симптомов и признаков. Такие оценки можно выражать в виде чисел различными способами, в зависимости от конкретной ситуации. Когда такие оценки получены, их можно преобразовать в апостериорные распределения вероятностей по всему списку рассматриваемых заболеваний, используя теорему Байеса. Такие процедуры позволяют извлекать информацию более эффективно, чем любые другие методы, предусматривающие использование человеческого суждения, которые когда-либо применялись. Более того, они позволяют получать данные, в первом приближении сравнимые с результатами теоретически оптимальных вычислений в случаях, когда такие вычисления возможны.

Описаны эксперименты с практической системой обработки информации, названной устройством вероятностной обработки информации PIP (Probabilistic Information Processor). Система PIP до настоящего времени не использовалась в экспериментах, связанных с принятием медицинского решения, но такие исследования должны быть проведены. Системы такого типа легче оценивать при их использовании в медицинской диагностике, чем во многих других реальных ситуациях принятия решения, потому что при диагностике точность работы системы можно сравнивать с объективно правильной и доказанным медицинским диагнозом.

Приемлемость выходных данных системы PIP оценивается человеком (в области медицинской диагностики и лечения — врачом). Существует тесная связь между выходными данными системы PIP и входными данными, необходимыми для выбора лечения в соответствии с принципом максимизации ожидаемой величины.

Практические системы обработки информации, такие, как PIP, могут быть использованы врачами для обслуживания пациентов в будущем, после: а) разработки более полных алгоритмов диагностического процесса и б) создания удобных входных и выходных устройств и обеспечения экономически выгодного для врача машинного обслуживания.

5. Приблизиться к созданию алгоритмов (программ для вычислительных машин), моделирующих последовательность действий врача, когда он ставит диагноз, можно путем изучения самого врача в роли человека, принимающего решение. Одна из концепций, на которых базируются исследования в этой области, заключается в том, что человеческое поведение может быть смоделировано машинной программой; более того, некоторые исследователи полагают, что при использовании информационного подхода можно получить

более адекватное понимание процессов. Деревья решений, сети решений и блок-схемы диагностического процесса разработаны на основе именно этих исследований.

6. Исследования по машинным средствам диагностики (см. п. 4) и исследования диагностического процесса, осуществляемого врачом (см. п. 5), тесно связаны и фактически могут дополнять друг друга. Мы уже начинаем понимать, каким образом это происходит (в гл. 3 описаны некоторые точки соприкосновения этих двух областей).

Следует сделать некоторые замечания относительно материала, который не обсуждается в этой книге, но может рассматриваться как составная часть процесса принятия медицинского решения. Существует большая литература по вопросам машинного анализа электрокардиограмм и автоматизации клинических лабораторных исследований. Рассмотрение этих областей уведет нас слишком далеко в сторону от основной темы книги, хотя упомянутые вопросы, несомненно, важны. Здесь мы не обсуждаем также такие вопросы, как точность диагностических процедур, последовательный сбор данных и процессы последовательного анализа, несмотря на то что в течение длительного времени эти темы пользуются заслуженным вниманием в медицинских исследованиях. Имеется много прекрасных книг по этим вопросам, и мы не хотим их хотя бы в какой-то мере недооценить. На протяжении всей книги, где только возможно, мы излагаем идейную сторону и анализ, основанный на теории байесовского статистического вывода. В некоторых местах приведены замечания, сопоставляющие байесовские и традиционные статистические методы.

Байесовские статистические методы, как они изложены в гл. 1, представляют собой отличный от общепринятого подход к теории статистического вывода; для него можно сформулировать три основные идеи: 1) вероятности являются упорядоченными мнениями; 2) статистика (или любой вид обработки информации) связана с пересмотром мнений в свете новой информации и 3) теорема Байеса из теории вероятностей является оптимальным формальным правилом, которое указывает, как именно должен быть сделан такой пересмотр [21]. Эти идеи доказали свою полезность в исследованиях по медицинской диагностике как для теории, так и для приложений. В дополнение к этому, принцип правдоподобия, специально выделенный в байесовских статистических методах, устанавливает, что правила, определяющие окончание сбора входных данных, не зависят от интерпретации данных. Применение этого принципа имеет важное значение для медицинской диагностики, так как выбор наиболее эффективных диагностических тестов из большого числа возможных может оказаться существенным фактором с точки зрения экономических аспектов медицинского обслуживания и обследования состояния здоровья населения.

В гл. 2 излагаются некоторые исследования и приложения байесовских статистических методов в медицинской диагностике, опубликованные в течение приблизительно семи последних лет. Примеры, обсужденные в гл. 2, снабжены библиографическим списком (приложение 1), в котором материалы сгруппированы по областям медицины (например, исследования по врожденным порокам сердца, по функции щитовидной железы), а также по применяемой методике анализа (например, исследования по условно-вероятностным методам, по процедурам сравнения). Исследования, в которых используется дискриминантная функция и факторный анализ, включены в приложение 1, чтобы придать законченный характер библиографии по электронно-вычислительным методам диагностики. Для тех исследователей, которые пожелают начать изучение медицинской диагностики с применением байесовской теории статистического вывода, в приложении 3 представлена полная машинная программа для диагностики заболеваний щитовидной железы. Эта программа была разработана М. Вильямсом и Л. Т. Фитцджералдом в медицинском колледже Университета Флориды и используется в настоящее время как в этом, так и в некоторых других учреждениях.

Клиницист в роли человека, принимающего решение, стал объектом изучения врачей, заинтересованных в улучшении методов обучения студентов-медиков, и психологов, занимающихся проблемой решения и суждения. Последние исследования в этой области изложены в гл. 3. Для многих врачей более привлекательны деревья решения и блок-схемы, предназначенные для объяснения диагностического процесса, чем байесовские статистические методы первых двух глав; они могут ознакомиться с гл. 3, прежде чем ломать себе голову над гл. 1 и 2.

Связь исследований диагностического процесса с исследованиями по использованию байесовского статистического вывода становится ясной, если поставить вопрос, действует ли врач в процессе нахождения диагноза интуитивно по байесовскому методу. Если теорему Байеса, как она сформулирована в теории вероятности, принять за оптимальное формальное правило для пересмотра мнения в свете новой информации (а психологические исследования показали, что человека можно обучить действовать байесовским методом в специфических ситуациях), то будет иметь смысл исследование методов, с помощью которых врачи пользуются или могли бы пользоваться байесовской процедурой обработки информации как частью диагностического процесса. В результате открывается возможность разработки улучшенных стратегий диагностического процесса.

Гл. 4 начинается с обсуждения проблемы ошибок наблюдателя — предмета, хорошо знакомого читателям-медикам. Примеры с конкретными данными об ошибках наблюдателя были выбраны из

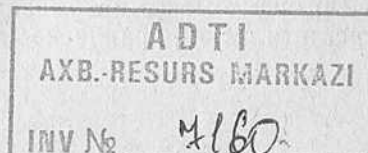
исследований по точности интерпретации снимков грудной клетки и мазков, окрашенных по методу Папаниколау. Точность интерпретации снимков грудной клетки была изучена более подробно, чем любой другой медицинский диагностический тест, и поэтому мы использовали именно этот пример в качестве медицинского объекта для обсуждения вопросов, связанных с проверкой гипотез, обнаружением сигналов и процессом решения.

В одной недавней работе по восприятию человеком звуковых и визуальных сигналов показано, что можно отделить факторы, которые влияют на отношение данного наблюдателя к задаче, от факторов, которые влияют на его чувствительность. Если наблюдателя просят сказать в условиях теста либо «да, сигнал был», либо «нет, сигнала не было», то его реакция в терминах правильных ответов или ложно положительных ответов (эквивалентных ложно положительным заключениям в медицинской диагностике) может быть представлена в виде рабочей характеристической кривой, или кривой ROC (Receiver Operating Characteristic Curve).

Рабочие характеристические кривые для тестов, связанных с обнаружением сигнала, сходны по виду с кривыми, которые мы можем получить путем преобразования обычных рабочих характеристических кривых для ложно положительных ответов в функции ложно отрицательных ответов при интерпретации снимков грудной клетки.

Преобразование такой кривой в характеристическую кривую истинно положительных ответов в функции ложно положительных ответов осуществляется при помощи матрицы решения.

Таким образом, рабочая характеристическая кривая имеет в медицинской диагностике, по-видимому, тот же смысл, что и в исследованиях по обнаружению сигналов. Отсюда следует, что мы обладаем методом для изучения диагностических критериев как самостоятельной и независимой меры, не связывая ее с врачебной интерпретацией критериев. Психофизические исследования в области обнаружения сигналов показали, что можно выделить две независимые меры: меру чувствительности наблюдателя и меру его критерия реакции. Последняя мера, оценивающая реакцию наблюдателя, т. е. интерпретацию врачом диагностических критериев, выражается через известное уже нам отношение правдоподобия, с которым мы впервые встречаемся в гл. 1 и 2. Таким образом, мы можем установить связь между проверкой гипотез и процессом принятия решения при интерпретации снимков грудной клетки и байесовским подходом к диагностике, изложенным в гл. 1 и 2. Важные положения, касающиеся использования логарифмических шансов, логарифмических отношений правдоподобия и матриц решения, проиллюстрированы данными, взятыми из исследования по интерпретации снимков грудной клетки.



Проблеме ошибок наблюдателя, которую мы излагаем в гл. 4, посвящено больше опубликованных работ, чем любому другому вопросу, рассмотренному в этой книге. В результате просмотра большого количества публикаций мы составили избранную библиографию, которая представлена в приложении 2. Стремление к более эффективному использованию медицинских данных пробудило новый интерес к исследованию ошибок наблюдателя, ибо, как указал Н. Винер [129], «для полуточных наук существует грамматика, значительно отличающаяся от грамматики, которая подходит для точных наук. Ошибки наблюдения являются частью самого наблюдения, и любое разделение между точными значениями и ошибками в случае полуточных наук приведет к методологии, которая в действительности менее точна в отношении результатов, которые мы хотели бы получить, чем более широкий, хотя и менее точный, метод». Я думаю, эти слова Винера в области медицины означают следующее: ошибки наблюдателя являются в настоящее время составной частью медицинской диагностики и могут рассматриваться как часть некоторого вероятностного процесса, количественно определяющего нашу неопределенность в отношении диагноза.

В гл. 5 продолжено обсуждение процессов решения и описано исследование, которое иллюстрирует возможные разногласия по поводу правил решения в диагностике. Этот материал основан на анализе разногласий среди врачей относительно предоперационного диагноза и последующего лечения одиночного узла легкого. Мы показываем, что матрицы решения и рабочие характеристические кривые должны определяться как составная часть исследований по выработке диагностических критериев и что использование матриц решений и рабочих характеристических кривых поможет избежать неверных правил решения.

В результате процесса решения должны быть получены диагнозы, которые можно использовать в качестве наиболее подходящих исходных данных для выбора оптимального лечения. Концепция *максимальной ожидаемой величины* и связанная с ней концепция *минимальных ожидаемых потерь* представляются важными при выборе лечения, но пока еще мы не знаем таких медицинских исследований, которые могут быть приведены для демонстрации применения этих концепций.

В заключительной части гл. 5 мы приводим некоторые рассуждения, которые, я надеюсь, являются «разумными предположениями» относительно будущих исследований, связанных с разработкой диагностических критериев и выбором оптимального лечения. Здесь мы рассматриваем также вопросы о том, какую помощь в диагностике можно ожидать от математических машин при использовании системы RIP.

Во время моего пребывания в клиниках Калифорнийского университета в Сан-Франциско (в связи с изучением проблем радиоло-

гии) я принимал участие в исследовании по определению надежности данных рентгенограмм грудной клетки при диагностике новообразований в легких. Проект, которым руководили Дж. Йерашелми и Л. Г. Гарланд, впервые заставил меня обратить внимание на проблему ошибок наблюдателя. Позднее, после перехода в Национальные институты здравоохранения, я совместно с Робертом С. Ледли начал серию исследований, которые завершились публикацией некоторых идей [53], касающихся группы математических методов, которые, как нам казалось, будут полезны при исследовании вопросов медицинской диагностики. Были изучены процедуры логического исключения и условно-вероятностный метод, использующий теорему Байеса; эти процедуры, как показано в нашей книге, тесно связаны между собой с точки зрения байесовской стратегии.

Недавняя работа группы исследователей помогла выяснить некоторые проблемы применения теории байесовских статистических выводов в медицинской диагностике. Особенно полезные предложения были сделаны Г. Р. Уорнером, Г. С. Лодвиком и К. М. Вильямсом.

Мне очень приятно выразить признательность профессору Б. де Финетти за проявленный интерес к этой работе и полезные замечания, и это дает мне возможность вспомнить случай, имевший место в недалеком прошлом. Профессор де Финетти много сделал для философского и математического обоснования байесовских статистических методов, введя в 30-х годах идею субъективного определения вероятности. Случай, о котором мне хочется рассказать, произошел на XI Международном конгрессе по радиологии в Риме летом 1965 г. Мы с д-ром Лодвиком устроили выставку, иллюстрирующую применение теоремы Байеса в рентгенографической диагностике костных опухолей. Выставочный стенд включал небольшую цифровую вычислительную машину, с помощью которой рентгенологи могли проверить свою диагностическую проницательность. Профессор де Финетти посетил выставку; мы были рады возможности продемонстрировать практическое применение в медицинской диагностике тех идей, которые профессор де Финетти предложил около тридцати лет назад.

Т. Г. Бердсолл и У. У. Питерсон из отделения электротехники Мичиганского университета предложили вариант теории обнаружения сигналов, который инициировал развитие психофизических приложений, представленных в гл. 4. Я обязан им и У. П. Тэннеру младшему, в чьей лаборатории в Мичиганском университете были реализованы многие психофизические приложения этой теории. Я также выражаю признательность многим исследователям, которые продолжают развивать эту интересную область.

В нескольких важных аспектах сделал вклад в эту книгу У. Эдвардс. Он вскрыл явление консерватизма при обработке ин-

формации человеком, а его оригинальное предложение — система вероятностной обработки информации (РІР) — представляет основной исходный материал соответствующей главы. Кроме того, были исключительно полезными его замечания и обсуждение некоторых вопросов еще в рукописных набросках. Наконец, я очень благодарен своему секретарю Ш. А. Найден за тяжелый труд, терпение и внимание при работе над рукописью.

Л. Ластед

Портленд, шт. Орегон

БАЙЕСОВСКАЯ ТЕОРИЯ СТАТИСТИЧЕСКОГО ВЫВОДА И ЕЕ ПРИМЕНЕНИЕ В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ

*Медицина — это наука неопределенности
и искусство вероятности¹⁾.*

Одной из основных причин этой неопределенности является все увеличивающееся разнообразие проявлений одного и того же заболевания.

Сэр Вильям Ослер²⁾

1.1. ВВЕДЕНИЕ

Выдающийся английский врач сэр Джордж В. Пикеринг писал в 1960 г.: «Диагноз является вопросом вероятности, и это слишком хорошо знают те из нас, кто проследил судьбу своих пациентов вплоть до морга. Прогноз — это также вопрос вероятности, и, рассматривая возможные методы лечения, мы должны основывать наше решение на знании вероятностей». Хотя многие клиницисты готовы были согласиться с этим утверждением, не было ясно, как можно использовать вероятность для постановки медицинского диагноза. Эта трудность, возможно, возникла из-за классического понятия вероятности как относительной частоты или как гипотетического ее предела. Для многих врачей казалось бессмысленным применение понятия относительной частоты к конкретному пациенту.

Некоторые авторы делают различие между разными типами вероятностей [36]. Мы предпочитаем начать с понятия вероятности, развитого на основе представления о степени уверенности, и вывести из него вероятность как предел частоты. В этой книге под вероятностью понимается ее субъективная мера. Такое определение вероятности делает это понятие вопросом мнения и, таким образом, пригодно в качестве врачебной оценки относительно некоторого конкретного состояния пациента. Субъективная мера вероятности используется врачами для анализа дискретного набора симптомов данного пациента. Вероятности служат для количественного выражения неопределенности, а субъективные меры вероятности позволяют выразить степень уверенности относительно данного признака, симптома или диагноза. Субъективная мера вероятности дает возможность клиницисту выразить, насколько он уверен в клиническом

¹⁾ Bean W. B. (Ed.), *Aphorisms from His Bedside Teaching and Writings*, New York, Henry Schuman, 1950, p. 125.

²⁾ *Teaching and thinking*, in: *Aequanimitas with Other Adresses*, 2nd ed., Philadelphia, Blakiston, 1930, p. 123.

заключении, в терминах относительных шансов или отношений правдоподобия (об этом мы будем говорить в дальнейшем). Относительные шансы и отношения правдоподобия выражаются числами.

Все вероятности можно рассматривать как условные. Это означает, что они выражают, насколько вы уверены относительно такого события, как симптом или диагноз, и что степень уверенности обусловлена всем тем, что вы уже знаете относительно этого события. Субъективная мера условных вероятностей является вероятностным выражением знания, приобретенного из опыта. Одно из полезных математических построений, известное как формула Байеса, позволяет при помощи субъективной меры условных вероятностей совместно использовать наблюдаемые данные и известную ранее информацию в целях решения задачи дифференциальной диагностики. Субъективная мера вероятности, субъективная мера условной вероятности и формула Байеса будут рассмотрены более детально при обсуждении байесовской стратегии принятия решения.

Существует несколько точек зрения на использование байесовской стратегии, и применимость этого подхода для статистического вывода в настоящее время является спорной. Мы занимаем байесовскую и субъективно-вероятностную позицию, основанную на трех идеях: 1) вероятности — это упорядоченные мнения, 2) статистика (или любой способ обработки информации) связана с пересмотром мнения в свете новой информации и 3) формула Байеса из теории вероятностей является формально оптимальным правилом осуществления пересмотра мнения. Хороших учебников по методам Байеса нет, а имеющаяся литература по этому вопросу представляет собой разрозненный материал. Тем не менее читатель может найти подробное математическое изложение теоремы Байеса и прекрасные литературные ссылки в статьях Корнфилда [11—13] и Эдвардса [18, 21]. Мы особенно признательны Эдвардсу за материал, используемый нами при дальнейшем изложении. Полное математическое рассмотрение стратегии Байеса приведено в работе [89].

1.2. СУБЪЕКТИВНАЯ МЕРА ВЕРОЯТНОСТИ

Субъективная мера вероятности — это число, заключенное между 0 и 1 и выражающее степень уверенности некоего идеализированного субъекта в истинности данного утверждения. Смысл идеализации субъекта заключается в следующем: назначаемая им вероятность того, что имеет место одно из двух взаимоисключающих событий, должна быть равна сумме назначаемых им вероятностей для каждого из этих событий в отдельности. Такое свойство аддитивности является существенной частью теории вероятностей, но для людей оказывается слишком трудным делом постоянно вести себя в соответствии с этим свойством. Практическое определение

субъективной меры вероятности может быть дано следующим образом: вероятность события A для вас равна сумме денег, которую вы готовы заплатить мне сейчас в обмен на мое обещание заплатить вам один доллар, если событие A произойдет. В теории полезности деньги заменяются мерой полезности. Важная мысль, лежащая в основе этого определения вероятности, заключается в том, что вероятность — это упорядоченное мнение.

В этой книге все вероятности рассматриваются как субъективные меры. Это не исключает того, что субъективные меры вероятности могут быть частотами, отношениями числа благоприятных к общему числу возможностей или числами, полученными в результате любого другого вычислительного процесса, который представляется подходящим для выработки суждения в конкретной ситуации. Важно иметь ясное понимание различия между идеализированными в указанном выше смысле субъективными мерами вероятностей и обычными субъективными вероятностями, которые, как правило, являются результатом выбора, сделанного человеком. Термины «субъективная мера вероятности» и «субъективная вероятность» служат для того, чтобы подчеркнуть это различие; в этом вопросе мы следуем Эдвардсу, Линдману и Сэвиджу [21].

Даже если у вас имеется упорядоченное мнение по такому вопросу, как диагноз, оно может радикально отличаться от мнения вашего коллеги. Термин «упорядоченное» указывает на то, что сумма вероятностей рассматриваемых событий равна единице. Обычно в результате применения формулы Байеса ваше мнение и мнение коллеги будет так преобразовано в ряд соответствующих наблюдений, т. е. признаков, симптомов и лабораторных исследований, что они станут почти неразличимыми.

Это приближенное слияние первоначально расходящихся мнений является, как указал Эдвардс с соавторами [21], одной из причин того, что выводы из достаточно адекватных данных считаются «объективными». Субъективной мере вероятности иногда противопоставляется тот довод, что научные или какие-либо другие знания не могут быть просто мнением. И все же нет четкой грани, отделяющей, например, предположение о том, что многие случаи рака у людей вызваны вирусами, от мнения, что многие случаи рака вызваны курением, или от «знания», что многие случаи рака вызваны радиацией.

Ваше мнение относительно диагноза может отличаться от мнения вашего коллеги, потому что он и вы располагаете разными объемами соответствующей информации. Пример, приведенный в гл. 2 (разд. 2.6.2), показывает, что Нуджент, работающий в Солт-Лейк-Сити и Лиддл из Нэшвилла, шт. Теннесси, собрали исходные данные для диагноза синдрома Кушинга, которые в определенной степени расходятся. Необходимо отметить, что в традиционной статистике не используется широко теорема Байеса; быть может, это объяс-

няется тем, что эта теорема служит для пересмотра вероятностей на основе новой информации, причем пересматривать относительные частоты не имеет смысла.

1.3. УСЛОВНЫЕ ВЕРОЯТНОСТИ И ТЕОРЕМА БАЙЕСА

Математическое определение условной вероятности симптома S при данном заболевании D таково:

$$P(S|D) = \frac{P(S \cap D)}{P(D)}, \text{ причем исключается случай } P(D) = 0. \quad (1)$$

Символ S обозначает любые сведения о пациенте в терминах симптомов, признаков и диагностических тестов. Все то, что мы обозначаем через S , будем называть одним термином — «симптомы», понимая под этим комплекс симптомов, признаки, лабораторные данные и т. д. Символ D обозначает заболевание, комплекс из нескольких заболеваний или в некоторых случаях отсутствие проявлений заболевания. Это является существенным упрощением с точки зрения обозначений для огромного множества клинических вариаций, как указано в гл. 2.

В терминах субъективных мер вероятностей уравнение (1) можно интерпретировать как сумму денег, которую вы готовы заплатить в обмен на доллар, который будет уплачен вам, если у пациента имеется комплекс симптомов S , с дополнительным условием, что сделка не состоится при отсутствии заболевания D . При изучении врожденных пороков сердца величина, определяемая уравнением (1), может рассматриваться как вероятность цианоза при условии, что имеет место дефект межпредсердной перегородки со стенозом легочной артерии; эта величина равна вероятности совместного существования цианоза и дефекта межпредсердной перегородки со стенозом легочной артерии, деленной на вероятность дефекта межпредсердной перегородки со стенозом легочной артерии, причем эти свойства рассматриваются для данного пациента с точки зрения данного исследователя в данном госпитале. Фактически любая вероятность должна иметь индекс, связанный с именем человека или группы людей, чье мнение она описывает.

Обозначение вида $P(S|D)$ читается следующим образом: «Вероятность S при данном D ». Использование вертикальной черты, пожалуй, оберегает нас от ошибок, которые могут возникнуть, если мы просто будем говорить о «вероятности симптомов S » безотносительно к «данной» информации о D . Простые алгебраические преобразования уравнения (1) приводят к основной формуле Байеса:

$$P(D|S) = \frac{P(S|D)P(D)}{P(S)} \quad (2)$$

при условии, что величины $P(D)$ и $P(S)$ не равны нулю. Если в уравнении (1) поменять ролями символы S и D , то новое соотношение можно выразить в симметричной форме:

$$\frac{P(S|D)}{P(S)} = \frac{P(D|S)}{P(D)}. \quad (3)$$

Интерпретация уравнения (3) заключается в том, что вероятностное соотношение между S и D равно соотношению между D и S . Де Финетти [14], формулируя это соотношение, придает ему следующий смысл: вероятность S при данном D изменяется в том же смысле и с той же мерой, как и вероятность D при данном S .

В уравнении (2) $P(D)$ — это первоначальная, или априорная, вероятность некоторого заболевания D . Это, по сути дела, также условная вероятность, хотя ее условность и не подчеркнута в обозначении; все вероятности в действительности условны. Рассматриваемая вероятность D условна по отношению ко всей предыдущей информации о D до того, как у пациента был обнаружен комплекс S . Аналогично $P(D|S)$ — конечная, или апостериорная, вероятность D — представляет собой вероятность D , условную по отношению ко всем предыдущим сведениям, включая данные S . $P(S|D)$ формально есть вероятность того, что данные S будут обнаружены, если заболевание D действительно имеет место. Эта величина выражает существенность данных S для обнаружения заболевания D . Получение значений $P(S|D)$ для каждого S и D является важным шагом при использовании теоремы Байеса в медицинской диагностике. В статистических приложениях величину $P(S|D)$ получают при помощи вычислений, основанных на так называемой статистической модели (например, на предположении о том, что наблюдаемые данные подчиняются нормальному распределению); в других приложениях величину $P(S|D)$ или какое-либо другое число, выполняющее ту же функцию (как мы скоро увидим при рассмотрении отношения правдоподобия), можно определить непосредственно из субъективного суждения.

Редкому заболеванию или редкому комплексу симптомов могут соответствовать низкие значения вероятностей, входящих в уравнение (2), если эти вероятности вычисляются на основе относительных частот. Тогда вычисление $P(D|S)$ может дать результат, вводящий в заблуждение. Удастся еще рассматривать вероятности порядка 0,1, но такое низкое значение, как 0,0009, может оказаться вообще неприемлемым, если только количество наблюдений не является очень большим, и, следовательно, редкость этого заболевания обычно делает невозможным определение вероятностей.

Ниже дан пример, иллюстрирующий применение формулы Байеса при редком заболевании или редком комплексе симптомов.

Заболевание	$P(D)$	$P(S_1 D)$	$P(D S_1)$
D_1	0,50	0,10	0,05
D_2	0,001	0,90	0,0009
D_3	0,90	0,001	0,0009

Для этой конкретной популяции заболевание D_1 встречается в 50% случаев, а редкое заболевание D_2 — в 0,1% случаев. Тем не менее комплекс симптомов S_1 встречается у 10% пациентов с заболеванием D_1 , тогда как тот же комплекс можно встретить в большом процентном отношении (90%) у пациентов с редким заболеванием D_2 . Комплекс симптомов S_1 редко встречается с некоторым обычным заболеванием D_3 . Вычисление величины $P(D|S)$ с использованием формулы Байеса дает обманчивые результаты для обоих заболеваний D_2 и D_3 .

Редкие события все же происходят, и врач всегда должен быть готовым встретить редкий комплекс симптомов или редкое заболевание. Было бы неуместно отказываться от использования формулы Байеса при диагностике на вычислительной машине только потому, что вычисления могут ввести врача в заблуждение, когда имеет место редкое заболевание. Вычисления, которые дают такую низкую вероятность диагноза $P(D|S)$, как 0,0009, должны немедленно возбудить у врача подозрение, что происходит нечто необычное; он должен исследовать этот случай низкой вероятности. В гл. 2 (разд. 2.2) мы объясним, как Вильямс и его коллеги рассматривали редкий случай врожденного отсутствия связывающего глобулина щитовидной железы в работе по машинной диагностике заболеваний щитовидной железы.

Важно знать как можно больше об априорном распределении редкого заболевания или редкого комплекса симптомов. Читатель может познакомиться с интересным обсуждением редких событий и принципом устойчивой оценки в статье [21]. С точки зрения распределения вероятностей редкие события более уместно описывать при помощи пуассоновских процессов.

Таким образом, в тех случаях, когда вероятность появления события в опыте мала, можно использовать распределение Пуассона. Это распределение хорошо известно и часто называется «законом редких событий». Обычный путь проверки, является ли данный процесс пуассоновским, заключается в том, что определяют, подчиняется ли число благоприятных случаев распределению Пуассона.

Распределение Пуассона выражается следующей формулой:

$$P_n = \frac{m^n}{n!} e^{-m}.$$

Другими словами, P_n — вероятность того, что при среднем числе событий m на интервале измерения число действительно наблюдаемых событий на этом интервале равно n .

Если редким событием является диагноз D и мы хотим получить величину $P(D)$ из распределения Пуассона, то приведенная выше формула будет иметь следующий смысл: $P_n(D)$ — это вероятность того, что при среднем числе m случаев постановки диагноза D на интервале измерения (например, m случаев в год) число случаев, когда заболевание D будет действительно иметь место, равно n (n случаев в год). Если вы наблюдали в клинике в среднем трех пациентов в год с саркомой кости для определенной возрастной группы, определенного пола и т. д., то вероятность того, что вы встретите пять таких случаев в этом году, равна

$$P_5(D) = \frac{3^5 \cdot e^{-3}}{5!} = 0,1.$$

Тот же подход можно использовать, если под редким событием понимать комплекс симптомов S и определять из распределения Пуассона величину $P(S|D)$. Изложение вопросов использования распределения Пуассона можно найти в учебниках по теории вероятностей ¹⁾.

Вероятность $P(S)$ в знаменателе уравнения (2) обычно не представляет большого интереса. Она может быть вычислена или вовсе исключена из уравнения следующим образом. Предположим, что D_i — это одно из набора единственно возможных и несовместных заболеваний. Определение вероятности требует, чтобы выполнялось равенство

$$\sum_i P(D_i|S) = \sum_i P(D_i) = 1.$$

Далее, используя уравнение (2), получим

$$P(S) = \sum_i P(S|D_i) P(D_i). \quad (4)$$

Выбор группы несовместных заболеваний является важным, но в значительной степени произвольным. Этот выбор должен всегда оставлять место для категории «некоторое другое объяснение». Это означает, что в некоторых случаях диагноз остается «неопределенным»; в медицине такая ситуация действительно встречается довольно часто, как показывает внимательное изучение протоколов клиничко-патологических конференций. При принятии решения в этом случае обычно изучают ситуацию и основываются на отрицании выбора «некоторого другого объяснения» путем неформальной

¹⁾ Например, см. Lindley D. V., Introduction to probability and statistics, Part 1, Probability, Cambridge, Cambridge U. P., 1965, chapt 2.

оценки того, достаточно ли соответствует наблюдаемым фактам какое-либо из заболеваний, перечисленных в списке, чтобы отклонить категорию «некоторого другого объяснения». Включение «неопределенной» категории в дифференциальную диагностику необходимо для того, чтобы «замкнуть круг» рассматриваемых диагнозов.

Существуют приложения теоремы Байеса, в которых группа рассматриваемых событий непрерывна. Это, например, случай, когда рассматриваются кривые разведения красителя при диагностике врожденных пороков сердца.

В частности, если параметр или набор параметров λ имеет априорную плотность распределения вероятности $u(\lambda)$ и если x — это случайная переменная (или набор случайных переменных, например набор измерений), для которой $v(x|\lambda)$ — плотность распределения x при данной величине λ , то апостериорная плотность вероятности λ при данной величине x равна

$$u(\lambda|x) = \frac{v(x|\lambda)u(\lambda)}{v(x)}. \quad (5)$$

Четыре вероятности в уравнении (2), а именно $P(D|S)$, $P(S|D)$, $P(D)$ и $P(S)$, являются субъективными мерами вероятности. Это не означает, что они не могут быть частотами, отношениями благоприятных возможностей ко всем возможностям, или числами, полученными прямо из суждения врача. Однако, как указано в работе [21], некоторые вероятности, если можно так выразиться, более субъективны, чем другие. В определенных приложениях может оказаться, что практически все врачи имеют весьма согласованное мнение о величине $P(S|D)$, например, в случае вероятности цианоза при наличии дефекта межпредсердной перегородки и стеноза легочной артерии. При этом значение $P(S|D)$ становится общепринятым. Это обычно происходит, когда величина $P(S|D)$ определяется простой моделью, которую ученые принимают как приближенное описание их мнения о рассматриваемой ситуации. Традиционный пример такой статистической модели — это извлечение шара из урны, относительно которой известно, что она содержит несколько шаров белых и черных. Если серию шаров извлекают из урны, причем после каждого извлечения шар кладут обратно в урну и как следует ее встряхивают, то большинство людей согласится, по крайней мере в качестве первого предположения, что вероятность вынуть определенную последовательность K (такую, например, как черный, белый, черный, черный) в предположении, что в урне было B черных и W белых шаров, равна

$$\left(\frac{B}{B+W}\right)^b \left(\frac{W}{B+W}\right)^w,$$

где b — число черных и w — число белых шаров в последовательности K .

Для большинства заболеваний общепринятая модель $P(S|D)$ может включать наблюдения случайных выборок из хорошо определенной популяции, подобно предыдущему примеру. Но точное определение популяции может оказаться трудным или невозможным делом, и получить выборку, случайные свойства которой удовлетворяют вас и вашего скептически настроенного коллегу, может быть достаточно сложно. В некоторых случаях, и пожалуй даже в большинстве случаев, при оценке $P(S|D)$ далеко не все врачи приходят к общему соглашению. При этом нельзя достигнуть единого заключения, потому что не существует данных относительно $P(S|D)$ для большинства заболеваний. Ледли и Ластед [53] считают, что причина, по которой медицинский опыт выражается в форме величины $P(S|D)$, заключается в следующем: эта условная вероятность относительно независима от местных окружающих факторов, таких, как географический район и время года, а зависит в первую очередь от физиолого-патологических аспектов самого комплекса заболеваний. Таким образом, изучение процесса заболевания как некоторой причины, вызывающей появление различных комплексов симптомов, может быть выражено в виде условных вероятностей следующего типа: вероятности появления некоторого комплекса симптомов у пациента при условии, что последний страдает данным комплексом заболеваний.

Некоторые данные, подтверждающие эту точку зрения и приведенные в статье [132], взяты из работы по диагностике заболеваний щитовидной железы, выполненной в Отделении ядерной медицины Боннского университета (ФРГ). Авторы обследовали сто пациентов, страдающих гиперфункцией щитовидной железы, и сравнили частоты клинических признаков и симптомов с результатами, опубликованными в статье [32] сотрудниками отделения радиологии медицинского колледжа Университета Флориды. Соответствующие значения $P(S|D)$ приведены на фиг. 1. Исследования, проведенные в ФРГ и во Флориде, дали сходные результаты.

Тем не менее общепринятые модели не бывают совершенными и часто вообще недостижимы. Что же можно сделать в такой ситуации? Эдвардс с соавторами [20, 21] показали, что в некоторых статистических ситуациях, таких, например, с которыми мы встречаемся в медицинской диагностике, уместно положить в основу общепринятой модели, объясняющей величину $P(S|D)$, мнения специалистов высокой квалификации. Это означает, что опытный клиницист может дать ценную информацию относительно величины $P(S|D)$, используя свое клиническое суждение. Психологические исследования [18] по использованию байесовской процедуры в ситуациях, где нужно принимать решение, показали, что люди могут научиться оценивать величину $P(S|D)$ в виде отношений правдоподобия или шансов. Этот вопрос будет вкратце рассмотрен ниже (гл. 4). Лодвик [62] подтвердил это положение, воспользовавшись оценкой величины

Клинические признаки, или симптомы (S), для гиперфункции щитовидной железы (D)	Частоты $P(S D)$ [31, 32]	Частоты $P(S D)$ [132]
Нервозность	0,915 (108/118)	0,925 (195/210)
Чувствительность к теплу	0,742 (72/97)	0,746 (100/134)
Потливость	0,678 (59/87)	0,747 (109/146)
Увеличение аппетита	0,605 (69/114)	0,579 (84/145)
Потеря веса	0,836 (112/134)	0,846 (187/221)
Гиперкинез	0,755 (83/110)	0,688 (117/170)
Теплая влажная кожа	0,708 (80/110)	0,760 (130/171)
Легкий тремор	0,871 (108/124)	0,898 (193/215)
Летаргия	0,001 (0/119)	0,102 (16/157)
Чувствительность к холоду	0,051 (5/98)	0,094 (13/138)
Пониженная потливость	0,001 (0/98)	0,001 (0/157)
Потеря аппетита	0,133 (15/133)	0,160 (25/156)
Увеличение веса	0,023 (3/133)	0,032 (7/218)
Замедленные движения	0,018 (2/109)	0,012 (2/171)
Сухая грубая кожа	0,009 (1/112)	0,011 (2/185)
Отек лица	0,008 (1/127)	0,026 (5/195)
Глазные симптомы	0,300 (30/100)	0,446 (87/195)

Ф и г. 1.

Частоты признаков и симптомов для гиперфункции щитовидной железы. Значения частот можно использовать в качестве величин $P(S|D)$. Для гиперфункции щитовидной железы $P(D) = 0,276$ (269/974) [132]. Для гиперфункции щитовидной железы $P(D) = 0,198$ (174/879); для нормальной функции щитовидной железы $P(D) = 0,662$ (582/879); для пониженной функции щитовидной железы $P(D) = 0,140$ (123/879) [31].

$P(S|D)$, даваемой рентгенологом при диагностике костных опухолей.

Вероятность $P(D)$ — это априорная вероятность заболевания до того, как собраны симптоматические данные. Необходимость оценки априорных вероятностей $P(D)$ является существенным источником противоречий при применении формулы Байеса в медицинской диагностике. Нужно отметить, что в общих работах по теории статистических решений, где вместо D рассматривается некоторая гипотеза H , критика применения формулы Байеса сосредоточивается на трудности оценки величины $P(H)$. Некоторые исследователи пробовали обойти эту трудность, используя метод отношения правдоподобия (который мы обсудим ниже), поскольку в нем не применяется величина $P(D)$. Лишь немногие авторы пытались использовать в качестве источника $P(D)$ клиническое суждение. Таким образом, представляется важным рассмотреть априорную вероятность $P(D)$ более подробно.

Вспомним, что в ходе нашего рассмотрения величина $P(D)$ была определена как субъективная мера вероятности, хотя это не исключает использование частот (или относительных частот) заболевания, имевших место в данной больнице или клинике. Например, величина $P(D)$ для пониженной функции щитовидной железы по вашему опыту в вашей больнице может быть равна 0,140. Конечно, у вашего коллеги величина $P(D)$ для пониженной функции щитовидной железы может иметь другое значение, но это не значит, что вы и он не должны использовать значения, которые получил каждый из вас. Действительно, должно быть особенно интересно и показательно попытаться определить причины, по которым ваше значение отличается от значения коллеги. В работе [53] указано, что сомножитель $P(D)$ представляет собой величину, учитывающую влияние местных факторов — географического местоположения, сезонного влияния, возникновения эпидемий и т. д. Нуджент с соавторами [75] и Лиддл отметили такие различия для $P(D)$ при диагностике синдрома Кушинга. Нуджент в Метаболической клинике медицинского колледжа Университета Юты нашел значение $P(D)=0,25$ для синдрома Кушинга, тогда как Лиддл из Эндокринологической клиники Вандербилтского университета, штат Теннесси, приводит значение $P(D)=0,19$. Такое расхождение в значениях $P(D)$ можно было ожидать, принимая во внимание «индивидуализированность» диагнозов в Юте и Вандербилте.

Рассмотрим пример использования формулы Байеса и данных, приведенных в работе [75].

Для предполагаемого пациента, обследованного в эндокринологической клинике, получены следующие данные: пол женский, возраст 40 лет, общее ожирение, красный дермографизм, диастолическое кровяное давление выше 105, гирсутизм и понижение толерантности к глюкозе; никаких других признаков, которые указывали бы на синдром Кушинга, не имеется.

Используя частоты, приведенные в таблице к фиг. 2, вычислим вероятность того, что у предполагаемого пациента имеется синдром Кушинга. Заметим, что если симптом отсутствует, то при вычислении вместо $P(S|D)$ используется величина $[1-P(S|D)]$. Для упрощения расчетов симптомы с S_{12} по S_{19} опускаем.

$$P(D_1 | S_i) = \frac{P(D_1) [1 - P(S_1 | D_1)] P(S_2 | D_1) [1 - P(S_4 | D_1)] [1 - P(S_6 | D_1)] \times [1 - P(S_8 | D_1)] [1 - P(S_7 | D_1)] P(S_8 | D_1) P(S_9 | D_1) \times [1 - P(S_{10} | D_1)] P(S_{11} | D_1)}{\text{числитель} + P(D_2) [1 - P(S_1 | D_2)] P(S_2 | D_2) [1 - P(S_4 | D_2)] \times [1 - P(S_6 | D_2)] [1 - P(S_7 | D_2)] P(S_8 | D_2) P(S_9 | D_2) \times [1 - P(S_{10} | D_2)] P(S_{11} | D_2)}$$

$$P(D_1 | S_i) = \frac{(0,25) (0,73) (0,03) (0,35) (0,18) (0,42) (0,48) (0,46) (0,39) (0,62) (0,50)}{\text{числитель} + (0,75) (0,94) (0,62) (0,93) (0,69) (0,70) (0,76) (0,22) \times (0,17) (0,83) (0,29)} < 0,01.$$

Таким образом, на основании вычислений приходим к выводу, что у пациента вряд ли имеется синдром Кушинга. Хотя такой расчет мо-

Признаки и симптомы (S)	Частоты $P(S D_1)$ для пациентов с синдромом Кушинга (D_1)	Частоты $P(S D_2)$ для пациентов, не имеющих синдрома Кушинга (D_2)
S1 Остеопороз	0,27	0,06
S2 Ожирение туловища	0,90	0,29
S3 Общее ожирение	0,03	0,62
S4 Слабость	0,65	0,07
S5 Полнокровие	0,82	0,31
S6 Количество лейкоцитов 11 000 на mm^3 или больше	0,58	0,30
S7 Угри	0,52	0,24
S8 Дермографизм (красный или багровый)	0,46	0,22
S9 Диастолическое кровяное давление 105 мм или выше	0,39	0,17
S10 Отек (пастозность)	0,38	0,17
S11 Гирсутизм	0,50	0,29
S12 Экхимозы	0,53	0,06
S13 Содержание калия в сыворотке 3,6 mEq/l или меньше	0,25	0,04
S14 Скудные менструации	0,72	0,51
S15 Головные боли	0,41	0,37
S16 РОЭ 49 или выше	0,37	0,32
S17 Пол женский	0,65	0,77
S18 Ненормальный уровень связывающего глобулина	0,88	0,77
S19 Возраст 35 лет или меньше	0,55	0,52

Ф и г. 2.

Частоты признаков и симптомов для синдрома Кушинга. Для пациентов с синдромом Кушинга $P(D_1) = 0,25$. Для пациентов, не имеющих синдрома Кушинга, $P(D_2) = 0,75$ [75].

жет быть выполнен на простом арифмометре, очевидно, использование электронной вычислительной машины оказало бы существенную помощь в случае большого количества признаков и целой группы больных.

Как же следует поступить, если, по вашему мнению, невозможно определить значение $P(D)$ на основе общепринятой модели в терминах случайных выборок из хорошо определенной популяции? Можете ли вы сказать, например, что, согласно вашему опыту, величина $P(D)$ для гиперфункции щитовидной железы в вашей клинике равна 0,2? Ответ утвердителен: «да».

Но возникает следующий вопрос: раз значение $P(D)$ может быть результатом клинического суждения, не является ли эта величина не только неопределенной, но и переменной и, следовательно, бесполезной для ваших целей? Дело в том, что, хотя априорные вероятности часто неопределенны и переменны, они совсем не обязательно бесполезны. Сами они часто имеют лишь незначительное влияние на результаты, полученные по формуле Байеса. Интересно рассмотреть с точки зрения байесовской теории, почему это происходит.

Признаки и симптомы (S)	Частоты для пациентов с синдромом Кушинга (D_1)	Шансы для пациентов с синдромом Кушинга, $\Omega(D_1) = \frac{P(S D_1)}{1-P(S D_1)}$	Частоты для пациентов, не имеющих синдрома Кушинга (D_2)
	$P(S D_1)$	$\Omega(D_1)$	$P(S D_2)$
S 2 Ожирение туловища	0,90	9,0 к 1	0,29
S18 Ненормальный уровень связывающего глобулина	0,88	7,3 к 1	0,77
S 5 Полнокровие	0,82	4,5 к 1	0,31
S14 Скудные менструации	0,72	2,5 к 1	0,51
S 4 Слабость	0,65	1,8 к 1	0,07
S17 Пол женский	0,65	1,8 к 1	0,77
S 6 Количество лейкоцитов 11 000 на mm^3 или больше	0,58	1,4 к 1	0,30
S19 Возраст 35 лет или меньше	0,55	1,2 к 1	0,52
S12 Экхимозы	0,53	1,1 к 1	0,06
S 7 Угри	0,52	1,1 к 1	0,24
S11 Гирсутизм	0,50	1,0 к 1	0,29
S 8 Дермографизм (красный или багровый)	0,46	0,85 к 1	0,22
S15 Головные боли	0,41	0,70 к 1	0,37
S 9 Диастолическое кровяное давление 105 мм или выше	0,39	0,64 к 1	0,17
S10 Отек (пастозность)	0,38	0,61 к 1	0,17
S16 РОЭ 49 мм или выше	0,37	0,58 к 1	0,32
S 1 Остеопороз	0,27	0,37 к 1	0,06
S13 Содержание калия в сыворотке 3,6 $мэкв/л$ или меньше	0,25	0,33 к 1	0,04
S 3 Общее ожирение	0,03	0,03 к 1	0,62

Ф и г. 3.

Частоты признаков и симптомов для синдрома Кушинга. Признаки и симптомы (S) пациентов с синдромом Кушинга расположены в порядке уменьшения частоты. При отсутствии данных о частотах опытный клиницист может указать шансы.

Если наблюдения в определенном смысле точны относительно априорного распределения, на которое они опираются, то вид и свойства этого априорного распределения оказывают незначительное влияние на распределение апостериорное. Мостеллер и Уоллес [71] пришли к этому заключению на основании применения формулы Байеса при исследовании проблемы спорного авторства известных статей «The Federalist Papers». Они пишут: «Априорные распределения не играют важной роли. Хотя выбор начальных констант (выбор априорных распределений) и имеет значение, они становятся не

Признаки и симптомы (S)	Отношение правдоподобия $L(D_1; S)$ для пациентов с синдромом Кушинга, $L(D_1; S) = P(S D_1)/P(S D_2)$
S 4 Слабость	9,28
S12 Экхимозы	8,83
S13 Содержание калия в сыворотке 3,6 мэкв/л или меньше	6,25
S 1 Остеопороз	4,50
S 2 Ожирение туловища	3,10
S 5 Полнокровие	2,64
S 9 Диастолическое кровяное давление 105 мм и выше	2,29
S10 Отек (пастозность)	2,23
S 7 Угри	2,17
S 8 Дермографизм (красный или багровый)	2,09
S 6 Количество лейкоцитов 11 000 на мм ³ или больше	1,93
S11 Гирсутизм	1,72
S14 Скудные менструации	1,40
S16 РОЭ 49 мм или выше	1,15
S18 Ненормальный уровень связывающего глобулина	1,14
S15 Головные боли	1,11
S19 Возраст 35 лет или меньше	1,06
S17 Пол женский	0,84
S 3 Общее ожирение	0,05

Ф и г. 4.

Отношения правдоподобия для признаков и симптомов синдрома Кушинга в порядке убывания их величины. Отношения правдоподобия могут быть вычислены из данных о частотах. Отношение правдоподобия — это отношение двух частот. При отсутствии данных о частотах отношения правдоподобия (или логарифмические отношения правдоподобия) могут быть указаны опытным клиницистом. Таблицы частот, отношений правдоподобия и шансов для многих комплексов симптом—заболевание могут быть изданы в виде справочников для врачей.

очень существенными, если мы приблизимся к распределению, определяемому достаточно хорошим набором экспериментальных данных. Отсюда можно заключить, что подчеркивание трудностей и даже невозможности выбора априорных распределений является слабым аргументом, когда критикуют применимость формулы Байеса». Мостеллер и Уоллес говорили об априорных распределениях для генерирующих процессов, но это утверждение применимо также к априорным распределениям для гипотез.

Чтобы рассмотреть практическую ситуацию, предположим, что байесовская модель используется при составлении программы вычислительной машины для обработки поступающей информации о пациентах с заболеваниями щитовидной железы в эндокринологической клинике. Информация в виде величины $P(S|D)$ задается врачом, который уже получил значение $P(D)$ из какого-либо источника, как мы обсуждали выше. Вычисления производятся в соответствии с формулой Байеса, и врач получает на выходе машины результат для каждого пациента в виде значений $P(D|S)$ для следующих дифференциальных диагнозов: гиперфункции щитовидной железы, пониженной функции щитовидной железы, нормальной функции щитовидной железы. В дальнейшем полученные значения $P(D|S)$ для каждого пациента добавляются к предыдущим данным для вычисления нового значения $P(D)$. Очевидно, какими бы ни были априорные вероятности $P(D)$ в любой большой клинике в начале дня или недели, они будут быстро перекрыты массой приходящей информации. Вот почему первоначальное значение $P(D)$ для гиперфункции щитовидной железы могло быть 0,2 или 0,3 [начальные оценки $P(D)$ будут более или менее подходящими до тех пор, пока они не особенно близка к 0 или 1] с пренебрежимо малым влиянием на $P(D|S)$.

Улучшение значений $P(D)$ с помощью величины $P(D|S)$ легко осуществляется программой вычислительной машины и обычно используется в вычислительной диагностике сердечных заболеваний, заболеваний щитовидной железы и костных опухолей. Сошлемся только на один пример. В работе [32] опубликована программа вычислительной машины для диагностики заболевания щитовидной железы, которая «накапливает опыт» путем улучшения матрицы основных данных после каждого диагноза¹⁾. Это позволяет заново вычислять величины $P(D)$ и $P(S|D)$. Упомянутая программа составлена так, что она может быть использована другими исследователями (приложение 3), причем исследователь может сам составить матрицу симптомы — заболевания с данными, полученными менее чем из ста случаев.

¹⁾ Метод и программы самообучения диагностических систем опубликованы ранее в советской литературе. См. *Экспериментальная хирургия и анестезиология*, № 4, 1964.—Прим. перев.

Заполнение матрицы симптома — заболевания величинами $P(S|D)$ является ключом к использованию формулы Байеса. На фиг. 5 приведена матрица, разработанная Лодвиком [61] для диагностики первичных костных опухолей.

Вид опухоли	Частота $P(D)$	Возраст		Локализация				Размер		Поражен-ная суб-станция				
		$P(S D)$		$P(S D)$		$P(S D)$		$P(S D)$		$P(S D)$				
		до 21 года	21—30	31 и больше	малые и плоские	длинные	эпифиз	пластинка	метафиз	диафиз	1—6 см	6 см и больше	кость	хрящ
Гигантоклеточная опухоль	15	20	35	45	20	80	99	01	100	20	40	60	01	00
Хондробластома	05	75	20	05	20	80	100	50	75	00	90	10	00	30
Хондромиксоидная фиброма	03	50	35	15	30	70	30	20	100	25	85	15	00	02
Хондросаркома	17	25	25	50	35	65	40	01	85	65	20	80	05	65
Фибросаркома	10	20	20	60	05	95	55	00	90	65	20	80	25	02
Остеосаркома	25	65	25	10	10	90	30	05	95	75	15	85	98	05
Параостеальная саркома	05	20	35	45	00	100	30	01	100	50	25	75	100	05
Опухоль Юинга	15	70	25	05	35	65	20	05	85	90	15	85	00	00
Саркома ретикулярных клеток	05	10	25	65	20	80	50	01	85	80	15	85	00	00

Фиг. 5.

Упрощенная матрица априорных вероятностей для первичных костных опухолей. Каждое число отражает вероятное количество случаев, в которых будет обнаружен признак в выборке из ста случаев. Некоторые очень высокие или очень низкие значения вероятностей (100, 99, 01, 00) по существу используются как булевы функции для исключения определенных категорий заболевания.

На начальном этапе построения матрицы, если наличие признака при данном заболевании сомнительно или признак вообще отсутствует, в соответствующем квадрате лучше поставить 1, а не 0¹⁾. Такой подход позволяет избежать трудностей, связанных с делением на нуль при последующих вычислениях. В матрице, показанной на фиг. 5, такой прием не был использован. Отсутствие данных приводит к новым проблемам. Вопрос о том, как поступать при отсутствии данных или в каких случаях отсутствие данных можно принять за «данные», нуждается в дополнительном изучении.

В процессе накопления данных симптома — заболевания для некоторых пациентов будет получена неполная или частичная информация. Частичная информация относится к тем случаям, когда

¹⁾ При этом, естественно, вместо соответствующего признака рассматривается дополнительный. — Прим. перев.

неизвестно, имели ли пациент определенный симптом или комплекс симптомов. В этих случаях матрица может быть улучшена путем добавления единицы по данному входу для каждого известного симптома и добавлением нуля по данному входу при отсутствии симптома.

Требуют рассмотрения следующие два вопроса, касающиеся использования величин $P(D)$ и $P(S|D)$. Это предположения о том, что применение формулы Байеса в медицинской диагностике требует, чтобы заболевания были взаимоисключающими и чтобы различные симптомы были независимы. С проблемами первого типа справиться нетрудно, потому что, как мы заметили ранее, распределение $P(D)$ по отдельным заболеваниям хотя и важно, но в значительной степени произвольно. В работе [125] приведен пример, который иллюстрирует это положение: «Невозможно вычислить вероятность комбинации заболеваний по данным о частотах симптомов для каждого из заболеваний в отдельности. Для иллюстрации этого утверждения рассмотрим два заболевания — стеноз легочной артерии и дефект межпредсердной перегородки. Так как ни одно из этих заболеваний не вызывает цианоза, представляется невозможным предсказать просто из частот цианоза в отдельных заболеваниях, что у пациента, страдающего обоими заболеваниями, может быть цианоз. Таким образом, стеноз легочной артерии плюс дефект межпредсердной перегородки нужно рассматривать как новое заболевание». Уорнер устранил границу, разделяющую эти два заболевания, для того чтобы сформировать их комбинацию, которую он затем рассматривает как новое заболевание. Будущие исследования в других областях медицины могут обнаружить заболевания, которые, наоборот, следовало бы разъединить.

В статье Уорнера [124] одно из предположений заключается в том, что при данном заболевании симптомы не зависят друг от друга. Уорнер тщательно отобрал данные, для которых вычислял величины $P(S|D)$. Для данной матрицы заболеваний предположение о независимости может быть проверено при помощи критерия « χ -квадрат», если можно получить достаточные данные относительно совместного наличия симптомов при каждом заболевании. Основываясь на упомянутой работе Уорнера, некоторые исследователи решили, что применение формулы Байеса в медицинской диагностике обязательно требует предположения о независимости симптомов. Но дело не в этом. Корнфилд [11] совершенно ясно показал это при обсуждении теоремы Байеса: «Во-вторых, не было сделано никаких предположений относительно того, что компоненты многомерного события независимы. Так, для состояний θ_2 и θ_3 в примере с подбрасыванием монеты мы имеем полную зависимость между первым и вторым бросанием. Мы подчеркиваем это очевидное обстоятельство только потому, что в медицинской литературе, по-видимому, существует широко распространенное ошибочное мнение, согласно

которому применение формулы Байеса к проблеме диагностики требует предположения о независимости различных симптомов». Симптомы S_1 и S_2 считаются независимыми для определенного заболевания D , если условная вероятность S_1 при данном S_2 и данном D равна безусловной вероятности S_1 для всех S_2 при данном D .

Тем не менее предыдущее обсуждение не следует интерпретировать так, как будто все вопросы об условной независимости симптомов уже решены или проблемы, которые вытекают из этих вопросов, не связаны с серьезными трудностями. Это будет далеко от истинного положения вещей; в действительности некоторые из наиболее полезных результатов для медицинской диагностики могут быть получены при изучении проблем, связанных с условной независимостью симптомов. В работе [125] показано преимущество группирования взаимоисключающих симптомов, с тем чтобы не более чем один симптом присутствовал в любом наборе. Например, врач может выбрать только один симптом из набора:

цианоз умеренный,
цианоз резкий (с пальцами в виде «барабанных палочек»),
цианоз непостоянный,
цианоз диссоциированный.

Уменьшение количества данных путем комбинирования взаимосвязанных симптомов при первичных опухолях костей было осуществлено в работе [61] при помощи системы градации опухолей. Этот метод подробно описан в гл. 2 (разд. 2.4). Тщательный поиск групп или наборов симптомов во всех областях медицины должен привести к объединению симптомов в комбинации, с которыми будет проще иметь дело, чем со списками симптомов, которые мы имеем в настоящее время.

Для дискретных симптомов S и заболеваний D общее выражение теоремы Байеса может быть записано в следующем виде:

$$P(D|S_n, S_{n-1}, \dots, S_1) = \frac{P(S_n|D, S_{n-1}, S_{n-2}, \dots, S_1) P(D|S_{n-1}, S_{n-2}, \dots, S_1)}{P(S_n|S_{n-1}, S_{n-2}, \dots, S_1)}. \quad (6)$$

Такая форма не требует условной независимости симптомов S , но величины

$$P(S_n|D, S_{n-1}, S_{n-2}, \dots, S_1) \text{ и } P(S_n|S_{n-1}, S_{n-2}, \dots, S_1)$$

очень трудно получить из соответствующих частот или оценить на основании клинического суждения. Как было отмечено раньше, член $P(S_n|S_{n-1}, S_{n-2}, \dots, S_1)$ часто может быть исключен.

Два симптома, S_1 и S_2 , относящиеся к заболеванию D , не могут быть, вообще говоря, независимы друг от друга, поскольку истинность или ложность гипотезы о возможности некоторого заболевания, являющегося причиной симптомов S_1 и S_2 , приводит к тому, что

эти симптомы, если только они действительно диагностически значимы, становятся взаимозависимыми. Тем не менее иногда удается найти такое разбиение заболеваний, для которого симптомы S_1 и S_2 будут условно независимыми¹⁾. Тогда для некоторого заболевания D_i

$$P(S_1|D_i, S_2) = P(S_1|D_i) \quad \text{и} \quad P(S_2|D_i, S_1) = P(S_2|D_i).$$

В том и только в том случае, если нам удастся определить группу заболеваний, для которых условная независимость существует для всех симптомов и комплексов симптомов, теорема Байеса может быть записана в простой форме уравнения (2):

$$P(D|S) = \frac{P(S|D)P(D)}{P(S)}.$$

Насколько серьезно следует относиться к тому, что невозможно доказать независимость симптомов? Один из ответов заключается в следующем: не нужно подходить к этому настолько серьезно, чтобы отказаться от использования методов Байеса в приложении к медицинской диагностике. Вы можете считать, что симптомы условно независимы, даже когда у вас есть причины предполагать, что это не так. Поступая таким образом, можно сэкономить машинное время, но все же не следует относиться к такому положению вещей слишком безответственно. Нужно постоянно иметь в виду возможность столь значительных нарушений условной независимости, что они могут привести к большим ошибкам в диагностике и лечении. Тогда стоит сделать дополнительную попытку исследовать ситуацию. В этом случае может оказать помощь система, основанная на использовании электронной вычислительной машины, такая, например, как «Система вероятностной обработки информации» (PIP), описанная в гл. 5.

Лодвик [62] пришел к аналогичным заключениям в своей работе по диагностике костных опухолей. Он утверждает: «Наша версия правила Байеса предполагает, что ожидаемые признаки независимы, т. е. что информация, относящаяся к наличию или отсутствию одного признака, не дает информации относительно вероятности наличия или отсутствия любого другого. Тем не менее нельзя доказать, что признаки независимы. Можно лишь определить корреляцию между признаками в данной выборке. Отсутствие корреляции просто указывает на возможность независимости; оно не может быть принято в качестве доказательства. Таким образом, мы оказываемся на теоретически шаткой основе, применяя теорему Байеса в нашей интерпретации. Однако в практическом отношении мы проверили байесовский подход, и он действует».

¹⁾ Это обнаружено и строго доказано в работе М. Л. Быховского «К вопросу о вероятностной зависимости признаков в проблеме диагностики», *Экспериментальная хирургия и анестезиология*, № 2, 1967. — Прим. перев.

Хотя байесовский подход и эффективен, Темплтон с соавторами [117] подчеркивают необходимость улучшенных оценок условной вероятности $P(S|D)$ наблюдения определенной комбинации симптомов при данном заболевании. Эти авторы показали, что, хотя в результате вычислений, основанных на предположении о независимости симптомов, наибольшая вероятность приписывается большей части (0,78) доказанных правильных диагнозов, эта вычисленная наибольшая вероятность точно не отражала шансы постановки пациенту правильного диагноза.

Бранк и Лер [4] предложили метод определения улучшенных оценок апостериорной вероятности заболевания $P(D|S)$, вычисленной с помощью теоремы Байеса в случае, когда не предполагается независимость Грама — Шмидта; в нем предположение о независимости симптомов заменяется более слабым предположением о существовании линейной зависимости. Для реализации метода нужно небольшое число верифицированных диагнозов; он был проверен при диагностике врожденных пороков сердца на данных Темплтона с соавторами.

Метод, используемый Бранком и Лером, заключается в следующем. Выберем некоторый важный симптом, представляющий собой случайную переменную, такую, например, как возраст, и возьмем эту переменную в качестве первой в новой группе случайных переменных. Те симптомы, которые сильно коррелированы с возрастом, отбрасываются, и выбирается симптом, слабо связанный с возрастом. Некоторая линейная функция переменной возраста вычитается из второй переменной так, что эти две переменные становятся некоррелированными. Затем отбрасываются те случайные переменные (симптомы), которые являются существенно линейными комбинациями этих двух переменных. Далее, выбирается третья переменная и линейная комбинация первых двух вычитается из третьей так, что эти три переменные становятся некоррелированными и т. д. Таким образом, исходный набор коррелированных случайных переменных заменяется новым набором некоррелированных случайных переменных. Этот новый набор некоррелированных переменных приводит к улучшенной оценке условных вероятностей для $P(S|D)$, так что диагноз врожденного порока сердца, которому приписывалась наибольшая вероятность, согласуется с правильным диагнозом в 86% случаев. «Предыдущий» метод, использованный Темплтоном с соавторами, согласуется в 78% случаев.

Бранк и Лер подробно изложили процесс ортогонализации Грама — Шмидта, а обсуждение, сопутствующее математическим выкладкам, может стимулировать дальнейший интерес к поискам методов группирования коррелированных переменных (симптомов).

1.4. ПРАВДОПОДОБИЕ И ШАНСЫ

Правдоподобие некоторой гипотезы (диагноза заболевания) представляет собой вероятность (при данной гипотезе) действительного наблюдения результатов эксперимента (симптомов). Член $P(S|D)$, который мы рассматривали выше, по сути дела и есть правдоподобие. Теорема Байеса говорит, что конечная вероятность гипотезы $P(D|S)$ пропорциональна ее начальной вероятности $P(D)$, помноженной на ее правдоподобие $P(S|D)$.

Может показаться, что правдоподобие и вероятность — просто два разных слова, определяющие одну и ту же величину, поэтому нужно рассмотреть этот вопрос подробнее. В некоторых задачах, обычно дедуктивного характера, мы имеем дело с вероятностями различных комплексов симптомов S_1, S_2, \dots при одном и том же данном заболевании D_1 , так что переменной является величина $P(S_i|D_1)$. В других задачах, обычно индуктивных, имеет место некоторое событие или один и тот же комплекс симптомов S и (как в случае применения теоремы Байеса) мы интересуемся вероятностью S_i при данных различных заболеваниях D_1, D_2, \dots , так что переменной является величина $P(S_i|D_i)$. Следовательно, в зависимости от того, что является переменной в вероятностной функции, используется термин «вероятность $P(S_i|D_1)$ » или «правдоподобие $P(S_i|D_i)$ ». При этом мы говорим о вероятности события (симптом S) или о правдоподобии гипотезы (заболевание D). Одно из различий между вероятностью и правдоподобием заключается в том, что если симптомы S_i являются единственно возможными и взаимоисключающими, то сумма вероятностей равна единице, если же заболевания D_i являются единственно возможными и взаимоисключающими, то сумма правдоподобий обычно не равна единице.

Было показано [21], что наиболее важную роль при вычислениях по формуле Байеса играет отношение правдоподобия (т. е. отношение двух правдоподобий)

$$\frac{P(S|D_i)}{P(S|D_j)} \quad (i \neq j). \quad (7)$$

Эта величина может выражать, например, отношение вероятности комплекса симптомов при заболевании астмой к вероятности того же комплекса симптомов у контрольной группы здоровых людей. Особенно удобная модификация теоремы Байеса для некоторых приложений имеет вид отношения шансов к правдоподобию.

Для двух заболеваний D_1 и D_2 и одного и того же комплекса симптомов S теорема Байеса может быть дважды записана так:

$$P(D_1|S) = \frac{P(S|D_1) P(D_1)}{P(S)} \quad (8)$$

и

$$P(D_2|S) = \frac{P(S|D_2)P(D_2)}{P(S)}. \quad (9)$$

Разделив уравнение (8) на уравнение (9), получим

$$\frac{P(D_1|S)}{P(D_2|S)} = \frac{P(S|D_1)}{P(S|D_2)} \cdot \frac{P(D_1)}{P(D_2)}. \quad (10)$$

Член $P(S|D_1)/P(S|D_2)$ есть отношение правдоподобия $L(D_1; S)$, как следует из уравнения (7). Отношения правдоподобия для признаков и симптомов в синдроме Кушинга приведены в таблице на фиг. 4.

Если пациент (возраст 16 лет) имеет первичную опухоль кости, каково отношение правдоподобия (на основании возраста) для хондробластомы относительно хондросаркомы? Искомое отношение правдоподобия можно получить, используя данные, приведенные на фиг. 5, следующим образом:

$$\frac{P(S|D_1)}{P(S|D_2)} = \frac{P(\text{возраст до 21 года} | \text{хондробластома})}{P(\text{возраст до 21 года} | \text{хондросаркома})} = \frac{75}{25} = \frac{3}{1}.$$

Тщательное определение признаков, симптомов и результатов лабораторных исследований является очень важным во всех областях медицины. Этот вопрос подробно обсуждается в гл. 2. Здесь важно отметить, что определение симптомов S менее критично, когда они используются для вычисления отношения правдоподобия, чем просто для вычисления условной вероятности. Основное требование заключается в том, что симптом S должен иметь одно и то же определение для заболеваний, которые сравнивают, а именно для D_1 и D_2 .

Отношение правдоподобия было использовано при разделении пациентов на группы и в гематологии. Эти экспериментальные работы обсуждаются в гл. 2 (разд. 2.5.1 и 2.5.2).

Величина $P(D_1)/P(D_2)$ представляет собой начальное значение относительных шансов Ω_0 и является просто отношением начальной вероятности D_1 к начальной вероятности D_2 . Лодвик считает, что начальные шансы костносаркомных клеток по отношению к ретикулярным клеткам равны $25/5=5/1$. Здесь слово «шансы» означает в точности то же самое, что и на скачках или гонках.

Уравнение (10) можно записать следующим образом:

$$\Omega_1 = L(D_1; S) \Omega_0. \quad (11)$$

Если пациент в возрасте до 21 года имеет первичную костную опухоль, каковы шансы, что это скорее костная саркома, чем ретикулярные клетки?

$$\begin{aligned}\Omega_1 &= \frac{P(\text{возраст до 21 года} \mid \text{костная саркома})}{P(\text{возраст до 21 года} \mid \text{ретикулярные клетки})} \times \frac{P(\text{костная саркома})}{P(\text{ретикулярные клетки})} = \\ &= \frac{65}{10} \cdot \frac{25}{5} \approx \frac{32}{1}.\end{aligned}$$

Этот пример показывает, как начальные шансы модифицируются отношением правдоподобия.

Заметим, что умножение величин $P(S|D_1)$ и $P(S|D_2)$ на любую константу не изменяет величины конечных шансов. Этот факт называется принципом правдоподобия, и на нем основаны важные приложения теоремы Байеса в медицинской диагностике и лечении. Согласно принципу правдоподобия, все наблюдения, приводящие к одной и той же функции правдоподобия, должны иметь своим следствием одно и то же заключение.

Математическая связь между шансами и вероятностями может быть выражена следующим образом. Шансы $\Omega(D)$ в пользу заболевания D связаны с вероятностью $P(D)$ заболевания D и вероятностью $1 - P(D)$, т. е. вероятностью отсутствия заболевания D , условием

$$\Omega(D) [1 - P(D)] = P(D).$$

Следовательно,

$$\Omega(D) = \frac{P(D)}{1 - P(D)}, \quad (12)$$

и можно записать

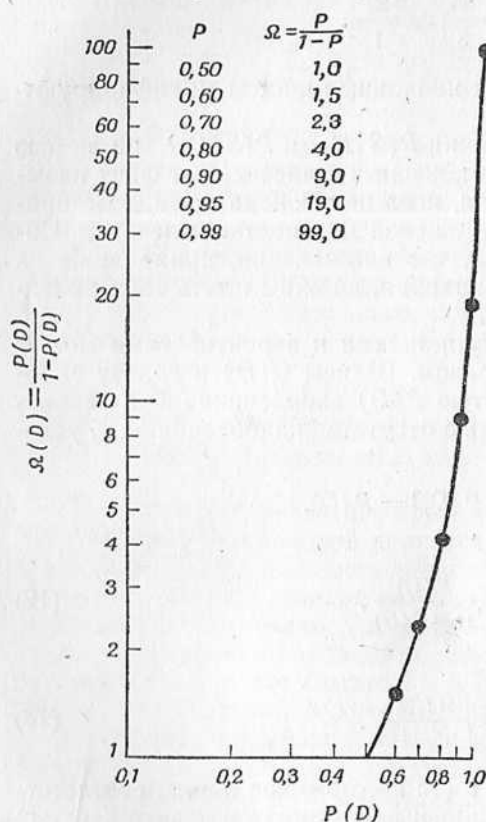
$$P(D) = \frac{\Omega(D)}{1 + \Omega(D)}. \quad (13)$$

Например, когда шансы равны 1 (1 : 1 — обычное пари), соответствующая вероятность равна половине; вероятность в девять десятых соответствует шансам 9 (9 : 1). На фиг. 3 приведены шансы для признаков и симптомов при синдроме Кушинга.

Экспериментальные исследования, проведенные Эдвардсом [18] и Стивенсом [105], показали, что удобно использовать логарифмическую шкалу для иллюстрации процессов суждения или решения, когда применяется метод соответствия или сравнения. Эдвардс [18] нашел, что операторы быстро схватывают идею логарифмических шансов или логарифмического отношения правдоподобия и что в большинстве ситуаций для человека легче оценить шансы и отношение правдоподобия, чем условную вероятность. Кривая шансов в функции вероятности приведена на фиг. 6 (в логарифмических шкалах).

Определение величин, которые могут быть использованы в качестве правдоподобия $P(S|D)$, является наиболее важным шагом при

применении теоремы Байеса к ситуациям, связанным с процессом принятия решения в диагностике. В некоторых случаях может оказаться более разумным представлять эту информацию в виде отношений правдоподобия или шансов.



Фиг. 6.
Вероятность в функции шансов (логарифмический масштаб).

суждение, утверждая: «В пользу первичной плевральной опухоли говорят очевидные плевральные затруднения из-за опухоли и отрицательные аргументы для рака бронхов. Против плевральной опухоли говорит статистическая вероятность». В соответствии с последними данными, использованными Фляйшнером, шансы примерно равны 5 : 1, то есть 5 — за рак бронхов и 1 — за мезотелиому плевры или брюшины.

«Мой вывод заключается в том, что у пациента имеется асбестоз с пульмональным фиброзом, хронический бронхит, эмфизема и двустороннее плевральное известкование. У него была злокачествен-

отношений правдоподобия или шансов.

Может быть, врачи и не очень часто выражают предпочтительные диагнозы в виде шансов [85], однако мы нашли упоминание о следующем случае. Это произошло во время медицинской конференции 22 августа 1958 г. в Стронговском мемориальном госпитале, Рочестер, шт. Нью-Йорк. Профессор Брэдфорд, председатель секции педиатрии, использовал в своем докладе концепцию шансов и закончил его фигуральным замечанием: «В этом случае ставьте деньги на вирус Echo № 9».

Доктор Фляйшнер использовал вероятность и шансы в недавнем обсуждении клиничко-патологических случаев; тем не менее он двигался к правильному диагнозу путем отбрасывания статистических шансов в пользу очевидности, которая, как он чувствовал, перевешивала шансы. Он использовал свою субъективную меру вероятности и свое клиническое

ная пульмональная или плевральная опухоль, со средостенной лимфоаденопатией и прорастанием во втором межреберье справа. Отбрасывая статистическую вероятность и основывая мой вывод скорее на рентгенологической очевидности, я отдаю небольшое диагностическое предпочтение злокачественной мезотелиоме перед раком бронхов».

Анатомический диагноз оказался следующим: 1) мезотелиома плевры справа с прорастанием в средостение, грудную клетку и перикард; 2) пульмональный асбестоз; 3) двусторонний плевральный фиброз и известкование.

Глава 2

НЕКОТОРЫЕ ИССЛЕДОВАНИЯ ПО ИСПОЛЬЗОВАНИЮ ТЕОРЕМЫ БАЙЕСА И ВЫЧИСЛИТЕЛЬНЫХ МАШИН В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ

Проблемы заболевания являются более сложными и трудными, чем любые другие, которые приходится решать тренированному уму.

Сэр Вильям Ослер¹⁾

2.1. ВВЕДЕНИЕ

Доктор С. Кэбот, один из основателей клинико-патологических конференций, назвал дифференциальную диагностику «очень опасным делом — опасным для репутации врачебной мудрости. Я полагаю, именно из-за этой опасности так мало было написано по дифференциальной диагностике и так много по общей». Замечание, сделанное доктором Кэботом, остается верным и сегодня; возможно, оно только усилено опасностями, связанными с обсуждением дифференциальной диагностики при использовании математических моделей и вычислительных машин.

В сериях интересных статей под общим названием «Медицинская диагностика: настоящее, прошлое и будущее» Энгл и Дэвис [22—24] указали, что диагностика является настолько установившейся частью медицинской практики, что лишь немногие медики задумываются над вопросами: существует ли общее соглашение относительно того, что называть диагностикой; является ли диагностика искусством, наукой или и тем и другим одновременно; является ли диагноз основной предпосылкой в проблеме лечения и даже являются ли диагнозы четко определенными классами заболеваний?

Энгл заметил также: «Диагностика тесно связана с тем, что мы называем заболеванием. В настоящее время нет единой концепции заболевания и, следовательно, нет никакого единого и в то же время гибкого и практического понятия диагностики. Важно иметь в виду некоторые обстоятельства, связанные с понятием «заболевание», которые создают особенно трудные проблемы в диагностике. Во-первых, не все заболевания имеют общепринятые определения. Во-вторых, степень уверенности, с которой врач может поставить диагноз, значительно изменяется от заболевания к заболеванию, так как не все заболевания конкретны и четко отличаются одно от другого. В-третьих, наши определения некоторых заболеваний непрерывно

¹⁾ On the educational value of the medical society, in: *Aequanimitas with Other Addresses*; 2nd. ed., Philadelphia, Blakiston, 1930, p. 331.

изменяются, в результате чего не существует фиксированного числа заболеваний. Наконец, проявления некоторых даже хорошо описанных заболеваний также изменяются».

Приведенные выше замечания выделяют круг сравнительно старых проблем, связанных с медицинской диагностикой. Мы действительно нуждаемся в лучшем определении признаков, симптомов и заболеваний, и нам хотелось бы лучше понять логические процессы диагностики. Поэтому в последнее время усилия исследователей были направлены на то, чтобы дать ответы на эти вопросы с помощью математических моделей и вычислительных машин. Эти исследования, по-видимому, группируются в двух основных областях, ни одна из которых не имеет достаточно четкой границы.

В одной области исследований, которую можно назвать исследованием процесса постановки диагноза, основное внимание уделяется логическому процессу клинициста. Мы рассматриваем этот вопрос в следующей главе, используя деревья решений и блок-схемы. Другая область исследований связана главным образом с усовершенствованием определений симптомов, признаков и заболеваний и отысканием функциональных связей между ними. Эта область стала известна как исследования по машинной диагностике; она и является предметом настоящей главы.

Упомянутые две области дополняют друг друга, поскольку исследования диагностического процесса представляют собой попытки понять и предсказать то, что может сделать человек, ставший диагнозом, и поскольку исследования по машинной диагностике определяются как попытки сконструировать математические или статистические модели медицинской диагностики, которые обеспечат наиболее точное выполнение этой работы. Отыскание функциональных связей между признаками, симптомами и диагнозами при изучении машинной диагностики может дать ключ к лучшим методам постановки диагноза человеком. Аналогично лучшее понимание процессов человеческого суждения при диагностике может помочь исследователю создать более точную машинную диагностическую логику. Интересные статьи, в которых обсуждаются диагностические модели и сам диагностический процесс, можно найти в трудах конференции, проведенной в Мичиганском университете в мае 1963 г. [41, 42].

Некоторые из ранних попыток выразить связи типа симптом — заболевание в наиболее удобной форме завершились разработкой таблиц, которые использовал врач, отыскивая наилучшие совпадения между симптомами пациента и той или иной таблицей. Английский врач Ф. А. Нэш [72] описал механическое устройство, напоминающее счетную линейку, которое дает возможность медику сравнивать различные комбинации из восьмидесяти двух признаков и симптомов для того, чтобы выбрать наиболее приемлемый диагноз из возможных 337 заболеваний. Практическую ограниченность устройств такого типа легко оценить, если принять во внимание, что

даже для одной узкой области, такой, как офтальмология, имеется (согласно данным французского офтальмолога Пэша [83, 84]) около тысячи «болезней» и около двух тысяч признаков и симптомов только для роговой оболочки глаза. Поэтому Пэша и другие исследователи, например Липкин и Харди [60] и Эрдманн [25], начали разработку механических методов, предусматривающих использование перфокарт. Сначала применялись карты с краевой перфорацией и спицы, позже — перфокарты и сортировальные машины. Эти системы были по существу детерминистскими и использовали процессы сравнения. Разработки в области медицинской диагностики требуют рассмотрения большого числа переменных, поэтому заметного прогресса в этом направлении нельзя было ожидать до тех пор, пока исследователи в этой области медицины не получили возможность использовать электронные вычислительные машины. Это произошло в начале 60-х годов.

Для целей диагностики с помощью вычислительных машин были применены процедуры различных типов. Они включают метод дискриминантных функций, факторный анализ, машинные методы сравнения и модели условной вероятности. Примеры процедур каждого типа описаны в работах, перечисленных в приложении 1.

В настоящее время невозможно сконструировать достаточно общий алгоритм (т. е. написать программу вычислительной машины для любого диагностического процесса), но для некоторых хорошо определенных областей медицины эта задача может быть решена весьма успешно. Условно-вероятностные модели сейчас усиленно изучаются, и мы отобрали работы трех исследователей, чтобы проиллюстрировать использование метода условной вероятности в диагностике заболеваний щитовидной железы, врожденных пороков сердца и опухолей кости. Общим для этих исследований является хорошо описанная группа заболеваний, для которых проводится диагностика, хорошо определенное множество признаков, симптомов и лабораторных исследований и прочная статистическая база, полученная из хорошо отобранной популяции.

2.2. МАШИННАЯ ДИАГНОСТИКА ЗАБОЛЕВАНИЙ ЩИТОВИДНОЙ ЖЕЛЕЗЫ

Оверолл с соавторами [76—82] подробно описал результаты применения нескольких математических моделей, разработанных для диагностики функции щитовидной железы по соответствующим признакам, симптомам и лабораторным данным больного. Мы включили полный список ссылок, потому что Вильямс и Оверолл являются единственными исследователями, которые опробовали несколько различных типов диагностических моделей и сравнивали их эффективность. Байесовская условно-вероятностная модель, которая была выбрана в конечном итоге, дала правильную классификацию состо-

яния тиреоидного обмена (т. е. успешно предсказала терапевтический результат) для 97,2% последних пятисот пациентов с подозрением на заболевание щитовидной железы, которых обследовали в медицинском колледже Университета Флориды [32].

Полезно рассмотреть возможные аргументы в пользу выбора условно-вероятностной модели по сравнению с другими моделями, например с методом дискриминантных функций. На первоначальный выбор модели может оказать влияние точка зрения исследователя на проблему диагностики. Так, если он чувствует, что диагностика является по существу задачей классификации, он может выбрать дискриминантный анализ, который был предложен Р. Фишером [30] для проблем таксонометрии. Постановка задачи такова. Дан индивидуум, у которого проведено p измерений t_1, t_2, \dots, t_p , возможно, сильно коррелированных между собой, и нужно решить на основе этих измерений, к какой из двух возможных групп принадлежит индивидуум.

