

Доказательная радиология: основные принципы и подходы к ее реализации

И.П. Королюк

Самарский государственный медицинский университет

За последние 20–25 лет лучевая диагностика значительно обогатилась большим числом новых методов получения изображения внутренних органов. В медицинской практике широкое распространение получили такие высокотехнологические методы, как компьютерная томография (КТ), магнитно-резонансная томография (МРТ), ультразвуковая биолокация. Постепенно набирает силы новый метод радионуклидной диагностики – позитронная эмиссионная томография (ПЭТ). Получают распространение цифровые методы рентгенологической визуализации – цифровая рентгенография, флюорография, маммография. В практику здравоохранения входят электронные средства передачи и хранения лучевых изображений: телерадиология, радиологические информационные системы – RIS, системы передачи и архивирования лучевых данных – PACS. При этом использование новых способов получения изображения представляет лучевому диагносту принципиально новый вид образов органов в виде трехмерной графики, виртуальной эндоскопии, искусственного моделирования изображений.

Так, по данным Медицинской консультативной комиссии по ценообразованию США (MedPAC) [1], с 1999-го по 2002 г. ежегодный прирост лучевых исследований составил 10,1%, в то время как другие методы диагностики возросли лишь на 5,2%. За этот же период среди всех лучевых исследований высокотехнологические методы выросли на 15–20%.

Следует отметить, что на волне новых технологических решений далеко не всегда корректно учитываются такие обстоятельства, как экономический фактор и диагностическая эффективность их внедрения в здравоохранение.

Кроме того, нужно признать, что практическая деятельность лучевого диагноста пока еще отстает от технического оснащения этой

специальности. Причин этому несколько. Это и ограниченный доступ к современным средствам электронной информации, и не критичный подход к научным публикациям в медицинских журналах, и излишняя степень доверия к данным из учебников и монографий, которые уже в силу чисто технических причин их подготовки и издания существенно отстают от новейших научных результатов. Большинство лучевых диагностов все еще слабо представляют себе, как совместить научную доказательность опубликованных результатов нового метода исследования с другими видами информации, в частности с ее стоимостью, предпочтительностью использования и обстоятельствами применения, что в целом определяет достоверность получаемой информации.

Все методы в лучевой диагностике можно разделить на две группы. Первую из них составляют *аппаратно-зависимые системы*, например КТ или МРТ. Их результат в значительной степени зависит от физико-технических факторов получения диагностического изображения. Ко второй группе относятся *операторозависимые системы*, например УЗИ. В этих системах диагностическое изображение в значительной степени связано с умением и навыками врача, который исследует пациента (в данном случае методом биолокации). В подобных условиях специалисту, работающему в сфере лучевой диагностики, чрезвычайно трудно получить надежную и достоверную информацию об используемых им методах, максимально отвечающих условиям медицинской услуги “высокое качество – эффективная стоимость”.

В идеале лучевой диагност должен иметь такую стратегию исследования, чтобы получить необходимую и достоверную диагностическую информацию в точно назначенный срок, в нужном месте и в необходимом формате. Эта ин-

формация необходима, чтобы принять верное клиническое, научное и управленческое решение.

Для того чтобы отвечать этим требованиям, специалист должен быть ознакомлен с современной текущей информацией, свежими диагностическими решениями своих коллег и с максимальной оптимистичностью осваивать мастерство новых технологий. При этом необходимо максимально использовать возможности существующих локальных компьютерных архивов медицинских изображений (например, через систему PACS), а также накопленные клиничко-лучевые наблюдения в базах данных ведущих медицинских центров мира.

Существует ряд обстоятельств, характерных для деятельности специалиста в области лучевой диагностики. Многие из них носят чисто субъективный характер и, следовательно, подвержены индивидуальной интерпретации.

Во-первых, в тех случаях, когда мы встречаемся с редкими или недостаточно часто встречающимися симптомами и, успешно их интерпретируя, в дальнейшем интуитивно (подсознательно) экстраполируем получаемые успешные данные на последующие аналогичные клинические наблюдения, что в принципе далеко не всегда корректно. Подобная особенность мышления свойственна многим наблюдателям. Это свойство в работе лучевого диагноста в свое время удачно подметил [2].

Например, при выявлении на КТ метастазов в головной мозг при злокачественной опухоли надпочечника мы делаем вывод, что КТ головного мозга является методом выбора при обследовании всех других пациентов, имеющих риск развития данного заболевания. Это конечно же ошибочное заключение.

Во-вторых, *и это очень важно(!)*, мы обучаемся (и обучаем) лучевой диагностике на основе собственного опыта и опыта своих учителей, что в принципе понятно и объяснимо. Однако необходимо учесть, что и те и другие способы получения знаний не всегда объективны. Врач черпает сведения из учебников и руководств, а также используя периодическую медицинскую печать, которая в подавляющем большинстве не содержит точно выверенных научных данных, основанных на современных доказательных технологиях. Например, [3] в статье было показано, что среди 142 журнальных публикаций, посвященных МРТ головного мозга, 113 не отвечают требованиям научного анализа, 28 – лишь частично соответствуют ему и лишь 1(!) работа выполнена на безукоризненно научном уровне. Выводы по-

давливающего числа статей, следовательно, не являются строго научно достоверными.

Для лучевого диагноста, имеющего ограниченный доступ к современной научной литературе и весьма скромные познания в критическом анализе медицинских публикаций с позиции научной доказательности, возникает весьма трудная задача – сориентироваться в правомерности и целесообразности использования тех или иных рекомендуемых лучевых методов в повседневной практической работе.

К этому можно также добавить еще один существенный фактор, который возник в последнее время – значительно возросшее число медицинских журналов, книг, методических рекомендаций, многие из которых быстро устаревают и не имеют содержательных и доказательных медицинских сведений. Это особенно характерно для России, где в последнее время появилось большое количество журналов (медицинских и “полумедицинских”), не проходящих серьезной экспертной и рецензионной оценки.

Группа авторов [3] предложили систематизированную оценку научно-практической значимости всех публикаций в медицинских журналах. Вкратце она такова. Первую группу (А) составляют методологически строгие, выверенные научные исследования, имеющие достоверность обобщающего характера (то есть результаты этих исследований применимы при всех условиях), рандомизированную группировку наблюдений, проспективную проверку диагностического теста (которая всегда более достоверна, чем ретроспективная) и, главное(!), имеют подтверждение полученных данных единым безупречно точным эталонным методом диагностики, так называемым *золотым стандартом*.

Нижнюю группу публикаций относят к группе D, которая содержит работы с большим числом ошибок в построении исследований, так называемых дыр в дизайне исследования. Промежуточное положение занимают публикации в журналах, относящиеся к группам B и C. В этих статьях имеется различная степень погрешностей в построении научного исследования и анализа его данных. Поэтому их данные должны оцениваться критически. Необходимо отметить, что публикуемые во многих радиологических журналах “случаи из практики”, хотя сами по себе и являются интересными, не могут служить основанием для освоения практической деятельности.

