

© С.А. Трущелёв, 2014
УДК 616.89-072.85

Для корреспонденции

Трущелёв Сергей Андреевич – кандидат медицинских наук, доцент, ведущий научный сотрудник отделения информатики и системных исследований в психиатрии ФГБУ «Московский научно-исследовательский институт психиатрии» Минздрава России

Адрес: 107076, г. Москва, ул. Потешная, д. 3

Телефон: (495) 963-76-26, (495) 963-09-97

E-mail: sat-geo@mail.ru

С.А. Трущелёв

Условия применения диагностических тестов в психиатрии (аналитический обзор)

Diagnostic tests in psychiatry:
conditions of application
(analytical review)

S.A. Truschelev

This literature review highlights the conditions of application of psychometric diagnostic tools in psychiatry. With psychometrics being one of the components in psychiatric diagnosis in modern psychiatry, a point is raised about the need for expert certification in this field and basic operating characteristics of diagnostic tools are discussed that should be reflected in instructional-technical documentation.

Key words: psychiatry, psychometrics, clinical psychology, diagnosis, methodology

ФГБУ «Московский научно-исследовательский институт психиатрии»
Минздрава России
Moscow Research Institute of Psychiatry

Обзор литературы посвящен условиям применения диагностических психометрических инструментов в психиатрии. В связи с тем что психометрия в современной психиатрии является одним из компонентов диагностического процесса, поднимается вопрос о необходимости сертификации специалистов в этой области и рассматриваются основные операционные характеристики диагностических инструментов, которые должны быть отражены в инструктивно-технологической документации.

Ключевые слова: психиатрия, психометрия, клиническая психология, диагностика, методология

Подход, связанный с использованием в клинической и исследовательской практике формализованных методов оценки состояния здоровья, иногда подвергается критике. Однако ни одно исследование, претендующее на научную значимость, не обходится без применения тестов, шкал, опросников или анкет. Во многом благодаря разработке методов количественных и качественных измерений психической жизни психология и психиатрия стали научными дисциплинами. Теперь современную психиатрию невозможно представить без психометрии – науки о способах измерения феноменов психической деятельности человека [11, 14, 26].

Психометрические методики как диагностические тесты в психиатрии применяют в разных целях: массовое обследование населения для выявления больных или лиц с высоким риском того или иного заболевания (скрининг), поиск сопутствующих болезней, уточнение диагноза (собственно диагностическое исследование), оценка эффективности лечения [25, с. 21; 32, с. 11; 37, с. 90]. Динамические патопсихологическое и нейропсихологическое исследования имеют особое значение при принудительном лечении, поскольку позволяют верифицировать клинические решения об изменении тактики психофармакотерапии и правомерности перехода к следующему этапу

лечения [3, с. 5]. Диагностические тесты используют в научно-исследовательской деятельности и клинической практике для оценки глубины и изменчивости психического расстройства, для определения конкретных задач по лечению и разработке программ медицинской реабилитации [8, с. 158; 24, с. 400; 30, с. 57]. Расширение такого подхода в диагностике объясняется рядом обстоятельств: во-первых, изменения личности больного обладают большой нозологической специфичностью и могут быть использованы для решения вопросов дифференциальной диагностики; во-вторых, результаты анализа преморбидных свойств личности могут оказаться полезными в установлении причин происхождения ряда заболеваний (и не только психических, но и психосоматических, например, язвенной болезни, гипертонической болезни и др.); в-третьих, характеристика личностных изменений в период болезни обогащает представления о патогенетических механизмах; в-четвертых, изучение особенностей личности больного очень важно для построения рационального комплекса лечебно-профилактических и реабилитационных мероприятий [2, с. 12].

Многие методы, применяемые в психиатрии для решения экспертных задач диагностики и лечения, в известной мере субъективны. Об этом свидетельствуют результаты исследования вопросов диагностики и терапии биполярного аффективного расстройства, полученные С.Н. Мосоловым и Е.Г. Костюковой [15]. В последние годы это заболевание является одной из наиболее широко обсуждаемых медицинских проблем в связи как со значительной распространенностью, так и с трудностями его диагностики и лечения. Обобщая литературные данные, авторы установили, что пациентам с биполярным расстройством ошибочный диагноз первоначально ставят в 73% случаев, а правильный диагноз определяется в среднем через 8 лет от начала болезни, чаще после обследования и наблюдения у трех разных специалистов.