Первоначальный метод Фишера был выведен без учета вероятности ошибочной классификации, и, поскольку ошибки неправильной классификации являются важным аспектом медицинской диагностики, необходимо дальнейшее рассмотрение этого вопроса. При этом обсуждении мы будем следовать доводам, выдвинутым Корнфилдом [12].

При каких обстоятельствах исходный метод Фишера приводит к минимальной ошибке классификации? Ответ может быть получен прямым приложением результата, известного в литературе по математической статистике как лемма Неймана — Пирсона [74]. Существуют два возможных типа ошибочной классификации. Индивидуум, принадлежащий к группе 1, может быть неправильно отнесен к группе 2. Обозначим вероятность такой классификации символом α . Аналогично обозначим через β вероятность отнесения к группе 1 индивидуума, принадлежащего к группе 2. Лемма Неймана — Пирсона указывает при очень общих условиях, как минимизировать β при заданном значении α . Это требует рассмотрения многомерных функций плотности вероятностей.

Предположим, что имеется единственная переменная t_1 , которая выражает уровень щавелево-уксусной трансминазы в сыворотке крови (SGOT). В каждой из двух изучаемых популяций, скажем популяции пациентов с заболеванием печени и популяции здоровых людей, мы не можем ожидать, что все индивидуумы имеют одинаковое значение t_1 . Мы наблюдаем два распределения частот относительно t_1 , по одному для каждой популяции, различающихся своим расположением и формой и имеющих некоторую область пересечения. Распределение частот для t_1 в 1-й популяции имеет математическое представление в виде функции плотности вероятности, скажем $f_1(t_1)$, причем $f_1(t_1)h_1$ — приближенное выражение для относительной части популяции со значением измеряемой величины между

t_1 и t_1+h_1 . Для двух измерений мы имеем двумерную функцию плотности вероятности $f_1(t_1, t_2)$, где $f_1(t_1, t_2) h_1 h_2$ — приближенная часть популяции со значением первой измеряемой величины между t_1 и t_1+h_1 и второй измеряемой величины между t_2 и t_2+h_2 . Распространение этих определений на p измерений очевидно, и мы будем обозначать многомерные функции плотности вероятности вектора t_1, t_2, \dots, t_p для двух популяций через $f_1(t)$ и $f_2(t)$.

Лемма Неймана — Пирсона утверждает, что для минимизации значения β при данном α нужно вычислить специальную функцию измерений t , а именно отношение правдоподобия $L(t)$:

$$L(t) = \frac{f_1(t)}{f_2(t)}. \quad (1)$$

Тогда индивидуум должен быть отнесен к популяции 1, если $L(t)$ больше или равно некоторой константе c , и к популяции 2, если $L(t)$ меньше этой константы. Константа c определяется условием, что для индивидуума, принадлежащего к популяции 1, имеет место равенство

$$\text{Вероятность } [L(t) < c] = \alpha. \quad (2)$$

Чем больше значение линейного дискриминанта, тем больше мы уверены, что индивидуум принадлежит к популяции 2; чем меньше его значение, тем более вероятно, что он принадлежит к популяции 1. Но существуют промежуточные значения, которые вызывают сомнения при классификации. Это возвращает нас к рассмотрению не задачи классификации, а вероятности того, что при данном t пациент принадлежит к популяции 2. Если априорная вероятность этого события равна P_0 , то апостериорная вероятность, скажем P_2 , определяется теоремой Байеса

$$P_2 = \frac{f_2(t) P_0}{P_0 f_2(t) + (1 - P_0) f_1(t)} = \left[1 + \frac{L(t)(1 - P_0)}{P_0} \right]^{-1} = \left[1 + L(t) \frac{1}{\Omega_0} \right]^{-1}. \quad (3)$$

Таким образом, апостериорная вероятность зависит от отношения правдоподобия $L(t)$ и отношения априорных вероятностей или априорных шансов. С этим вопросом мы уже встречались в гл. 1 в связи с уравнениями (9) — (11), и вычисления вероятностей для синдрома Кушинга, рассмотренные в гл. 1 (разд. 1.3), были сделаны на основе уравнения, имеющего вид уравнения (3) этой главы. Согласно Корнфилду [11, 12], классификация является, таким образом, особым случаем, когда все индивидуумы, у которых апостериорные вероятности превышают определенное значение, относятся к популяции 2. Вид уравнения (3) также предполагает, что диагностические процессы, основанные либо на отношениях правдоподобия, либо на апостериорных вероятностях, тесно связаны, а не диаметрально противоположны, как иногда считают [9]. Обобщение дискриминационной задачи различения между двумя группами на случай различ-

ния между тремя и более группами, а также особенности, связанные с большим числом переменных, также рассмотрены Корнфилдом [12].

Вильямс и его сотрудники подготовили машинную программу для диагностики заболеваний щитовидной железы, которая использует байесовскую условно-вероятностную модель типа рассмотренной в гл. 1. Благодаря любезности Л. Т. Фитцджералда и К. М. Вильямса мы имеем возможность обсудить эту программу и привести полное изложение этой диагностической системы (см. приложение 3).

При осмотре пациента в радиоизотопной лаборатории медицинского колледжа Университета Флориды врач собирает данные о признаках, симптомах и результатах лабораторных исследований в соответствии с перечнем, приведенным ниже:

Возраст	Летаргические движения	Величина основного обмена
Пол	Чувствительность к холоду	СБИ ¹⁾ , ЭБИ ²⁾ или T_4
Раса	Пониженная потливость	6-часовое накопление I^{131}
Нервозность	Пониженный аппетит	24-часовое накопление I^{131}
Чувствительность к теплу	Увеличение веса	Хроматографическая фракция
Повышенная потливость	Увеличенная летаргия	иодсодержащих аминокислот
Повышенный аппетит	Сухая жесткая кожа	Уровень связывающего глобулина щитовидной железы
Потеря веса	Отечность лица	
Гиперкинез	Боли в щитовидной железе	
Теплая влажная кожа	Глазные признаки	
Легкое дрожание пальцев		

Эти данные переносятся из специальной таблицы истории болезни пациента на перфокарты (разметка перфокарт приведена в приложении 3) и вводятся в вычислительную машину. При помощи программы, представленной в блок-схемах (приложение 3), и матрицы условных вероятностей машина вычисляет вероятность пониженной функции щитовидной железы, нормальной функции щитовидной железы или гиперфункции щитовидной железы — сначала только на основании признаков и симптомов, затем только на основании результатов лабораторных исследований и, наконец, на основании всех имеющихся данных. Все три типа результатов выводятся на печать, как показано на фиг. 8.

Как отмечают Вильямс и его сотрудники, существенное свойство программы заключается в том, что для каждого пациента с подозрением на заболевание щитовидной железы автоматически рассматривается возможность замаскированной пониженной функции щитовидной железы. Программа также включает накапливающую матрицу, в которую можно вводить признаки, симптомы и результаты

¹⁾ СБИ (PBI) — связанный с белком иод сыворотки.

²⁾ ЭБИ (BEI) — экстрагируемый бутанолом иод сыворотки.

Вводится 0, если признак отсутствует, и 1, если признак присутствует. Если данных нет, то в соответствующем месте оставляется пропуск.

Столбец	Данные	Столбец	Данные
1—5	Идентификация пациента Номер Фамилия пациента		
6—7	Возраст	8—	Пол (0 — мужской, 1 — женский)
9—	Раса (0 — белый, 1 — цветной)	10—	
11—	Недавнее возникновение нервозности	19—	Недавнее возникновение летаргии
12—	Недавнее усиление чувствительности к теплу	20—	Недавнее возникновение чувствительности к холоду
13—	Недавнее возникновение повышенной потливости	21—	Недавнее возникновение пониженной потливости
14—	Недавнее возникновение повышенного аппетита	22—	Недавнее понижение аппетита
15—	Недавнее понижение веса	23—	Недавнее повышение веса
16—	Гиперкинез	24—	Летаргические движения
17—	Теплая влажная кожа	25—	Сухая жесткая кожа
18—	Легкое дрожание пальцев	26—	Отечность лица
27—	Глазные признаки (опущение века, лагофтальм, экзофтальм)	28—	Недавнее возникновение болей или припухлости в щитовидной железе

Столбец	Данные	Здесь проставляются точные лабораторные данные
29	Основной обмен (%)	< 40 (1); - 40 ÷ 0 (2); 0 ÷ + 20 (3); + 21 ÷ + 45 (4); > + 45 (5)
30	СБИ (%)	< 2,3 (1); 2,3 ÷ 4,0 (2); 4,1 ÷ 7,0 (3); 7,1 ÷ 9,6 (4); > 9,6 (5)
	ЭБИ или T_4 (%)	< 1,8 (1); 1,8 ÷ 3,2(2); 3,3 ÷ 5,6 (3); 5,7 ÷ 7,7 (4); > 7,8 (5)
31	6-часовое накопление I^{131} (%)	< 4 (1); 4 ÷ 6 (2) (2); 7 ÷ 20 (3); 21 ÷ 37 (4); > 37 (5)
32	24-часовое накопление I^{131} (%)	< 6 (1); 6 ÷ 13 (2); 14 ÷ 34 (3); 35 ÷ 55 (4); > 55 (5)
33	Исследование хроматографической функции иодсодержащих аминокислот (%)	< 25 (1); 25 ÷ 27 (2); 27 ÷ 34 (3); 34 ÷ 36 (4); > 36 (5)
34	Уровень связывающего глобулина щитовидной железы (в T_4 %)	< - 12 (1); 12 ÷ 20 (2); > 20 (3)
35	Принимал ли пациент в течение последней недели антитиреоидные лекарства (тиоурацил, итрумил, пропилтиоурацил, тапазол и т. д.), тиоцианаты, перхлораты, сульфанамиды, изонацид, PAS? 0 — нет, 1 — да.	
36	Назначались ли пациенту в течение последних двух недель высушенные препараты щитовидной железы, триодтирозин (цитомел), тироксин, глпак, медиопак, неопакс, урокон, миокон, тиксокон, ренографин? 0 — нет, 1 — да.	

Ф и г. 7.

Список данных по щитовидной железе для машинной диагностики с использованием программ IV и V [31].

Столбец	Данные
37	Назначались ли пациенту в течение последних трех недель какие-либо препараты йода, в том числе иодоформ Люголя, иодные мази или иодные препараты с витаминами или успокаивающими веществами? 0 — нет, 1 — да.
38	Принимал ли пациент в течение последних трех лет диодраст, холографин, телепак, приодакс, дионозил, сальпикс? 0 — нет, 1 — да.
39	Принимал ли пациент в течение последнего года липодол, висциодол, этиодол, понтопак, иодохлорол? 0 — нет, 1 — да.
40	Ставили ли ранее пациенту диагноз заболевания щитовидной железы? 0 — нет, 1 — да.
41—43	Если ответ на последний вопрос положительный, проставить числа, указанные в скобках: разлитый зоб (250), нетоксичный узловатый зоб (251), тиреотоксикоз с зобом или без зоба (252), микседема и кретинизм (253), другие заболевания (тиреоидит) (254), карцинома (194).
44	Обнаружена ли у пациента при пальпации и (или) сканировании нормальная железа или нет? 0 — нормальная, 1 — ненормальная. Если железа нормальная, то на следующие четыре вопроса отвечать не следует.
45	Обнаружено ли при пальпации или сканировании диффузное увеличение железы? 0 — нет, 1 — да.
46	Обнаружены ли при пальпации и (или) сканировании множественные узлы железы? 0 — нет, 1 — да.
47	Обнаружен ли при пальпации и (или) сканировании одиночный узел железы? 0 — нет, 1 — да.
48	Если имеется одиночный узел, то накапливает ли он I^{131} меньше, чем остальная часть железы (тогда поставить 1), или он обладает такой же активностью, как и остальная железа (тогда поставить 0)?
49	Если имеется одиночный узел, то накапливает ли он I^{131} больше, чем остальная часть железы (1 — да)?
50	Соответствует ли результат сканирования послеоперационным данным? 0 — нет, 1 — да.
51	Соответствует ли результат сканирования данным после тиреоидита? 0 — нет, 1 — да.

Фиг. 7. Продолжение

лабораторных исследований пациента после того, как диагноз установлен. Редкое заболевание, такое, как врожденное отсутствие связывающего глобулина щитовидной железы, может быть довольно просто включено в машинную программу при помощи соответствующей подпрограммы [32].

Одним из основных процессов при организации машинной диагностики является проверка первоначальной матрицы при помощи другой группы больных с верифицированными диагнозами. Для системы диагностики заболеваний щитовидной железы исходная матрица была построена на основании данных о 879 пациентах, которые были разделены на три группы в результате наблюдения за их реакцией на лечение в течение шести месяцев; к этим группам были отнесены соответственно пациенты, имеющие пониженную функцию щитовидной железы, нормальную функцию щитовидной железы или гиперфункцию щитовидной железы. Матрица была проверена путем пропускания через систему двух аналогичным образом классифицированных групп в количестве 268 и 210 больных.

Виды обследования		Вероятность, %	Признаки, имеющиеся у больного
По признакам и симптомам	Пониженная функция щитовидной железы	0	17
	Нормальная функция щитовидной железы	100	
	Гиперфункция щитовидной железы	0	
По лабораторным данным	Пониженная функция щитовидной железы	0	2
	Нормальная функция щитовидной железы	9	
	Гиперфункция щитовидной железы	91	
По всем данным	Пониженная функция щитовидной железы	0	
	Нормальная функция щитовидной железы	0	
	Гиперфункция щитовидной железы	100 (замаскиро- ванная)	
	Отсутствие заболевания щитовидной железы	0	
	Разлитый зоб (250)	0	
	Нетоксичный узловатый зоб (251)	0	
	Тиреотоксикоз (252)		100
	Микседема и кретинизм (253)		0
	Другие заболевания (тиреоидит) (254)		0
	Карцинома щитовидной железы (194)		0

Ф и г. 8.

Машинный диагноз функционального состояния щитовидной железы и дифференциальный диагноз для пациента с кодом 50 000 [32].

Среди преимуществ машинной диагностики авторы работы [32] отмечают экономический фактор: «В процессе диагностики заболеваний щитовидной железы существует некоторая оптимальная последовательность лабораторных исследований, различающихся по стоимости, сложности и возможной опасности для пациента». Определение оптимальных последовательностей диагностических исследований для различных областей медицины является важной задачей будущих разработок.

2.3. МАШИННАЯ ДИАГНОСТИКА ВРОЖДЕННЫХ ПОРОКОВ СЕРДЦА

В 1961 г. Уорнер с соавторами [124] опубликовал результаты исследований по использованию теории Байеса и цифровых вычислительных машин для диагностики врожденных пороков сердца. В течение пяти лет Уорнер и его сотрудники накопили значительный

опыт машинной диагностики. Имея это в виду, интересно разобрать их работу и отметить некоторые более поздние их высказывания.

При разработке системы были рассмотрены пятьдесят три симптома и тридцать пять заболеваний [125]. Их перечень приведен на фиг. 9 и 10.

Для диагностики используется расширенная формула Байеса [127]

$$P_{(D_i | a_1, a_2, \dots, a_n)} = \frac{P_{D_i} \prod_{j=1}^{j=n} (P_{(S_j | D_i)})^{a_j} (1 - P_{(S_j | D_i)})^{1-a_j}}{\sum_{i=1}^{i=k} (\text{Числитель})_i}, \quad (4)$$

где $a_j=1$, если имеет место j -й симптом, и $a_j=0$, если j -й симптом отсутствует.

Уравнение (4) утверждает, что вероятность того, что пациент страдает заболеванием D_i , если он имеет набор симптомов a_1, a_2, \dots, a_n , где a_j принимает значения 1 или 0 в зависимости от наличия или отсутствия симптомов S_j , равна априорной вероятности заболевания D_i в рассматриваемой популяции, умноженной на произведение членов, связанных с каждым из симптомов. Каждый такой член P_{S_j} при данном D_i есть вероятность j -го симптома, встречающегося при заболевании D_i . В этом случае в расчетную формулу входит величина P_{S_j} , так как показатель степени другого сомножителя (1 минус вероятность) становится равным нулю и, следовательно, сам сомножитель равен единице. С другой стороны, если симптом отсутствует, величина a_j равна нулю, и при вычислении будет использовано дополнение к вероятности (второй сомножитель).

Врач не всегда может установить наличие или отсутствие симптома с уверенностью, необходимой для исключаящего решения, т. е. для выбора нуля или единицы. Поэтому Уорнер предложил следующее уравнение:

$$P_{(D_i | a_1, a_2, \dots, a_n)} = \frac{P_{D_i} \prod_{j=1}^{j=n} [a_j P_{(S_j | D_i)} + (1 - a_j) (1 - P_{(S_j | D_i)})]}{\sum_{i=1}^{i=k} (\text{Числитель})_i}, \quad (5)$$

где a_j — вероятностная оценка каждого симптома.

В уравнении такого вида величины a не обязательно должны быть только единицей или нулем, а могут принимать любое значение от нуля до единицы и представляют собой вероятностную оценку врача, выражающую его уверенность относительно наличия или отсутствия симптома у больного. Если a равно 0,5, что указывает на

равную вероятность наличия и отсутствия симптома, то этот симптом не будет влиять на диагноз, поскольку вероятность его и дополнение к этой вероятности имеют равный вес. Преимущество уравнения (5) состоит в том, что оно позволяет использовать любые данные (симптоматику), которые могут быть представлены с вероятностной шкалой от нуля до единицы. Так, Уорнер сообщает, что в его лаборатории используется машинная программа, которая вычисляет вероятность того, что данная электрокардиограмма принадлежит к одной из десяти выделенных электрокардиографических форм. Данные ЭКГ и результаты этих вычислений объединяются при помощи уравнения (5) с оцениваемыми врачом вероятностями для симптоматики, извлекаемой из истории болезни, физического и рентгеновского обследования. Таким образом, при помощи формулы (5) производятся окончательные вычисления для дифференциальной диагностики врожденных пороков сердца.

В этой математической модели медицинской диагностики приняты определенные предположения. Во-первых, Уорнер допустил, что симптомы не зависят друг от друга при данном заболевании. Это предположение мы уже обсуждали в гл. 1 (разд. 1.3). Второе предположение, сделанное Уорнером, состоит в том, что рассматриваемые заболевания являются взаимоисключающими. Это предположение также было рассмотрено нами в гл. 1 (разд. 1.3). В перечне симптомов (фиг. 9) скобки отделяют группы взаимоисключающих симптомов так, что в каждом наборе может присутствовать не более одного симптома из группы. Правила действий для групп исключающих друг друга симптомов мы проиллюстрируем на симптомах 5, 6, 7 и 8, которые представляют различные формы цианоза. Если у пациента обнаружен любой из этих симптомов, то по определению у него не может быть никакого другого симптома из этой группы. Таким образом, было бы ошибкой использовать дополнение к вероятности легкого цианоза как добавочную информацию, если у пациента обнаружен резкий цианоз, так как последний с необходимостью исключает легкую форму. В терминах математической модели, если присутствует симптом 5, то $S_5=1$, $S_6=-1$, $S_7=-1$, $S_8=-1$.

Наконец, Уорнер принял, что случайный механизм появления симптомов при каждом из заболеваний стационарен (т. е. не изменяется со временем). Если в определенной области применения этот случайный механизм изменяется со временем, то матрица вероятностей симптомов — заболевания должна уточняться через интервалы времени, достаточно короткие по сравнению с периодом колебаний механизма образования распределения симптоматики для рассматриваемого заболевания. Например, симптомы пневмонии частично зависят от этиологического агента, и характеристики этого агента могут изменяться с изменением средств антибиотической терапии.

Список симптомов¹⁾

1	1. Априорная частота	—24.	Систолический шум во 2-м межреберье слева
		—25.	Пансистолический шум во 2-м межреберье слева
Возраст		—26.	Мезосистолический шум во 2-м межреберье слева
— 2.	[Менее 1 года 1—20 лет Более 20 лет	—27.	Непрерывный шум во 2-м межреберье слева
— 3.		—28.	Систолический шум во 2-м межреберье справа
— 4.		—29.	Диастолический шум во 2-м межреберье справа
	Данные физического обследования	—30.	Протодиастолический шум
— 5.	[Цианоз умеренный Цианоз сильный (с пальцами в виде барабанных палочек)	—31.	Диастолический шум в сочетании с непрерывным
— 6.		—32.	Шум громче 3/6 (10 мм) ²⁾
— 7.	Цианоз перемежающийся	—33.	[Усиленный легочный компонент тона II Ослабленный легочный компонент тона II
— 8.	Цианоз дифференциальный	—34.	
— 9.	Положение больного «на корточках»	—35.	Фиксированное расщепление тона II
—10.	Пульс на бедренной артерии слабее, чем на плечевой		
	Фонокардиограмма		Данные ЭКГ
—11.	[Систолический шум на верхушке Пансистолический шум на верхушке Мезосистолический шум на верхушке	—36.	Фибрилляция предсердия или большая зубчатая волна P
—12.		—37.	[Отклонение оси вправо (более 110°) Отклонение оси влево (менее 0°)
—13.		—38.	
—14.	[Диастолический шум на верхушке Протодиастолический шум на верхушке	—39.	Зубец R больше 1,2 мв в отведении V ₁
—15.		—40.	rR' или qR в отведении V ₁
—16.	Поздний диастолический шум на верхушке	—41.	Зубец R больше 2,5 мв в отведении V ₆
—17.	[Систолический шум в 4-м межреберье слева Пансистолический шум в 4-м межреберье слева Мезосистолический шум в 4-м межреберье слева	—42.	Инверсия волны T в отведении V ₆
—18.			Рентгенологические данные
—19.		[Непрерывный шум в 4-м межреберье слева Диастолический шум в 4-м межреберье слева	—43.
—20.	—44.		[Периферические сосуды увеличены Периферические сосуды уменьшены
—21.	[Пандиастолический шум в 4-м межреберье слева Протодиастолический шум в 4-м межреберье слева	—45.	

¹⁾ Квадратными скобками охвачены взаимоисключающие признаки.

²⁾ Отклонение фонокардиограммы 10 мм при чувствительности 1/5.

Фиг. 9.

Список симптомов для диагностики врожденных пороков сердца [125].

- | | | | |
|------|--------------------------------------|------|--------------------------|
| —46. | [Усиленный рисунок корней | —50. | [Увеличение аорты |
| —47. | [Ослабленный рисунок корней | —51. | [Уменьшение аорты |
| —48. | [Главная легочная артерия увеличена | —52. | Кардиомегалия |
| —49. | [Главная легочная артерия не видна | —53. | Сердце в форме восьмерки |

Фиг. 9. Продолжение

Список болезней

- | | |
|--|--|
| 1. Норма | 18. Незаращение боталлова протока |
| 2. Дефект межпредсердной перегородки | 19. Легочный артерио-венозный свищ |
| 3. Дефект межпредсердной перегородки со стенозом легочной артерии | 20. Врожденная митральная болезнь |
| 4. Дефект межпредсердной перегородки с легочной гипертензией | 21. Первичное заболевание миокарда |
| 5. Атриовентрикулярное сообщение | 22. Аномальное отхождение коронарных артерий |
| 6. Частичное аномальное впадение легочных вен | 23. Врожденное заболевание аорты |
| 7. Полное аномальное впадение легочных вен | 24. Дефект межжелудочковой перегородки с легочным кровотоком $\leq 1,4$ системного кровотока |
| 8. Трикуспидальная атрезия (без транспозиции) | 25. Коарктация аорты |
| 9. Болезнь Эбштейна | 26. Общий артериальный ствол |
| 10. Межжелудочковый дефект с клапанным стенозом легочной артерии | 27. Транспозиция магистральных сосудов |
| 11. Межжелудочковый дефект с инфундибулярным стенозом легочной артерии | 28. Гипертрофический субаортальный стеноз |
| 12. Клапанный стеноз легочной артерии с градиентом 40 мм | 29. Отсутствие дуги аорты |
| 13. Инфундибулярный стеноз легочной артерии с градиентом 40 мм | 30. Дефект межжелудочковой перегородки с легочным кровотоком $> 1,4$ системного кровотока |
| 14. Атрезия легочной артерии | 31. Дефект межжелудочковой перегородки с легочной гипертензией |
| 15. Периферический стеноз легочной артерии | 32. Незаращение боталлова протока с легочной гипертензией |
| 16. Легочная гипертензия ¹⁾ | 33. Трикуспидальная атрезия с транспозицией |
| 17. Аорто-пульмональный свищ | 34. Стеноз легочной артерии с градиентом < 40 мм Hg |
| | 35. Разрыв синуса Вальсальва |

¹⁾ Среднее давление в легочной артерии больше или равно среднему системному артериальному давлению.

Фиг. 10.

Список заболеваний для диагностики врожденных пороков сердца [125].

Программа Уорнера, которая включает аналоговые данные ЭКГ и другую информацию, использовалась при изучении качества врачебной и машинной диагностики [119, 125, 126]. Программа вычислительной машины и врач начинают с одних и тех же наблюдений, так что при их сравнении оценивается логика, а не точность наблюдений.

Для оценки качества диагностики как врача, так и вычислительной машины, использовали две статистические меры. Первая представляла собой относительное число случаев, в которых вычислительная машина или врач поставили правильный диагноз (он подтверждался последующими исследованиями). Второй мерой качества диагностики являлась средняя величина вероятности, придаваемая правильному диагнозу программой вычислительной машины или врачом.

Уорнер пришел к заключению, что различия врачебной интерпретации входных данных представляют собой наиболее существенный источник ошибок. Ему также удалось показать, что наблюдатель делает два вида ошибок: ошибки в распознавании симптомов и логические ошибки в результате получения неправильных заключений из точных данных. Обнадеживающим фактом является то, что через некоторый период времени качество работы как программы вычислительной машины, так и врачей улучшается, насколько можно об этом судить по упомянутым выше двум мерам качества [125].

2.4. МАШИННАЯ ДИАГНОСТИКА ПЕРВИЧНЫХ ОПУХОЛЕЙ КОСТИ

Для большинства рентгенологов наблюдение пациента с первичной костной опухолью может оказаться редким событием, и поэтому врачу трудно накопить даже в течение всей жизни значительный опыт по рентгенологической диагностике костных опухолей. Работы Лодвика [61, 62, 64] по рентгенологической диагностике первичных костных опухолей на вычислительной машине показывают, что можно создать систему, которая окажет рентгенологу существенную помощь, когда он сталкивается с этой проблемой.

Постановка диагноза на вычислительной машине требует систематизированного анализа врачом рентгенограммы костной опухоли. Система должна быть внутренне последовательной и удовлетворять дополнительным требованиям, состоящим в том, что она может быть легко освоена и использована другими врачами. Врач должен сам осуществлять просмотр рентгеновского снимка, так как в настоящее время еще невозможен автоматический анализ рентгенограмм

при помощи комбинированного электронного сканирующе-вычислительного устройства. Как мы указывали в лекции Мемориального фонда в 1959 г. [66], в принципе можно создать устройство со сканирующим электронным лучом для считывания рентгенограммы грудной клетки; однако, после того как получены данные с пленки, не известно, какие операции нужно выполнять над ними, чтобы осуществлять различение между нормальными и аномальными снимками грудной клетки. Проблемы распознавания образов все еще остаются слишком трудными. Тем не менее автоматический анализ рентгенограмм остается важной целью, к достижению которой прилагаются энергичные усилия, и для этого есть веские причины. Это интересная в научном отношении область, имеющая, кроме того, большое экономическое значение. Увеличивающийся объем рентгенологических исследований и относительное сокращение рентгенологов требуют разработки устройства для анализа рентгенограмм, которое могло бы осуществлять по крайней мере отделение патологических случаев от нормальных. Тогда врач мог бы сконцентрировать свое внимание в большей степени на интерпретации рентгенограмм патологических случаев. Но прежде чем автоматический анализ рентгенограмм сможет стать реальностью, нужно предпринять некоторые промежуточные шаги, такие, как подробное описание рентгенограмм в терминах типа изображений, описательных слов и измерений. Работа Лодвика является некоторым продвижением в этом направлении.

Система, используемая Лодвиком, основана на методе анализа костных опухолей, развитом Лодвиком и Джонсоном из ортопедического отдела Института патологии Вооруженных сил. Авторы признают, что в одной опухоли может быть представлено много типов клеток, и предлагают систему классификации по преобладающему типу клеток. Этот подход к классификации приводит к следующим основным типам опухолей:

- остеосаркома,
- фибросаркома,
- хондросаркома,
- хондробластома,
- хондромиксоидная фиброма,
- хондрома,
- гигантоклеточная опухоль,
- саркома ретикулярных клеток,
- полиморфная саркома,
- миелома,
- саркома Юинга,
- саркома надкостницы,
- остеохондрома.

Лодвик и Тернер [63] издали брошюру, в которой представлена специальная форма для рентгенологического анализа (фиг. 11) и

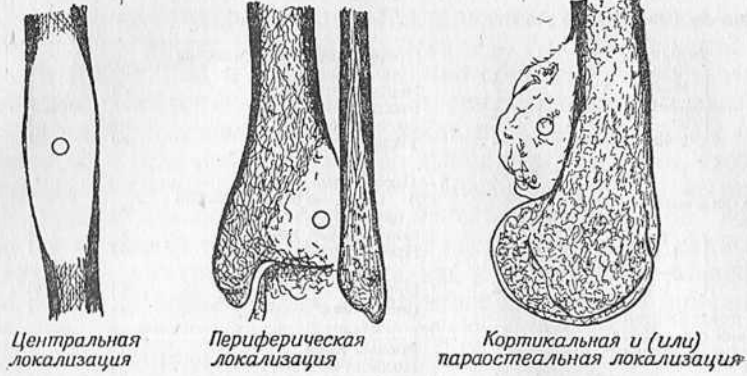
Номер случая (1-6)				4. Минерализация межклеточного вещества опухоли	Столбец	Да	Нет	
1. Клинические данные								
Пол	Мужчина	7	1 0	Пропускает рентгеновские лучи	35	1	0	
Возраст	0-9	8	1 2	Непрозрачное	36	1	0	
	10-19	9	1 2	Рыхлое	37	1	0	
	20-29	10	1 2	Бугристое	38	1	0	
	30-39	11	1 2	Твердое	39	1	0	
	>40	12	1 2					
Задержка диагноза в месяцах				5. Деструкция кости				
	0	13	1 2	В виде «географической карты»	40	1	0	
	1-2	14	1 2	«Изъеденная молью»	41	1	0	
	3-11	15	1 2	Проникающая	42	1	0	
	>12	16	1 2	Перелом	43	1	0	
5-летняя продолжительность жизни после установления диагноза				Смещение	44	1	0	
		17	1 0	Ровный край	45	1	0	
2. Размер				Дольчатый край	46	1	0	
	0-30 мм	18	1 2	Неровный край	47	1	0	
	31-60	19	1 2	Сглаженный край	48	1	0	
	61-90	20	1 2	Острый край	49	1	0	
	91	21	1 2	Смазанный край	50	1	0	
		22	1 2	Инвазия края 1 см	51	1	0	
				Инвазия края более 1 см	52	1	0	
3. Локализация				6. Пенетрация поверхностного слоя				
	Центральная	23	1 2	Нет	53	1	2	
	Эксцентрическая	24	1 2	Частичная	54	1	2	
	Надкостница или поверхностный слой		25	1 2	Полная	55	1	2
		Таз или крестец	26	1 2	6. Проплиферация кости	56	1	0
	Другие плоские кости, включая ребра, череп и нижнюю челюсть		27	1 2	Склеротическая щель	57	1	0
		Небольшие кости, включая надколенник	28	1 2	В виде крапинок	58	1	0
	Трубчатые кости, включая ключицу		29	1 2	Эндостальная	59	1	0
		Подразделение трубчатых костей			Гиперостальная	60	1	0
	Суставная поверхность	30	1 0		Трабекулярная	61	1	0
	Эпифиз	31	1 0		Опорная	62	1	2
Надкостница	32	1 0		Поверхностный слой не увеличен в объеме	63	1	2	
Метафиз	33	1 0		Увеличен на 10 мм	64	1	2	
Диафиз	34	1 0		Увеличен на 11 мм и больше	65	1	2	
				Нет треугольника Кодмана	66	1	2	
				1 треугольник Кодмана	67	1	2	
				2 треугольника Кодмана	68	1	2	
				3 и более треугольников Кодмана	69	1	2	
				Нет периостаза	70	1	2	
				Пластинчатый периостаз	71	1	2	
				Аморфный периостаз	72	1	2	
				Нет игольчатости (гладкая)	73	1	2	
				Ровная	74	1	2	
				Щетинообразная	75	1	2	
				Бархатистая				

Фиг. 11.

Форма, используемая при градации и диагностике первичных костных опухолей при помощи вычислительной машины. В машине не учитываются данные ни о задержке в диагнозе, ни о 5-летнем сроке жизни пациента. Отрицания записываются как «0» в тех разделах, где отсутствие признака имеет существенное самостоятельное значение. Отрицания записываются как «2» для взаимоисключающих признаков [63].

ЛОКАЛИЗАЦИЯ

Относительно продольной оси



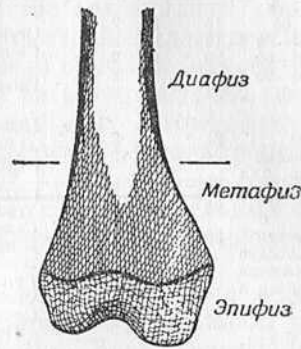
Центральная локализация

Периферическая локализация

Кортикальная и (или) параостеальная локализация

По отношению к тазу

Расположение вдоль костей



Ф и г. 12.

- а** — положение относительно продольной оси кости (выбрать один вариант):
- | | |
|---|--------|
| центральное | 13 1 2 |
| эксцентрическое | 14 1 2 |
| кортикальная и (или) параостеальная локализация | 15 1 2 |
- б** — относительно специфических костей (выбрать один вариант):
- | | |
|-------------------------------|-------------------------------|
| (16—17) рассматриваемая кость | 12 — нижняя челюсть |
| 01 — малая берцовая | 13 — верхняя челюсть |
| 02 — лучевая | 14 — лонная |
| 03 — ключица | 15 — ребра |
| 04 — лопатка | 16 — крестец |
| 05 — череп | 17 — большая берцовая |
| 06 — бедро | 18 — локтевая |
| 07 — позвоночник | 19 — кости кисти |
| 08 — плечо | 20 — кости стопы |
| 09 — подвздошная | 21 — кости туловища |
| 10 — седалищная | 22 — кости верхней конечности |
| 11 — коленная чашечка | 23 — кости нижней конечности |
- в** — подразделения, имеющие отношение ко всем видам кости (можно выбрать больше одного варианта):
- | | |
|-----------------|--------|
| эпифиз | 18 1 0 |
| пластинка роста | 19 1 0 |

тщательно отобранные иллюстрации, поясняющие важнейшие типы рентгенологических заключений (фиг. 12). Распознавание костных опухолей и их диагностика требуют, чтобы принимаемые решения были четкими, особенно тогда, когда это касается дифференциации между костной и хрящевой опухолями.

Данные, собранные в соответствии с указанной формой рентгенологического анализа, переносятся на перфокарты и вводятся в вычислительную машину для расчета по формуле Байеса. Лодвик составил вектор значений априорных вероятностей $P(D)$ для девяти типов опухолей и матрицу условных вероятностей симптома — заболевания $P(S|D)$ (фиг. 13).

Были использованы две матрицы значений условных вероятностей симптома — заболевания. Первая включает значения $P(S|D)$, полученные в виде частот на основании материала из Регистра Кодмана для костных сарком, литературных данных и архива Лодвика. Вторая матрица основана на частотах, модифицированных его собственным опытом. После проведенной диагностики ряда верифицированных случаев Лодвик [62] пришел к заключению, что вторая матрица дает лучшие результаты (приблизительно на 5%), чем матрица, основанная только на фактическом материале.

В результате байесовского анализа получается дифференциальный диагноз в виде перечня возможных костных опухолей в порядке убывания их вероятностей от наибольшей к наименьшей. Было отмечено, что точность правильного определения гистологического типа и рентгенологической градации опухоли составляет 90% [62].

Предположим, что вы не согласны с матрицей вероятностей Лодвика. Должны ли вы построить вашу собственную матрицу? Конечно,

g — подразделения, имеющие отношение к трубчатым костям (можно выбрать больше одного варианта):

суставная поверхность	20	1	0
метафиз	21	1	0
диафиз	22	1	0

д — конец трубчатой кости:

проксимальная половина	24	1	0
дистальная половина	23	1	0

Положение относительно продольной оси кости. Положение центра опухоли относительно продольной оси кости является очень важным фактором в идентификации опухоли. Опухоль, возникающая из кортикального слоя (или параостеального в смысле локализации), обычно имеет центр на или вне кортикального слоя.

Положение относительно таза. Хорошо известно, что определенные виды опухолей имеют специфические локализации. Например, хондробластома почти всегда обнаруживается по крайней мере частично внутри эпифиза. Гигантоклеточные опухоли также возникают в локализациях, примыкающих к эпифизной пластинке. На рисунке показана эпифизная локализация в этой сложной кости.

Положение вдоль кости. Рисунок длинной кости дан в основном для того, чтобы показать границу метафиза. Для целей данного анализа метафизом считается та часть кости, которая охватывает отделенную пунктиром зону между эпифизной пластинкой и диафизом [63].

Вид опухоли	Частота P(D)	P(S D)																	
		Возраст			Локализация				Размер		Пораженная субстанция		Градация						
		до 21 года	21—30	31 и больше	Малые и плоские	Длинные	Место поражения длинной кости				— см	6 см и больше	кость	хрящ	I-A	I-B	I-C	II	III
							эпифиз	диафиз	метафиз	платинка									
Гигантоклеточная опухоль . . .	15	20	35	45	20	80	90	01	100	20	40	60	01	00	15	35	50	00	00
Хондробластома	05	75	20	05	20	80	100	50	75	00	90	10	00	30	50	35	15	00	00
Хондромиксоидная фиброма . .	03	50	35	15	30	70	30	20	100	25	85	15	00	02	85	15	01	00	00
Хондросаркома	17	25	25	50	35	65	40	01	85	65	20	80	05	65	15	20	25	25	15
Фибросаркома	10	20	20	60	05	95	55	00	90	65	20	80	25	02	00	10	40	30	20
Остеосаркома	25	65	25	10	10	90	30	05	95	75	15	85	98	05	00	00	10	30	60
Параостеальная саркома	05	20	35	45	00	100	30	01	100	50	25	75	100	05	15	25	55	05	00
Опухоль Юинга	15	70	25	05	35	65	20	05	85	90	15	85	00	00	00	05	10	20	65
Саркома ретикулярных клеток	05	10	25	65	20	80	50	01	85	80	15	85	00	00	00	00	20	30	50

Фиг. 13.

Упрощенная матрица априорных вероятностей для первичных опухолей кости. Каждое число отражает вероятное количество случаев, в которых будет присутствовать данный признак в выборке из 100 случаев. Некоторые очень высокие или очень низкие значения вероятностей (100, 99, 01, 00) по существу используются как булевы функции для исключения определенных категорий заболевания. Например, опухоль, попадающая в радиологическую градацию III, не может быть гигантоклеточной опухолью, хондробластомой, хондромиксоидной фибромой или параостеальной саркомой. Аналогично опухоль, включающая диафиз длинной кости, не может быть хондробластомой [61].

	I-A	I-B	I-C	II	III
Деструкция кости (если присутствует)	«Географическая» Поверхностный слой не пенетрирован	«Географическая» Поверхностный слой не пенетрирован	«Географическая» Возможна деструкция типа «изъеденная молью» в периферической зоне 1 см Поверхностный слой пенетрирован	«Изъеденная молью» Поверхностный слой пенетрирован	Проникающая Поверхностный слой пенетрирован
Пролиферация кости	Должна быть склеротическая щель; может быть увеличен корковый слой на 1 см или менее над поверхностью кости Опорная перестройка кости Нет треугольника Кодмана	Может быть склеротическая щель; должен быть увеличен корковый слой больше чем на 1 см Опорная перестройка кости Нет треугольника Кодмана	(Если присутствует) Может быть обширная Опорная перестройка кости Один треугольник Кодмана	(Если присутствует) Крапчатая Нет опорной перестройки Один или больше треугольников Кодмана	(Если присутствует) Крапчатая Нет опорной перестройки Два или больше треугольников Кодмана
Минерализация межклеточного вещества (если присутствует)	Рыхлая или твердая	Рыхлая или твердая	Рыхлая или твердая	Рыхлая, твердая или бугристая	Бугристая или непрозрачная
Размер	До 6 см	Любой размер	Любой размер	Больше 6 см	Больше 6 см
Предсказываемый 5-летний срок выживаемости	95—100%	40—95%	25—60%	15—35%	4—15%

Фиг. 14.

Прототипы рентгенологических градаций первичных опухолей кости [61].

но, если у вас имеются исходные данные. Тогда у вас будет возможность сравнить вашу модель с моделью Лодвика. В тех случаях, когда встретятся различия, полезно будет определить их причины. В конечном счете должна быть разработана общепринятая модель для анализа костных опухолей, которая в сжатом виде представляет опыт специалистов и может быть использована практическими врачами. Реальный выигрыш от такой модели заключается в том, что она приведет к общему соглашению, ибо общее соглашение по таким вопросам, как определение признаков, симптомов, диагноза и априорных вероятностей, гораздо важнее, чем то, чтобы эти определения были «абсолютно точными», хотя нужно стремиться к максимально возможной точности.

Рентгенологическая градация костных опухолей использовалась Лодвиком [61] как средство для сжатия информации и исключения зависимых переменных; эта процедура находится еще в экспериментальной стадии, но мы решили ее здесь описать, так как она, по-видимому, является методом, который имеет общее значение при разработке машинной диагностики.

Многие изменения, которые являются результатом роста опухоли в кости, преимущественно отражают скорость роста и агрессивность поведения опухоли и лишь косвенно связаны с гистологическим типом опухоли. Прототипы пяти рентгенологических градаций представлены в таблице на фиг. 14. Эти виды разрушения кости и костных прорастаний, лежащие в основе рентгенологических градаций, встречаются не только в первичных костных опухолях, но обнаружены также во вторичных новообразованиях, при инфекциях, дисплазии и при местном повреждении и восстановительной реакции почти любого типа.

Используя рентгенологические градации, приведенные на фиг. 14, можно создать таблицу истинности [64] для обработки входных данных по логической схеме, действующей по принципу «все или ничего»; при этом последовательно исключаются невозможные категории, пока не останется только одна градация. Эта процедура предполагает, что связанная с переменными условная вероятность равна либо 1, либо 0. Таблица истинности для рассматриваемых типов прорастания кости приведена на фиг. 15.

Если возникают разногласия в данных, касающихся рентгенологических градаций, то предлагается отдавать предпочтение различным видам рентгенологических заключений в соответствии с приведенной ниже последовательностью:

1. Виды деструкции кости.
2. Пенетрация кортикального слоя кости.
3. Изъеденные контуры края.
4. Виды костных пролифераций.
5. Количество треугольников Кодмана.
6. Размер расширенного кортикального слоя кости.

Переменная	Градация				
	I-A	I-B	I-C	II	III
1. Склерозированный край	1	0	2	2	2
2. Увеличенная в объеме оболочка	0	1	0	2	2
3. Один треугольник Кодмана	2	2	0	0	0
4. Два треугольника Кодмана	2	2	2	0	0
5. Три или больше треугольников Кодмана	2	2	2	2	0
6. Опорная перестройка кости	0	0	0	2	2
7. Наличие обесцвеченных участков	2	2	2	0	0

Фиг. 15.

Таблица истинности для пролиферативной реакции кости. Если определенная градация требует присутствия некоторого вида пролиферации кости, то в соответствующей клетке матрицы стоит 1. Если данная градация требует отсутствия определенного вида пролиферации, то в соответствующей клетке стоит цифра 2. Если для данной градации не имеет значения наличие или отсутствие данного типа пролиферации, то в соответствующей клетке стоит 0. Таблицы истинности обрабатываются по строкам в последовательности слева направо и сверху вниз [64].

7. Формы костной опухоли.

8. Размер опухоли.

В следующем примере показано, как эти предпочтения используются на практике. При обследовании обнаружено: вид разрушения кости — «географический», склерозированный край, незатронутый кортикальный слой кости, распространяющийся менее чем на 1 см, опухоль более 6 см в диаметре. В соответствии с таблицей на фиг. 14 все эти данные, за исключением размера опухоли, подходят к рентгенологической градации I-A. Размер опухоли в перечне предпочтений расположен в самом конце и, следовательно, не принимается во внимание.

Рентгенологическая градация с использованием таблиц истинности напоминает сходную процедуру, которая по предложению Ледли и Ластеда [54] может быть использована для решения вопроса о том, какое следующее исследование нужно провести у больного, и для определения минимального набора необходимых исследований на каждом этапе диагностики.

2.4.1. МЕТОД ОПРЕДЕЛЕНИЯ МИНИМАЛЬНО НЕОБХОДИМЫХ ДОПОЛНИТЕЛЬНЫХ ТЕСТОВ

Для описания этого метода мы воспользуемся конкретным примером. На фиг. 16 приведен набор комплексов симптома — заболевания, включающий четырнадцать симптомов и восемь заболеваний.

В этом примере симптомы разделены на группы (анамнез, физические исследования, лабораторные исследования крови и исследования костного мозга), расположенные в порядке возрастания трудности их выполнения. Заметим, что симптом $S(1)$ в действительности не вносит ничего в диагноз, так как в каждом столбце против этого признака стоит единица и, следовательно, по этому признаку невозможно провести различие между столбцами. Предположим, для примера, что пациент имеет и симптом $S(2)$ и симптом $S(3)$. Тогда диагноз на этом этапе заключается в том, что пациент имеет комплекс заболеваний, связанный с одним из первых одиннадцати столбцов. Таким образом, на основе имеющихся симптомов мы исключаем из рассмотрения остальные столбцы, как показано на фиг. 16.

		Номера столбцов: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16														
Анамнез		S(1)	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111	1111
→ S(2)			1111	1111	1111	1111	1111	1111	1111	1000	1000	1000	1000	1000	1000	1000
→ S(3)			1111	1111	1111	1111	1111	1111	1111	1110	1110	1110	1110	1110	1110	1110
Физические исследования		→ S(4)	1111	1111	1111	1111	1111	1111	1111	1000	1000	1000	1000	1000	1000	1000
→ S(5)			1100	0000	0000	0000	0000	0000	0000	0111	0111	0111	0111	0111	0111	0111
→ S(6)			1100	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
Исследования крови		S(7)	11	1101	01	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110
→ S(8)			11	1110	11	1100	1100	1100	1100	1100	1100	1100	1100	1100	1100	1100
→ S(9)			11	1010	01	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001	1001
→ S(10)			11	0100	00	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
→ S(11)			11	10000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
→ S(12)			11	100	100	100	100	100	100	100	100	100	100	100	100	100
Исследования костного мозга		S(13)	1011	0000	0111	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
S(14)			100	1110	0000	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110	1110
D(1)			0000	0000	1000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
D(2)			0101	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
D(3)			0010	0100	0001	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000
D(4)			0001	1000	0000	0100	0100	0100	0100	0100	0100	0100	0100	0100	0100	0100
D(5)			0000	1000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
D(6)			0000	0100	0111	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
D(7)			1000	0010	0011	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000
D(8)			0000	0001	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Фиг. 16.

Логический базис для комплекса симптомы — заболевания, состоящего из четырнадцати симптомов и восьми заболеваний. После каждой серии диагностических тестов исследуются возможные диагнозы для определения необходимых тестов на следующем этапе [54].

Предположим, что мы рассматриваем симптомы $S(4)$ и $S(5)$, и у пациента есть $S(4)$ и нет $S(5)$; при этом остается только семь столбцов в качестве возможных диагнозов, а именно столбцы 3, 4, 5, 6, 7, 8 и 9 (фиг. 16). Выбор между возможными исследованиями крови $S(7)$, $S(8)$, $S(9)$, $S(10)$, $S(11)$ и $S(12)$ сделать не так легко. Как будет показано ниже, необходимы только три из этих шести исследований, а именно $S(9)$, $S(10)$ и $S(12)$. Предположим, что пациент имеет положительные результаты по исследованиям $S(9)$ и $S(12)$ и отрицательный по исследованию $S(10)$. Тогда, как пока-

зано на фиг. 16, по этому признаку невозможно провести различие между столбцами. Предположим, для примера, что пациент имеет и симптом $S(2)$ и симптом $S(3)$. Тогда диагноз на этом этапе заключается в том, что пациент имеет комплекс заболеваний, связанный с одним из первых одиннадцати столбцов. Таким образом, на основе имеющихся симптомов мы исключаем из рассмотрения остальные столбцы, как показано на фиг. 16.

Из симптомов физических исследований $S(4)$, $S(5)$ и $S(6)$ только два являются необходимыми, так как все три не могут осуществить различие между оставшимися столбцами лучше, чем любые два.

зано на фиг. 16, остается только один столбец, а именно столбец 5. Таким образом, исследования костного мозга не являются необходимыми и у пациента имеется комбинация из двух заболеваний $D(4)$ и $D(5)$. Предположим, с другой стороны, что при физическом исследовании пациент имел и симптом $S(4)$ и симптом $S(5)$. Тогда останутся только столбцы 1 и 2; исследования крови их не различают, и возникнет необходимость в проведении исследования костного мозга $S(13)$.

2.4.2. МЕТОД МИНИМИЗАЦИИ КОЛИЧЕСТВА ИССЛЕДОВАНИИ

Теперь очевидно, что на любой стадии исследования пациента минимальное количество исследований следующего порядка трудности может быть определено при помощи процесса, описанного выше. Можно показать, что проблема выбора минимального количества исследований на любой стадии тесно связана с задачей минимизации булевых функций. В реальных ситуациях, где количество исследований и возможных столбцов в таблице гораздо больше, помощь машины становится существенной. Тем не менее полезно иметь приближенные правила для выбора минимума исследований в относительно простых случаях.

Рассмотрим, например, выбор между исследованиями крови $S(7)$, $S(8)$, $S(9)$, $S(10)$, $S(11)$ и $S(12)$, когда остаются только столбцы 3, 4, 5, 6, 7, 8 и 9. Первый шаг — это подсчет количества единиц, встречающихся в этих столбцах для соответствующих строк:

	34	5678	9	
$S(7)$	11	1101	0	5
$S(8)$	11	1110	1	6
→ $S(9)$	11	1010	0	4
$S(10)$	11	0100	0	3
$S(11)$	10	0000	1	2
$S(12)$	10	1100	1	4

} Количество единиц

Следующий шаг при построении минимального набора заключается в выборе такого исследования, которое разделяет все столбцы примерно на две равные части (в нашем случае это исследование должно иметь 3 или 4 единицы). Этому требованию удовлетворяет исследование $S(9)$. Далее рассмотрим те столбцы, которые по $S(9)$ имеют единицы, и будем искать следующее исследование, которое разделяет их поровну; то же самое сделаем для столбцов, которые по $S(9)$ содержат нули, и если теперь удастся в качестве этого исследования выбрать то же самое, что и для единиц в столб-

цах по $S(9)$, то тем лучше. Затем процесс повторяется. Например, пусть мы имеем:

Столбцы, разделенные исследованием $S(9)$	$\left\{ \begin{array}{l} S(7) \\ S(8) \\ \rightarrow S(10) \\ S(11) \\ S(12) \end{array} \right.$	$\begin{array}{cccc ccc} 3 & 4 & 5 & 7 & 6 & 8 & 9 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{array}$
Столбцы, разделенные исследованием $S(10)$	$\left\{ \begin{array}{l} S(7) \\ S(8) \\ S(11) \\ \rightarrow S(12) \end{array} \right.$	$\begin{array}{cccc ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 3 & 4 & 5 & 7 & 6 & 8 & 9 \end{array}$

Как следует из таблицы, здесь необходимы только три исследования: $S(9)$, $S(10)$ и $S(12)$.

Это вовсе не означает, что выбранные три исследования наиболее эффективны, если мы хотим определить, имеет ли пациент некоторый определенный комплекс заболеваний. Например, предположим, что нам нужно узнать, имеет ли пациент комплекс заболеваний столбца 9, т. е. $D(1)$, и ни одного из других заболеваний (фиг. 16). Тогда процесс отыскания количества необходимых исследований, наиболее «близкого» к минимальному, выглядит следующим образом. Перепишем рассматриваемый подмассив так, чтобы столбец 9 состоял из одних единиц, вводя отрицание симптомов:

	3	4	5	6	7	8	9	
$\overline{S(7)}$	0	0	0	0	1	0	1	} 4 Количество единиц в столбцах 3, 5, 6 и 7 для каждой строки
$\overline{S(8)}$	1	1	1	1	1	0	1	
$\overline{S(9)}$	0	0	0	1	0	1	1	
$\overline{S(10)}$	0	0	1	0	1	1	1	
$\overline{S(11)}$	1	0	0	0	0	0	1	
$\overline{S(12)}$	1	0	1	1	0	0	1	

Количество нулей: 3 5 3 3 3 4

Теперь подсчитаем нули в столбцах; для строк подсчитаем единицы, которые встречаются в столбцах с наименьшим количеством нулей. Столбцы 3, 5, 6 и 7 имеют по три нуля каждый; это наименьшее количество нулей. Строка для $S(7)$ имеет только одну единицу в этих столбцах, строка для $S(8)$ — четыре единицы и т. д. Теперь выберем в качестве первого исследования то, которое имеет наименьшее количество таких единиц. В нашем примере имеется выбор из $S(7)$, $S(9)$ и $S(11)$. Выбрав $S(7)$, вычеркнем соответствующую строку и все столбцы с нулями в этой строке; затем повторим этот процесс на оставшемся массиве:

	7	9	
$S(8)$	1	1	2
$S(9)$	0	1	1
$S(10)$	1	1	2
$S(11)$	0	1	1
$S(12)$	0	1	1

Количество нулей 3

Совершенно очевидно, что второе исследование мы можем выбрать из $S(9)$, $S(11)$ и $S(12)$. Таким образом, для того чтобы определить, страдает ли пациент заболеванием $D(1)$, необходимо провести только два исследования крови.

Существует, конечно, много других ситуаций, для которых можно придумать специальные методы минимизации. Например, вместо выбора минимального количества исследований на данном этапе диагностики может оказаться желательным минимизировать какую-либо характеристику или цену, связанную с этими исследованиями. Мы не нашли работ, в которых бы экспериментально исследовались правила, определяющие оптимальную последовательность исследований (т. е. лабораторных тестов, ЭКГ, рентгеновских процедур) для данного пациента¹⁾. Это важный вопрос, и он заслуживает дальнейшей разработки.

2.5. ПРАВДОПОДОБИЕ И ОТНОШЕНИЕ ПРАВДОПОДОБИЯ

Среди ранних экспериментов по применению машинных средств медицинской диагностики заслуживают обсуждения работы Липкина и Харди [60] и Липкина [58, 59], поскольку эти ученые предложили концепцию правдоподобия для дифференциальной диагностики заболеваний крови и показали, как эту концепцию можно реализовать на механических устройствах для сортировки перфокарт (позже — при помощи численных методов сравнения на электронных вычислительных машинах), чтобы получить определенный диагностический индекс и расположить диагнозы в порядке убывания этого индекса. Эта методика позволяла устанавливать дифференциальный диагноз.

Прежде чем продолжать чтение этого параграфа, читателю было бы полезно вернуться к обсуждению понятия правдоподобия, приведенному в гл. 1 (разд. 1.4). Важное положение, которое нужно всегда помнить, заключается в том, что правдоподобие комплекса симптомов S при данном заболевании D пропорционально $P(S|D)$. Следовательно, умножение $P(S|D_1)$ и $P(S|D_2)$ на любую

¹⁾ По этому вопросу см. [П1—А 5], а также Вишневский А. А., Артоблевский И. И., Быховский М. Л., Принципы построения диагностических машин, Вестник АМН СССР, № 2, 1964. — Прим. перев.

константу K не повлияет на конечные шансы. Это можно выразить в виде отношения шансы — правдоподобие:

$$\Omega_1 = \frac{P(D_1|S)}{P(D_2|S)} = \frac{P(S|D_1)K}{P(S|D_2)K} \frac{P(D_1)}{P(D_2)} = L(D_1; S) \Omega_0. \quad (6)$$

Рассмотрим следующий пример применения принципа правдоподобия, не относящийся к медицине. Предположим, что у нас есть две гипотезы относительно монеты: что на одной ее стороне «орел», а на другой «решка» и что на обеих сторонах «орел». Пусть априорные вероятности этих двух гипотез равны, т. е. каждая равна $1/2$. Предположим теперь, что монету подбросили десять раз и она каждый раз падала «орлом». Тогда правдоподобия двух гипотез равны $2^{-10} = 1/1024$ (для гипотезы, что монета с «орлом» и «решкой») и 1 (для гипотезы, что монета с двумя «орлами»).

Правдоподобие связано с теоремой Байеса, которая в действительности устанавливает, что апостериорная, или конечная, вероятность гипотезы (заболевания) пропорциональна его априорной, или начальной, вероятности, умноженной на правдоподобие. Следовательно, в примере с подбрасыванием монеты, приведенном выше, апостериорные вероятности пропорциональны $1/1024$ и 1. Итак, апостериорная вероятность того, что монета с двумя «орлами», равна $1024/1025$, т. е. это событие практически достоверно.

Дальнейшее рассмотрение функций правдоподобия и максимального правдоподобия увело бы нас в сторону от рассматриваемых проблем. Читатель найдет интересную трактовку этих вопросов в работах Корнфилда [13] и Линдли [57].

2.5.1. ПРИМЕР ПРИМЕНЕНИЯ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ В ГЕМАТОЛОГИИ

В работе Липкина и Харди [60] сравнивается комплекс симптомов пациента, у которого имеется заболевание крови, с набором объективных критериев, которые были определены для того, чтобы установить правдоподобие диагноза. Все данные, описывающие двадцать шесть заболеваний крови, перечислены в главном кодовом списке, часть которого приведена на фиг. 17.

Каждому виду входной информации для каждого из заболеваний приписывается некоторое числовое значение, или вес (фиг. 18).

Так как вес по существу представляет собой правдоподобие, то такой процесс в конечном счете приписывает субъективную меру вероятности $P(S|D)$ каждому виду информации в соответствии со следующими указаниями [60]: «Если дано описание заболевания, припишем положительный вес данным, способствующим постановке правильного диагноза, а отрицательный вес — данным, которые ей не способствуют. Пусть важность данного вида входной информации о заболевании определяет величину веса. Каждый вид ин-

АНАМНЕЗ

	Симптомы	Код №
Пол	{ Мужской	0
	{ Женский	1
Раса	Негр	2
Беременность		3
Возраст	{ Меньше 5 лет	4
	{ Меньше 4 недель	5
	{ Больше 5 лет	6
Семейная гемолитическая анемия в анамнезе		7
Расстройство пищеварения, изжога, постоянное чувство тяжести в эпигастрии		8
Отрыжка, метеоризм		8

АНАЛИЗ ПЕРИФЕРИЧЕСКОЙ КРОВИ

	Симптомы	Код №
Гематокрит	{ Меньше 37 см ³ на 100 см ³	29
	{ 37—54	30
	{ Больше 54	31
		33
Количество лейкоцитов	{ Меньше 5·10 ³ на мм ³	
	{ 5—10·10 ³	34
	{ Больше 10·10 ³	35
	{ Больше 60·10 ³	36
тромбоцитов	{ Меньше 190·10 ³ на мм ³	37
	{ 190—400·10 ³	38
	{ Больше 400·10 ³	39

Ф и г. 17.

Выдержка из главного списка признаков и их кодов для диагностики гематологических заболеваний [58].

формации может, таким образом, иметь различный вес для каждого из заболеваний. Пусть каждое описание заболевания имеет все виды информации в составе главного кода, а каждый вид информации для каждого заболевания — положительный или отрицательный числовой вес». Приписанные значения весов находятся в диапазоне от 20 до —20, а сумма всех весов для каждого из заболеваний дает полный положительный и полный отрицательный вес.

В соответствии с указанными выше правилами показатели состояния пациентов сравнивались с показателями, описывающими заболевания. Математически эта процедура сводится к сравнению суммы взвешенных диагностических критериев пациентов с суммой взвешенных диагностических критериев заболевания; таким образом, определяется степень близости состояния данного пациента к данному заболеванию. Взвешенное среднее данных пациента опре-

деляется по формуле

$$W.A. = \frac{\sum wX}{\sum w}, \quad (7)$$

где

- $W.A.$ — взвешенное среднее (отношение правдоподобия);
- w — вес данного симптома, признака, лабораторного исследования и т. д.;
- X — величина, равная 1 для каждого симптома, имеющегося у пациента, и равная 0 для каждого отсутствующего симптома;
- $\sum wX$ — сумма отдельных величин X для данного пациента, умноженных на соответствующие веса w ;
- $\sum w$ — сумма весов всех характеристик для данного заболевания.

Таким образом, если пациент имеет все симптомы, признаки, и т. д., которые характеризуют заболевание, то окончательное взвешенное среднее $W.A.$, которое Липкин называет *диагностическим индексом*, будет равно 1,00. Если пациент не имеет ни одной характеристики, соответствующей заболеванию, то диагностический индекс равен нулю. Диагностический индекс, вычисленный для семи случаев, приведен в таблице на фиг. 19. Числовые данные показывают, что в случае, когда диагностический индекс приближается к 0,2, при дифференциальной диагностике следует уже считаться с возможностью данного заболевания.

Дифференциальный диагноз для конкретного пациента осуществляется путем сравнения данных этого пациента с данными из материала верифицированных классических случаев (фиг. 19). Используемый метод представляет собой процедуру сравнения, которая реализуется вычислительной машиной. Ответ машины имеет вид списка дифференциальных диагнозов заболеваний; в дополнение к этому машина выдает на печать исследования, которые могут потребоваться в дальнейшем.

В работе [59] Липкин отметил трудности, возникающие при выборе функции правдоподобия для целей дифференциальной диагностики. Он указывает: «Если бы «возможные случаи» были связаны с некоторыми аналитическими функциями распределения значений параметров, то можно было бы найти функцию правдоподобия, которая связывает отношение правдоподобия (приведенное к некоторому стандарту) с интервалом возможных значений параметров. Но исторически описания «возможных случаев» (заболеваний) не сложились как аналитические функции распределения значений параметров. Поэтому нельзя выработать и функцию правдоподобия».

Хотя функции правдоподобия не были проверены в применении к медицинской диагностике, мы возвращаемся к рассмотрению принципа правдоподобия, чтобы изложить метод отношения прав-

Сравнение взвешенных средних клинических диагнозов для семи случаев

Случай	Диагностический индекс (взвешенные средние, положительные)	Клинический диагноз
1	Хроническая лимфоидная лейкемия 0,41 Острая гемолитическая анемия 0,28 Идиопатическая тромбоцитопеническая пурпура 0,27 Гемофилия 0,26	Хроническая лимфоидная лейкемия Приобретенная гемолитическая анемия
2	Агранулоцитоз 0,49 Хроническая лимфоидная лейкемия 0,45 Язвенный стоматит 0,21 Наследственный сфероцитоз 0,19	Хроническая лимфоидная лейкемия Агранулоцитоз
3	Мальтийская анемия 0,28 Гемофилия 0,23 Острая гемолитическая анемия 0,22 Множественная миелома 0,22	Мальтийская анемия Нетромбоцитопеническая пурпура
4	Гипохромная микроцитарная анемия 0,31 Идиопатическая тромбоцитопеническая пурпура 0,27 Множественная миелома 0,19 Мальтийская анемия 0,15	Идиопатическая тромбоцитопеническая пурпура Анемия острая и хроническая
5	Множественная миелома 0,37 Апластическая анемия 0,34 Агранулоцитоз 0,30 Хроническая лимфоидная лейкемия 0,09	Множественная миелома Агранулоцитоз
6	Агранулоцитоз 0,47 Хроническая лимфоидная анемия 0,21 Эозинофильная лейкемия 0,19 Моноцитарная лейкемия 0,18	Агранулоцитоз Макроцитарная анемия
7	Агранулоцитоз 0,47 Апластическая анемия 0,21 Хроническая лимфоцитарная лейкемия 0,19 Острая постгеморрагическая анемия 0,19	Макроцитарная анемия Нейтропения Тромбоцитопения

Ф и г. 19.

Диагностические индексы, полученные для семи случаев, каждый из которых включает несколько гематологических аномалий. Заболевания, приведенные в перечне диагностических индексов, сравниваются с клиническими диагнозами [59].

доподобия, который использовался при автоматическом многомерном разделении и в диагностике. Вспомним, что, согласно принципу правдоподобия, все наблюдения (симптомы), приводящие к одной и той же функции правдоподобия, должны приводить к одному и тому же заключению (заболеванию). В последующем обсуждении отношения правдоподобия в задачах диагностики мы следуем работе [21].

Два возможных комплекса симптомов S и S^1 (не обязательно являющиеся результатом одного и того же исследования) могут иметь одинаковое (потенциальное) влияние на наше мнение о разделении заболеваний D_i . Иными словами, величина $P(D_i|S)$ может быть равна $P(D_i|S^1)$ для любого i . Когда же именно S и S^1 с очевидностью эквивалентны? Оставляя в стороне теоретическую возможность равенства нулю некоторых из величин $P(D_i)$, мы можем сказать, что это имеет место, если для некоторой положительной константы k и для всех i

$$P(S^1 | D_i) = kP(S | D_i). \quad (8)$$

Но уравнение (8) означает, что

$$P(D_i | S) = \frac{P(S | D_i) P(D_i)}{P(S)} = \frac{P(S^1 | D_i) P(D_i)}{P(S^1)} = P(D_i | S^1) \quad (9)$$

независимо от того, каковы априорные вероятности $P(D_i)$, так как

$$P(S^1) = \sum P(S^1 | D_i) P(D_i) = k \sum P(S | D_i) P(D_i) = k P(S).$$

Этот вывод и есть принцип правдоподобия: два (потенциальных) комплекса симптомов S и S^1 имеют одинаковую степень важности, если выполняется уравнение (8).

2.5.2. ПРИМЕР ПРИМЕНЕНИЯ ОТНОШЕНИЯ ПРАВДОПОДОБИЯ В ЗАДАЧЕ АВТОМАТИЧЕСКОГО МНОГОМЕРНОГО РАЗДЕЛЕНИЯ

В работах Коллена с соавторами [9, 10] используется метод отношения правдоподобия, основанный на процедуре проверки гипотез, предложенной Нейманом в качестве математической базы, для автоматического многомерного разделения. Термин «выделение заболевания» обозначает осуществление теста, который с достаточным правдоподобием обнаруживает заболевание, если оно действительно имеет место, для того чтобы отделить лиц, которые, вероятно, имеют это заболевание, от лиц, которые, вероятно, его не имеют. Термин «многомерное разделение» обозначает совокупность тестов для разделения некоторого множества заболеваний. Автоматическое многомерное разделение — это расширенная концепция использования автоматического и полуавтоматического оборудова-

Номер вопроса						Пациенты, страдающие астмой		Пациенты, не страдающие астмой		Отношение правдоподобия $L=\theta = \frac{P(S D_1)}{P(S D_0)}$	
1	2	3	4	5	6	№	$P(S D_1)$	№	$P(S D_0)$		
Нет	Нет	Нет	Да	Нет	Нет	0	0	16)	0,03095	0,13535	0
Нет	Нет	Нет	Нет	Да	Нет	0	0	6	0,01161		0
Нет	Нет	Нет	Да	Нет	Да	0	0	20	0,03868		0
Нет	Нет	Да	Нет	Нет	Да	0	0	5	0,00967		0
Нет	Нет	Да	Да	Нет	Нет	0	0	3	0,00580		0
Да	Нет	Нет	Да	Нет	Нет	0	0	3	0,00580		0
Да	Нет	Нет	Нет	Да	Нет	0	0	1	0,00193		0
Нет	Нет	Нет	Да	Да	Нет	0	0	1	0,00193		0
Нет	Нет	Да	Да	Нет	Да	0	0	5	0,00967		0
Нет	Нет	Нет	Да	Да	Да	0	0	2)70	0,00387		0
Да	Нет	Нет	Да	Нет	Да	0	0	1	0,00193		0
Да	Нет	Да	Нет	Нет	Да	0	0	1	0,00193		0
Да	Нет	Нет	Да	Да	Нет	0	0	1	0,00193		0
Нет	Нет	Да	Да	Да	Нет	0	0	1	0,00193		0
Да	Нет	Да	Да	Да	Нет	0	0	1	0,00393		0
Да	Нет	Да	Да	Нет	Да	0	0	1	0,00193		0
Да	Нет	Да	Нет	Да	Да	0	0	1	0,00193		0
Нет	Нет	Нет	Нет	Нет	Да	1	0,00435	24	0,04642		0,094
Нет	Нет	Нет	Нет	Нет	Нет	16	0,6957	353	0,68279		0,102
Нет	Нет	Нет	Нет	Да	Да	1	0,00435	8	0,01547		0,281
Нет	Нет	Да	Нет	Нет	Нет	5	0,02174	33	0,06383	0,341	
Да	Нет	Нет	Нет	Нет	Нет	1	0,00435	5	0,00967	0,500	
Да	Нет	Нет	Нет	Да	Да	1	0,00435	1	0,00193	2,25	

Нет	Да	Да	Нет	Нет	Да	3	0,01304	2	0,00387	3,37	
Нет	Да	Да	Да	Нет	Нет	2	0,00870	1	0,00193	4,50	
Нет	Да	Нет	Нет	Нет	Нет	21	0,09130	7	0,01354	6,74	
Нет	Да	Да	Нет	Да	Да	3	0,01304	1	0,00193	6,74	
Нет	Да	Да	Да	Нет	Да	3	0,01304	1	0,00193	6,74	
Нет	Да	Нет	Нет	Нет	Да	11	0,04783	3	0,00580	8,24	
Да	Да	Нет	Нет	Нет	Нет	23	0,10000	4	0,00774	12,92	
Нет	Да	Да	Нет	Нет	Нет	18	0,07826	2	0,00387	20,23	
Да	Да	Да	Нет	Да	Да	11	0,04783	1	0,00193	24,73	
Да	Да	Нет	Да	Да	Да	15	0,06522	1	0,00193	33,72	
Нет	Да	Нет	Нет	Да	Нет	1	0,00435	0	0	∞	
Нет	Да	Нет	Да	Нет	Нет	1	0,00435	0	0	∞	
Нет	Нет	Да	Нет	Да	Да	1	0,00435	0	0	∞	
Нет	Да	Да	Нет	Да	Нет	1)95	0,00435	0,41301	0	0	∞
Да	Да	Нет	Нет	Да	Нет	1	0,00435	0	0	∞	
Да	Да	Нет	Да	Нет	Нет	2	0,00870	0	0	∞	
Да	Да	Нет	Нет	Нет	Да	3	0,01304	0	0	∞	

Фиг. 20.

Относительное количество пациентов с бронхиальной астмой $P(S|D_1)$ и без бронхиальной астмы $P(S|D_0)$, которые ответили «да» на шесть диагностических вопросов. Значения отношения правдоподобия, равные 0 и ∞, можно исключить, начиная построение матрицы симптомы — заболевания с 1 в каждой клетке [9].

Номер вопроса						Пациенты, страдающие астмой		Пациенты, не страдающие астмой		Отношение правдоподобия $L=0 = \frac{P(S D_1)}{P(S D_0)}$
1	2	3	4	5	6	№	$P(S D_1)$	№	$P(S D_0)$	
Нет	Да	Нет	Да	Нет	Да	5	0,02174	0	0	∞
Нет	Да	Нет	Нет	Да	Да	5	0,02174	0	0	∞
Да	Да	Да	Нет	Нет	Нет	14	0,06087	0	0	∞
Да	Да	Да	Да	Нет	Нет	1	0,00435	0	0	∞
Да	Да	Нет	Да	Да	Нет	1	0,00435	0	0	∞
Нет	Да	Нет	Да	Да	Да	1	0,00435	0	0	∞
Нет	Нет	Да	Да	Да	Да	1	0,00435	0	0	∞
Да	Да	Да	Нет	Нет	Да	4	0,01734	0	0	∞
Да	Да	Нет	Да	Нет	Да	5	0,02174	0	0	∞
Да	Да	Нет	Нет	Да	Да	8	0,03478	0	0	∞
Да	Да	Да	Да	Да	Нет	1	0,00435	0	0	∞
Да	Да	Да	Да	Нет	Да	3	0,01304	0	0	∞
Нет	Да	Да	Да	Да	Да	3	0,01304	0	0	∞
Да	Да	Да	Да	Да	Да	33	0,14348	0	0	∞
Итого						230	1,00	517	1,00	

Фиг. 20. Продолжение.

ния, электронного или механического, для определения, достаточно ли правдоподобно наличие заболевания или нужно провести дополнительное диагностическое исследование. В этом контексте под заболеванием мы понимаем любое отклонение от «нормы», имеющее «характерные» симптомы.

Тесты для многомерного разделения, использованные Колленом, являются частью периодического полного обследования состояния здоровья населения. Мы здесь не рассматриваем тесты и подробности машинной обработки данных, так как читатель может найти их в работах [9, 10]. Использование метода отношения правдоподобия будет проиллюстрировано на примере выделения бронхиальной астмы, проведенного Колленом.

Были выбраны шесть дихотомических вопросов из вопросника, заполненного 230 пациентами с клиническим диагнозом бронхиальной астмы и 517 случайно выбранными пациентами, которые не страдали этим заболеванием. Этот набор проверочных вопросов приведен ниже; как указывает Коллен, они «не могут считаться ни наиболее диагностически важными, ни обязательно независимыми друг от друга».

Имели ли вы в последний год или два года	Да	Нет
1) затрудненное дыхание, из-за которого вы просыпались?		
2) затрудненное дыхание с астматическими хрипами?		
3) откашливание желтой или зеленой мокроты?		
Имели ли вы в последний год или два года стесняющее чувство или боль в груди,		
5) которая будит вас ночью?		
4) когда сердиты или возбуждены?		
6) которая длится больше чем пять минут?		

Было подсчитано отношение пациентов из астматической и неастматической групп, которые ответили «да» на эти шесть вопросов, для определения частного от деления количества пациентов с каждым из комплексов симптомов на общее количество пациентов в группе. Результаты приведены на фиг. 20. Возможное число комбинаций ответов на шесть вопросов равно $2^6 = 2^6 = 64$. В рассматриваемой конкретной популяции встретились только пятьдесят пять комбинаций. С увеличением n обработка по всем комбинациям становится все сложнее и возникает необходимость применения вычислительной машины.

Количество ответов «да» на эти вопросы в группе астматиков $P(S|D_1)$ больше, чем в группе неастматиков $P(S|D_0)$. Из фиг. 20 видно, что ответы неастматиков встречаются в основном в верхней части таблицы, а ответы астматиков — в нижней ее части. Кроме того, определенные комбинации ответов «да» чаще можно найти у страдающих астмой, чем у лиц, у которых этого заболевания нет. Было вычислено отношение правдоподобия $L(D_1; S)$ по условным

вероятностям $P(S|D_1)$ и $P(S|D_0)$ для каждой из пятидесяти пяти комбинаций симптомов:

$$L(D_1; S) = \frac{P(S|D_1)}{P(S|D_0)} = \theta \text{ (обозначение Коллена)}. \quad (10)$$

С байесовской точки зрения отношения правдоподобия, представленные на фиг. 20, вычислены некорректно. Они являются просто отношениями числа случаев наличия астмы у пациентов с определенной группой симптомов к числу случаев отсутствия астмы у пациентов с той же группой симптомов. Такие частоты являются удобными, но не всегда оптимальными оценками отношений правдоподобия. Например, на фиг. 20 наблюдение любой из первых восемнадцати групп симптомов с отношением правдоподобия, равным нулю, будет означать абсолютную уверенность в том, что у пациента не может быть астмы, тогда как наблюдение любой из последних двадцати одной группы с отношением правдоподобия, приближающимся к бесконечности, будет означать абсолютную уверенность в том, что у пациента с такими симптомами есть астма.

Любое из этих заключений нереально; абсолютная уверенность недостижима. Один из способов избежать таких нелепостей, как нулевые или бесконечные отношения правдоподобия, состоит в том, чтобы начинать построение каждой матрицы симптом — заболевание по крайней мере с одного наблюдения в каждой позиции. Это отражает идею, которой последователи Байеса часто придают важное значение: пока не наблюдался какой-либо симптом или заболевание, имеет место полная неопределенность относительно связи между ними. Для относительно редких симптомов и заболеваний оказываются полезными более сложные и тонкие методы оценки отношений правдоподобия по наблюдаемым частотам.

Полученные отношения правдоподобия (θ) сводятся в таблицу в порядке возрастания их значений, как показано на фиг. 21. Вычисление отношения правдоподобия не зависит от предположения о том, что симптомы встречаются независимо. Тем не менее нам представляется, что на данной ступени знания о связи признаков и симптомов не следует заранее предполагать, что имеется зависимость между признаками или, наоборот, что они полностью независимы, и каждое предположение, связанное с зависимостью, должно быть специально изучено. Используя данные, собранные группой Коллена, Баэр [2] сравнил вероятности обнаружения симптома без предположения независимости с вероятностями, вычисленными в предположении, что симптомы статистически независимы. Полученные значения для одного и того же симптома существенно отличаются друг от друга, но Баэр не продолжил свои исследования, чтобы показать, что точность диагностики может быть улучшена, если отказаться от допущения о независимости симптомов.

Отношение правдоподобия θ	Пациенты, страдающие астмой		Пациенты, не страдающие астмой		Пациенты, страдающие астмой $\geq \theta$		Пациенты, не страдающие астмой $< \theta$				
	№	%	№	%	чувствительность (истинно положительные случаи)	α (ложно отрицательные случаи)	специфичность (истинно отрицательные случаи)	100-специфичность (ложно положительные случаи)			
									№	%	№
Область для отрицательных случаев	0	0	70	13,5	230	100,0	0,0	0	0,0	100,0	
	0,09	1	24	4,6	230	100,0	0,0	70	13,5	86,5	
	0,10	16	7,0	353	68,3	229	99,6	0,4	94	18,2	81,8
	0,28	1	0,4	8	1,5	213	92,6	7,4	447	86,5	13,5
	0,34	5	2,2	33	6,4	212	92,2	7,8	455	88,0	12,0
Область для положительных случаев	0,50	1	0,4	5	1,0	207	90,0	10,0	488	94,4	5,6
	2,25	1	0,4	1	0,2	206	89,6	10,4	493	95,4	4,6
	3,37	3	1,3	2	0,4	205	89,1	10,9	494	95,6	4,4
	4,50	2	0,9	1	0,2	202	87,8	12,2	496	95,9	4,1
	6,74	27	11,7	9	1,7	200	87,0	13,0	497	96,1	3,9
	8,24	11	4,8	3	0,6	173	75,2	24,8	506	97,9	2,1
	12,92	23	1,0	4	0,8	162	70,4	29,6	509	98,5	1,5
	20,23	18	7,8	2	0,4	139	60,4	39,6	513	99,2	0,8
	24,73	11	4,8	1	0,2	121	52,6	47,3	515	99,6	0,4
	33,72	15	6,5	1	0,2	110	47,8	52,2	516	99,8	0,2
	∞	95	41,3	0	0	95	41,3	58,7	517	100,0	0,0

Фиг. 21. Использование отношения правдоподобия для установления положительных и отрицательных областей при выделении бронхиальной астмы [9].

Чем больше значение отношения правдоподобия θ , тем более правдоподобно, что пациент страдает астмой. Чем меньше значение θ , тем более правдоподобно, что у пациента нет этого заболевания. Необходимо выбрать такое отношение правдоподобия, которое отделяет пациентов с астмой от пациентов, не имеющих этого заболевания. Выбор его в большой степени произволен и делается на основе определенных критериев, устанавливаемых врачом. Определение этого отношения правдоподобия осуществляется методом статистической проверки гипотез. Этот вопрос будет обсужден в гл. 4 в связи с ошибками наблюдателя, а здесь мы ограничимся только кратким замечанием по этому поводу.