Оригинальная группировка медицинских публикаций по их научной и практической значимости приведена в работе [4].

Доказательный анализ медицинской литературы относится к важному разделу деятельности лучевого диагноста. Следствием такого подхода к анализу литературы является корректная профессиональная деятельность лучевого диагноста. Достаточно подробно данный вопрос описан в работе [5]. Здесь же уместно упомянуть некоторые детали подхода к правильному подбору, чтению и анализу медицинских публикаций.

Во-первых, необходимо строго ограничить круг обязательно читаемой литературы. Это относится как к бумажным, так и электронным изданиям. Существует понятие *релевантной* публикации, то есть публикации, посвященной изучаемой проблеме. Среди обилия существующих журналов и книг по медицине имеется своеобразное “ядро” изданий, в котором сосредоточена основная масса публикаций по конкретной теме. Это “ядро” журналов нужно обязательно читать. Так, в радиологии ежегодно издается около 40 высококачественных рецензируемых медицинских журналов, в которых публикуется свыше 4000 научных статей. Далее публикации “распыляются” по другим изданиям по закону, предложенному Г. Бредфордом и носящему его имя. Согласно этому закону, число *релевантных* статей “распыляется” по журналам в соответствии со следующей закономерностью: $n_1 : n_2 : n_3 = 1 : (1a) : (1a)^2$, где 1 – число статей в журналах, входящих в “ядро”; а – константа, равная 5. Следовательно, если число “ядерных” журналов составляет 15, то круг журналов с *релевантными* статьями будет рассеиваться следующим образом: 15 : 75 : 5625. Отсюда следует вывод, что кроме “ядерных” журналов остальные можно только просматривать либо пользоваться другими, альтернативными источниками информации.

Во-вторых, необходимо выработать в себе правило четко разграничивать первичную – оригинальную информацию и вторичную – вспомогательную. Отдавать предпочтение нужно первой из них.

В-третьих, любое научное сообщение должно быть четко структурировано. Это – признак высокой культуры автора и, возможно, высокой ценности публикации. Стандартная структура научного сообщения должна иметь следующие разделы:

– *Введение.*

– *Материал и методы исследования.*

– *Результаты исследования.*

– *Обсуждение.*

– *Выводы.*

– *Список литературы.*

Современные научные журналы высокой культуры требуют обязательной подробной рубрикации научных сообщений.

В-четвертых, следует максимально широко использовать электронные средства для доступа к информации, в частности Интернет, электронные базы данных, компакт-диски.

В-пятых, следует активно посещать научные форумы, конференции, съезды, на которых необходимо уделять внимание не только докладываемым материалам, но и личности докладчика, его умению ориентироваться в рассматриваемом вопросе. Это поможет в дальнейшем составить собственное мнение о ценности его публикаций в медицинских изданиях.

Исходя из сказанного, можно сформулировать требование к подготовке лучевого диагноста: *основой для получения знаний должна стать только научно выверенная литература, в которой полученные данные (и знания) основаны на принципах так называемой доказательной медицины.*

Создание и развитие концепции *доказательной медицины (ДМ)* достаточно полно отражены в ряде публикаций [4, 5, 6]. Источниками этого нового научного направления в медицине послужили ранее опубликованные работы по исследованию информативности медицинских диагностических решений [7, 8, 9, 10]. Сущность ДМ состоит в том, что клинические и организационные решения в сфере здравоохранения должны опираться только на **достоверную и современную информацию**.

*Доказательная медицина, таким образом, – это концепция организации медицинских знаний, основанная на строгих научных данных. При этом личный опыт, авторитет коллег и литературные сведения имеют лишь вторичное, подчиненное значение. Главенствующим принципом принятия врачебного и управленческого решения является только объективный факт. Отсюда вытекает другое (зарубежное) название этой науки – **evidence-based medicine (EBM)** – медицина, основанная на фактах.*

В настоящее время ДМ включена в учебные планы ряда медицинских университетов, например Самарского государственного медицинского университета, Сибирского государственного медицинского университета,

McMaster University (Ontario, Canada, и др. По требованию ВАК все диссертационные исследования должны базироваться на принципах ДМ.

Термин *evidence-based medicine* был предложен в начале 90-х годов группой канадских ученых Университета McMaster под руководством G. Guyatt [11]. В других странах, например, США, получил распространение термин *evidence-based practice* – *EBP* (*практика, основанная на доказательствах*). Предложен также еще один термин – *evidence-based decision making* – *EBDM* (*принятие решения, основанное на фактах*). Этот термин отражает принятие врачебного решения, включающего в себя инструменты для принятия такого решения, исходя из приемлемости научных и практических знаний, почерпнутых в научной литературе, относящейся к группе наиболее достоверных исследований, а также с учетом интересов здоровья пациентов и организации здравоохранения в целом. Иногда для данного направления в медицине употребляют также термин “*доказательное здравоохранение*” (*evidence-based health care* – *EBHC*) [12].

В последнее время за рубежом получило новое направление ДМ – **доказательная радиология** (*evidence-based Radiology* – *EBR*) [12]. Материалы EBR регулярно публикуются в профильных специализированных медицинских журналах (*Radiology*, *Am. J. Roentgenol.*, *Europen Radiol.*, *J. Comput. Assist. Tomogr.*, *Med. Decis. Making* и др.). В ряде медицинских центров и университетов, например, в Harvard Medical School, открыты центры по доказательной радиологии. Успешно и содержательно функционируют по доказательной радиологии порталы в Интернете: *HealthImaging.com*, *Evidencebasedradiology.net* и др.

Предпосылками для введения доказательной медицины и доказательной радиологии (лучевой диагностики) как научного знания явились несколько факторов, присущих медицинской деятельности. Главным при этом выступает ряд обстоятельств.

1. *Диагноз в лучевых исследованиях всегда в некоторой степени неопределенен и поэтому должен выражаться через вероятности.*

2. *В каждом клиническом наблюдении всегда заложена предвзятость.*

3. *В любом лучевом исследовании в той или иной степени обязательно присутствуют случайные ошибки.*

4. *Принятие управленческого решения в организации управления службой лучевой диагностики всегда сопряжено с учетом политических и экономических обстоятельств, а также личностных свойств руководителя.*

5. *Информационные потоки медицинских, в том числе лучевых, данных, как правило, слабо структурированы.*

В идеале лучевой диагност должен иметь приемлемую стратегию исследования, обеспечивающую актуальную, достоверную и необходимую информацию, чтобы руководствоваться ею при принятии клинических, научных и административно-управленческих решений. Эта информация должна быть приемлемой в нужное время, в нужном месте и в нужном формате. Она должна включать в себя непредвзятое отношение к имеющимся данным, базироваться на самых современных технологиях и стимулировать мастерство исполнения диагностических процедур.