Все настойчивее веление клинической практики – применение объективных и чувствительных методов и методик диагностики болезней и психических нарушений [38, с. 75]. Эта потребность связана с существенным увеличением глобального бремени психических болезней [7, 19] и слабо-разработанной системой ранней диагностики [16, 18, 38]. Перед научным сообществом стоит актуальная проблема по поиску высококачественных диагностических инструментов, пригодных для использования в клинической практике [3, с. 10; 24, с. 400; 28, с. 147]. Отсюда закономерно следует вопрос, по каким критериям можно судить о качестве диагностических тестов? В связи с этим **целью** сообщения стал обзор современных требований, предъявляемых к психодиагностическим методикам и применяемых в психиатрии.

Требования к внедрению диагностических методик

Диагностический инструмент (тест, шкала), как любая другая медицинская технология, должен сопровождаться инструктивно-технологической документацией, в которой должны быть отражены его операционные характеристики. Такие инструменты должны обладать репрезентативностью и надежностью, их необходимо валидизировать в заданной области применения. Эффективность диагностического инструмента по выявлению определенных патологических признаков болезни должна характеризоваться данными о чувствительности, специфичности, прогностической ценности положительного и отрицательного результата.

Кроме того, в инструкции по применению конкретной методики обязательно должна приводиться характеристика пациентов, включенных в исследование, и обосновываться точка разделения больных и здоровых. Исходная группа больных должна быть исчерпывающим образом определена и охарактеризована. В нее должны включаться пациенты с детально описанной болезнью (ее стадия, форма и т.д.). Так, если не учтена фаза болезни, то нельзя напрямую использовать результаты исследования. Источник подбора больных чрезвычайно важен, так как это определяет, в каких условиях результаты исследования могут быть применимы. Систематические ошибки при отборе больных могут возникнуть вследствие доступности диагностики (diagnostic access bias). Необходимо еще помнить о том, что вследствие низкой доступности эффективных средств диагностики и лечения у людей, малообеспеченных или проживающих в отдаленных районах, возможно худшее течение заболевания и редкое его выявление [5].

Если в инструкции не приведены эти данные, то квалифицированные пользователи должны провести стандартизационные исследования в своей области, разработать частные нормы, направить их в головную методическую организацию и получить одобрение на использование диагностического теста. Процедура обработки результатов должна включать обоснованные методы статистического анализа и стандартизации, а выводы (диагностические суждения) должны сопровождаться указанием значений методов точечной и/или интервальной оценки и вероятностного уровня их статистической значимости [4, 17, 18, 20, 21].

Стандартизация – это единообразие процедуры проведения теста и оценки его результатов. Следовательно, необходимо обеспечить выработку единых требований к процедуре тестирования. Стандартизация процедуры подразумевает унификацию инструкций, бланков обследования, способов регистрации результатов, условий проведения обследования [1].

Воспроизводимость представляет собой важнейшее условие успешности внедрения и использования диагностических методов в клинической практике и научных исследованиях. В широком смысле под воспроизводимостью понимаются надежность и сопоставимость интерпретации результатов метода, которые оцениваются путем сравнения результатов исследования одного и того же объекта разными специалистами или различными способами. Такие работы возможны только с участием координирующей (головной) научно-методической организации.

Результативность метода диагностики в значительной мере зависит от мастерства обеспечивающего специалиста, который должен обладать необходимыми компетенциями и навыками работы в этой области. Стремительное развитие специализированных областей знания, усложнение самих диагностических методов и средств их обеспечения требуют от специалистов соответствия высокому качеству предоставляемой услуги и наличия способностей в применении профессиональных знаний и умений. Вместе с тем потребитель предоставляемой услуги должен быть уверен в качестве результата диагностики.