В упрощенном виде метод проверки гипотез для выбора между альтернативными гипотезами связан со следующей задачей: дан результат наблюдения и две гипотезы, каждая из которых определяет распределение наблюдений; нужно сделать наилучший выбор между двумя гипотезами. Нейман и Пирсон отметили, что понятие «наилучший» в любом конкретном случае не имеет определенного смысла, но что можно избрать некое правило поведения, и если строго следовать этому правилу, то оно по отношению к большому числу случаев приведет к максимизации некоторых ожидаемых величин. Такое правило (лемма Неймана — Пирсона) заключается в вычислении правдоподобия результата наблюдения относительно каждой из этих двух гипотез, определении их отношения и принятии первой гипотезы, если отношение превышает некоторую константу. Это правило является достаточно общим и имеет глубокий смысл; в нем ничего не говорится о нормальности законов распределения или равенстве дисперсий, но число альтернативных гипотез обязательно должно быть равно двум. Это делает его удобным для процедур разделения, но, пожалуй, менее полезным в медицинской диагностике, потому что оно ничего не говорит об априорной вероятности гипотезы. В процедурах выделения заболевания это правило используется для установления отношений правдоподобия, которые определяют область с максимальным правдоподобием обнаружения заболевания, если оно имеет место, и с ограниченным правдоподобием ложного обнаружения заболевания.

Шесть вопросов, используемых для обнаружения пациентов, страдающих бронхиальной астмой, подразделили две данные популяции на четыре категории, представленные на фиг. 21: истинно положительная (группа пациентов, действительно страдающих бронхиальной астмой), истинно отрицательная (группа пациентов, действительно не имеющих бронхиальной астмы), ложно положительная (пациенты, которые не имеют бронхиальной астмы, но которые были отнесены к страдающим этим заболеванием согласно ответам на шесть вопросов) и ложно отрицательные (пациенты, которые страдают бронхиальной астмой, но которые были отнесены к группе не имеющих этого заболевания согласно ответам на шесть

вопросов). Ложно отрицательные случаи называются «ошибками первого типа» или ошибками типа 1, а процентная доля ложно отрицательных ответов обозначается термином «альфа-уровень» (α). Ложно положительные случаи называются «ошибками второго типа» или ошибками типа 2, и процентная доля ложно положительных ответов обозначается термином «бета-уровень» (β).

В медицинской практике ложно отрицательные ответы обычно интуитивно считаются более важными ошибками, которых следует избегать. Это мнение, по-видимому, основано на каком-то неписаном законе (с другой точкой зрения читатель может ознакомиться в статье Шеффа [97]). Результатом практики является то, что врач устанавливает некоторый приемлемый для него альфа-уровень, чтобы ограничить число ложных отрицаний. Каким образом он это делает, обычно невозможно выразить ни в терминах экономических факторов, ни каким-либо другим количественным способом.

Например, в случае туберкулеза для величины α будет выбрано, вероятно, значение 1,0% из-за «серьезных последствий» этого заболевания, тогда как для бронхиальной астмы можно допустить достаточно большое значение α , скажем 10%, потому что ошибка в диагнозе не влечет за собой немедленных серьезных осложнений. Как показано на фиг. 21, Коллен выбрал альфа-уровень 10%. При этом он одновременно установил оперативный уровень в 5,6% для ложно положительных ответов и уровень отношения правдоподобия $\theta = 0,50$. В гл. 4 мы рассмотрим некоторые последствия выбора альфа-уровня.

Хотя фиксация альфа-уровня обычно считается средством установления «наилучшей» критической области для процедуры разделения, Корнфилд [13] недавно указал альтернативное определение «наилучшего», заключающееся в минимизации линейной функции двух типов ошибок $\lambda\alpha + \beta$, где величина λ служит мерой нежелательности или ценой ошибок первого типа (ложно отрицательные ответы) относительно ошибок второго типа (ложно положительные ответы). Рубин и Коллен [95] использовали такой метод в своих более поздних исследованиях процедур разделения для пациентов с грыжей и с язвой двенадцатиперстной кишки.

Пока еще не опубликованы результаты оценки эффективности автоматических процедур разделения, основанных на некоторой группе тестов и методе отношения правдоподобия, но общий подход, по-видимому, приведет к успеху. Для того чтобы выбрать «наилучшие» диагностические критерии, их нужно тщательно исследовать и «переисследовать». Метод Неймана может быть эффективным как процедура разделения при определенных условиях, и все же он может оказаться недостаточным для того, чтобы дать глубокое понимание ряда вопросов, интересных для медицинской диагностики. К таким вопросам относятся роль условной независимости симптомов и априорных вероятностей симптомов и заболе-

ваний и связь между заболеваниями (т. е. являются ли они взаимоисключающими или сгруппированными, должны ли они быть разделены на более подробные нозологические формы). Допустим, что мы игнорируем априорные вероятности $P(D)$ при диагностике; как мы можем определить степень уверенности в диагнозе $P(D|S)$ после того, как получены величины $P(S|D)$, $L(S_1;D)$ или $\Omega(S_1;D)$ на основе симптоматики пациента, если мы не знаем, какова была степень уверенности в данном диагнозе, прежде чем были проведены эти наблюдения?

2.6. РАЗРАБОТКА ДИАГНОСТИЧЕСКИХ КРИТЕРИЕВ

В заключение этой главы нам хочется подчеркнуть необходимость пересмотра и переоценки диагностических критериев, которые используются в каждой области медицины. Процедуры принятия решения и вычислительные машины могут быть полезными для медицинской диагностики, если только они помогут в разработке «наилучших» диагностических критериев. В поисках «наилучших» критериев мы не должны быть связаны мышлением в терминах традиционной классификации заболеваний; нужно исследовать возможности применения понятий максимальной ожидаемой величины и минимальных ожидаемых потерь. Эти концепции обсуждаются в гл. 4 (разд. 4.6). В существующих ныне системах классификации заболеваний, безусловно, имеются таксонометрические несообразности. Не лучше ли будет классифицировать заболевание в соответствии со сходством совокупностей признаков, симптомов и лабораторных данных? Не перечеркнет ли это традиционные представления о том, что заболевание классифицируют в соответствии с этиологией для решения вопроса о лечении или прогнозе? Существуют ли шкалы заболеваний, которые перечеркивают границы медицинских специальностей? Что произойдет, если будут найдены диагностические критерии, общие для большого числа областей медицины? Хотя все это — интригующие вопросы, не следует ожидать больших изменений в классификации заболеваний в ближайшем будущем, ибо предстоит еще многое сделать по разработке диагностических критериев и систематическому сбору надежных данных на основе этих критериев, прежде чем мы приблизимся к ответам на эти вопросы.

Каким же образом нужно приступить к этим работам? Необходимо ли развернуть обширные совместные исследования на национальной основе? Мы думаем, что нет. Совместные исследования могут потребоваться со временем, но уже сейчас один исследователь или небольшая группа исследователей могут продвигаться в указанном направлении, не дожидаясь национальных усилий.

Осуществление проекта начинается с определения признаков и симптомов и конструирования матрицы симптомы — заболевания,

аналогичной матрице, приведенной на фиг. 13. Признаки и симптомы используются для определения типа классификации заболеваний. Затем классификация применяется для группировки пациентов. Но поскольку мы первоначально определили признаки и симптомы на основе наблюдений над отдельными пациентами, следует проявить осторожность, чтобы не попасть в порочный круг определений, классификаций и наблюдений над отдельными пациентами. Классификации всегда являются неполными; мы должны регулярно пересматривать определения симптомов на основе наблюдений над пациентами, чтобы проверять правильность классификаций. Для этого требуется изучение отдельных пациентов; тем самым оправдываются публикации с описанием отдельных случаев.

Матрица симптомы — заболевания составляется после того, как симптомы и заболевания определены. Априорная вероятность заболевания $P(D)$ может быть вычислена на основе вашего медицинского архива; условную вероятность $P(S|D)$ можно взять в виде частоты, полученной также на основе архивного материала клиники, или оценить, исходя из клинического суждения и имея в виду, что величина $P(S|D)$ будет уточняться по мере накопления данных; наконец, если данные для некоторого элемента матрицы отсутствуют, то в соответствующей позиции ставится 1.

Для иллюстрации методов разработки диагностических критериев мы выбрали три примера: работу группы исследователей в одном учреждении, работу исследователей из двух разных учреждений и работу национального профессионального общества.

2.6.1. РАЗРАБОТКА ДИАГНОСТИЧЕСКИХ КРИТЕРИЕВ ИССЛЕДОВАТЕЛЯМИ В ОДНОМ УЧРЕЖДЕНИИ

Вильсон и его сотрудники [130] из отделения рентгенологии Медицинского центра Университета Миссури разработали критерии для отделения доброкачественной язвы желудка от злокачественного перерождения язвы. По литературным данным был составлен список из семидесяти клинических, лабораторных и рентгенологических признаков. В соответствии с этим списком были установлены диагностические критерии. Для обеспечения объективности обследования пациента были созданы штриховые рисунки рентгенологических аномалий, иллюстрирующие принятые определения критериев. Каждая область медицины нуждается в разработке методов, которые наилучшим образом определяли бы принятые диагностические критерии (например, для области рентгенологии, по-видимому, эффективным оказалось использование штриховых рисунков в дополнение к словесному определению).

После того как были описаны диагностические критерии и собраны материалы историй болезни, на основе которых проводилась разработка, Вильсон получил свои данные об относительных частоте

Дифференциальный диагноз между доброкачественной язвой желудка и злокачественной язвой желудка
 Для доброкачественной язвы $P(D_1) = 0,634$, для злокачественной язвы $P(D_2) = 0,366$

	Присутствует	Отсутствует	Нет информации	Симптомы и клинические показатели (S)	Частоты		
					Доброкачественная язва, $P(S D_1)$	Злокачественная язва, $P(S D_2)$	
1	1	2	2				
2	1	2	2	Возраст 20—54	0,473	0,224	
3	1	2	2	Возраст 55—63	0,338	0,328	
4	1	0	2	Возраст 64—99	0,149	0,440	
5	1	0	2	Симптомы, существующие более 6 месяцев	0,819	0,667	
6	1	0	2	Язвенный анамнез	0,605	0,219	
7	1	0	2	Симптомы непроходимости	0,525	0,455	
				Кислотность желудочного сока	0,630	0,300	
Рентгенологические показатели							
8	1	0	2	Ниша меньше 1 см	0,246	0,096	
9	1	2	2	Контуры ниши	0,595	0,121	
10	1	2	2	Контуры ниши	0,130	0,707	
11	1	2	2	Контуры ниши	0,265	0,155	
12	1	2	2	Форма язвы	0,672	0,148	
13	1	2	2	Форма язвы	0,209	0,426	
14	1	2	2	Форма язвы	0,015	0,391	
15	1	2	2	Положение основания язвы по отношению к стенке желудка	0,316	0,123	
16	1	2	2	а) меньше 6 мм в глубину	0,107	0,640	
17	1	2	2	б) 6 мм и больше в глубину	0,455	0,018	
18	1	0	2	в) проникает до серозной оболочки	0,725	0,223	
19	1	0	2	Радиальная конвергенция складок	0,756	0,278	
20	1	0	2	Радиальная конвергенция складок, достигающая кратера язвы	0,930	0,346	
21	1	0	2	Проекция верхушки язвы вне контуров желудка	0,538	0,664	
22	1	0	2	Отношение ширины к глубине равно или больше 1	0,148	0,044	
23	1	0	2	В язве присутствует воздушно-жидкостный уровень	0,070	0,487	
				Стенка желудка утолщена в месте язвы больше чем на 3 мм			
24	1	2	2	Характеристики опухолевидного образования	0,876	0,198	
25	1	2	2	Характеристики опухолевидного образования	0,065	0,457	
26	1	0	2	Язва в центре опухолевидного образования	0,935	0,350	
27	1	0	2	Линия Хемптона	0,154	0,009	
28	1	0	2	Симптом Кармана	0,140	0,133	
29	1	0	2	Перистальтика вялая	0,254	0,475	
30	1	0	2	Перистальтика отсутствует	0,014	0,355	
31	1	0	2	Дефект наполнения на рентгеновском снимке	0,015	0,488	
32	1	0	2	Ригидность стенки желудка	0,054	0,576	
33	1	0	2	Деформация луковицы 12-перстной кишки	0,369	0,129	
34	1	0	2	Укорочение малой кривизны	0,012	0,143	
35	1	0	2	Рельеф складок неразличим	0,010	0,126	
36	1	0	2	Более чем 2/3 желудка вовлечены в процесс	0,010	0,124	
37	1	0	2	Болезненность в области язвы при рентгенологическом исследовании	0,679	0,036	
38	1	0	2	При рентгенологическом исследовании патология не обнаружена	0,350	0,027	
39	1	0	2	Болезненность в эпигастральной области при пальпации	0,836	0,633	
40	1	0	2	Пальпируемая опухоль	0,013	0,231	

Фиг. 22.

Частоты симптомов, признаков и рентгенографических наблюдений для отделения доброкачественной язвы желудка от злокачественной [131].

Инструкции для пользования таблицей: Выделите только подходящие переменные. Специально укажите, какая информация недоступна. Для переменных, заключенных в скобки, выберите только один вариант. Признаки и симптомы можно переписать в порядке уменьшения частот для доброкачественной язвы. Когда будут получены вероятностные данные по многим заболеваниям, врачу можно будет представить набор вероятностных таблиц симптомы — заболевания в виде справочника.

Частоты	Доброкачественная язва, $P(S D_1)$	Злокачественная язва, $P(S D_2)$	
1. Возраст 20—54	0,634	0,366	
2. Возраст 55—63	0,473	0,224	$P = 0,00001$ (100,0%)
3. Возраст 64—99	0,338	0,328	$P = 0,47304$ (52,7%)
4. Симптомы, существующие более 6 мес.	0,149	0,440	$P = 0,00000$ (100,0%)
5. Язва в анамнезе	0,819	0,687	$P = 0,98849$ (98,8%)
6. Симптомы непроходимости	0,605	0,219	$P = 0,00000$ (100,0%)
7. Кислотность желудочного сока	0,525	0,455	$P = 0,19391$ (80,6%)
8. Кислотность желудочного сока	0,630	0,300	$P = 0,00029$ (100,0%)
9. Ширина ниши меньше 1 см	0,246	0,096	$P = 0,00067$ (99,9%)
10. Характеристика ниши а) ровная	0,595	0,121	$P = 0,00000$ (100,0%)
11. Характеристика ниши б) неровная или нечеткие контуры	0,130	0,707	$P = 0,00000$ (100,0%)
12. Характеристика ниши в) волнообразный контур с амплитудой 1 мм	0,265	0,155	$P = 0,01592$ (98,4%)
13. Форма язвы а) правильной формы	0,672	0,148	$P = 0,00000$ (100,0%)
14. Форма язвы б) неправильной формы	0,209	0,426	$P = 0,00004$ (100,0%)
15. Форма язвы в) асимметричной формы	0,015	0,391	$P = 0,00000$ (100,0%)
16. Основание язвы а) глубина меньше 6 мм	0,316	0,123	$P = 0,00008$ (100,0%)
17. Основание язвы б) глубина 6 мм и больше	0,107	0,640	$P = 0,00000$ (100,0%)
18. Основание язвы в) проникает до серозной оболочки	0,455	0,018	$P = 0,00000$ (100,0%)
19. Радиальная конвергенция складок	0,725	0,223	$P = 0,00000$ (100,0%)
20. Радиальная конвергенция складок, достигающая кратера язвы	0,756	0,278	$P = 0,00000$ (100,0%)
21. Проекция верхушки язвы вне контура желудка	0,930	0,346	$P = 0,00000$ (100,0%)
22. Отношение (ширина язвы)/(глубина язвы) ≥ 1	0,538	0,664	$P = 0,02391$ (97,6%)
23. Язва имеет воздушно-жидкостный уровень	0,148	0,044	$P = 0,00312$ (99,7%)
24. Стенка желудка утолщена в месте язвы больше чем на 3 мм	0,070	0,487	$P = 0,00000$ (100,0%)

24. Характеристика опухолевидного образования. Гладкое или угловатое	0,876	0,198	$P = 0,00000$ (100,0%)
25. Характеристика опухолевидного образования. Неровные или нечеткие контуры	0,065	0,457	$P = 0,00000$ (100,0%)
26. Язва в центре опухолевидного образования	0,935	0,350	$P = 0,00000$ (100,0%)
27. Линия Хемптона	0,154	0,009	$P = 0,00001$ (100,0%)
28. Симптом Кармана	0,140	0,133	$P = 0,52903$ (52,9%)
29. Перистальтика вялая	0,254	0,475	$P = 0,00105$ (99,9%)
30. Перистальтика отсутствует	0,014	0,355	$P = 0,00000$ (100,0%)
31. Дефект наполнения	0,015	0,483	$P = 0,00000$ (100,0%)
32. Ригидность стенки желудка	0,054	0,576	$P = 0,00000$ (100,0%)
33. Деформация луковицы 12-перстной кишки	0,369	0,129	$P = 0,00001$ (100,0%)
34. Укорочение малой кривизны	0,012	0,143	$P = 0,00002$ (100,0%)
35. Рельеф складок неразличим	0,010	0,126	$P = 0,00002$ (100,0%)
36. Более 2/3 желудка вовлечены в процесс	0,010	0,124	$P = 0,00002$ (100,0%)
37. Болезненность в области язвы при рентгенологическом исследовании	0,679	0,036	$P = 0,00000$ (100,0%)
38. При рентгенологическом исследовании патология не обнаружена	0,350	0,027	$P = 0,00025$ (100,0%)
39. Боли в эпигастральной области	0,836	0,633	$P = 0,99255$ (99,3%)
40. При физическом обследовании пальпируется опухоль	0,013	0,231	$P = 0,00022$ (100,0%)

Фиг. 23.

Частоты для симптомов, признаков и рентгенографических данных при отделении доброкачественной язвы желудка от злокачественной. Фактор « P » указывает, насколько данные частоты значимы для дифференцирования доброкачественной язвы от злокачественной. Заметим, что признак Кармана (№ 28), например, не существует при дифференцировании доброкачественной язвы от злокачественной. [Так как $P(S | D_1)$ почти равно $P(S | D_2)$, и, следовательно, фактор P близок к 0,5. —Прим. пер.] [131].

тах $P(D)$ и $P(S|D)$ для доброкачественных и злокачественных язв. Затем была проанализирована новая серия верифицированных историй болезни с использованием этих априорных вероятностей и теоремы Байеса. При первом исследовании были выделены семнадцать статистически важных переменных, а в дальнейшем список был расширен до тридцати одной переменной (фиг. 22). С другой стороны, тридцать четыре переменные были исключены, хотя они и упоминались в литературе.

По поводу работы Вильсона можно сделать два замечания. Во-первых, проведенное исследование помогло выявить некоторые несоответствия в определениях диагностических критериев для доброкачественной и злокачественной язв желудка. Во-вторых, Вильсон сделал «общеупотребительной» свою модель для разделения доброкачественной и злокачественной язв желудка, представив диагностические критерии, поясняющие рисунки и априорные вероятности $P(D)$ и $P(S|D)$. Другие рентгенологи могут теперь принять его критерии, а также величины $P(D)$ и $P(S|D)$. Если же этого не произойдет, то можно будет по крайней мере выяснить большую часть причин разногласий, так как предложенный тип «общепринятой модели» позволяет выявить важные различия довольно быстро. Следующим шагом является пересмотр диагностических критериев и повторение соответствующих наблюдений, до тех пор пока не будет достигнута хотя бы приближенная согласованность первоначально расходившихся мнений. Результатом должно быть общее соглашение между рентгенологами относительно набора «наилучших» диагностических критериев и значений вероятностей $P(S|D)$ для дифференциации доброкачественной и злокачественной язв желудка.

2.6.2. РАЗРАБОТКА ДИАГНОСТИЧЕСКИХ КРИТЕРИЕВ ИССЛЕДОВАТЕЛЯМИ ИЗ ДВУХ РАЗЛИЧНЫХ УЧРЕЖДЕНИЙ

Если исследователи из двух различных клиник хотят объединить данные для получения вероятностей $P(S|D)$, то нужно, чтобы они пришли к соглашению о диагностических критериях и изучили возможные различия в рассматриваемых популяциях пациентов. Опыт Нуджента [75] в Метаболической клинике медицинского колледжа Университета Юты иллюстрирует эти положения. Нуджент изучал диагностику синдрома Кушинга. Для увеличения объема исходных данных он решил объединить свой материал с материалом Лиддла из Эндокринологической клиники Вандербилтского университета. Исследователи достигли согласия о диагностических критериях, за исключением ряда различий, относительно которых Нуджент [75] замечает: «...критерии, использованные Лиддлом в его раннем сообщении, отличались от приведенных в табл. 2 лишь тем, что он рассматривал слабость и экхимоз как присутствующие

признаки на основании убедительного анамнеза и не требовал объективной демонстрации этих признаков; в Университете Юты надежными признаками считали только объективно обнаруженные пурпур или слабость, и эти признаки записывались как присутствующие только в тех случаях, когда они были обнаружены при физическом обследовании».

Были рассмотрены и другие различия в популяции в Юте и Вандербилте, такие, как влияние на величину гематокрита, которое может оказывать высота местоположения клиники (Солт-Лейк-Сити, шт. Юта, находится на высоте 1500 м; Нешвилль, шт. Теннесси, — на высоте 150 м). После учета этих различий можно было вычислить величину $P(S|D)$ по объединенным данным и проверить первоначальные вероятности на новой группе пациентов. Нуджент указал, что априорная вероятность $P(D)$ составляла 0,25 для Юты и 0,19 для Вандербилта. Это могло оказать влияние на вычисления $P(D|S)$ по формуле Байеса и, по-видимому, могло придать «индивидуальный» характер диагнозам в Юте и Вандербилте.

Нуджент [75] заключает свою работу замечанием: «Местные условия, влияющие на исходные популяции и интерпретацию признаков (принятых в этой работе), определяют частоты заболеваний и их признаков в других эндокринологических клиниках. Врач в данной эндокринологической клинике при определении частоты заболеваний и их признаков может отдать предпочтение своему собственному опыту. Если он не располагает нужной информацией, то он может взять в качестве первой оценки данные относительно частот, используемые в этой работе. Наш опыт применения изложенного метода вычисления вероятности синдрома Кушинга подтверждает необходимость более объективных критериев при идентификации клинических признаков».

2.6.3. РАЗРАБОТКА ДИАГНОСТИЧЕСКИХ КРИТЕРИЕВ ПРОФЕССИОНАЛЬНЫМ ОБЩЕСТВОМ

Некоторые диагностические критерии могут быть полезными для одного врача или небольшой группы врачей, но в какой-то момент времени возникает необходимость в согласованных критериях для большой группы медиков. Возможным механизмом для достижения такого соглашения может быть профессиональное общество. Несколько обществ уже предпринимали совместные исследования для определения диагностических критериев. Среди этих исследований мы хотели бы отметить опыт Американской ассоциации по ревматизму в разработке диагностических критериев для ревматоидного артрита.

В 1954 г. на рабочих конференциях Американской ассоциации по ревматизму было высказано требование о разработке более четких диагностических критериев для ревматоидного артрита. Был

назначен комитет [93], который впоследствии предложил критерий классификации для определенных, вероятных и возможных ревматоидных артритов, а затем добавил список исключений. Первоначальные исследования включали оценку чувствительности и специфичности диагностических критериев. Эти критерии широко использовались в течение приблизительно двух лет, и члены комитета получили много замечаний от коллег. Стало очевидным, что критерии недостаточно точны. В 1959 г. были опубликованы результаты пересмотра диагностических критериев 1958 г.; была добавлена четвертая категория — классический ревматоидный артрит.

Пересмотренные критерии (1959 г.) приведены ниже. Заметим, что для точного определения классической, определенной, вероятной и возможной групп использовано минимальное число критериев, причем все выделенные подгруппы из пяти симптомов одинаково важны. Существенное значение имеет фактор времени (например, в критериях классического ревматоидного артрита с первого по пятый включительно признаки или симптомы, связанные с суставами, должны проявляться непрерывно в течение по крайней мере шести недель). Исключающие критерии помогают довольно точно отделять пациентов, которых в дальнейшем не нужно рассматривать как имеющих ревматоидный артрит.

Классический ревматоидный артрит

Этот диагноз требует наличия семи из приведенных ниже критериев. В критериях с первого по пятый суставные признаки или симптомы должны проявляться непрерывно в течение по крайней мере шести недель. (Любой из признаков, перечисленных в разделе «Исключения», указывает на то, что пациент не принадлежит к этой категории.)

1. Утренняя скованность.
2. Боль при движении или чувствительность по крайней мере в одном суставе (наблюдается врачом).
3. Опухание (утолщение мягких тканей или жидкость, а не только костное разрастание) по крайней мере в одном суставе (наблюдается врачом).
4. Опухание (наблюдается врачом) по крайней мере еще одного сустава (любой промежуток времени между двумя суставными атаками, когда отсутствуют жалобы, может быть не больше трех месяцев).
5. Симметричное опухание сустава (наблюдается врачом) с одновременным поражением этого же сустава по обеим сторонам тела (двустороннее поражение фаланговых, пястно-фаланговых или предплюсно-фаланговых сочленений учитывается при наличии абсолютной симметрии). Поражение концевой фаланговой сустава не удовлетворяет этому критерию.

6. Подкожные узелки (наблюдаются врачом) на костных выступах, на разгибательных поверхностях или в околосуставных областях.

7. Типичные для ревматоидного артрита изменения, найденные при рентгенологическом обследовании (которые должны включать по крайней мере костную декальцинацию, расположенную либо в самом суставе, либо большей своей частью вокруг пораженных суставов, а не только дегенеративные изменения). Дегенеративные изменения не исключают пациентов из любой группы, классифицированной как ревматоидный артрит.

8. Положительная реакция агглютинации — обнаружение «ревматоидного фактора» любым методом, который в двух лабораториях был положительным не более чем в 5% случаев для контрольной нормы, или положительная реакция агглютинации со стрептококком.

9. Скудный муциновый преципитат из синовиальной жидкости (с хлопьями и мутным раствором).

10. Характерные гистологические изменения в синовиальной оболочке, включающие три или больше из следующих наблюдений: выраженная ворсинчатая гипертрофия; пролиферация поверхностных синовиальных клеток, часто с отграничением; заметная инфильтрация клеток хронического воспаления (при преобладании лимфоцитов или клеток плазмы) с тенденцией формирования «лимфоидных узелков»; отложение плотного фибрина на поверхностях или интерстициальное отложение; наличие центров клеточного некроза.

11. Характерные гистологические изменения в узелках, где наблюдаются грануломатозные участки с центральными зонами клеточного некроза, окруженные гистиоцитами, а также периферический фиброз и хроническая воспалительная клеточная инфильтрация, преимущественно периваскулярная.

Определенный ревматоидный артрит

Этот диагноз требует наличия пяти из вышеуказанных критериев. В критериях с первого по пятый суставные признаки или симптомы должны быть непрерывными в течение по крайней мере шести недель. (Любой из признаков, перечисленных в разделе «Исключения», указывает на то, что пациент не принадлежит к этой категории.)

Вероятный ревматоидный артрит

Этот диагноз требует наличия трех из вышеуказанных критериев. По крайней мере в одном из критериев с 1-го по 5-й суставные признаки или симптомы должны быть непрерывными хотя бы в

течение шести недель. (Любой из признаков, перечисленных в разделе «Исключения», указывает на то, что пациент не принадлежит к этой категории.)

Возможный ревматоидный артрит

Этот диагноз требует наличия двух из следующих критериев, причем общая продолжительность суставных симптомов должна быть по крайней мере три недели. (Любой из признаков, перечисленных в разделе «Исключения», указывает на то, что пациент не принадлежит к этой категории.)

1. Утренняя скованность.
2. Чувствительность или боль при движении (наблюдается врачом) с повторными или непрерывными болями в течение трех недель (в анамнезе).
3. Опухание суставов в прошлом или при осмотре.
4. Подкожные узелки (наблюдаются врачом).
5. Повышенная скорость оседания эритроцитов или С-реактивный белок.
6. Иррит.

Исключения

1. Типичная сыпь диссеминированной красной волчанки (в форме «бабочки», закупорка волосяного мешочка и наличие областей атрофии).
2. Высокая концентрация клеток красной волчанки (четыре или больше в двух мазках, приготовленных из гепаринизированной крови, стоявшей в термостате не свыше 2 час).
3. Гистологическое доказательство узелкового периартериита с сегментарным некрозом артерий в сочетании с узелковой лейкоцитарной инфильтрацией, распространяющейся периваскулярно и с тенденцией к эозинофилии.
4. Слабость шеи, туловища и мышц глотки или мышечная припухлость, характерная для дерматомиозита.
5. Явная склеродермия (не ограничивающаяся пальцами).
6. Клиническая картина, характерная для ревматической атаки с летучими поражениями суставов и наличием эндокардита, особенно если она сопровождается подкожными узелками, краевой эритемой или хореей. (Повышенный титр антистрептолизина не исключает диагноза ревматоидного артрита.)
7. Клиническая картина, характерная для подагрического артрита с острыми приступами опухания, красноты и боли в одном или более суставах, особенно если колхицин оказывает благоприятное действие.
8. Подагрические отложения в суставах.

9. Клиническая картина, характерная для острого инфекционного артрита бактериального или вирусного происхождения с наличием острого источника инфекции или в тесной связи с болезнью инфекционного происхождения; озноб; лихорадка; острое поражение суставов, обычно сначала летучего характера (особенно если имеются микроорганизмы в суставной жидкости или наблюдается реакция на лечение антибиотиками).

10. Туберкулезные палочки в суставах или гистологически доказанный суставной туберкулез.

11. Клиническая картина, характерная для синдрома Рейтера с уретритом и конъюнктивитом, сопровождающаяся острым поражением суставов, обычно сначала летучего характера.

12. Клиническая картина, характерная для синдрома плечо — рука: одностороннее поражение плеча и руки с диффузным опуханием руки и последующей атрофией и контрактурами.

13. Клиническая картина, характерная для гипертрофической легочной остеоартропатии с «барабанными палочками» (пальцы) и (или) гипертрофический периостит вдоль диафизов длинных костей, особенно при наличии поражения легких.

14. Клиническая картина, характерная для нейроартропатии, с уплотнением и деструкцией костей, вовлеченных суставов и сопутствующими признаками.

15. Гомогентизиновая кислота в моче, обнаруживаемая макроскопически при алкалинизации.

16. Гистологическое доказательство саркоида или положительная реакция Квейма.

17. Множественная миелома, доказанная по заметному увеличению плазменных клеток в костном мозгу или белка в моче по Бенсу—Джонсу.

18. Повреждения кожи, характерные для узловой эритемы.

19. Лейкемия или лимфома с характерными клетками в периферической крови, в костном мозге или тканях.

20. Агаммаглобулинемия.

Через два года после того, как были введены исправленные критерии, Кобб [7] опубликовал сообщение о полученном опыте, выдержки из которого мы приводим, чтобы проиллюстрировать некоторые из проблем и достижений, отмеченных членами Комитета:

«Когда собрался Комитет, было немедленно достигнуто соглашение о том, что нужно сделать попытку перечислить все известные проявления ревматоидного артрита. В результате был составлен огромный список, из которого было сразу же вычеркнуто большое число пунктов по разным причинам. Например, признак утомления был недостаточно диагностически значимым. Дюпюитреновская контрактура, по-видимому, встречается слишком редко, чтобы оказать помощь в диагностике, а парестезии, видимо, были

слишком подвержены субъективным вариациям наблюдателя, чтобы иметь реальную диагностическую ценность.

Вторым вопросом, по которому Комитет быстро пришел к соглашению, была важность разделения данного заболевания на несколько подгрупп в соответствии с тяжестью заболевания. Действительно, при выборе категорий «определенный», «вероятный», «возможный» был использован принцип совместного учета диагностической ясности и тяжести заболевания. Наряду с этим стояла цель — сделать категорию «определенный», насколько это возможно, специфичной, причем допускался риск исключить из нее некоторые случаи, которые следовало бы включить, если бы мы действовали наилучшим образом. С другой стороны, для категории «возможный» была принята максимальная чувствительность, хотя допускался риск включения некоторых случаев, которые в действительности должны были быть исключены. Такой подход представлялся разумным, так как означал, что в группе «определенный» может оказаться только очень немногих лиц, не страдающих ревматоидным артритом, а группа «возможный» должна исключать относительно немногих человек, страдающих ревматоидным артритом, хотя она может включать значительное число тех, у кого нет этого заболевания. Оправдание этого последнего пункта состоит в том, что, когда имеются критерии «возможного» ревматоидного артрита, возможность этого диагноза должна всегда учитываться, даже если она может быть исключена при дальнейшем исследовании.

Третье, что сделал Комитет, — это принял принцип статистических решений при отборе наиболее полезных критериев; правда, этот вопрос решался медленно, так как стало казаться, что статистические и клинические суждения приводят к сходным заключениям. К счастью, удалось получить некоторые материалы по историям болезни, перенесенные на перфокарты; эти материалы были связаны с исследованиями, проводимыми в Питтсбурге. Так как исследование проводилось тремя врачами, имевшими, по-видимому, сходные мнения, было решено дополнить его, собирая истории болезни пациентов у нескольких врачей, которые, как было известно, в основном интересуются ревматическими заболеваниями. Каждого врача попросили выбрать из своей клиники или частной практики наиболее свежие истории болезни пяти пациентов с определенным ревматоидным артритом, пяти пациентов с вероятным ревматоидным артритом и пяти пациентов без очевидных признаков ревматоидного артрита. На основании этих историй болезни была заполнена подробная таблица симптоматики для дальнейшего изучения. В общем было получено 332 хорошо описанных случая из 19 городов.

Процедура, используемая при анализе этого материала, состояла просто в вычислении чувствительности и специфичности каждо-

го признака или группы признаков, как это обычно делают при оценке нового лабораторного исследования. Подробности работы хорошо описаны в докладе Комитета (1957 г.).

Поскольку комбинации признаков, сходные с теми, которые выбираются в качестве критериев, могут встретиться при других ревматических заболеваниях, было решено ввести список признаков, исключающих соответствующую историю болезни из дальнейших исследований, хотя этот случай может быть описан и исследован отдельно. Этот дополнительный список симптомов был разработан главным образом на основе клинического опыта и не являлся предметом какой-либо статистической оценки.

В конце концов было достигнуто соглашение, что предложенные критерии должны быть опубликованы и испытаны в течение двух лет. За этот период были собраны замечания специалистов, пользовавшихся предложенными критериями, и описание пересмотренной методики было опубликовано в 1959 г. Принципиальным изменением явилось добавление еще более ограниченной категории «классический ревматоидный артрит». Важно, чтобы продолжался регулярный пересмотр этих критериев в свете новых знаний, ибо в противном случае они могут в действительности замедлить прогресс, способствуя созданию предубеждений, из-за которых новые открытия будут восприниматься с трудом. Может быть, разумно было бы, чтобы следующий пересмотр был осуществлен Комитетом, составленным преимущественно из новых членов, но таких, которые достаточно связаны с этой областью, чтобы обеспечивалась преемственность.

Дополнительные выгоды, связанные с установлением критериев

Кроме полезных результатов от проведенной разработки, которые были отмечены выше, принятие указанных критериев приведет к некоторым дополнительным выгодам. Во-первых, многие из специалистов, занимающихся обучением студентов-медиков, нашли, что эти критерии полезны как некая база при обсуждении заболевания. Они оказались полезными даже для тех специалистов, которые полностью или частично не согласны с конкретными критериями, так как они при этом все же могут служить основой для дальнейшей работы и, кроме того, облегчают выражение различных точек зрения.

Это сразу же приводит нас ко второму полезному результату, а именно возникновению дискуссии. Так, возник спор относительно того, что же такое ревматоидный артрит. Конечно, до тех пор пока спор не переходит в стадию эмоциональной защиты той или другой точки зрения, он полезен и приводит к существенному прояснению представлений.

С третьим результатом члены Комитета столкнулись с самого начала, хотя в процессе использования установленных критериев он проявлялся довольно медленно. Здесь мы имеем в виду преподнесенный нам урок точности; ведь мы должны постоянно учитывать, насколько точными являются наши диагностические средства. Комитет был шокирован, когда обнаружилось, что 30% случаев, которые тот или иной врач мог бы назвать определенным ревматоидным артритом, не удовлетворяли оптимальному набору критериев, которые установил Комитет. Разумеется, частично это объяснялось тем, что некоторые наблюдения, относящиеся к части пациентов из выборки, не были записаны. Тем не менее, если отбросить те случаи, где информация была пропущена, все же оставались случаи, которые большинство врачей рассматривали бы как определенный ревматоидный артрит, но которые не удовлетворяли выбранным критериям, и, наоборот, имело место небольшое число случаев, которые удовлетворяли критериям, но которые большинство врачей не рассматривали бы как определенный ревматоидный артрит.

Четвертый результат состоит в том, что, работая над этой проблемой, мы кое-что узнали о методологии. Прежде всего стало очевидным преимущество того, что в нашем распоряжении имелась и некоторая общая популяция и клиническая выборка. С одной стороны, специфичность проявления наилучшим образом может оцениваться на выборке из общей популяции, потому что специфичность связана с относительной частью пациентов, не страдающих заболеванием и не имеющих рассматриваемых его проявлений. С другой стороны, чувствительность, которая определяется относительной частью пациентов, страдающих заболеванием и имеющих рассматриваемые его проявления, лучше оценивается в клинической выборке, потому что выборка из общей популяции должна была бы быть очень большой, чтобы содержать достаточно случаев для адекватной оценки этого параметра. Добавим далее, что только после проведения трудоемкого анализа всей этой массы данных при помощи простого сортировочного устройства (для подготовки нужных кросс-табуляций) мы смогли оценить достоинства быстродействующих вычислительных машин. Наконец, нужно отметить, что не были предприняты попытки приписать относительные веса различным проявлениям (признакам). Для этого можно было бы обратиться к математическим методам факторного анализа или дискриминантных функций.

Часто обсуждалось желание получить точный диагноз, которое возводится на пьедестал в академической медицине. Но, с другой стороны, подчеркивалась значительная неточность наших сегодняшних диагностических методов и была выражена надежда, что в конце концов различные мнения приведут к более ясному пониманию рассматриваемого заболевания и, может быть, позволят нам

уменьшить «области сомнения». Следовало бы помнить, что в каждодневной практике мы лечим людей, а не болезни. Это означает, что быстродействующая вычислительная машина не угрожает искусству медицины, ее применение скорее предоставит возможность совершенствовать это искусство, так как больше человеческих умов будет освобождено для решения других задач в результате упрощения диагностических критериев».

Глава 3

ИСПОЛЬЗОВАНИЕ БЛОК-СХЕМ И ДЕРЕВА РЕШЕНИЙ В ДИАГНОСТИЧЕСКОМ ПРОЦЕССЕ

Практика медицины — это искусство, а не ремесло; призвание, а не коммерческое дело, призвание, требующее тренировки не только вашего ума, но и сердца.

Сэр Вильям Ослер¹⁾

Человеческая натура не может познать таинства искусства, не имея опыта.

Сократ²⁾

3.1. ВВЕДЕНИЕ

В этой главе мы обращаем наше внимание на исследования, несколько отличающиеся по характеру от работ, которые мы рассматривали в предыдущих главах. В первой группе этих исследований изучается сам клиницист и его роль как человека, принимающего решение. В них показано, что интуиция клинициста может быть предметом научного исследования, и утверждается, что если не признать, что диагноз является результатом познавательной деятельности, то нет возможности передавать диагностический опыт начинающим клиницистам. С этими исследованиями тесно связаны два других, в которых показано, что для обучения медицинской диагностике могут быть разработаны алгоритмы в терминах машинной программы для системы, использующей обычные медицинские истории болезни, и для вычислительной схемы, основанной на системе Сократа. Эти исследования в свою очередь ведут к рассмотрению роли, которую могут играть в медицинской диагностике методы распознавания образов и таксонометрические методы разделения на группы.

3.2. РЕШЕНИЕ ЗАДАЧИ ДИАГНОСТИКИ

Клиницист в роли человека, принимающего решение, стал предметом изучения медиков [1, 91], заинтересованных в улучшении методов медицинского образования, и психологов, исследующих процессы решения проблем и суждения [50, 103, 133]. Интерес к человеку, принимающему решение и выступающему в роли клинического преподавателя, был выражен недавно Адамсом [1] на конференции по диагностическим процессам, когда он заметил, что «вопросы

¹⁾ The master word in medicine, in: *Aequanimitas with Other Addresses*, 2nd. ed., Philadelphia, Blakiston, 1930, p. 368.

²⁾ *Dialogues*, i. v. 201.

преподавателя, обращенные к студентам, могут представлять типы вопросов, которые он задает себе, когда сам решает проблему. Возможно, это открывает путь для изучения способа, которым пользуется преподаватель в процессе решения задачи. Мы должны будем приложить все усилия, чтобы узнать, что происходит в мозгу преподавателя. Один из важных вопросов, по-видимому, — это вопрос о том, что мы можем узнать о процессах, которые в действительности имеют место в мозгу человека, решающего проблемы медицинской диагностики, путем изучения машинных процессов или каким-то другим способом?»

Один подход к вопросу, заданному Адамсом, очевидно, находится в области психологических исследований, которые связаны с мышлением, суждением и решением проблем и используют машинные методы для обработки информации. Кляйнмунтц [48] настолько подробно изучил познавательную деятельность интерпретатора психологических тестов, что смог переложить эту деятельность на язык вычислительной машины. Эта система правил принятия решения, реализованная на вычислительной машине, во многих сложных ситуациях оказалась более точной, чем лучшие из интерпретаторов.

Диагноз может рассматриваться как специфический пример процесса решения проблем, когда клиницист сталкивается с массивом данных, тщательно анализирует этот массив, отбрасывая некоторые аспекты данных, и дополняет имеющуюся информацию новыми данными. Решение проблемы состоит из этиологической формулировки, классификации пациента по таксонометрическим категориям и предписания лечения. Важно иметь в виду возможность того, что, хотя решение проблем диагностики для каждого врача является в высшей степени индивидуальной процедурой, может иметь место сходное поведение при решении этой проблемы для различных клинических специальностей, несмотря на разнообразие тестов и клинических средств.

Для дальнейшего обсуждения мы выбрали определенный подход, в котором с целью моделирования процессов принятия решения клиницистами используются существующие машинные языки для обработки информации. Такой информационно-вычислительный подход является одним из способов рассмотрения психологической деятельности, а использование машинных языков обработки информации дает исследователю возможность описать точно и в явном виде системы познавательных структур, которые он строит на основании наблюдаемого поведения [90]. Это направление исследований имеет дело с процессами и функциями и связано с тонкой структурой поведения.

Концепция моделирования поведения человека с помощью машинной программы, возможно, недостаточно знакома читателю. Ньюэлл и Саймон [73, 99] представили убедительные доказательст-

ва того, что электронная вычислительная машина может выполнять сложный комплекс процессов, которые очень напоминают процессы, наблюдаемые при мышлении. Эти исследователи собирали данные у испытуемых, которых просили «думать вслух» во время решения проблем. Комментарии испытуемых записывали на магнитофонную ленту, а затем пытались составить машинные программы, которые моделировали бы поведение в соответствии с этими записями. Авторы рассматривают программу, способную моделировать поведение, как теоретическую систему психологических процессов и структур, лежащих в основе поведения. Предполагают, что такие теоретические построения равносильны теориям, выраженным словами или математическими символами, и подчиняются такой же теории адекватности [90]. Важность использования этого подхода к изучению процесса принятия клинического решения объясняется тем, что он заставляет исследователя вполне строго определять рассматриваемые процессы.

Поскольку метод «мышления вслух» оказался полезным при анализе процесса поведения, он был применен Кляйнмунтцем [50] для исследования процесса медицинской диагностики. Метод «мышления вслух» помог доказать, что решение проблем может быть представлено как последовательность операций, которые могут иметь иерархическую связь и наилучшим образом иллюстрируются при помощи «дерева решений». Саймон [98] отметил, что «дерево будущих поведений» было независимо использовано в таких разнообразных приложениях при попытках построить основу теории поведения, что оно может стать действительным ядром нового поведенческого направления в науке.

Недавнее замечание Тейлора [116] может служить полезным резюме относительно перспектив этого подхода к изучению диагностических процессов решения: «Мое собственное убеждение заключается в том, что более адекватное понимание процессов мышления может быть достигнуто путем использования подхода, применяемого при обработке информации».

Я верю, что использование мышления вслух неоднократно показало свою плодотворность и что интроспективный анализ полученного опыта может служить полезным источником гипотез, касающихся рассматриваемых процессов. Но я также уверен, что такие гипотезы обязательно должны быть предметом тщательной проверки. Один из методов проверки состоит в моделировании — путем представления выдвинутой гипотезы в форме машинной программы и сравнения поведения программы с поведением человека».

Кляйнмунтц объединил эвристический и информационный подходы с методом «мышления вслух» при исследовании интерпретации геометрических контуров человеком [49], а также процесса принятия решения клиницистом-неврологом [50]. Здесь полезно отметить различие между тремя типами машинных программ для

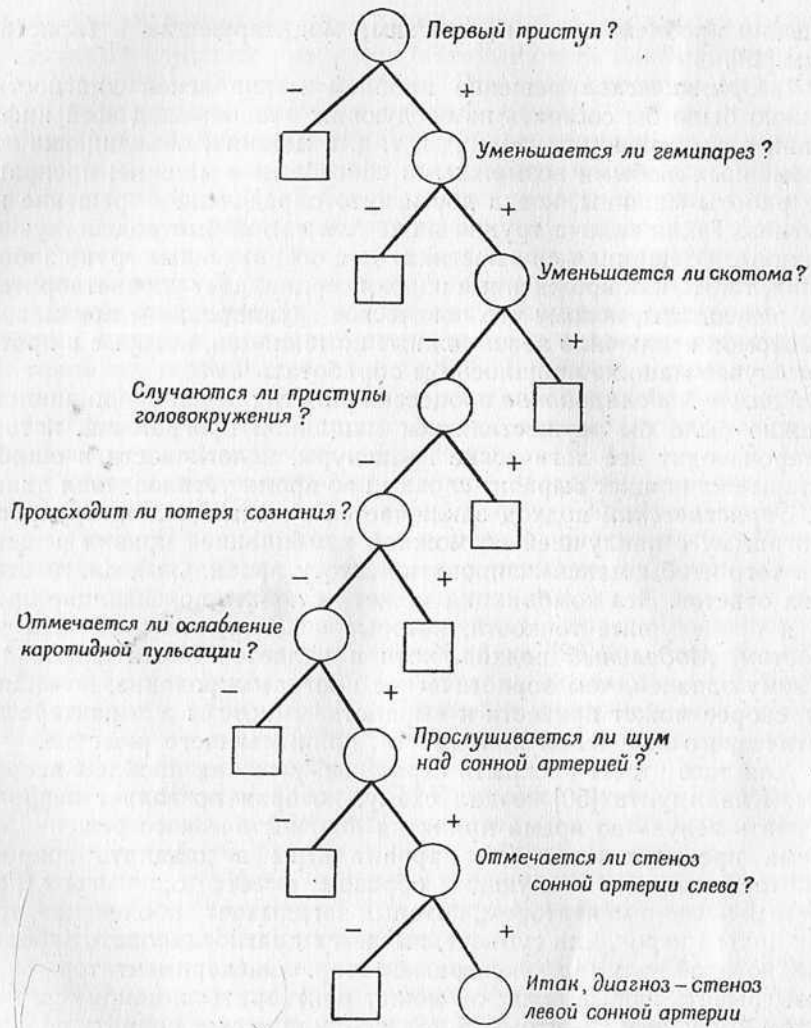
решения проблем: алгоритмическим, моделирующим и эвристическим [49].

Алгоритмическое решение проблем медицинской диагностики должно было бы состоять из следующих этапов: ввод всей информации о симптомах, признаках и т. д. в машину; объединение этих переменных любыми возможными способами в машине; прекращение работы машины, когда достигнуто определенное решение или диагноз. Такая задача трудна даже для самой быстродействующей машины. «Машинная диагностика» для ограниченных групп заболеваний, таких, как врожденные пороки сердца, дает удовлетворительные результаты, потому что логическое группирование признаков и симптомов исключило невозможные комбинации, которые в противном случае машине пришлось бы обработать ¹⁾.

Прямое моделирование процесса принятия решения клиницистом должно было бы осуществляться машинной программой, которая воспроизводит все логические процедуры, нелогичности и ошибки, которые клиницист выразит словами во время установления диагноза. *Эвристический* подход заключается в попытке сконструировать программу с наилучшей возможной комбинацией правил решения, для того чтобы максимизировать частоту правильных диагностических ответов. Эта комбинация может включать поясняющие правила и процедурные тонкости, которые используются опытным диагностом. *Модельный* подход, хотя и является более широкой по объему задачей, чем эвристическое программирование, по-видимому, скорее может привести к выработке гипотезы о характере аналитического поведения клинициста, принимающего решение.

Для того чтобы раскрыть стратегию решения проблем неврологом. Кляйнмунтц [50] создал схему, которая позволяет неврологу «думать вслух» во время принятия диагностического решения. Эта схема представляет собой вариант игры в двадцать вопросов, в которую играют следующим образом: имеется один игрок, называемый экспериментатором, который загадывает заболевание, тогда как другой игрок, или субъект, пытается диагностировать заболевание, которое задумал экспериментатор. Экспериментатор может разыгрывать любые роли: он может притвориться, например, пациентом, имеющим симптомы S_1 , S_2 и S_3 , или может принять роль невролога, который думает об определенном заболевании, характеризующемся симптомами S_1 , S_2 и S_3 . Работа диагноста в любом случае заключается в том, чтобы выяснить наличие определенных симптомов или признаков. Он может, если считает нужным, запросить результаты определенных лабораторных исследований. Необходимо, чтобы экспериментатор был опытным неврологом и мог отвечать на

¹⁾ Существующие системы для диагностики врожденных пороков сердца построены не по указанному здесь методу перебора возможных комбинаций признаков. Такой перебор даже в рамках возможных комбинаций оказался бы для машины невыполнимой задачей. — *Прим. ред.*



Ф и г. 24.

Структура дерева диагностической игры невролога.

Исходная информация: внезапная левосторонняя центральная скотома и правосторонний гемипарез у пациента пятидесяти пяти лет. Процесс принятия диагностического решения может быть отчетливо представлен в виде структуры типа дерева, изображающей поисковую стратегию диагноста. Кружками отмечены узлы, в которых врач может либо поставить дифференциальный диагноз, либо избрать продолжение диагностического процесса путем запроса дополнительной информации. Плюсовые и минусовые ветви, исходящие из узлов, обозначают наличие или отсутствие предыдущего симптома [50].

вопросы субъекта; он должен уметь распознать приемлемость многих симптомов, признаков и лабораторных исследований, которые, возможно, связаны с определенным заболеванием. Вопросы субъекта и ответы экспериментатора записываются на магнитофон для дальнейшего анализа и использования при разработке машинной программы.

Структура дерева решений для некоторой конкретной диагностической игры показана на фиг. 24. Между точкой, в которой невролог задал первый вопрос («Первый ли это случай?»), и точкой, где ставится диагноз, находится девять тестовых узлов и восемь двоичных ветвей — пять положительных и три отрицательных.

Число ветвей на дереве решений, очевидно, зависит от опыта и изобретательности диагноста. Для одной и той же диагностической задачи один врач может использовать дерево решений с четырьмя ветвями, тогда как другому понадобится сорок ветвей. Изменяя количество исходной информации, передаваемой субъекту, и контролируя виды и количество информации, полученной во время игры, можно провести систематическое изучение диагностических деревьев решения, определить структуру оптимального дерева решения и оценить влияние различных признаков и симптомов на точность диагностики.

Кляйнмунтц [50] предложил пять гипотез относительно диагностического процесса невролога-клинициста, последняя из которых заключается в следующем: «В поисковой стратегии невролога используется и кратковременная и долговременная память». Он замечает: «Когда диагносты уже продвинулись вниз приблизительно на половину пути по некоторому определенному диагностическому дереву и их в этой точке просят вспомнить или воспроизвести всю информацию, которую они накопили до этой точки, они, по-видимому, не могут вспомнить те данные, которые не подтверждают определенную дифференциальную диагностическую гипотезу. Другими словами, они могут вспомнить только информацию, относящуюся к определенному диагнозу, и, по-видимому, выборочно забывают не связанные с ним данные».

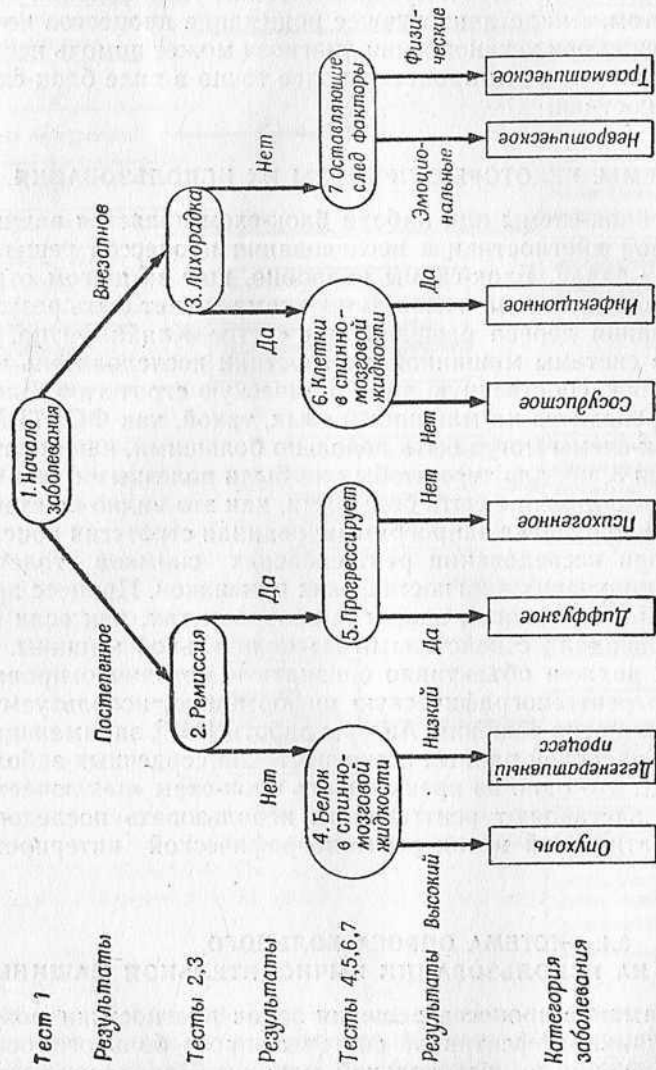
Гипотеза, предложенная Кляйнмунтцем, легла в основу теории решения неврологических диагностических задач, которая может быть проверена, потому что ее компоненты можно задавать в терминах машинной программы. Эта теория хороша лишь в том смысле, что предусматриваемый ею способ обработки неврологической информации неотличим от действий человека, ставящего диагноз. Кляйнмунтцу удалось убедить некоторых клиницистов в том, что решение диагностических проблем является познавательной деятельностью и что она может быть изучена в строгих лабораторных условиях. Это важное достижение, потому что исследования процесса принятия решения во всех областях медицинской диагностики будут успешны только в том случае, если в них примут участие врачи.

Существует ли экспериментальное доказательство связи между деревьями решений в медицинской диагностике и использованием теоремы Байеса? Пока нет. Поможет ли байесовский подход, который использует вероятностно-информационную систему обработки данных, выбирать узлы на дереве решений? Возможно, поможет, и этот вопрос необходимо изучить.

Дерево решения является не единственной моделью, представляющей диагностический процесс. Кляйнмунтц [50] и Уортман [133] недавно приступили к исследованию нескольких других моделей. Одна из них — «дискриминационная сеть», предложенная Фейгенбаумом [26], — приведена на фиг. 25. Эта сеть используется для определения категории неврологических заболеваний; можно заметить, что в ней каждый тест уменьшает число возможных исходов.

Хотя эксперименты по диагностике установили существование некоторой схемы деления заболеваний на категории, не было ясно, чем пользовались врачи, ставившие диагноз: структурой типа деревьев или моделями типа сетей. В некоторых экспериментах диагност не продолжали процесс разбиения на категории до конечного узла, а пытались на некотором этапе сразу определить искомые заболевания. Такое поведение согласуется с гипотезой [50] о том, что диагност полностью не использует преимущество, даваемое малоинформативными или отрицательными данными, для упрощения диагностического дерева всех допустимых гипотез. Такие действия врача связаны с различием между одновременной и последовательной стратегиями поиска, как это имеет место и в задачах формирования понятий [3], и, по-видимому, определяются самым распространенным явлением консерватизма в оценке и использовании вероятностей. То, что характеристики человека близки к характеристикам оптимального устройства обработки вероятностной информации, пожалуй, неудивительно, но в некоторых ситуациях, как показали эксперименты Филлипса, Нейса и Эдвардса [86], отклонения от оптимальности достаточно велики. Работа этих исследователей, возможно, поможет объяснить, почему диагност не способен «выжать все соки» из данных о пациенте, которые он обрабатывает. Этот вопрос будет более подробно обсужден в главе, посвященной ошибкам наблюдателя.

Теперь у некоторых читателей может возникнуть вопрос о полезности исследований диагностического процесса и о том, связаны ли эти исследования с проектами машинной диагностики, которые обсуждались в предыдущей главе. Как мы указывали раньше, эти две области исследования дополняют друг друга, если исследование процесса решения задач диагностики ставится для того, чтобы понять и предсказать, что может сделать человек, ставящий диагноз, и если исследования машинной диагностики определяются как попытка создать математические или статистические модели медицинской диагностики, которые сделают возможным наиболее точное



Ф и г. 25.

Дискриминационная сеть, используемая для определения категорий неврологических заболеваний. Диагностический процесс в неврологии может быть представлен как последовательность тестов. Каждый тест в приведенной выше сети приводит к уменьшению числа возможных исходов [133].

выполнение этой работы. Открытие функциональных связей между признаками, симптомами и диагнозами при разработке машинной диагностики может дать ключ к лучшим методам установления диагноза клиницистом. Аналогично лучшее понимание процессов человеческого суждения при установлении диагноза может помочь исследователю воспроизвести эти процессы более точно в виде блок-схем машинной диагностики.

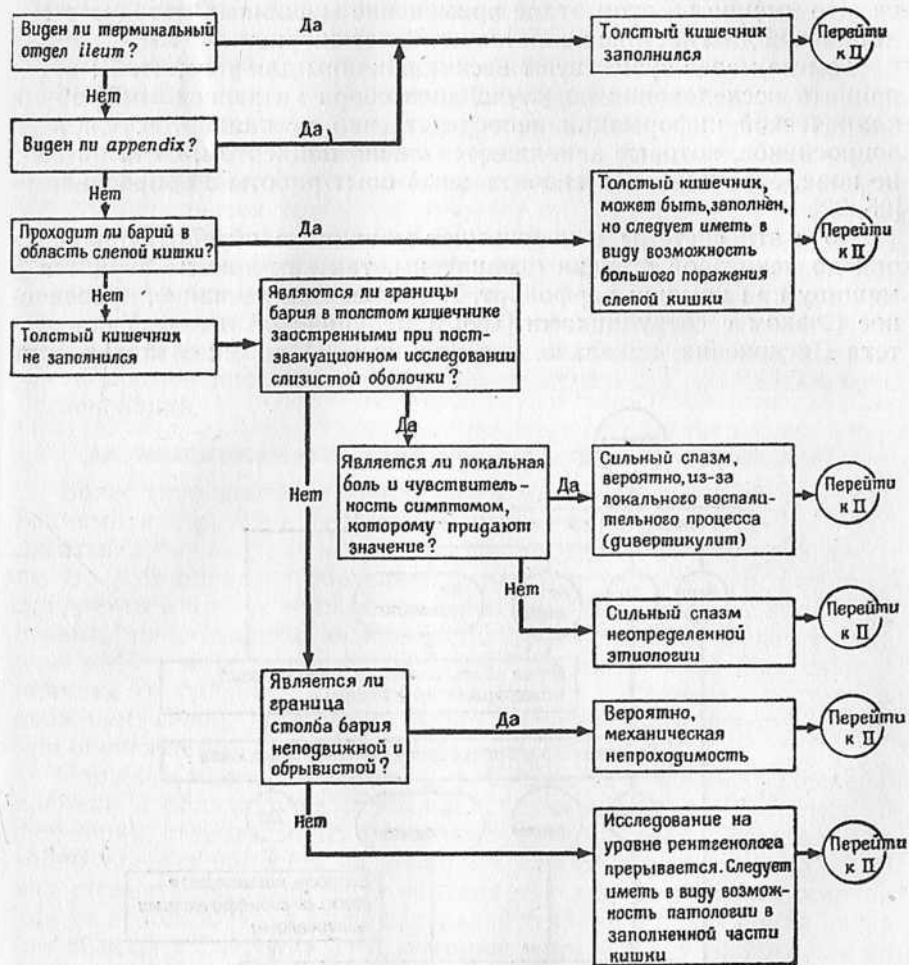
3.3. БЛОК-СХЕМЫ. НЕКОТОРЫЕ ПРИМЕРЫ ИХ ИСПОЛЬЗОВАНИЯ

Разработка блок-схемы или набора блок-схем является важной частью машинной диагностики и исследования процессов решения диагностических задач. Блок-схемы подробно, шаг за шагом отражают процесс постановки диагноза. Блок-схема может быть результатом исследования дерева решения или «сети» Кляйнмунтца, но при разработке системы машинной диагностики исследователь может создавать свою собственную диагностическую стратегию. Блок-схема затем переводится на машинный язык, такой, как ФОРТРАН или IPL-V. Блок-схемы могут быть довольно большими, как показано в приложении 3, но, для того чтобы они были полезными, они вовсе не обязательно должны быть большими, как это видно из схемы на фиг. 26. Здесь приведена запрограммированная стратегия поиска, используемая при исследовании рентгеновских снимков толстой кишки с целью получения диагностических признаков. Процесс принятия решения рентгенологом запрограммирован так, как если бы его действия совпадали с действиями вычислительной машины.

Рентгенолог должен объективно опознать и методично проанализировать всю рентгенографическую информацию, используемую им в процессе принятия решения. Авторы работы [117], занимающиеся машинной обработкой данных и диагностикой сердечных заболеваний, отмечают, что одно из преимуществ блок-схем «заключается в том, что они заставляют рентгенолога использовать последовательный систематический метод рентгенографической интерпретации».

3.4. СИСТЕМА ОПРОСА БОЛЬНОГО, ОСНОВАННАЯ НА ИСПОЛЬЗОВАНИИ ВЫЧИСЛИТЕЛЬНОЙ МАШИНЫ

Лучшее понимание процесса решения задач диагностики может привести к созданию эффективной системы опроса больного, основанной на применении вычислительной машины. Заполнение истории болезни забирает у врача много времени, и из-за ограниченности во времени он часто допускает ошибки в историях болезни. Анамнез и заполнение истории болезни являются тем этапом диагностического процесса, когда у врача есть возможность установить непосредственную связь с пациентом. Этот аспект взаимоотношений



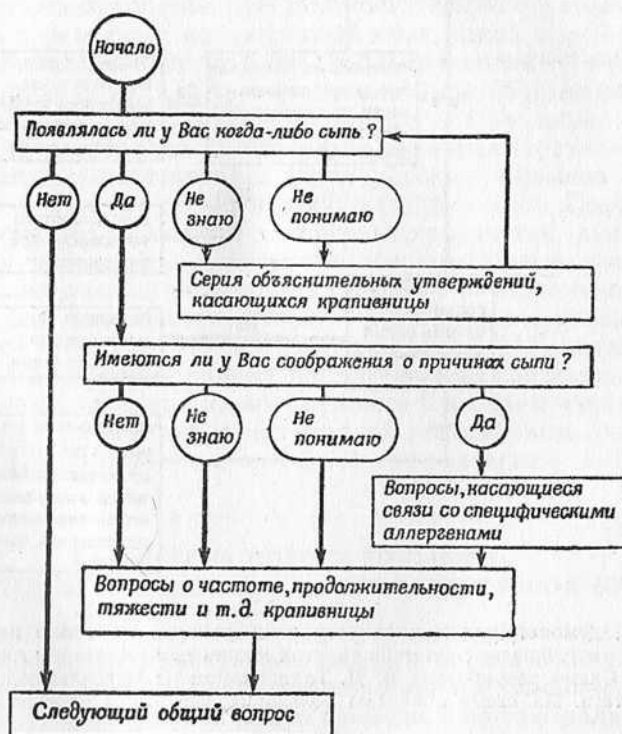
Фиг. 26.

Блок-схема, демонстрирующая запрограммированную стратегию поиска при исследовании рентгеновских снимков толстой кишки для выявления диагностических признаков. Схема разработана В. Д. Тадденхэмом из Пенсильванского университета. На схеме показаны только первые этапы программы.

врача и пациента почти все представители медицины считают неподходящей областью для любого вида автоматизации. Нам кажется, что именно на этом этапе применение машинных методов в медицинской диагностике является наиболее спорным.

Тем не менее существуют веские причины для того, чтобы предпринять исследования по улучшению сбора и записи подробной клинической информации непосредственно от пациентов. Система вопросников, которые заполняются лично пациентом, сама по себе не нова. Уже накоплен значительный опыт работы с вопросниками [25, 92].

Хотя эти системы и используют машинную обработку данных, они до некоторой степени ограничены, так как данные вводятся в машину при помощи перфокарт. Недавнее исследование, проведенное Слэком с сотрудниками [100] в медицинской школе Университета Висконсина, показало, что клинический анамнез может быть



Ф и г. 27.

Часть блок-схемы для машинной системы медицинского опроса, включающая симптомы аллергии. Схема отражает некоторые логические аспекты, использованные при составлении вопросников [100].

введен в машину (LINC) непосредственно от пациента. Машинный анамнез, в котором используется вопросник, относящийся к симптомам аллергии, создан в основном для последующей врачебной диагностики; машинная программа оказалась равноценной врачу, проводящему опрос, в отношении гибкости и полноты и в то же время превзошла врача в отношении логичности, четкости и экономичности. Часть блок-схемы этой системы приведена на фиг. 27. Пациент в функциональном отношении «сопряжен» с вычислительной машиной. Это означает, что вопросы предъявляются пациенту на экране электроннолучевой трубки, а отвечает он при помощи клавиатуры типа пишущей машинки, которая непосредственно связана с управляющим блоком вычислительной машины. Полученные анамнезы хранятся на магнитной ленте и могут быть быстро выведены на печать по требованию врача. Этот вид проведения опроса находится в ранней стадии разработки, но гибкость и потенциальные возможности применения таких систем дают основание для дальнейших исследований.

3.5. МАШИННАЯ «СИСТЕМА СОКРАТА» ДЛЯ ДИАГНОСТИКИ

Более глубокое понимание связей между симптомами и заболеваниями и стратегии решения диагностических задач может дать богатый материал для использования машин в обучении диагностике. Во всех областях образования давно приводятся убедительные аргументы в пользу машинного обучения. Они связаны с преимуществами, вытекающими из индивидуального подхода. Ведется большая работа по машинному обучению в средних и высших школах, начиная со сравнительно простого уровня и кончая достаточно сложным. Очень интересный обзор различных применений машин при обучении был опубликован Сапсом [109].

Сложные машинные системы, работающие в режиме разделения времени и обладающие огромными возможностями обработки информации, становятся все более доступными, однако серьезное внимание следует обратить на проблему выбора материала для обучения студента. Недавно была описана обучающая машина, основанная на использовании вычислительной техники и получившая название «Система Сократа» [115]; мы опишем именно эту систему для иллюстрации активной роли, которую могут играть и вычислительная машина, и студент при изучении медицинской диагностики.

В рассматриваемой системе вычислительная машина ставит перед студентом диагностическую задачу, а затем указывает схему, по которой студент продвигается к решению и формирует свои собственные концепции. Процесс занятия гибок и находится почти полностью под контролем обучающегося. Студент не просто отвечает поочередно на вопросы, он может также их задавать. Основная цель системы — это не просто выдача на печать ответов, а повышение эффективности процесса обучения. Вычислительная машина во-

влекает студента в «разговор», во время которого он пытается решить поставленную задачу. Разговор осуществляется с помощью электрической пишущей машинки (в будущем большинство людей должно научиться хорошо печатать). Студент печатает вопрос или утверждение, после чего вычислительная машина печатает ответ, замечание или, возможно, встречный вопрос.

Для иллюстрации применения «Системы Сократа» в работе [115] рассмотрен гипотетический случай, придуманный сотрудниками Гарвардского центра здравоохранения. Хотя этот случай гипотетичен, читатель, несомненно, заметит непосредственную связь данного метода с исследованиями процесса решения задач диагностики, проведенными Кляйнмунтцем, и убедится в необходимости определения условно-вероятностных связей между заболеваниями и признаками, симптомами и лабораторными исследованиями. Словарь студента для данной задачи состоит из сорока вопросов и тридцати пяти повествовательных утверждений. Ниже этот пример приведен полностью ¹⁾.

«Пожалуйста, напечатайте ваше имя, дату и наклонную черту.
МАЙРА БРИН, АВГУСТ 1963/

Середина декабря. В конце дня в больницу был доставлен пациент. Это хорошо развитый, упитанный 34-летний негр-шахтер. Его привезла жена, которая сообщила, что он в течение четырех дней страдал «гриппом». Последние три дня она давала ему много жидкости и аспирина каждые 6 часов, но не заметила ни малейшего признака улучшения. Сегодня ему внезапно стало хуже. Последние два дня он жаловался на жар и боль в животе. Сегодня утром у него был сильный озноб, настолько сильный, что у него стучали зубы.

Как врач, принявший больного, вы должны поставить диагноз для этого случая. Чтобы получить информацию, необходимую для постановки диагноза, вам нужно выполнить физическое обследование больного и назвать все необходимые, по вашему мнению, лабораторные исследования; они будут проведены по вашему указанию. Это осуществляется путем ввода вопросов в вычислительную машину. Набор вопросов, которые вы можете задать, содержится в словаре, который вам выдан. Машина будет реагировать на каждый вопрос ответом или замечанием.

Когда вы закончите обследование и лабораторные тесты, напечатайте: «продолжим». На этом «кончается рабочий день». Прежде чем вы «пойдете домой», вы можете оставить указания для ночной сестры провести вторую серию лабораторных исследований с тем, чтобы результаты были готовы для вас «завтра». Разумеется, вторая серия будет состоять лишь из тех исследований, которые вы сочтете необходимыми.

¹⁾ Реплики студента напечатаны прописными буквами, а вычислительной машины — строчными. — *Прим. перев.*

Завтра утром вам будут представлены результаты некоторых первых лабораторных исследований, другие не будут готовы до конца дня. На основе информации, которую вы имели возможность получить, вы должны поставить диагноз. Если вы чувствуете, что нет достаточных оснований для заключительного диагноза, вы можете провести новое полное физическое обследование. Можно также запросить результаты второй серии лабораторных исследований.

Вы можете поставить диагноз в любой момент, напечатав название одного из заболеваний, имеющихся в словаре программы. Иногда машина будет обсуждать с вами определенные диагностические возможности. Если вам будут заданы вопросы, вы должны ответить на них, напечатав диагноз или подходящее утверждение из последнего раздела словаря. Начинайте физическое обследование, как только вы к этому готовы.

ФИЗИЧЕСКОЕ ОБСЛЕДОВАНИЕ?

Хорошо. Что именно вы хотели бы узнать?

ОБСЛЕДОВАНИЕ ГРУДНОЙ КЛЕТКИ?

В медицине обычно принято начинать обследование с проверки общего состояния пациента.

ОБЩЕЕ СОСТОЯНИЕ?

Сильно истощен, лежит в постели на правом боку. Выглядит тяжело больным, в поту, сохраняет сознание, однако состояние это неустойчиво. Пароксизмальный кашель с выделением небольшого количества мокроты с прожилками крови ржавого цвета. Дышит с трудом.

ОФТАЛЬМОЛОГИЧЕСКОЕ И ОТОЛАРИНГОЛОГИЧЕСКОЕ ИССЛЕДОВАНИЯ?

Проверьте сначала все важнейшие жизненные показатели!

ТЕМПЕРАТУРА?

103.¹⁾

КРОВЯНОЕ ДАВЛЕНИЕ?

150/84.

ПУЛЬС?

100 в минуту. Временами экстрасистола.

ЧАСТОТА ДЫХАНИЯ?

20 в минуту, поверхностное.

[В этом месте из диалога ради краткости были исключены несколько вопросов и ответов, в том числе разговор об исследовании органов брюшной полости, грудной клетки, глаз и уха -- горла -- носа.]

КАШЕЛЬ?

Есть, вчера усилился.

БОЛЬ В СПИНЕ?

Нет.

¹⁾ 103° F = 39,5° C. — Прим. перев.

ИССЛЕДОВАНИЕ СЕРДЦА?

Перкуторно не увеличено, но имеется слабый систолический шум на верхушке. В остальном норма.

ИССЛЕДОВАНИЕ ОРГАНОВ БРЮШНОЙ ПОЛОСТИ?

Живот мягкий, замечена колющая боль в нижнем правом квадранте, временами распространяющаяся на правую боковую область.

ГРИПП.

Информация пока еще недостаточна для какого-либо заключительного диагноза. Вы невнимательны!

КОСТНО-МЫШЕЧНАЯ СИСТЕМА?

В основном нормальная, за исключением общей слабости и болей.

КОНЕЧНОСТИ?

Нормальные, за исключением того, что ногтевые ложа, кажется, цианотичны, но это трудно определить из-за цвета кожи пациента.

ИССЛЕДОВАНИЯ ПРЯМОЙ КИШКИ?

Вы невнимательны! Пациент слишком тяжело болен, чтобы можно было проводить это исследование.

ЛАБОРАТОРНЫЕ ДАННЫЕ?

Хорошо. Какие лабораторные исследования, по вашему мнению, нужно провести?

ЭРИТРОЦИТЫ?

Они будут подсчитаны. Результаты вам пришлют как можно скорее.

ЛЕЙКОЦИТЫ?

Анализ будет сделан.

[Здесь опущены запросы таких лабораторных исследований, как анализ мочи, рентгенография и культура крови, и обещание машины выдать результаты исследований.]

МАЗОК И ПОСЕВ СЛИЗИ ИЗ НОСА И ГЛОТКИ?

Эти анализы сейчас будут заказаны. Когда вы окончите запрашивать все нужные вам лабораторные данные, скажите, что вы готовы «продолжить».

ПРОДОЛЖИМ.

Вам могут понадобиться повторные лабораторные исследования. Может быть, вы хотите предупредить ночную сестру, чтобы она заказала лабораторные исследования, которые вы захотите посмотреть завтра?

ДА.

Хорошо. Это ускорит дело. Сегодня вы больше ничем не сможете помочь пациенту. Давайте закончим на сегодня. Может быть, вам будет легче установить диагноз завтра, после того как будут доставлены первые результаты лабораторных исследований и вы еще раз осмотрите пациента. Теперь хорошо выспитесь, а завтра, когда вы придете на работу, скажите мне «ПРОДОЛЖИМ».

ПРОДОЛЖИМ.

Сейчас утро, и вы делаете обход больных. Вы подошли к вашему пациенту, и ночная сестра вам докладывает: «Температура оставалась высокой, но пациент провел сравнительно спокойную ночь. Кашель уменьшился, но еще держится. Время от времени в мокроте заметны следы крови. У пациента все еще сохраняется боль в животе и одышка. Начата вторая серия лабораторных исследований». Некоторые результаты первой серии уже получены: анализ мочи — недостаточно жидкости для анализа; эритроциты — 5 миллионов; лейкоциты — 15 500; лейкоцитарная формула крови — сдвиг влево. Когда будут получены остальные результаты, вам об этом сообщат. Прежде чем начинать повторное обследование, может быть, полезно будет рассмотреть диагностический смысл уже имеющейся информации. Вы, возможно, считаете, что еще рано прийти к какому-то определенному заключению. Однако, быть может, вы готовы уже сейчас поставить диагноз? Вы готовы?

НЕТ.

Хорошо. Тем не менее у вас имеется большой объем информации, по крайней мере достаточный для того, чтобы рассмотреть некоторые диагностические варианты (может быть, несколько вариантов одновременно). Есть ли у вас какие-либо предложения?

ДА.

Назовите одно из них.

ИНФАРКТ ЛЕГКОГО.

Хорошо. Плевральная боль, одышка, лихорадка, а также признаки и симптомы, полученные в результате физического обследования, безусловно, соответствуют инфаркту легкого. Думаете ли вы, что озноб и присутствие мокроты ржавого цвета также подтверждают этот диагноз?

НЕТ.

Правильно. Хотя количество лейкоцитов соответствует этому диагнозу, не так ли?

НЕТ.

Неправильно. Количество лейкоцитов соответствует вашему диагнозу. Это еще одно подтверждение вашего диагноза, но мы рассмотрели только одну возможность. Есть ли у вас еще какие-нибудь предположения?

НЕТ.

В этом случае мне хотелось бы обсудить возможность вирусной пневмонии. Тахикардия, лейкоцитоз, учащенное дыхание, озноб, кровянистая мокрота и острая плевральная боль — все это указывает на такой диагноз, не правда ли?

ДА.

Вы сбиты с толку. Но в том, что я только что сообщила, не было ни капли истины; на самом деле все как раз наоборот. Вам следует заново просмотреть курс элементарной медицины, доктор.

[Здесь опущены сообщения о результатах некоторых лабораторных исследований, запрошенных накануне.]

ИНФАРКТ ЛЕГКОГО.

Да, я помню, вам раньше понравился этот диагноз. Безусловно, необходимо принять во внимание эту возможность... Но обдумайте его еще раз, учитывая запись в истории болезни об ознобе, высокой температуре и довольно-таки большом количестве эритроцитов.

ОСТРЫЙ ПИЕЛОНЕФРИТ.

Очень хорошо. Вы должны иметь в виду острый пиелонефрит. Этот диагноз совместим со многими наблюдаемыми у пациента признаками и симптомами, но не со всеми. Не понадобятся ли вам некоторые признаки дизурии, чтобы удостовериться в этом диагнозе? А как же насчет симптомов, касающихся дыхания, в анамнезе и при обследовании? Вам нужно провести переоценку фактов.

ФРИДЛЕНДЕРОВСКАЯ ПНЕВМОНИЯ.

Мудрая мысль, но в данном случае этот диагноз не годится. Ведь вы не наблюдали ни обычной для фридлендеровской пневмонии желеобразной мокроты, ни диплобациллы Фридлендера. Пока отложите этот диагноз.

[Здесь опущено обсуждение некоторых физических обследований, а также неприемлемые диагностические утверждения.]

ПЛЕВРАЛЬНЫЙ ВЫПОТ.

Хорошее предложение, однако следует соблюдать осторожность. Какие у вас есть доказательства за или против этого диагноза?

ЕСТЬ ПОЛОЖИТЕЛЬНЫЕ АРГУМЕНТЫ В ПОЛЬЗУ ДРУГОГО ЗАБОЛЕВАНИЯ.

Имеются ли такие аргументы?

ЭТО ВОЗМОЖНЫЕ, ХОТЯ И НЕ РЕШАЮЩИЕ АРГУМЕНТЫ.

Раз вы не уверены, продолжайте обследование, пока у вас не появятся твердые доказательства.

ЭРИТРОЦИТЫ?

Эритроциты не показательны в такой ситуации. Это исследование будет неоправданным затруднением для пациента.

ЛЕЙКОЦИТЫ?

22 500.

СТАФИЛОКОККОВАЯ ПНЕВМОНИЯ.

Неплохо. Но где же возбудители заболевания? Когда вы выделите изолированный золотистый стафилококк в значительных количествах в пробах жидкости выделений пациента, дайте мне знать.

ПНЕВМОКОККОВАЯ ПНЕВМОНИЯ.

Вы абсолютно уверены?

ДА.

Вы не можете быть абсолютно уверены, пока у вас нет неоспоримых доказательств. Что поможет вам удостовериться в этом диагнозе?

ДАнные, ПОЛУЧЕННЫЕ ПРИ ПОСЕВЕ ИЗ МАЗКА МОКРОТЫ.

Хорошо. Что-нибудь еще?

НЕТ.

Я думаю, вам могут также понадобиться результаты из рентгенографического обследования. Начинайте, и мы посмотрим, что из этого получится.