Необходимо добавить, что диагностическая информация находится в постоянном развитии в соответствии с накоплением новых сведений в лучевых базах данных и интегрированных лучевых коммуникационных системах (например, PACS). Необходимо учитывать и то обстоятельство, что при разумном подходе всегда можно найти оптимальное взаимоотношение между наилучшим использованием информации, ее ценой, приемлемостью и обстоятельствами (контекстом) использования.

В работе группы экспертов по доказательной радиологии [13] приведена таблица (табл. 1), содержащая уровни доказательности исследований и рекомендации для лучевой диагностики.

К сожалению, ни в лучевой диагностике, ни в медицинской диагностике в целом не фигурируют, за редким исключением, такие термины, как *эффективность* (*идеальная, реальная, экономическая*), *достоверность*, *результативность*, *действенность методов*, *доступность диагностики*, *воспроизводимость* и *сходимость результатов исследований*, а также некоторые другие понятия доказательной медицины. Лишь в редких научных публикациях, адресованных практическому врачу, присутствуют такие окончательные критерии итогов проведенных исследований, как *чувствительность*, *специфичность*, *точность*, *прогностическая значимость положительных и отрицательных результатов*.

В ДМ термин “сила метода” (*efficacy*) определяет меру успеха данного диагностического

Таблица 1. Уровни доказательности и класс рекомендаций: диагностические изображения

Уровень доказательности	Тип исследования
1. Идеальный	Контролируемая серия наблюдений с подходящим спектром последовательно подобранных пациентов, каждый из которых обязательно подлежит как диагностическому, так и стандартному референтному тесту
2. Сильный	Контролируемая серия наблюдений либо с непоследовательным подбором пациентов, либо с ограниченным (суженным) спектром изучаемых индивидуальностей; каждый из пациентов обязательно подлежит диагностическому и стандартному референтному тесту
3. Умеренный	Неконтролируемая серия наблюдений с подходящим спектром последовательно подобранных пациентов, но без референтного стандартно теста, который мог бы быть использован для сравнения
4. Слабый	Неконтролируемая серия наблюдений, в которой был использован референтный стандарт, исследована диагностическая точность силы теста; однако мнение экспертов не содержит подробной критической оценки
5. Очень слабый	Наблюдения из практики; исследование технической силы – <i>efficacy</i> (идеальной эффективности) новой технологии получения изображения

теста, примененного в идеальных технологических условиях его использования влиять на принятие врачебного решения относительно диагноза и лечения пациента. В лучевой диагностике *efficacy* иногда называют *внутренней оценкой* метода лучевой диагностики. Иногда для обозначения данного критерия диагностического теста применяют термин “идеальная эффективность” [5].

Характеристика теста, обозначаемая как *эффективность (effectiveness)*, предполагает использование данного метода в обычных (усредненных) условиях его применения. Понятно, что *effectiveness* всегда будет ниже, чем *efficacy*. Отметим, что подавляющее большинство лучевых диагностов работают именно на уровне *effectiveness*. Некоторые авторы [5] для обозначения этого понятия используют также термин “реальная эффективность”.

В работе [13] приводится иерархическая модель оценки силы метода (*efficacy*) (идеальной эффективности). Ниже приводим эту модель в полном изложении, так как она весьма полезна лучевому диагносту при оценке используемого (или внедряемого) метода.

1. Техническая *efficacy*.

- Разрешение изображения (в пар линий/мм).
- Передача модуляций функциональных изменений.
- Границы серой шкалы, величина пятнистости.
- Нечеткость изображения.

– Параметры компьютеризированного изображения.

2. Диагностическая точность *efficacy*.

- Соотношение ошибочных и правильных диагнозов в серии наблюдений.
- Диагностическая точность (процент верных диагнозов в серии наблюдений).
- Чувствительность*, специфичность, прогностичность положительного и отрицательного ответов теста в определении клинической проблемы.

– Определение области под характеристической (ROC) кривой.

3. Диагностическое осмысливание *efficacy*.

- Число (процент) наблюдений в серии, в которых используемый тест оказался “полезным” для постановки диагноза.
- Энтропия изменений в распределении вероятности дифференциальной диагностики.
- Различия в клинической субъективности устанавливаемого диагноза.
- Вероятности до и после проведения диагностического теста.

– Эмпирически субъективное разделение теста на позитивный и негативный ответы в серии наблюдений.

4. Терапевтическая *efficacy*.

- Число (процент) использованных исследований, оказавшимся “полезными” в планировании лечения в серии наблюдений.
- Процент случаев медицинских или хирургических процедур, которых удалось избе-

* Этот и другие термины будут расшифрованы в дальнейшем.

жать в результате использования диагностического теста.

– Число (процент) случаев запланированной терапии, которые были изменены после выполнения диагностического теста (наличие ретроспективного заключения в клинической истории болезни).

– Число или процент проспективно прослеженных случаев, у которых изменилось лечение после выполнения диагностического теста.

– Полезность теста по оценке пациента.

5. Результат теста для пациента (patient outcome efficacy).

– Процент больных с улучшением состояния с применением и без применения диагностического теста.

– Смертность, которую удалось избежать после получения информации о результатах диагностического теста.

– Изменения в качестве жизни за ожидаемую ее продолжительность.

– Ожидаемая оценка информативности теста для определения качества жизни за один год.

– Стоимость информации от выполнения диагностического теста для качества жизни за один год.

– Оценка пациентом пользы выполненного теста (моделирование Марков, “время обмена”).

6. Социальная efficacy.

– Оценка “успех–стоимость” с социальной точки зрения общества.

– Анализ “эффективность–стоимость” с социальной точки зрения общества.

Действенность метода (*efficiency*). Этот термин относится к мастерству, с которым используются ресурсы для решения данной клинической проблемы. В частности, действенность процедуры получения диагностического изображения считается высокой, когда наибольшее число истинно положительных и истинно отрицательных результатов получено при наименьших затратах, имея в виду ресурсы, деньги и время. Иногда этот критерий диагностического теста называют *экономической, или затратной, эффективностью* [5]. Наконец, необходимо учитывать, какова *доступность* (*availability*) данного метода диагностики для людей, которым необходимо выполнить данное исследование. Имеются в виду финансовые и организационные аспекты.