Наличие головной научно-методической организации, в которой аттестуется методика и ведется банк данных о ее применении, а также организованы подготовка и переподготовка сертифицированных пользователей – одно из важных требований, предъявляемых к качеству диагностических тестов, используемых в медицинской практике.

Методика, не удовлетворяющая перечисленным выше требованиям, считается не пригодной для использования в медицинской практике. Неаттестованную методику можно применять только в научно-исследовательских целях, но и при этом ее необходимо использовать в сравнении с аттестованной (с эталонным методом). Опубликованные результаты исследования или копию отчета о научно-исследовательской работе необходимо направить в головную научно-методическую организацию, которая в свою очередь обязана периодически публиковать результаты обобщения поступивших данных и своевременно вносить изменения в методические рекомендации и инструкции.

Стандартный критерий диагностики

Для того чтобы оценить диагностическую эффективность метода исследования, используют его способность распознавать заболевание в сравнении со стандартным критерием диагностики, эталоном, или так называемым «золотым стандартом»

диагностики. В последние годы активно внедряется термин «стандартный критерий» – понятие, которое присваивают методике (тесту), обладающей наивысшей информативностью в разумных условиях и максимально эффективно отражающей состояние пациента [10, с. 213]. Так уж заведено, что в современной медицинской практике к таким стандартам относят высокотехнологичные (генетические, гистологические, биохимические и др.) методы исследования, аутопсию (результаты вскрытия), клинически явные признаки болезни. Иногда под «золотым стандартом» понимают совокупность методов, максимально полно и быстро решающих вопросы диагностики. Однако в большинстве случаев такие диагностические методы малодоступны из-за сложности и уникальности применяемых медицинских технологий.

Диагностические тесты, используемые в клиническом обследовании, не идеальны, а результаты исследования всегда содержат вероятность ошибки распознавания болезни. Как правило, для любого клинического исследования имеется несколько диагностических альтернатив различной степени точности. Однако широкое применение тех или иных методов обследования ограничено различиями доступности, стоимости, безопасности, а также факторами времени и здравого смысла. Тем не менее принято для конкретной болезни выделять эталонный метод диагностики – наиболее точный и эффективный метод, с помощью которого можно установить болезнь или обосновать ее отсутствие. Поскольку эталонный метод не обладает 100% точностью, то с развитием научного знания и технологий его могут сменить другим, более точным.

Характеристика теста, обозначаемая как эффективность, предполагает использование данного метода в обычных (усредненных) условиях его применения. Понятно, что эффективность всегда будет ниже, чем сила метода. В связи с этим для обозначения этого понятия рекомендуют термин «реальная эффективность».

При применении эталонного метода (стандартного критерия) диагностики необходимо приводить ссылку на регламентирующий документ и указывать его основные операционные характеристики для избранного объекта исследования.

В клинической практике при не вполне ясном течении болезни, как правило, требуется расширение диагностических мероприятий. В таких случаях, если недоступен по каким-то причинам эталонный метод диагностики, прибегают к методикам, обладающим наилучшими информационными показателями воспроизводимости в данной области. Порой

эти методы имеют большое диагностическое значение в распознавании болезни и являются весомым обоснованием применения эталонного метода. Например, известны случаи болезни Ниманна–Пика – редкого мультисистемного наследственного ауто-сомно-рецессивного заболевания, относящегося к группе лизосомных болезней накопления. Болезнь представляет значительные диагностические сложности, обусловленные широким фенотипическим полиморфизмом психических и соматических нарушений, может дебютировать в любом возрасте [9]. «Золотым стандартом» диагностики при этом заболевании признан молекулярно-генетический анализ (метод прямого секвенирования гена *NPC1*). Однако такое исследование проводят только в крупных специализированных медико-диагностических центрах. В связи с этим, чтобы иметь достаточно обоснованное решение для направления на это исследование, применяют весьма доступный тест – Индекс вероятности болезни Ниманна–Пика типа С (www npc-si.ru/works). Основываясь на результатах, полученных при валидации, авторы теста выделили прогностические категории вероятности риска болезни. Так, при показателе 70 баллов и более специфичность теста составляет 95%, чувствительность – 80%. Это соответствует высокой степени вероятности болезни. В таком случае врач обязан направить пациента в ближайший центр для незамедлительного проведения генетического тестирования. Индекс вероятности болезни Ниманна–Пика типа С не относится к психометрическим методикам; это медицинский калькулятор. Разработка этого диагностического инструмента – хороший пример информационной поддержки врачебной деятельности.