РЕНТГЕН?

Еще не готов. Попробуйте посмотреть результаты исследования крови.

ФОРМУЛА КРОВИ?

Заметный сдвиг влево. Хотя у вас имеются не все лабораторные данные, сейчас уже достаточно информации, чтобы поставить предположительный диагноз. Попробуйте.

ОСТРЫЙ ПИЕЛОНЕФРИТ.

Довольно трудно придерживаться этого диагноза без соответствующих указаний в анамнезе или признаков дизурии. Кроме того, боль переходит из боковой области в нижнюю часть живота. Вы все еще рассматриваете возможность почечного заболевания? Пересмотрите выводы.

ТУБЕРКУЛЕЗНАЯ ПНЕВМОНИЯ.

Есть ли у вас еще какие-нибудь соображения по поводу диагноза? Неожиданный скачок вверх лейкоцитов и формула крови говорят против последнего диагноза, и к тому же в мокроте не обнаружены туберкулезные палочки.

ИССЛЕДОВАНИЕ МОЧИ?

Исследование показало присутствие белка, небольшое количество гиалиновых цилиндров, немного лейкоцитов, единичные эритроциты.

РЕНТГЕН?

Рентгенография обозначила область уплотнения в правой нижней доле легкого, указывающую на легочный процесс; наступил как раз подходящий момент, чтобы вы пересмотрели основные выводы и поставили другой предположительный диагноз.

НАЙДЕНА ЛИ ЖИДКОСТЬ?

Нет.

МАЗОК МОКРОТЫ?

Грамположительные палочки.

СТРЕПТОКОККОВАЯ ПНЕВМОНИЯ.

Хорошо, но как вы можете отличить ее от любой другой бактериальной пневмонии? Подождите, пока у вас будет бесспорное доказательство.

ВИРУСНАЯ ПНЕВМОНИЯ.

А как же насчет озноба, мокроты ржавого цвета, лейкоцитоза и т. д. — всех доказательств острого лихорадочного бактериального процесса? Вы ошибаетесь, доктор.

ПОСЕВ МОКРОТЫ?

Имеется пневмококк.

СТРЕПТОКОККОВАЯ ПНЕВМОНИЯ.

Невероятно! Вы пренебрегли очевидными данными, которые должны были указать правильный диагноз.

ПНЕВМОКОККОВАЯ ПНЕВМОНИЯ.

Очень хорошо. Теперь это очевидно, хотя в общей картине есть некоторые признаки и симптомы, вводящие в заблуждение. Вы пришли к правильному диагнозу.

[В этом примере студент делал много диагностических предположений. Затраченное время (большая часть которого была использована студентом на принятие решений) составило около часа. В другом случае, при решении этой же задачи более проницательным студентом, правильный диагноз был получен быстрее. Соответствующий диалог приведен ниже; в нем опущены предварительные вопросы. Вычислительная машина, распознав более способного студента, становится менее требовательной, и ее замечания носят более мягкий характер.]

Прежде чем начинать повторное обследование, может быть, полезно будет рассмотреть диагностический смысл уже имеющейся информации. Вы, возможно, считаете, что еще рано прийти к какому-то определенному заключению. Однако, быть может, вы готовы уже сейчас поставить диагноз? Вы готовы?

НЕТ.

Хорошо. Тем не менее у вас имеется большой объем информации, по крайней мере достаточный для того, чтобы рассмотреть некоторые диагностические варианты (может быть, несколько вариантов одновременно). У вас есть какие-либо предложения?

ДА.

Назовите одно из них.

ПНЕВМОКОККОВАЯ ПНЕВМОНИЯ.

Хорошо. Общие признаки легочного процесса налицо. Конечно, вы не выделили возбудителя. Вы думаете, боль в животе и боку подтверждает предположение о пневмококковой пневмонии?

НЕТ.

Правильно. Хотя количество лейкоцитов соответствует этому диагнозу, не так ли?

ДА.

Я тоже так думаю, однако дальнейший анализ должен подтвердить этот вывод. Но мы рассмотрели только одну возможность. Есть ли у вас еще какие-нибудь предположения?

НЕТ.

В таком случае я бы хотела поговорить об инфаркте легкого. К этому диагнозу привлекает внимание наличие сильной лихорадки, мокрота ржавого цвета и озноб; вызывает сомнение одышка, плевральная боль и признаки уплотнения — вы согласны?

НЕТ.

Вы совершенно правы. Я просто хотела проверить, насколько вы внимательны. Теперь давайте посмотрим, как чувствует себя пациент сегодня. Когда будете готовы, начинайте физическое обследование.

[Здесь опущены запросы об общем состоянии и жизненных показателях.]

ПНЕВМОКОККОВАЯ ПНЕВМОНИЯ.

Да, я помню, вам раньше понравился этот диагноз. Вы абсолютно уверены?

НЕТ.

Что поможет вам удостовериться в этом диагнозе?

ДАННЫЕ, ПОЛУЧЕННЫЕ ПРИ ПОСЕВЕ ИЗ МАЗКА МОКРОТЫ.

Хорошо. Что-нибудь еще?

ДАННЫЕ РЕНТГЕНОВСКОГО ОБСЛЕДОВАНИЯ.

Хорошо. Рентгенография обнаружила область уплотнения в правой нижней доле легкого, указывающую на легочный процесс. Лабораторные данные показали наличие грамположительных микроорганизмов в мазке мокроты и наличие пневмококка в посеве. Была ли у вас мысль исключить такие возможности, как аппендицит и острый пиелонефрит? Вы явно не были сбиты с толку ссылкой на боль в брюшной полости. Во всяком случае, вы поставили правильный диагноз».

Около тридцати часов потребовалось врачу и программисту, чтобы подготовить этот медицинский случай. Большая часть времени была затрачена на выбор такой задачи, которая давала бы полную и в то же время достаточно экономную демонстрацию системы. Еще тридцать часов были затрачены на программирование и канцелярскую работу. Используя этот опыт и уже разработанные компилирующие программы, можно составить и запрограммировать аналогичные примеры за значительно меньшее время.

Наши сведения об условных вероятностях $P(D)$ и $P(S|D)$, сетях решений и обучающих машинах постепенно накапливаются; тем не менее в двух интересных областях исследований, примыкающих к этим проблемам, существуют вопросы, на которые пока нет ответов: является ли медицинская диагностика процессом распознавания образов и могут ли быть развиты методы группирования, которые помогут объединить связанные симптомы и заболевания?

3.6. МЕДИЦИНСКАЯ ДИАГНОСТИКА И ПРОЦЕСС РАСПОЗНАВАНИЯ ОБРАЗОВ

Перед нами стоит все тот же мучительный вопрос: как врач ставит диагноз? Мы можем предположить, правда, не имея для этого достаточных оснований, что в начале диагностического процесса врач использует процедуру распознавания образов, чтобы быстро

сосредоточиться на группе возможных заболеваний (процедура разделения для дифференциальной диагностики), а затем обращается к своей памяти, «извлекая» субъективные меры вероятности, чтобы поставить окончательный диагноз. Невозможно создать алгоритм для описания всего диагностического процесса врача, и будет справедливым заметить, что нам в значительной мере недостает понимания этого диагностического процесса. Тем не менее проблемы распознавания образов в медицине остаются в числе наиболее интересных в научном смысле и пользуются довольно высоким приоритетом в связи с тем, что для процедур обследования большое значение имеют экономические факторы (например, необходимо сканирующе-вычислительное устройство для автоматической обработки рентгеновских снимков грудной клетки).

«Распознавание образов» можно рассматривать в терминах машинного моделирования как создание моделей восприятия формы. Воспринимаемыми формами могут быть печатные буквы или цифры, рукописные знаки, группы симптомов, рентгенограммы или хромосомные контуры. С 1957 г. накоплен большой и интересный опыт по исследованию процессов классификации [94] и восприятия образов. Читатель может ознакомиться с обзором [120], содержащим большое количество материалов по этому вопросу из области технических приложений; этот обзор может быть с успехом использован учеными, работающими в области психологии и медицины.

По-видимому, медицинский диагноз отчасти все же является процессом распознавания образов, хотя в приложении методов распознавания геометрических форм к медицинским проблемам не было достигнуто большого прогресса. Тем не менее мы бы хотели обсудить некоторые примеры такого подхода.

В статье [66] мы указывали, что можно фотосканировать рентгеновский снимок грудной клетки, используя ручное сканирующее устройство, но еще нет технических средств для отделения нормальных снимков от аномальных. Таким образом, в 1959 г. не было возможности осуществлять автоматический анализ рентгенограмм; нет такой возможности и в настоящее время. Причина этого — то, что мы не можем выработать критерии для образов, видимых на снимках грудной клетки, например, типа «мелкоочагового инфильтрата» или «петрификата», чтобы создать алгоритм (программу вычислительной машины), который будет отделять нормальные снимки грудной клетки от аномальных. Существенным предварительным шагом в этом направлении являются исследования, в которых была предпринята попытка установить диагностические критерии, коррелированные с некоторыми типовыми штриховыми рисунками, как это делал Лодвик (см. гл. 2).

Многие исследователи продолжают работу по использованию методов распознавания образов для анализа рентгеновских снимков грудной клетки и хромосом. Меерс с сотрудниками [69] начали раз-

работку системы автоматического сканирования рентгеновских снимков грудной клетки при помощи электроннолучевого растра. Они нашли способ сканирования и преобразования малых участков рентгеновского снимка грудной клетки в цифровые данные, которые вводят в память вычислительной машины. Затем изображение преобразуется в матричную форму и выводится на экран электроннолучевой трубки. Предусмотрена возможность регулировки характеристик рентгенографического изображения с целью такого изменения уровней контрастности, чтобы определенные детали изображения выделялись более рельефно. В последующих работах эти ученые сообщили об успешном распознавании контуров рентгенограмм грудной клетки методами сканирования [70]; благодаря этому может быть вычислено отношение размеров сердца и грудной клетки.

В статье [68] изложен опыт работы с системой CYDAC (цитофотометрическая система преобразования данных), которая содержит две независимые системы микроскопов: механическое сканирующее устройство и сканирующее устройство с электроннолучевой трубкой. Электрический выходной сигнал каждого из этих устройств после коррекции и выбора участка преобразуется в двоичные числа и хранится на магнитной ленте для машинного анализа. Это исследование отражает основную проблему, которая возникает при преобразовании оптической информации в цифровую — проблему чрезмерно большого объема данных.

Чтобы определить степень разрешения, которая потребуется для сканирующего устройства, авторы разделили поле сканирования на регулярно размещенные оптически однородные участки, которые были названы элементарными рамками. Используя переменный размер рамки, они обнаружили, что воспроизведение тонкой структуры улучшается с уменьшением размера рамки даже тогда, когда размер рамки становится меньше теоретических пределов оптического разрешения. Для изображений, показывающих тонкие морфологические детали, потребовалось 4800 рамок, чтобы представить лишь малую часть всей интересующей области. В результате оказалось, что при выборе микроскопического поля в $25 \cdot 25 \text{ мк}$ и разрешающей способности в $0,1 \text{ мк}$ понадобится объем памяти в 500 000 бит. Имея это в виду, можно подсчитать тот поистине огромный объем памяти, который потребуется для автоматического анализа рентгенограмм.

В работах [51] и [52] сообщается о быстродействующей системе автоматического анализа биомедицинских изображений, основанной на применении сканирующего устройства FIDAC (ввод данных с пленки в цифровую вычислительную машину) и соответствующей системы машинного программирования FIDACSYS.

Устройство FIDAC использовалось главным образом для высокоскоростного сканирования и анализа фотомикрографического материала (подсчет и идентификация хромосом). Кроме того, что эта

Группа заболеваний	Диагностическая функция группы	Заболевание	Веса симптомов								
			1	2	3	4	5	6	7	8	
Почки (1) RENAL	P (N + R) + + H(RM+NE)	Нефрит (11) Nephritis	I	N, E, P, H, R	S				T		
		Нефроз (12) Nephrosis	I	N, E, P, H, R	S	∅					
		Неврома (13) Nephroma		N, E, P, H, R		∅			A, M		
Нервная система (2) NERVIOUS SYSTEM	U (M + A + R + + ∅ + E)	Неврит (21) Neuritis		E, S	N, U, R, I			T			
		Невроз (22) Neurosis	I	E, S, ∅	N, U, R						
		Неврома (23) Neuroma		E	N, U, R, ∅			A, M			
Глаз, ухо, горло, нос (3) EYE, EAR, NOSE, THROAT	2TP (E + N) + + 3I (2T + R + + N + ∅ + H + + 25)	Синусит (31) Sinusitis	I, S				N, T			U	
		Ринит (32) Rhinitis	I	S, T	R, N			H			
		Тиннит (33) Tinnitus	I, S				N, T				U
		Ирит (34) Iritis		I, T							R, S
		Отит (35) Otitis		I, T							∅, S
		Ретинит (36) Retinitis	I, S, T				N		E, R		
Костно-мышечная система (4) MUSCULO-SKELETAL SYSTEM	2R + D + Y + + M(I + 25) + + 2N (∅ + E)	Артрит (41) Arthritis	I, S, T,				A				
		Миозит (42) Myositis	I, ∅, S, T						HR		
											M

Желудочно-кишечный тракт (5) GASTRO-INTESTINAL SYSTEM	G + L + C + + A [A + MT + + P (S + T)] + 35	Остеит (43) Osteitis		I, S, T			E, ∅				
		Тендонит (44) Tendonitis	I, T			∅, N, S	E			D	
		Глоссит (51) Glossitis	I	S, T			∅	G, L			
		Колит (52) Colitis	I	S, T			∅	G, L			
		Гастрит (53) Gastritis	I, S, T				R	A, G			
		Гепатит (54) Hepatitis	I, S, T			A, E	H	P			
		Гепатома (55) Hepatoma			A, E, M, ∅, T		H, P				
Илеит (56) Ileitis	I, S, T									E L	

Фиг. 28.

Диагностическая модель типа автомата Тьюринга. Схема отражает важные аспекты диагностической модели. Числа в круглых скобках, следующие за названием группы заболеваний, например RENAL (1), относятся к порядку, в котором группы заболеваний проверяются по диагностическому дереву; диагностическая функция группы состоит из перечисленных булевых комбинаций ключевых букв, которые эквивалентны диагностическим критериям для группы заболеваний. Круглые скобки обозначают логическое «и»; знак «+» читается как логическое «или»; число в круглых скобках, следующее за названием заболевания, например NEPHRITIS (11), указывает на номер диагностической программы автомата (D. A.) (Lusted L. B., Computer techniques in medical diagnosis, In: Stacy and Waxman [Eds.], Computers in Biomedical Research, New York, Academic, 1965, v. 1, pp. 319—338.)

система способна сканировать изображения с периодом меньше 0,2 сек на рамку, она характеризуется большим разрешением, чем оптический микроскоп для изображений с размерами 16 и 35 мм. Было также обнаружено, что скорость и сканирующее разрешение ограничены временем цикла и объемом ферритовой памяти, применяемой быстродействующей вычислительной машиной. Устройство FIDAC способно идентифицировать каждую хромосому в соответствии с размером и другими критериями классификации и может подсчитать общее число хромосом за время около 20 сек.

Разработки в области распознавания образов могут быть применимы и к образам, состоящим из слов или групп слов, например признакам и симптомам. Это можно проиллюстрировать некоторыми работами, проведенными несколько лет назад с целью создания распознавательного теста, который можно было бы запрограммировать для вычислительной машины и передать студентам-медикам. Нужно было установить, пользовались ли эти студенты для диагностики хорошо определенным алгоритмом. Предполагалось, что это можно сделать с помощью статистического сравнения результатов работы студентов и машинных программ.

В качестве модели можно попробовать идентифицировать некоторое «слово», играющее роль диагноза, путем анализа кодовых букв, которые вводятся как исходный материал. В реальной диагностической ситуации каждая буква должна представлять собой признак или симптом определенного пациента. С точки зрения моделирования представляется вполне целесообразным использовать названия заболеваний как «образы» и английские буквы в названии заболевания как «симптомы».

Модель распознавания образов включает названия двадцати двух заболеваний, разбитых на пять групп (фиг. 28). «Симптомы» представляются семнадцатью буквами, встречающимися в названиях этих заболеваний, и эти буквы (симптомы) взвешиваются для каждой группы заболеваний. Проблема взвешивания симптомов исключительно трудна; есть надежда, что проверка модели и на студентах и на врачах поможет выяснить, как используются веса человеком, ставящим диагноз.

Фактически мы отбираем один диагноз (слово) из множества, включающего двадцать два диагностических слова; в качестве входных данных выбираем от одной до четырнадцати букв, располагая их в случайном порядке; задаем схему подсчета очков. Кроме того, имеется дерево решений, сходящееся в процессе поиска к группе заболеваний. Выбор группы заболеваний основывается на определенных минимальных критериях в виде булевых комбинаций ключевых букв (например, NPH для группы почечных заболеваний, HP или GS для группы желудочно-кишечных заболеваний).

Вторая часть машинной программы применяется к выходным данным первой части программы, классифицирующей буквенный

(симптомный) вход по принципу его возможной связи с некоторым списком заболеваний. На этой стадии каждое заболевание рассматривается как «шаблон» и отдельно сравнивается с набором входных букв. Во-первых, наличие «неподходящих» букв используется для отбрасывания невозможных диагнозов (например, в диагнозе *Neuritis* нет буквы «п»). Во-вторых, каждой имеющейся в наличии букве приписывается определенное количество очков с учетом частоты, с которой эта буква встречается в данной группе заболеваний; при этом «наиболее важным» буквам приписывается максимальное количество очков. В вычислительную машину, подсчитывающую очки, введены пороги на основе двадцатибалльной системы, устанавливающие для каждого заболевания следующие градации по количеству очков: *A* — от 16 до 20 очков, *B* — от 10 до 15 очков, *C* — от 0 до 9 очков, *RO* — исключение. Окончательно на печать выдается список заболеваний, которые «могут» быть представлены введенными в машину буквами, с подробным подсчетом очков для каждого. Здесь нет непосредственного использования вероятностей, хотя система градации по очкам в какой-то мере отражает вероятностный подход.

Программа для вычислительной машины создана Коффином с соавторами [8] и является приложением работы Стэла с соавторами [104] в области автоматов Тьюринга. Программа состоит приблизительно из 25 000 тьюринговских пятизначных слов, и обработка входной цепочки из пяти букв для постановки диагноза обычно занимает от восьми до двенадцати секунд машинного времени вычислительной машины типа SDS-920.

Ниже для иллюстрации приведен пример с пояснениями:

1. Входная цепочка пробита на одной перфокарте:

$$: 0001 Q \quad \frac{\$ \dots \$}{\text{до } 35} \quad *H \emptyset MNP* \quad \frac{\$ \dots \$}{\text{до } 20}$$

а. Двоеточие (:) обозначает начало цепочки.

б. Четыре числа, следующие за двоеточием, указывают, что должен быть реализован последующий диагностический автомат (D. A.). Эти четыре числа занимают область «автоматического вызова».

в. Символ Q обозначает начало «рабочего поля памяти» процессора цепочки.

г. Буквы (симптомы), например H \emptyset MNP, перфорируются между звездочками.

д. Знаки \$ указывают области памяти для подсчетов и промежуточных результатов.

2. После того как цепочка введена в машину, числа из «области вызова» используются как ключ для поиска диагностического автомата D. A. № 1. D. A. № 1 обнаруживается и вводится в оперативную память. D. A. № 1 — это автомат типа дерева решений, который

определяет, может ли быть поставлен диагноз и к какой категории заболеваний он принадлежит. Если диагноз не может быть поставлен (т. е. не может быть выбрана ни одна из групп заболеваний), то на печать выдается «NO DX» и машина останавливается.

3. На этом этапе выводится на печать категория заболевания. В приведенном выше примере печатается слово RENAL и в область вызова записывается идентифицирующее число первого заболевания в группе почечных. Цепочка будет теперь выглядеть так:

: 0011 Q\$.....\$ *HØMNP*, RENAL. \$\$.....\$.

(Заметим, что комбинация «PN» во входной цепочке оказалась достаточной чтобы идентифицировать группу заболеваний RENAL.) В действительности на печать выдается лишь часть цепочки, заключенная между запятой и знаком периода.

4. В память вызывается D. A. № 11 (диагностический автомат сравнения шаблонов для заболевания NEPHRITIS), и цепочка подвергается специальному тесту для слова NEPHRITIS. Присутствие буквы M в цепочке исключает слово NEPHRITIS. В этом месте цепочка будет выглядеть так:

: 0011 Q\$*HØMNP*, NEPHRITIS RØ. \$ \$

5. Следующим вызывается D. A. № 12 для проверки диагноза NEPHRØSIS, который опять исключается из-за присутствия буквы M в цепочке. В этом месте цепочка будет выглядеть так:

: 0012 Q\$.....\$ *HØMNP*, NEPHRØSISRØ, \$.....\$

6. Следующим вызывается D. A. № 13 для проверки диагноза NEPHRØMA; поскольку ни одна из входных букв непосредственно не исключает диагноз NEPHRØMA, осуществляется подсчет очков.

Буквы *HØMNP* дают 16 очков, и этому диагнозу присваивается градация A. Цепочка теперь выглядит так:

: 0013 Q\$.....\$ *HØMNP*, NEPHRØMA A. \$.....\$

7. Оператор вычислительной машины видит только часть цепочки теста каждый раз, когда осуществляется печать. Последовательная обработка цепочки этого примера на машине имеет такой вид:

Печать	Время печати
HØMNP	Во время ввода цепочки в машину
RENAL	При завершении анализа дерева решений, выделяющего группу заболеваний
NEPHRITIS RØ	При завершении работы диагностического автомата NEPHRITIS
NEPHRØSIS RØ	При завершении работы диагностического автомата NEPHRØSIS
NEPHRØMA A	При завершении работы диагностического автомата NEPHRØMA
RUN CØMPLETE	

У нас нет еще прямых доказательств того, что врач использует в диагностическом процессе образы признаков и симптомов, но это

вовсе не значит, что мы должны прекратить исследование преимуществ, которые можно было бы получить от объединения симптомов и заболеваний в смысловые пучки и группы. Группу симптомов легче запоминать и обрабатывать, чем отдельные симптомы. Мы пытаемся в некотором смысле «переупаковать» медицинские данные в такие пакеты, которые будут более компактны и информативны.

Признаки, симптомы и заболевания могут быть сгруппированы при помощи классификационных или таксонометрических процессов; некоторые исследователи считают и собственно диагностику таксонометрическим процессом. Тем не менее опыт клинициста должен подсказать ему именно то, что продемонстрировала экспериментальная работа — диагноз не является ни просто процедурой классификации, ни просто таксонометрической процедурой.

Давайте вернемся ненадолго назад, чтобы пересмотреть некоторые положения, касающиеся диагностического процесса, который представлен на фиг. 29 в виде блок-схемы.

Познавательные процессы, которые использует врач, когда он продвигается по ступеням этой или сходной блок-схемы, нам не известны. Было предложено много гипотез, и мы рассмотрели некоторые из них в предыдущем изложении. Цель врача — определить, что случилось с данным пациентом, принять решение о лечении, сделать прогноз и осуществить выбранный метод лечения. Мы предполагаем, что первые этапы диагностического процесса могут быть процедурой распознавания образов и что врач, возможно, использует диагностические деревья решений или сети, чтобы сформулировать предположительный диагноз. На этой стадии он может воспользоваться классификационными схемами. Кляйнмунтц [50] обнару-



Фиг. 29.
Блок-схема диагностического процесса [42].

жил, что невролог использует по крайней мере две, а иногда и три классификационные процедуры при решении задачи диагностики в неврологии. Он использует классификацию по этиологическим признакам, по анатомической локализации повреждения и в конце концов по синдрому заболевания. В своем окончательном решении относительно диагноза он может воспользоваться субъективной мерой вероятности.

Итак, подчеркнем следующее положение: классификация не является ни «диагностическим процессом», ни конечным пунктом в диагностике. Правда, это важная часть диагностического процесса, связанная с лечением отдельного пациента. Поэтому нам необходимо теперь обсудить вопросы таксонометрии применительно к проблеме диагностики.

3.7. КЛАССИФИКАЦИЯ ЗАБОЛЕВАНИЙ. МЕТОДЫ ГРУППИРОВАНИЯ И ВЫЧИСЛИТЕЛЬНАЯ ТАКСОНОМЕТРИЯ

Эффективная, точная классификация заболеваний может быть создана только на основе хорошо определенных и общепринятых признаков, симптомов, лабораторных исследований и заболеваний. Классификация также должна быть основана на надежных данных. Читателю, наверное, уже надоело постоянное возвращение к этой теме, но мы повторяем ее в различных вариантах для того, чтобы подчеркнуть следующее важное положение: сложные математические методы и вычислительные машины не снимают ответственность, лежащую на враче, который должен решить, какие именно данные потребуются, собрать надежные данные и осмыслить результаты анализа данных.

Чтобы обсуждать вопросы классификации, мы должны иметь наглядное представление сгруппированных, или объединенных в пучки симптомов и заболеваний. Мы считаем полезным использовать для этой цели диаграммы Венна, но это в большой степени вопрос личного предпочтения. Фиг. 30 взята из работы по прогнозированию ревматической лихорадки Файнштейна [29] — известного пропагандиста использования диаграмм Венна в медицине [27—29]. Диаграммы Венна являются полезным инструментом для изучения заболевания, так как обладают уникальным свойством обнаруживать и выделять пересекающиеся наборы симптомов. Круги на фиг. 30 показывают процентное количество пациентов с острой ревматической лихорадкой, у которых наблюдались проявления артрита, артралгии, хореи, кардита и острого кардита. Области пересечения указывают количество пациентов, имеющих больше чем один симптом. Пересекающиеся круги определяют образы в пространстве симптомов. Граница круга определяется симптоматическими критериями, которые устанавливаются врачом, например критериями, выбран-

Слишком большое количество пересекающихся кругов создает беспорядочную картину, и может оказаться необходимым какой-то другой метод визуального представления результатов группирования. Мы советуем ознакомиться с работой [102], где рассматриваются методы дендрограмм (древовидных структур) и таксонометрических моделей.

В рамках настоящей книги невозможно подробное изложение вопросов классификации и вычислительной таксонометрии. Мы рекомендуем читателю посмотреть интересные статьи по таксонометрии заболеваний, представленные в книге [41]. Полезна также работа [102], посвященная разработке и сравнению различных методов вычислительной таксонометрии. Ниже мы приводим лишь резюме важнейших положений.

Некоторые врачи могут спросить: «Зачем вообще классифицировать?» В работе [101] указано: «Даже лучшие классификации всегда искусственны. В частности, устанавливать классификацию и давать названия группам не очень полезно хотя бы потому, что классификации заболеваний, установленные одним человеком, часто оказываются неадекватными и поэтому переклассифицируются и переименовываются другим исследователем».

Один из доводов в пользу классификации заключается в том, что «образ» заболевания полезен при передаче информации о заболеваниях. Если бы у нас не было образа заболевания общего вида (например, «артрит») или более специального вида (например, «туберкулез легких»), то нам, наверное, пришлось бы выдумать названия или коды для того, чтобы врачи могли их использовать, говоря об этих заболеваниях. Концепция *полезности* является центральной при рассмотрении классификации и выбора естественного образа заболевания [16, 101]. Всегда возникает вопрос: «Классификация с какой целью?» Классификация бессмысленна, если не установлена цель, ради которой она создается. Естественные классификационные группы — это такие группы, которые выделяют наибольшее число свойств и, следовательно, наиболее полезны. Но кто должен определить полезность? Врач.

Врач не может уклониться от этой проблемы выбора. Как указано в работе [16], если мы имеем дело с элементами некоторого множества реальных объектов — людей, заболеваний, признаков или каких-либо других, то любой объект обладает бесконечным числом свойств. «Сколько бы свойств вы ни определили, всегда есть по крайней мере бесконечный счетный набор других свойств, которых вы не можете знать и действительно не знаете. В этом случае людям, занимающимся распознаванием образов и установлением медицинского диагноза, приходится полагаться на свое собственное суждение, свою собственную интуицию и вспомогательные методы статистики, математики и т. п., которые помогают им выбрать небольшой конечный набор свойств множества реальных

объектов и справиться с поставленной задачей. Должен признаться, что весьма трудно определить полезность, но мы не можем обойтись без нее».

В настоящее время мы еще не в состоянии решить вопрос о полезности, и эта проблема заслуживает постоянного обсуждения, потому что полезность определенных признаков и симптомов изменяется вместе с состоянием медицинских знаний. Сейчас лабораторные исследования используются так широко, что некоторые врачи предлагают создать классификацию заболеваний на базе биохимических исследований, хотя это встречает сопротивление большинства клиницистов. Освященная веками клинико-патологическая конференция испытывает все возрастающее влияние изменяющейся ситуации в медицине и концепции полезности. В предисловии к «Трудам клинико-патологических конференций в Массачусетском главном госпитале» Б. Каствлман [6] следующими словами охарактеризовал сложившееся положение: «Критики называют клинико-патологическую конференцию игрой «угадай-ка», но это несправедливо, если разбираемый случай выбран разумно. Правильно выбранный случай приводит к обсуждению дифференциального диагноза и клиническим замечаниям, предлагающим правильный диагноз. Этот выбор — не простое дело, потому что нужно достигнуть золотой середины между очень сложной проблемой, которая не может быть диагностирована на клинической основе, и очевидным случаем, который не вызывает сомнений. Поскольку диагностические методы и искусство диагностов все более совершенствуются, все реже встречаются подходящие проблемы для обсуждения дифференциальной диагностики. Например, все случаи подострого бактериального эндокардита ранее считались интересными для обсуждения на клинико-патологической конференции, но сегодня заслуживают представления лишь случаи с необычными осложнениями».

Постановка проблемы полезности означает, что классификационная система всегда является в некотором смысле незавершенной, но из этого вовсе не следует, что каждый отдельный пациент не важен или что изучение индивидуального случая устарело. Прежде чем мы будем классифицировать индивидуального пациента, нужно определить классы и подклассы заболеваний. Следует помнить, однако, что сама классификационная система может быть определена как результат наблюдения большого количества отдельных пациентов. Из-за взаимодействия между системой классификации и классификацией индивидуального пациента и из-за зависимости этого взаимодействия от состояния наших знаний мы должны непрерывно пересматривать нашу классификацию, основываясь на данных, собранных об отдельных пациентах за любой период времени [42].

До сих пор в наших обсуждениях мы не рассматривали фактор времени. И не потому, что мы предпочли игнорировать динамиче-

ский аспект заболеваний, т. е. изменение симптомов со временем, а потому, что мы можем лишь очень немного предложить в смысле возможных процедур логического анализа. Нам неизвестно, как обращаться с этими проблемами. Мы позволим себе обратить внимание на два обстоятельства.

Во-первых, при вычислении или оценке априорной условной вероятности $P(S|D)$ мы допускаем, что статистический характер выборки данных не изменяется или, иными словами, что истинные частоты симптомов не изменяются в течение рассматриваемого периода времени. Следовательно, данные должны уточняться через промежутки времени, короткие по сравнению с периодом изменений в симптоматическом образе заболевания. Для отдельного пациента с таким заболеванием, как стеноз легочной артерии, симптомы которого изменяются медленно, такое предположение может быть обоснованным; однако оно может не быть таковым для пациента с быстро развивающейся недостаточностью кровообращения.

Во-вторых, диаграмма Венна, изображенная на фиг 30, представляет собой симптомный образ в определенное время. Размеры и относительные положения кругов будут изменяться с течением времени. Каким образом можно учесть эти динамические соотношения? Возможно, в будущем удастся демонстрировать их на телевизионном экране, управляемом машинной программой. Но прежде чем мы сможем воспользоваться этой технологией, нам придется проделать большую предварительную работу по изучению комплексов симптомов.

В исследовании [102] по вычислительной таксонометрии обсуждены типы весовых коэффициентов, использованные при вычислении сходства; сравниваются методы группирования по одной связи, средней связи, расстоянию от контрольного узла и при помощи факторного анализа. В работе [67] показано, что методы вычислительной таксонометрии, предложенные в работе [102], были полезными при группировании 99 пациентов с заболеваниями сердца на основе 129 признаков, симптомов и лабораторных исследований. Что же касается более подробного ознакомления с факторным анализом, то читателю следует обратиться к работам [38] и [102].

ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЯ И ОШИБКИ НАБЛЮДАТЕЛЯ В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ

И что ты смотришь на сучок в глазе брата твоего, а бревно в твоём глазе не чувствуешь?... Вынь прежде бревно из твоего глаза, и тогда увидишь, как вынуть сучок из глаза брата твоего.

Евангелие от Матфея, 7 : 3—5

4.1. ВВЕДЕНИЕ

Это изречение из Библии было использовано Л. Г. Гарландом [35] как введение к обсуждению точности диагностических процедур. Характерна та напряженная атмосфера, которая создается при обсуждении ошибок наблюдателя в медицине. Нетрудно привести некоторые примеры. Так, опытные интерпретаторы рентгенограмм грудной клетки склонны пропустить (рассматривая как отрицательные) около 30% снимков, положительных в смысле рентгенологических признаков заболевания, и интерпретировать как положительные около 2% отрицательных. Далее, при оценке пар последовательных рентгенограмм с целью определения изменений в состоянии заболевания один опытный врач склонен не соглашаться с другим приблизительно в одном из трех случаев, а при пересмотре материала склонен не соглашаться с самим собой в одном из пяти случаев. Как указал Гарланд, «аналогичные ошибки встречаются во многих областях клинической практики. Действительно, если можно было бы обследовать все области медицины, то такое явление оказалось бы достаточно общим». Это горькая пилюля, которую приходится глотать врачу, и он часто старается избежать этого. Его реакция обычно такова: «Ну что же, это не касается моей каждодневной работы».

Если мы считаем, что медицинский диагноз — это вопрос вероятности, то какой же подход должны мы выбрать при представлении фактов студентам-медикам и стажирующимся врачам? Возможно, нашей целью должна быть замена безопасности мышления, которая обеспечивается продвижением по хорошо определенному, проторенному пути, новым видом безопасности, основанной на признании неясности, неопределенности и открытого выбора [43, 44].

В предыдущих главах мы показали, что диагноз является процессом решения при наличии неопределенности. Мы обсудили диагностический процесс как дерево решений или сеть с узлами, которые представляют собой проверочные пункты гипотезы. Дерево или сеть являются моделью последовательной стратегии проверки гипо-

тезы врачом, и он, возможно, способен улучшить свою деятельность в многогипотезной ситуации (дифференциальная диагностика), сокращая количество принимаемых к рассмотрению диагнозов настолько быстро, насколько позволяют данные. Чтобы выполнить такое сокращение гипотез, врачу нужно знать, какие тесты или комбинации тестов будут наиболее полезны на каждом уровне диагностического дерева или сети. На этот вопрос нельзя дать самый общий ответ, и он нуждается в дальнейших исследованиях.

Мы привели несколько примеров диагностических критериев, выбранных при помощи байесовских методов, и указали на необходимость дальнейших исследований, которые таксонометрически объединят признаки, симптомы и заболевания. В этой исследовательской области могут оказаться полезными вычислительные машины. В качестве диагностических критериев нужно рассматривать группы признаков и симптомов, потому что лишь очень немногие заболевания могут быть диагностированы только по одному признаку или тесту. Например, невозможно выбрать какой-то один признак, который отделил бы доброкачественную язву желудка от злокачественной язвы желудка. Для этого врач должен использовать группу признаков.

В этой главе мы обсудим некоторые следствия, вытекающие из выбора диагностических критериев врачом и из его отношения к диагнозу или предубеждения (оптимистического или пессимистического). Изложение начинается с обзора некоторых исследований по определению ошибок наблюдателя в диагностической рентгенологии и цитологии. Эти исследования приводят к рассмотрению вопросов, связанных с проверкой гипотез и правилами решения, и к сопоставлению ошибок наблюдателя при диагностике с результатами недавних психофизических исследований по обнаружению и распознаванию сигнала человеком-наблюдателем.

4.2. ОШИБКИ НАБЛЮДАТЕЛЯ ПРИ РЕНТГЕНОЛОГИЧЕСКОЙ ДИАГНОСТИКЕ

В начале второй мировой войны были проведены исследования для оценки относительной эффективности различных рентгенографических методов для массового обследования с целью распознавания туберкулеза легких. Было обнаружено, что размер носителя информации — фотопленка шириной 35 мм или размерами 14"×17" — был менее важен, чем уровень вариаций наблюдателя [135]. Пять опытных врачей, интерпретировавших серии снимков на пленке с разными размерами, пропустили приблизительно 25% положительных снимков в каждой серии; когда те же снимки рассматривались повторно спустя три месяца, один и тот же интерпретатор был склонен изменить свое мнение приблизительно в одном из каждых пяти положительных случаев. Диагностические

критерии для оценки «положительный» не были определены. По-видимому, ожидалось, что интерпретаторы будут анализировать снимки «так, как они это обычно делают» для рентгенографического обнаружения заболевания легких, предположительно активного. Результаты исследования удивили многих врачей, и две группы опытных рентгенологов решили продолжить изучение этого вопроса [35]. Они пришли к выводу, что упомянутое явление действительно имеет место. Впоследствии были опубликованы работы [35, 135, 136], в которых был установлен тот факт, что двойная интерпретация пленок значительно уменьшила число пропущенных положительных снимков. Были проведены бактериологические и (или) клинические исследования, чтобы вынести окончательное заключение о том, является ли действительно «положительным» пациент, названный «рентгенологически положительным».

Анализ серии интерпретаций снимков грудной клетки в каждом из этих исследований показал, что возможны четыре альтернативы: интерпретация «отрицательный» (нет заболевания) для «истинно отрицательного» случая, или нормы; интерпретация «положительный» (имеет место заболевание) для «истинно положительного» случая, или больного пациента; интерпретация «отрицательный» для «истинно положительного» пациента и интерпретация «положительный» для «истинно отрицательного» пациента. В медицинской литературе было достигнуто соглашение обозначать отрицательную интерпретацию для «истинно положительного» пациента как «ложно отрицательную». Ложно отрицательную интерпретацию при рентгенографических исследованиях часто называют «пропуском», и это понятие эквивалентно ошибке типа 1 (ошибке первого рода, α -ошибке) в теории проверки гипотез. Утверждение, объявляющее больного человека здоровым, представляет собой ошибку ложно отрицательного типа.

Положительную интерпретацию для «истинно отрицательного» пациента часто называют «гипердиагностикой», и это понятие соответствует ошибке типа 2 (ошибке второго рода, β -ошибке) в теории проверки гипотез. Отнесение здорового человека к группе больных означает ошибку ложно положительного типа. Все эти термины используются в литературе, и иногда в связи с ними возникает путаница.

Другая трудность заключается в том, что ошибка может быть вычислена по-разному. Представление ошибок в виде процентного числа ложно положительных случаев и процентного числа ложно отрицательных случаев приводит к преувеличению важности ложно положительных ошибок, когда заболевание имеет относительно небольшое распространение в общей популяции. Этот вопрос будет обсужден более подробно в дальнейшем. Для рассматриваемого примера мы используем метод оценки ошибок, предложенный Гарландом [35], в следующем виде: если серия из n рентгенограмм со-

держит десять положительных и девять отрицательных снимков и рентгенолог пропускает три положительных снимка (три ложно отрицательные ошибки) и гипердиагностирует два отрицательных снимка (две ложно положительные ошибки), то можно считать, что он допустил ошибку в 5% случаев. С другой стороны, поскольку эта серия из ста рентгенограмм исследуется с целью обнаружения пациентов, страдающих заболеванием, рентгенолог, не распознавший три из десяти положительных снимков, допускает 30% ложно отрицательных ошибок. Мы будем использовать этот последний критерий для оценки ложно отрицательных ошибок. Аналогично число ложно положительных ошибок равно двум из девяти отрицательных снимков, т. е. их доля составляет приблизительно 2%. Суммарная доля ошибок, подсчитанная таким образом, равна 32%.

Как ложно положительные, так и ложно отрицательные интерпретации являются ошибками наблюдателя и представляют собой некоторую «цену потерь» для системы диагностики в смысле стоимости дополнительных исследований, волнений и неудобств для пациента. Поэтому важно найти такое соотношение ложно положительных и ложно отрицательных интерпретаций, которое обеспечило бы минимальную «цену» для системы. Чтобы определить эту «цену» для любого диагностического теста или процедуры, необходимо найти для них соотношение между ложно положительными и ложно отрицательными ошибками. Одним из удобных приемов является вычерчивание кривых разделения или рабочих характеристических кривых для каждой диагностической процедуры. Ниже мы опишем, как было установлено соотношение между ложно отрицательными и ложно положительными интерпретациями при разделении рентгеновских снимков грудной клетки.

Авторы работ [134] и [136] при помощи остроумной серии экспериментов провели сравнение независимых результатов нескольких последовательных интерпретаций, чтобы установить количество ложно отрицательных и ложно положительных ошибок при разделении рентгенограмм грудной клетки с точки зрения наличия или отсутствия туберкулеза. В одном из этих исследований [136] каждому снимку, представляющему собой фотофлюорограмму грудной клетки на 70-миллиметровой пленке (обследована большая популяция студентов колледжей, всего 15 000 человек), было дано десять независимых интерпретаций. Каждую фотофлюорограмму просматривали восемь разных рентгенологов, шесть из которых выполнили по одному просмотру и двое — по два независимых просмотра (всего десять просмотров каждого снимка). Во втором исследовании [135] снимки просматривали пятьдесят человек, причем эта группа интерпретаторов состояла наполовину из рентгенологов и наполовину из специалистов в области легочных заболеваний; каждый провел в среднем три тысячи просмотров по случайной

схеме, обеспечивающей десять интерпретаций каждого снимка. Интерпретаторов просили указать как «положительные» те фотофлюорограммы с потенциально значимыми затемнениями в легких, для которых они хотели бы получить вторичный снимок. Каждый студент, чья фотофлюорограмма была интерпретирована как «положительная» в двух или более из десяти просмотров, был вновь вызван для получения стандартной (14"×17") рентгенограммы. Если только один просмотр дал «положительный» ответ, то фотофлюорограмма пересматривалась группой из трех интерпретаторов. Исходя из их решения, студента либо вновь вызывали для съемки стандартной рентгенограммы, либо не вызывали. Все рентгенограммы (14"×17") были интерпретированы независимо тремя экспертами. Если по крайней мере два из трех просмотров были «положительными», то предпринималось дальнейшее клиническое обследование. Если же только один из просмотров был «положительным», то эксперты при вторичном просмотре решали, нужно ли вызвать студента для клинического обследования. В результате проведенного исследования 249 студентов были идентифицированы как имеющие затемнения в легких и требующие дальнейшего клинического обследования. Эти студенты считались «положительными по рентгену». После дополнительных обследований, таких, как туберкулиновые пробы и бактериологические исследования, только восемь студентов (или 3,2% «положительных по рентгену» случаев) были признаны страдающими активной формой туберкулеза. Нужно заметить, что относительно четверых из этих студентов было известно, что они раньше имели активную форму туберкулеза, так что обследование выявило четыре ранее неизвестных случая (таким образом, их частота составляет приблизительно 0,26 на каждую тысячу).

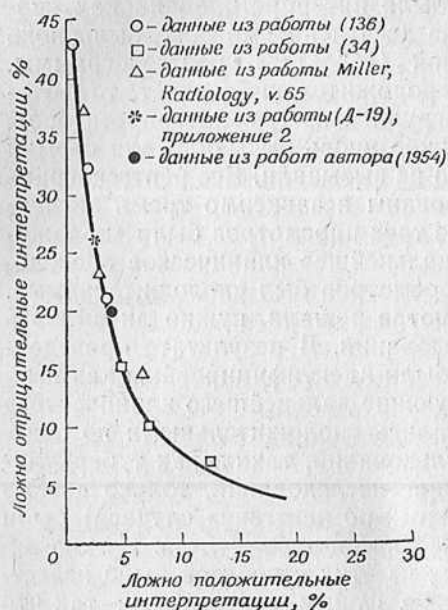
Ниже приведены результаты интерпретации фотофлюорограмм при проведении этого исследования.

Общее количество 70-миллиметровых снимков	14 867
Количество снимков, названных «положительными» одним или большим числом интерпретаторов	1 307
Количество снимков, признанных «положительными по рентгену»	249
Средний процент ложно отрицательных интерпретаций для всех интерпретаторов	39
Средний процент ложно положительных интерпретаций для всех интерпретаторов	1,2

Для пятидесяти интерпретаторов 39% ложно отрицательных ответов в действительности означают 736 ошибок типа 1, а 1,2% ложно положительных ответов означают 1216 ошибок типа 2. Таким образом, низкий процент ложно положительных интерпретаций может быть довольно обманчивым с точки зрения действительного

числа встречающихся случаев, когда рассматриваемое заболевание имеет относительно низкую частоту в общей популяции, подвергающейся проверке.

В приведенном выше исследовании [136, 135] оценивалось значение



Фиг. 31.

Рабочая характеристическая кривая для интерпретации фотофлюорограмм в случае туберкулеза легких. Ясно видна обратная пропорциональность между процентом ложно отрицательных и процентом ложно положительных интерпретаций. Врач может выбирать рабочую точку на кривой в зависимости от того, какую он придает относительную «цену» ложно отрицательным случаям по отношению к ложно положительным (т. е. насколько он хочет быть оптимистичным или пессимистичным в своей оценке) [66].

стическим отношением, а затем со значительно более пессимистическим отношением. Пессимистическое отношение отражает большую осторожность со стороны рентгенолога, когда он предпочитает следующую линию поведения: «если сомневаешься, называй снимок положительным», или «лучше обезопасить себя, чем потом сожалеть». Такое поведение увеличит число обнаруженных истинно поло-

ного двойного просмотра (т. е. независимые суждения двух различных интерпретаторов при отборе «положительных» случаев). Существуют два способа объединения этих суждений. Случай можно называть «положительным» либо только тогда, когда оба интерпретатора с этим согласны, либо только тогда, когда по крайней мере один интерпретатор считает случай «положительным». Первый метод увеличивает число ложно отрицательных ошибок, но уменьшает число ложно положительных по сравнению с экспериментами, в которых просмотр осуществляется одним интерпретатором, тогда как второй метод уменьшает число ложно отрицательных ошибок, но увеличивает число ложно положительных. Данные, полученные таким способом, представлены графически на фиг. 31.

На этом графике отмечены также точки, полученные независимо Гарландом [34], который попытался оценить влияние отношения интерпретатора к поставленной задаче; для этого он просил рентгенологов интерпретировать группу снимков дважды: сначала с оптимистическим отношением, а затем со

жительных случаев, но оно повлечет за собой и увеличение процента ложно положительных ошибок.

На фиг. 31 мы указали также результаты независимых исследований по интерпретации фотофлюорограмм грудной клетки, проведенных Стрэдлингом и Джонстоном в Великобритании, а также отметили точку, характеризующую автора этой книги как интерпретатора (20% ложно отрицательных, 4% ложно положительных ошибок) по материалам одного из указанных ранее исследований [136]. Важно отметить, что все точки приблизительно лежат на одной кривой, которая может быть использована как рабочая кривая при интерпретации фотофлюорограмм грудной клетки и которая демонстрирует интересную обратную зависимость между ложно отрицательными и ложно положительными интерпретациями.

Большинство диагностических тестов не было изучено так полно, как интерпретация рентгенограмм грудной клетки. Может возникнуть вопрос о методе, используемом для построения рабочей характеристической кривой, поскольку мы не определили точно диагностические критерии. Конечно, и выборки популяций были различны в каждом исследовании. Мы предлагаем вниманию читателя еще один пример, который достаточно хорошо документирован. Это описание попытки разработать автоматическое устройство для цитологического выделения злокачественных опухолей.

4.3. ОШИБКИ НАБЛЮДАТЕЛЯ В ЦИТОЛОГИИ РАКА

Это исследование было предпринято в надежде освободить квалифицированного цитодиагноста от рутинного труда по разделению мазков, окрашенных по методу Папаниколау. Целью работы было создание сканирующе-вычислительного устройства для автоматического разделения мазков эксфолиативных клеток женских половых органов в программе обнаружения рака [39, 118]. Кроме собственно конструирования электронно-сканирующего микроскопа и вычислительного устройства, необходимо было произвести количественную оценку диагностических критериев, которые использует цитолог, чтобы определить, является ли взятый у пациента микропрепарат, содержащий клетки, положительным или отрицательным в отношении рака. Упомянутое автоматическое оборудование нельзя было сконструировать без этих критериев, поскольку его эффективность зависела от того, насколько точно удастся смоделировать процесс разделения, осуществляемый цитологом.

Была осуществлена широкая программа измерений для установления критериев, используемых цитологом, и выражения этих критериев в количественном виде. Пробы, взятые у пятисот женщин, были подготовлены на микроскопических слайдах. Слайды были разделены на две группы: (1) случайная выборка из группы нормальных пациентов без очевидных признаков злокачественности

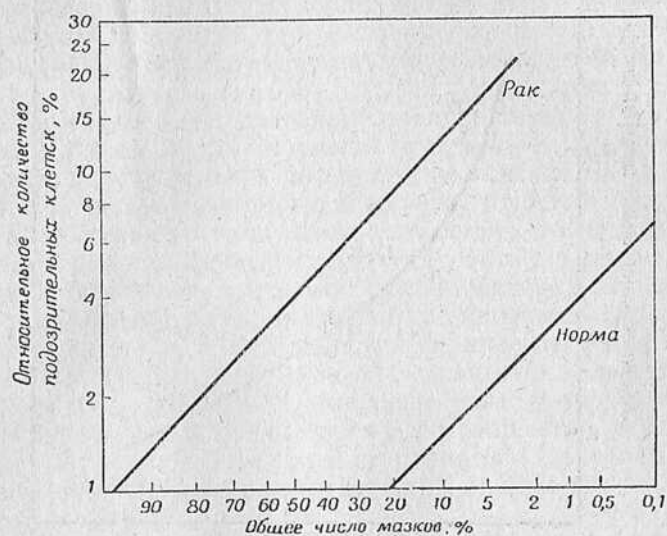
и (2) мазки, взятые у пациентов с диагнозом «рак шейки матки», подтвержденным биопсией или оперативным вмешательством. В начале исследования было обнаружено, что цитолог использовал в качестве первичного критерия процентное содержание клеток с большими плотными ядрами. На основе этих результатов было разработано несколько простых критериев для проверки подозрительных клеток; эти критерии выражались через измеряемые величины, а именно размеры клеток и плотности. Затем была определена эффективность разделения по каждому критерию или комбинации критериев [40]. Ниже представлен один из исследованных наборов критериев:

1. Ядро должно иметь размер меньше 20 мк по большому диаметру.
2. Поглощение света ядром должно превышать 50%.
3. Контраст между ядром и цитоплазмой должен превышать некоторую заданную величину.
4. Коэффициент поглощения цитоплазмы не должен превышать 0,5.
5. Резкие изменения в форме изображения являются основанием для отказа от принятия решения.

Проведенный анализ показал, что независимо от того, как были выбраны критерии, всегда встречалось некоторое число проб, взятых у подтвержденных раковых пациентов, которые содержали очень немного клеток, соответствующих выбранным критериям (ложно отрицательный результат), тогда как пробы, взятые у некоторых здоровых пациентов, имели большее количество таких клеток (ложно положительный результат). Таким образом, эти две популяции — раковые больные и здоровые пациенты — пересекаются по отношению к выбранным критериям или любой комбинации критериев, использованных для обнаружения рака. Из этого следует, что независимо от того, какая именно комбинация критериев выбрана, некоторые пациенты, у которых нет рака, будут классифицированы как имеющие рак (ложно положительные), а некоторые из пациентов, больных раком, будут классифицированы как здоровые (ложно отрицательные). Более строгие критерии, возможно, помогли бы отобрать больше истинно положительных раковых пациентов, но только ценой увеличения количества ложно положительных случаев. В работе [39] приведен пример на основе реальных данных описанного выше исследования, иллюстрирующий это положение.

Данные для двух популяций — группы здоровых людей (истинно отрицательные случаи) и группы подтвержденных раковых пациентов (истинно положительные случаи) — были представлены по отдельности в графическом виде, чтобы показать распределение процента проб, содержащих определенное количество подозрительных клеток. Полученные кривые распределения оказались сильно

асимметричными, но их удалось достаточно хорошо аппроксимировать при помощи функции логарифмического нормального распределения. На фиг. 32 показаны две интегральные кривые распределения соответственно для нормальной и раковой популяций в логарифмическом вероятностном масштабе, аппроксимирующие реальные данные. Для этих кривых использованы следующие критерии отбора клеток: 1) диаметр ядра $d \geq 10$ мк и 2) оптический коэффициент поглощения $\epsilon \geq 0,5$. Как показывают кривые, если мы



Фиг. 32.

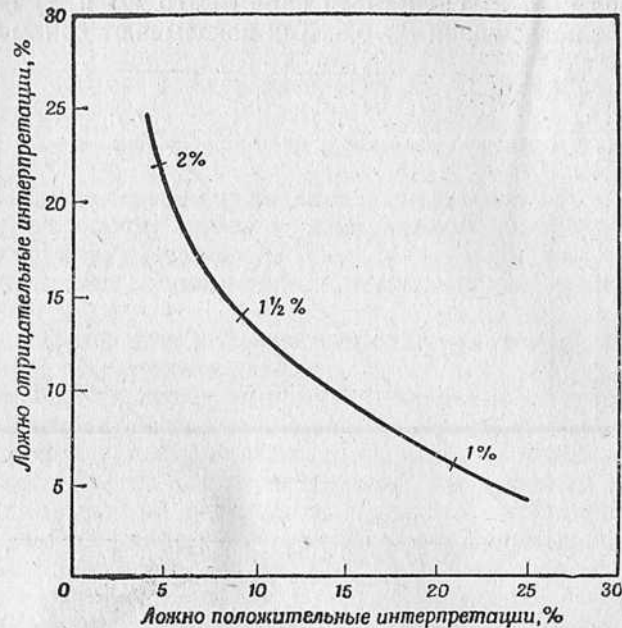
Интегральные кривые распределения количества мазков, окрашенных по методу Папаниколау, в которых процентное содержание клеток с параметрами $d \geq 10$ мк, $\epsilon \geq 0,5$ превышает заданное число [39].

потребуем, чтобы по крайней мере 2% клеток в пробе превысили минимальные значения $d=10$ мк и $\epsilon=0,5$, то будут пропущены 22% раковых пациентов (ложно отрицательные) и 5% здоровых людей будут отнесены к больным (ложно положительные).

Если же вместо этого мы потребуем, чтобы только 1% клеток превышал критерии, то мы пропустим лишь 6% раковых пациентов (ложно отрицательные), но 20% здоровых будут классифицированы как больные раком (ложно положительные). Такая обратная зависимость между количествами ложно отрицательных случаев и ложно положительных случаев иллюстрируется графиком на фиг. 33. На этой кривой отмечены три различных значения процентного содержания подозрительных клеток. В зависимости от значения, выбираемого врачом для процентного содержания клеток, которые должны превышать минимальные критерии, он может

управлять программой разделения, сдвигая ее в любую точку на кривой, и получать указанные процентные содержания ложно отрицательных и ложно положительных случаев.

На основе описанного выше метода авторы работы [40] построили серию рабочих характеристических кривых (мы обычно называем их кривыми разделения), сходных с кривой, приведенной на



Фиг. 33.

Рабочая характеристическая кривая разделения раковых больных по мазкам, окрашенным по методу Папаниколау. Данные взяты из кривых распределения, представленных на фиг. 32. Ясно видна обратная пропорциональность между процентом ложно отрицательных и процентом ложно положительных интерпретаций. Кривая имеет форму, аналогичную форме кривой на фиг. 31 для интерпретации рентгеновских снимков грудной клетки [39].

фиг. 33. Каждая кривая представляет определенную группу критериев, и если какая-либо кривая располагается целиком ниже и левее любой другой кривой, то ее критерии дают более высокие рабочие характеристики в смысле меньшего числа ложно отрицательных и ложно положительных случаев. После того как найдена кривая, представляющая нижнюю левую границу указанного семейства кривых, больше не существует простой и строгой основы для выбора «лучшей» из каких-либо двух точек на этой кривой, поскольку каждая из точек лучше другой только в отношении вероятности ошибок определенного типа, т. е. либо ложно отрицательных,

либо ложно положительных случаев. Таким образом, перед врачом одновременно стоят две задачи: выбор диагностических критериев и выбор «отношения» к разделению, или рабочей точки на характеристической кривой разделения.

В работе [39] было найдено, что в определенных пределах кривые разделения для различных критериев сходны по виду при условии, что пробы собирались и препарировались аналогичными методами. Таким образом, данный конкретный тест при использовании определенных критериев будет всегда давать некоторое количество ложно положительных и ложно отрицательных случаев. Граничная линия между отрицательными и положительными популяциями может быть в некотором смысле произвольно задана врачом. Некоторые врачи, не желающие пропустить ни одного пациента, больного раком, захотят провести эту линию по 1%-ному уровню подозрительных клеток (более пессимистическая точка зрения), тогда как другие врачи, которые озабочены большим количеством ложно положительных пациентов, требующих последующего обследования и биопсии тканей, могут установить граничную линию по 2% или даже и более высокому уровню подозрительных клеток. Решение врача о положении граничной линии фактически является суждением о «цене», и в медицине сейчас имеется лишь немного объективных критериев, которыми врач мог бы руководствоваться при решении этой проблемы. К вопросам, связанным с ожидаемой «ценой» решения, мы будем все время возвращаться в остальной части книги. Врач все время будет сталкиваться с проблемой оценки его «личных» критериев в отношении ожидаемой «цены» при конкретных ситуациях.

По-видимому, почти все диагностические проблемы могут быть изучены так же тщательно, как в приведенных примерах исследования рентгеновских снимков грудной клетки для обнаружения туберкулеза легких и цитологического разделения здоровых людей и больных раком. Из-за пересечения истинно отрицательных и истинно положительных популяций возникает проблема неопределенности, и очень трудно найти тест или группу тестов, которые полностью разделяли бы эти две популяции. Теперь уже должно быть более очевидно, почему необходимо искать группы симптомов и тестов, которые как раз и обеспечат наиболее эффективное разделение двух или большего числа популяций. Как мы указывали ранее, полезным инструментом, помогающим врачу в поиске эффективных групп симптомов, является вычислительная машина. Работа [88] иллюстрирует это положение. Ее авторы использовали машинную обработку данных (система CYDAC) для анализа типа клеток. Выбрав двадцать один независимый триплет из имевшихся параметров (измерения в области ядра, цитоплазмы и определение ядерно-цитоплазменного контраста), они смогли достигнуть полного разделения в ограниченной популяции, составленной из эозино-

филов, лимфоцитов, моноцитов и нейтрофилов. Это означает, что исследователям удалось найти такие критерии, при помощи которых популяции клеток разделяются «почти полностью».

Мы хотим перейти теперь к рассмотрению вопросов, связанных с проверкой гипотез, процессами принятия решения и обнаружением сигнала, чтобы объединить некоторые методы, которые, по нашему мнению, будут полезны при разработке диагностических критериев и при изучении процесса принятия решения в медицине.

4.4. ПРОВЕРКА ГИПОТЕЗ, ОБНАРУЖЕНИЕ СИГНАЛА И ПРОЦЕССЫ ПРИНЯТИЯ РЕШЕНИЯ

Статистическая проверка гипотез используется для осуществления выбора между двумя альтернативными возможностями, и врачи часто полагают, что проверка гипотез применяется скорее для выделения заболеваний или разработки новых лабораторных тестов, чем для поиска диагностических критериев. В самом простом случае проблема, с которой мы сталкиваемся в процессе проверки гипотез, состоит в следующем: дан результат некоторого теста и две гипотезы, каждая из которых определяет функцию распределения результата этого теста; нужно сделать наилучший выбор между этими двумя гипотезами. Применительно к медицинскому разделению или выбору диагностического теста это утверждение может быть перефразировано следующим образом: дана произвольная совокупность пациентов и два состояния — норма или патология, — каждое из которых определяет функцию распределения результатов теста; нужно сделать наилучший выбор между этими двумя состояниями. Термин «норма» здесь используется в смысле «непатологическое состояние», поскольку это последнее определяется специальным диагностическим тестом или набором принятых для этой цели диагностических критериев.

Надежность теста или набора критериев, используемых для отделения здоровых людей от больных, часто определяют при помощи таких характеристик, как чувствительность и специфичность. Чувствительность — это способность теста дать положительный ответ, когда исследуемый пациент действительно болен, или «истинно положителен» по отношению к рассматриваемому заболеванию;

$$\text{Чувствительность, \%} = \frac{\text{количество пациентов, страдающих данным заболеванием, с положительным результатом теста}}{\text{количество всех пациентов, страдающих данным заболеванием и подвергшихся тесту}} \times 100.$$

Специфичность — это способность теста дать отрицательный ответ, когда исследуемый пациент не страдает заболеванием, или «истинно отрицателен» по отношению к рассматриваемому заболеванию:

$$\text{Специфичность, \%} = \frac{\text{количество пациентов, не страдающих данным заболеванием, с отрицательным результатом теста}}{\text{количество всех пациентов, не страдающих данным заболеванием и подвергшихся тесту}} \times 100.$$

Как указывает Вечю [121], несмотря на то, что какой-либо диагностический тест при предварительной оценке известных групп людей, страдающих и не страдающих данным заболеванием, может иметь высокую эффективность, нужно быть осторожным, применяя эти результаты к произвольным группам, так как при относительно малом распространении данного заболевания в общей популяции увеличивается число ложно положительных ошибок. Автор предлагает использовать понятие предположительного значения теста, которое определяется следующим образом:

$$\begin{aligned} &\text{Предположительное значение, положительный результат теста} = \\ &= \frac{\text{количество пациентов, страдающих данным заболеванием, с положительным результатом теста}}{\text{общее количество пациентов с положительным результатом теста}} \times 100. \end{aligned}$$

$$\begin{aligned} &\text{Предположительное значение, отрицательный результат теста} = \\ &= \frac{\text{количество пациентов, не страдающих данным заболеванием, с отрицательным результатом теста}}{\text{общее количество пациентов с отрицательным результатом теста}} \times 100. \end{aligned}$$

Вечю показал полезность этой формулы при обнаружении диабета в общей популяции, а Рассел [96] расширил концепцию Вечю о предположительном значении и ввел индекс надежности клинического диагноза. Оптимальный индекс определяется отношением правдоподобия — отношением истинно положительных случаев к ложно положительным случаям в процентах. Этот метод будет описан в разд. 4.6. Несостоятельность теста в смысле чувствительности приводит к ложно отрицательным результатам, а в смысле специфичности — к ложно положительным результатам. Изложенное сведено в таблицу на фиг. 34.

Рабочая характеристическая кривая (фиг. 31 и 33) полностью описывает некоторую серию таких матриц, поскольку дополнения к двум данным относительным величинам (ложно отрицательный результат и ложно положительный результат) в каждой точке кривой — это относительные величины в двух других клетках матрицы.

$$\text{Чувствительность, \%} = \frac{\text{количество истинно положительных пациентов}}{\text{общее количество пациентов, страдающих данным заболеванием и подвергшихся тесту}} \times 100.$$

Таким образом, чувствительность — это вероятность правильного диагноза для положительного случая (для больного, страдающего данным заболеванием).

		Результаты теста		
		положительные	отрицательные	
Категория заболевания	заболевание	Истинно положительные случаи	Ложно отрицательные случаи (ошибки 1-го типа, α -ошибки)	Общее количество больных лиц, подвергшихся тесту
	норма (отрицание заболевания)	Ложно положительные случаи (ошибки 2-го типа, β -ошибки)	Истинно отрицательные случаи	Общее количество здоровых лиц, подвергшихся тесту
		Общее число положительных случаев	Общее число отрицательных случаев	

Фиг. 34.

Матрица четырех различных исходов при вынесении решения относительно двух возможных заболеваний. (Предполагается, что одно из заболеваний соответствует «норме».)

$$\text{Специфичность, \%} = \frac{\text{количество истинно отрицательных пациентов}}{\text{общее количество пациентов, не страдающих данным заболеванием и подвергшихся тесту}} \times 100.$$

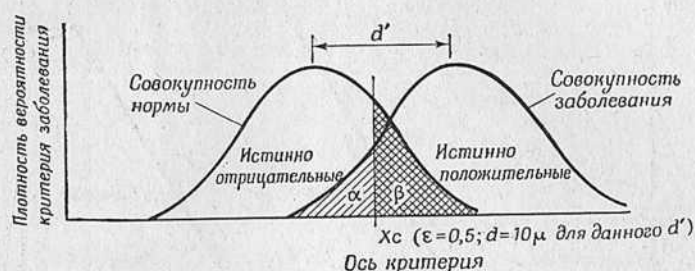
Таким образом, специфичность — это вероятность правильного диагноза для отрицательного случая (для больного, не страдающего данным заболеванием).

$$100\text{-чувствительность, \%} = \alpha, \% = \frac{\text{количество ложно отрицательных пациентов}}{\text{общее количество пациентов, страдающих данным заболеванием и подвергшихся тесту}} \times 100;$$

$$100\text{-специфичность, \%} = \beta, \% = \frac{\text{количество ложно положительных пациентов}}{\text{общее количество пациентов, не страдающих данным заболеванием и подвергшихся тесту}} \times 100.$$

Связи между этими четырьмя состояниями — истинно отрицательным, истинно положительным, ложно отрицательным и ложно положительным — показаны на фиг. 35, где введены две гипотетические плотности распределения. Теперь предположим, что мы выбрали в качестве критериев для отбора злокачественных клеток те, которые были использованы в последнем примере (см. разд. 4.3): диаметр клетки $d \geq 10 \text{ мк}$ и коэффициент поглощения $\epsilon \geq 0,5$.

На фиг. 35 представлены графики плотности вероятности критерия заболевания (ось Y) для двух совокупностей, при этом выбор критерия определяет относительное положение двух показанных кривых. Ось X называется осью критерия, или осью наблюдения, поскольку врач (которого можно называть «наблюдателем») выбирает на ней рабочую точку X_c , которой он намерен пользоваться. Так, при одном и том же наборе критериев он может уменьшать вероятность ложно положительных случаев, смещая эту точку



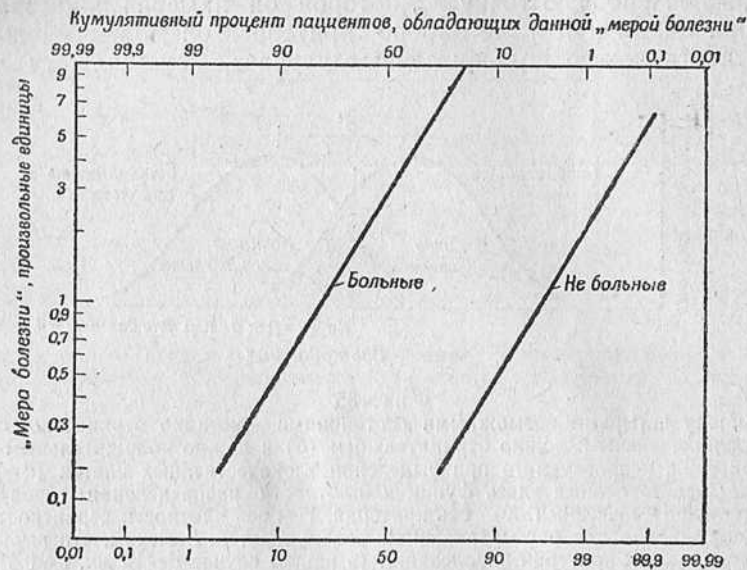
Фиг. 35.

Связь между четырьмя возможными состояниями — истинно отрицательным, истинно положительным, ложно отрицательным (α) и ложно положительным (β) — для критерия, используемого при выделении злокачественных клеток ($d=10$ мк, $\epsilon=0,5$). Представленные здесь функции плотности распределения вероятностей являются гипотетическими. X — ось критерия, Y — ось плотности вероятности критерия заболевания для рассматриваемых совокупностей. Заметим, что выбор конкретного значения критерия заболевания (в нашем случае $d=10$ мк, $\epsilon=0,5$) определяет относительное расположение кривых, или параметр d' . На оси критерия врач выбирает и использует некоторую рабочую точку. Видно, что перемещение прямой X_c по оси X изменяет относительное количество ложно отрицательных и ложно положительных случаев. Для того чтобы увеличить количество ложно отрицательных случаев, прямую X_c нужно сдвинуть вправо, а для увеличения количества ложно положительных случаев — влево.

вправо. Вероятность ложно положительных случаев определяется площадью, ограниченной кривой плотности распределения для совокупности здоровых лиц справа от вертикальной прямой, проходящей через точку X_c . При перемещении этой прямой вправо увеличивается вероятность ложно отрицательных случаев, так как вероятность ложно отрицательных случаев представляется площадью, ограниченной кривой распределения для совокупности больных лиц слева от прямой, проходящей через точку X_c . Если врач предпочитает консервативное поведение, чтобы выявлять возможно большее количество раковых больных, то он может выбрать точку X_c (двигаясь влево) так, чтобы количество ложно отрицательных случаев составило не более 1%.

Для решения вопроса о том, какое из двух заболеваний является наиболее вероятным, можно использовать метод проверки гипотез

тез. На фиг. 36 и 37 в сжатом виде представлены соотношения между двумя рассматриваемыми совокупностями. Следует иметь в виду, что здесь термин «больной» можно интерпретировать как «страдающий заболеванием D_1 », а термин «не больной» — как «страдающий заболеванием D_2 ».

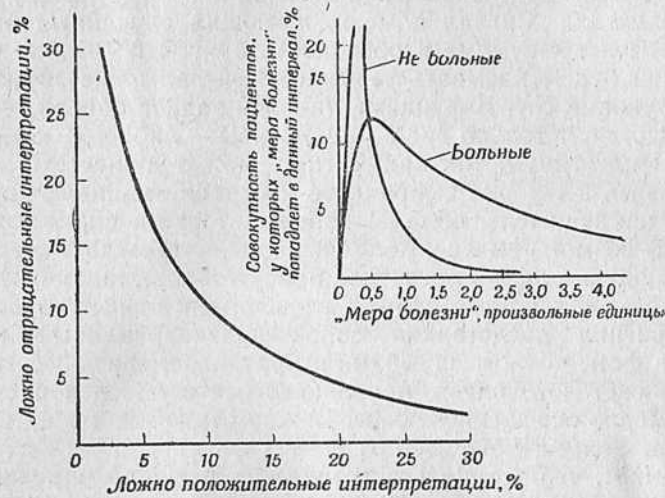


Фиг. 36.

Разделение совокупности лиц на две группы — «больные» и «не больные» (норма) на основе диагностического критерия для данного заболевания. Графики представлены в логарифмическом масштабе [66].

Мы предполагаем, что совокупности как истинно отрицательных случаев, так и истинно положительных случаев однородны и что эти две группы пересекаются в соответствии с некоторой «мерой болезни», или набором диагностических критериев, которые используются для определения наличия заболевания. Хорват с соавторами [39, 40] показал, что кривые распределения по отношению к некоторому выбранному критерию могут быть в значительной степени несимметричными. Но во многих случаях удастся получить хорошее приближение при помощи логарифмической нормальной функции распределения. На фиг. 32 представлены кривые, соответствующие реальным данным, а на фиг. 36 — кривые для гипотетических условий. На фиг. 37 представлены кривые фиг. 36 (справа вверху) в обычном, а не логарифмическом масштабе; здесь хорошо видна их асимметрия. Ось X можно рассматривать как ось критерия в том же смысле, как и на фиг. 35. Если прямую X_c перемещать вдоль

оси X , то для каждого ее положения можно вычислить процентное количество ложно отрицательных и ложно положительных случаев и результаты этих вычислений представить в виде графика. Рабочая характеристическая кривая, полученная в результате этой процедуры, показана на фиг. 37, слева внизу. Эта кривая имеет тот же смысл, что и кривые на фиг. 31 и 33.



Фиг. 37.

Справа сверху: плотность распределения вероятностей для совокупности, представленной на фиг. 36, в обычном масштабе; видна асимметричность кривых распределения. Слева внизу: рабочая характеристическая кривая — относительное количество ложно отрицательных случаев в функции относительного количества ложно положительных случаев. Кривая получена на основе функций плотности распределения, изображенных справа сверху, путем перемещения системы диагностических критериев («меры болезни») вдоль оси X (как это объяснено на фиг. 35) [66].

Результаты изучения ошибок наблюдателя при фотофлюорографии грудной клетки и интерпретации цитологических изменений при раке дают основание предположить, что можно отделить факторы, влияющие на диагностические критерии врача, от факторов, которые влияют на его субъективное отношение к тесту как наблюдателя, т. е. человека, принимающего решение. Если дальнейшее изучение этого вопроса в других областях медицины подтвердит эти выводы, то, возможно, удастся найти объективные критерии для некоторых ожидаемых величин, используемых при диагностическом решении, — величин, которые сейчас еще остаются субъективными и трудно выразимыми. Прежде чем планировать такие исследования, полезно ознакомиться с литературой по обнаружению и распознаванию сигнала человеком. Серия статей под редакцией Светса [111] является хорошим вводным материалом для этой цели.

4.5. ПРОБЛЕМА ОБНАРУЖЕНИЯ СИГНАЛОВ

Фундаментальная задача обнаружения сигнала заключается в следующем: в течение некоторого фиксированного интервала времени наблюдается некоторое явление; на основании этого наблюдения принимается решение о том, содержит ли рассматриваемый интервал только фон, т. е. сигнал помехи, или он содержит также и полезный сигнал. Сигнал помехи, имеющий случайный характер, будем называть «шумом» и обозначать буквой N ; второй возможный случай будем называть термином «сигнал плюс шум» и обозначать буквами SN . В фундаментальной задаче обнаружения сигнала существуют только две альтернативы — либо только шум, либо сигнал плюс шум, так как сигнал может присутствовать или отсутствовать в течение выбранного интервала наблюдения. Таким образом, наблюдатель, обладающий некоторыми априорными сведениями о возможном ансамбле сигналов, утверждает после каждого наблюдения либо «да, сигнал присутствует», либо «нет, сигнал отсутствует». В опубликованных экспериментальных работах полезный сигнал представлял собой кратковременный звуковой сигнал на фоне «белого шума» или кратковременную вспышку небольшого светового пятна, появляющегося в определенном месте на равномерно освещенном фоне. Важно отметить, что сигнал всегда наблюдается на фоне шума.

Напомним, что теория статистических решений определяет оптимальное поведение в такой ситуации, когда необходимо выбрать между двумя альтернативными статистическими гипотезами, основываясь на наблюдаемом явлении. Проблема обнаружения сигнала представляет собой как раз такую задачу статистического решения, и она требует применения метода проверки статистических гипотез. В психофизике существует мнение, что такой подход к процессу обнаружения сигнала может быть применен не только к «идеальному наблюдателю», но и к человеку-наблюдателю.

Светс [111] на основании своей экспериментальной работы приходит к выводу, что «отделение факторов, которые влияют на отношение наблюдателя к процессу, от факторов, которые влияют на его чувствительность, является основным достижением психофизического приложения теории статистических решений».

Это утверждение опирается на новые положения, напоминающие те, которые были получены при изучении ошибок наблюдателя, исследующего фотофлюорограмму грудной клетки или интерпретирующего цитологические изменения при раке. Какие данные лежат в основе утверждения Светса? Являются ли эти данные полезными для врача, изучающего ошибки наблюдателя в процессе установления медицинского диагноза? Мы рассмотрели некоторые из них; по всей вероятности, они могут оказаться полезными. Хочется подчеркнуть, что читатель все же должен довольно осторожно

но проводить непосредственные аналогии между психофизическими исследованиями и изучением ошибок наблюдателя в медицинской диагностике; тем не менее этот путь является обнадеживающим для дальнейших исследований.

4.6. ЗАМЕЧАНИЯ О ПРОЦЕССЕ ПРИНЯТИЯ РЕШЕНИЙ ПРИ ВОСПРИЯТИИ СИГНАЛА ЧЕЛОВЕКОМ

Исследование проблемы поведения наблюдателя в радиолокационных системах, выполненное 15 лет назад, привело к тому, что теория статистических решений трансформировалась в теорию обнаружения сигналов. Эта последняя носит характер общей теории, и ее общность позволила некоторым исследователям предположить, что она пригодна и для описания процесса обнаружения сигналов человеком-наблюдателем [111]. В частности, эта теория, по-видимому, может служить основой для реалистического описания поведения человека-наблюдателя в задачах восприятия. Как указывают авторы работы [114], наибольший интерес представляют свойства теории, позволяющие отделить критерий реакции, которым пользуется наблюдатель, осуществляя суждение в процессе восприятия, от факторов, влияющих на его чувствительность как обнаружителя сигнала. Благодаря возможности разделения этих двух аспектов деятельности наблюдателя теория может оказаться весьма полезной как для специалиста, изучающего индивидуальные и социальные процессы восприятия (сюда можно включить также медицинскую диагностику и лечение), так и для специалиста, изучающего сенсорные функции. Авторы указанной работы считают, что установление стандарта, при помощи которого можно было бы оценивать поведение человека-оператора, существенно помогло бы при описании и интерпретации экспериментальных результатов и было бы чрезвычайно полезным при планировании новых экспериментов.

Возможности применения теории обнаружения сигналов для оценки упомянутого выше критерия реакции описаны в работе [114] в терминах «фундаментальной задачи обнаружения».

«Экспериментатор указывает интервал времени наблюдателю, и последний должен решить, имел ли место сигнал в течение этого интервала. Предполагается, что каждый интервал содержит случайную помеху — шум, который либо является свойством среды, либо создается оборудованием экспериментатора, генерирующим полезные сигналы, либо умышленно вводится экспериментатором, либо просто является шумовым свойством сенсорной системы. Некоторые интервалы содержат, кроме шумового фона, специальный сигнал. Ответ наблюдателя ограничен только этими двумя классами событий: он говорит либо «да» (сигнал имеет место), либо «нет» (имеет место только шум). Заметим, что он не должен

говорить, видит ли он (или слышит) сигнал; он сообщает только, что в данных конкретных обстоятельствах он предпочитает принять решение, что сигнал имеет место, или решение, что сигнал отсутствует.

Относительно характера сенсорного возбуждения принимаются два предположения. Первое состоит в том, что оно непрерывно изменяется. Вследствие постоянного присутствия шума оно изменяется во времени и тогда, когда сигнал отсутствует, и от одной реализации сигнала к другой, хотя полезный сигнал номинально остается одним и тем же. Второе предположение состоит в том, что сенсорное возбуждение (поскольку оно влияет на ответ наблюдателя) может быть представлено как одномерная переменная. Согласно теории, наблюдатель полностью осведомлен о вероятностях каждого возможного состояния возбуждения, которое может иметь место на интервале наблюдения, содержащем только шум, и на интервале наблюдения, содержащем шум и полезный сигнал, и он основывает свой ответ на отношении этих двух вероятностей¹⁾, т. е. на отношении правдоподобия. Отношение правдоподобия, полученное на любом интервале наблюдения, является действительным числом, отличным от нуля, и, следовательно, может быть представлено на одномерной числовой оси.

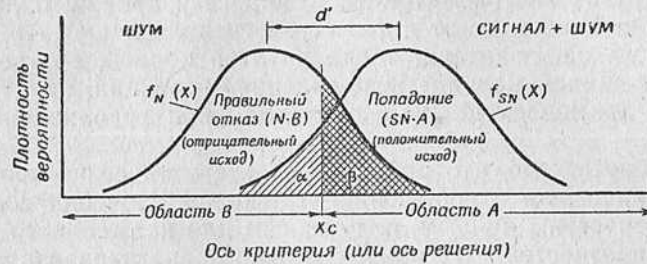
Предполагается, что ответ наблюдателя после окончания интервала наблюдения зависит от того, превышает или нет полученное на этом интервале отношение правдоподобия некоторую критическую величину — критерий реакции. Упомянутый критерий, как полагают, устанавливается наблюдателем в соответствии с преследуемой им целью при обнаружении сигнала и другими факторами, зависящими от конкретной ситуации. Если он желает максимизировать количество правильных ответов, то критерий должен зависеть от априорной вероятности появления сигнала на данном интервале. Если же наблюдатель хочет максимизировать общий доход²⁾, то его критерий должен зависеть от этой вероятности, а также от платежной матрицы, соответствующей четырем возможным исходам процесса решения. В любом случае критерий реакции, используемый наблюдателем, может быть выражен в виде величины отношения правдоподобия. Таким образом, решение наблюдателя для данного интервала основывается не только на воспринимаемой им на этом интервале сенсорной информации, но и на предварительных сведениях различного типа, а также на его собственной мотивации.