Важными критериями, характеризующими методы лучевой диагностики, являются такие

критерии, как *достоверность, обобщаемость, воспроизводимость и сходимость* результатов исследования. *Достоверность* (*validity*) лучевого метода показывает, в какой степени полученные результаты исследования соответствуют истинному положению вещей (конкретно – состоянию пациента). Этот показатель отражает *точность* (*accuracy*) диагностики. Для оценки достоверности проведенного лучевого исследования (а также некоторых других критериев диагностики, о чем будет сказано ниже) данное лучевое исследование сравнивают с эталонным методом, который носит название *золотой стандарт*. *Золотой стандарт* – это метод исследования, с максимальной точностью отражающий состояние исследуемого органа. В качестве такого метода чаще всего выступают данные биопсии или аутопсии, иногда другие, в том числе лучевые, методы диагностики, например коронарография, мультиспиральная КТ (например, 64 или 256 срезов), электронно-лучевая томография в диагностике ишемической болезни сердца. Диагноз, основанный на золотом стандарте, носит название *референтный*. Необходимо отметить, что золотой стандарт не всегда имеет абсолютно точное значение. Например, ишемическая болезнь сердца даже в выраженной стадии своего развития может сочетаться с нормальной коронарограммой.

Достоверность лучевого метода диагностики (*валидность* – *validity*) – это соответствие результатов исследования истинному положению вещей. Достоверность лучевого исследования, характерная для данной группы больных, в силу ряда обстоятельств может оказаться не вполне корректной при обследовании другой группы пациентов. Поэтому для характеристики достоверности данного исследования в другой группе больных (или популяции в целом) служит другой критерий – *обобщаемость* диагностического метода (*generalizability*). Иногда его называют *внешней достоверностью* (*external validity*) в отличие от достоверности, описанной выше, которую иногда называют *внутренней достоверностью* (*internal validity*).

Воспроизводимость (*reliability*) методов лучевой диагностики обозначает свойство теста иметь идентичные результаты, полученные разными специалистами, в различных лечебных учреждениях (городах и странах). Важно помнить, что лучевой тест, имеющий высокую достоверность, может оказаться мало воспроизводимым. В лучевой диагностике такими

Вклад в решение клинических проблем

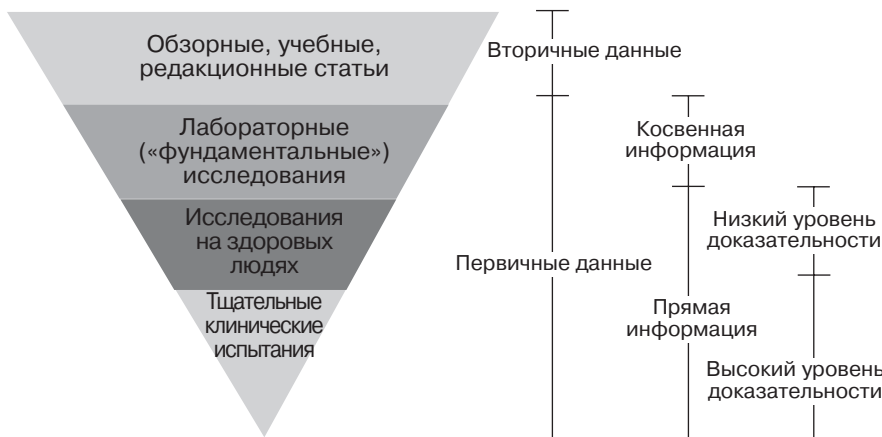


Рис. 1. Характеристика литературы по научной проблеме: относительная ценность различных статей применительно к поиску ответа на клинической вопрос (Флетчер Р. и соавт., 1998).

свойствами обладают, например, рентгенография грудной клетки, *in vitro* диагностика, некоторые виды МРТ и гамма-сцинтиграфии, которые выполнены в различных лечебных учреждениях. Соотношение между достоверностью и воспроизводимостью по данным диагностических исследований [4] показано на рис. 1.

Ошибки в лучевой диагностике, как и в медицинской диагностике вообще, принято подразделять на *рандомизированные и систематические*.

Рандомизированные, или случайные, ошибки возникают, прежде всего, вследствие неправильных (или недостаточно точных) заключений специалиста, проводящего лучевые исследования. Вариабельность анализа легочного рисунка различными врачами на основании анализа на рентгенограммах грудной клетки хорошо известна в медицинской практике. Достаточно часты рандомизированные ошибки в *in vitro* диагностике, при визуальном анализе КТ и МРТ. Существующие методы статистической обработки медицинских данных позволяют снизить влияние рандомизированных ошибок при выполнении серии лучевых процедур. В медицине принято считать, что частота рандомизированных ошибок ниже 5% от всех проведенных исследований ($p < 0,05$) свидетельствует о достаточной точности данного метода диагностики.

Систематические ошибки (или смещения — bias) отражают неправильное выполнение исследований, ошибки в настройке аппаратуры, неадекватный подбор больных, некорректно выполненный анализ полученных данных. Лучевая диагностика, судя по публикациям и

собственному опыту, изобилует систематическими ошибками. Особенно часто систематические ошибки возникают от неправильной настройки аппаратуры (например, юстировки излучателей и приемников изображения, некорректного подбора датчиков УЗИ, неполноценной настройки приемников излучения в гамма-камере или КТ). Необходимо отметить, что в последнее время настройка аппаратуры в лучевой диагностике в большинстве своем осуществляется автоматически с помощью компьютерных программ. Однако сбои в программном обеспечении компьютеров хорошо известны специалистам, в том числе работающим в сфере лучевой диагностики.

При проведении лучевого исследования, как и при выполнении любого диагностического теста вообще, могут встретиться следующие варианты врачебных заключений.

1. Интерпретация “положительный” у больных с наличием заболевания. Это *истинно положительные случаи (TP, true positive)*.

2. Интерпретация “отрицательный” у пациентов без заболевания. Это — *истинно отрицательные случаи (TN, true negative)*.

3. Интерпретация “положительный” у пациентов с отсутствием заболевания (у здоровых лиц). Это *ложноположительные случаи (FP, false positive) (ошибки I типа, или α -ошибки, “ложные тревоги”*. В практической медицине риск совершить такого рода ошибки каждой врач устанавливает по-своему. Однако общепринятой критической величиной ложной тревоги является величина α -риска, равная 0,05, или 5%. Этот уровень обозначается в медицинской статистике как уровень статистически значимого показателя. Он обозначает

ется как $p \leq 0,05$. Практически это означает, что врач ошибается в 1 случае из 20.

4. Интерпретация “отрицательный” у больных с заболеваниями. Это *ложноотрицательные случаи (FN, false negative) (ошибки II типа, пропуски, или β -ошибки)*.

Как ложноположительные, так и ложноотрицательные ошибки представляют собою совокупность ошибок диагностики и выдвигают основное требование перед врачом-диагностом — их минимизации.

Для характеристики информативности диагностического метода исследования служат объективные параметры, именуемые *операционными характеристиками теста*.

К операционным характеристикам метода диагностики относятся:

- чувствительность (Se , sensitivity),
- специфичность (Sp , specificity),

Другая группа показателей информативности относится к *вспомогательным критериям*, так как их значимость непостоянна и зависит от распространенности заболевания (*преваленса*).