Операционные характеристики диагностического теста

Достижения в области биостатистики в целом позволяют получить представление о реальной эффективности диагностических тестов, однако многие практикующие специалисты и исследователи часто полагаются на неполные данные о показателях психометрических тестов и их воспроизводимости [35]. Воспроизводимость диагностического теста характеризуют объективными показателями, именуемыми операционными характеристиками. К ним относят чувствительность и специфичность [6, 22, 23]. Эти показатели рассчитывают на ос-

нове группировки первичных данных в 4-польной таблице сопряженности (табл. 1). К вспомогательным критериям информативности относят точность, вероятность болезни, прогностическую ценность теста.

Из табл. 1 следует, что с помощью стандартного критерия диагностики («золотого стандарта») выявлено количество больных, которое указано в ячейке A+C, и здоровых – в ячейке B+D. С помощью теста сравнения выявлено больных – A+B и здоровых – C+D, истинно-положительные результаты – A (true positive, TP), истинно-отрицательные – D (true negative, TN), ложноположительные – B (false positive, FP), ложноотрицательные – C (false negative, FN). Термины «положительный результат теста» и «негативный результат теста» используют в их обычном значении – для обозначения наличия или отсутствия заболевания соответственно.

Интерпретация понятия «отрицательный» у пациентов с наличием заболевания следует понимать как ложноотрицательный случай, пропуск диагностики болезни. В теории проверки гипотез такие пропуски называют ошибками первого рода, или α -ошибками. Интерпретация понятия «положительный» у здоровых пациентов – ложноположительный случай, случаи «ложной тревоги», гипердиагностики. В теории проверки гипотез такие пропуски относятся к ошибкам второго рода, или β -ошибкам.

Чувствительность (sensitivity, Se) – показатель способности теста на выявление истинно больных; указывает на долю лиц с заболеванием, имеющих положительный результат диагностического теста. Формула для расчета: $Se = A / (A + C)$. Показатель изменяется от 0 до 1. Показатель может выражаться в процентах. Чем ближе полученная оценка чувствительности к 1 (или 100%), тем лучше проверяемый тест диагностирует наличие болезни у пациентов. Высокочувствительные тесты нередко именуют идентификаторами. Тесты с высокой чувствительностью рекомендуется применять на ранних этапах диагностического процесса (для скрининга), когда требуется сузить круг предполагаемых заболеваний. Высокочувствительные тесты редко пропускают случаи заболевания. Необходимо отметить, что высокочувствительные тесты часто вызывают «ложную тревогу» и вследствие этого возникает необходимость дальнейшего обследования.

Специфичность (specificity, Sp) – доля лиц без заболевания, имеющих отрицательный результат

Таблица 1. Форма группировки первичных данных, полученных в сравнительном исследовании

Тест сравнения	Заболевание (диагностировано по стандартному критерию диагностики)		Итого по строке
	имеется	отсутствует	
Положительный	A	B	A+B
Отрицательный	C	D	C+D
Всего по графе	A+C	B+D	A+B+C+D

теста (т.е. доля здоровых, признанных здоровыми). Формула для расчета: $Sp=D/(B+D)$. Показатель изменяется от 0 до 1. Показатель может выражаться в процентах. Чем ближе полученная оценка специфичности к 1 (или 100%), тем лучше проверяемый тест распознает отсутствие болезни у пациентов. Высокоспецифичные тесты редко дают ложноположительные результаты, т.е. обладают высокой селективностью. Чем выше специфичность метода, тем надежнее с его помощью подтверждается заболевание, следовательно, он более эффективен. Высокоспецифичные тесты называются в диагностике дискриминаторами. Такие методы эффективны на втором этапе диагностики, когда круг предполагаемых заболеваний сужен и необходимо с большой уверенностью доказать наличие болезни. Отрицательным фактором высокоспецифичного теста является то, что его использование сопровождается весьма значительным числом пропусков заболевания.