Рассмотрим далее некоторую вероятность, характеризующую переменное отношение правдоподобия, в частности вероятность

¹⁾ То есть отношении вероятности наблюдаемого возбуждения при условии, что на интервале наблюдения имеет место только шум, к вероятности этого возбуждения при условии, что на интервале наблюдения имеют место шум и полезный сигнал. — *Прим. перев.*

²⁾ В терминах теории игр. — *Прим. перев.*

различных значений отношения правдоподобия, имеющих место при каждом из двух классов возможных событий — «только шум» и «сигнал плюс шум»; обозначим соответствующие плотности вероятности через $f_N(x)$ и $f_{SN}(x)$. Таким образом, мы имеем два распределения вероятностей. При этом одно из них, соответствующее сигналу с шумом, будет иметь большее среднее значение (фиг. 38).



Фиг. 38.

Функции плотности распределения вероятностей для шума и совокупности «сигнал плюс шум». Ввиду того что X — непрерывная переменная, на графике представлена функция плотности вероятности, а не функция вероятности величины X . Теорема отсчетов, используемая в теории электрических цепей, позволяет провести квантование непрерывных функций, с которыми мы имеем дело в психофизике, и получить дискретные величины для приложения статистической теории решений. Взаимное расположение представленных плотностей вероятности определяется интервалом d' (эта величина определена в тексте), который остается относительно постоянным при изменении критерия реакции и при различных экспериментальных процедурах. Критерий реакции наблюдателя отмечается на оси X , так что возможные исходы его решения определяются соответствующей точкой X_c . Возможны следующие четыре исхода: правильное отрицание (истинно отрицательный исход), попадание (истинно положительный исход), промах (ложно отрицательный исход, область α), ложная тревога (ложно положительный исход, область β).

Если наблюдатель следует описанной выше процедуре, т. е. отвечает, что сигнал присутствует, когда отношение правдоподобия превышает некоторый критерий (X_c на фиг. 38), и отвечает, что имеется только шум, когда отношение правдоподобия меньше этого критерия, то из соответствующей матрицы возбуждения — реакция, содержащей четыре возможных исхода, можно получить две независимые меры: меру критерия реакции наблюдателя и меру его чувствительности.

Выделение этих двух мер осуществляется при помощи анализа с использованием рабочей характеристики. Если мы потребуем от наблюдателя изменять свой критерий при переходе от одной группы попыток к другой и если для каждого критерия на одной оси координат отметим относительное количество ответов «да», когда имеет место сигнал (относительное число «попаданий»), а на другой оси координат — относительное количество ответов «да», когда присутствует только шум (относительное число «ложных тревог»),

то с изменением критерия мы получим кривую, которая показывает, что относительное количество попаданий является неубывающей функцией относительного количества ложных тревог. Эта рабочая характеристическая кривая полностью описывает последовательные матрицы возбуждение — реакция, так как дополнения к этим двум относительным величинам представляют собой содержимое двух других клеток матрицы. Конкретная кривая, полученная таким методом, зависит от параметров сигнала и шума и от чувствительности наблюдателя; точка на этой кривой, соответствующая любой данной матрице возбуждение — реакция, представляет критерий, применяемый наблюдателем при реализации этой матрицы.

Было обнаружено, что рабочие характеристические кривые человека-наблюдателя с хорошей степенью приближения соответствуют теоретическим кривым, полученным для нормального распределения вероятностей. Эти кривые можно охарактеризовать одним параметром: разностью между средними значениями распределения «сигнал плюс шум» и распределения «шум», деленной на среднеквадратическое отклонение распределения шума. Этот параметр обозначается через d' . Кроме того, тангенс угла наклона касательной к любой точке рабочей кривой равен критерию отношения правдоподобия, который обеспечивает реализацию этой точки».

Индекс обнаруживаемости d' данного сигнала для данного наблюдателя можно определить как разность между средними значениями двух функций плотности вероятности, отнесенную к величине их среднеквадратического отклонения (фиг. 38):

$$d' = \frac{M_{f_{SN}(x)} - M_{f_N(x)}}{\sigma_{f_N(x)}}. \quad (1)$$

Величина d' представлена, как обычно, в виде нормированного отклонения x/σ . Предполагается, что плотности вероятности $f_N(x)$ и $f_{SN}(x)$ являются нормальными и отличаются только своими средними значениями, а иногда и среднеквадратическими отклонениями. Это предположение основано на экспериментальных результатах, представленных в виде рабочих характеристических кривых наблюдателя, воспринимающего сигнал.

Процесс наблюдения можно описать в терминах отношения правдоподобия

$$\lambda(X) = \frac{f_{SN}(X)}{f_N(X)}, \quad (2)$$

причем оптимальный критерий всегда может быть представлен в виде некоторого значения Y величины $\lambda(X)$. Прежде чем обсуждать вопрос о критическом значении Y , необходимо рассмотреть условные вероятности возможных исходов процесса решения.

Наблюдатель выбирает некоторое подмножество из всех наблюдений, а именно критическую область A (фиг. 38), так что любое наблюдение из этого подмножества приводит его к принятию гипотезы SN , когда он отвечает, что имеет место сигнал. Все другие наблюдения представляют собой дополнительное подмножество B ; любое из них приводит к отрицанию гипотезы SN или, что эквивалентно, к принятию гипотезы N , т. е. эти две гипотезы являются взаимоисключающими и единственно возможными. Критическая область A в соответствии с фиг. 38 содержит значение X , лежащее справа от точки, отвечающей данному значению критерия X_c .

Решение наблюдателя может иметь четыре исхода: наблюдатель может сказать «да» или «нет», и в каждом из этих случаев ответ может быть правильным или неправильным. Таким образом, решением может быть: «попадание» ($SN \cdot A$, или совпадение гипотезы SN и области A , т. е. «истинно положительный» случай); правильное отрицание ($N \cdot B$, или «истинно отрицательный» случай); «промах» ($SN \cdot B$, или «ложно отрицательный» случай, α); ложная тревога ($N \cdot A$, или «ложно положительный» случай, β).

Если априорная вероятность появления сигнала и параметры распределения (фиг. 38) фиксированы, то выбор критического значения X_c полностью определяет вероятность различных исходов. Ясно, что эти четыре вероятности являются взаимосвязанными. Увеличение вероятности попадания $p(SN \cdot A)$ (перемещение линии X_c влево) может быть достигнуто только за счет увеличения вероятности ложной тревоги (ложно положительного случая) $p(N \cdot A)$ и уменьшения двух других вероятностей — вероятности истинно отрицательного результата $p(N \cdot B)$ и вероятности ложно отрицательного результата $p(SN \cdot B)$. Это по сути дела то же самое явление, которое мы наблюдали при интерпретации фотофлюорограмм грудной клетки.

При психофизическом изучении процесса обнаружения сигналов особый интерес представляют две условные вероятности, а именно условная вероятность решения «да» $p_{SN}(A)$ в случае, когда сигнал имеет место (истинно положительный ответ):

$$p_{SN}(A) = \frac{p(SN \cdot A)}{p(SN)}, \quad (3)$$

и условная вероятность решения «да» в случае, когда имеет место только шум (ложно положительный ответ):

$$p_N(A) = \frac{p(N \cdot A)}{p(N)}. \quad (4)$$

В этих формулах $p(SN)$ — априорная вероятность появления сигнала и $p(N) = 1 - p(SN)$ — априорная вероятность того, что имеет место только шум.

По двум величинам $p_{SN}(A)$ и $p_N(A)$, полученным экспериментально, можно определить величину d' , воспользовавшись известными таблицами, дающими площадь, ограниченную кривой нормального распределения¹⁾ [111]. Вычислительная процедура состоит в вычерчивании графиков по точкам $p_N(A)$ и $p_{SN}(A)$ на бумаге в масштабе вероятности и (или) вероятности, нормированной по стандартным отклонениям.

Выбор оптимального критерия для $\lambda(X)$ требует рассмотрения вопроса об ожидаемой величине (обозначим ее через EV) для данного решения.

4.6.1. ВЫБОР СПОСОБА ДЕЙСТВИЯ И ОЖИДАЕМАЯ ВЕЛИЧИНА В ПРОЦЕССЕ РЕШЕНИЯ

Для большинства процессов принятия решения в условиях неопределенности или риска известен только один формальный принцип, который можно серьезно положить в основу выбора оптимального действия. Это принцип максимизации ожидаемой величины. Принцип минимакса и связанная с ним «концепция сожаления» не заслуживают серьезного рассмотрения как основа для принятия решений в реальных условиях. Для того чтобы использовать принцип максимизации ожидаемой величины, мы должны располагать четырьмя видами информации [18, 19]. Во-первых, должен быть сформулирован набор или список рассматриваемых действий (или симптомов, если это касается диагностики), которые должны быть хорошо определены. Во-вторых, должен быть сформулирован единственно возможный набор взаимоисключающих состояний рассматриваемого явления (гипотез или диагнозов), которые также должны быть хорошо определены. В-третьих, для каждой комбинации состояния и действия должно быть определено некоторое число, называемое полезностью (доходом); доход — это просто суждение человека, осуществляющего решение, относительно предпочтительности данной конкретной комбинации действия и состояния. В-четвертых, должна быть определена вероятность каждого состояния.

Если обозначить через w_{ij} полезность, связанную с j -м действием (S), когда действительно имеет место i -е состояние (D), то ожидаемая величина EV для j -го действия определяется как

$$EV_j = \sum_i P(D_i | S) w_{ij}. \quad (5)$$

Выражение (5) представляет собой уравнение осреднения, и, следовательно, принцип максимизации ожидаемой величины выражает просто тот факт, что мы должны вычислить средний доход

¹ Таблицы интегральной функции нормального распределения. — Прим. перев.

для каждого действия и выбрать такое действие, при котором этот средний доход наибольший.

Множитель $P(D_i|S)$ в уравнении (5) представляет собой вероятность дифференциального диагноза, полученную в результате вычислений при помощи теоремы Байеса, как было показано на примерах в предыдущих главах. Мы уже рассматривали требования, связанные с информацией первого, второго и четвертого видов, но нам еще не известны числа, относящиеся к информации третьего вида. Полезность представлена в уравнении (5) членом w_{ij} . Нам не встречалась ни одна работа в медицинской литературе, где была бы описана попытка выработать величины доходов для реальных условий, связанных с медицинской диагностикой и лечением. В разд. 4.7 будет показано, как можно определить величины доходов по данным, относящимся к интерпретации снимков грудной клетки. Но сейчас понятия ожидаемого дохода и ожидаемых потерь мы обсудим в терминах теории принятия решений.

Матрица полезности (доходов) и матрица потерь эквивалентны с точки зрения стратегии. Любое действие, которое является хорошим по отношению к одной матрице, является также хорошим и по отношению к другой. Приведем пример матрицы полезности:

		Категория диагностического решения врача	
		болен	не болен
Категория состояния пациента	болен	W_{11}	W_{12}
	не болен	W_{21}	W_{22}

В этой матрице величины полезности обозначены через W_{ij} . Матрицу потерь можно связать с матрицей полезности, вычитая каждый элемент матрицы полезности из наибольшего элемента в данной строке. Преимущество матрицы потерь состоит в том, что вместо четырех чисел полезности нам понадобятся только два числа.

		Полезность Категория решения		Потери Категория решения	
		болен	не болен	болен	не болен
Категория состояния пациента	болен	a	b	0	$a-b$
	не болен	c	d	$c-d$	0

Теперь можно получить ожидаемую величину для четырех возможных исходов процесса решения.

Обозначая через V положительные индивидуальные оценки и через K стоимости, или отрицательные оценки, получим

$$EV = V_{SN \cdot A} \cdot p(SN \cdot A) + V_{N \cdot B} \cdot p(N \cdot B) - K_{SN \cdot B} \cdot p(SN \cdot B) - K_{N \cdot A} \cdot p(N \cdot A). \quad (6)$$

Если подставить в уравнение (6) априорные и условные вероятности, т. е. вместо величины $p(SN \cdot A)$ подставить произведение $p(SN) \cdot p_{SN}(A)$, то максимизация величины EV будет эквивалентна максимизации выражения

$$EV_{\max} = [p_{SN}(A) - Y \cdot p_N(A)]_{\max}, \quad (7)$$

где

$$Y = \frac{p(N)}{p(SN)} \cdot \frac{(V_{N \cdot B} + K_{N \cdot A})}{(V_{SN \cdot A} + K_{SN \cdot B})}. \quad (8)$$

В связи с полученными выражениями необходимо изучить природу величин V и K . Нужно внимательно рассмотреть эти переменные, чтобы определить, с функциями какого типа мы имеем дело. Можно показать [111], что величина Y равна отношению правдоподобия $\lambda(X)$, соответствующему оптимальному критерию X_c на фиг. 38; величина Y просто взвешивает соотношение числа попаданий и ложно положительных случаев. Из уравнения (8) следует, что величина Y определяется априорной вероятностью сигнала и шума, априорной вероятностью только шума и оценками, связанными с отдельными исходами решения. Уравнение (8) можно трактовать как отношение вероятностей состояний при данных сигналах в связи с уравнениями (10) и (11) гл. 1, имея в виду, что величина Y представляет собой критическое отношение правдоподобия $L(D_1; S)$, а $p(N)/p(SN)$ — обратное отношение априорных шансов Ω_0 . При этом уравнение (8) можно записать в виде

$$X_c = Y = \frac{1}{\Omega_c} \cdot \frac{(V_{N \cdot B} + K_{N \cdot A})}{(V_{SN \cdot A} + K_{SN \cdot B})}. \quad (9)$$

Это выражение можно связать с отношением правдоподобия, использованным нами в процедуре медицинского разделения, описанной в гл. 2, разд. 5. Заметим, что величины V и K в числителе и знаменателе являются вполне определенными числами, если задана матрица потерь.

Возвращаясь к фиг. 38, отметим, что оптимальная точка разделения $X = X_c$ на оси X соответствует такому значению абсциссы, при котором отношение ординаты функции $f_{SN}(X)$ и ординаты функции $f_N(X)$ представляет собой некоторое заданное число, а именно Y . Напомним, что величина Y равна величине $\lambda(X) = f_{SN}(X)/f_N(X)$, соответствующей оптимальному критерию.

В симметричном случае, когда обе априорные вероятности равны и все четыре оценки для исходов также равны, $Y=1$ и оптимальное значение X_c достигается в точке, где $f_{SN}(X) = f_N(X)$, т. е. где две кривые пересекаются.

Важно отметить, что при любой цели, которую преследует наблюдатель в процессе обнаружения сигнала, и для любой системы параметров, которые могут характеризовать условия обнаружения (таких, как априорные вероятности и оценки, связанные с различными исходами решения), оптимальный критерий может быть выражен через одну величину Y , т. е. критическое значение отношения правдоподобия.

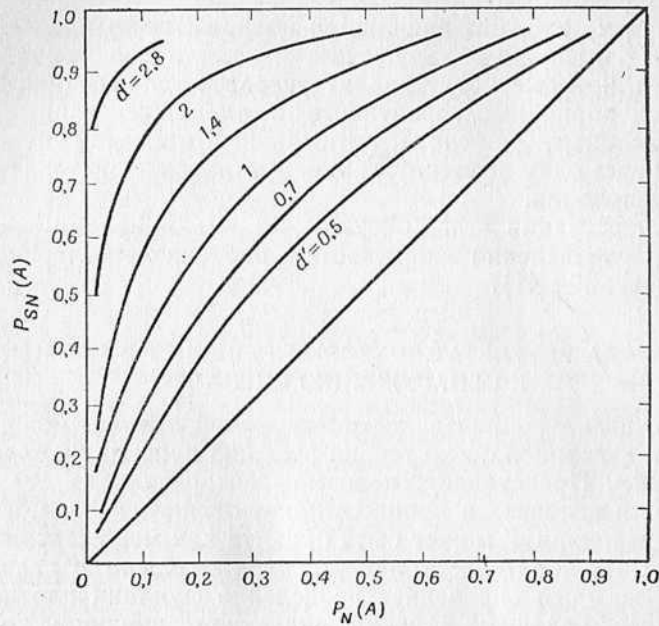
Для иллюстрации этого обсуждения мы приводим некоторые примеры распределений вероятности и рабочих характеристических кривых (фиг. 39—41).

4.6.2. СВОДКА РЕЗУЛЬТАТОВ ТЕОРИИ РЕШЕНИЙ В ПРИМЕНЕНИИ К ОБНАРУЖЕНИЮ СИГНАЛОВ

Рассмотрение вопросов психофизической теории обнаружения сигналов и связанной с ней теории решений приводит к следующим заключениям. Представляется возможным различать два аспекта деятельности человека в процессе обнаружения сигнала. Упомянутая выше величина d' может быть принята как мера сенсорных возможностей наблюдателя. Количественное значение d' зависит от критерия, который определяет разделение функции плотности вероятности наблюдаемых явлений. Значение d' инвариантно относительно различных экспериментальных процедур и критериев реакции.

Вторая мера λ отражает влияние таких переменных, как установка, отношение к процессу или мотивация наблюдателя. Величина λ является отношением правдоподобия, которое определяет выбор наблюдателя между двумя гипотезами. Наблюдатель пользуется критерием решения, основанным на отношении правдоподобия, поскольку он является оптимальным, если преследуются такие цели: 1) максимизация ожидаемой величины для решений, 2) минимизация риска, 3) определение апостериорной вероятности, 4) максимизация процента правильных решений и 5) удержание на определенном уровне частоты ошибок для некоторой альтернативы и максимизация количества правильных решений для другой альтернативы (последний подход был использован Колленом в процедурах отбора, описанных в гл. 2). Оптимальное значение λ может быть определено в виде некоторого критического числа, которое отражает оценки наблюдателя, или «стоимости» по отношению к каждому из четырех возможных исходов процесса решения. Для каждой выбранной точки на оси решения определена пара исходов процесса решения, например количество истинно положительных

случаев и количество ложно положительных. При перемещении критического значения $\lambda = Y$ по оси решений реализуется рабочая характеристическая кривая в виде функции количества истинно

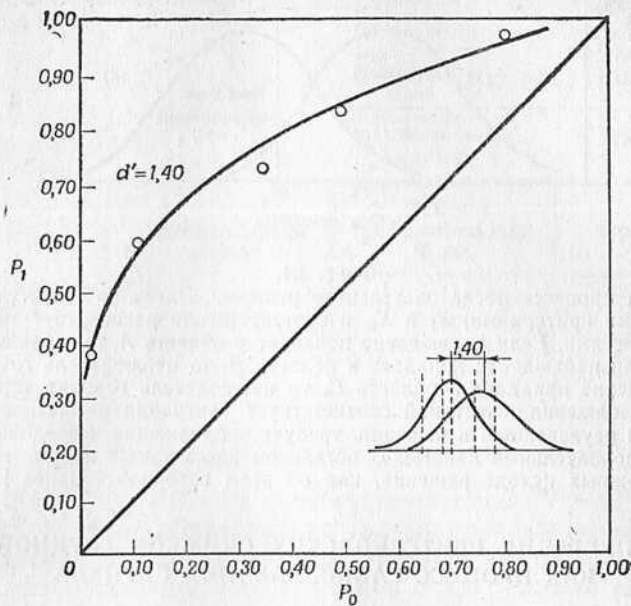


Фиг. 39.

Характеристические кривые (ROC) наблюдателя. Эти кривые представляют собой зависимость $p_{SN}(A)$ (условная вероятность истинно положительного исхода) от $p_N(A)$ (условная вероятность ложно положительного исхода) с величиной d' в качестве параметра. Кривые построены в предположении, что функции плотностей вероятности $f_N(X)$ и $f_{SN}(X)$ являются нормальными и имеют равные дисперсии. Видно изменение формы и положения кривых с увеличением d' , т. е. с увеличением расстояния между двумя функциями плотностей вероятности. Чтобы проверить предположение о том, что функции плотности вероятности относительно оси решения являются нормальными, можно построить кривые ROC в осях, градуированных в соответствии с нормальным распределением. Если функции плотности вероятности данного наблюдателя являются нормальными, то эмпирические точки кривой ROC, вычерченной относительно этих осей, лучше всего аппроксимируются прямой линией. Кривые для различных d' , представленные на фиг. 39, превращаются в прямые, если их вычертить относительно осей нормального распределения [114].

положительных случаев в зависимости от количества ложно положительных. Конкретное расположение точек на рабочей характеристической кривой зависит от взаимного расположения двух функций плотности вероятности. Величину Y можно вычислить по рабочей характеристической кривой, как тангенс угла наклона касательной к этой кривой в данной рабочей точке принятия решения.

Наклон касательной к характеристической кривой в рассматриваемой точке соответствует отношению правдоподобия, которое принимал наблюдатель в качестве своего критерия. Таким образом,

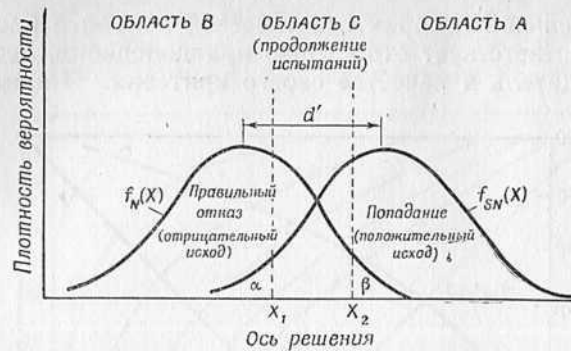


Фиг. 40.

Кривая ROC для данных, полученных в визуальных экспериментах. Заметим, что кривая с параметром d' не симметрична относительно отрицательной диагонали, как это имеет место для таких кривых на фиг. 39. Кривая такого типа характеризует случай, когда наблюдатель располагает неадекватной информацией относительно частоты и фазы сигнала, как, например, в случае сигнала белого света. Асимметрия функции плотности распределения вероятности, изображенной на внутреннем рисунке, отражает эту ситуацию. Такого же типа кривые имеют место в экспериментах типа «да — нет», в процедуре вынужденного выбора и в процедуре разделения [112].

мы можем решать обратную задачу — по рабочей характеристической кривой наблюдателя определять критерий, которым этот наблюдатель пользовался.

Вернемся теперь к рассмотрению интерпретации рентгенограмм грудной клетки, чтобы выяснить, можно ли в этом случае определить значения d' и λ , используемые в этой диагностической ситуации. В частности, мы хотели бы выяснить, сможет ли такой подход помочь в определении тех оценок значимости, которые наблюдатель (врач) может приписывать четырем возможным исходам решения: истинно положительному, истинно отрицательному, ложно положительному и ложно отрицательному.



Фиг. 41.

Иллюстрация процесса последовательного решения. Здесь наблюдатель руководствуется двумя критериями, X_1 и X_2 , и соответственно располагает тремя критическими областями. Если наблюдение попадает в область A, то наблюдатель говорит «да». Если наблюдение попадает в область B, то наблюдатель говорит «нет». Если наблюдение попадает в область C, то наблюдатель говорит «продолжить». Область продолжения испытаний соответствует ситуации в диагностике, когда врач, будучи неуверенным в решении, требует продолжения исследований, чтобы получить дополнительные данные. В остальном здесь также имеют место четыре возможных исхода решения, как об этом говорилось ранее [113].

4.7. ИНТЕРПРЕТАЦИЯ РЕНТГЕНОВСКИХ СНИМКОВ ГРУДНОЙ КЛЕТКИ КАК ПРОЦЕСС РАСПОЗНАВАНИЯ СИГНАЛА

Предположим, что функция шума образуется несколькими источниками: некоторым суммарным шумом (рассеивание рентгеновских лучей, зернистость пленки, случайные движения глаза), который приводит к расплывчатым очертаниям изображения на рентгеновском снимке грудной клетки, и «шумом», вносимым нормальной популяцией. Плотность вероятности распределения этой функции шума, $f_N(D)$, неизвестна. Предположим, далее, что существует вторая функция, аналогичная функции «сигнал плюс шум», где сигнал — это заболевание, рассматриваемое в процессе разделения или в диагностическом исследовании. Плотность распределения вероятности этой функции обозначим через $f_{SN}(D)$; она также неизвестна.

При исследовании процесса интерпретации рентгеновских снимков грудной клетки мы начинаем с доступных нам данных и действуем в обратном направлении, чтобы посмотреть, можно ли оценить плотности распределения $f_N(D)$ и $f_{SN}(D)$, которые лежат в основе наблюдаемых результатов. Точки кривой на фиг. 31, представляющие количество ложно отрицательных случаев (FN) в функции количества ложно положительных случаев (FP), можно преобразовать в точки кривой на фиг. 42, так чтобы они представляли количество истинно положительных случаев (TP) в функции

количества ложно положительных случаев (FP). Каждая точка кривой на фиг. 31 может быть подвергнута такому преобразованию путем построения матрицы решений.

Матрица решений для точки $FN=20\%$, $FP=4\%$ (фиг. 31) имеет следующий вид:

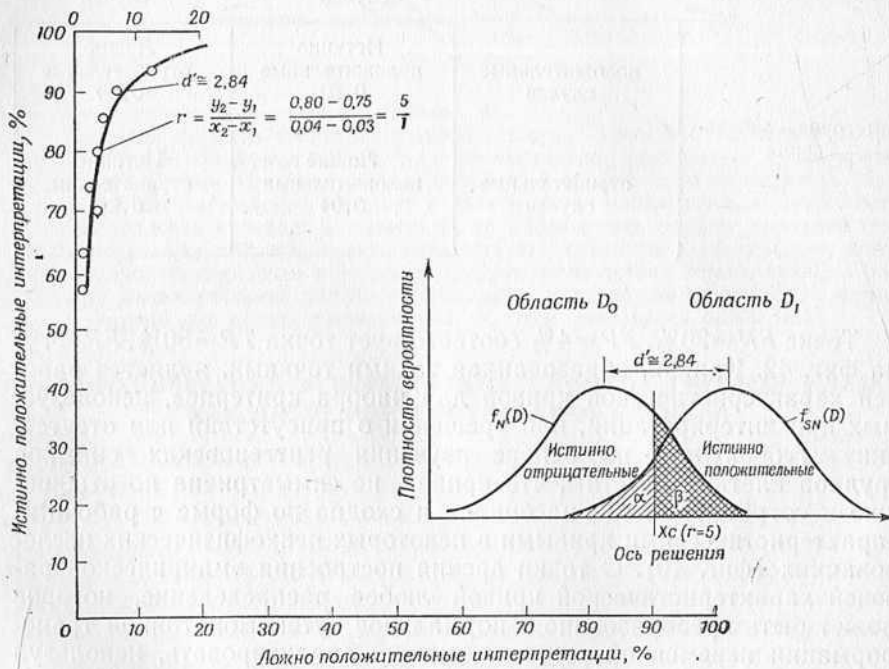
		Категория решения	
		положительные случаи	отрицательные случаи
Категория заболевания	положительные случаи	Истинно положительные 0,80	Ложно отрицательные 0,20
	отрицательные случаи	Ложно положительные 0,04	Истинно отрицательные 0,96

Точке $FN=20\%$, $FP=4\%$ соответствует точка $TP=80\%$, $FP=4\%$ на фиг. 42. Кривая, образованная такими точками, является рабочей характеристической кривой для набора критериев, используемых при интерпретации, или «решении о присутствии или отсутствии» туберкулеза на основе изучения рентгеновских снимков грудной клетки. Заметим, что кривая не симметрична по отношению к «отрицательной диагонали» и сходна по форме с рабочими характеристическими кривыми в некоторых психофизических исследованиях (фиг. 40). С точки зрения построения эмпирической рабочей характеристической кривой любое распределение, которое может быть преобразовано в нормальное путем монотонной трансформации переменной решения, можно анализировать, используя предположение о нормальности функций распределения по оси наблюдения. Тем не менее к таким предположениям следует относиться с осторожностью.

Рабочая кривая имеет монотонно уменьшающийся наклон касательной, и это означает, что мы пользуемся подходящим правилом решения, даже если диагностические критерии и правило решения не были точно определены. Любая рабочая характеристическая кривая, которая не имеет монотонно убывающего наклона, основана на неверном правиле решения. Это значит, что наблюдатель принимает первую гипотезу, когда отношение правдоподобия превышает определенное значение, и все же принимает вторую гипотезу, когда отношение правдоподобия становится еще выше. Любая инверсия в наклоне характеристической кривой указывает на то, что путем

перестановки некоторых точек на оси решения можно улучшить качество работы.

Некоторая информация может быть получена при изучении выборки из общей популяции. В работе [35] проиллюстрированы практические аспекты исследований, связанные с массовым обследованием для выявления туберкулеза легких. Указано, что при обследовании 200 000 взрослых людей в США среднее количество истинно положительных случаев острого туберкулеза легких равнялось



Ф и г. 42.

Рабочая характеристическая кривая для процедуры просмотра фотофлюорограмм грудной клетки с целью обнаружения туберкулеза легких. Данные фиг. 31 были представлены в виде зависимости процента ложно положительных случаев от процента истинно положительных. Наклон кривой монотонно убывает; это указывает на то, что применялось корректное правило решения. Ни в одном из представленных исследований не были определены диагностические критерии для d' . Мера d' может быть получена из таблиц (см. текст), а величина Y вычисляется как наклон кривой в любой выбранной рабочей точке. Эти данные использовались для того, чтобы воссоздать возможный вид функции плотности распределения шума $f_N(D)$ (помехи среды и т. д.), и сигнала $f_{SN}(D)$ — диагностического критерия для рассматриваемого заболевания. После того как экспериментальные точки были вычерчены в нормальных координатах, оказалось, что они лучше всего аппроксимируются прямой линией. Таким образом, предположение о нормальности функций плотности вероятности наблюдателя $f_N(D)$ и $f_{SN}(D)$, введенное для простоты, подтверждается экспериментально.

100. При одноразовой интерпретации 200 000 снимков грудной клетки опытный рентгенолог в среднем обнаруживает семьдесят и пропускает тридцать истинно положительных случаев (30 ложно отрицательных ответов). Кроме того, один интерпретатор «находит» приблизительно 2196 положительных случаев (ложно положительные ответы) из 199 900 истинно отрицательных. На основе этих данных можно получить следующие вероятности:

Априорная вероятность истинно положительных случаев $p(SN) = P(D_1) = \frac{100}{200\,000} = 0,0005$

Априорная вероятность истинно отрицательных случаев $p(N) = 1 - p(SN) = 1 - P(D_1) = 0,9995$

Условная вероятность истинно положительных случаев $p_{SN}(D_1) = P_{TP}(D_1|S) = \frac{70}{100} = 0,7$

Условная вероятность ложно положительных случаев $p_N(D_1) = P_{FP}(D_1|S) = \frac{2196}{199\,900} = 0,01$

Мера d' может быть определена из условных вероятностей для истинно положительных $p_{SN}(D_1)$ и ложно положительных $p_N(D_1)$ случаев при помощи таблиц, представленных в работе [111]. Для $p_{SN}(D_1) = 0,7$ и $p_N(D_1) = 0,01$ значение d' равно 2,84. Это значение представляется подходящим с точки зрения формы и расположения рабочей кривой, показанной на фиг. 42, если основываться на опубликованных Светсом кривых для $d' \approx 2,8$. Следовательно, можно предположить, что наше значение d' аппроксимируется величиной 2,8, хотя нам не известна точная функция плотности распределения.

Меру Y можно определить по рабочей характеристической кривой. Она равна тангенсу угла наклона касательной к кривой в определенной точке и представляет собой оптимальное отношение правдоподобия, используемое как критическое значение на оси решения. В нашем примере интерпретации группы рентгеновских снимков грудной клетки величина Y вычисляется по наклону касательной в точке, соответствующей относительному числу истинно положительных случаев 0,8 и относительному числу ложно положительных случаев 0,04. Эта величина равна $m = (y_2 - y_1) / (x_2 - x_1)$, где $y_2 = 0,8$, $x_2 = 0,04$ и $y_1 = 0,75$, $x_1 = 0,03$, т. е. $m = (0,8 - 0,75) / (0,04 - 0,03) = 5/1 = 5 = Y = X_c$.

Грубое приближение функций $f_N(D)$ и $f_{SN}(D)$ можно построить из оценок априорных вероятностей, условных вероятностей, величины d' и величины Y , как показано в правом нижнем углу на фиг. 42.

Ранее было введено понятие ожидаемой величины решения (EV) и было отмечено, что ожидаемая величина должна быть максимизирована в отношении целей или полезностей, которые имеет в виду наблюдатель.

Максимизация величины EV эквивалентна максимизации выражения

$$p_{SN}(D_1) - Y \cdot p_N(D_1). \quad (10)$$

В примере интерпретации рентгеновских снимков грудной клетки

$$p_{SN}(D_1) = 0,7,$$

$$p_N(D_1) = 0,01.$$

Итак, мы имеем

$$p_{SN}(D_1) - Y \cdot p_N(D_1) = (0,7) - Y \cdot (0,001),$$

и в качестве переменной, которую нужно максимизировать, остается только отношение правдоподобия

$$Y = \frac{p(N)}{p(SN)} \cdot \frac{(V_{N \cdot D_0} + K_{N \cdot D_1})}{(V_{SN \cdot D_1} + K_{SN \cdot D_0})}, \quad (11)$$

где V — положительная индивидуальная оценка (полезность) и K — цена, или отрицательная индивидуальная оценка (потери). Поскольку $p(N)$ и $p(SN)$ — априорные вероятности, значение Y максимизируется путем варьирования V и K .

Чтобы продемонстрировать связь четырех возможных решений, построим следующую матрицу:

		Категория решения	
		положительные случаи	отрицательные случаи
Категория заболевания	положительные случаи	Истинно положительные, полезность $V_{SN \cdot D_1} = V_{TP}$	Ложно отрицательные, цена $K_{SN \cdot D_0} = K_{FN}$
	отрицательные случаи	Ложно положительные, цена $K_{N \cdot D_1} = K_{FP}$	Истинно отрицательные, полезность $V_{N \cdot D_0} = V_{TN}$

Уравнение (11) можно переписать в терминах FN (ложно отрицательные случаи) и FP (ложно положительные случаи) следующим образом:

$$Y = \frac{p(N)}{p(SN)} \frac{(\text{полезность}_{\text{истинно отрицательные}} + \text{цена}_{\text{ложно положительные}})}{(\text{полезность}_{\text{истинно положительные}} + \text{цена}_{\text{ложно отрицательные}})}$$

Если мы подставим в эту формулу значения величин Y , $p(N)$ и $p(SN)$ из рассмотренного примера интерпретации фотофлюорограмм грудной клетки, то получим

$$5 = \frac{0,9995 (V_{TN} + K_{FP})}{0,0005 (V_{TP} + K_{FN})},$$

$$\frac{(V_{TN} + K_{FP})}{(V_{TP} + K_{FN})} \approx \frac{5}{2000} = \frac{1}{400} \text{ единиц.} \quad (12)$$

Такой подход должен помочь более четко сформулировать идеи врача об относительных значениях полезности и цен возможных четырех исходов процесса решения. Если $V_{TN} = V_{TP} = 1$ единиц (единица пользы в теории полезности), то уравнение (11) принимает вид

$$\frac{1 + K_{FP}}{1 + K_{FN}} \approx \frac{1}{400} \text{ единиц.}$$

Если цены K_{FN} и K_{FP} относительно велики по сравнению со значениями полезности V_{TN} и V_{TP} , то можно записать

$$\frac{K_{FP}}{K_{FN}} = \frac{1}{400} \text{ единиц} \text{ или } \frac{K_{FN}}{K_{FP}} = \frac{400}{1} \text{ единиц.} \quad (13)$$

Возможная интерпретация уравнения (13) для рассматриваемого примера состоит в том, что мы готовы заплатить за ложно отрицательный случай в четыреста раз большую цену, чем за ложно положительный. Другими словами, мы готовы получать или платить по четыреста ложно положительных случаев за каждый возможный пропущенный положительный случай (т. е. за каждый ложно отрицательный). Ложно положительные случаи требуют дальнейших исследований, которые имеют дополнительную стоимость для диагностической системы. Принцип «лучше обезопасить себя, чем потом сожалеть» характеризует консервативное отношение к процессу решения, которое благоприятствует принятию относительно большего числа ложно положительных диагнозов, чем ложно отрицательных. Но готовы ли мы заплатить за четыреста новых исследований, связанных с ложно положительными решениями, лишь бы не пропустить этот положительный случай? Должны ли мы принять какое-то другое отношение, такое, как 200 к 1, или 50 к 1, или 10 к 1? Насколько консервативными мы хотим быть? Такие вопросы врач должен задавать себе, и мы хотим предложить некоторые подходы, которые помогут ему найти ответы на них в будущем.

Читатель может подставить в приведенные выше уравнения другие значения вероятностей, полезностей и цен, чтобы охаракте-

ризовать свою собственную деятельность. Кроме того, в качестве объекта исследования можно взять заболевание, имеющее относительно более широкое распространение, чем туберкулез. Для врача очень важно учитывать относительные значения полезностей и цен в их связи с исходами его решения. Необходимо ввести какую-то единицу значения полезности. Линдер [56] при обсуждении состояния здоровья американского населения указал, что медицинской науке все еще не хватает «единицы обращения» (такой, как доллар в экономике), которая, возможно, позволила бы разработать некоторый единый индекс здоровья. Каким образом можно было бы выразить значения полезности и цены альтернативных действий в медицинской практике в «денежных» единицах? Автор работы [97] предполагает, что одним из решений может быть использование платежных уровней при нетрудоспособности и страховании жизни; они дают сравнительные значения «цен» смерти и постоянной или временной нетрудоспособности различной степени. Хотя этот подход не отражает все обстоятельства, учитываемые врачом при вынесении решения (боль и страдания пациента не могут быть втиснуты в эти рамки), он все-таки включает многие из основных факторов. Такой подход дал бы возможность создать более реалистическую матрицу «дохода» из абсолютных значений, которая позволит затем определить относительное значение «стоимости» того или другого вида лечения (или отказа от лечения) при помощи уравнения ожидаемых величин.

Читатель может не согласиться с этим, заметив, что от врачей обычно не требуются суждения в терминах безусловных альтернатив, таких, как «да» или «нет», больной или здоровый, и что предыдущее обсуждение методов проверки гипотез не очень реалистично. В этом возражении есть доля истины, но психофизические эксперименты [111] показали, что процедуры оценки значимости дают значения d' , неотличимые от значений, полученных в процедуре типа «да» — «нет». Экспериментальные результаты подтвердили осуществимость процедуры оценки; например, они показали, что требование к наблюдателю руководствоваться одновременно пятью критериями решения в задаче обнаружения не является бессмысленным. Испытуемые были вполне способны осуществлять наблюдения.

Однако процедуры оценки довольно редко использовались при изучении процесса принятия решения в медицинской диагностике и при лечении. Эти методы требуют дальнейшего изучения. В работе [46] применялась процедура оценки для изучения адекватности прогностического суждения кардиолога в случаях инфаркта миокарда. Цель исследования заключалась в том, чтобы определить, можно ли надежно измерить прогностическое суждение кардиолога, используя метод оценки.

4.8. ПРОЦЕДУРА ОЦЕНКИ ЗНАЧИМОСТИ, ИСПОЛЬЗУЕМАЯ ДЛЯ ИЗУЧЕНИЯ ПРОГНОСТИЧЕСКОГО СУЖДЕНИЯ ПРИ ИНФАРКТЕ МИОКАРДА

Были исследованы прогностические суждения для двадцати случаев инфаркта миокарда. Для того чтобы определить, какая информация должна быть предоставлена врачам, осуществляющим соответствующие суждения, была проведена серия интервью с кардиологами и терапевтами. Их спрашивали, какая информация им нужна для того, чтобы иметь возможность немедленно высказать определенное прогностическое суждение. Немедленный прогноз сводился к предсказанию возможности выживания при данной атаке болезни. На основе полученной информации были просмотрены истории болезни больных с инфарктом миокарда в четырех больницах. Принимались во внимание двадцать пять переменных, и в соответствии с ними были отобраны двадцать историй болезни, хотя ни в одном из этих случаев не удалось получить всей желаемой информации. Были взяты те истории болезни, которые ближе всего соответствовали пожеланиям опрошенных специалистов в отношении необходимой информации. Данные для каждого пациента записывались на отдельном листе.

Процедура теста заключалась в том, что исследователи попросили двадцать кардиологов расположить эти истории болезни в порядке их предпочтительности с точки зрения возможности немедленного прогноза. На основе этих построений были выработаны числовые оценки (порядковая оценка с применением операций нормирования), причем надежность полученных значений была 0,97 (при максимальной возможной величине 1,00). В этом эксперименте кардиологи оказались способными сделать весьма точные прогностические суждения.

Один из интересных результатов этой работы [46] состоит в том, что количество времени, затрачиваемое кардиологами для вынесения прогностического суждения, не повлияло на качество суждения. Это открытие находится в противоречии с выводом, сделанным при исследовании диагностики фотофлюорограмм грудной клетки, которая проводилась с помощью процедур типа «да» — «нет»; в последней работе количество времени, затраченное на решение, по-видимому, было связано с характером суждений [39]. Эти результаты позволяют предположить, что при разделении фотофлюорограмм грудной клетки врачи использовали процессы поиска и распознавания образов, тогда как кардиологи, делавшие прогностические суждения об инфаркте миокарда, использовали «какой-то другой» процесс. Эти вопросы требуют дальнейшего исследования.

4.9. ИЗЛОЖЕНИЕ ТЕОРИИ ОБНАРУЖЕНИЯ СИГНАЛОВ НА БОЛЕЕ ОБЩЕМ (БАЙЕСОВСКОМ) ЯЗЫКЕ

Читатель, возможно, заметил некоторое сходство между обсуждаемым в этой главе процессом принятия решения, основанным на теории обнаружения сигналов (психофизические приложения), и байесовским подходом, который был изложен в гл. 1 и 2. Оба подхода могут быть описаны как «байесовские», и, следовательно, полезно будет исследовать некоторые их сходства и различия. Очевидное различие заключается в том, что для представления приблизительно одних и тех же понятий используются два вида обозначений. Для того чтобы избежать путаницы, мы приводим на фиг. 43 таблицу эквивалентных терминов.

С одной стороны, различия между подходами исследователей, применяющих байесовские методы, и исследователей, пользующихся методами теории обнаружения сигналов, носят философский характер. Эта разница во мнениях иллюстрируется тем фактом, что ссылки на теорему Байеса или «байесовский подход» редко встречаются в психофизической литературе, связанной с теорией обнаружения сигналов. По крайней мере одно объяснение этому дано Грином и Светсом [37] в их книге «Теория обнаружения сигналов и психофизика». При обсуждении правила Байеса они заметили: «В философии науки возник большой спор об этом правиле. Разногласия сводятся к вопросу о том, как могут быть определены априорные вероятности; и действительно, поскольку мир находится только в одном состоянии, имеет ли смысл думать о вероятности того, что существуют различные состояния мира. Мы будем считать, что априорные вероятности являются просто частью той информации, которая задается при постановке задачи принятия решения, и таким образом исключим возможные возражения».

Различие мнений по поводу использования априорных вероятностей, априорных правдоподобий или априорных шансов может в действительности быть не столь значительным, как кажется. Существует изоморфизм между байесовской теорией решения и теорией обнаружения сигнала в психофизическом приложении.

Цель этого параграфа — выявить идентичность двух указанных концепций, несмотря на различия в обозначениях, подчеркнуть большую универсальность идей, используемых специалистами по теории обнаружения сигналов, и показать роль некоторых специальных предположений, принимаемых этими исследователями [17].

Мы начнем с того, что рассмотрим проблему обнаружения сигнала с точки зрения общей теории решения. Пусть стимул или сигнал D_1 появляется либо на фоне шума, либо отсутствует, так что имеется только шумовой фон D_0 . Наблюдатель знает, какова вероятность присутствия стимула $P(D_1)$, и может реализовать только две ответные реакции: он может сообщить, что сигнал имел место или что

сигнала не было. Каждой возможной комбинации действия и состояния соответствует «платеж» (который может быть положительным или отрицательным). Ниже приведена соответствующая платежная матрица:

		Диагностическая категория решения врача		
		положительный случай	отрицательный случай	
Категория состояния (заболевания) пациента	болен	w	x	$w > x$
	не болен	y	z	$z > y$

Мы принимаем, что лучше дать правильный ответ, чем ошибиться, т. е. $w > x$ и $z > y$. Отметим далее, что идеализированный «оптимальный» наблюдатель выберет действие, которому соответствует самая высокая ожидаемая величина. Но вероятности состояний, которые влияют на его выбор, имеют характер не априорных, а апостериорных вероятностей $P(D_i|S)$, находящихся в его распоряжении после окончания наблюдения. Поскольку исследуются только две гипотезы, достаточно рассмотреть апостериорные шансы $P(D_1|S)/P(D_0|S)$. Приближение величины шансов к бесконечности соответствует области, в пределах которой лучше сказать, что стимул имеет место, а приближение величины шансов к нулю соответствует области, в пределах которой лучше сказать, что стимула нет. Легко показать, что существует некоторое значение шансов, которое определяет границу между этими двумя областями, т. е. такое значение, при котором оба действия одинаково привлекательны. Обозначим это значение шансов через Ω_c , а соответствующее значение вероятности $P(D_1|S)$ — через P_c .

Тот факт, что два действия одинаково привлекательны, означает следующее равенство:

$$P_c w - P_c x = (1 - P_c) z - (1 - P_c) y. \tag{14}$$

Перегруппировав члены, получим:

$$P_c (w - x) = (1 - P_c) (z - y), \tag{15}$$

$$\frac{P_c}{1 - P_c} = \frac{z - y}{w - x} = \Omega_c.$$

Теперь по формуле Байеса получаем

$$\Omega_c = L_c \Omega_0, \tag{16}$$

Байесовские формулы		Формулы теории обнаружения сигнала (психофизика)
$P(D_1)$	Априорная вероятность заболевания D_1	$= p(SN)$
$P(D_0) = 1 - P(D_1)$	Априорная вероятность заболевания D_0	$= p(N) = 1 - p(SN)$
$P(D S)$	Условная вероятность заболевания D при данных симптомах S	$= p_{SN}(D)$
$P_{TP}(D_1 S)$	Условная вероятность истинно положительного случая D_1 при данных симптомах S	$= p_{SN}(D_1)$
	$\approx \frac{\omega}{\omega + x}$	
$P_{FP}(D_1 S)$	Условная вероятность ложно положительного случая D_1 при данных симптомах S	$= p_N(D_1)$
	$\approx \frac{y}{y + z}$	
	где ω , x , y и z приведены в нижеследующих матрицах решения	

Категория состояния (заболевания) пациента	Диагностическая категория решения врача		Категория сигнала	Сигнал плюс шум		Только шум	
	положительные ответы	отрицательные ответы		имеет место	не имеет места	ω	x
{ болен не болен	ω	x	{ имеет место не имеет места	y	z	ω	x
	y	z		y	z	y	z

$$d' = \frac{M_{D_1} - M_{D_0}}{\sigma_{D_0}}$$

d' можно вычислить по $P_{TP}(D_1|S)$ и $P_{FP}(D_1|S)$

Таблицы, предложенные в работе [111], могут быть использованы для определения d'

Критическое отношение правдоподобия L_c

$$L_c = Y = \frac{1}{Q_0} \cdot \frac{V_1}{V_0} = \frac{P(D_0)}{P(D_1)} \cdot \frac{V_{TN} + K_{FP}}{V_{TP} + K_{FN}}$$

$$d' = \frac{M_{SN} - M_N}{\sigma_N}$$

d' можно вычислить по $p_{SN}(D_1)$ и $p_N(D_1)$

Использование таблиц, предложенных в работе [111]

Критическое отношение правдоподобия X_c

$$X_c = Y = \frac{p(N)}{p(SN)} \cdot \frac{V_{\text{истинно отриц}} + K_{\text{ложная тревога}}}{V_{\text{попадание}} + K_{\text{промах}}}$$

Фиг. 43.

Таблица эквивалентных терминов для байесовского подхода к процессу принятия решения и подхода, основанного на теории обнаружения сигнала (психофизика). Здесь D_1 и D_0 — два возможных состояния пациента; врач имеет два возможных выбора. Он может сказать: «да, заболевание D_1 имеет место», или «нет, заболевание D_1 не имеет места». Это аналогично ситуации, возникающей при обнаружении сигнала, когда наблюдатель сообщает «да, сигнал присутствует (SN)», или «нет, сигнал отсутствует (N)».

где L_c — критическое отношение правдоподобия, соответствующее преобразованию Ω_0 в Ω_c . Итак,

$$L_c = \frac{z-y}{w-x} \frac{1}{\Omega_0}. \quad (17)$$

Если субъект проводит наблюдение, дающее некоторое отношение правдоподобия L , то при $L > L_c$ он скажет, что стимул присутствовал, а при $L < L_c$ — что стимул отсутствовал; за исключением различия в обозначениях, это идентично выбору величины L_c в соответствии с уравнением (17).

Использование величины L_c в действительности несколько обманчиво. Человек, принимающий решение, не основывает его на значении L_c . Он основывает свое решение на собственном апостериорном мнении, которое зависит от его априорного мнения и от наблюдения или наблюдений, которые он провел. Это положение становится более очевидным, если рассмотреть ситуацию, когда наблюдатель может (если захочет) сделать несколько наблюдений. В такой ситуации именно апостериорное мнение управляет решением, принимаемым после завершения всех наблюдений.

Все же разумно разделить факторы, управляющие решением, на такие переменные, которые могут быть известны до наблюдения, а именно Ω_0 , w , x , y и z , и переменную L , которая может быть известна только после наблюдения. Специалисты по теории обнаружения сигнала особенно заинтересованы в рассмотрении процесса наблюдения, отделенном от процесса принятия решения, для которого наблюдения являются входными данными. Поэтому они предпочитают разделить проблему решения таким образом, чтобы сосредоточить внимание на критическом отношении правдоподобия L_c .

Конечно, все эти замечания еще более применимы к решению о наличии или отсутствии заболеваний, принимаемому на основе наблюдения некоторого симптома или набора симптомов, чем о наличии или отсутствии сигнала.

4.9.1. РАБОЧИЕ ХАРАКТЕРИСТИЧЕСКИЕ КРИВЫЕ НАБЛЮДАТЕЛЯ

Для любого значения апостериорных шансов $\Omega[(D_1/D_0)|S]$ будет существовать вероятность ошибки, если наблюдатель утверждает, что гипотеза D_1 истинна. Эта вероятность равна $1 - P(D_1|S)$. Существует также вероятность ошибки, если наблюдатель утверждает, что истинна гипотеза D_0 . Однако более интересными, чем эти безусловные вероятности ошибок, являются две условные вероятности $P(L > L_c | D_1)$ и $P(L > L_c | D_0)$.

Если $L > L_c$, то наблюдатель утверждает, что гипотеза D_1 истинна; следовательно, первая из этих вероятностей есть вероятность истинно положительного ответа, а вторая — вероятность ложно положительного ответа. С увеличением L_c обе эти вероятности возраста-

ют, а с уменьшением L_c обе уменьшаются. Если отложить значения $P(L > L_c | D_0)$ по оси X и значения $P(L > L_c | D_1)$ — по оси Y , то получаемый график представляет собой рабочую характеристическую кривую наблюдателя или, сокращенно, ROC — кривую, которую мы уже обсуждали ранее. Кривые ROC являются удобным средством для изучения качества работы любого механизма, вырабатывающего информацию.

В теории обнаружения сигналов принято считать шумом некоторое явление, вызывающее распределение возможных значений некоторой переменной, обычно обозначаемой через x ; считается, что собственно сигнал создает такое же распределение, сдвинутое в сторону больших значений x . Удобно принять, что оба эти распределения нормальны и имеют одинаковые дисперсии. При таких предположениях единственной подходящей переменной, которая указывает, насколько хорошо наблюдатель может отличать сигнал от шума, является расстояние между средними значениями этих двух кривых распределений, измеренное в единицах их общего стандартного отклонения. Это расстояние обозначим через d' . В типичных исследованиях по обнаружению сигнала величина d' является зависимой переменной. При этом обычная процедура заключается в получении кривой ROC или набора таких кривых с последующим определением по ним d' . Каждое значение d' однозначно определяет некоторую кривую ROC. Заметим, однако, что использование таких понятий, как L_c и кривые ROC, ни в коей мере не связано с довольно специальными и часто недостаточно обоснованными предположениями о нормальности кривых распределения и равенстве стандартных отклонений. Эти предположения являются решающими для любого анализа, использующего такие величины, как d' , но вовсе не необходимы для более общих приложений теории решения. Это особенно важно, если (как это часто случается в медицинской диагностике) диагностические признаки рассматриваются как дискретные, а не непрерывно распределенные переменные.

В гл. 1 и 2 мы подробно не обсуждали проблемы оценки величин $P(S|D)$, а лишь указали, что такие оценки часто бывают существенно субъективными. Однако иногда они являются все же достаточно объективными. Если для каждого симптома, набора симптомов и интересующего нас заболевания получить таблицу частот, с которыми каждый симптом наблюдался у пациентов, действительно страдающих рассматриваемым заболеванием, то такие частоты симптомов можно использовать для оценки $P(S|D)$. Однако при такой оценке возникает ряд проблем. Во-первых, исследуемая популяция должна быть однородна с той, по отношению к которой принимаются решения. Это требование часто вызывает трудности. Известно, например, что популяция больных с сердечными заболеваниями в одной клинике может резко отличаться от популяции больных с сердечными заболеваниями в другой клинике того же

города. Во-вторых, возникает особый случай, когда необходима вероятность симптома в нормальной популяции, но требуемые для этого наблюдения часто не могут быть проведены. Например, сомнительным является использование частоты болей в области грудной клетки у больных с дерматологическими заболеваниями, не страдающих сердечными заболеваниями, в качестве оценки частоты болей в области грудной клетки для нормальной популяции.

Мы уже отмечали причины, по которым врачи должны прийти к соглашению в отношении определения симптомов и заболеваний. Очевидно, различия в определениях затрудняют интерпретацию таблиц симптома — заболевания.

Своеобразная статистическая проблема при разработке таблиц симптома — заболевания возникает из-за того, что частоты некоторых симптомов и заболеваний в действительности очень малы. Если не проявить в этом вопросе осторожность, то могут иметь место абсурдные заключения, основанные на низких частотах. Во избежание таких заключений следовало бы в первую очередь начинать составление таблицы симптома — заболевания с 1, а не с 0 в каждой клетке, чтобы исключить ситуации с отношением правдоподобия, равным 0 или ∞ , как в случае, представленном на фиг. 20. С байесовской точки зрения, существуют веские причины для такого подхода. Но в случае редких событий этого оказывается недостаточно, хотя такой метод и является важным начальным этапом.

Подробный байесовский анализ этой проблемы требует рассмотрения биномиальных, бета- и бета-биномиальных распределений или их многомерных эквивалентов. Существо вопроса заключается в том, что при подготовке таблицы симптома — заболевания или таблицы частот заболеваний нужно начинать с априорного распределения, отражающего все, что вы знали, прежде чем были собраны данные. Среди многих способов реализации представления о таком априорном распределении существует один метод, который особенно хорошо подходит к задачам такого типа. Мы имеем в виду, что нужно начинать с создания воображаемой таблицы симптома — заболевания, отражающей, какое число пациентов, имеющих различные комбинации интересующих симптомов и заболеваний, пришлось бы просмотреть, чтобы вынести то априорное суждение, которое вы сейчас имеете, в предположении, что вы начинали с полного неведения. Конечно, если ваши априорные суждения были сформированы в результате действительного осмотра пациентов, то на этот вопрос ответить легче. Если же ваши априорные суждения основаны на незначительном материале, почерпнутом из медицинских учебников или из историй болезни, то построить такой воображаемый набор априорных частот может оказаться более трудным делом. Все же вам, возможно, удастся оценить, по крайней мере грубо, какой набор действительных наблюдений симптомов и заболеваний приведет вас к сходным мнениям, если бы вы начали с полного неведения.

Первоначальное использование единицы в каждой клетке матрицы симптомы — заболевания является предельным случаем применения этой идеи — предельным случаем, в котором полное неведение представлено однородным распределением по всем возможностям. Общепринятая практика начинать с нуля в каждой клетке матрицы соответствует неоднородному распределению. Это — распределение, в котором заложена высокая степень априорной уверенности в том, что какие-то данные всегда встретятся или что все люди окажутся имеющими какое-то одно заболевание или нормальными. Подобная ситуация возникает при подбрасывании монеты, если вы начинаете с сильным предубеждением о том, что у монеты либо два «орла», либо две «решки», но не знаете, какое именно из этих двух предположений верно. Особенно важно принять во внимание априорные мнения, когда у вас имеются веские априорные причины считать, что рассматриваемое событие — редкое (или что редким является противоположное ему событие). Поэтому следует иметь в виду, что таблица воображаемых наблюдений, как и таблица действительных наблюдений, должна начинаться с единицы в каждой клетке, если только вы не считаете, что этот исход невозможен.

После того как построена таблица воображаемых априорных наблюдений, вы собираете ваши действительные наблюдения и добавляете их к воображаемым априорным наблюдениям. Тогда относительные частоты, основанные на суммах действительных и воображаемых наблюдений, являются формально подходящими оценками значений $P(S|D)$, если вы составляете таблицу симптомы — заболевания, и соответственно оценками значений $P(D)$, если вы пытаетесь установить частоту или априорную вероятность заболевания.

Очевидно, чем больше отношение общего числа воображаемых наблюдений к общему числу действительных наблюдений, тем более важную роль будут играть ваши априорные суждения при получении конечного результата. При консервативном подходе указанное отношение должно быть небольшим, вероятно, не больше 0,1. Если ваши априорные суждения действительно весомы и требуют относительно большого количества воображаемых наблюдений для их представления, то единственный способ сохранить значение отношения 0,1 или меньше — это убедиться, что количество действительных наблюдений, которые вы используете, также велико. Этот статистический результат просто выражает следующий очевидный интуитивный факт: чем сильнее ваши априорные мнения, тем больше доказательств потребуется, чтобы изменить их, если они неверны.

4.10. «ПЛАВАНИЕ В НИЛЕ»

В заключение обсуждения вопросов логического анализа, по-видимому, уместным будет рассказать одну историю из медицинской практики, которая произошла в современном «реальном мире», где

врачи сталкиваются с неопределенностью редких заболеваний и неубедительными тестами. Большинство из нас признает то обстоятельство, что врачу приходится решать, что он может сделать для пациента, несмотря на эту неопределенность. Это нелегко, и врач в большинстве случаев удивительно хорошо справляется с этой задачей. Излагаемая ниже история была выбрана для того, чтобы подчеркнуть, насколько трудной может оказаться проблема решения. Мы вовсе не стремимся критиковать тех врачей, которые здесь упоминаются.

История «Плавание в Ниле» была опубликована Б. Руше в «Нью-Йоркере» (The New Yorker) 7 ноября 1964 г. Мы получили разрешение поместить здесь выдержки из этой публикации.

Плавание в Ниле

Вернон Берри, двадцати пяти лет, из Джерси-Сити, в первый раз заболел во время своего путешествия вокруг света, находясь в районе Коулун в Гонконге в конце февраля 1958 г. Он проснулся с мучительной головной болью, воспаленным горлом и ознобом. От врача в Коулуне он получил четыре укола пенициллина по поводу «острого бронхита» и, чувствуя себя лучше, отправился в Сингапур. Второй раз он почувствовал себя больным приблизительно три дня спустя по дороге из Сингапура в Бангкок, когда его ноги начали зудеть и отекать. Зуд, казалось, распространялся по всему телу; начали болеть суставы. Врачи, наблюдавшие за ним в течение нескольких последующих дней, лечили его симптоматически, полагая, что у него аллергическая реакция, вероятно, на пенициллин. Хотя он немного поправился, у него оставалась припухлость вокруг глаз и постоянный умеренный понос.

Берри продолжал путешествие автобусом, поездом и самолетом через Бирму, Восточный Пакистан и Индию в Дамаск. В самолете на пути в Дамаск развилось новое ощущение — свинцовая боль в пояснице и боках, которая распространилась по всему животу и перешла в острую дергающую боль. Врач в Дамаске нашел, что у Берри поднялась температура (выше 100° F), и обнаружил в моче следы крови. На основе этих данных он предположил почечную инфекцию.

Немногим более недели провел Берри в частной лечебнице в Дамаске, после чего, несколько поправившись, 14 апреля отправился в Каир. Из Каира он поднялся вверх по Нилу, чтобы провести три дня в Луксоре, отдыхая среди памятников древней культуры, и поплавать в Ниле. Из Египта он продолжил путешествие в Бейрут, Анкару, Афины и оттуда в Цюрих, чтобы начать шести-восьмидневное путешествие по Европе. Аллергическая припухлость вокруг глаз периодически появлялась, но в остальном он чувствовал себя обнадеживающе хорошо. Когда он был в Вене, у него случился при-

ступ более в животе, но обследование урологом и соответствующий анализ не обнаружили отклонений от нормы. Успокоенный, Берри продолжал путешествовать и приехал домой в Джерси-Сити в начале августа.

Берри остался в Джерси-Сити и через одну-две недели нашел работу чертежника в одной архитектурной фирме. Он жил с родителями, встречался со старыми друзьями и вел обычный образ жизни. Аллергическая припухлость вокруг глаз сохранялась, а зуд и отеки на других частях тела продолжали то появляться, то исчезать. К тому же в начале сентября он начал чувствовать неясные перемежающиеся приступы тошноты. Это чувство особенно сильно проявлялось непосредственно перед едой. Он говорил себе, что эти явления скоро пройдут, а когда они не проходили, он старался не обращать на них внимания. В конце недели он сдался и вызвал для осмотра домашнего врача. Врач внимательно осмотрел его и вынужден был признаться, что находится в тупике. Он сказал, что во всех отношениях, кроме одного, Берри был, как обычно, совершенно здоров. Исключение составляла формула крови. Было обнаружено патологическое увеличение одного из видов белых клеток крови — эозинофилов. Это подсказывало предположение о наличии инфекции. Возможности были слишком многочисленны, чтобы о них говорить, но консультация специалиста могла бы помочь разобраться в сложившейся ситуации. Доктор предложил Берри обратиться к знакомому терапевту. Несколько дней спустя, после второго тщательного осмотра, Берри опять столкнулся с неясным результатом. Не считая заметно высокого количества эозинофилов, все тесты и исследования, проведенные терапевтом, дали отрицательные результаты. Терапевт тем не менее не был полностью дезориентирован. У него создалось, как он сказал, определенное впечатление, или «подозрение». Эозинофилия, напомнил он домашнему врачу, характерна в случае трихинеллеза. То же можно сказать и о припухлости аллергического типа вокруг глаз, и об уртикарных расчесах. Не могла ли причиной жалоб пациента быть слабая инфицированность трихинеллами? Домашний врач передал этот очень тонкий диагноз Берри, добавив, что он в общем склонен согласиться с ним.

Не существует специфичного лечения трихинеллеза, и в подавляющем большинстве случаев время оказывается лучшим лекарством. Целебные силы времени ничего не сделали для Берри. Хотя припухлость вокруг глаз спала и периодическая сыпь и расчесы в значительной степени перестали возвращаться, тянущая тошнота во время еды и сомнительная эозинофилия не исчезли — в течение недель, месяцев, года. Часть неприятностей удалось устранить осенью 1959 г. Домашний врач Берри обнаружил, что его тошнота была вызвана кишечными паразитами *Ascaris Lumbricoides*. «Эти паразиты, — мужественно продолжал доктор, — были также возможной причиной того, что, казалось, говорило о трихинеллезе». Но это

было не все. Аскаридоз в отличие от трихинеллеза быстро реагирует на лечение. В качестве лекарства был выбран пиперазин-цитрат; обычно оказывается достаточно одной дозы. Так было и в случае Берри. В течение недели неприятные ощущения в желудке перестали его беспокоить, и он почувствовал себя почти здоровым. Единственным, что вызывало беспокойство, было количество эозинофилов. Оно упало, но все еще было ненормально высоким.

Примерно в середине ноября Берри постигло новое огорчение. Ему стало больно мочиться. Он терпел эту загадочную боль в течение нескольких нелегких дней, а затем отправился на прием к врачу. Врач выслушал его жалобы с некоторым облегчением, так как обследование, которое обнаружило воспаление предстательной железы, подтвердило его первое впечатление. У Берри был простатит. Затем врач начал лечение. Оно включало массаж простаты, назначение сульфисоксазола и обнадеживающий прогноз. Он надеялся, что к концу недели состояние Берри значительно улучшится. Но этого не случилось. Боли по-прежнему продолжались. Было проведено цистоскопическое обследование мочевыводящих путей, и результаты оказались неясными. Врач не видел причины изменять свой диагноз или лечение. Он был убежден, что периодический массаж простаты даст в конце концов желаемый эффект.

В скором времени его надежды оправдались. Спазмы в животе у Берри стали повторяться значительно реже и, наконец, в начале 1960 г. совсем прекратились. За этим последовало многообещающее изменение в его жалобах по поводу мочеиспускания. На некоторое время Берри показалось, что его беды почти кончились. Но так казалось лишь некоторое время. Облегчение, как вскоре стало очевидным, было не более чем облегчением. По временам боль возвращалась. К концу 1960 г. Берри потерял всякую надежду на излечение и свыкся с мыслью вести жизнь больного хроническим простатитом.

Состояние Берри как хронического инвалида длилось почти целый год. Оно внезапно закончилось одним декабрьским вечером 1961 г., когда он заметил кровь в моче. На следующее утро крови прибавилось и она стала темнее. Мучимый страхом, он позвонил домашнему врачу. Врач велел Берри немедленно явиться к нему и затем направил его к урологу. Уролог задал несколько острых вопросов и отвел в операционную для нового цистоскопического обследования. В девять часов утра на следующий день в воскресенье, 17 декабря, Берри положили в Мемориальный госпиталь в Манхэттене с диагнозом: «подозрение на рак мочевого пузыря».

Берри провел в Мемориальном госпитале шесть дней. 22 декабря он был выписан оттуда и передан на попечение д-ра Гарри Моста, главы отделения профилактической медицины Нью-Йоркского университета и специалиста по тропическим болезням. Запись в его истории болезни гласила:

«Физическое обследование при поступлении: пациент мужского пола, белый, правильного телосложения, упитанный с..... минимально увеличенной и неболезненной простатой.

При микроскопическом исследовании мочи обнаружено несколько эритроцитов и несколько лейкоцитов в поле зрения. Гемоглобин 15 Г%. Гематокрит 42%. Количество лейкоцитов 5900 с 4% эозинофилов при первом разделении и 20% эозинофилов при втором. Подсчет эозинофилов показал 98 эозинофилов в куб. мм. Сахар крови находился в пределах нормы. При цитологическом исследовании мочи злокачественных клеток не обнаружено...

Были проведены цистоскопия и биопсия. Мочевой пузырь был нормальной емкости и формы. По всей слизистой пузыря имелись многочисленные маленькие, почти точечные возвышения без каких-либо видимых изменений нормального лимонно-желтого цвета слизистой. Эти образования, по-видимому, были локализованы в подслизистом слое. В верхней части задней стенки пузыря имелись неправильные, частично сливающиеся красноватые приподнятые области, поверхность которых по своему виду представляла нечто среднее между сосочковой опухолью и буллезным отеком, но ни для одного из этих образований она не являлась характерной. Эти области занимали общую площадь в несколько квадратных сантиметров и были полностью удалены трансуретральным способом.

Патоморфологическое исследование обнаружило несомненные доказательства шистосоматоза».

«Я очень рад, что Берри передали мне», — сказал д-р Мост. «Ничто не могло бы заинтересовать меня больше, чем этот случай. Я думал, что смогу помочь ему, и мне не терпелось попробовать. Это, конечно, было первое мое желание, но не единственное. Мне не терпелось поговорить с ним — узнать все, что возможно, об этом случае, ибо я обладал достаточным опытом, чтобы понимать, что мы имеем дело с важным фактом. Это, казалось, подтверждало развитие новой и потенциально довольно зловещей тенденции. Попробую объяснить, что я имею в виду. Шистосоматоз всегда являлся наиболее редким из редких заболеваний. Это заболевание в нашей стране не распространено и не может иметь места из-за отсутствия улиток, являющихся переносчиками инфекции. Те немногие случаи, которые наблюдались в Соединенных Штатах, были «ввезены» извне. Много американских солдат вернулись домой с Филиппин после второй мировой войны с шистосоматозом японским, и определенное количество иммигрантов из Пуэрто-Рико с шистосоматозом Менсона обнаруживали в восточных приморских городах каждый год, вот и все. Так дело обстояло до недавнего времени. Теперь оно, по-видимому, принимает новый оборот. Появился новый вид жертвы. Это не солдат, находившийся в эндемической области, и не местный житель, подвергавшийся опасности заражения с детства, Это турист — человек, просто проезжающий места, где возможна инфекция.

Берри был четвертым из тех, кого мне пришлось наблюдать за период времени меньше двух лет. Остальные были: врач, его жена и их друг — холостяк, которые предприняли трехнедельный карибский круиз зимой 1960 г. Они подверглись заражению — это могло случиться только таким образом — в пруду с проточной водой на Св. Лючия, одном из Виндвардских островов. Первым заболел врач. В течение нескольких дней он думал, что у него грипп. Затем он был госпитализирован, и им занялся его коллега. Большое количество эозинофилов указывало на паразитарную инфекцию, и лечащий врач послал список клинических данных работникам здравоохранения различных островов, где побывали больные, с просьбой высказать свои диагностические предположения. Ответы, которые он получил, более всего склонялись в пользу тифозной лихорадки, анкилостомоза, аскаридоза, амёбной дизентерии и стронгилодоза. Возможность шистосоматоза никогда не рассматривалась, пока я не занялся этим вопросом. Меня попросили о консультации, зная мой интерес к тропической медицине. Шистосоматоз был подсказан мне внимательным обзором маршрута, а серии тестов подтвердили мое подозрение. Биопсия прямой кишки указывала на такую разновидность, как шистосоматоз Менсона. Это было как раз спустя две недели после начала заболевания. Определенные обстоятельства — классические клинические проявления, недавний визит в эндемическую область в анамнезе и первоклассное госпитальное оборудование — все это позволяло считать, что заболевание можно было бы диагностировать несколько раньше. Два других случая были сначала весьма легкими, симптомы не имели особенно яркого выражения. Если бы не было очевидной связи с больным врачом, они бы прошли незамеченными неопределенно долго. Так это и было в случае с Берри. И все же ключевые моменты — посещение эндемической области и эозинофилия — были налицо и в этом случае. Все, что было необходимо, — это лишь осознание самой возможности такого заболевания.

Беда в том, что возможность шистосоматоза просто не приходит в голову американскому врачу. Это вполне понятно, или, вернее, таково было обычное положение вещей. По моему мнению, теперь дело обстоит уже не так. И Берри, и врача, и его жену, и их друга можно рассматривать просто как изолированные случаи, но я так не думаю. Я считаю, что скорее всего есть и другие случаи, о которых мы просто не слышали. И вот почему. Американский турист изменился за последние несколько лет. Существует целое новое поколение. Наши туристы теперь в основной своей массе — это не только богатые или обеспеченные люди из нескольких больших городов. Они теперь представляют все слои населения и едут из разных мест страны. Причиной, конечно, являются более дешевые и быстрые воздушные путешествия. Теперь больше туристов, и они больше путешествуют и посещают более отдаленные районы, включая и

тропики. И с каждым годом все увеличивается возможность того, что некоторые из них подвергнутся заражению серьезным тропическим заболеванием. Шистосоматоз — это только одно такое заболевание, и еще далеко не самое худшее. Я не говорю о чуме, холере и тому подобных вещах. Я имею в виду такие редкие заболевания, как кала-азар и африканский трипаносомоз, сонная болезнь и злокачественная трехдневная малярия. Злокачественная трехдневная малярия является очень серьезным заболеванием с высокой смертностью, да и другие заболевания могут иметь фатальный исход, но у них есть свои обнадеживающие особенности. Они все поддаются лечению. Их можно успешно лечить и излечивать. Они могут быть излечимы, если их вовремя диагностировать. Но сначала нужно хотя бы заподозрить их возможность. О них нужно помнить. Б. Магрейт из Ливерпульского отдела тропической медицины предложил важное в современных условиях дополнение к обычному опросу пациента. Он предлагает, чтобы в некоторый момент беседы доктор спрашивал: «Где вы бывали?» Я поддерживаю эту идею.

Магрейт считает, что врач услышит ответ и обратит на него внимание. И это самое важное. Такой подход мог бы многое изменить в случае с Берри. Я понимаю, что там были усложняющие факторы. Ранние аллергические проявления шистосоматоза были стерты еще более ранней реакцией на пенициллин, и, конечно, наличие кишечного паразита было наиболее несчастливой случайностью. Тем не менее, поскольку врачи, которые наблюдали больного, знали с самого начала, что Берри только что вернулся из длительного путешествия по одной из стран, наименее удовлетворительных в смысле гигиенических условий, трудно сказать, что его случай рассматривали достаточно внимательно. Во всяком случае, так мне показалось после бесед с Берри в процессе лечения.

Я впервые увидел Берри приблизительно спустя месяц после того, как он выписался из Мемориального госпиталя. Было бы лучше, если бы это случилось раньше, но тут уж ничем нельзя было помочь. Я хотел быть уверенным в том, что именно я должен лечить, прежде чем приступать к лечению. Лечение шистосоматоза зависит в некоторой степени от типа (или типов) заболевания. Я знал, что Берри совершил кругосветное путешествие, поэтому не была исключена возможность, что у него более чем один тип паразита. Сначала нужно было посмотреть анализы, по которым был поставлен клинический диагноз, но доставка их потребовала некоторого времени. Затем хотелось исследовать некоторые пробы стула и мочи на наличие микроскопических яиц, которые помогают идентифицировать тип (или типы) паразитов, а это опять означало отсрочку. В конце концов начали лечение 29 января. К этому времени я уже получил результаты исследований проб, и они не содержали неприятных сюрпризов. Они просто подтвердили первоначальные данные, полученные при биопсии мочевого пузыря. У Берри был моче-

половой шистосоматоз, и только один этот тип заболевания. Где именно он получил его, довольно трудно сказать. Он посетил дюжину различных мест в эндемических областях Египта и Среднего Востока и в большинстве этих мест он мылся в ваннах, наполняемых вручную водой из какого-нибудь, возможно, зараженного источника. Правда, у меня есть подозрение, что он подхватил болезнь в Египте — в Луксоре, когда плывал в Ниле.

Лечение шистосоматоза включает применение какого-либо препарата сурьмы. Виннокислый антимонил-калий (рвотный камень), наиболее сильнодействующий переносимый препарат, обычно используется для лечения японского шистосоматоза. Вводится он внутривенно, и эта процедура является исключительно тонкой. Оплешность может вызвать настоящую беду. Более того, наблюдались серьезные и даже фатальные токсические реакции на лекарство. Менее сильный препарат, содержащий сурьянонатриевую соль и называемый стибофеном, эффективен при мочеполовом шистосоматозе и при шистосоматозе Менсока, и его можно вводить внутримышечно».

Итак, 29 января Берри получил инъекцию 5 куб. см стибофена, за которой последовали еще шестнадцать инъекций через регулярные промежутки времени, пока общая доза не достигла 85 куб. см. 22 февраля, в день семнадцатой инъекции, у него был приступ тошноты, его вырвало, и поэтому инъекции были прекращены. Последующая серия исследований показала значительно более прозрачную мочу и устойчивое снижение количества эозинофилов. Новая цистоскопия была проведена 22 августа и не показала наличия острого воспаления.

Семь месяцев спустя, 13 марта 1963 г., была проведена последняя цистоскопия, и на этот раз можно было утверждать, что Берри был излечен окончательно.

ПРЕДПОЛОЖИТЕЛЬНОСТЬ И УВЕРЕННОСТЬ — НЕКОТОРЫЕ ДАЛЬНЕЙШИЕ ИССЛЕДОВАНИЯ

Разнообразие является законом жизни, и так же, как никакие два лица не бывают абсолютно похожи, никакие два организма не могут быть совершенно одинаковыми и никакие два индивидуума не реагируют одинаково и не ведут себя одинаково при необычных условиях, которые мы называем заболеванием. В этом и состоит основная трудность при обучении врача.

Сэр Вильям Ослер¹⁾

5.1. ВВЕДЕНИЕ

После постановки диагноза врач должен принять решение о лечении. Принятие решения в медицине служит основой для действия. Врач должен решить, как лечить пациента; может быть, он вообще не нуждается в лечении. Но отказать пациенту в лечении особенно трудно, поскольку пациент обычно ожидает, что врач должен «что-то сделать». Часто выбор лечения бывает относительно простым делом — прямым применением общепринятых и доступных терапевтических мер, связанных с определенным диагнозом.

По-видимому, столь же часто выбор лечения включает рассмотрение и оценку сложной конфликтной ситуации, которая зависит не только от установленного диагноза, но также и от моральных, этических, социальных, терапевтических и экономических аспектов, касающихся данного пациента, его семьи и общества, в котором он живет. Решение о выборе лечения должно основываться на рассмотрении дополнительных данных и величин, хотя, как мы показали, оно является интегральной частью общего динамического процесса диагноз — лечение (фиг. 29, гл. 3). Решение о лечении для пациента с мочеполовым шистосоматозом, которое было описано в истории «Плавание в Ниле», иллюстрирует это положение. У врача был набор соединений сурьмы: рвотный камень — наиболее сильный, но еще переносимый препарат (обычно вводимый внутривенно), и сурьмянонатриевая соль — менее сильное лекарство, которое может быть введено внутримышечно. Рвотный камень, как известно, иног-

¹⁾ On the educational value of the medical society, in: *Aequanimitas with Other Addresses*, 2nd. ed., Philadelphia, Blakiston, 1930, p. 331.

да вызывает серьезные токсические реакции; поэтому врач выбрал менее токсичную и менее сильную сурьмянонатриевую соль.

Логический анализ, связанный с решением о лечении, рассматривается Ледли и Ластедом в работах [53—55] как «второй комплексный процесс выработки суждения», осуществляемый после решения задачи диагноза. Кроме того, мы утверждали, что полезно отделить проблему стратегии лечения от проблемы оценки значимости (критериев), потому что проблема стратегии может быть исследована как вероятностная задача, тогда как проблема выбора оценок значимости включает такие «неуловимые», или субъективные, понятия, как моральные и этические нормы, которые должны быть (по крайней мере в настоящее время) предоставлены суждению врача. Мы считаем, что решение о выборе лечения можно рассматривать как одну из следующих трех проблем: 1) решение о лечении при определенности, 2) решение о лечении при риске и 3) решение о лечении при неопределенности. В задачах первого типа может быть использовано линейное программирование; к задачам второго и третьего типа может быть применена концепция ожидаемой величины. В работе [55] динамически повторяющийся процесс диагноз — лечение был исследован методом динамического программирования. Мы не уверены в том, что какие-либо из этих предложений были проверены на практике.

После опубликования наших работ мы получили ряд замечаний врачей, психологов, математиков и других специалистов, которые выдвигали доводы за и против отделения диагностического решения от решения о лечении. Один из авторов [97] заметил: «Можно было бы возразить, что лучше разработать метод для систематического определения абсолютных показаний относительно лечения или отсутствия лечения, каким бы грубым он ни был, чем целиком отдать эту проблему на усмотрение врача с его изящными, но все же непонятными процессами суждения... Возможно, трудность, встретившаяся Ледли и Ластеду, заключается не в том, что оцениваемые факторы «неуловимы», а в том, что они, по-видимому, выражены в несоизмеримых единицах».

На основе исследований, в которых выявляется довольно слабая связь между диагностическим решением и выбором лечения, можно было бы изучить следующие вопросы: использует ли врач для диагноза набор критериев и оценок значимости, которые существенно отличаются от критериев и оценок для выбора лечения; если он действительно использует два разных набора критериев и оценок, то может ли он осуществлять корреляцию между ними? Стремится ли врач при вынесении диагностического решения только точно описать заболевание без учета влияния его решения на последующее лечение? В работе [20] изложено исследование по диагностике и лечению одиночного узла легкого, иллюстрирующее некоторые из этих положений.