Преваленс (Ps) — это вероятность определенного заболевания, или проще, его частота встречаемости среди изучаемой группы людей или популяции в целом.

От преваленса следует отличать *инцидент (In)* — вероятность нового заболевания в рассматриваемой группе людей (популяции) за определенный промежуток времени, чаще за один год.

К вспомогательным критериям информативности относятся:

- точность (Ac , accuracy),
- прогностичность положительного результата ($+VP$, positive predictive value),
- прогностичность отрицательного результата ($-VP$, negative predictive value).

Чувствительность (Se) — это пропорция правильных положительных результатов теста среди всех больных. Определяется по формуле:

$$Se = \frac{TP}{D_+} \cdot 100\%$$

где Se — чувствительность; TP — верно положительные случаи; D_+ — число больных искомым заболеванием.

Чувствительность априори показывает, какова будет доля больных, у которых данное исследование даст положительный результат. Чем выше чувствительность теста, тем чаще с его помощью будет выявляться заболевание,

тем, следовательно, он более эффективен. В то же время если такой высокочувствительный тест оказывается отрицательным, то наличие заболевания маловероятно. Поэтому их следует применять для исключения заболеваний. В силу этого высокочувствительные тесты нередко именуют *идентификаторами*. Тесты с высокой чувствительностью рекомендуется применять на ранних этапах диагностического процесса, когда требуется сузить круг предполагаемых заболеваний. Необходимо также отметить, что высокочувствительный тест дает много “ложных тревог”, что требует дополнительных финансовых затрат на дальнейшее обследование.

Специфичность (Sp) — это пропорция правильных отрицательных результатов теста среди здоровых пациентов. Данный показатель определяется по формуле

$$Sp = \frac{TN}{D_-} \cdot 100\%$$

где Sp — специфичность; TN — истинно отрицательные случаи; D_- — здоровые пациенты.

Определив специфичность, можно априори предполагать, какова доля здоровых лиц, у которых это исследование даст отрицательный результат. Чем выше специфичность метода, тем надежнее с его помощью подтверждается заболевание, тем, следовательно, он более эффективен. Высокоспецифичные тесты называются в диагностике *дискриминаторами*. Тестирование эффективно на втором этапе диагностики, когда круг предполагаемых заболеваний сужен и необходимо с большой уверенностью доказать наличие болезни. Отрицательным фактором высокоспецифичного теста является тот факт, что его использование сопровождается весьма значительным числом пропусков заболевания.

Из сказанного следует очень важный практический вывод, который состоит в том, что в медицинской диагностике желателен тест, который был бы *априорно* и высокоспецифичен, и высокочувствителен. Однако в реальности этого достичь нельзя, так как повышение чувствительности теста неизбежно будет сопровождаться потерей его специфичности и, наоборот, повышение специфичности теста сопряжено со снижением его чувствительности. Отсюда следует вывод: чтобы создать оптимальную диагностическую систему, нужно найти компромисс между показателями чувствительности и специфичности, при ко-

торых финансовые затраты на обследование будут оптимально отражать баланс между рисками “ложных тревог” и пропуска заболеваний.

Точность (Ac) – это пропорция правильных результатов теста среди всех обследованных пациентов. Она определяется по формуле:

$$Ac = \frac{TP + TN}{D_+ + D_-} \cdot 100\%$$

где *Ac* – точность; *TP* – истинно положительные решения; *TN* – истинно отрицательные решения; *D₊* – все здоровые пациенты; *D₋* – все больные пациенты.

Точность, таким образом, отражает, сколько всего правильных ответов получено в результате испытаний данного теста.

Для правильного понимания диагностической эффективности методов исследования важную роль играют критерии *апостериорной* вероятности – прогностичность положительного и отрицательного результатов. Именно эти критерии показывают, какова вероятность заболевания (или его отсутствия) при известном результате исследования. Нетрудно видеть, что апостериорные показатели имеют большее значение, чем априорные.

Прогностичность положительного результата (+VP) – это пропорция правильно положительных случаев среди всех положительных значений теста. Данный показатель определяется по формуле:

$$+PV = \frac{TP}{TP + FN} \cdot 100\%$$

где *+PV* – прогностичность положительного результата; *TP* – истинно положительные случаи; *FN* – ложно отрицательные случаи.

Прогностичность положительного результата, таким образом, напрямую показывает, насколько велика вероятность болезни при положительных результатах исследования.

Прогностичность отрицательного результата (-VP) – это пропорция верно отрицательных случаев среди всех отрицательных решений. Этот критерий определяется по формуле:

$$-PV = \frac{TN}{TN + FP} \cdot 100\%$$

где *-PV* – прогностичность отрицательного результата; *TN* – истинно отрицательные случаи; *FP* – ложно положительные случаи.

Данный критерий, таким образом, показывает, насколько велика вероятность того, что пациент здоров, если результаты теста отрицательные.

Если операционные характеристики – чувствительность и специфичность – не зависят от частоты заболевания, то прогностичность теста – положительная и отрицательная – напрямую связана с *преваленсом*. Чем выше преваленс, тем выше прогностичность положительного результата. Прогностичность тестов также связана с их операционными характеристиками – чувствительностью и специфичностью. Чем выше чувствительность теста, тем выше прогностическая ценность его отрицательного результата. Прогностичность положительного результата в основном зависит от специфичности. Низкоспецифичные методы сопровождаются возникновением большого числа ложно положительных решений. Это приводит к снижению прогностичности положительных результатов теста.

При этом *параллельное* применение нескольких тестов обеспечивает их *большую чувствительность*, а следовательно, и более высокую прогностическую ценность отрицательного результата. Вместе с тем *снижается специфичность и прогностическая ценность положительного результата теста*.

При *последовательном* применении диагностических тестов снижаются *чувствительность и прогностическая ценность отрицательного результата*, но вместе с тем *повышаются специфичность и прогностическая ценность положительного результата*. Таким образом, комбинация различных методик исследования, изменение их последовательности меняют совокупность операционных характеристик каждого теста в отдельности и *общую прогностичность* их результатов.

Из сказанного следует важный вывод доказательной медицины: *прогностические характеристики любого теста нельзя автоматически, без учета преваленса и ряда других обстоятельств, переносить на все лечебные учреждения*.

При использовании лучевого теста для скрининговых исследований систематические ошибки наиболее часто возникают вследствие некорректного подбора пациентов, подлежащих профилактическому исследованию. Данный вопрос особенно актуален в связи с боль-

шим распространением в нашей стране проверочной флюорографии грудной клетки и развешиваемой в настоящее время по национальному проекту “Здоровье” проверочной маммографии.