Для расчетов созданы многообразные средства автоматизации. Для примера приводим компьютерную программу (рис. 1), разработанную Rob Herbert (University of uOttawa, Canada, 2002). Переведенную версию калькулятора можно получить на сайте Московского научно-исследовательского института психиатрии (www.mniip.webservis.ru). С помощью этой программы на основе данных 4-польной таблицы (см. табл. 1) вычисляют показатели чувствительности и специфичности. При этом отметим, поскольку данные обычно приводят по результатам выборочных исследований, то в научных публикациях необходимо приводить не только показатели центральной тенденции (среднее, медиана), но и их доверительные интервалы (ДИ) [12].

Сравнение посредством ДИ возможно, но оно обладает меньшей мощностью, чем применение типовых критериев статистического анализа, т.е. будет завышена вероятность пропуска различия там, где оно есть. Такой метод сравнения лучше делать в тех

случаях, когда у исследователя нет полного набора оригинальных данных (например, при сравнении собственных данных с опубликованными).

Иногда возникает необходимость сравнения показателей чувствительности или специфичности одного теста с соответствующими данными сравнимого теста. В этих случаях подход таков: поскольку чувствительность – это показатель доли, то мы имеем полное право обрабатывать ее, как и любую другую долю. Это же правило относится к специфичности. Например, значения чувствительности теста можно сравнить методом хи-квадрат МакНемара на основе 4-польной таблицы (см. табл. 1). Аналогично со специфичностью. Если хи-квадрат покажет статистически значимое различие, следует проанализировать, за счет каких ячеек таблицы оно проявилось; это делается с помощью расчета отклонений по Фриману–Тьюки.

Вспомогательные критерии информативности

Точность (ассурасу, Ac) – доля правильных результатов теста в общем количестве результатов. Этот показатель рассчитывают по формуле: $Ac=A+D/A+B+C+D$, т.е. отношение суммы истинно-положительных и истинно-отрицательных результатов теста к общему количеству результатов. Показатель может выражаться в процентах. Применяют его для оценки эффективности методов измерения медико-биологических данных. Точность измерений – это соответствие результатов измерения истинному значению определяемой величины. Высокая точность измерения достигается при минимальных рандомизированных и систематических погрешностях. Чем они меньше, тем правильнее измерения. Иногда этот критерий называют диагностической эффективностью (diagnostic efficiency, De). Точность диагностического метода зависит от квалификации исследователя, воспроизводимости самого метода, используемого оборудования, выбранного крите-

ОПЕРАЦИОННЫЕ ХАРАКТЕРИСТИКИ ДИАГНОСТИЧЕСКОГО ТЕСТА			
Введите данные в эту таблицу:		«Золотой стандарт» положителен	«Золотой стандарт» отрицателен
Тест положителен		55	9
Тест отрицателен		12	40
Введите необходимую точность доверительного интервала (напр. 95%) 95			
РЕЗУЛЬТАТ:			
Чувствительность:	0,8209	ДИ: от 0,7125 до 0,8945	
Специфичность:	0,8163	ДИ: от 0,6864 до 0,9002	
Отношение правдоподобия для положительного результата:	4,469	ДИ: от 2,451 до 8,15	
Отношение правдоподобия для отрицательного результата:	0,219	ДИ: от 0,129 до 0,373	
Диагностическое отношение шансов:	20,8209	ДИ: от 7,835 до 52,958	

Рис. 1. Экранная форма компьютерной программы для вычисления операционных характеристик диагностического теста

Таблица 2. Способ расчета прогностической ценности теста по Killeen [33]

Показатель	Наличие болезни	Отсутствие болезни
Априорная вероятность	0,1	0,9
Претестовая вероятность положительного результата теста	0,95	0,05
Общая расчетная вероятность	$\frac{a}{0,1 \times 0,95} = 0,095$	$\frac{b}{0,9 \times 0,05} = 0,045$
Прогностическая ценность теста (PPV)	$\frac{a}{a+b}$ 0,68	$\frac{b}{a+b}$ 0,32

рия патологии, популяции, в которой данный тест используется.