5.2. ПРИМЕР РАЗНОГЛАСИЙ МЕЖДУ ВРАЧАМИ ПО ПОВОДУ ПРЕДОПЕРАЦИОННОЙ ДИАГНОСТИКИ И ПОСЛЕДУЮЩЕГО ЛЕЧЕНИЯ ОДИНОЧНОГО УЗЛА ЛЕГКОГО

Пятнадцать врачам-специалистам (5 рентгенологам, 5 торакальным хирургам и 5 терапевтам, специализирующимся по заболеваниям грудной клетки) были представлены предоперационные рентгенограммы грудной клетки и протокол, содержащий всю информацию, которую можно было получить до хирургического вмешательства, для пятидесяти пациентов. Все эти пятьдесят пациентов лечились в Леттерманском военном госпитале в Сан-Франциско, все они уже перенесли операцию (торакотомия), так что для каждого пациента имелся верифицированный диагноз. Цель исследования состояла в том, чтобы определить возможную точность предоперационного диагноза и изучить метод отбора пациентов для торакотомии.

Прежде чем перейти к обзору этого исследования, мы приведем некоторые его результаты. Затем мы рассмотрим некоторые возможные интерпретации этих результатов.

Врачи продемонстрировали свою способность предсказывать до операции, каким является данный одиночный узел легкого — доброкачественным или злокачественным. Но на вопрос о том, показана ли торакотомия, они постоянно отвечали утвердительно и советовали провести хирургическое удаление узла для дальнейшей диагностики. Эти данные показывают, что врачи не осознавали своей способности правильно диагностировать эти патологические изменения. Врачи могли ставить диагноз с большой степенью уверенности, а действовали так, как будто их диагнозы были только предположениями.

5.2.1. МАТЕРИАЛ И АНАЛИЗЫ

Основным материалом для исследования служила группа из пятидесяти двух пациентов, представленных к торакотомии для удаления одиночных легочных узлов или образований, дающих тени с ровными краями («штампованных» новообразований). Критериями для выделения «штампованных» новообразований были: 1) диаметр новообразования от 1 до 6 см, 2) округлая и яйцевидная форма с резко выраженными границами и 3) затемнение, окруженное со всех сторон легочной тканью нормального вида. В таблице на фиг. 44 представлены данные об окончательном гистологическом диагнозе, среднем возрасте и симптоматике.

Для окончательного исследования были отобраны пятьдесят пациентов (42 с доброкачественными и 8 со злокачественными ново-

Патологический диагноз	Число случаев	Общий процент	Средний возраст	Симптомы	
				да	нет
Гранулома					
Этиология не определена	16	30,8	37	0	16
Кокцидиома	9	17,3	38	0	9
Туберкулома	5	9,6	41	0	5
Гистоплазма	2	3,8	37	0	2
Хроническая очаговая пневмония	6	11,5	38	0	6
Гамартома	6	11,5	58	0	6
Злокачественные образования					
Бронхогенный рак	5	9,6	57	2	3
Рак альвеолярных клеток	3	5,8	50	1	2
Всего	52	99,9	44	3	49

Фиг. 44.

Данные по пятидесяти двум хирургическим удалениям единичных «штампованных» новообразований [20].

образованиями). Каждому из пятнадцати специалистов был выдан следующий вопросник:

1. Согласуется ли рассматриваемое новообразование с вашей интерпретацией «штампованного» новообразования?

2. Каков ваш предоперационный диагноз:

а) воспаление?

б) доброкачественная опухоль?

в) злокачественная опухоль?

г) аномалия?

3. Рекомендуете ли вы хирургическое удаление?

4. Каковы, с вашей точки зрения, показания к хирургическому удалению:

а) для диагностики?

б) для лечения локализованной опухоли?

в) для лечения локализованного воспалительного заболевания?

5. Какие исследования вы хотели бы провести в дополнение к снимкам грудной клетки и тем, что содержатся в протоколе?

Ниже кратко суммированы ответы на эти вопросы.

Вопрос 1.

Согласуется ли рассматриваемое новообразование с вашей интерпретацией «штампованного» новообразования?

Расхождение в мнениях специалистов показало, что отсутствует единое понимание термина «штампованное» новообразование. Существенно, что ни одно из злокачественных новообразований не рассматривалось как «штампованное» большинством специалистов. Следовательно, можно предположить, что большая часть «штампованных» теней имеет характеристики, отличные от характеристик доброкачественных образований, но при сравнении этих результатов с предоперационными диагнозами различия не были очевидны.

Вопрос 2

Каков ваш предоперационный диагноз?

Матрицы решения для пяти рентгенологов

Вопрос 3

Рекомендуете ли вы хирургическое удаление?

Матрицы решения для пяти рентгенологов

Категория решения

злокачественная опухоль

Истинно положительные случаи (TP)	Ложно отрицательные случаи (FN)
$\frac{30}{40} = 0,75$	$\frac{10}{40} = 0,25$
Ложно положительные случаи (FP)	Истинно отрицательные случаи (TN)
$\frac{8}{210} = 0,04$	$\frac{202}{210} = 0,96$

Категория заболевания

Категория решения

да

нет

Лечение (TR)	Лечение (FN)
$\frac{36}{40} = 0,90$	$\frac{4}{40} = 0,10$
Лечение (FP)	Лечение (TN)
$\frac{88}{210} = 0,42$	$\frac{122}{210} = 0,58$

Категория заболевания

Категория решения

злокачественная опухоль

Истинно положительные случаи (TP)	Ложно отрицательные случаи (FN)
$\frac{29}{40} = 0,72$	$\frac{11}{40} = 0,28$
Ложно положительные случаи (FP)	Истинно отрицательные случаи (TN)
$\frac{26}{210} = 0,12$	$\frac{184}{210} = 0,88$

Категория заболевания

Матрицы решения для пяти хирургов

Категория решения

да

нет

Лечение (TR)	Лечение (FN)
$\frac{32}{40} = 0,80$	$\frac{8}{40} = 0,20$
Лечение (FP)	Лечение (TN)
$\frac{153}{210} = 0,73$	$\frac{57}{210} = 0,27$

Категория заболевания

Матрицы решения для пяти терапевтов

Категория решения

злокачественная опухоль доброкачественная опухоль

Истинно положительные случаи (TP)	$\frac{26}{40} = 0,65$	Ложно отрицательные случаи (FN)	$\frac{14}{40} = 0,35$
Ложно положительные случаи (FP)	$\frac{24}{210} = 0,11$	Истинно отрицательные случаи (TN)	$\frac{186}{210} = 0,89$

Категория заболевания
 злокачественная опухоль доброкачественная опухоль

Категория решения

да нет

Лечение (TP)	$\frac{26}{40} = 0,65$	Лечение (FN)	$\frac{14}{40} = 0,35$
Лечение (FP)	$\frac{85}{210} = 0,41$	Лечение (TN)	$\frac{125}{210} = 0,59$

Категория заболевания
 злокачественная опухоль доброкачественная опухоль

Матрицы решения для большинства из 15 специалистов, пришедшего к единодушному мнению

Категория решения

злокачественная опухоль доброкачественная опухоль

Истинно положительные случаи (TP)	$\frac{85}{120} = 0,71$	Ложно отрицательные случаи (FN)	$\frac{35}{120} = 0,29$
Ложно положительные случаи (FP)	$\frac{58}{630} = 0,09$	Истинно отрицательные случаи (TN)	$\frac{572}{630} = 0,91$

Категория заболевания
 злокачественная опухоль доброкачественная опухоль

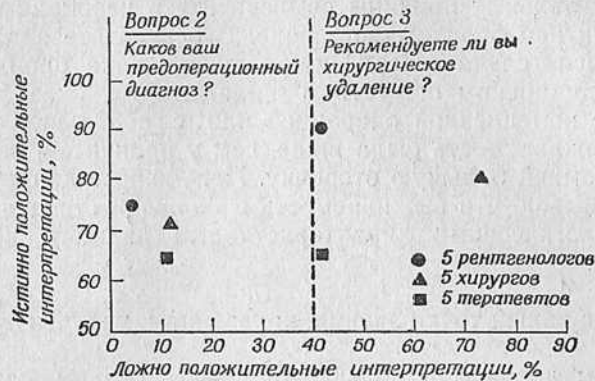
Категория решения

да нет

Лечение (TP)	$\frac{94}{120} = 0,78$	Лечение (FN)	$\frac{26}{120} = 0,22$
Лечение (FP)	$\frac{326}{630} = 0,52$	Лечение (TN)	$\frac{304}{630} = 0,48$

Категория заболевания
 злокачественная опухоль доброкачественная опухоль

Изучение представленных матриц решения выявляет разногласия между врачами в отношении правил решения. На основании этих матриц нельзя построить рабочие характеристические кривые.



Фиг. 45.

Рабочие характеристические точки для диагностики и лечения при исследовании единичных узлов в легком. По этим точкам невозможно построить кривые ROC, имеющие смысл. Это означает, что специалисты не пользовались одними и теми же правилами решения.

(фиг. 42, гл. 4), имеющие какой-либо смысл. Рентгенологи, хирурги и терапевты, очевидно, имели разные мнения по поводу правил решения в этом исследовании, тогда как в случае интерпретации фотофлюорограмм грудной клетки, приведенном в гл. 4, правила решения рентгенологов и терапевтов согласовались друг с другом.

Сравнение матрицы диагностического решения с матрицей выбора лечения для каждой из групп специалистов показывает, что, хотя точность диагноза для истинно отрицательных случаев (доброкачественные новообразования) была достаточно высокой, тем не менее у врачей не было столь же большой уверенности в суждении о лечении, и они советовали применить хирургическое вмешательство для многих доброкачественных новообразований. Большинство рентгенологов правильно выделило всю группу из сорока двух пациентов с доброкачественными новообразованиями, и все же они рекомендовали торакотомию шестнадцати из этих пациентов. Отбор тридцати двух из сорока двух пациентов с доброкачественными новообразованиями для торакотомии, сделанный хирургами, представляется на редкость высоким в процентном отношении, поскольку лишь трое из этих пациентов были диагностированы большинством хирургов как имеющие признаки злокачественности. То,

что хирурги отобрали большую часть пациентов для торакотомии, вероятно, связано с преобладающей в хирургии доктриной ранней резекции всех потенциально злокачественных новообразований. По крайней мере одно недавнее исследование [128] показало, что существуют убедительные причины сомневаться в необходимости срочной резекции при бронхогенном раке. В результате анализа более полутора тысяч случаев верифицированного рака бронхов, проведенного Медицинским обществом Филадельфии, оказалось, что у пациентов с немедленной операцией или с небольшой отсрочкой операции выживаемость была ниже, чем у пациентов, имевших перед торакотомией большую отсрочку. Полученные данные позволили сделать вывод, что выживаемость в большой степени предопределена биологическими характеристиками каждого конкретного случая рака.

Вопрос 4.

Каковы, с вашей точки зрения, показания к хирургическому удалению?

Основным доводом в пользу хирургического вмешательства в случае новообразований, которые позднее оказались воспалительными (туберкулома, гистоплазма, кокцидиома, хроническая очаговая пневмония), являлось стремление уточнить диагноз («для диагностики»), причем этот мотив значительно преобладал над желанием хирургического вмешательства «для лечения локализованной опухоли». Среди случаев, для которых впоследствии были доказаны злокачественные новообразования, рекомендация удаления «для лечения локализованной опухоли» преобладает над рекомендацией удаления «для диагностики» приблизительно в четыре раза, и эти случаи были правильно диагностированы до операции как злокачественные в таком же процентном отношении.

Эдвардс, Кокс и Гарланд пришли к выводу, что показание «для лечения локализованного воспалительного заболевания» не является веской причиной для торакотомии. Эти исследователи отметили также, что точка зрения многих специалистов в этой области, согласно которой точный предоперационный диагноз невозможен, является неверной. Следовательно, неубедительна доктрина, требующая операции для всех таких пациентов в целях диагностики.

Вопрос 5.

Какие диагностические исследования вы хотели бы провести в дополнение к снимкам грудной клетки и тем, что содержатся в протоколе?

Ответы были слишком обширными, и их трудно представить в виде таблиц. Большинство протоколов считалось адекватными для точной оценки. В нескольких случаях исследователи запрашивали такую дополнительную информацию, как ламинограммы, тест кожи, бронхоскопическое исследование, цитологические исследования мокроты и биопсия лимфоузлов в области *m. scalenus*.

На основе полученных данных Эдвардс, Кокс и Гарланд рекомендовали, чтобы пациенты, у которых обнаружен одиночный узел легкого, или «штампованная» тень, изучались терапевтично-хирургическо-рентгенологической конференцией специалистов по грудной клетке с участием двух специалистов в каждой специальной области (лучше было бы использовать группу из 15 специалистов, но это практически невозможно). Если пятеро из шести (более двух третей) врачей согласны, что новообразование является незлокачественным, то диагностическая торакотомия должна быть отложена и должно быть установлено внимательное наблюдение за пациентом. При этом можно ожидать правильный диагноз в 99% случаев злокачественных и в 90% случаев доброкачественных новообразований, причем большинство ошибок будет иметь место в относительно небольшой группе больных с гамартомой. Представляется оправданной отсрочка хирургического вмешательства, если четыре из шести (две трети или больше) специалистов согласны, что новообразование является незлокачественным. Важно, что такие терапевтично-хирургическо-рентгенологические конференции по грудной клетке будут основывать свои мнения и решения в большей степени на диагнозах относительно злокачественности или незлокачественности, чем на индивидуальных решениях за или против хирургического вмешательства.

5.2.2. КОММЕНТАРИИ ПО ПОВОДУ ИССЛЕДОВАНИЯ НОВООБРАЗОВАНИЙ ТИПА «ШТАМПОВАННЫХ» ТЕНЕЙ

Мы выбрали для обсуждения исследование Эдвардса, Кокса и Гарланда, поскольку оно помогает сосредоточить внимание на нескольких важных вопросах, которые уже рассматривались в этой книге. Диагностические критерии во всех областях медицины должны подвергаться проверке, подобно тому как критерии для новообразований типа «штампованных» теней. Очень часто о новом диагностическом наблюдении или диагностическом тесте сообщают без предварительного исследования его полезности, не говоря уже об использовании для этой цели метода проверки гипотез.

Диагностические критерии и соглашение между врачами по поводу диагностических критериев — это важные факторы в медицинской практике. Исследование «штампованных» теней иллюстрирует трудности, которые могут возникнуть при рассмотрении специфических диагностических критериев. Однако существует расхождение мнений относительно критериев и тогда, когда применяются методы проверки гипотез. Матрицы решения и рабочие характеристические кривые являются полезным средством для обнаружения разногласий. В предыдущих главах мы подчеркнули важность внимательного отбора диагностических критериев и необходимость соглашения по поводу критериев. Усилия, затраченные на проверку имеющихся

и поиски новых методов отбора диагностических критериев, полностью окупятся.

Исследование «штампованных» теней иллюстрирует тот факт, что врачи могут оказаться способными осуществлять диагностику с высокой точностью, но они могут не осознавать эту способность и в результате этого не проявить соответствующую высокую точность в лечении. Правда, без проведения дальнейших исследований любые обобщения в этом вопросе не имеют смысла. Почти не вызывает сомнений, что врачи будут принимать более последовательные диагностические решения и решения о выборе лечения, если они будут знать, какую степень точности можно ожидать от их решений. Групповое мнение врачей использовалось именно как один из методов уравнивания неуверенности каждого отдельного врача.

Предположим, что благоприятное стечение обстоятельств дает нам возможность реорганизовать работу тех пятнадцати специалистов, которые принимали участие в исследовании «штампованных» теней; попросим их продолжить работу с определенными изменениями в процедуре.

Во-первых, пусть специалисты пересмотрят диагностические критерии для «штампованных» теней. Можем ли мы найти критерии, которые отделяли бы доброкачественные от злокачественных новообразований более эффективно, чем использованные ранее критерии? Здесь мы проверяем гипотезы «доброкачественное новообразование» и «злокачественное новообразование» и хотим найти критерии, которые дадут относительно большое разделение между распределениями плотностей вероятности гипотез. Это эквивалентно отысканию большого значения d' , поскольку вероятность правильного обнаружения (правильный диагноз) повышается с увеличением значения d' . Некоторые из специалистов могут предложить другой набор критериев, и группа диагностов может согласиться с ними после обсуждения. Далее, необходимо осуществить повторную проверку гипотез с этими новыми критериями. Нужно построить матрицы решения и вычертить рабочие характеристические кривые, чтобы убедиться, что наклон кривой монотонно убывает. Это убеждает нас в том, что специалисты используют верные диагностические правила. Может оказаться необходимым полностью повторить процедуру несколько раз.

Цель описанной процедуры заключается в том, чтобы помочь врачу быть более последовательным в диагностике и лечении. Это достигается при помощи процесса, который позволяет врачу пересматривать свои решения на основе новой информации и отделять вопросы, связанные с вероятностью диагноза и лечения («уловимые» факторы), от оценок значимости («неуловимые» факторы). Захочет он пересмотреть свои суждения о диагнозе и лечении в свете полученной новой информации или нет, — это вопрос отношения врача к проблеме.

Наша задача состоит в разработке методологии, помогающей врачу разобраться в своей непоследовательности и проанализировать сложные проблемы решения в задаче диагноз — лечение. Когда люди делают то, что им кажется естественным, они, по-видимому, оказываются весьма непоследовательными.

Как было показано в предыдущих главах, исследования по разработке средств врачебного решения в сложной ситуации диагноз — лечение все еще находятся в начальной стадии. Трудные проблемы решения возникают также в экономических, политических и военных ситуациях. В этих областях уже разработаны некоторые методы стратегии, помогающие более аналитически подходить к решению возникающих задач. Некоторые из этих методов с определенной модификацией можно применить к принятию решения в медицине. Имея это в виду, мы в последующих параграфах этой главы кратко изложим общие идеи моделирования, теории полезности и некоторые проекты, связанные с дальнейшим изучением рассматриваемой проблемы.

5.3. МОДЕЛИРОВАНИЕ И ТЕОРИЯ ПОЛЕЗНОСТИ

Трудность анализа сложных проблем объясняется тем, что в них участвует так много различных факторов и переменных, что человеческое мышление с его ограниченными возможностями обработки информации не в состоянии уследить за всеми переменными и всеми возможными связями между ними. Поэтому полезно сконструировать модель рассматриваемой сложной проблемы, ибо модель, хотя она является неполным описанием, все же помогает человеку, решающему проблему или принимающему решение, иметь более определенное представление об этой проблеме. Моделирование заставляет человека, решающего проблему, точно установить, что именно он знает, и это помогает ему увидеть, что он может определить (назовем это конкретной частью задачи) и что он не может определить (назовем это субъективной частью задачи). Модель может представлять собой математическое выражение, такое, как теорема Байеса, или диаграмму, такую, как дерево решений, или словесное описание образа, такое, как «мелкоочаговый инфильтрат в правой верхней легочной доле» (по изображению на рентгеновском снимке грудной клетки); наконец, образ или рисунок сами по себе тоже являются моделью объекта.

В работе [55], которая была опубликована несколько лет назад, при обсуждении медицинской диагностики и современных методов принятия решения были предложены модели для трех ситуаций при принятии решения о лечении: 1) решение о лечении при определенности, 2) решение о лечении при риске и 3) решение о лечении при неопределенности. Эти модели соответствуют теории игр и решений, развитой в работе [65].

5.3.1. РЕШЕНИЕ О ЛЕЧЕНИИ ПРИ ОПРЕДЕЛЕННОСТИ ¹⁾

Под «решением о лечении при определенности» мы понимаем выбор лечения, когда точно известно, какой комплекс заболеваний имеется у пациента, и когда прогнозы для альтернативных лечений также известны. В простейшем виде мы можем это записать так:

	D_j
T_1	EV_{1j}
T_2	EV_{2j}

где D_j — комплекс заболеваний, который имеется у пациента, T_1 и T_2 — альтернативные лечения и EV_{1j} и EV_{2j} — соответствующие оценки значимости этих видов лечения, связанные с прогнозом. Обычно выбирается то лечение, которому соответствует наибольшая величина оценки значимости. В более сложном случае проблема решения при определенности может включать *линейное программирование*. Например, предположим, что в нашем распоряжении два возможных метода лечения пациента, больного раком, — лекарственные средства и лучевая терапия — и что в величинах каких-то общих единиц эффективности (например, связанных с действием на животное определенного веса) лекарство содержит 1000 единиц на унцию, а облучение дает 1000 единиц в минуту. Предположим далее, что пациент должен получить для выздоровления по крайней мере 3000 единиц эффективности. Однако оба лечения токсичны для больного, причем (опять в каких-то общих единицах токсичности) лекарство содержит 400 единиц на унцию, тогда как облучение дает 1000 единиц в минуту. Предположим, что пациент не должен получить больше 2000 единиц токсичности. Наконец, в величинах дискомфорта, испытываемого пациентом, предположим, что лекарство в три раза хуже на унцию, чем облучение в минуту. Задача состоит в определении доз, которые удовлетворяют критериям эффективности излечения и токсичности и минимизируют дискомфорт. Если X_1 — количество унций лекарства, которое нужно прописать, и X_2 — количество минут облучения, которое больной должен получить, то, для того чтобы вылечить пациента, мы должны обеспечить условие

$$1000X_1 + 1000X_2 \geq 3000.$$

Но, принимая во внимание токсичность, необходимо выполнить условие

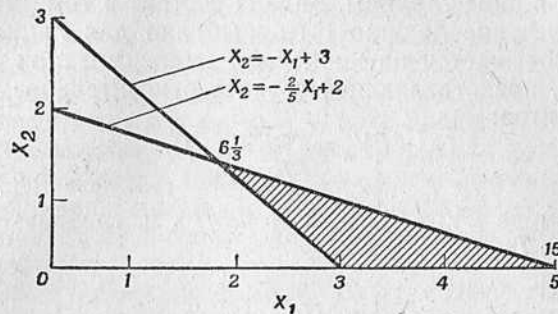
$$400X_1 + 1000X_2 \leq 2000.$$

¹⁾ Выдержки из работы Ледли и Ластеда [55].

Задача состоит в том, чтобы определить значения X_1 и X_2 , которые удовлетворяют этим неравенствам и минимизируют величину

$$D = 3X_1 + X_2.$$

Поскольку, кроме того, $X_1 \geq 0$ и $X_2 \geq 0$, решение находится в области, указанной на фиг. 46 штриховкой. Ввиду того что решение должно совпадать с одной из трех вершин выделенного треугольника, мы проверяем значения $3X_1 + X_2$ в каждой из трех точек $(3, 0)$,



Фиг. 46.

Графическое решение задачи выбора между лекарственным лечением и облучением [55].

$(5, 0)$ и $5/3, 4/3$; соответствующие значения D равны 9, 15 и 6,3. Таким образом, показано комбинированное лечение, состоящее из приема $5/3$ унций лекарства и $4/3$ минуты облучения.

Задача, которую мы только что решали, конечно, нереалистична, но лишь из-за упрощений, присущих любому иллюстративному примеру. Другие, более близкие к действительности задачи, включают множество факторов и неравенств. Мы хотим подчеркнуть, что в реальных ситуациях, даже в случае принятия решения при определенности, возникает много важных задач максимизации и минимизации, с которыми можно справиться при помощи линейного программирования.

5.3.2. РЕШЕНИЕ О ЛЕЧЕНИИ ПРИ РИСКЕ ¹⁾

Под «решением о лечении при риске» мы подразумеваем выбор метода лечения, когда альтернативные диагнозы были поставлены не абсолютно точно, а с известными вероятностями. Предположим, что в такой ситуации известно значение EV_{hj} для каждого альтернативного лечения T_h при каждом из возможных диагнозов D_j .

¹⁾ См. примечание на стр. 198.

Например, предположим, что пациенту были поставлены два возможных диагноза D_2 и D_3 с соответствующими вероятностями $P_2 = 5/7$ и $P_3 = 2/7$. Допустим далее, что существует лечение T_1 , которое эффективно в 90% случаев при комплексе заболеваний D_2 (т. е. $EV_{12} = 90/100$) и эффективно в 30% случаев при комплексе заболеваний D_3 (т. е. $EV_{13} = 30/100$). С другой стороны, пусть существует альтернативное лечение T_2 , которое эффективно лишь в 10% случаев при D_2 , но зато в 100% случаев при D_3 (т. е. $EV_{22} = 10/100$ и $EV_{23} = 100/100$). Такая ситуация представлена на фиг. 47 в виде таблицы. Задача состоит в том, чтобы определить, какое лечение должно быть выбрано для пациента. Метод решения основывается на концепции максимизации ожидаемой величины EV , представляющей среднее относительное число вылеченных пациентов:

$$EV_k = \sum_j EV_{kj} \cdot P_j.$$

В нашем примере $EV_1 = 51/70$, $EV_2 = 25/70$, и, следовательно, выбирается лечение T_1 .

	D_2 $P_2 = 5/7$	D_3 $P_3 = 2/7$
T_1	$EV_{12} = 90/100$	$EV_{13} = 30/100$
T_2	$EV_{22} = 10/100$	$EV_{23} = 100/100$

Фиг. 47.

Таблица оценок описанного примера для различных комбинаций лечение — заболевание [55].

	D_2 $P_2 = 5/7$	D_3 $P_3 = 2/7$
T_1	+5	-10
T_2	-5	+8

Фиг. 48.

Таблица оценок другого примера для различных комбинаций лечение — заболевание [55].

До сих пор мы рассматривали оценку значимости (качества лечения) по отношению к данному комплексу заболеваний в виде величины, непосредственно связанной с эффективностью лечения заболевания. Однако ситуация не всегда бывает столь простой; например, некоторые виды хирургического вмешательства связаны с определенным риском. Если операция успешна, пациент будет излечен или ему будет принесена польза; если же операция окажется неудачной, пациент может умереть. Следовательно, оценку значимости, связанную с этим лечением, определить труднее. В качест-

ве иллюстрации возможности более абстрактных оценок предположим, что оценки методов лечения были выбраны в виде некоторых чисел в интервале между -10 и $+10$, как показано на фиг. 48. В результате вычислений, аналогичных предыдущим, находим $EV_1=5/7$ и $EV_2=-9/7$; следовательно, выбираем лечение T_1 .

Два положения требуют дальнейшего обсуждения. Во-первых, мы рассмотрели нашу проблему с точки зрения множества пациентов, каждый из которых страдал заболеванием D_2 или D_3 , и нашли, как нужно выбирать лечение, которое максимизирует количество излеченных пациентов. В частной практике врач обычно имеет дело с одним пациентом. Однако после некоторого размышления можно прийти к выводу, что при максимизации ожидаемого числа излеченных людей мы также максимизируем вероятность того, что каждый *отдельный* пациент будет излечен; следовательно, мы можем применять наши результаты и к случаю одного пациента.

Во-вторых, заметим, что мы не рассматривали проблему принятия решений как оценку значимости различных комбинаций лечение — заболевание в целом. Мы отделили проблему *стратегии* от проблемы *решения об оценках значимости*. Здесь рассматривается только проблема стратегии.

Следует отметить и некоторые другие обстоятельства. Мы должны быть уверены, что заболевания D_j являются взаимоисключающими и единственно возможными заболеваниями, т. е. что пациент не может иметь более одного заболевания D_j , но в то же время пациент должен обязательно иметь одно из этих заболеваний. Если некоторое D_j представляет собой все оставшиеся возможные комплексы заболеваний после того как был осуществлен логический анализ, то *эти условия автоматически выполняются*. Аналогично лечения T_k должны быть взаимоисключающими и единственно возможными, т. е. каждый раз может быть применено не более чем одно лечение, но одно из рассматриваемых лечений должно быть обязательно применено. Другими словами, нужно рассмотреть все возможные лечебные процедуры, которые можно применить, и записать их в таком виде, чтобы каждый раз можно было выбирать только одну такую процедуру.

5.3.3. РЕШЕНИЕ О ЛЕЧЕНИИ ПРИ НЕОПРЕДЕЛЕННОСТИ

В работе [55] мы предложили, чтобы при выборе лечения при неопределенности использовалась «смешанная стратегия». Однако де Финетти [15] указал, что в условиях решения о лечении при неопределенности, по-видимому, более удобно будет использовать теорию «максимизации ожидаемой полезности», чем подход, основанный на «смешанной стратегии». Теория полезности не проверялась в действительных условиях медицинской диагностики, и мы обнаружили

всего три публикации, в которых были сделаны серьезные попытки определить функции полезности, и то лишь по отношению к коммерческим проблемам [110]. Теория полезности ставит своей целью описание поведения человека в ситуациях, связанных с риском; основы этой теории были сформулированы в фундаментальной работе Неймана и Моргенштерна [123].

Полезность измеряют в произвольных единицах, называемых «единицами полезности». При помощи соответствующего опроса может быть определена связь между единицами полезности и, например, денежными единицами (долларами) для каждого индивидуума. Эта связь и определяет функцию полезности для данного индивидуума. В любом случае принятия решения, включающего элемент риска, человек выбирает ту альтернативу, которая максимизирует полезность в его понимании [33].

Относительные полезности для индивидуума являются измеримыми, а абсолютные — нет. Человек может приписать произвольные значения полезности, и при этом функция полезности определяется в соответствии со следующим правилом: если человек, принимающий решение, безразличен к выбору между двумя альтернативами, то ожидаемая полезность обеих альтернатив считается одинаковой.

Автор работы [110] делает вывод, что, хотя, согласно проведенным экспериментам, в коммерческих операциях существует тенденция отказаться от критерия ожидаемой величины, у нас нет другой подходящей основы для предложения или предсказания решений, которые включают элемент риска, если мы отвергнем теорию полезности. Читателю, интересующемуся этими проблемами, следует ознакомиться с указанной работой; он найдет в ней примеры функций полезности, полученных для решений в области деловых отношений; ни одна из них не является линейной. Пятнадцать специалистов, упомянутых при описании исследования новообразований типа «штампованных» теней, также могли бы отвергнуть критерии ожидаемой величины и теорию полезности, но эти концепции представляются подходящим отправным пунктом для обдумывания и исследования некоторых проблем, возникающих при принятии решения о лечении при неопределенности.

5.3.4. РАЗДЕЛЕНИЕ КОНКРЕТНЫХ И СУБЪЕКТИВНЫХ ФАКТОРОВ ¹⁾

Величины оценок V_{hj} могут быть получены из двух источников — субъективных суждений и знания вероятностей излечения. Статистический анализ, связанный с определением таких вероятностей, не вызывает трудностей; субъективные суждения, по определению, иг-

¹⁾ См. примечание на стр. 198.

норируют математический или статистический анализ. Таким образом, возникает проблема согласования этих источников при определении величины V_{kj} . Возможный подход к решению этой задачи изложен ниже.

В идеализированной ситуации правильное лечение должно приводить к излечению всех пациентов, страдающих данным заболеванием. Если же с данным лечением связана только определенная вероятность (значение которой меньше единицы, но больше нуля), излечения *известного* комплекса заболеваний, то должно иметь место одно из двух объяснений: либо лечение представляет собой не одно и то же в различных случаях применения (что противоречит нашему предположению), либо диагноз не определен достаточно точно. Первая ситуация может встретиться, если лечение состоит во введении лекарства, характеристики которого изменяются; этот случай мы рассматривать не будем. Вторая ситуация может встретиться, например, когда определенное заболевание вызвано двумя или большим числом видов бактерий, которые еще недостаточно хорошо различимы. Тогда лекарство может действовать на один вид бактерий, но не действовать на другой, и создается впечатление вероятностного характера излечения, тогда как ситуация в действительности отражает вероятностный характер вида бактерий, вызывающих заболевание. Эту ситуацию мы и рассмотрим. Следовательно, вероятность излечения может быть интерпретирована скорее как *вероятность, связанная с диагностическими возможностями*; при этом оценки, вытекающие из субъективных суждений, входят сами собой в расчетную схему.

Например, пусть первоначальные возможные диагнозы будут D_1 или D_2 с известными вероятностями P_1 и P_2 соответственно. Пусть вероятности излечения при воздействии лечения T_1 на D_1 и D_2 будут равны p_{11} и p_{12} соответственно. Таким образом, исходная матрица имеет вид

Вероятности	P_1	P_2
Комплексы заболеваний	D_1	D_2
T_1	$f_{11} (p_{11}, \text{суб})$	$f_{12} (p_{12}, \text{суб})$

где $f_{ji} (p_{ji}, \text{суб})$ представляет собой величину оценки, основанную и на значении p_{ji} и на субъективных суждениях. Теперь в соответствии с приведенным выше рассуждением мы введем вместо D_1 два комплекса заболеваний D_1' и D_1'' и вместо D_2 — два комплекса заболеваний D_2' и D_2'' . Тогда вероятности этих комплексов

будут соответственно равны $P_1 p_{11}$, $P_1(1-p_{11})$, $P_2 p_{12}$ и $P_2(1-p_{12})$. Таким образом, окончательная матрица будет выглядеть так:

Вероятности Комплексы заболеваний	$P_1 p_{11}$ D_1'	$P_1(1-p_{11})$ D_1''	$P_2 p_{12}$ D_2'	$P_2(1-p_{12})$ D_2''
T_1	EV'_{11}	EV''_{11}	EV'_{12}	EV''_{12}

где величины EV'_{ji} и EV''_{ji} теперь определяются чисто субъективными суждениями.

5.3.5. ОЦЕНКИ ЗНАЧИМОСТИ И ЭТИКА СИТУАЦИИ

В предыдущем обсуждении мы предлагали стратегию, которую врач мог бы использовать для более четкого осознания своего мышления при принятии решений в случаях, когда смешиваются вероятности и этические факторы. Ситуации, в которых врач может сделать четкое разделение между вероятностями и чисто субъективными оценками, встречаются действительно редко, и мы пользуемся понятием *чисто субъективного фактора* в смысле математической переменной, чтобы проиллюстрировать метод.

Важная роль оценок значимости при принятии медицинского решения вновь привлекла внимание ученых, так как биомедицинские исследования с развитием науки продолжают ставить все новые проблемы. В работе [107] этот вопрос сформулирован так: «Этические и моральные аспекты в сравнительно установившихся областях медицинской практики, по-видимому, привлекают мало внимания, но значительное внимание теперь концентрируется на новых этических проблемах, возникающих вследствие быстрого прогресса в биологии и медицине. Некоторые из таких наиболее запутанных проблем относятся к области трансплантации тканей и органов.

Дальнейшее развитие биологии и медицины вызовет в свою очередь новые этические вопросы, например этические аспекты установления разумных пределов для многократных пересадок пораженных органов и частей тела. Такие этические проблемы представляют собой нечто большее, чем просто интересную, но второстепенную тему; они влияют на основное направление дальнейших разработок».

Вопросы принятия медицинского решения при неопределенности и проблема оценок значимости тесно переплетаются между собой; Флетчер [33] пишет: «Как должна была поступить одна леди из Аризоны, когда она узнала несколько лет тому назад, что у нее, «возможно», родится дефективный ребенок из-за того, что она принимала талидомид? Она попросила суд принудить ее доктора и клинику искусственно прекратить беременность; суд отказал, к огорчению

судьи, поскольку закон запрещает все без исключения аборты, если нет специальных медицинских показаний».

Чем может руководствоваться врач для определения вероятностей и оценок значимости в такой ситуации? Что он должен сделать, когда его пациент, который умирает, рассматривается как возможный почечный донор, и группа врачей, занимающихся трансплантацией почек, требует провести предсмертные диагностические процедуры для того, чтобы взять пересаживаемую ткань в хорошем состоянии к моменту трансплантации?

В работе [107] Стикел пишет: «Все оценки значимости, которые имеют влияние на решение, обязательно совместно взвешиваются в процессе принятия решения», но, как он заметил позже, эти оценки могут носить конкретный или субъективный характер. «Конкретные оценки значимости,— говорит он,— могут быть определены количественно по денежной или какой-либо другой числовой шкале», тогда как «субъективные оценки значимости по своей сущности не могут быть выражены количественно, но о них можно судить в смысле относительного порядка важности. Порядок важности этих оценок может быть выражен формально и явно (хотя это и не имеет места в этических ситуациях) или может быть выражен в неявном виде в процессе решения. Решение обязательно включает определение различных факторов, которые взвешиваются в процессе оценки платы за решение. В процессе принятия решения человек учитывает установленный им для себя порядок важности оценок независимо от того, думает ли он когда-либо явно об этом порядке. В некоторых решениях входит смесь конкретных и субъективных оценок значимости, и такие решения могут подразумевать некоторую эквивалентность между конкретными и субъективными оценками, хотя такая эквивалентность и не устанавливается точно».

Мы согласны со Стикелом в том, что в настоящее время все оценки значимости обязательно совместно взвешиваются, но мы думаем, что достигли некоторого успеха в количественном подходе к определению конкретных оценок и в создании метода, при помощи которого они могут быть изучены отдельно от субъективных оценок. Конкретные оценки значимости — это, например, те оценки, которые влияют на отношение рентгенолога к интерпретации снимков грудной клетки. Его пессимизм или оптимизм отражены в величине отношения количества ложно положительных к количеству истинно положительных случаев при чтении снимков; этот фактор выражается отношением правдоподобия

$$Y = \frac{p(N)}{p(SN)} \frac{(\text{полезность}_{\text{истинно отрицательные}} + \text{цена}_{\text{ложно положительные}})}{(\text{полезность}_{\text{истинно положительные}} + \text{цена}_{\text{ложно отрицательные}})} .$$

Эти величины можно установить при помощи метода «подходящего опроса». Вы можете спросить врача: «Является ли для вас

полезность истинно положительного ответа такой же, как и полезность истинно отрицательного? Если нет, то как бы вы сравнили относительные значимости (в единицах полезности)?» «Сколько вы готовы заплатить (в единицах полезности) за ложно положительный случай по сравнению с ложно отрицательным случаем?» Как мы уже указывали, эти оценки частично характеризуются той рабочей точкой, которую вы выбираете на характеристической кривой ROC независимо от того, задумываетесь ли вы прямо об этом или нет.

Существует много областей медицины, где связи между полезностями (значимостями) и ценами в терминах единиц полезности или долларов могут быть определены путем «подходящего опроса». Но есть и такие области, где субъективные оценки, влияющие на принятие медицинского решения, не могут быть получены методом опроса. При этом возникает ряд вопросов: Какие оценки значимости конкретны, а какие субъективны? Как выразить субъективные оценки? Ответить на них можно лишь так: «Это все зависит от...» Это все зависит от ситуации, от индивидуального пациента и от *итогового результата* в смысле прав пациента и ответственности его врача. В работе [108] указано, что следует подходить к концепции большей пользы, всегда имея в виду три принципа: мы должны защищать жизнь; здоровье лучше, чем болезнь; мы должны обращаться с пациентами, как с людьми, а не просто как со средними статистическими случаями из медицинской практики. Врач должен принимать решения для каждого пациента на основе имеющихся данных. Ему можно рекомендовать руководствоваться ситуационным подходом к этим вопросам в том виде, как он был определен Флетчером [33] и затем рассмотрен Стикелом [107] применительно к проблеме трансплантации органов.

Чувства, которые руководили нами при написании этой книги, в некоторых отношениях находятся в согласии с чувствами, возникающими при чтении книги Флетчера [33]. Несколько выдержек из упомянутой книги иллюстрируют то, что представляется нам общим для этих исследований:

1. С развитием вычислительной техники возникает большое количество различных аналитических возможностей в области этики. Здесь, правда, мало интересного для юридической науки, но зато многое можно получить в отношении проблемы этики ситуаций.

2. Возможно, что если мы будем уметь приписывать числовые оценки «опасным» факторам в проблемах совести, то вычислительные методы, связанные даже с такими интимными вещами, как вопросы любви, смогут придать точность этой «старейшей области» этики. Характерным для ситуационной этики является стремление выразить качества в количественных терминах.

3. Этот контекстуальный ситуационный метод, предусматривающий обращение к клиническим случаям, этот способ принимать решения включает слишком много различных переменных, чтобы понравиться некоторым людям. Они предпочитают уцепиться за несколько хорошо известных констант, одобренных законом, и игнорировать все остальные факторы.

4. Теория неявно выраженных оценок делает позицию ее последователей несостоятельной, заставляя давать абсолютные запрещения определенным действиям вне зависимости от обстоятельств.

Замечание

Если перенести мысли, содержащиеся в предыдущей цитате, в сферу проблем принятия решений в области медицины, то они позволяют хорошо выразить главную идею этой книги. Принятие медицинского решения — процесс не детерминистский, а вероятностный и связанный с множеством переменных. Он ориентируется на пациента и имеет дело с проблемами пациента как индивидуума, используя субъективно-вероятностный подход. В нынешних условиях врачу все труднее и труднее защищать позицию, утверждающую, что диагноз верен «по определению», т. е. «диагноз верен, потому что я так говорю», и поэтому он поддерживает исследование процесса принятия медицинского решения; это исследование позволяет заглянуть вглубь все усложняющихся диагностических ситуаций, включающих в себя и вероятности, и оценки значимости.

Но описанный подход к процессу принятия решения может возбудить тревогу среди многих врачей, как отметил Гарланд в работе [35]. Он пишет: «Когда мы преподносим эти факты стажерам, мы сначала встречаемся с некоторым недоверием. После ознакомления с ограничениями в наблюдениях и в других областях некоторые стажеры совсем расстраиваются, повторяя слова одного из студентов Джонсона [43, 44], сказавшего: «Но весь мир не может же быть студнем», или «Похоже на то, как будто мой мир совсем развалился». Им не хочется думать, что специалисты в данной области не всегда все знают о своем предмете. Может быть (как говорит Джонсон), мы должны стремиться заменить безопасность мышления по хорошо определенным, знакомым канонам новым видом безопасности, основанным на допущении неясности, неопределенности и открытого выбора. Такой подход, разумно примененный, должен указывать путь ученому.

5. Исторически большинство людей в действительности были в большей или меньшей мере ситуационистами, но отличие современных условий заключается в том, что мы являемся ситуационистами в смысле рационального и открыто признаваемого метода.

Замечание

Многих врачей можно было бы включить в те группы, составляющие большинство людей, которые в действительности всегда являлись ситуационистами. Так, многие врачи используют субъективную меру вероятности, логарифмические шансы и логарифмические отношения правдоподобия в процессе принятия медицинского решения независимо от того, знают они об этом или нет.

Необходимо объединить научное направление, где используются субъективная мера вероятности и конкретные оценки значимости,

с «другим» миром — миром этики, представленным субъективными оценками значимости при принятии медицинского решения в отношении данного индивидуального пациента. При столкновении с проблемами этики мы предпочитаем подход, известный как «ситуационная этика». Стикел [107] очень хорошо резюмирует существующее положение: «Наконец в этом изложении был использован ситуационный подход с полным учетом критики, которая была выдвинута против него. Ситуационистов обвиняли, например, в безответственном отбрасывании или урезывании освященных временем моральных законов и этических принципов. К счастью, из нашего обсуждения, взятого в полном контексте, ясно, что ситуационный подход не только создает полезную «систему отсчета», относительно которой можно рассматривать моральные аспекты обсуждаемых проблем (трансплантация органов), но также усиливает чувство ответственности, которое сопутствует лучшим традициям медицины».

5.4. НАУЧНЫЕ ИССЛЕДОВАНИЯ И РАЗРАБОТКА ПРОЕКТОВ — НАСТОЯЩЕЕ И БУДУЩЕЕ

Выше мы отмечали необходимость использования новых математических процедур при исследовании процесса принятия решения в диагностике и планировании лечения. Математические процедуры, эффективные сами по себе, могут оказаться бесполезными при наличии неясной медицинской терминологии. Таким образом, весьма важно сделать медицинские термины настолько точно определенными, насколько это возможно для большого числа пользующихся ими специалистов, а в дальнейшем и улучшить использование медицинских терминов в литературе.

5.4.1. ИССЛЕДОВАНИЯ В ОБЛАСТИ МЕДИЦИНСКОЙ ТЕРМИНОЛОГИИ

Обсуждение медицинской терминологии, медицинского синтаксиса и многочисленных проектов, находящихся сейчас в стадии разработки, могло бы заполнить несколько томов. В настоящем разделе мы приведем примеры, которые, как нам кажется, хорошо иллюстрируют важность адекватного языка при исследовании процессов принятия решения в медицине.

Во-первых, важно, чтобы по каждой специальности группа экспертов создала тезаурус — список медицинских терминов и модифицирующих слов, которые бы согласились использовать специалисты в каждой области для изложения специальных вопросов в историях болезни и публикуемых научных статьях. Примером такого тезауруса является список приблизительно из 6200 слов, подготовленный д-ром Б. Г. Лэмсоном (Отделение патологии Клиник Калифорнийского университета в Лос-Анджелесе) для описания патологических состояний. Любой термин должен иметь одно и то же

значение в рамках данной определенной специальности независимо от того, используется ли он в истории болезни или в публикуемой статье. В работе «Проект исправления терминологии по диабету» (д-р А. Лазаров, отделение анатомии Университета Миннесоты) предпринята попытка сделать именно это. Когда-нибудь в будущем медицинские записи в историях болезни и медицинская литература будут непосредственно связаны между собой при помощи линий связи и машинной обработки. Создание систем обработки медицинской информации будет менее трудным, если мы уже теперь начнем составлять словарь для каждой медицинской специальности и уделять внимание правильному использованию терминов с целью уменьшения количества неясных понятий.

Во-вторых, каждая медицинская специальность нуждается в словаре — основном списке медицинских терминов, построенном в алфавитном порядке; он будет включать определение терминов, относительно которых достигнуто соглашение (группой специалистов одного профиля). На этом этапе члены такой специализированной группы столкнутся с необходимостью определения признаков и симптомов о данной специальности, и, конечно, все мы будем помогать им продвигаться в этом направлении. Например, каково определение термина «штампованная тень» для торакальных хирургов, рентгенологов и терапевтов? Было бы хорошо, если бы относительно определения такого термина, как «штампованная тень», используемого медиками нескольких специальностей, было достигнуто соглашение между различными группами врачей.

«Текущая медицинская терминология Американской медицинской ассоциации» представляет собой попытку установить словарь медицинских терминов для всей медицины.

В-третьих, нам необходимо накопить бóльший опыт по отношению к многочисленным комбинациям медицинских терминов, которые в медицинских работах используются в одних и тех же или сходных значениях. Проблемы в этой области огромны, ибо пока еще мало общих методов анализа структуры предложений. Может быть, медицине даже придется подождать осуществления соответствующих разработок в области лингвистики. Врачам часто кажется, что единственным способом медицинского описания является использование разнообразия слов и фраз. Возможно, удастся прийти к соглашению об определенных фразах или предложениях, которые будут использоваться в сообщениях о новых результатах, часто встречающихся в специальной области, например в рентгенологии. В настоящее время проводятся эксперименты для оценки целесообразности применения стандартных фраз в рентгенологии и патологии.

Каждой специальной области медицины необходима стандартная номенклатура для описания заболеваний. Имеющийся опыт в разработке «Стандартной номенклатуры в патологии» (SNOP)

является шагом в этом направлении в области патологии. Мы привлекаем внимание к SNOP, потому что она была создана для использования при записях историй болезни, обрабатываемых машиной.

Для каждой медицинской специальности полезно исследовать диагностические критерии, которые в ней используются, и, если необходимо, заново определить эти критерии. Вспомним обзор опыта Комитета Американской ассоциации по ревматизму, приведенный в гл. 2. Были определены критерии для ревматоидного артрита; критерии проверялись в течение двух лет и после этого были надлежащим образом модифицированы. Врачи продолжают пользоваться этими критериями; за исключением некоторых модификаций, которые поддерживают критерии на современном уровне, за последнее время они подвергались лишь небольшим изменениям. Ревматологи используют эти критерии на практике, и можно заключить, что усилия Комитета Американской ассоциации по ревматизму были не напрасны.

Для чего нужно разрабатывать тезаурус, составлять словарь и изучать структуру медицинских фраз? Один из ответов заключается в том, что эти проекты являются надежным способом упорядочения логической структуры медицинских знаний. Мы приобретаем все новые знания, но лишь очень немногие специалисты стремятся так организовать хранение этих знаний, чтобы они были доступны по первому требованию.

Как было отмечено в предисловии, исследования процесса принятия решения в медицине оказались возможными, потому что при наличии вычислительных машин можно ограничить разумными пределами усилия, затрачиваемые врачами для этой цели. Таким образом, врач может создать необходимый тезаурус, может определить признаки, симптомы и заболевания и может установить диагностические критерии. Затем он может подготовить свои медицинские данные и описания для машинной обработки. На основе медицинских данных, заложенных в машину, соответствующая машинная программа позволит легко отыскать нужные данные, чтобы выделить и подсчитать, например, количество «штампованных теней» в специализированной группе пациентов. Следовательно, мы можем надеяться вычислить частоты признаков, симптомов и заболеваний, которые нам нужны как основа для оценок условных вероятностей $P(S|D)$ и $P(D)$.

5.4.2. ПРИМЕРЫ СИМБИОЗА ВРАЧ — МАШИНА

Вычислительная машина является полезным устройством при рассмотрении некоторых структур. Лингвисты считают, что машина может служить хорошим хранилищем сложных структур и хо-

рошим помощником при «отладке» их представлений о структуре языка. Не лишним будет отметить, что в будущем, вероятно, биолог, патолог и рентгенолог смогут использовать машину для исследований и «отладки» их представлений о биологических структурах. Когда именно «в будущем» — это зависит от прогресса в лингвистике и в разработке специальных машинных языков; кроме того, это зависит от прогресса в деле приведения в порядок медицинской терминологии.

Для иллюстрации возьмем гипотетический пример из рентгенологии [47]. Разработан машинный язык, который позволяет рентгенологу описать то, что он видит на рентгеновском снимке грудной клетки. Предположим, что определенное количество информации о рентгенологических исследованиях грудной клетки уже подготовлено и заложено в машинную память рентгенологами. Рентгенолог говорит вычислительной машине (либо печатая на машинке, либо, в будущем, прямо посредством диктовки): «У этой 43-летней белой женщины на уровне первого ребра в правой верхней доле легкого наблюдается «пушистый» инфильтрат. Каковы наиболее подходящие диагнозы?»

Вычислительная машина пытается проанализировать смысл этого утверждения и ищет различные связи для «пушистого» (мелкоочагового) инфильтрата. Машина спрашивает (при помощи пишущей машинки или телевизионного экрана): «Пожалуйста, опишите, что вы имеете в виду, говоря «пушистый» инфильтрат. Похож ли он на «пушистый» комок хлопка или на «пушистое» облако?»

Рентгенолог отвечает: «Пожалуй, скорее на «пушистый» комок хлопка с довольно нечеткими краями и очень яркими полосками повышенной плотности по всему комку». Рентгенолог может почувствовать себя несколько неуверенно, ибо, хотя он научился узнавать «пушистый» инфильтрат, ему трудно дать точное описание структуры образа.

Давайте предположим, что работа продолжается, так как рентгенологу все еще трудно описать «пушистый» инфильтрат. Наконец, по различным описательным терминам, используемым рентгенологом, машинная программа находит изображение, заложенное в машинной памяти, которое может быть «пушистым» инфильтратом.

Машинная программа указывает: «Изображение того, что вы описали как «пушистый» инфильтрат, будет синтезировано на экране. Приемлемо ли оно? Если нет, то дайте, пожалуйста, дальнейшее описание». И тогда диалог будет продолжен.

Важная роль принадлежит процедуре синтеза изображения из словесного описания. Продолжая поиск деталей, машинная программа помогает рентгенологу в описании и «отладке» его представлений о структуре. Она помогает ему установить тесную

корреляцию между его языком и изображением. Корреляция словесного описания «пушистого» инфильтрата, синтезированного (вычислительной машиной) изображения такого инфильтрата и вида его на рентгеновском снимке грудной клетки проверяется на точность путем наложения синтезированного изображения на действительное изображение рентгеновского снимка. Рентгенолог видит, каково соответствие или несоответствие между изображениями, и может продолжать улучшать сходство между ними, переходя к дальнейшему описанию. В некоторый момент рентгенолог может решить, что изображения достаточно близки, и задать машине вопрос, какие данные имеются в ее памяти относительно «пушистого» инфильтрата. Если данных о «пушистом» инфильтрате в машинной памяти нет, рентгенологу придется добавить в машинную память свои данные по условным вероятностям $P(S|D)$ и $P(D)$ для всех заболеваний D , где, как известно из практики, встречается «пушистый» инфильтрат, и, кроме того, модифицировать машинную программу, чтобы включить в нее новые диагностические возможности. В следующий раз, когда рентгенолог обнаружит у больного «пушистый» инфильтрат, заложенные данные помогут ему (если он будет нуждаться в этой помощи) вспомнить условные вероятности $P(S|D)$ для мелкоочагового инфильтрата при бактериальной пневмонии, туберкулезе и т. д. и, наконец, вычислить апостериорную вероятность $P(D|S)$ для каждого заболевания, которое, с его точки зрения, заслуживает рассмотрения. При этом окончательный выбор диагноза осуществляется рентгенологом.

Рентгенологу предстоит бурные времена. Ему придется больше думать об описании новообразований в терминах структуры и о значении структуры по отношению к нормальным процессам и патологии.

Изложенная выше история о рентгенологе, работающем с машиной, в действительности не произойдет ни завтра, ни, пожалуй, в ближайшие годы. Мы привели этот пример, чтобы проиллюстрировать подход к использованию машин в исследованиях, а со временем и в клинической практике. Слишком часто мы читаем или слышим о вычислительной машине, ставящей диагноз, или об исследованиях, которые сравнивают диагностическую точность вычислительной машины с точностью врачей. В конечном счете машина может действовать только на основе машинной программы и медицинских данных, причем и программа и данные в действительности разрабатываются и задаются врачом. Машинное программирование не очень хорошо знакомо большинству специалистов нашего поколения, хотя многие студенты средних специальных и высших школ уже изучают программирование. Именно это незнание машинного программирования приводит к тому, что многие

люди приписывают вычислительной машине нереалистические способности.

Знаменитый ученый Джон фон Нейман отметил в одной из своих работ [122], что английский логик А. Тьюринг в 1937 г. доказал возможность разработки системы программ для вычислительной машины, которые заставят ее вести себя так, как если бы это была другая машина с другими специальными свойствами. «Важный результат Тьюринга заключается в том, что таким путем можно заставить первую машину имитировать поведение *любой* другой машины. Структура команд, которой она теперь вынуждена следовать, может полностью отличаться от команд, характеризующих первую машину, которая в действительности используется в этом процессе. Таким образом, структура команд, к которой обращаются, может в действительности иметь значительно более сложный характер, чем команды, присущие первой машине: каждая из команд второй машины может включать выполнение нескольких операций машины, которая рассматривается как первичная».

Из этого следует, что можно разработать систему программ, которая заставит некоторую математическую машину вести себя так, как будто она является другой специфической машиной, например мозгом. Далее, можно заключить, что некоторая система программ может вынудить машину имитировать поведение мозга при постановке диагноза, причем машине не нужно (и скорее всего она не будет этого делать) использовать тот же самый процесс (каков бы он ни был), который используется мозгом человека при постановке того же диагноза.

Теоретически можно прийти к выводу, что машина может быть запрограммирована для осуществления диагностического процесса и постановки диагноза. В действительности же, как мы уже указывали, в настоящее время еще невозможно создать общую машинную программу диагностического процесса, и разговоры о машине, «ставящей диагноз», не принесут пользы, если при этом неизменно придается большее значение машинной программе (алгоритму) по сравнению с врачебным мышлением (клиническим суждением).

Много убедительных доводов можно представить в защиту экспериментов, которые сопоставляют совокупность клинического суждения врача и алгоритма (машинной программы) с чисто клиническим суждением. Мы уверены, что в настоящее время имеется достаточно оснований, чтобы предпринимать такие эксперименты, и предлагаем некоторые доказательства этого мнения, основанные на исследовании системы вероятностной обработки информации (PIP), которое было проведено Эдвардсом и др. [18, 19, 86].

5.4.3. НЕКОТОРЫЕ ЭКСПЕРИМЕНТЫ В ОБЛАСТИ ПРИНЯТИЯ МЕДИЦИНСКОГО РЕШЕНИЯ С ИСПОЛЬЗОВАНИЕМ СИСТЕМЫ ВЕРОЯТНОСТНОЙ ОБРАБОТКИ ИНФОРМАЦИИ (PIP)

Эдвардс и его сотрудники [86] были первыми, кто выполнил эксперименты, сравнивающие возможности людей с методом Байеса при обработке информации. Они первыми обнаружили тот факт, что человек является консервативным устройством обработки информации, не способным выделить из полученных данных скрытую в них определенную содержательную информацию. Эдвардс [17] предложил идею системы вероятностной обработки информации (PIP), которая использовала бы некоторый автоматический механизм комбинирования информации, освобождающий человека от легко механизированной части этой работы. По его предложению, человек должен оценивать отношение правдоподобия, т. е. отношение вероятности некоторых данных при одной гипотезе к их вероятности при другой гипотезе (если данные независимы, а гипотезы являются взаимоисключающими), а вычислительная машина должна синтезировать такие оценки по всем данным и гипотезам с тем, чтобы получить текущее распределение вероятностей всех гипотез на основе всех имеющихся данных.

Два основных результата, полученные в исследовании по проблеме симбиоза суждения человека и вычислений по теореме Байеса [18, 19], заключаются в консерватизме суждения человека и в том, что система PIP обрабатывает информацию эффективнее, чем ее соперник. Консерватизм был обнаружен в многочисленных и разнообразных экспериментах по принятию решения, включающих военные ситуации, связанные с приказами и управлением. Консерватизм отдельных лиц проявлялся в том, что они осуществляли только небольшие изменения в оценках вероятностей при переходе от одних данных к другим, даже когда байесовские вероятности указывали на значительные изменения. Лица, принимающие решения, не были настолько уверены, насколько им позволила бы теорема Байеса.

Система вероятностной обработки информации PIP по существу является системой совместной работы людей, оценивающих вероятности, с байесовским процессором, который вычисляет вероятности по формуле Байеса (вычислительная машина, запрограммированная для таких вычислений). На выходе системы получается непрерывно уточняемая оценка приемлемости рассматриваемой гипотезы.

При этом возникает интересный вопрос: «Не лучше ли было бы, чтобы человек оценивал априорные вероятности и отношения правдоподобия, а затем машина, запрограммированная по теореме Байеса, преобразовывала эти вероятности и отношения правдоподобия в апостериорные вероятности?» Накоплено значительное

количество фактов в пользу утвердительного ответа на этот вопрос, хотя многие другие связанные с ним вопросы остаются пока нерешенными [18, 19].

Прежде чем описать экспериментальные схемы, использующие системы вероятностной обработки информации РИР для изучения процессов принятия решения в медицине, мы приведем основные результаты недавней работы Эдвардса [19] по использованию систем РИР для диагностической обработки информации:

1. Существуют два способа придания количественного характера качественным данным, и кумулятивный характер теоремы Байеса делает ее применимой для обоих этих способов.

а) Если имеющихся данных недостаточно и для улучшения оценок используются повторные наблюдения, то каждое новое наблюдение можно обрабатывать непосредственно после его поступления, с тем чтобы использовать результат предыдущего вычисления в качестве необходимой априорной вероятности.

б) Байесовский метод применяется к информации, поступающей по нескольким входным линиям. При использовании теоремы Байеса не имеет значения, подобен или нет в качественном физическом смысле каждый элемент информации предыдущему элементу.

в) Таким образом, полезность теоремы Байеса для обработки информации определяется тем, что она формально является оптимальным способом объединения информации от одного или нескольких источников.

2. Независимость элементов информации требует специального рассмотрения. Корреляция между двумя элементами обычно возникает из-за того, что оба они связаны с каким-то основным процессом, который их создает. С точки зрения теоремы Байеса можно различить два случая.

а) Более обычный случай имеет место, когда этот основной процесс представляет собой одну из рассматриваемых гипотез (D).

б) Другой (и более сложный) случай возникает, когда корреляция между данными (S_1 и S_2) отражает нечто иное, чем истинность одной из рассматриваемых гипотез (D). В этом случае условная независимость не имеет места, и мы вынуждены либо заключить, что величина корреляции достаточно мала, чтобы ею можно было пренебречь, либо использовать любую из возможных процедур для устранения влияния взаимосвязи на конечное апостериорное распределение вероятностей.

3. Использование теоремы Байеса в системе РИР имеет несколько дополнительных преимуществ.

а) Она автоматически разделяет данные в соответствии с рассматриваемым процессом, отфильтровывает шум, сохраняя при этом полезную информацию, и автоматически взвешивает каждый элемент информации в соответствии с его важностью.

б) Применение теоремы требует хранения минимального объема входных данных. Как только данные обработаны, они могут быть «стерты» (поскольку это касается их использования по методу Байеса), если только они условно независимы. Если же существует возможность нарушения условной независимости, необходимо сохранять в памяти все данные, для того чтобы иметь возможность установить, имели ли место такие нарушения независимости на самом деле.

в) На выходе системы мы получаем как раз те данные, которые необходимы для выбора типа действия на основе максимизации ожидаемой величины.

4. Единственными величинами, которые необходимо вводить в систему РІР, являются $P(S|D)$, $L(D_1; S)$ или аналогичные величины. Их значения могут быть вычислены (как частоты событий) или получены на основе суждения специалиста.

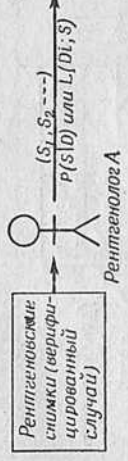
5. В основе системы РІР лежит конкретная гипотеза, заключающаяся в том, что человек может служить в качестве генератора величин $P(S|D)$ или $L(D_1; S)$. Другими словами, человека можно научить оценивать эти вероятности (или величины, связанные с ними) с достаточной точностью, чтобы их можно было положить в основу принятия решений, даже когда эти вероятности не могут быть вычислены никакой другой процедурой.

6. Как только принято предположение, что отношения правдоподобия $L(D_1; S)$ будут вводиться в систему на основе человеческого суждения, можно легко разработать детали системы РІР.

Эксперименты трех возможных типов, связанные с изложенным выше подходом, показаны схематически на фиг. 49, 50 и 51. Термин «байесовский процессор» используется для обозначения универсальной цифровой вычислительной машины, запрограммированной для вычисления величины $P(D|S)$ на основе теоремы Байеса. Врач может вводить исходные данные на перфокартах, при помощи пишущей машинки, световым пером на экране и т. д., а выходные величины могут быть выведены на печатающее устройство машины, отпечатаны на пишущей машинке или представлены визуально на экране. На фиг. 51 изображена схема эксперимента (насколько нам известно, подобный эксперимент пока не проводился), в котором используется цифровая вычислительная машина для расчета ожидаемых величин $EV_j = \sum_i P(D_i|S) V_{ij}$.

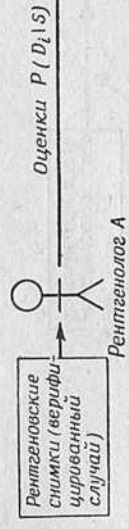
По-видимому, не следует пытаться заранее предсказать результаты предложенных выше экспериментов или каких-либо экспериментов аналогичного типа в других областях медицины со сложными ситуациями, возникающими в процессе принятия решения. Результаты исследований по вероятностной обработке информации, приведенные в работах [45] и [86], показывают, что оценка человеком апостериорных вероятностей альтернативных гипотез $P(D|S)$ хуже оценки апостериорных вероятностей, полученной

Вариант 1



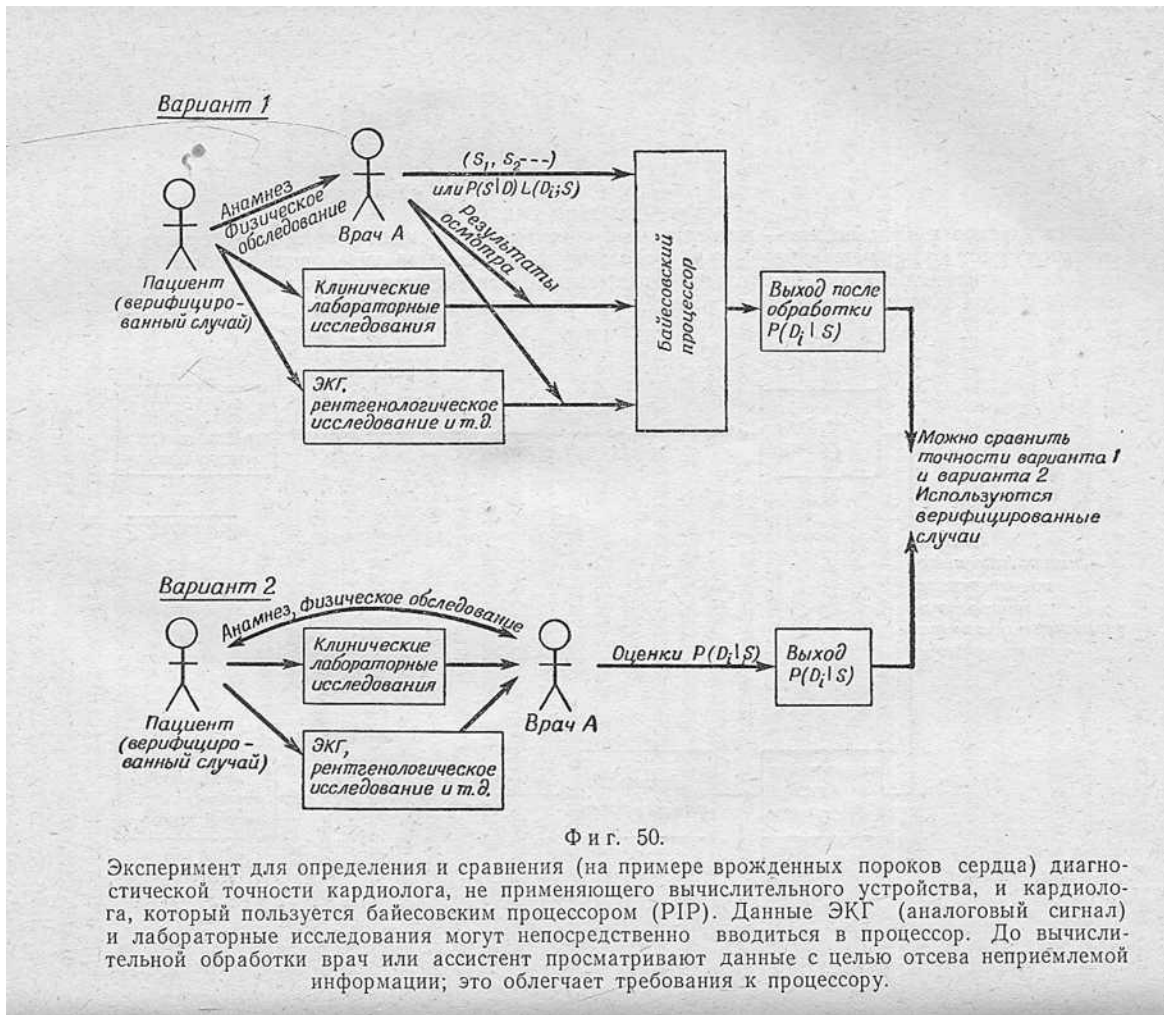
Можно сравнить точности варианта 1 и варианта 2. Используются верифицированные случаи

Вариант 2



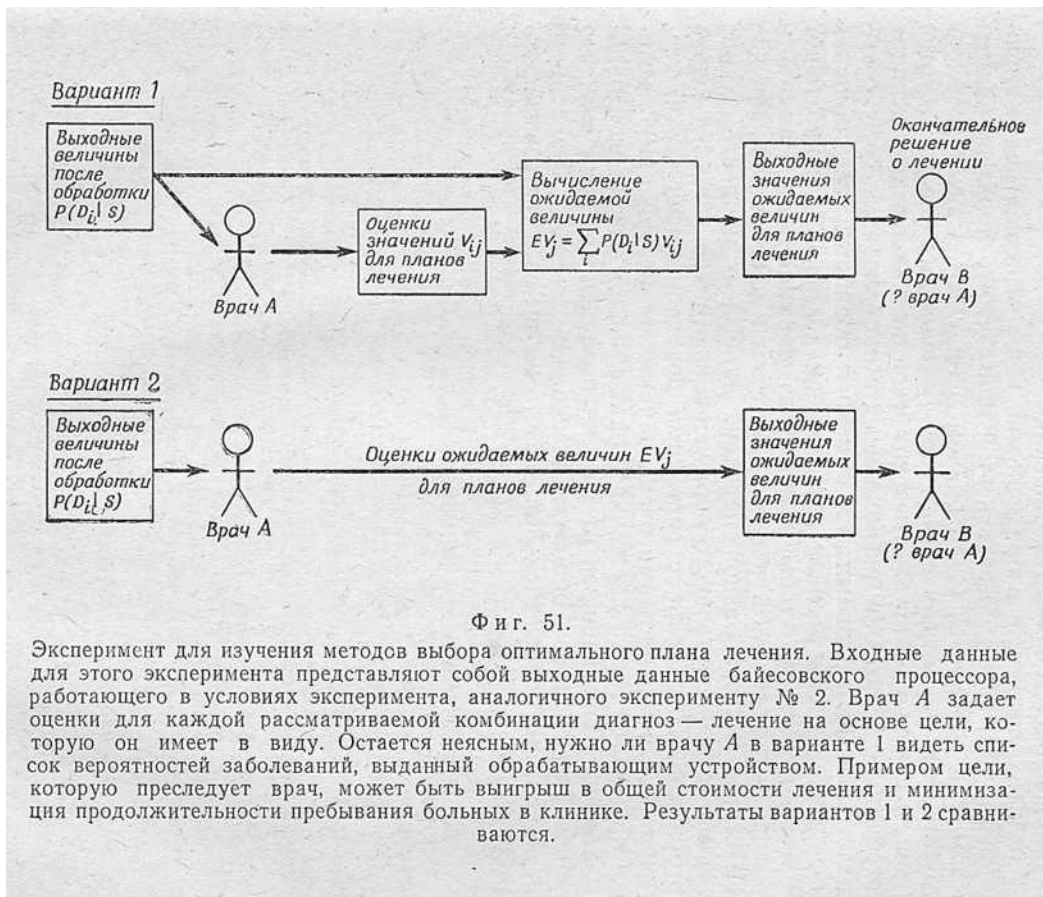
Ф и г. 49.

Эксперимент для определения и сравнения (на примере первичных опухолей кости) диагностической точности рентгенолога, не применяющего вычислительного устройства, и рентгенолога, который пользуется байесовским процессором (РПР). Величины $P(S|D)$ или $L(D_i; S)$ могут быть заданы либо рентгенологом А, либо рентгенологом — «экспертом» в области первичных опухолей кости.



Фиг. 50.

Эксперимент для определения и сравнения (на примере врожденных пороков сердца) диагностической точности кардиолога, не применяющего вычислительного устройства, и кардиолога, который пользуется байесовским процессором (РР). Данные ЭКГ (аналоговый сигнал) и лабораторные исследования могут непосредственно вводиться в процессор. До вычислительной обработки врач или ассистент просматривают данные с целью отсева неприемлемой информации; это облегчает требования к процессору.



Фиг. 51.

Эксперимент для изучения методов выбора оптимального плана лечения. Входные данные для этого эксперимента представляют собой выходные данные байесовского процессора, работающего в условиях эксперимента, аналогичного эксперименту № 2. Врач А задает оценки для каждой рассматриваемой комбинации диагноз — лечение на основе цели, которую он имеет в виду. Остается неясным, нужно ли врачу А в варианте 1 видеть список вероятностей заболеваний, выданный обрабатывающим устройством. Примером цели, которую преследует врач, может быть выигрыш в общей стоимости лечения и минимизация продолжительности пребывания больных в клинике. Результаты вариантов 1 и 2 сравниваются.

при помощи теоремы Байеса с использованием отношения правдоподобия $P(S|D_1)/P(S|D_2)$, выбранного человеком. Эти результаты приводят к заключению, что полезно будет предпринять исследование систем РИР в целом ряде областей медицины, где осуществляются процессы принятия решения.

Как мы указывали в предисловии, наша основная цель заключалась в том, чтобы сказать в этой книге что-то полезное врачам и тем, кто собирается стать врачом в будущем. Сейчас хотелось бы выразить эту мысль по-другому. Было написано много статей и много сказано об автоматических медицинских регистрирующих устройствах, о создании региональных и национальных сетей с вычислительными машинами для обмена и обработки медицинской информации. Все эти идеи хороши, однако врачам нужно будет много потрудиться, чтобы выработать терминологию, составить словари и определить признаки, симптомы и заболевания, прежде чем станет целесообразной передача данных о пациенте. У нас есть много вычислительного оборудования и всего лишь несколько алгоритмов. Положение с системами передачи информации, терминологией и алгоритмами очень напоминает историю [106], которая произошла много лет назад с одним членом английской палаты лордов. Запыхавшийся от спешки посланец сообщил ему, что прокладка кабеля в Индию закончена. Он лениво и без всякого энтузиазма посмотрел и сказал: «И что же вы собираетесь *сказать* Индии?».

ЛИТЕРАТУРА

1. A d a m s W. R., Studies of teaching in the diagnostic process, in: Jacquez, John A. (Ed.), *The Diagnostic Process*, Proceedings of a Conference held at the University of Michigan, May 1963, 1964, pp. 367—385.
2. B a e r R. M., A note on computer diagnosis, *IEEE Transactions on Biomedical Engineering*, **BME-11**, 8—12 (Jan.—April 1964).
3. B r u n e r J. S., G o o d n o w J. J., A u s t i n G. A., *A Study of Thinking*, N. Y., Wiley, 1956.
4. B r u n k H. D., L e h r J. L., An improved Bayes' method for computer diagnosis, in: Proceedings of Conference on the Use of Computers in Radiology, University of Chicago Center for Continuing Education, Oct. 22, 1966 (в печати).
5. Case Records of the Massachusetts General Hospital (Case 4—1967), *New Eng. J. Med.*, 276, 230—237 (1967).
6. C a s t l e m a n B., D u d d l e y H. R. (Ed.), in: Proceedings of Clinicopathological Conference of the Massachusetts General Hospital, London, Churchill Ltd., 1960.
7. C o b b S., On the development of diagnostic criteria, *Arthritis Rheum.*, **3**, 91—95 (Feb. 1960).
8. C o f f i n R. W., G o h e e n H. E., S t a h l W. R., Simulation of a Turing machine on a digital computer, in: Proceedings of AFIPS Fall Joint Computer Conference, 24, 35—43, 1963 (American Federation of Information Processing Societies).
9. C o l l e n M. F., R u b i n L., N e y m a n J. et al., Automated multiphase screening and diagnosis, *Amer. J. Public Health*, **54**, 741—750 (May 1964). Copyright 1964, by the American Public Health Association, Inc.
10. C o l l e n M. F., R u b i n L., D a v i s L., Computers in multiphase screening, in: Stacy R. W. and Waxman B. D. (Eds.), *Computers in Biomedical Research*. New York, Academic, 1965, vol. 1, ch. 14, pp. 339—352.
11. C o r n f i e l d J., Bayes' theorem, in: Proceedings of 6th IBM Medical Symposium, 1964, pp. 163—196, (a).
12. C o r n f i e l d J., Discriminant functions, in: Proceedings of 6th IBM Medical Symposium, 1964, pp. 567—189, (b).
13. C o r n f i e l d J., Sequential trials, sequential analysis and the likelihood principle, *Amer. Statistician*, **20(2)**, 18—23 (April 1966).
14. D e F i n e t t i B., Foresight: Its logical laws, its subjective sources, *Annales de l'Institut Henri Poincaré*, **7**, 1—68, 1937 (English translation available in Kyberg and Smokler: *Studies of Subjective Probability*, New York, Wiley, 1964, pp. 95—158).
15. D e F i n e t t i B., of University of Rome (личное сообщение), June 27, 1966.
16. E d e n M., The taxonomy of disease, in: Jacquez, John A. (Ed.), *The Diagnostic Process*, Proceedings of Conference held at The University of Michigan, May 1963—1964, pp. 47—49.
17. E d w a r d s W., Dynamic decision theory and probabilistic information processing, *IEEE Transactions on Human Factors in Electronics*, **4**, 59—73 (1962).
18. E d w a r d s R., Introduction to Special Issue on Revision of opinions by men and man-machine systems, *IEEE Transactions on Human Factors in Electronics*, **HFE-7**, 1—6, March 1966, (a).

19. Edwards W., Nonconservative Probabilistic Information Processing Systems, Report ESD-TR-66-404, Dec. 1966, (b) (Engineering Psychology Laboratory, Institute of Science and Technology, The University of Michigan, Ann Arbor.)
20. Edwards W. M., Cox R. S., Jr., Garland L. H., The solitary nodule (coin lesion) of the lung, An analysis of 52 consecutive cases treated by thoracotomy and a study of preoperative diagnostic accuracy, *Amer. J. Roentgen*, 88, 1020—1042 (Dec. 1962).
21. Edwards W., Lindman H., Savage L. J., Bayesian statistical inference for psychological research, *Psychol. Rev.*, 70, 193—242 (May 1963).
22. Engle R. L., Jr., Davis B. J., Medical diagnosis: Present, past and future. I. Present concepts of the meaning and limitations of medical diagnosis, *Arch. Int. Med. (Chicago)*, 112, 512—519 (Oct. 1963).
23. Engle R. L., Jr., Medical diagnosis: Present, past and future. II. Philosophical foundations and historical development of our concepts of health, disease and diagnosis, *Arch. Int. Med. (Chicago)*, 112, 520—529 (Oct. 1963).
24. Engle R. L., Jr., Medical diagnosis: Present, past and future. III. Diagnosis in the future, including a critique on the use of electronic computers as diagnostic aids to the physician, *Arch. Int. Med. (Chicago)*, 112, 530—543 (Oct. 1963), (b).
25. Erdmann A. J., Jr., Data processing and the interpretation of symptoms. I. Results, in: Jacquez J. A. (Ed.), *The Diagnostic Process, Proceedings of a Conference held at The University of Michigan, May 1963—1964*, pp. 215—31.
26. Feigenbaum E. A., *An Information Processing Theory of Verbal Learning*, Report № P-1817 (Oct.), The Rand Corporation, Santa Monica, 1959.
27. Feinstein A. R., Boolean algebra and clinical taxonomy. I Analytic synthesis of the general spectrum of a human disease, *New. Eng. J. Med.*, 269(18), 929—938 (Oct. 31, 1963).
28. Feinstein A. R., Symptoms as an index of biological behaviour and prognosis in human cancer, *Nature (London)*, 209(5020), 241—245 (Jan. 15, 1966).
29. Feinstein A. R., Stern E. K., Spagnuolo M., The prognosis of acute rheumatic fever, *Amer. Heart J.*, 68(6), 817—834 (Dec. 1964), The C. V. Mosby Company.
30. Fisher R. A., The use of multiple measurements in taxonomic problems, *Ann. Eugen*, 7, 179—188 (1936).
31. Fitzgerald L. T., Williams C. M., *Computer Diagnosis of Thyroid Disease, 1964*. (From Department of Radiology, University of Florida College of Medicine Gainesville, Florida.)
32. Fitzgerald L. T., Overall J. E., Williams C. M., A computer program for diagnosis of thyroid disease, *Amer. J. Roentgen*, 97, 901—905 (August 1966).
33. Fletcher J., *Situation Ethics, The New Morality*, Philadelphia, Westminster, 1966.
34. Garland L. H., Scientific evaluation of diagnostic procedures, *Radiology*, 52, 309—327 (March 1949).
35. Garland L. H., Studies on the accuracy of diagnostic procedures, *Amer. J. Roentgen*, 82(1), 25—38 (July 1959).
36. Good I. J., Kinds of probability, *Science*, 129, 443—447 (Feb. 20, 1959).
37. Green D. M., Swets J. A., *Signal Detection Theory and Psychophysics*, N. Y. Wiley, 1966.
38. Harman H. H., *Modern Factor Analysis*, Chicago. U. of Chicago, 1960, p. 469.