При оценке результатов скрининга следует, прежде всего, учитывать принципы подбора пациентов, подлежащих скринингу: первичный или вторичный скрининг, избирательный скрининг (группы риска) или тотальное обследование всего населения. Интуитивно принято оценивать эффективность скрининга по соотношению количества выявленных случаев заболевания при скрининге к количеству больных, обратившихся к врачу с жалобами на данное заболевание.

Однако это не совсем правильно. Систематические ошибки (смещения) при оценке программы скрининга могут зависеть от многих факторов. Так, повторный скрининг той же группы пациентов через определенный промежуток времени выявит меньшее количество больных, чем первичный. В научной литературе такой вид обследования носит наименование *скрининг заболеваемости* в отличие от предыдущего – *скрининг распространенности*.

Важным фактором оценки эффективности скрининга является субъективность оценки врачом результатов исследования. Это относится в особенности к тем случаям, когда сопоставляются два метода исследования (например, цифровая и пленочная маммография). При анализе таких результатов второй исследователь не должен знать результаты первого. Расхождения в диагностических решениях двух врачей в таких случаях могут достигать 62% (14).

При сравнительном анализе двух методов лучевой диагностики не следует пользоваться различными золотыми стандартами, один из которых применялся к больным, *пациентам*, а другой к – здоровым обследуемым. В таких случаях могут возникнуть завышенные оценки результативности метода, которые носят название *верификационные смещения*.

При организации скрининга должны быть приняты во внимание следующие требования к методам исследования:

- высокая чувствительность,
- низкая стоимость,
- воспроизводимость результатов,
- безопасность,
- доступность.

Поскольку скрининг проводится среди преимущественно здорового населения, при

оценке его результатов следует учитывать низкий преваленс заболевания в обследуемой группе людей. Поэтому для скрининга должны быть использованы высокочувствительные методы, при которых отрицательный результат высокопрогностичен. В то же время положительный результат наблюдается у всех больных с искомым заболеванием.

Так как прогностичность положительного результата зависит преимущественно от специфичности метода, которая обычно невысока у чувствительных методик, при скрининге возникает проблема ложноположительных результатов. Такие результаты приводят, во-первых, к неблагоприятным воздействиям на больных и, во-вторых, порождают проблему дополнительного углубленного обследования пациентов, что ложится достаточно тяжелым бременем на органы здравоохранения. Причем это бремя носит как финансовый, так и человеческий (профессиональный) фактор. Поэтому, приступая к скринингу, организаторы здравоохранения должны быть уверены в том, что могут обеспечить доступные методы углубленного обследования на приемлемом финансовом уровне.

В организации скрининга существует феномен, получивший название *порочный круг скрининга*. Этому подвержены все страны, так как в его основе лежит, главным образом, человеческий фактор. Суть порочного круга такова: для получения хороших результатов выбирается высокочувствительный метод. Применение такого метода приводит к большому числу ложноположительных результатов. Недостаточное их отсеивание (или уточнение) создает субъективное впечатление о высоком преваленсе заболевания. Это мнение доводится до сведения руководителей здравоохранения, которые выделяют дополнительные средства на проведение скрининга и возможность лечения предполагаемых заболеваний. Дополнительные финансовые вливания в скрининг еще больше увеличивают выход ложноположительных результатов, и далее, все идет по кругу, раскручиваясь по спирали до тех пор, пока не иссякнут источники дополнительного финансирования. Для разрыва порочного круга необходимо модифицировать скрининг: изменить уровень принятия диагностического решения или использовать другой метод скрининга.

Еще одно важное обстоятельство, которое нужно учитывать при выполнении сравнительных исследований, состоит в том, что рет-

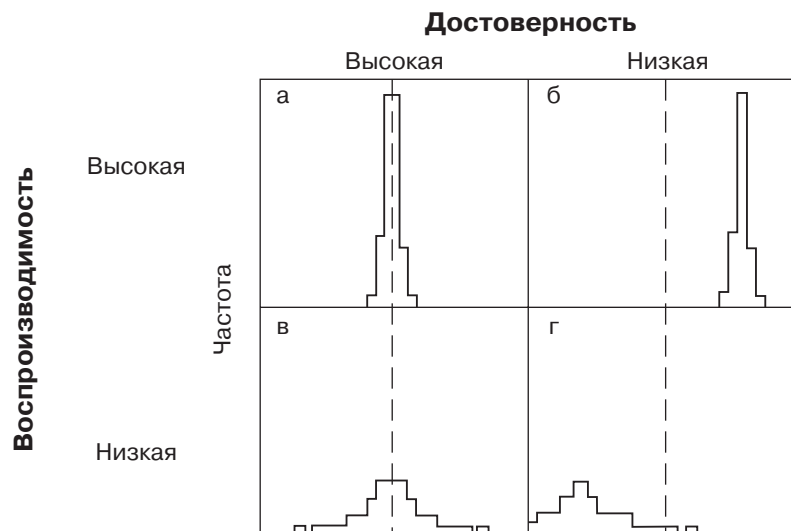


Рис. 2. Соотношение между достоверностью и воспроизводимостью диагностических исследований: а – высокая достоверность и высокая воспроизводимость; б – низкая достоверность и высокая воспроизводимость; в – высокая достоверность и низкая воспроизводимость; г – низкая достоверность и низкая воспроизводимость. Пунктирные линии показывают истинные значения диагностического теста.

роспективные исследования чисто психологически всегда более субъективны, чем проспективные наблюдения. Поэтому в доказательной медицине последние исследования являются более предпочтительными. В целом необходимо также отметить, что при небольших по объему сравниваемых выборках изучаемых наблюдений, тем более при таких, которые имеют склонность к смещению результатов, итоговые выводы получаются более оптимистичными. Как правило, последующие репрезентативные исследования, выполненные с учетом требований доказательной медицины, вносят необходимые, порою существенные, коррективы.

Теория доказательной медицины показала, что разграничение групп пациентов по состоянию здоровья на норму и патологию условно и зависит от точки разделения этих состояний в зависимости от субъективных качеств исследователя – его решительности или осторожности (рис. 3), а также от других предпосылок – внешних и внутренних. Если точка X на оси принятия решений сместится в сторону большей решительности или в противоположную сторону – осторожности, то изменятся взаимоотношения между всеми четырьмя показателями информативности: истинно положительными, истинно отрицательными, лож-

ноположительными и ложноотрицательными результатами.

С целью определения операционных характеристик теста результаты его применения сводят в определенную таблицу, называемую *матрицей решений* (табл. 2). При этом в структуре исследуемых пациентов должны быть обязательно представлены две группы наблюдений. Первую группу составляют пациенты с точно установленным (референтным) диагнозом. Вторую группу должны составить пациенты, у которых это заболевание, безусловно, отсутствует. Поясним сказанное на следующем примере.

Предположим, разрабатывается новый метод электронной (цифровой) флюорографии. Следует дать оценку его информативности, то есть определить его *операционные характеристики*. Для этой цели подбираются больные с безупречно и точно установленным диагнозом какого-либо заболевания (например, туберкулеза). Допустим, всего подобрано по 100 пациентов каждой группы. В первой группе больных туберкулезом флюорографический тест оказался положительным у 88 пациентов, а у 12 человек он отрицательный. Из второй группы пациентов здоровыми признаны 94 человека, у 6 пациентов возникло подозрение на туберкулез, и они отправлены на

Таблица 2. Распределение пациентов по наличию у них заболевания и результатам теста

Данные флюорографии	Туберкулез		Всего
	присутствует	отсутствует	
Положительные	88	4	92
Отрицательные	12	96	108
Всего	100	100	200

Таблица 3. Распределение больных и здоровых пациентов по категориям решений

Фактическое состояние	Категории решений					Всего
	1	2	3	4	5	
Туберкулез имеется	4	16	20	40	20	100
Туберкулез отсутствует	30	44	9	13	4	100

дальнейшее обследование. На основании полученных данных составляется матрица решений (см. табл. 2).

Результаты вычислений по данным, изложенным в таблице, позволяют определить диагностическую информативность, то есть определить чувствительность (Se), специфичность (Sp), точность (Ac), вероятность положительного (+ VP) и отрицательного ответов (- VP):

$$Se = \frac{88}{100} \cdot 100\% = 88\%, \quad Sp = \frac{96}{100} \cdot 100\% = 96\%,$$

$$Ac = \frac{184}{200} \cdot 100\% = 92\%, \quad +VP = \frac{88}{92} \cdot 100\% = 96\%,$$

$$-VP = \frac{96}{108} \cdot 100\% = 89\%.$$

В доказательной медицине для сравнительной оценки методов лучевой диагностики, в частности, различных способов визуализации внутренних органов получил распространение метод анализа *ROC-кривых* – операционных кривых наблюдателя (*Receiver Operation Characteristic curves*). Этот метод позволяет провести сравнительную оценку показателей различных способов визуализации органов, сопоставить между собой технические характеристики различных устройств для получения медицинских изображений (например, мониторы), сравнить эффективность выявления патологического состояния организма при использовании различных альтернативных диагностических систем. *Принципом ROC-анализа является сопоставление чувствительности и специфичности по уровню ложноположительных решений.*

ROC-кривые представляют собою график зависимости истинно положительных от ложноположительных решений, принимаемых наблюдателем при оценке результатов тестов. Диагностическими тестами могут быть также медицинские диагностические изображения, когда часть их них содержит в себе искомый сигнал (патологию), а часть – не содержит (норма).

Наиболее распространенным способом построения ROC-кривых, создающим условия

Таблица 4. Расчет точек построения ROC-кривой

Ось	Значение точек ROC-кривой				
Ордината	0,2	0,6	0,8	0,96	1,0
Абсцисса	0,04	0,17	0,26	0,7	1,0

для количественной их оценки, является группировка полученных данных по 5-балльной оценке. С этой целью подбираются две группы пациентов: 1) без патологических изменений (здоровые лица) и 2) содержащие их (больные).

В качестве примера обратимся к флюорографии больных туберкулезом и условно создадим две группы пациентов: в одной из них будут 100 здоровых лиц, в другой – 100 больных туберкулезом. Маркировка флюорограмм закрыта для исследования (слепой, или рандомизированный, метод). Наблюдатели дают оценку обнаруженным изменениям по следующим категориям.

1. Туберкулез абсолютно отсутствует.
2. Туберкулез, вероятно, отсутствует.
3. Равновероятное наличие или отсутствие туберкулеза.
4. Вероятное наличие туберкулеза.
5. Определено наличие туберкулеза.

Затем составляется таблица, содержащая в себе категории оценок относительно истинного положения дел: в качестве примера приведена табл. 3.

Для построения точек кривой необходимо вычислить их ординаты. При этом ось ординат будет соответствовать вероятности истинно положительных решений, ось абсцисс – вероятности ложноположительных решений. После некоторых преобразований параметров ось абсцисс можно обозначить как “чувствительность” метода, ось ординат – как “1 – специфичность”. Для построения первой точки на оси ординат в качестве истинно положительного решения рассматривается решение с максимальной уверенностью (категория 5), то есть 20 наблюдений. Это число относится к числу всех патологических изменений, то есть к 100. Получается значение первой точки ординаты – 0,2 (табл. 4). Вероятность ложноположительного решения по этой же, 5-й,

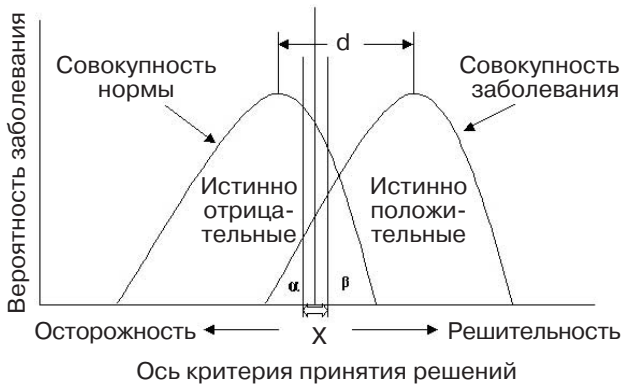


Рис. 2. Связь между 4 состояниями – истинно положительными, истинно отрицательными, ложно положительными (β) и ложноотрицательными (α) результатами. X – точка на оси критерия принятия решения для разделения больных на норму и патологию, утолщенная прозрачная стрелка – диапазон (зона) принятия решений для разделения нормы и патологии; d – критерий распознавания сигнала (болезни).

категории получается отнесением числа 4 (из второй строки) к числу всех здоровых людей, то есть к 100. Получается значение точки на оси абсцисс, равное 0,04, то есть начало графика ROC-кривой. Подобным образом строим точки ординаты и абсциссы для остальных категорий решений. Каждая последующая ячейка таблицы рассчитывается как сумма текущей и всех предыдущих категорий. Иными словами, значение в 4-й категории рассчитывается как сумма 5-й и 4-й: 3-й категории – сумма 5-й, 4-й и 3-й и т. д. В итоге получаем таблицу 3 со значением всех точек для построения ROC-кривой (рис. 3).

ROC-анализ применяется в лучевой диагностике для сравнения различных методов визуализации, например КТ, МРТ, УЗИ и др. (рис. 4).