39. Horvath W. J., The effect of physician bias in medical diagnosis, *Behav. Sci.*, 9, 334—340 (1964).
40. Horvath W. J., Toles W. E., Bostrom R. C., Quantitative measurements of cell properties of Papanicolaou smears as criteria for screening, in: Transactions of First International Cytology Congress, Chicago, American Cancer Society, 1956, pp. 371—97.
41. Jacques J. A. (Ed.), The Diagnostic Process, Proceedings of Conference held at The University of Michigan, May 1963. Ann Arbor, Malloy Lithographing, Inc., 1964, 391 pp., (a) (Library of Congress Catalog № 64, 21210).
42. Jacques J. A., The diagnostic process: problems and perspectives, in: Jacques J. A. (Ed.), The Diagnostic Process, Proceedings of Conference held at The University of Michigan, May 1963, 1964, pp. 23—37, (b).
43. Johnson M. L., A course on factors influencing scientific judgment, *J. Med. Educ.*, 30(7), 391—397 (July 1955), (a).
44. Johnson M. L., Observer error; its bearing on teaching, *Lancet*, 269(6887), 422—424, (August 27, 1955), (b).
45. Kaplan R. J., Newman J. R., Studies in probabilistic information processing, *IEEE Transactions on Human Factors in Electronics*, HFE-7, 49—63 (1966).
46. King D. J., Manegold R. F., Consistency of cardiologists' prognostic judgments in cases of myocardial infarction, *Amer. J. Cardiol.*, 15, 26—32 (Jan. 1965).
47. Kirsch R., The hypothetical example is derived from some ideas presented by Mr. Kirsch (National Bureau of Standards) at a Conference on the Use of Computers in Radiology, University of Chicago Center for Continuing Education, Oct. 21, 1966.
48. Kleinmuntz B., MMPI decision rules for the identification of college maladjustment: A digital computer approach, *Psychol. Monogr.* 77 (1963), (a).
49. Kleinmuntz B., Profile analysis revisited: A heuristic approach, *J. Counsel. Psychol.*, 10(4), 315—324 (1963), (b).
50. Kleinmuntz B., Diagnostic problem solving by computer, *Jap. Psychol. Res.*, 7, 189—194 (1965).
51. Ledley R. S., High-speed automatic analysis of biomedical pictures, *Science*, 146 (3641), 216—223 (Oct. 9, 1964).
52. Ledley R. S., Ruddle F. H., Chromosome analysis by computer, *Sci. Amer.*, 214 (4), 40—46 (1966).
53. Ледли Р. С., Ластед Л. Б., Объективные основания диагноза, Кибернетический сборник, вып. 2, ИЛ. М., 1961.
54. Ledley R. S., Lusted L. B., The use of electronic computers in medical data processing, *IRE Transactions on Medical Electronics*, ME-7, 31—47 (Jan. 1960).
55. Ледли Р. С., Ластед Л. Б., Медицинская диагностика и современные методы выбора решения, Сб. «Математические проблемы в биологии», изд-во «Мир», 1966.
56. Linderg F. E., The health of the American people, *Sci. Amer.*, 214, 21—29 (June 1966).
57. Lindley D. V., Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II. Inference, Cambridge, Cambridge U. P., 1965, ch. 5 and 7.
58. Lipkin M., The role of data processing in the diagnostic process, in: Jacques J. A. (Ed.), The Diagnostic Process, Proceedings of Conference held at The University of Michigan, May 1963, 1964, pp. 255—73, (a).
59. Lipkin M., The likelihood concept in differential diagnosis, *Perspect Biol. Med.*, 7, 485—497 (Summer 1964), (b).
60. Lipkin M., Hardy J. D., Mechanical correlation of data in differential diagnosis of hematological diseases, *JAMA*, 166(2), 113—125 (Jan. 11, 1958).

61. Lodwick G. S., A probabilistic approach to the diagnosis of bone tumors, *Radiol. Clin. N. Amer.*, 3, 487—497 (Dec. 1965).
62. Lodwick G. S., Computer aided diagnosis in radiology, A research plan, *Invest. Radiol.*, 1, 72—80 (Jan.—Feb. 1966).
63. Lodwick G. S., Turner A. H., Jr., Computer Aided Diagnosis of Primary Bone Tumors, Dec. 1965 (Available from Dr. Gwilym S. Lodwick, Department of Radiology, University of Missouri School of Medicine, Columbia, Missouri 65202).
64. Lodwick G. S., Turner A. H., Jr., Lusted L. B., Templeton A. W., Computer-aided analysis of radiographic images, *J. Chronic Dis.*, 19(4), 485—496 (April 1966).
65. Льюис Р. Д., Райфа Х., Игры и решения, ИЛ, М., 1961.
66. Lusted L. B., Logical analysis in roentgen diagnosis, *Radiology*, 74, 178—93 (Feb. 1960).
67. Manning R. T., Watson L., Signs, symptoms and systematics, *JAMA*, 198, 1180—1184 (Dec. 12, 1966).
68. Mendelsohn M. L., Kolman W. A., Bostrom R. C., Initial approaches to the computer analysis of cytophotometric fields, *Ann. N. Y. Acad. Sci.*, 115, 998—1008 (July 31, 1964).
69. Meyers P. H., Becker H. C., Sweeney J. W., et al., Evaluation of a computer-retrieved radiographic image, *Radiology*, 81, (2), 201—206 (1963).
70. Meyers P. H., Nice C. M., Becker H. C., et al., Automated computer analysis of radiographic images, *Radiology*, 83(6), 1029—1034 (1964).
71. Mosteller F., Wallace D. L., Inference and Disputed Authorship, The Federalist, Reading, Addison—Wesley, 1964.
72. Nash F. A., Differential diagnosis, an apparatus to assist the logical faculties, *Lancet.*, 266(6816), 874—875 (April 17, 1954).
73. Newell A., Simon H. A., Computer simulation of human thinking, *Science*, 134, 2011—2017 (Dec. 22, 1961).
74. Neyman J., Pearson E. S., On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society*, London, (Series A), 231, 289—337 (1933).
75. Nugent C. A., Warner H. R., Dunn J. T., et al., Probability theory in the diagnosis of Cushing's syndrome, *J. Clin. Endocr.*, 24(7), 621—627 (1964).
76. Overall J. E., Hollister L. E., Criteria for evaluating diagnostic decision making, in: Mountain, C. (Ed.), Proceedings of Second Houston Symposium on Biomathematics, Springfield, Thomas, 1964.
77. Overall J. E., Williams C. M., Models for medical diagnosis, *Behav. Sci.*, 6, 134—141 (April 1961), (a).
78. Overall J. E., Williams C. M., Models for medical diagnosis: Factor analysis. I. Theoretical, *Medizinische Dokumentation*, 5, 51—56 (April 1961), (b).
79. Overall J. E., Williams C. M., Models for medical diagnosis: Factor analysis. II Experimental, *Medizinische Dokumentation*, 5, 78—80 (July 1961), (c).
80. Overall J. E., Williams C. M., Conditional probability program for diagnosis of thyroid function, *JAMA*, 183, 307—313 (Feb. 2, 1963), (a).
81. Overall J. E., Williams C. M., Comparison of alternative models for thyroid diagnosis, in: Proceedings of Third San Diego Symposium for Biomedical Engineering, La Jolla, California, 1963, pp. 141—6, (b), (8848 La Jolla Shores Drive).
82. Overall J. E., Williams C. M., A computer procedure for the diagnosis of thyroid functioning, in: Enslein, K. (Ed.), Data Acquisition

- and Processing in Biology and Medicine, New York, Pergamon, 1964, vol. 2, pp. 261—276.
83. Paycha F., Critique scientifique des démarches intellectuelles du diagnostic. (Critical study of intellectual procedures in diagnosis.) *Presse Med.* 63, 1753—1756 (Dec. 17, 1955).
 84. Paycha F., Diagnosis, therapeutics, prognosis and computers, *IRE Transactions on Medical Electronics*, ME-7, 288—290 (Oct. 1960).
 85. Peckham R. H., Betting odds in medical diagnosis, *Amer. J. Med. Sci.*, 253, 36—37 (Jan. 1967).
 86. Phillips L. D., Hays W. L., Edwards W., Conservatism in complex probabilistic inference, *IEEE Transactions on Human Factors in Electronics*, HFE-7, 7—18 (1966).
 87. Pickering, Sir George W., In conclusion: the physician, in: *Controlled Clinical Trials — A Symposium*, Oxford, Blackwell, 1960, pp. 163—167.
 88. Prewitt J. M., Mendelsohn M. L., The analysis of cell images, *Ann. N. Y. Acad. Sci.*, 128, 1035—1053 (Jan. 31, 1966).
 89. Raiffa H., Schlaifer R., *Applied Statistical Decision Theory*, Boston, Harv. Bus. School, 1961.
 90. Reitman W. R., Information-processing models in psychology, *Science*, 144, 1192—1198 (June 5, 1964).
 91. Rimoldi H. J. A., Testing and analysis of diagnostic skills, in: Jacques, John A. (Ed.), *The Diagnostic Process*, Proceedings of Conference held at The University of Michigan, May 1963, 1964, pp. 315—43.
 92. Rome H. P., Automation techniques in personality assessment, The problem and procedure, *JAMA*, 182, 72 (Dec. 15, 1962).
 93. Ropes M. W., Bennett G. A., Cobb S., et al., Proposed diagnostic criteria for rheumatoid arthritis, *Ann. Rheum. Dis.*, 16, 118—125 (March 25, 1957).
 94. Rosen C. A., Pattern classification by adaptive machines, *Science*, 156, 38—44 (April 7, 1967).
 95. Rubin L., Collen M. F., Goldman G. E., Frequency decision theoretical approach to automated medical diagnosis, in: *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, U. of Calif., 1967, vol. IV, pp. 867—886.
 96. Russell C. D., Diagnostic screening tests, *New Eng. J. Med.*, 275, 280 (August 4, 1966).
 97. Scheff T. J., Decision rules, types of error, and their consequences in medical diagnosis, *Behav. Sci.*, 8, 97—107 (1963).
 98. Simon H. P., *Administrative Behavior*, 2nd Ed., New York, Macmillan, 1957, p. XXIX.
 99. Simon H. A., Newell A., Information processing in computer and man, *Amer. Sci.*, 52, 281—300 (Sept. 1964).
 100. Slack W. V., Hicks G. O., Reed C. E., et al., A computer based medical history system, *New Eng. J. Med.*, 274(4), 194—198 (Jan. 27, 1966).
 101. Sokal R. R., Numerical taxonomy and disease classification, in: Jacques, John A. (Ed.), *The Diagnostic Process*, Proceedings of Conference held at The University of Michigan, May 1963, 1964, pp. 51—60.
 102. Sokal R. R., Sneath P. H., *Principles of Numerical Taxonomy*, San Francisco, Freeman, 1963.
 103. Sprosty P. J., The use of questions in the diagnostic problem solving process, in: Jacques, John A. (Ed.), *The Diagnostic Process*. Proceedings of Conference held at The University of Michigan, May 1963, 1964, pp. 281—308.
 104. Stahl W. R., Coffin R. W., Goheen H. E., Simulation of

- biological cells by systems composed of string processing finite automata, in: *Proceedings of AFIPS Spring Joint Computer Conference*, 25, 89—102 (1964). (American Federation of Information Processing Societies.)
105. Stevens S. S., On the operation known as judgment, *Amer. Sci.*, 54(4), 385—401 (Dec. 1966).
 106. Stewart W. H., Health and the New Technologies, Presented at the Interuniversity Communications Council (EDUCOM) Conference, Durham, North Carolina, April 15, 1966.
 107. Stickel D. L., Ethical and moral aspects of transplantation, *Monogr. Surg. Sci.*, 3, 267—301 (1966).
 108. Stumpf S. E., Some moral dimensions of medicine, *Ann. Int. Med.*, 64(2), 460—470 (Feb. 1966).
 109. Suppes P., The uses of computers in education, *Sci. Amer.*, 215(3), 207—220 (Sept. 1966).
 110. Swalm R. O., Utility theory — insights into risk taking, *Harvard Business Rev.*, 44, 123—136 (1966).
 111. Swets J. A. (Ed.), Signal Detection and Recognition by Human Observers, N. Y., Wiley, 1964.
 112. Swets J. A., Is there a sensory threshold? *Science*, 134, 168—77 (July 21, 1961), Copyright 1961 by the American Association for the Advancement of Science.
 113. Swets J. A., Green D. M., Sequential observations of signal and noise. in: Cherry, Colin (Ed.), Information Theory, Butterworth & Co. (Publishers) Limited, 1961.
 114. Swets J. A., Tanner W. P., Jr., Birdsall T. G., Decision processes in perception, *Psychol. Rev.*, 68, 301—340 (Sept. 1961).
 115. Swets J. A., Feurzeig W., Computer-aided instruction, *Science*, 150, 572—576 (Oct. 29, 1965), Copyright 1965 by the American Association for the Advancement of Science.
 116. Taylor D. W., Cognitive processes in solving algebra word problems. in: Kleinmuntz B. (Ed.), Problem Solving, N. Y., Wiley, 1966, p. 120.
 117. Templeton A. W., Lehr J. L., Simmons C., The computer evaluation and diagnosis of congenital heart disease using roentgenographic findings, *Radiology*, 87, 658—670 (Oct. 1966).
 118. Tolles W. E., Horvath W. J., Bostrom R. C., A study of the quantitative characteristics of exfoliated cells from the female genital tract. II. Suitability of quantitative cytological measurements for automatic prescreening, *Cancer*, 14, 455—468 (May—June 1961).
 119. Toronto A. F., Veasy L. G., Warner H. R., Evaluation of a computer program for diagnosis of congenital heart disease, *Prog. Cardio Dis.*, 5, 362—377 (Jan. 1963).
 120. Uhr L., «Pattern recognition» — computers as models for form perception, *Psychol. Bull.*, 60, 40—73 (Jan. 1963).
 121. Vecchio T. J., Predictive value of a single diagnostic test in unselected populations, *New Eng. J. Med.*, 274, 1171—1173 (May 26, 1966).
 122. vonNeumann J., The Computer and the Brain, New Haven, Yale, 1958.
 123. vonNeumann J., Morgenstern O., Theory of Games and Economic Behavior, Princeton, Princeton, 1947.
 124. Warner H. R., Toronto A. F., Veasy L. G., et al., A mathematical approach to medical diagnosis., Application to congenital heart disease, *JAMA*, 177, 177—183 (July 22, 1961).
 125. Warner H. R., Toronto A. F., Veasy L. G., Experience with Bayes theorem for computer diagnosis of congenital heart disease *Ann. N. Y. Acad. Sci.*, 115, 558—567 (July 31, 1964).
 126. Warner H. R., Pryor T. A., Johnson A. S., On-line classi-

- fication of vector-cardiograms for diagnosis of congenital heart disease. Supplement № II to *Circulation*, XXXII (4), 215, (Oct. 1965).
127. Warner H. R., Stauffer W. M., Some techniques for computer assisted diagnosis of heart disease, in: Proceedings on Automated-Data Processing in Hospitals, Elsinore, Denmark, April—May 1966, p. 362—74.
 128. Weiss W., Boucot K. R., Cooper D. A., The survival of men with measurable proved lung cancer in relation to growth rate, *Amer. J. Roentgen*, 98, 404—415 (Oct. 1966).
 129. Wiener N., Nonlinear prediction and dynamics, in: Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability, 1954—1955, Berkeley, U. of Calif., 1956, vol. 3, pp. 247—252.
 130. Wilson W. J., Templeton A. W., Turner A. H., Jr., et al., The computer analysis and diagnosis of gastric ulcers, *Radiology*, 85, 1064—1073 (Dec. 1965).
 131. Wilson W. J., личное сообщение, April 11, 1967.
 132. Winkler C., Reichertz P., Kloss G., Computer diagnosis of thyroid diseases, Comparison of incidence data and considerations of the problem of data collection, *Amer. J. Med. Sci.*, 253, 27—34 (1967).
 133. Wortman P. M., Representation and strategy in diagnostic problem solving, *IEEE Transactions on Human Factors in Electronics*, 8, 48—53 (1966).
 134. Yerushalmy J., Statistical problems in assessing methods of medical diagnosis with special reference to X-ray techniques, *Public Health Rep.*, 62, 1432—1449 (Oct. 1947).
 135. Yerushalmy J., Reliability of chest radiography in the diagnosis of pulmonary lesions, *Amer. J. Surg.*, 89(1), 231—240 (Jan. 1955).
 136. Yerushalmy J., Harkness J. T., Cope J. H., Kennedy B. R., Role of dual reading in mass radiography, *Amer. Rev. Tuberculosis*, 61, 443—464 (April 1950).
 137. 1958 Revision of Diagnostic Criteria for Rheumatoid Arthritis, *Arthritis Rheum.*, 2, 16—20 (1959).

ПРИЛОЖЕНИЕ 1

ЛИТЕРАТУРА ПО МАТЕМАТИЧЕСКИМ И СТАТИСТИЧЕСКИМ ПРОЦЕДУРАМ, ИСПОЛЬЗУЕМЫМ В МАШИННОЙ ДИАГНОСТИКЕ

I. ТЕОРЕМА БАЙЕСА И УСЛОВНЫЕ ВЕРОЯТНОСТИ

А. Врожденные пороки сердца

1. Warner H. R., Rotonto A. F., Veasy L. G., Stephenson R., A mathematical approach to medical diagnosis. Application to congenital heart disease, *JAMA*, 177, 177—183 (July 22, 1961).
2. Warner H. R., Toronto A. F., Veasy L. G., Experience with Bayes' theorem for computer diagnosis of congenital heart disease, *Ann, NY Acad. Sci.*, 115, 558—567 (July 31, 1964).
3. Warner H. R., Pryor T. A., Johnson A. S., On-line classification of vectorcardiograms for diagnosis of congenital heart disease, Supplement № II to *Circulation*, XXXII(4), 215, (Oct. 1965).
4. Амосов Н. М., Шкабара Е. А., Опыт диагностики с помощью диагностической машины, *Экспериментальная хирургия и анестезиология*, 4, 15—22 (1961).
5. Быховский М. Л., Вишневский А. А., Харнас С. Ш., Вопросы построения диагностического процесса при помощи математических машин. *Экспериментальная хирургия и анестезиология*. 4, 3—15 (1961).
6. Gustafson J. E., The computer — for use in private practice. In Proceedings of 5th IBM Medical Symposium, 99—111 (Oct. 1963).
7. Gustafson J. E., Throckmorton T. D., The computer as a consultant, *Med. Times*, 93, 115—119 (Feb. 1965).
8. Takahashi K., Logic of diagnosis and its processing by computer — with respect to congenital heart diseases and brain tumors, in: Proceedings on Automated-Data Processing in Hospitals, Elsinore, Denmark, April—May 1966, pp. 477—498.
9. Templeton A. W., Lehr J. L., Simmons C., The computer evaluation and diagnosis of congenital heart disease, using roentgenographic findings, *Radiology*, 87, 658—670 (Oct. 1966).

Б. Приобретенные пороки сердца

1. Bruce R. A., Computer diagnosis of heart disease, in: Proceedings of 5th IBM Medical Symposium, pp. 77—78 (1963).
2. Bruce R. A., Cox A. R., Medical diagnosis by computer, *Northwest Med.*, 60, 1075—1082 (Nov. 1961).
3. Bruce R. A., Yarnall S. R., Computer-aided diagnosis of cardiovascular disorders, *J. Chronic Dis.*, 19, 473—478 (1966).

В. Электрокардиография

1. Kimura E., Mibukura Y., Miura S., Statistical diagnosis of electrocardiogram by Theorem of Bayes, *Jap. Heart. J.*, 4, 469—488 (Sept. 1968).

Г. Синдром Кушинга

1. Nugent C. A., Warner H. R., Dunn J. T., Tyler F. H., Probability theory in the diagnosis of Cushing's syndrome, *J. Clin. Endocr.*, 24, 621—627 (1964).

Д. Функция щитовидной железы

1. Fitzgerald L. T., Overall J. E., Williams C. M., A computer program for diagnosis of thyroid disease, *Amer. J. Roentgen*, 97, 901—905 (August 1966).
2. Overall J. E., Williams C. M., Conditional probability program for diagnosis of thyroid function, *JAMA*, 183, 307—313 (Feb. 2, 1963).
3. Silink K., A new mathematical approach to endocrinological diagnosis, *Prog. Biocybernetics*, 1, 20—35 (1964).

Е. Желудочно-кишечный тракт

1. Rinaldo J. A., Scheinok P., Rupe C. E., Symptom diagnosis, A mathematical analysis of epigastric pain, *Ann. Int. Med.*, 59, 145—154 (August 1963).
2. Wilson W. J., Templeton A. W., Turner A. H., Jr., Lodwick G. S., The computer analysis and diagnosis of gastric ulcers, *Radiology*, 85 1064—1073 (Dec. 1965).

Ж. Костные опухоли

1. Lodwick G. S., Haun C. L., Smith W. E., Keller R. F., Robertson E., Computer diagnosis of primary bone tumors, *Radiology*, 80, 273—275 (Feb. 1963).
2. Lodwick G. S., Computer analysis of tumor roentgenograms, in: Proceedings of 5th IBM Medical Symposium, pp. 213—224 (Oct. 1963).
3. Lodwick G. S., Turner A. H., Jr., Lusted L. B., Templeton A. W., Computer-aided analysis of radiographic images, *J. Chronic. Dis.*, 19(4), 485—496 (April 1966).

З. Невропатология и психиатрия

1. Birnbaum A., Maxwell A. E., Classification procedures based on Bayes formula, *Appl. Statistics*, 9, 152—169 (1961).
2. Мисюк Н. С., Лепешинский Н. А., Лисковец О. А., Мاستикин А. С., Диагностика опухолей мозга при помощи ЭВМ «Урал-1». *Журнал неврологии и психиатрии*, 64, 453—458 (1964).
3. Overall J. E., Gorham D. R., A pattern probability model for the classification of psychiatric patients, *Behav. Sci.*, 8, 108—116 (1963).
4. Takahashi K., Logic of diagnosis and its processing by computer — with respect to congenital heart diseases and brain tumors, in: Proceedings on Automated-Data Processing in Hospitals, Elsinore, Denmark, April—May 1966, pp. 477—498.

II. ОТНОШЕНИЕ ПРАВДОПОДОБИЯ

А. Гематология

1. Lipkin M., The role of data processing in the diagnostic process, in: Jacquez, John A. (Ed.), The Diagnostic Process, Proceedings of a Conference

- held at The University of Michigan, May 1963, Ann Arbor, Malloy Lithographing, Inc., 1964, pp. 255—273.
2. Lipkin M., The likelihood concept in differential diagnosis, *Perspect. in Biology Med.*, 7, 483—497 (Summer 1964).

Б. Разделение на группы и диагностика

1. Collen M. F., Rubin L., Neyman J., et al., Automated screening and diagnosis, *Amer. J. Public Health.*, 54, 741—750 (May 1964).
2. Baer R. M., A Note on computer diagnosis, *IEEE Transactions on Biomedical Engineering*, BME-11, 8—12 (Jan.—April 1964).
3. Rubin L., Collen M. F., Goldman G. E., Frequency decision theoretical approach to automated medical diagnosis, in: Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, U. of Calif., 1967, vol. IV, pp. 867—886.

В. Перечень входных медицинских данных для диагностики (дальнейшее развитие Корнеллского перечня входных медицинских данных)

1. Brodman K., van Woerkom A. J., Computer-aided diagnostic screening for 100 common diseases, *JAMA*, 197, 901—905 (Sept. 12, 1966).

III. ДИСКРИМИНАНТНЫЕ ФУНКЦИИ

А. Диабет

1. Bogdanik T., The application of mathematical discrimination in the differential diagnosis of two types of diabetes, *Progr. Biocybernetics*, 1, 45—54 (1964).

Б. Функция щитовидной железы

1. Overall J. E., Williams C. M., Models for medical diagnosis: factor analysis, I. Theoretical, *Medizinische Dokumentation*, 5, 51—56 (April 1961), II. Experimental, *Medizinische Dokumentation*, 5, 78—80 (July 1961).
2. Wayne E. J., Clinical and metabolic studies in thyroid disease, *Brit. Med. J.* № 5165, 1—11 (Jan. 2, 1960) and № 5166, 78—90 (Jan. 9, 1960).

IV. ФАКТОРНЫЙ АНАЛИЗ

А. Функция щитовидной железы

1. Overall J. E., Williams C. M., Comparison of alternative computer models for thyroid diagnosis, in: Proceedings of Third San Diego Symposium for Biomedical Engineering, La Jolla, 1963, pp. 141—146.

Б. Психиатрия

1. Overall J. E., A configural analysis of psychiatric diagnostic stereotypes, *Behav. Sci.*, 8 (3), 211—219 (1963).

В. Использование Корнеллского перечня входных медицинских данных

1. van Woerkom A. J., Brodman K., Statistics for a diagnostic model, *Biometrics*, 17, 299—318 (June 1961).

Г. Урология

1. Zinsser H. H., Patient derived autodeinition of a disease: Pyelonephritis, in: Jacquez, John A. (Ed.), *The Diagnostic Process, Proceedings of a Conference held at The University of Michigan, May 1963, Ann Arbor, Malloy Lithographing, Inc., 1964, pp. 103—128.*

V. ПРОЦЕДУРЫ СРАВНЕНИЯ**А. Дерматология**

1. Nогins A. L., Computers in dermatology, *Arch. Derm.* (Chicago), **90**, 506—511 (Nov. 1964).

Б. Гематология

1. Lipkin M., Hardy J. D., Mechanical correlation of data in differential diagnosis of hematological diseases, *JAMA*, **166** (2), 113—125 (Jan. 11, 1958).
2. Lipkin M., Engle R. L., Jr., Davis B. J., et al., Digital computer as an aid to differential diagnosis, Use in hematologic diseases, *Arch. Int. Med.*, **108**, 56—72 (July 1961).

В. Рак легких

1. Feinstein A. R., Koss N., Prognosis by computers based on storage and retrieval of data for 5-year clinical course of 691 cases of lung cancer, *J. Clin. Invest.*, **45** (6), 1007 (1966) (abstract).

Г. Неврология

1. Штабцов В. И., Остроумов Ю. А., Опыт диагностики нервных болезней при помощи ЭЦВМ «Урал-2», *Журнал неврологии и психиатрии*, **64**, 515—519 (1964).

Д. Персоналистские оценки

1. Rome H. P., Automation techniques in personality assessment. The problem and procedure, *JAMA*, **182**, 1069—1072 (Dec. 15, 1962).
2. Pearson J. S., Swenson W. M., Rome H. P., et al., Further experience with the automated Minnesota Multiphasic Personality Inventory, *Proc. Mayo Clin.*, **39**, 823—829 (1964).
3. Swenson W. M., Pearson J. S., Psychiatry — Psychiatric Screening, *J. Chronic. Dis.*, **19**, 507 (1966).

Е. Хирургия

1. Peterson O. L., Barsamian E., Diagnostic performance, in: Jacquez, John A. (Ed.), *The Diagnostic Process, Proceedings of a Conference held at The University of Michigan, May 1963, Ann Arbor, Malloy Lithographing, Inc., 1964, pp. 347—362.*

ПРИЛОЖЕНИЕ 2

ИЗБРАННАЯ БИБЛИОГРАФИЯ ПО ОШИБКАМ НАБЛЮДЕНИЯ

А. Работы по общим проблемам ошибок наблюдения

1. Bean W. B., The natural history of error, Pseudodoxia endemica, *AMA Arch. Int. Med.*, 105, 184—193 (Feb. 1960).
2. Fletcher C. M., Observer error, *Lancet*, 43, 212 (Jan. 23, 1954).
3. Garland L. H., Studies on the accuracy of diagnostic procedures, *Amer. J. Roentgen*, 82 (1), 25—38 (July 1959).
4. Garland L. H., The problem of observer error, *Bull. N. Y. Acad. Med.*, 36, 570—584 (Sept. 1960).
5. Horvath W. J., The effect of physician bias in medical diagnosis, *Behav. Sci.*, 9, 334—340 (1964).
6. Kahn M. W., Clinical and statistical prediction revisited, *J. Clin. Psychol.*, 16, 115—118 (April 1960).
7. Lusted L. B., Logical analysis and roentgen diagnosis, *Radiology*, 74, 178—193 (Feb. 1960).
8. Meehl P. E., On validation of clinical procedures, *Canad. J. Psychol.*, 13, 102 (1959).
9. Nacke O., Wagner G., Bibliography on the topic: The role of error in medicine; quality control as a task of medical documentation, *Meth. Inform. Med.*, 3 (3/4), 132—150 (July-Oct. 1964). (Note: This is the most complete list of observer error articles that I have found.)
10. Saiger G. L., Observations on the probability of error in medical diagnosis, *Ann. Int. Med.*, 56, 860—864 (June 1962).
11. Scheeff T. J., Decision rules, types of error, and their consequences in medical diagnosis, *Behav. Sci.*, 8, 97—107 (1963).

Б. Ошибки наблюдателя в кардиологии, гастроэнтерологии и при диагностике легочных заболеваний

1. Butterworth J. S., Reppert E. H., Auscultatory acumen in the general medical population, *JAMA*, 174, 32—34 (Sept. 3, 1960).
2. Conn H. O., Smith H. W., Brodoff M., Observer variation in the endoscopic diagnosis of esophageal varices. A prospective investigation of the diagnostic validity of esophagoscopy, *New Eng. J. Med.*, 272, 830—834 (April 22, 1965).
3. Edwards W. M., Cox R. S., Jr., Garland L. H., The solitary nodule (coin lesion) of the lung, An analysis of 52 consecutive cases treated by thoracotomy and a study of preoperative diagnostic accuracy, *Amer. J. Roentgen*, 88, 1020—1042 (Dec. 1962).
4. Fletcher C. M., The clinical diagnosis of pulmonary emphysema, *Proc. Roy. Soc. Med.*, 45, 577 (1952).
5. King D. J., Manecold R. F., Consistency of cardiologists' prognostic judgments in cases of myocardial infarction, *Amer. J. Cardiol.*, 15, 27—32 (Jan. 1965).

6. Levine S. A., Notes on the graduation of the intensity of cardiac murmurs, *JAMA*, 177, 261 (July 29, 1961).
7. Toronto A. F., Veasy G., Warner H. R., Evaluation of a computer program for diagnosis of congenital heart disease, *Prog. Cardio. Dis.*, 5, 362—377 (Jan. 1963).
8. Warner H. R., Pryor T. A., Johnson A. S., On-line classification of vectorcardiograms for diagnosis of congenital heart disease, Supplement № II to *Circulation*, XXXII, (4), 215, Oct. 1965.

В. Ошибки наблюдателя в педиатрии

1. Bakwin H., Pseudodoxia pediatrica, *New. Eng. J. Med.*, 232, 691—697 (June 14, 1945).

Г. Ошибки наблюдателя в психиатрии

1. Birnbaum A., Maxwell A. E., Classification procedures based on Bayes' formula, *Appl. Statistics*, 9, 152 (1961).
2. Grosz H. J., Grossman K. G., The sources of observer variation and bias in clinical judgments, I. The item of psychiatric history, *J. Nerv. Ment. Dis.*, 138 (2), 105—113 (Feb. 1964).

Д. Ошибки наблюдателя в рентгенологии

1. Amberg J. R., Accuracy of roentgen diagnosis in carcinoma of the stomach, *Amer. J. Dig. Dis.*, 5, 259—263 (March 1960).
2. Arnois D. C., Silverman F. N., Turner M. E., et al., The radiographic evaluation of pulmonary vasculature in children with congenital cardiovascular disease, *Radiology*, 72 (5), 689—698 (May 1959).
3. Baker H. L., Jr., Hodgson J. R., Further studies on the accuracy of oral cholecystography, *Radiology*, 74, 239—245 (Feb. 1960).
4. Berkson J., Good C. A., Carr D. T., et al., Identification of „positives” in roentgenographic readings, *Amer. Rev. Resp. Dis.*, 81, 660—665 (May 1960).
5. Bland J. H., VanBuskirk F. W., Tampas J. F., et al., A study of roentgenologic criteria for rheumatoid arthritis of the cervical spine, *Amer. J. Roentgen*, 95, 949—954 (Dec. 1965).
6. Coon H. O., Greenspan R. H., Clemett A. R., et al., Balloon tamponade in the radiological diagnosis of esophageal varices, *Gastroenterology*, 51, 29—40 (Jan. 1966).
7. Cooley R. N., Agnew C. H., Rios G., Diagnostic accuracy of the barium enema study in carcinoma of the colon and rectum, *Amer. J. Roentgen*, 84, 316—331 (August 1960).
8. Etter L. E., Dunn J. P., Kammer A. G., et al., Gastroduodenal X-ray diagnosis: A comparison of radiographic techniques and interpretations, *Radiology*, 74, 766—770 (May 1960).
9. Frieden J., Shapiro J. H., Feinstein A. R., Radiologic evaluation of heart size in rheumatic heart disease. Studies in young patients, *Arch. Int. Med.*, 111, 44—50 (Jan. 1963).
10. Gricksman A. S., Chu F. C., Bane H. N., Nickson J. J., Quantitative and qualitative evaluation of skin erythema. II. Clinical study in patients on a standardized irradiation schedule, *Radiology*, 75, 411—415 (Sept. 1960).
11. Gurland J., Johnson R. O., How reliable are tumor measurements?, *JAMA*, 194, 973—978 (Nov. 29, 1965).

12. Knott J. M. S., Christie R. V., Radiological diagnosis of emphysema, *Lancet*, **1** (16), 881—883 (April 21, 1951).
13. Powers W. E., Holtz S., Ogura J., et al., Contrast examination of larynx and pharynx; accuracy and value in diagnosis, *Amer. J. Roentgen*, **86**, 651—660 (Oct. 1961).
14. Sloan R. D., Roentgenologic evaluation of mesenteric small intestinal obstruction; a statistical analysis, *Amer. J. Roentgen*, **82**, 978—984 (Dec. 1959).
15. Smith M. J., Errors in diagnostic radiology; on the basis of complacency, *Amer. J. Roentgen*, **94**, 689—703 (July 1963).
16. Sosman M. C., The Shattuck Lecture: the specificity and reliability of roentgenographic diagnosis, *New Eng. J. Med.*, **242**, (22) 849—855 (June 1, 1950).
17. Standjord N. M., Moseley R. D., Jr., Schweinefus R. L., et al., Gastric carcinoma: Accuracy of radiologic diagnosis, *Radiology*, **74**, 442—451 (March 1960).
18. Stein G. N., Martin R. D., Roy R. H., et al., Evaluation of the conventional roentgenographic techniques for demonstration of duodenal ulcer craters, *Amer. J. Roentgen*, **91**, 801—807 (April 1964).
19. Stradling P., Johnston R. N., Reducing observer error in a 70-mm chest-radiography service for general practitioners, *Lancet*, **1**, 1247 (June 18, 1955).

Е. Ошибки наблюдателя в задачах разделения на группы

1. Blumberg M. S., Evaluating health screening procedures. *Operations Res.*, **5**, 351, 1957.
2. Chiang C. L., Hodges J. L., Yerushalmy J., Statistical problems in medical diagnosis, in: *Proceedings of Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, U. of Calif., 1956, vol. IV, pp. 121—133.
3. Neyman J., Outline of statistical treatment of the problem of diagnosis, *Public Health Rep.*, **62**, 1449—1456 (Oct. 3, 1947).
4. Nissen-Meyer S., Evaluation of screening tests in medical diagnosis, *Biometrics*, **20**, 730—755 (Dec. 1964).
5. Spencer C. C., Bostrom R. C., Performance on the cytoanalyzer in recent clinical trials, *J. Nat. Cancer Inst.*, **29**, 267—276 (August 1962).

Замечание: это исследование ошибок наблюдателя при использовании экспериментального оборудования (комбинации сканирующего и вычислительного устройств), разработанного для разделения цитологических образцов.

ПРИЛОЖЕНИЕ 3

МАШИННАЯ ДИАГНОСТИКА ЗАБОЛЕВАНИЙ ЩИТОВИДНОЙ ЖЕЛЕЗЫ

Настоящий материал является частью книги «Машинная диагностика заболеваний щитовидной железы», написанной Т. Фитцджералдом и К. М. Вильямсом из Отдела рентгенологии медицинского колледжа Университета Флориды (часть II. «Методы использования вычислительных машин») [31].

ОБЩАЯ ИНФОРМАЦИЯ О ПРОГРАММАХ

Программы IV (стр. 241—257) и V (стр. 258—276) для заболеваний щитовидной железы содержат блоки проверки ошибок в листе входных данных и ошибок в перфокартах. Вычислительная машина осуществляет проверку на отрицательные входные данные и входные данные, не соответствующие определенным пределам. Выполняются также проверки на пропуск необходимой информации и на противоречивость входных данных. Входные данные, относящиеся к предыдущему диагнозу, проверяются на наличие неприемлемых величин. Если обнаружена ошибка, то машина печатает тип ошибки вместе с номером соответствующего пациента; затем машина рассматривает следующий случай.

Программы IV и V предусматривают выдачу результата в виде двойной печати, т. е. на каждой половине листа печатается полный диагноз.

Для кодирования симптомов пациента удобно использовать цифры «0» и «1». Однако при этом необходимо делать различие между пробивкой «0» и пустым местом. Для этого в применявшейся библиотечной подпрограмме XSIGNF машины IBM-709 пропуск читался как «-0».

Если не встретилось ни одного пациента с некоторым признаком, симптомом или лабораторным тестом, то в качестве входной величины в соответствующем месте матрицы симптомы — заболевания указывают значение 0,001, а не 0.

РАСПОЛОЖЕНИЕ ДАННЫХ НА ПЕРФОКАРТАХ ДЛЯ ПРОГРАММЫ IV (СМ. ТАБЛИЦУ)

- Карта 1. Основные частоты для пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 3F10.3.
- Карты 2—10. Вероятности для отсутствующих признаков и симптомов и для присутствующих признаков и симптомов в случае пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 12F6.3.
- Карты 11—17. Вероятности для пяти видов лабораторных исследований в случае пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 12F6.3.

Карты данных пациента, пробитые в соответствии с листом входных данных, вводятся вслед за картами данных, приведенными выше. После того как поставлен диагноз последнему пациенту, программа заканчивает работу.

Программа V для заболеваний щитовидной железы имеет накапливающую матрицу; это означает, что признаки и симптомы пациента вместе с лабораторными исследованиями вводятся в эту матрицу данных после того, как поставлен диагноз. Например, предположим, что пациенту был поставлен диагноз «пониженная функция щитовидной железы». Тогда накапливающаяся матрица «уточняется» следующим образом: к общему числу случаев, диагностированных до сих пор, добавляется единица и, кроме того, к общему числу пациентов, имеющих пониженную функцию щитовидной железы, также добавляется единица; таким образом, основным частотам придаются новые значения. Для каждого непустого признака и симптома (признак либо присутствует, либо отсутствует) добавляется единица к общему числу случаев с пониженной функцией щитовидной железы, для которых этот признак или симптом отмечен (либо он есть, либо его нет). Если некоторая позиция во входных данных является пустой, то не производится никаких операций. Для каждого присутствующего признака или симптома (на входе «1») добавляется единица к общему числу случаев с пониженной функцией щитовидной железы, для которых этот признак или симптом имеет место. Для каждого присутствующего лабораторного исследования добавляется единица к общему числу случаев с пониженной функцией щитовидной железы, где это лабораторное исследование имеет место, и, кроме того, добавляется единица к тем случаям с пониженной функцией щитовидной железы, которые попадают в один из пяти интервалов значений лабораторных исследований, соответствующий входным данным. Если в 72-м столбце карты данных пациента пробита цифра 3, то печатается матрица, накопленная с учетом этого случая. Если в 72-м столбце карты данных пациента пробита цифра 2, то накапливающаяся матрица печатается и пробивается в подходящем формате, чтобы ее можно было использовать в следующем расчете. Если же в 72-м столбце карты данных пациента пробита цифра 1, то накапливающаяся матрица только печатается.

Подпрограмма МАТРИЦА (стр. 277—280) «уточняет» матрицу данных после диагностики каждого нового случая. Подпрограмма ПЕЧАТЬ печатает накапливающуюся матрицу, а подпрограмма ПРОБИВКА пробивает накапливающуюся матрицу.

РАСПОЛОЖЕНИЕ ДАННЫХ НА ПЕРФОКАРТАХ ДЛЯ ПРОГРАММЫ V (СМ. ТАБЛИЦУ)

- Карты 1—17. Идентичны картам программы IV.
- Карта 18. Общее число случаев, диагностированных до сих пор, и число накопленных случаев пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 415.
- Карты 19—22. Число накопленных случаев с каждым признаком и симптомом, имеющим место для пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 1415.
- Карты 23—26. Число накопленных случаев для каждого признака и симптома либо присутствующего, либо отсутствующего для пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 1415.
- Карты 27—33. Число накопленных случаев в каждом из пяти интервалов лабораторного теста, имеющего место для пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 1216.
- Карта 34. Число накопленных случаев с лабораторным тестом, имеющим

Расположение данных на перфокартах для программы V

Номер карты	Столбец												
	10	20	30	40	50	60	70						
1	0,140	0,662	0,198										
2	0,944	0,999	0,867	0,971	0,875	0,999	0,999	0,986	0,153	0,187	0,742	0,647	
3	0,500	0,130	0,107	0,257	0,857	0,629	0,829	0,751	0,898	0,549	0,891	0,881	
4	0,779	0,950	0,841	0,990	0,820	0,850	0,962	0,911	0,970	0,950	0,085	0,258	
5	0,322	0,395	0,164	0,245	0,292	0,129	0,999	0,949	0,999	0,867	0,977	0,982	
6	0,991	0,992	0,700	0,056	0,001	0,133	0,029	0,125	0,001	0,001	0,014	0,847	
7	0,813	0,258	0,353	0,500	0,870	0,893	0,743	0,143	0,371	0,171	0,249	0,102	
8	0,451	0,109	0,119	0,221	0,050	0,159	0,010	0,180	0,150	0,038	0,089	0,030	
9	0,050	0,915	0,742	0,678	0,605	0,836	0,755	0,708	0,871	0,001	0,051	0,001	
10	0,133	0,023	0,018	0,009	0,008	0,300							
11	0,250	0,750	0,001	0,001	0,001	0,636	0,273	0,091	0,001	0,001	0,619	0,238	
12	0,143	0,001	0,001	0,609	0,250	0,141	0,001	0,001	0,667	0,333	0,001	0,001	
13	0,001	0,001	0,562	0,269	0,151	0,018	0,008	0,129	0,749	0,106	0,008	0,009	
14	0,099	0,774	0,108	0,009	0,009	0,119	0,733	0,129	0,011	0,033	0,066	0,836	
15	0,066	0,001	0,001	0,001	0,050	0,653	0,297	0,001	0,001	0,115	0,197	0,689	
16	0,001	0,001	0,001	0,299	0,701	0,001	0,008	0,016	0,205	0,772	0,001	0,001	
17	0,001	0,057	0,943										
18	879	123	582	174									

19	4	0	4	1	6	0	0	1	61	52	8	12	23	60	
20	67	55	5	129	55	72	31	164	32	35	68	17	51	3	
21	55	54	11	26	9	16	108	72	59	69	112	83	80	108	
22	0	5	0	15	3	2	1	1	30						
23	71	62	30	35	48	64	72	73	72	64	31	34	46	69	
24	75	74	35	348	321	289	305	364	293	293	308	341	321	290	
25	306	359	287	291	299	318	118	97	87	114	134	110	113	124	
26	119	98	89	113	133	109	112	127	100						
27	17	51	0	0	0	14	6	2	0	0	13	5			
28	3	0	0	39	16	9	0	0	14	7	0	0			
29	0	0	123	59	33	4	2	34	197	28	2	3			
30	33	257	36	3	4	54	335	59	5	2	4	51			
31	4	0	0	0	5	66	30	0	0	0	12	42			
32	0	0	0	20	47	0	1	2	26	98	0	0			
33	0	2	33												
34	68	22	21	64	21	219	263	332	457	61	101	61	67	127	35
35	0	0	0	0	0	0	0	0							

Карта 35. место для пониженной функции щитовидной железы, нормальной функции щитовидной железы и гиперфункции щитовидной железы соответственно, в формате 1515. Общее число новых случаев, обрабатываемых накапливающей программой и общее число случаев в каждом классе в соответствии с классификацией Всемирной организации здравоохранения, в формате 1415.

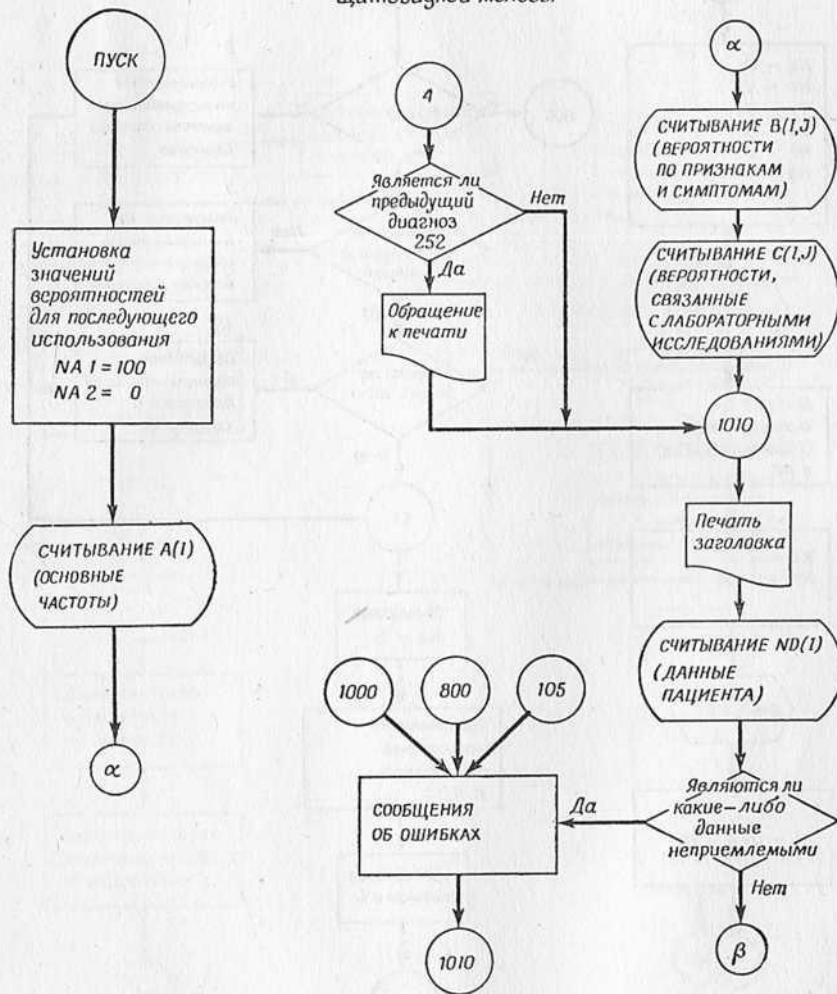
Ниже мы приводим блок-схемы программ IV и V для заболеваний щитовидной железы и подпрограмму МАТРИЦА. Подпрограммы КОНЕЦ, ПЕЧАТЬ и ПРОБИВКА весьма просты, и их блок-схемы здесь не приводятся.

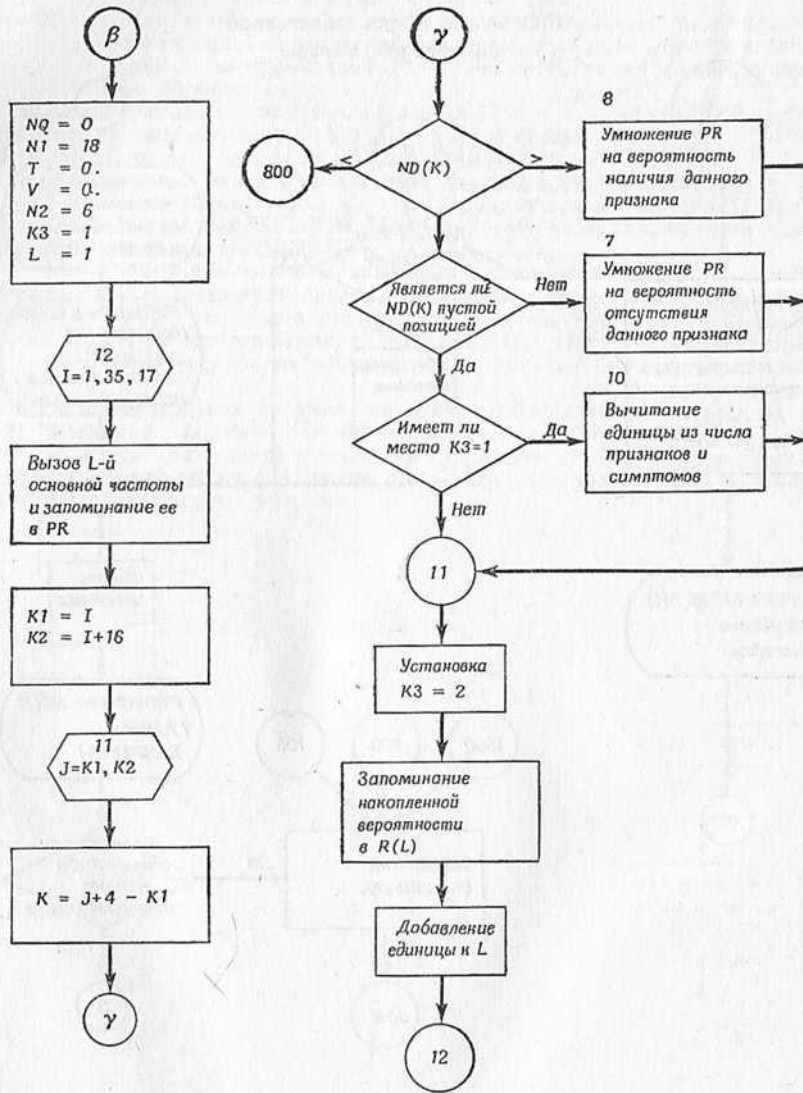
Вычислительный центр Университета Флориды использует отметку «r» в первой печатной позиции, чтобы вызвать пробивку элементов, определенных в команде ЗАПИСЬ НА ВЫХОДНОЙ ЛЕНТЕ. Эта управляющая отметка используется в подпрограмме ПРОБИВКА.

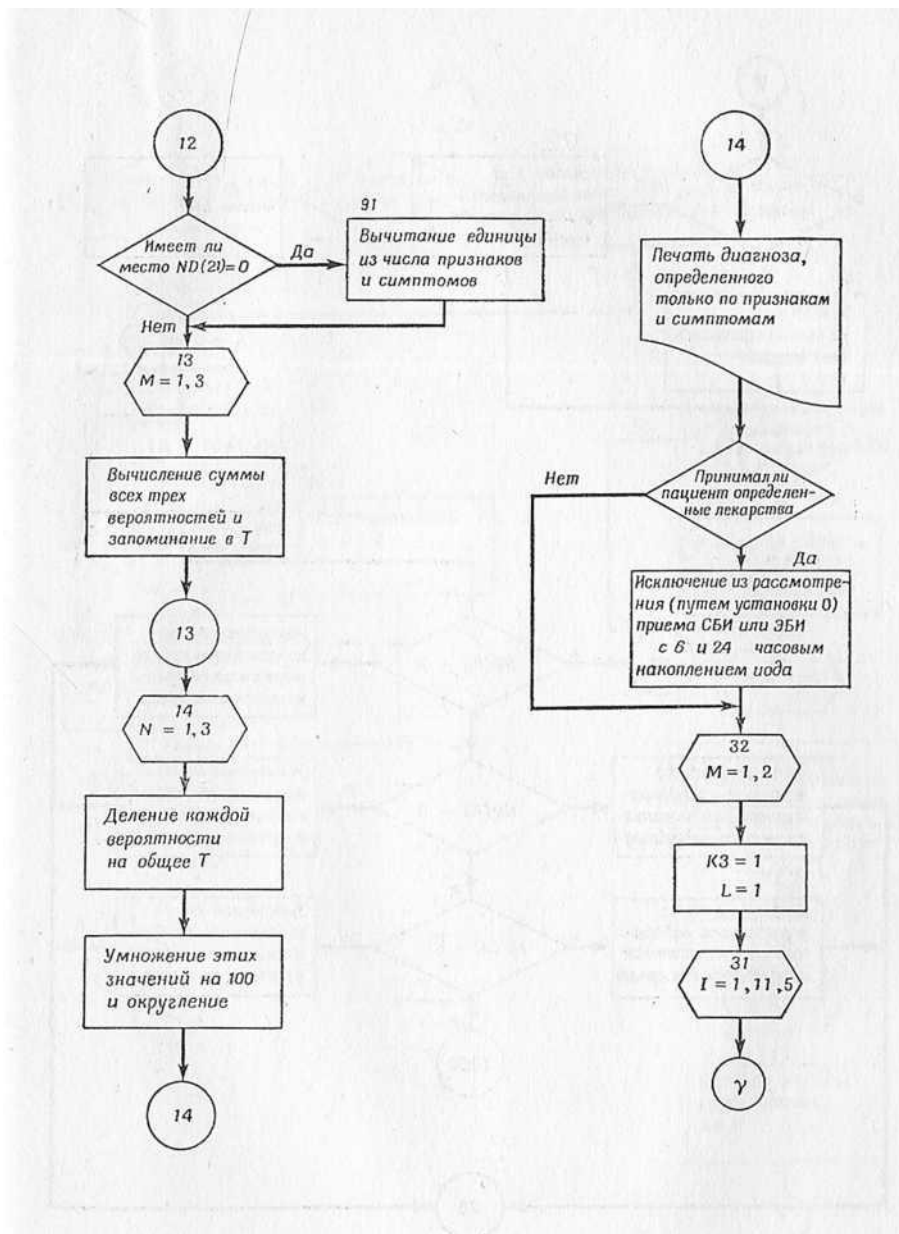
В случае ошибок в диагнозе необходимо исключить признаки, симптомы и лабораторные исследования неверно диагностированных случаев из накапливающейся матрицы. Это можно сделать при помощи программы ИСПРАВЛЕНИЕ ОШИБОК с подпрограммами ИСКЛЮЧЕНИЕ, ПРОБИВКА и ПЕЧАТЬ. При использовании исправляющей программы должны быть введены следующие данные:

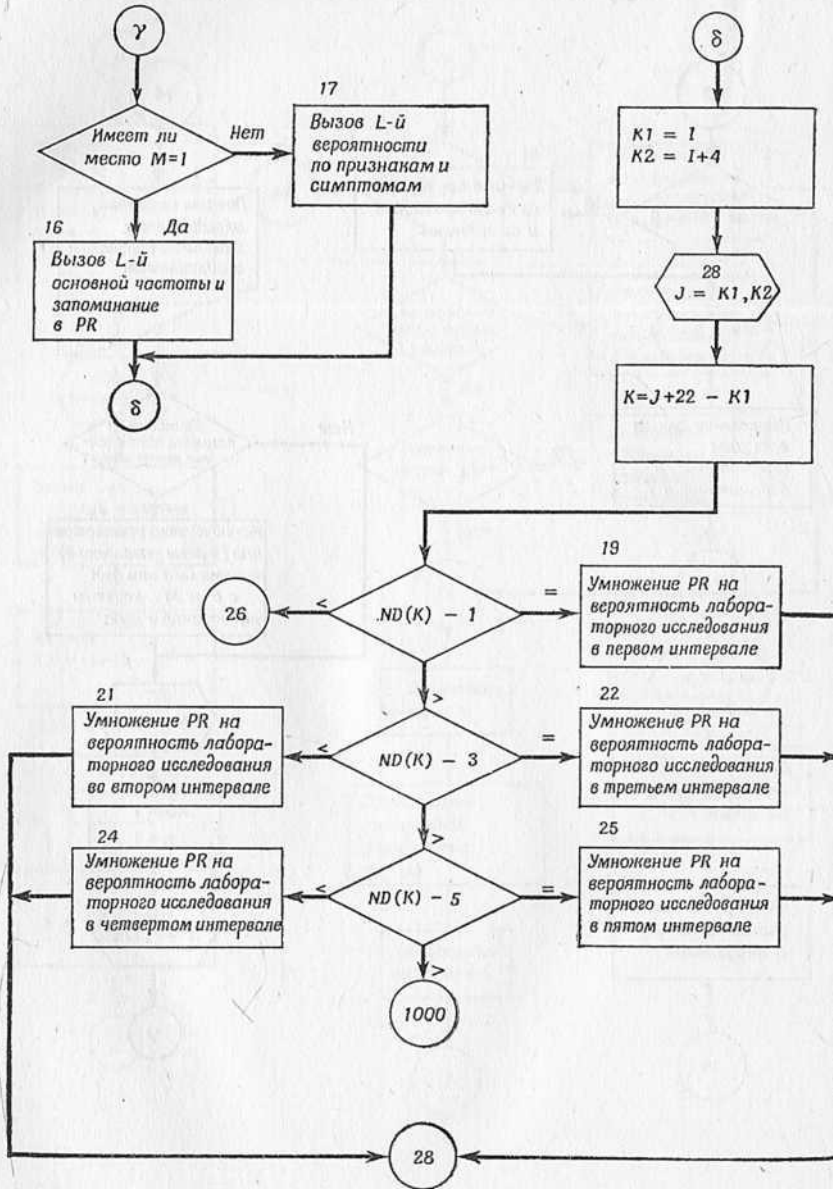
- 1) Накапливающаяся матрица, которая подлежит исправлению.
- 2) Машинный диагноз каждого пациента, который должен быть исключен, и соответствующая карта данных этого пациента. В 72-м столбце последней карты пробивается цифра 2, которая указывает, что должна быть напечатана и пробита исправленная матрица.

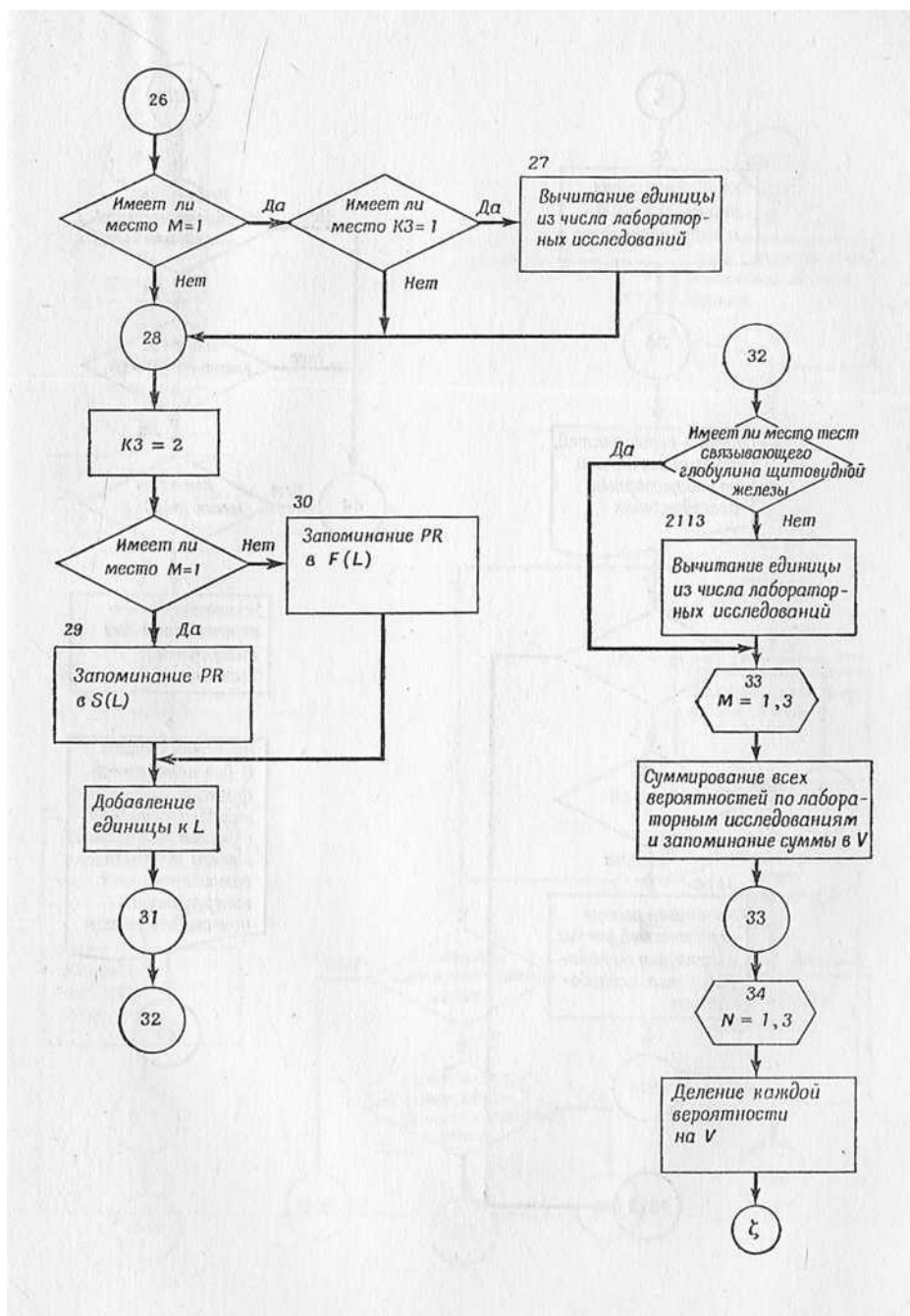
Программа IV для заболеваний щитовидной железы

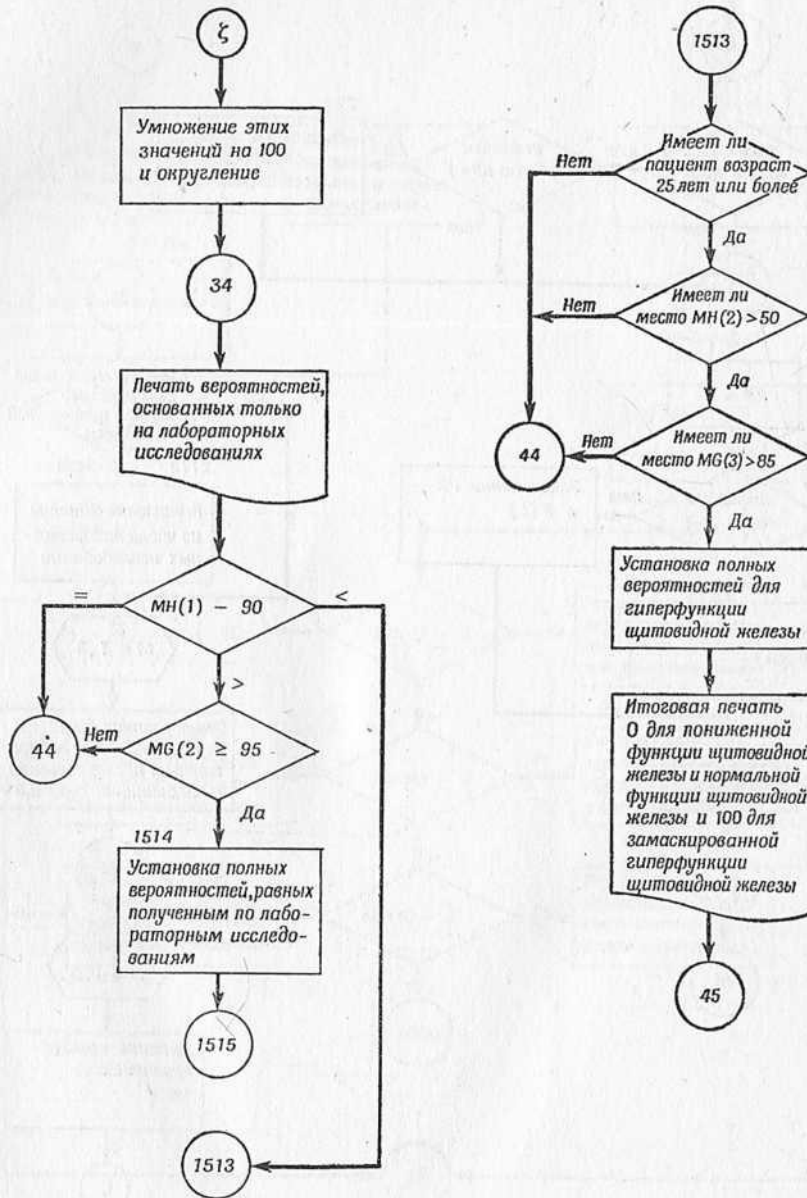


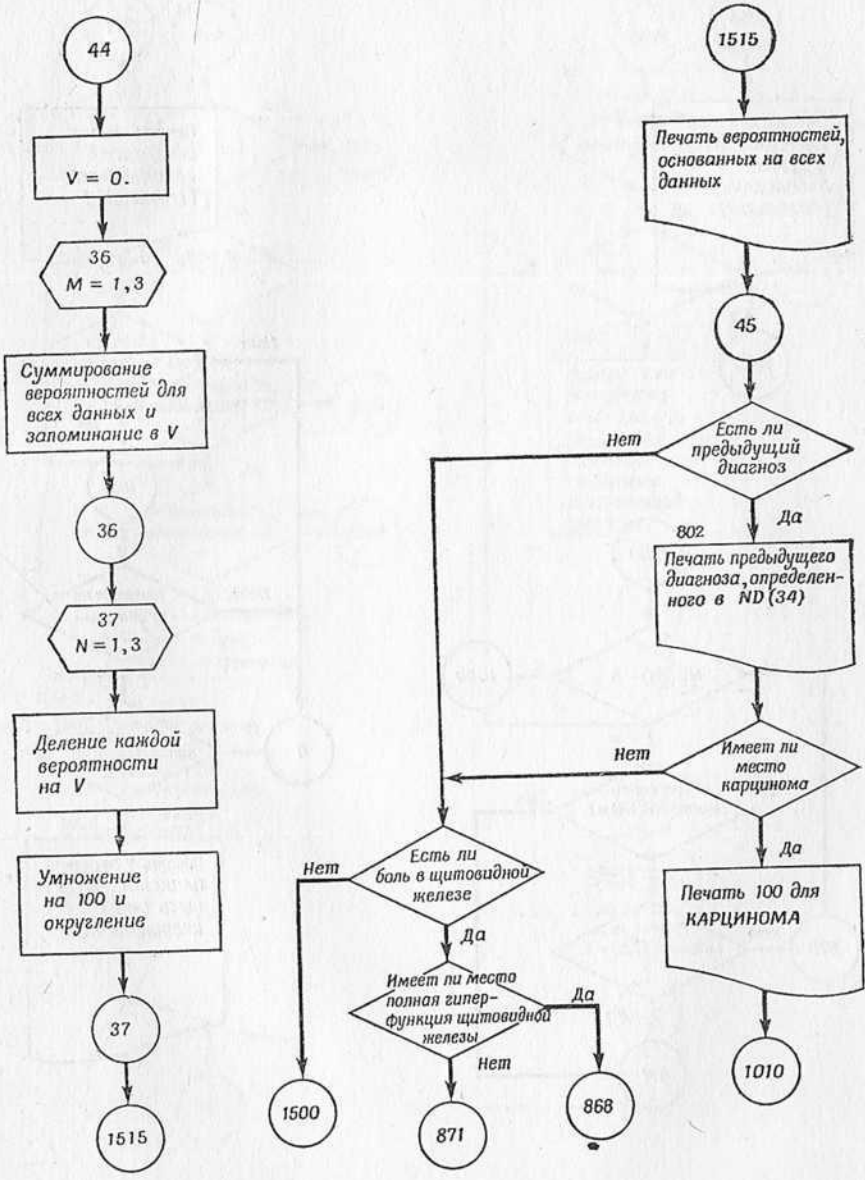


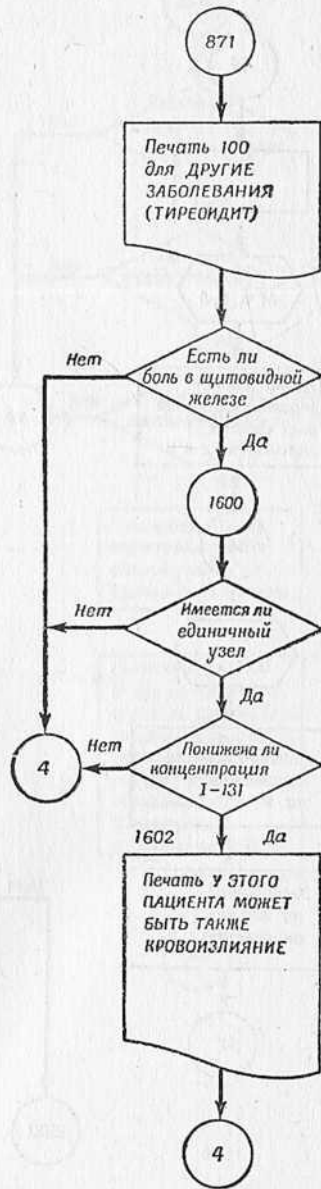
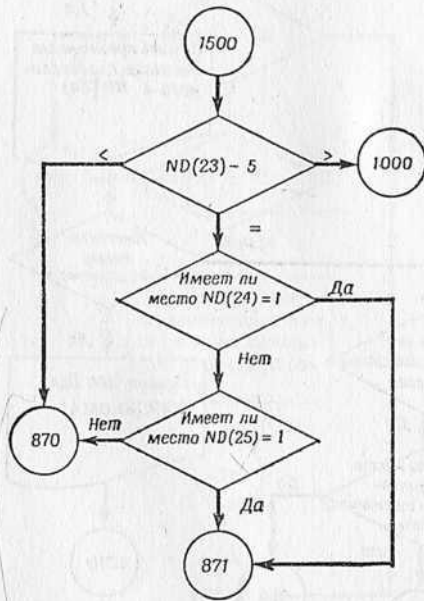


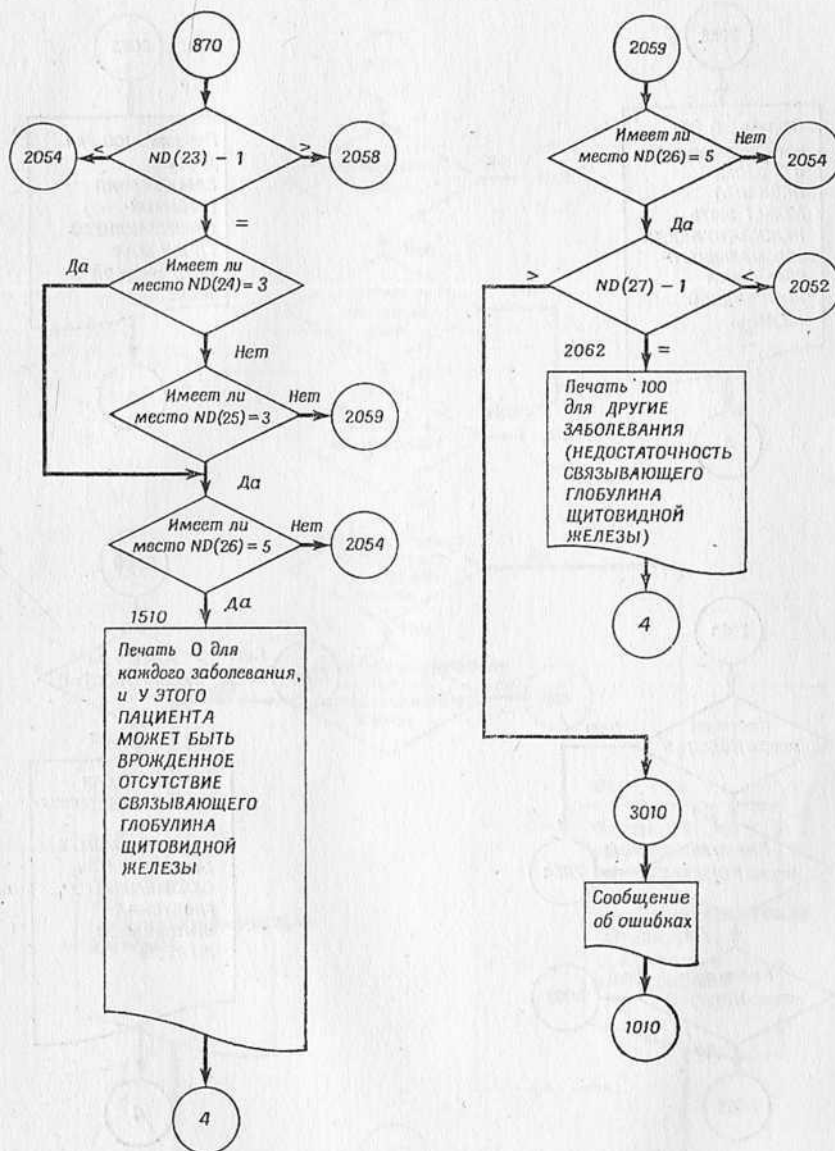


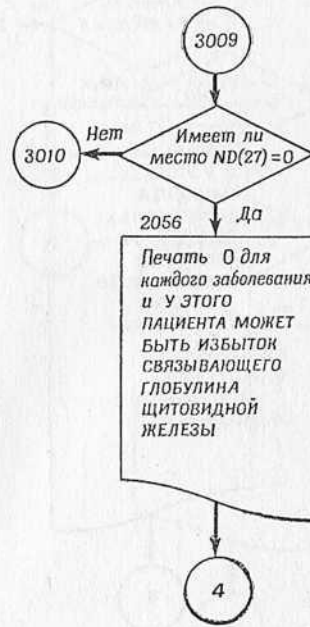
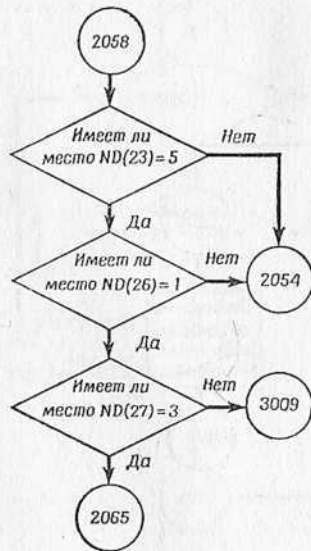
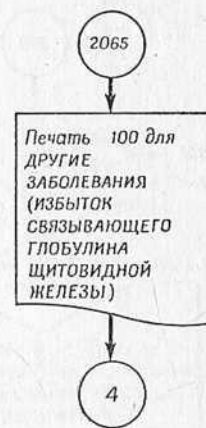
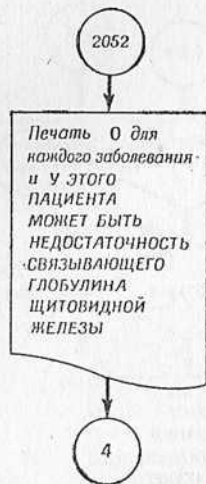


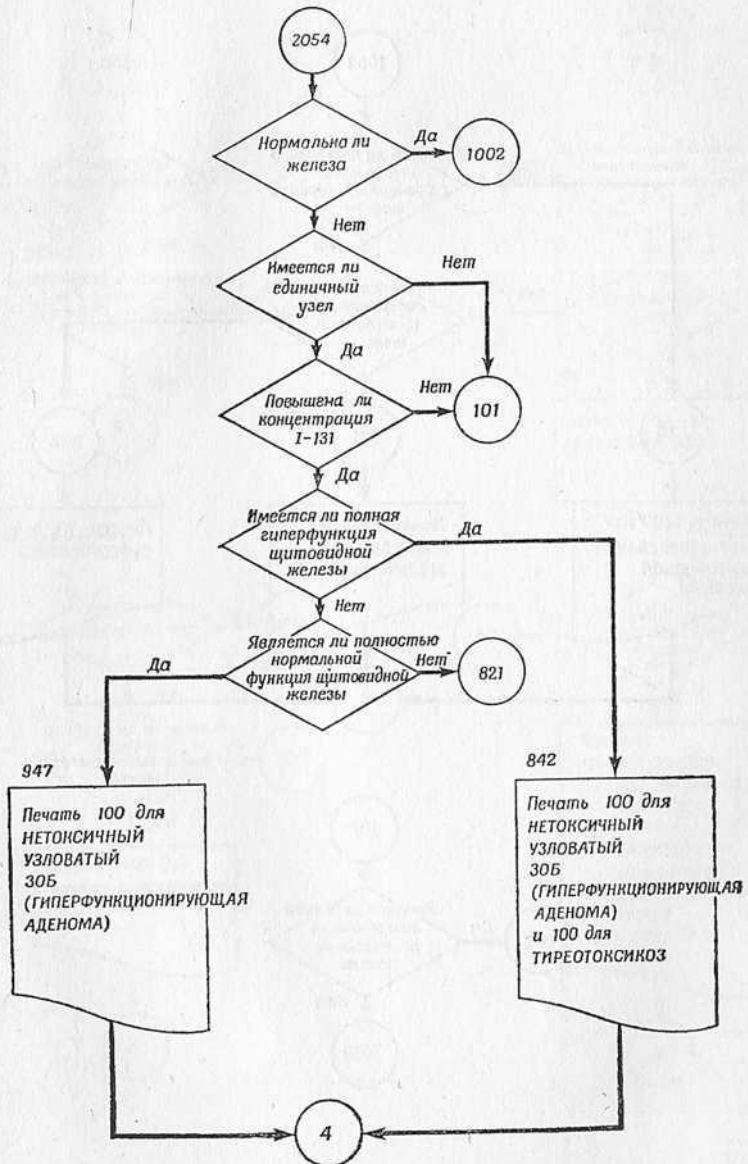


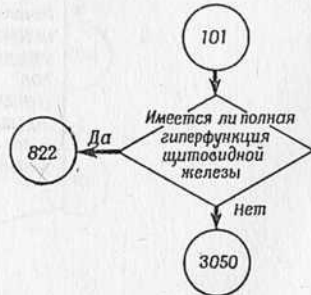
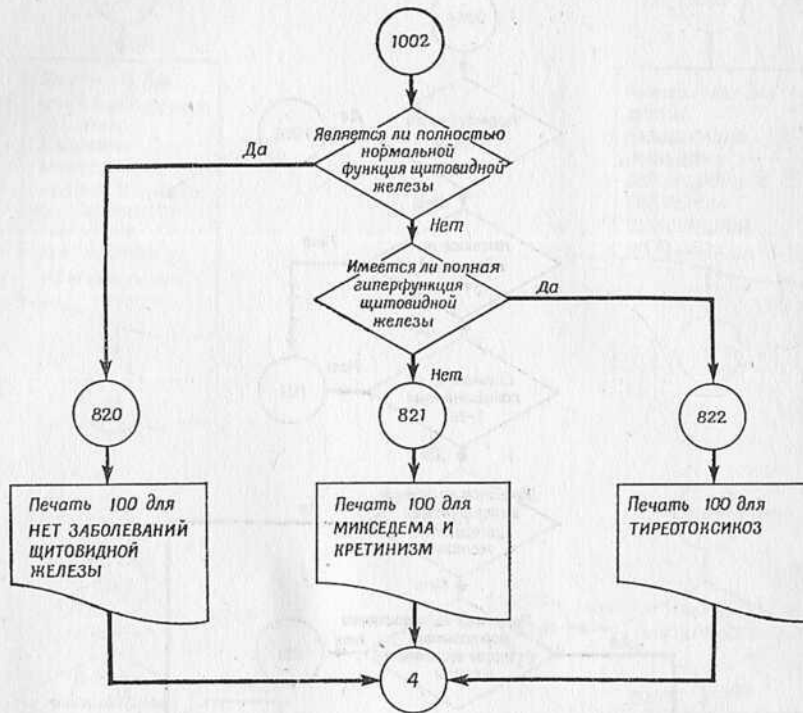


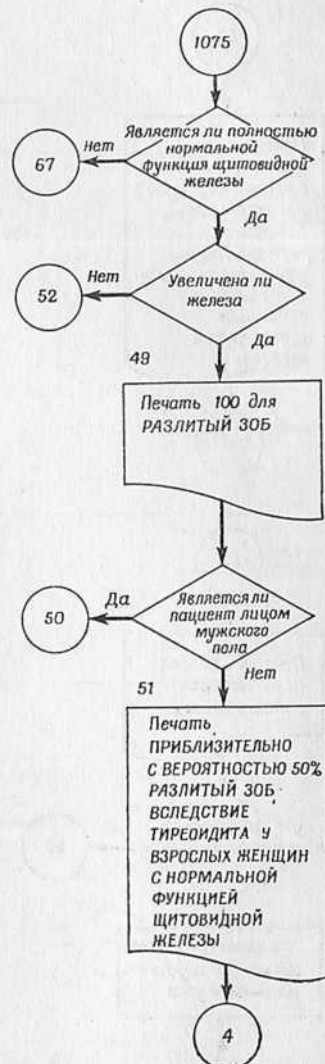
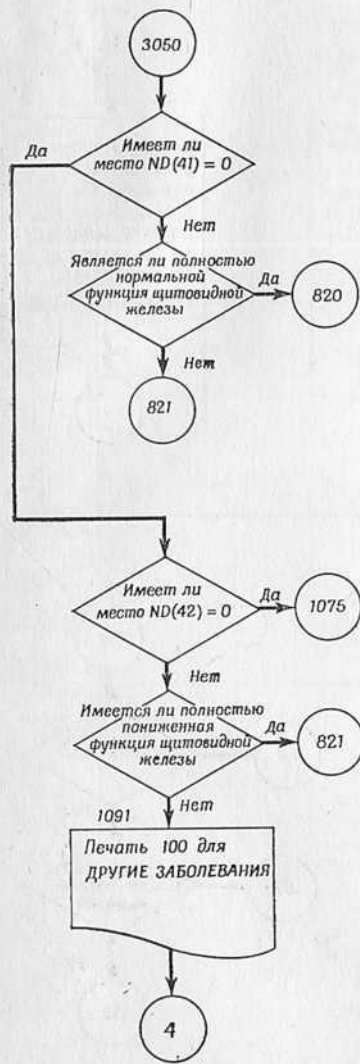