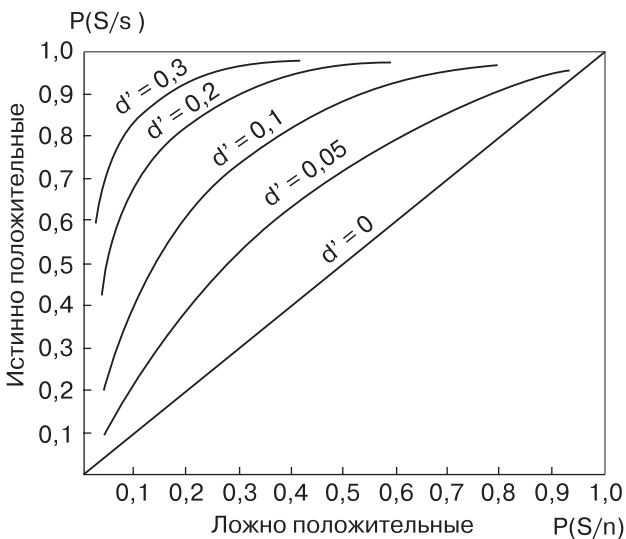


Рис. 4. Экспериментальные ROC-кривые, полученные при сравнительном изучении способности различных методов лучевой диагностики выявлять структурные изменения в органах. Наилучшей разрешающей способностью обладает метод, соответствующий кривой, расположенной в верхнем левом углу ($d = 0,3$).

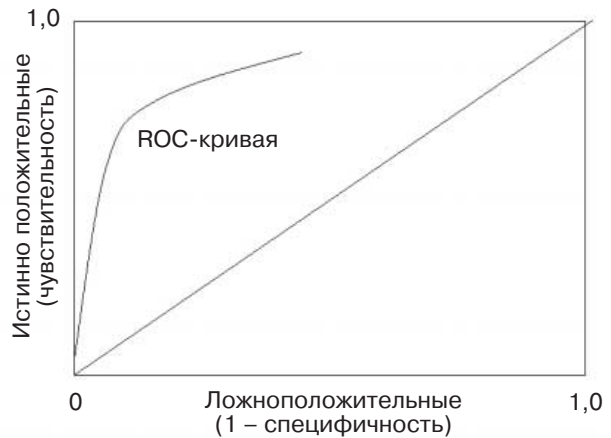


Рис. 3. ROC-кривая, полученная из таблицы 4.

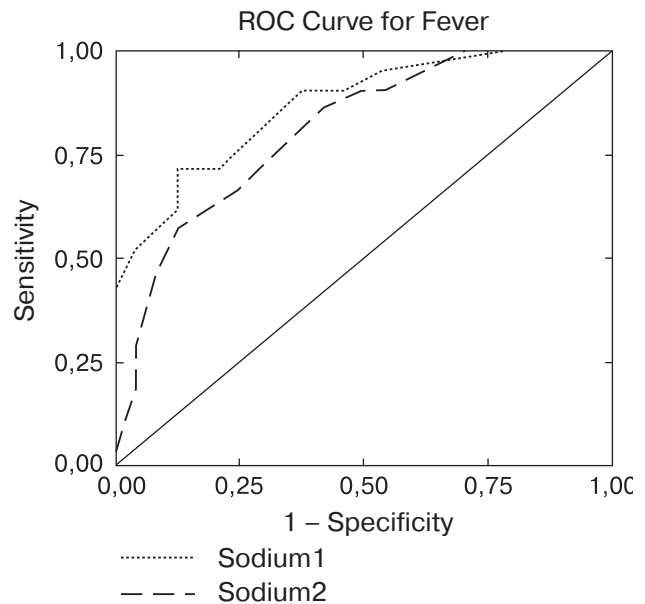


Рис. 5. ROC-кривые, полученные на компьютере с помощью программы NCSS.

При анализе ROC-кривых придерживаются следующего принципа: *чем ближе к левому верхнему углу координатной сетки расположена кривая, тем выше информативность исследуемого метода диагностики или лучше качество системы отображения данных*. Если кривая прилежит к диагонали (или совпадает), то информативность метода ничтожна. Необходимо отметить, что в качестве истинно положительных решений может выступать критерий “чувствительность”, а в качестве ложноположительных – критерий “1 – специфичность”.

Табличный принцип составления ROC-кривых не является единственным. В настоящее время некоторые статистические пакеты (например, NCSS) содержат программы для выполнения автоматизированного ROC-анализа (рис. 5).

В заключение необходимо остановиться на *мета-анализе как наивысшем методе доказательности в медицине, в том числе в лучевой диагностике*.

Метаанализ – это количественный анализ нескольких результатов одного и того же метода диагностики (или лечения), объединенных согласно существующим правилам этого анализа. Данный метод дает более мощный статистический результат, так как он получен в различных лечебных центрах и разными исследователями. Иногда его называют *многочетным анализом*.

Метаанализ содержит несколько этапов – обычно 4. На первом из них выявляются наиболее надежные публикации (построенные, как правило, на основе контролируемой рандомизации), которые затем подлежат обобщению и дальнейшему анализу. На втором этапе дается оценка качества исследований, приведенных в каждой публикации, их соответствие установленным критериям (подбор пациентов, характер проведенных исследований, особенности использованных методов и др.).

На третьем этапе производят обобщение всех результатов, то есть формируется одно общее исследование, в котором имеется несколько составляющих, имеющих свой “вес” и репрезентативность. На четвертом этапе осуществляется математическая обработка полученных объединенных результатов (чаще всего на основе регрессионного анализа) и все данные отображаются графически, таблично или в виде обобщенного показателя.

По мета-анализу в настоящее время накоплена большая литература. По данным MEDLINE число публикаций по этой пробле-

ме приближается к 5000, среди которых немало относится к лучевой диагностике (например, одна из последних публикаций – 15). Такие статьи представляют наибольший интерес для выработки коллективной стратегии и формирования общественного мнения по изучаемой проблеме.

В заключение целесообразно привести мнение авторитетного в этой области исследователя [16]. Приводим некоторые данные из его статьи, выдержанной в стиле компьютерной презентации.

Логика метаанализа

· Традиционные методы обзора данных сосредоточены в основном на *статистической значимости* тестирования.

· Значение тестирования не полностью коррелирует с этой задачей, так как:

– высока зависимость полученных данных от размеров группы наблюдений,

– истинные значения теста не соответствуют его “весу” в диагностической действительности.

· Метаанализ изменяет сосредоточенность полученных результатов по *направлению и амплитуде* эффекта изучаемого явления:

– Полученные данные представляют ли интерес во всех других наблюдениях?

– Направленность и амплитуда репрезентативна ли при изменении величины наблюдений?

Когда можно применить метаанализ?

Мета-анализ применим для коллективных исследований, которые:

· дают результаты в большей степени эмпирические, чем теоретические,

· в значительно большей степени продуцируют количественный результат, чем качественные находки,

· используют однородные конструкции исследований и взаимоотношений,

· получающиеся результаты могут быть сходными по статистическим формам (эффект размера выборки, коэффициент корреляции и др.),

· он дает сравнимые вопросы в руки исследователей.

Список литературы

1. <http://www.Evidence-Based/Radiology/A/Primer/for/Referring...>