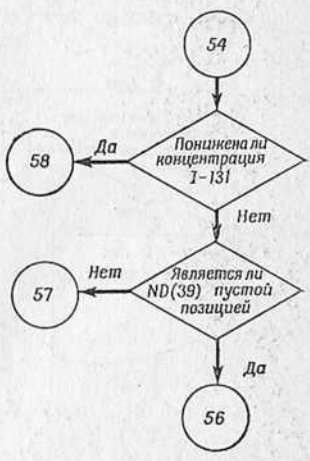
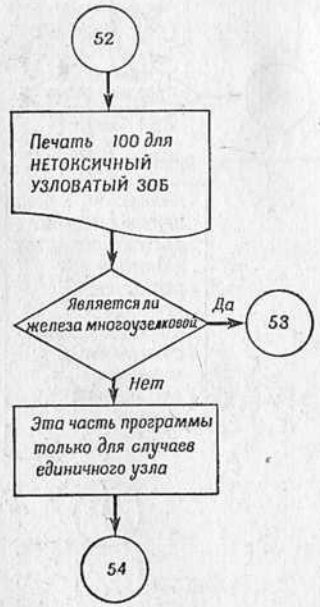
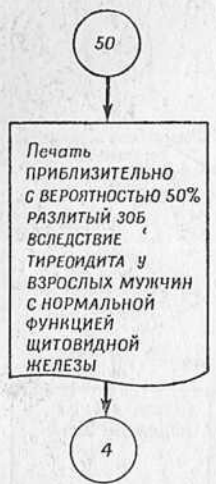


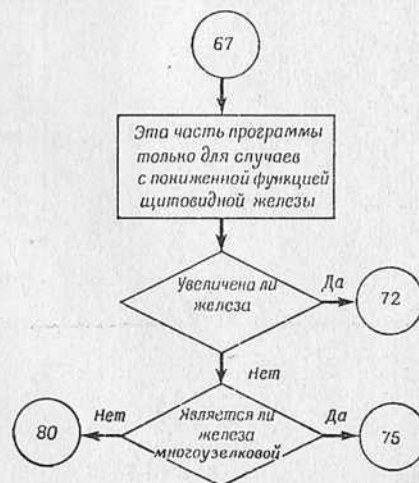
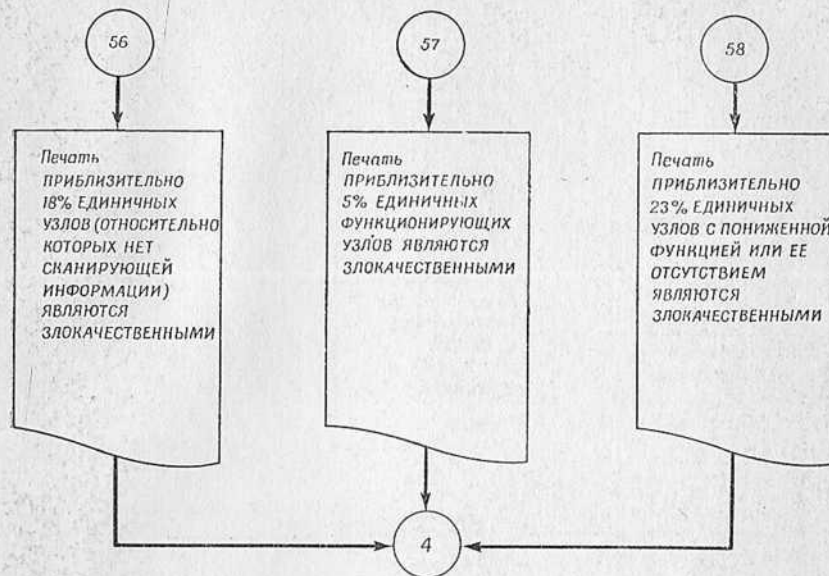


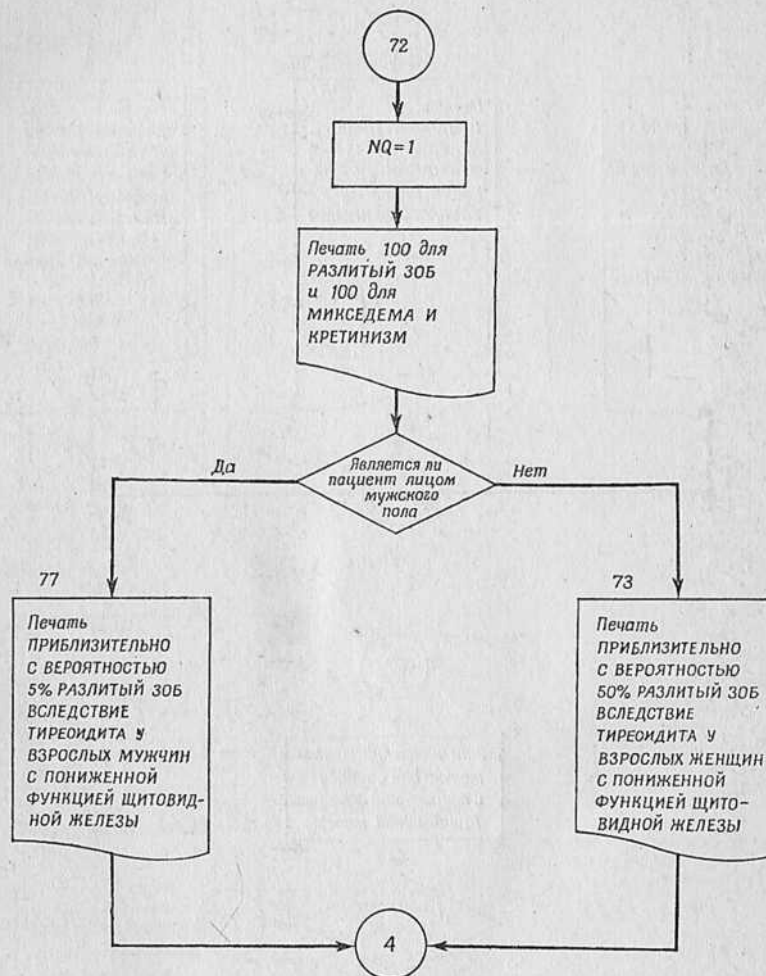


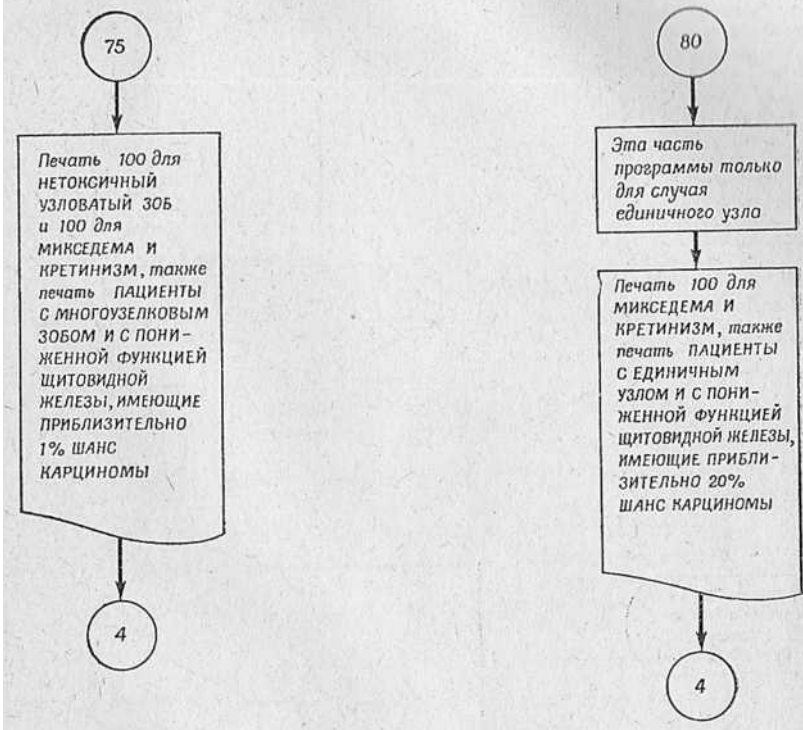




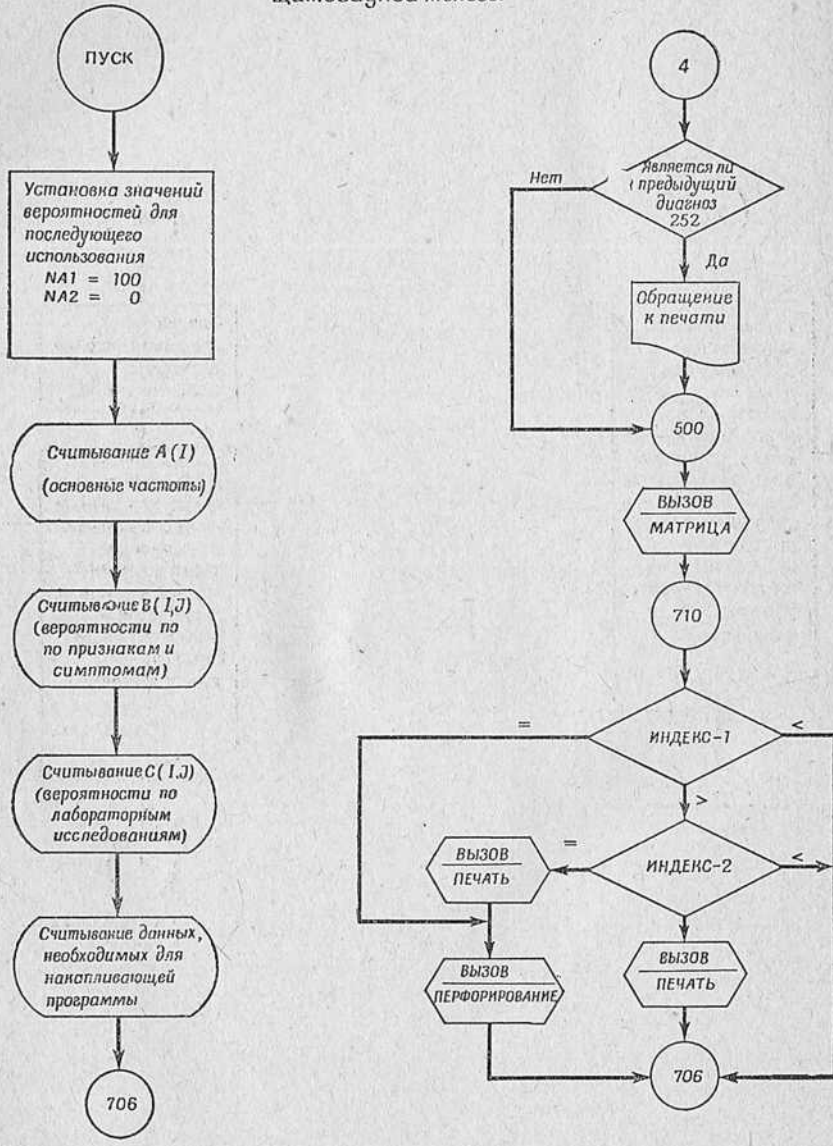


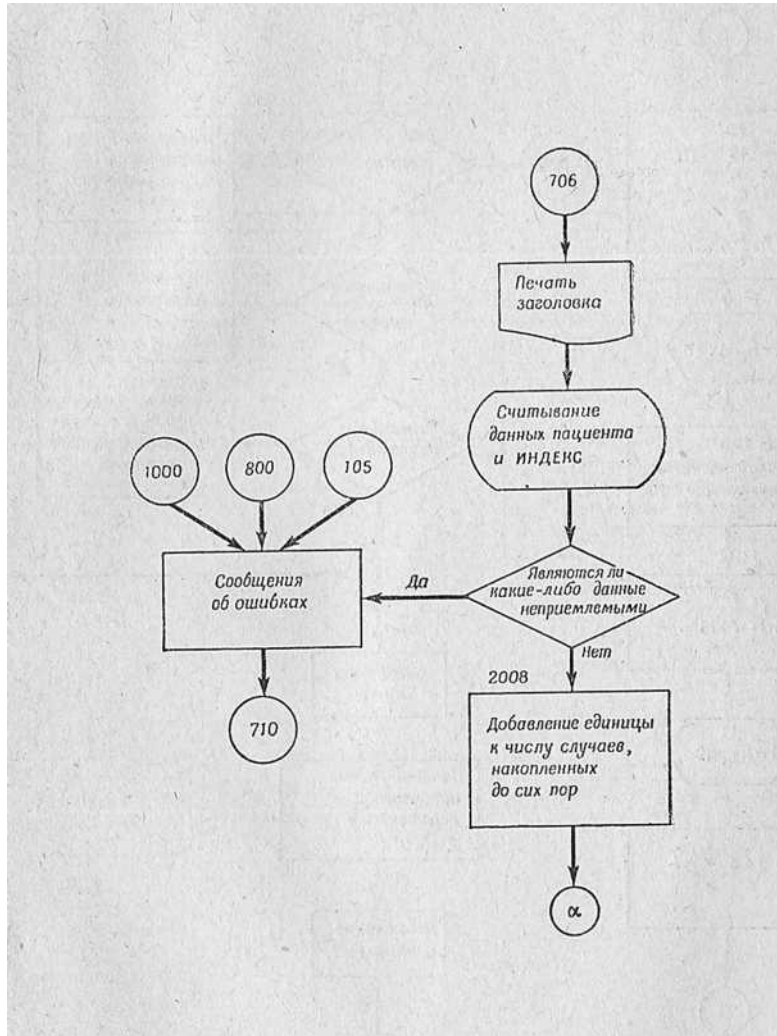


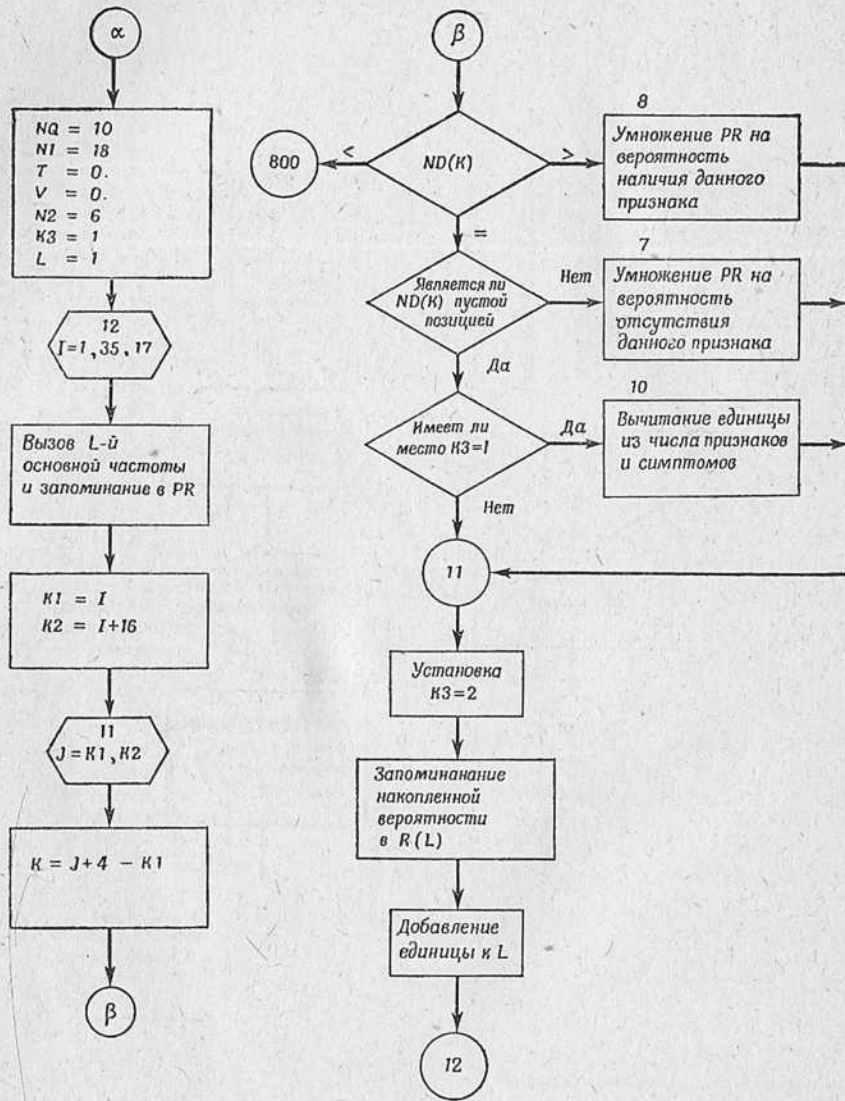


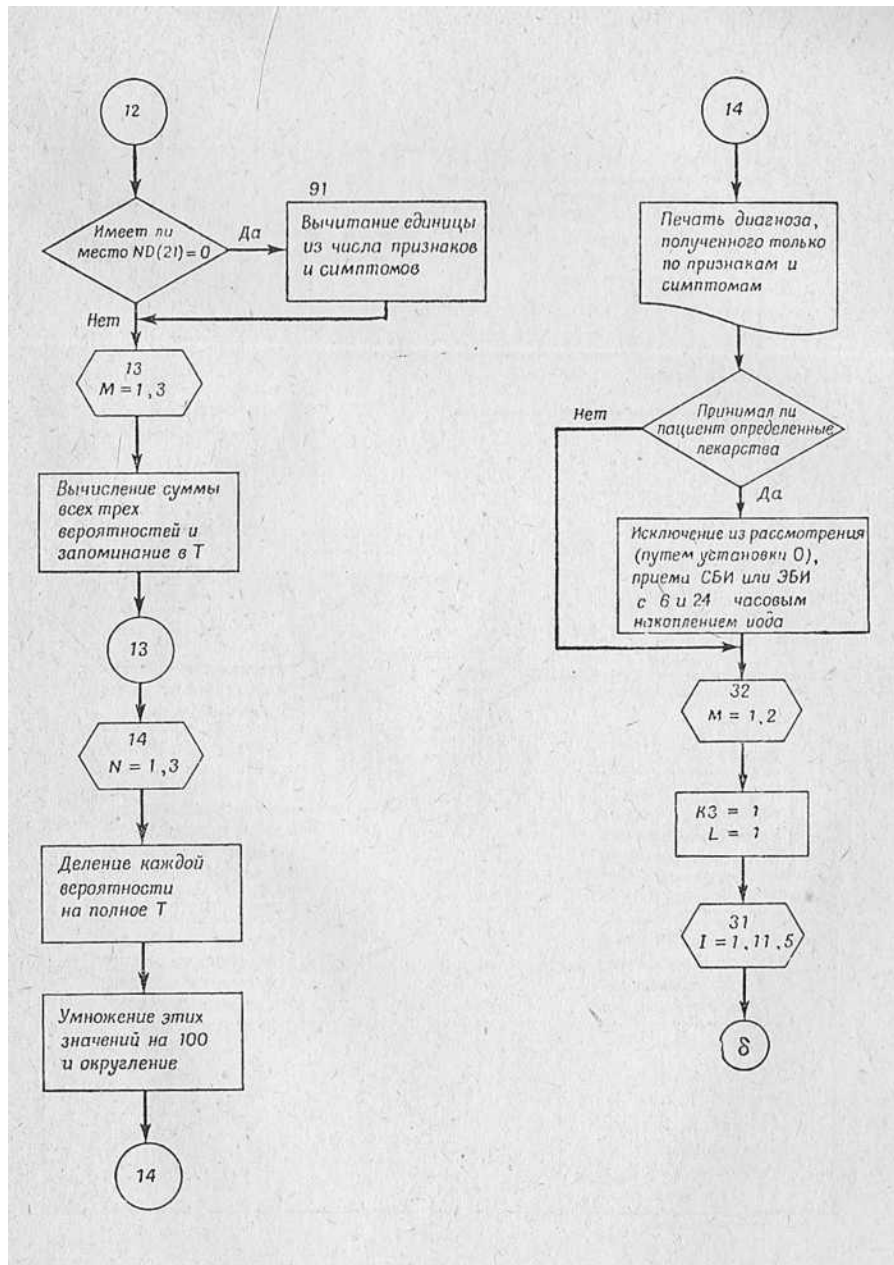


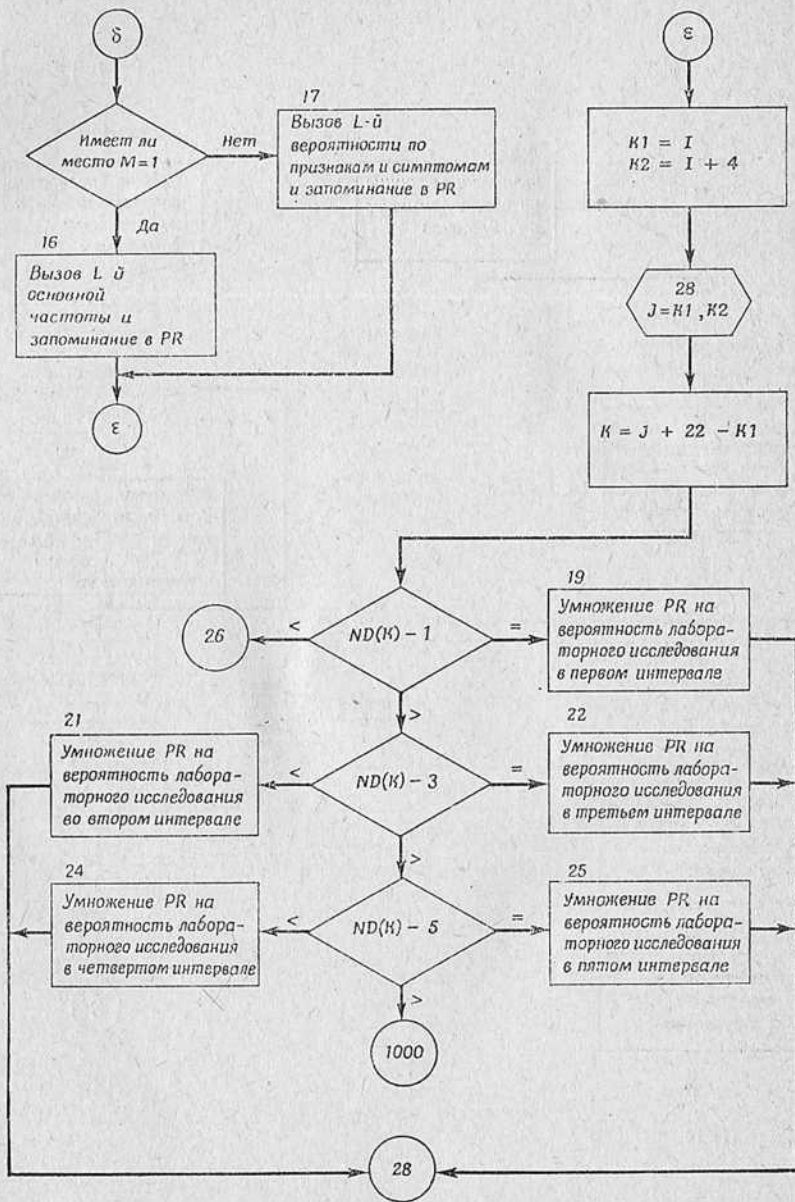
Программа V для заболеваний щитовидной железы

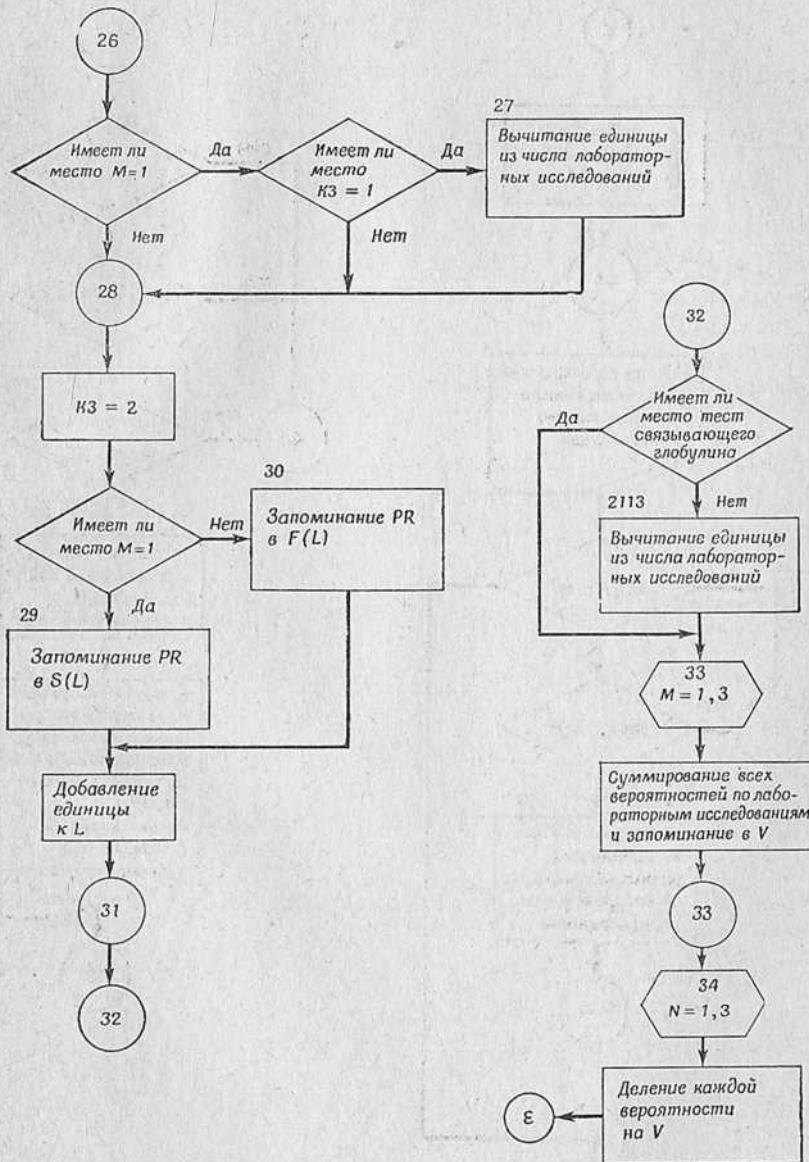


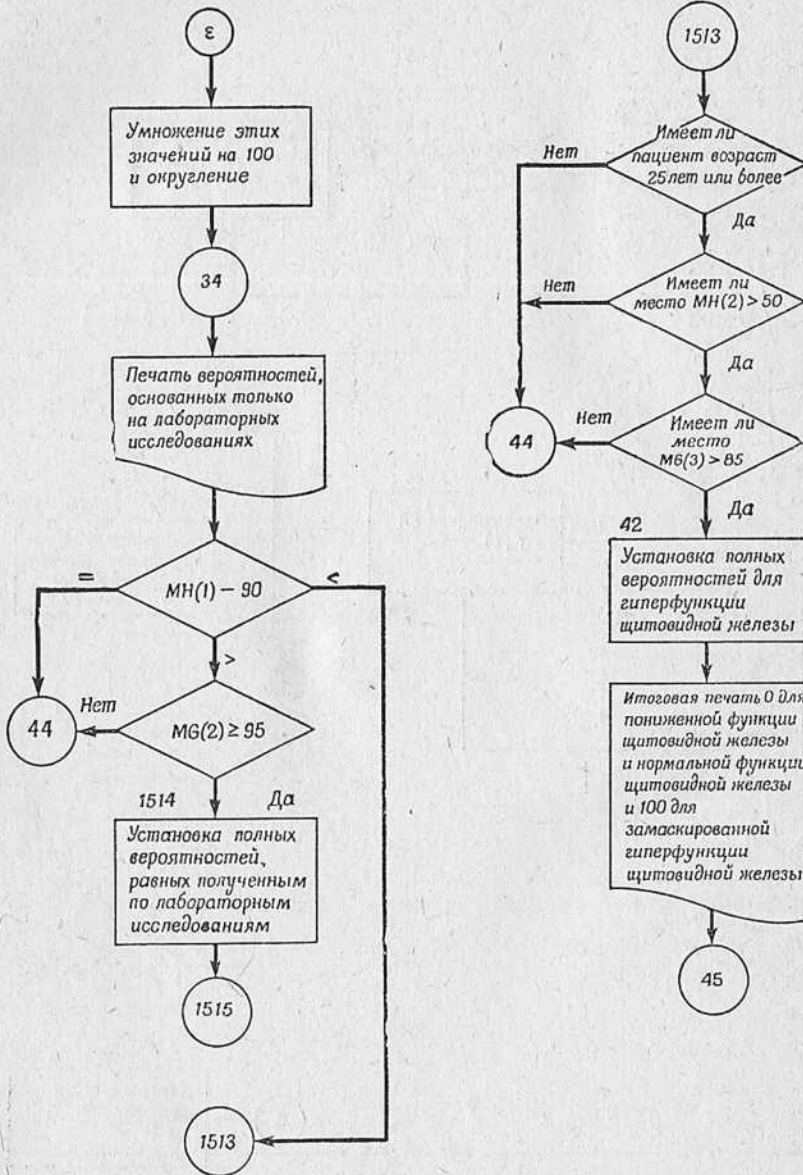


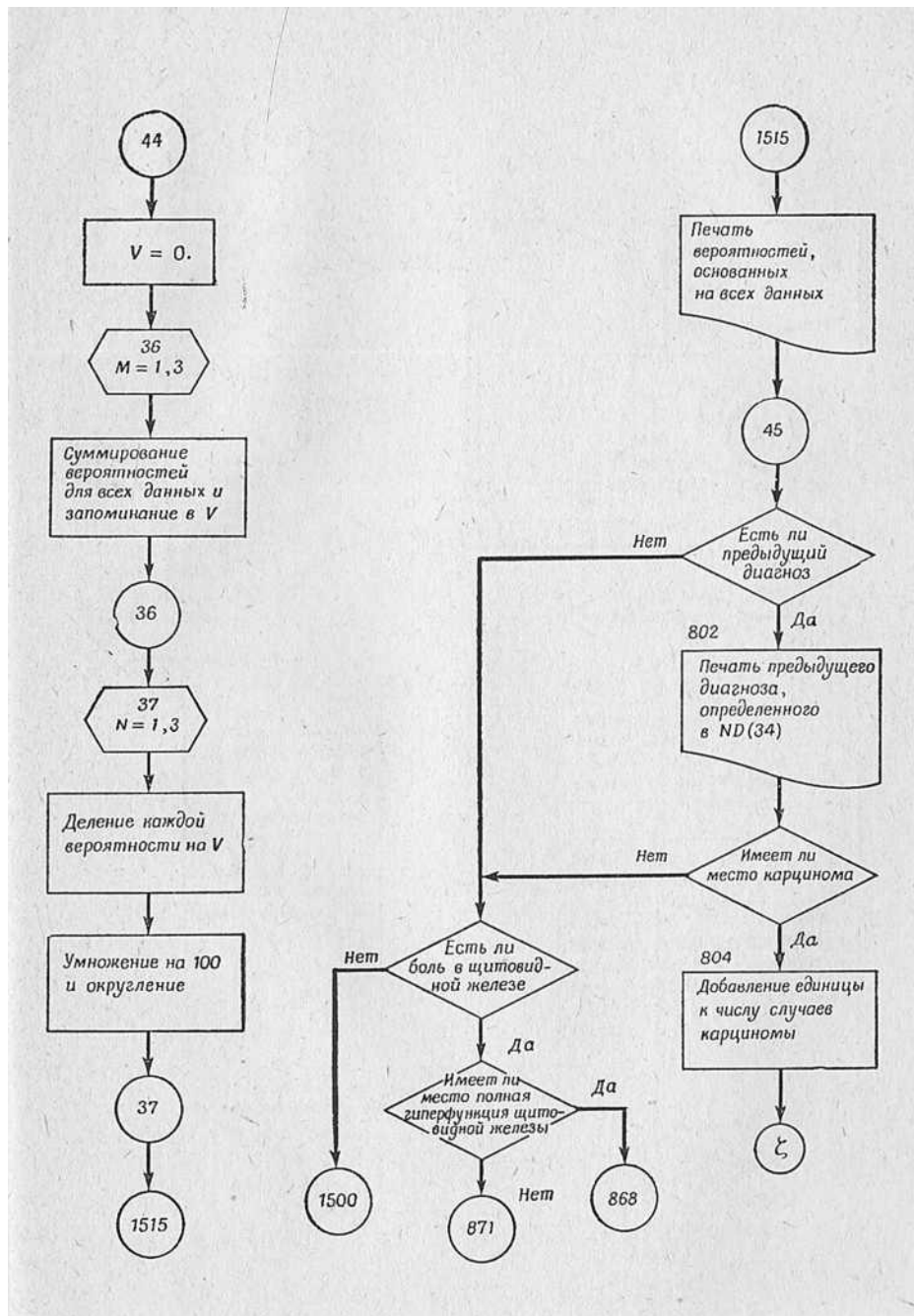


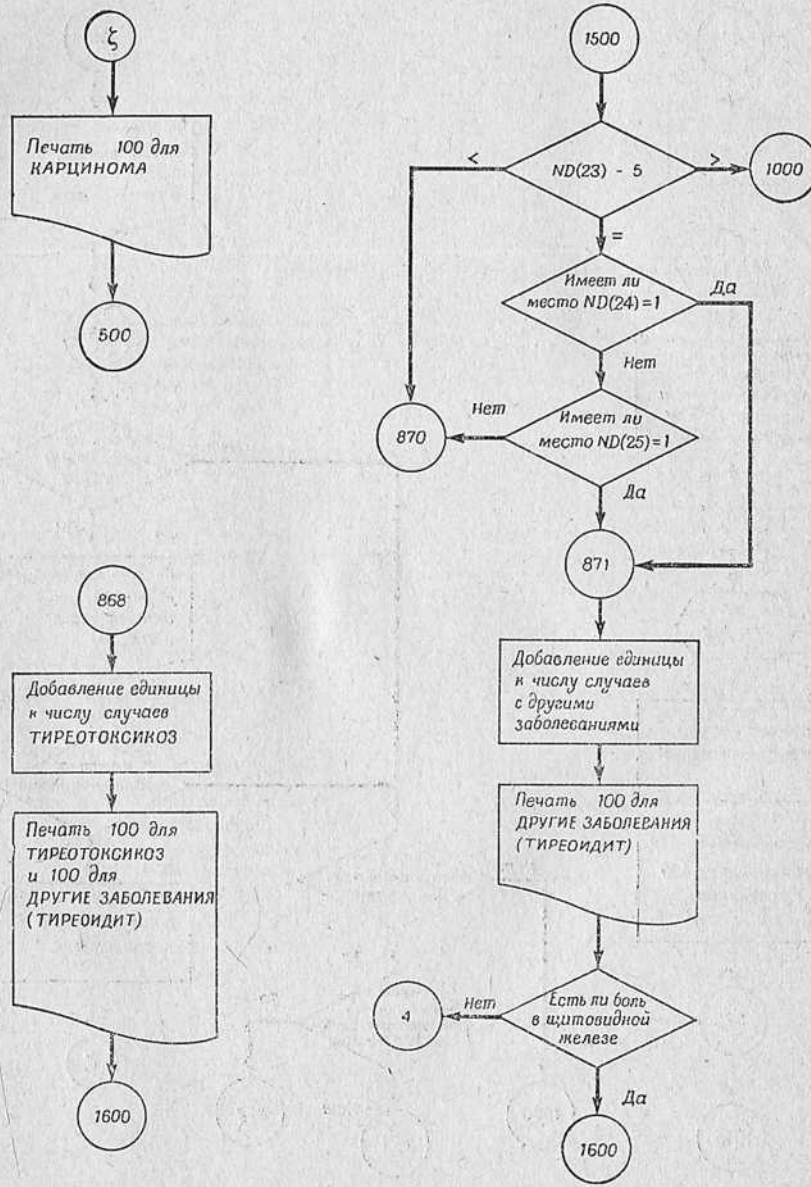




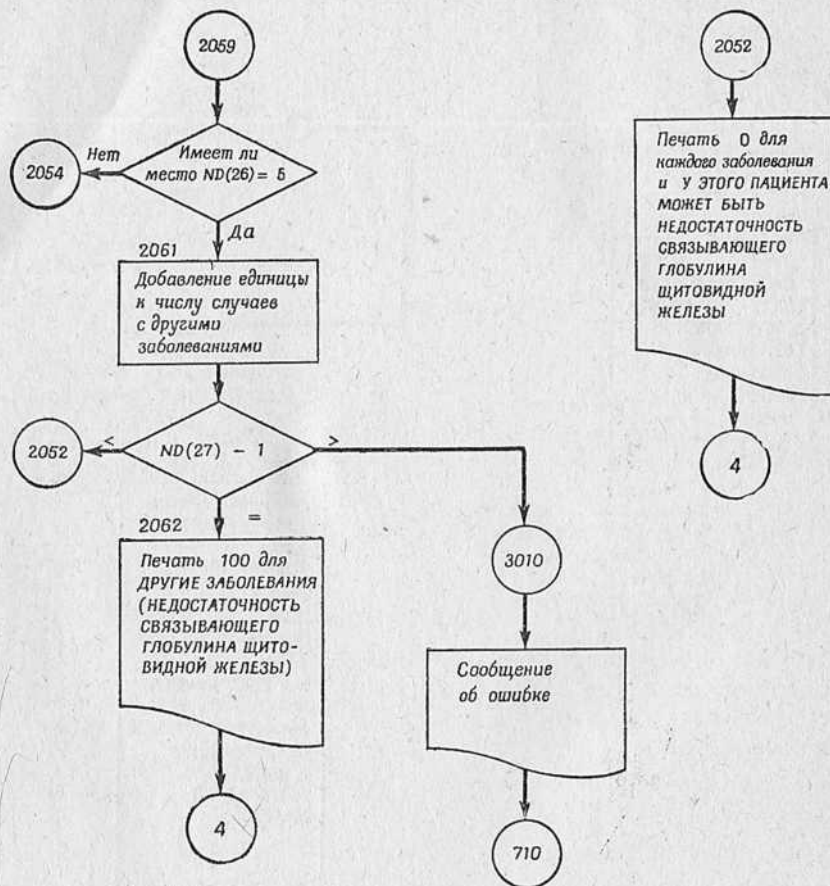


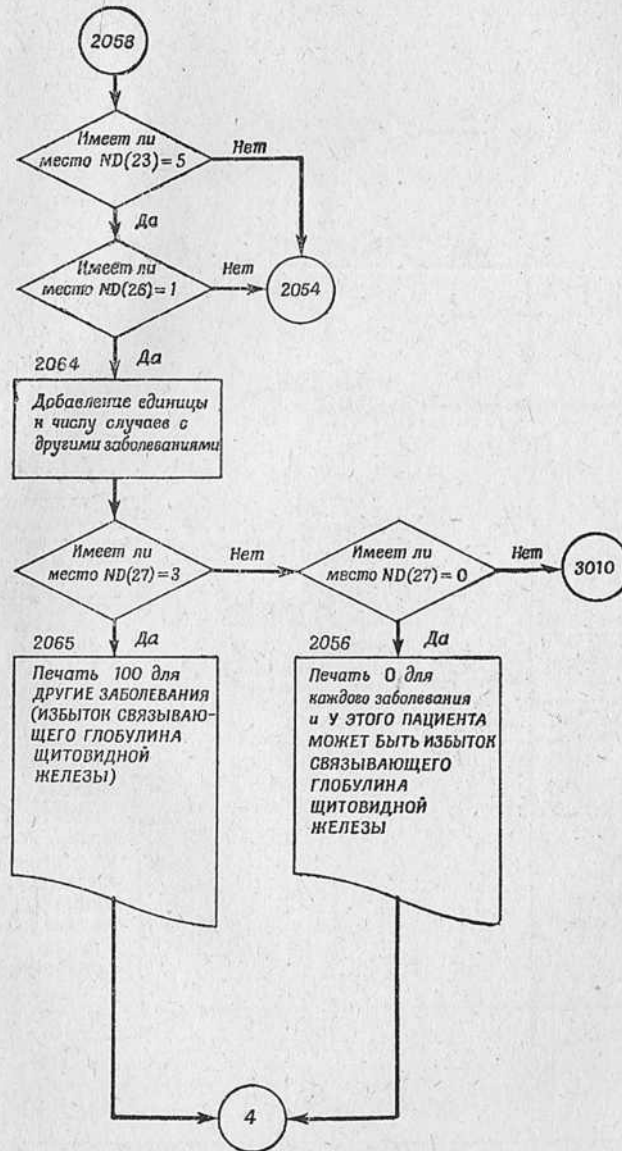


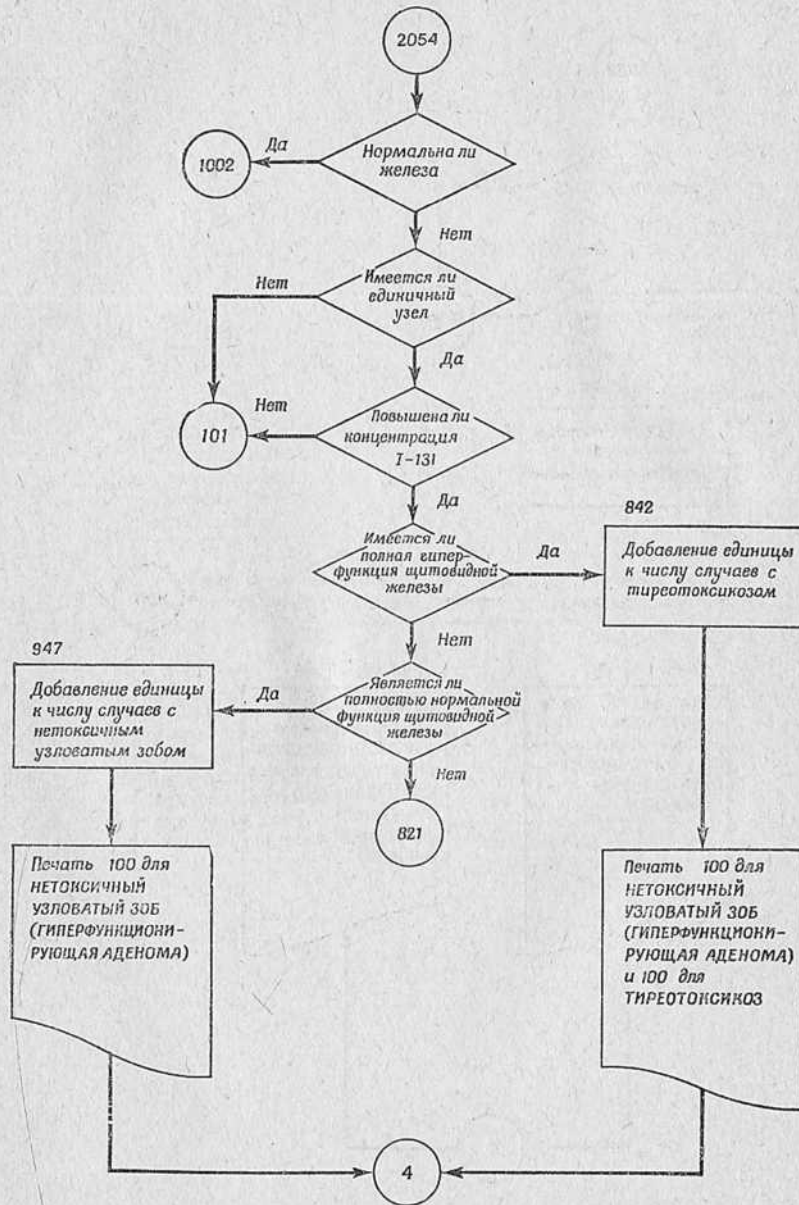


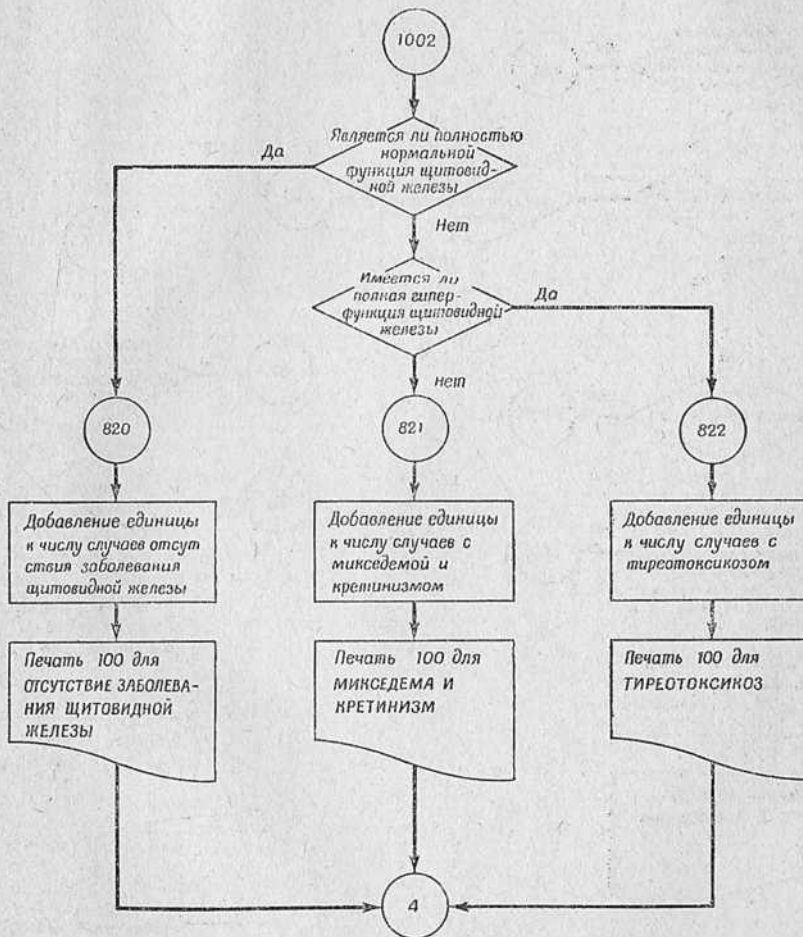


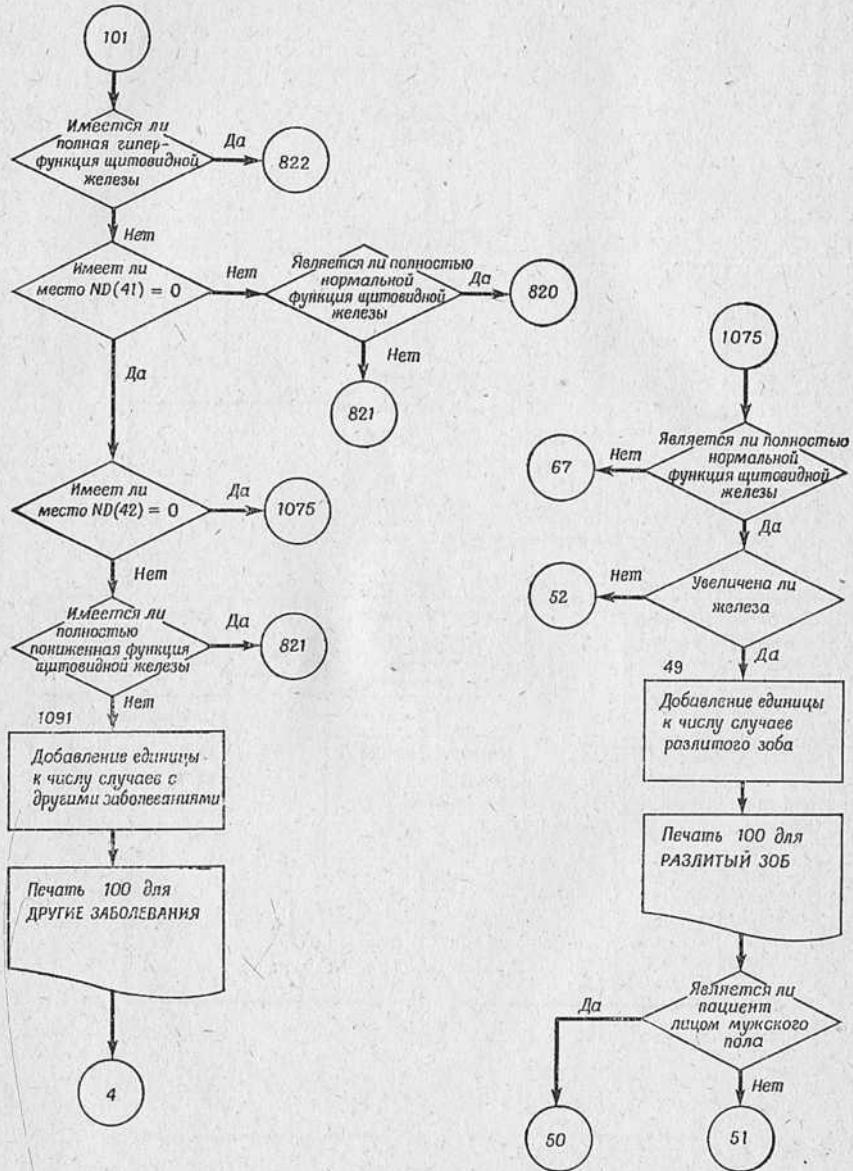


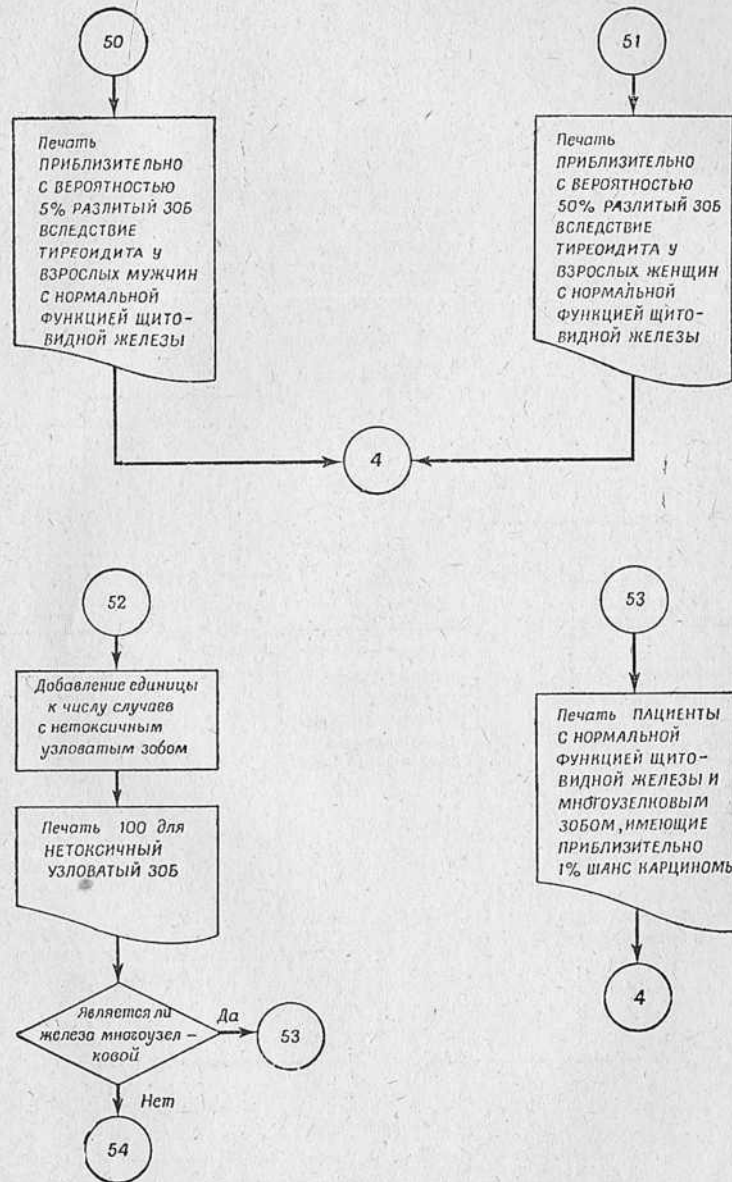


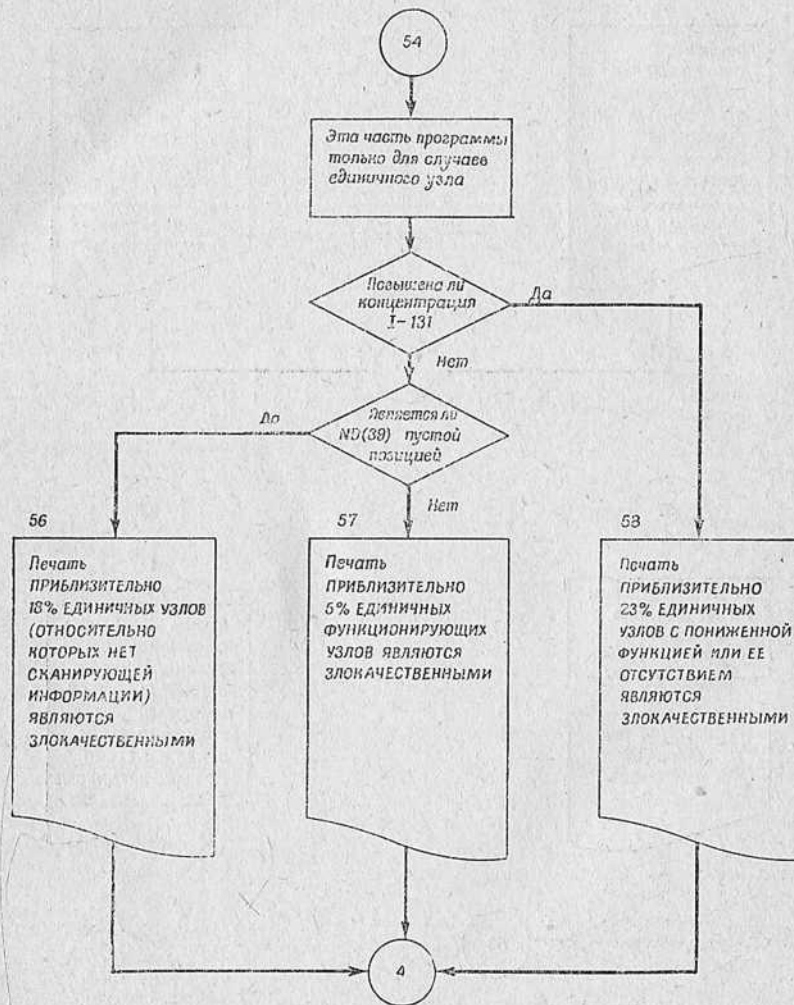


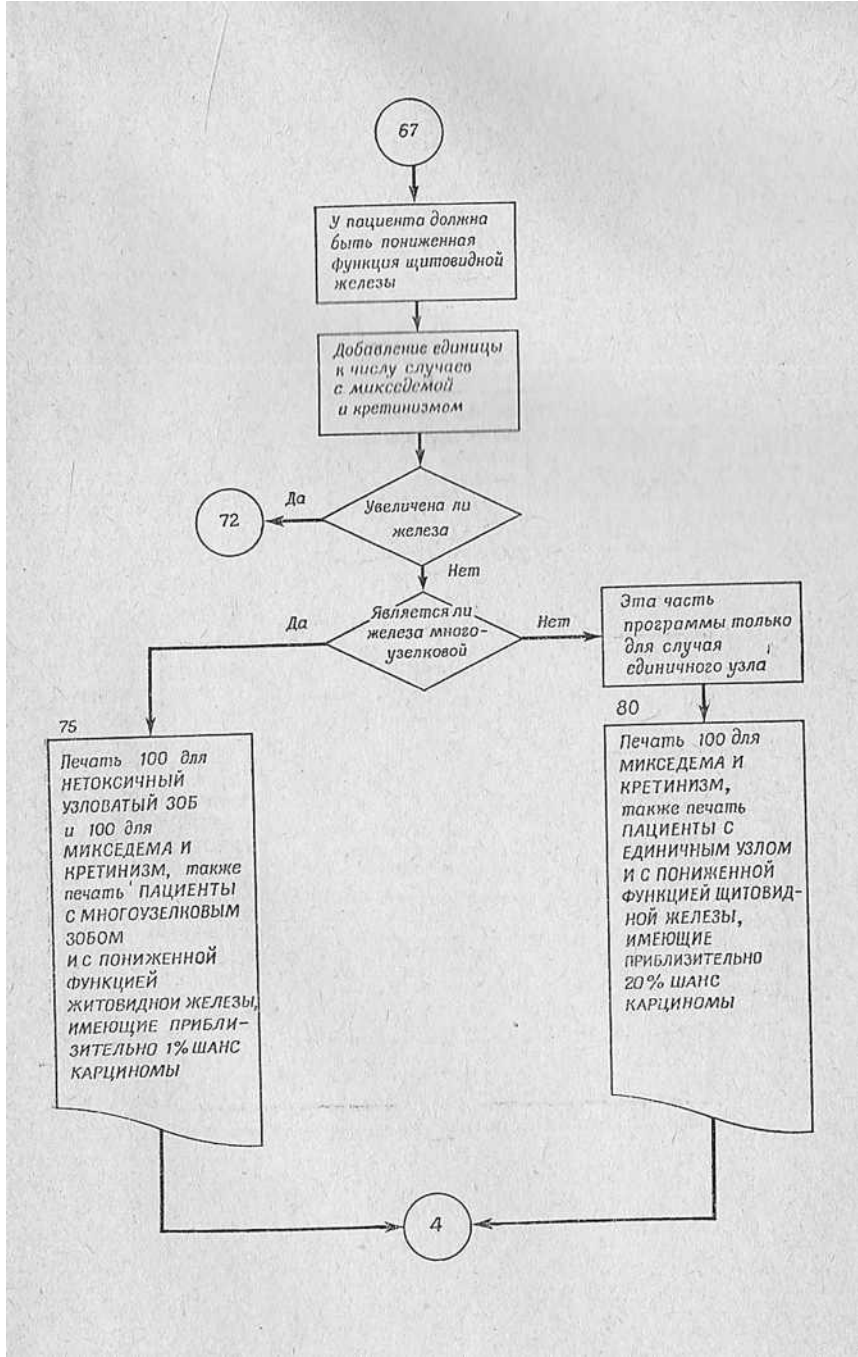


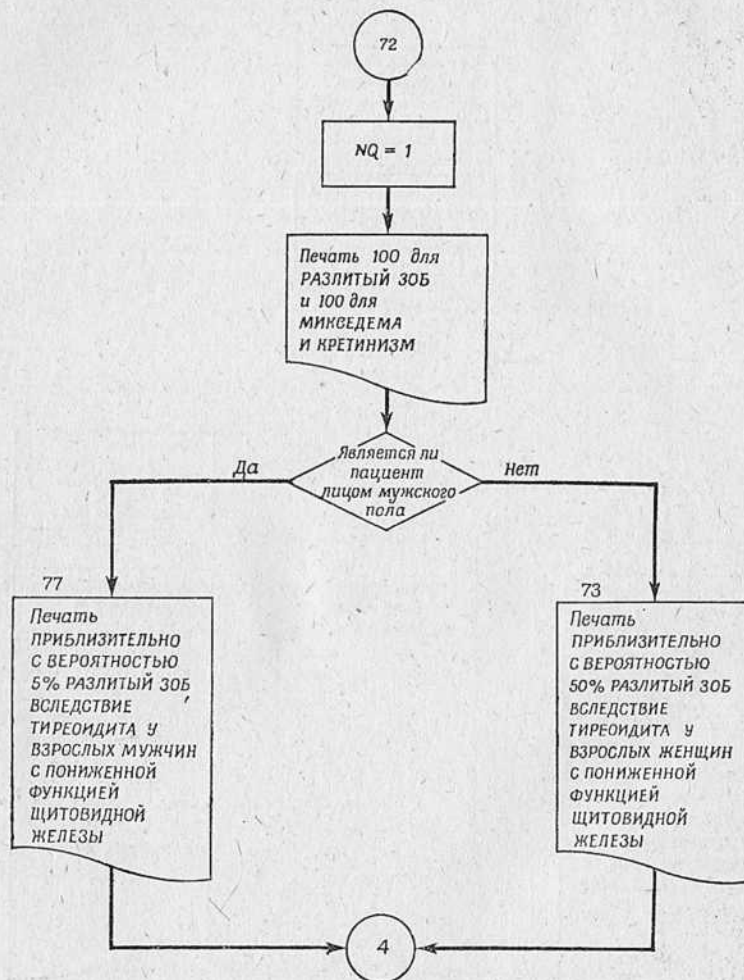




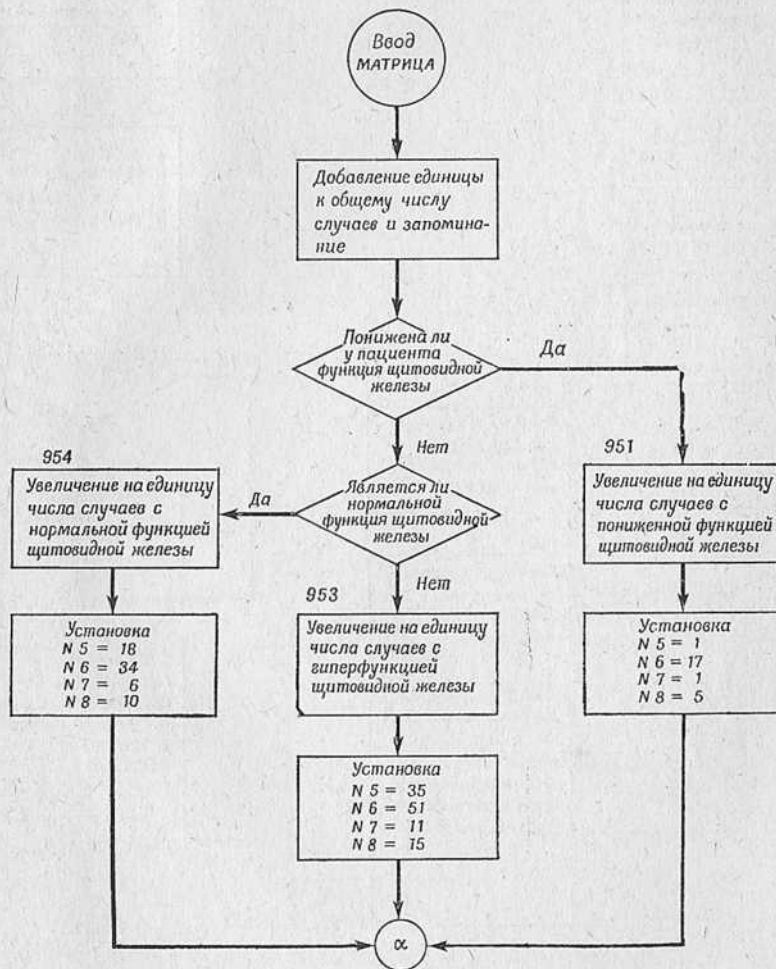


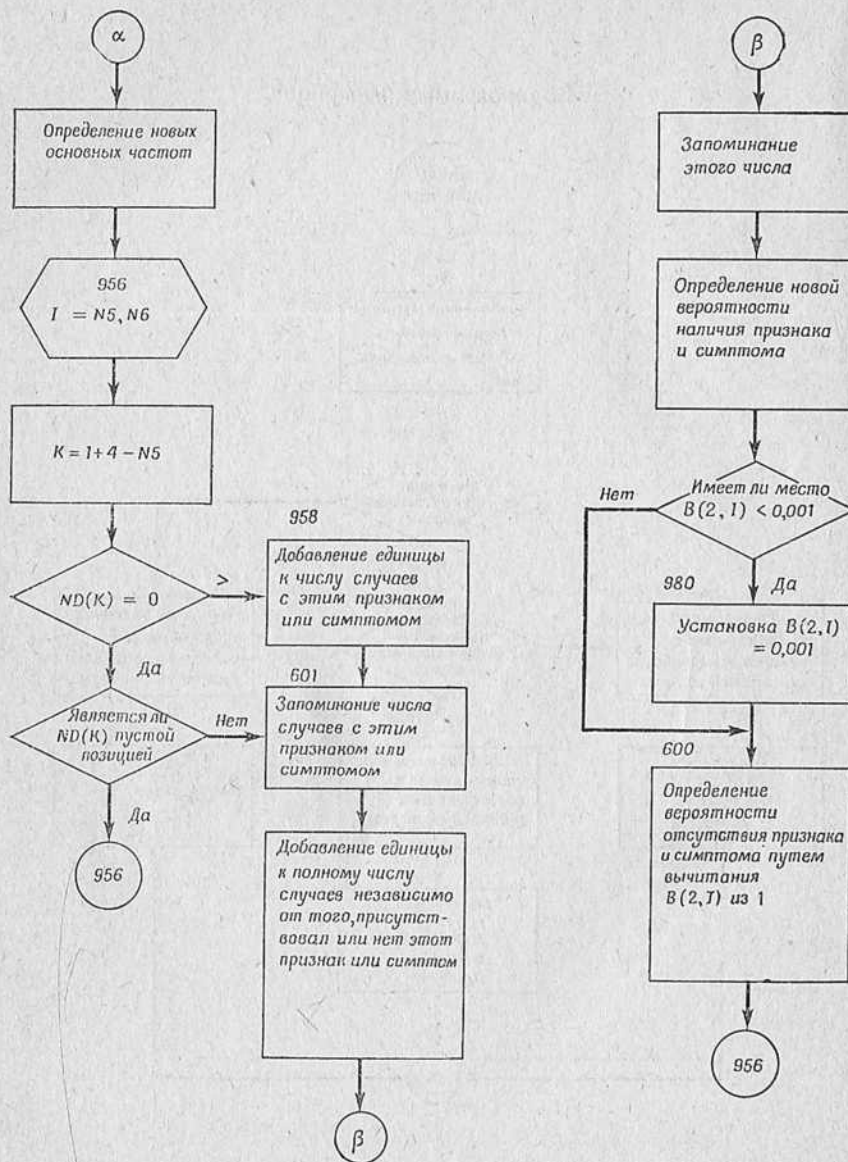


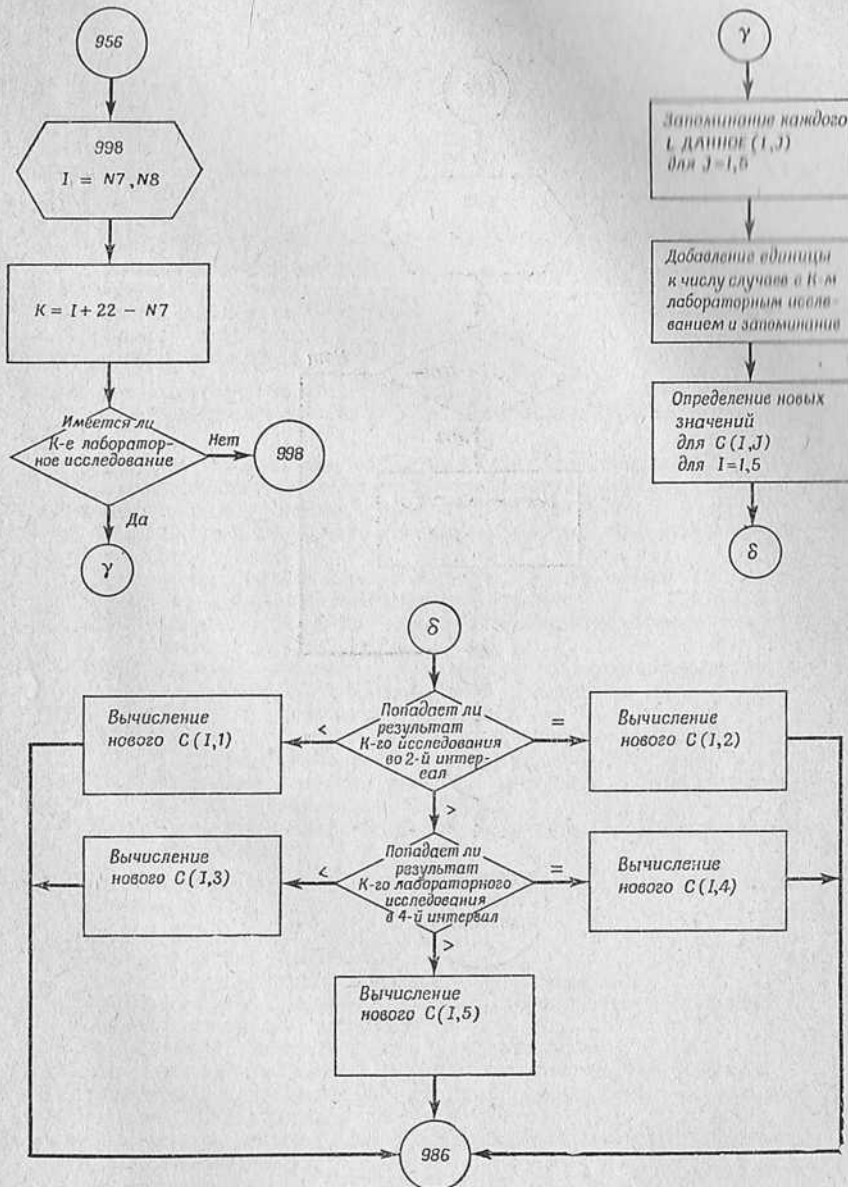


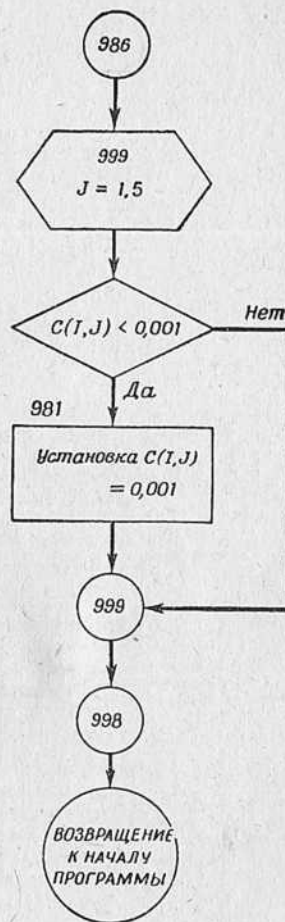


Подпрограмма „Матрица“







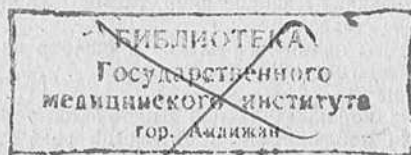


ОГЛАВЛЕНИЕ

Предисловие к русскому изданию	5
Предисловие	9
Глава 1. Байесовская теория статистического вывода и ее применение в медицинской диагностике	21
1.1. Введение	21
1.2. Субъективная мера вероятности	22
1.3. Условные вероятности и теорема Байеса	24
1.4. Правдоподобие и шансы	41
Глава 2. Некоторые исследования по использованию теоремы Байеса и вычислительных машин в медицинской диагностике	46
2.1. Введение	46
2.2. Машинная диагностика заболеваний щитовидной железы	48
2.3. Машинная диагностика врожденных пороков сердца	54
2.4. Машинная диагностика первичных опухолей кости	59
2.4.1. Метод определения минимально необходимых дополни- тельных тестов	67
2.4.2. Метод минимизации количества исследований	69
2.5. Правдоподобие и отношение правдоподобия	71
2.5.1. Пример применения отношения правдоподобия в гемато- логии	72
2.5.2. Пример применения отношения правдоподобия в задаче автоматического многомерного разделения	77
2.6. Разработка диагностических критериев	86
2.6.1. Разработка диагностических критериев исследователями в одном учреждении	87
2.6.2. Разработка диагностических критериев исследователями из двух различных учреждений	92
2.6.3. Разработка диагностических критериев профессиональным обществом	93
Глава 3. Использование блок-схем и дерева решений в диагностическом процессе	102
3.1. Введение	102
3.2. Решение задачи диагностики	102
3.3. Блок-схемы. Некоторые примеры их использования	110
3.4. Система опроса больного (анамнез), основанная на использо- вании вычислительной машины	110
3.5. Машинная «Система Сократа» для диагностики	113
3.6. Медицинская диагностика и процесс распознавания образов	121
3.7. Классификация заболеваний. Методы группирования и вычисли- тельная таксонометрия	130
Глава 4. Процессы принятия решения и ошибки наблюдателя в медицин- ской диагностике	135
4.1. Введение	135
4.2. Ошибки наблюдателя при рентгенологической диагностике	136
4.3. Ошибки наблюдателя в цитологии рака	141

4.4. Проверка гипотез, обнаружение сигнала и процессы принятия решения	146
4.5. Проблема обнаружения сигналов	152
4.6. Замечания по поводу процесса принятия решений при восприятии сигнала человеком	153
4.6.1. Выбор способа действия и ожидаемая величина в процессе решения	158
4.6.2. Сводка результатов теории решений в применении к обнаружению сигналов	161
4.7. Интерпретация рентгеновских снимков грудной клетки как процесс распознавания сигнала	164
4.8. Процедура оценки значимости, используемая для изучения прогностического суждения при инфаркте миокарда	171
4.9. Изложение теории обнаружения сигналов на более общем (байесовском) языке	172
4.9.1. Рабочие характеристические кривые наблюдателя	176
4.10. «Плавание в Ниле»	179
Глава 5. Предположительность и уверенность — некоторые дальнейшие исследования	187
5.1. Введение	187
5.2. Пример разногласий среди врачей по поводу предоперационной диагностики и последующего лечения одиночного узла легкого	189
5.2.1. Материал и анализы серий	189
5.2.2. Комментарии по поводу исследования новообразований типа «штампованных» теней	195
5.3. Моделирование и теория полезности	197
5.3.1. Решение о лечении при определенности	198
5.3.2. Решение о лечении при риске	199
5.3.3. Решение о лечении при неопределенности	201
5.3.4. Разделение конкретных и субъективных факторов	202
5.3.5. Оценка значимости и этика ситуации	204
5.4. Научные исследования и разработка проектов — настоящее и будущее	208
5.4.1. Исследования в области медицинской терминологии	208
5.4.2. Примеры симбиоза врач — машина	210
5.4.3. Некоторые эксперименты в области принятия медицинского решения с использованием системы вероятностной обработки информации (PIP)	214
Литература	221
Приложение 1	228
Приложение 2	232
Приложение 3	235

103 083



УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-278, 1-й Рижский пер., д. 2, издательство «Мир».

Л. ЛАСТЕД

**Введение в проблему принятия решений
в медицине**

Редактор *Л. Тимоир*
Художник *В. Карпов*
Художественный редактор *В. Варлашин*
Технический редактор *Н. Толстякова*
Корректор *В. Соколов*

Сдано в набор 8/II 1971 г. Подписано к печати 23/VI
1971 г. Бум. № 1. 60×90 1/16-9 бум. л. 18 печ. л.
Уч.-изд. л. 18,38. Изд. № 20/5865. Тираж 4300 экз.
Цена 2 р. 04 к. Зак. 100

ИЗДАТЕЛЬСТВО «МИР»
Москва, 1-й Рижский пер., 2

Ярославский полиграфкомбинат Главполиграфпро-
ма Комитета по печати при Совете Министров СССР.
Ярославль, ул. Свободы, 97.

ИЗДАТЕЛЬСТВО «МИР»

в 1971 г.

выпускает следующие книги:

Кибернетические проблемы бионики. Вып. 1. Анализ биологических прототипов, перев. с англ., 24 л.

В основу сборника положены материалы симпозиума по бионике, проведенного ВВС США (вышли отдельной книгой в 1968 г.). Дополнительно в сборник включены последние работы ряда известных американских и английских ученых (Маккаллок, Ферстер и др.) по той же тематике, опубликованные в различных изданиях.

В статьях данного выпуска рассмотрены принципы системной организации биологических объектов и методы моделирования биологических систем, широко представлены различные функциональные модели анализаторных систем, механизмов памяти и др.

Морис В. Наука об управлении. Байесовский подход,
Энглвуд-Клиффс, 1968, перев. с англ., 15 л.

О научном подходе к управлению написано немало книг. Один из наиболее важных этапов управления — процесс принятия решения — является трудным для формализации. Предлагаемая читателю небольшая книга посвящена применению байесовского подхода к этому заключительному этапу управления — стратегии и тактике принятия решений. Книга состоит из трех частей. В первой рассмотрены теоретические и методологические основы применения теоремы Байеса к исследуемым процессам. Во второй рассмотрены примеры применения, взятые из разных областей, в частности из области планирования капиталовложений и выпуска товаров, образования цен, и так называемые биномиальные задачи. В третьей части обсуждаются роль интуиции, выбор критерия, роль научного штата и другие проблемы.

Книга представляет интерес для специалистов самых разных профессий, интересующихся вопросами организации, управления, планирования, в том числе и конкретных работ, например планированием экспериментов.

ИЗДАТЕЛЬСТВО «МИР»

в 1972 г.

выпускает следующие книги:

Ланкастер Дж. Информационно-поисковые системы, перев. с англ., Нью-Йорк, 1968, 16 л.

Книга посвящена актуальным вопросам разработки, испытания и оценки информационно-поисковых систем (ИПС). Автор — видный американский специалист в области информации — принимал личное участие в крупных исследовательских проектах по разработке и оценке эффективности автоматизированных ИПС. В книге основное внимание сосредоточено не столько на технических, сколько на «интеллектуальных» проблемах информационного поиска, имеющих решающее значение для создания эффективно действующих ИПС — на теории и практике классификации и индексации, на стратегии поиска, взаимодействии с потребителем, на оценке эффективности информационно-поисковых систем.

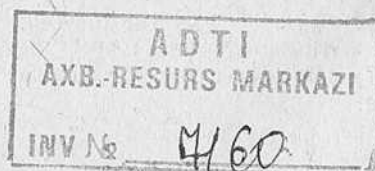
Удачно дополняя вышедшую в 1970 г. монографию Ч. Мидоу «Анализ информационно-поисковых систем», получившую признание у специалистов, данная книга может рассчитывать на еще более широкую читательскую аудиторию. Ясность и доступность изложения при сравнительно небольшом объеме делают ее ценным учебным пособием для молодых специалистов, студентов, аспирантов и «смежников», приходящих в область информатики из других областей, для работников библиотек, научно-исследовательских и проектных институтов, справочных и информационных центров промышленных предприятий и для всех, кто интересуется вопросами эффективного хранения и поиска научной, технической, медицинской и любых других видов информации.

Кибернетические проблемы бионики. Вып. 2. Синтез устройств и инженерные разработки и модели, перев. с англ., 24 л.

Как и в первом выпуске, в основу данной книги положены материалы симпозиума по бионике, опубликованные в 1968 г. и дополненные рядом работ видных английских и американских ученых (Паск, Джордж, Саймон и др.). Статьи этого выпуска посвящены работам в области искусственного интеллекта и создания распознающих устройств, а также инженерным моделям и их конкретным приложениям, в частности к задачам обработки информации на космических летательных аппаратах, управления различными системами космических аппаратов и др.

Тематика материалов сборника чрезвычайно разнообразна и охватывает ряд новых направлений исследований, не получивших еще должного освещения в отечественной научной и научно-технической литературе.

Сборники рассчитаны на математиков, инженеров и научных сотрудников, работающих в области автоматического управления, радиоэлектроники и вычислительной техники, а также психологов, специалистов в области эргономики и всех, кто интересуется проблемами современной кибернетики и бионики.



ВНИМАНИЮ ЧИТАТЕЛЯ!

Только предварительные заказы на книги в магазинах, торгующих научно-технической литературой, обеспечат своевременное получение интересующих вас книг.