Раньше качество (точность) диагностических исследований оценивали путем сравнения результатов явно больных людей и заведомо здоровых добровольцев. Естественно, что результаты отражали значительные выраженные различия между группами. Для реальной практики результаты таких исследований часто оказываются малоприменимыми, поскольку в скрытых случаях болезни дают случайную вероятность наличия болезни, а целью диагностических исследований часто является распознавание именно неявных симптомов заболевания.

Распространенность (вероятность болезни, probability, P) определяется как отношение числа лиц с наличием заболевания (или любого другого состояния) ко всей исследуемой популяции. Показатель рассчитывают по формуле $P = (A+C)/(A+B+C+D)$.

Показатель вероятности болезни до тестирования называют априорной (претестовой) вероятностью, а показатель, полученный после тестирования, называется прогностической ценностью, или апостериорной (посттестовой) вероятностью заболевания. Прогностическая ценность положительного результата теста связана с чувствительностью, специфичностью и распространенностью заболевания и рассчитывается по формуле $PPV = (Se \times P) / [(Se \times P) + (1 - Se) \times (1 - P)]$. Расчет показателя претестовой вероятности применяют при отборе наиболее эффективных диагностических тестов (методов исследований).

Распространенность заболевания влияет на прогностическую ценность теста, однако эта связь редко объясняется. В связи с этим А.А. Killeen [33] предлагает определять положительную прогностическую ценность теста на основе байесовского анализа (табл. 2). Расчет проводится исходя из показателей априорной вероятности болезни (a priori probability) и вероятности положительного результата теста (probability of a positive test result). Например, при априорной распространенности заболевания в 10% (0,1) и претестовой вероятности в 95% (0,95) вероятность положительного теста у больных составит 68% (0,68), а у здоровых – 32% (0,32).

Распространенность заболевания в популяции сильно влияет на положительную прогностическую

ценность диагностического теста – с увеличением распространенности заболевания его вероятность становится больше.

Прогностическая ценность теста

С практической точки зрения для оценки результатов исследования интерес представляет вероятность совпадения заключения с окончательным диагнозом. Для этих целей оцениваются показатели прогностичности.

Прогностическая ценность (predictive value) – способность диагностического теста определить наличие (отсутствие) заболевания при известном результате исследования. Показатель зависит как от чувствительности и специфичности теста, так и от распространенности заболевания в исследуемой популяции. Выделяют прогностичность положительного и отрицательного результата.

Прогностичность положительного результата (positive predictive value, PPV) – доля правильных положительных результатов диагностического теста. Формула: $PPV = A/(A+B)$. Показатель определяется как частота его совпадения с заболеванием. Он показывает, насколько велика вероятность наличия болезни (синдрома, симптома) при положительных результатах исследования.

Прогностичность отрицательного результата (negative predictive value, NPV) – доля правильных отрицательных результатов диагностического теста. Формула: $NPV = C/(C+D)$. Этот показатель определяется как частота его совпадения с отсутствием заболевания. Данный критерий показывает, насколько велика вероятность того, что пациент здоров, если результаты исследования отрицательные.

Чем чувствительнее тест, тем выше прогностическая ценность его отрицательного результата (т.е. возрастает уверенность врача в том, что отрицательные результаты исследования отвергают наличие заболевания). Наоборот, чем специфичнее тест, тем выше прогностическая ценность его положительного результата (т.е. врач может с большей уверенностью считать, что положительные результаты подтверждают предполагаемый диагноз). Поскольку распространенность заболевания влияет на прогностическую ценность диагностического метода, последняя неизбежно зависит и от условий его выполнения. Если положительные резуль-

таты даже высокоспецифичного метода получены в популяции с низкой вероятностью заболевания, то они окажутся преимущественно ложноположительными.

Производные критерии информативности

Перечисленные выше критерии информативности диагностики базируются на принципах дихотомических решений: «да»–«нет», «норма»–«патология». Однако хорошо известно, что в клинической практике врача не всегда удается классифицировать получаемые данные по такой схеме. В ряде случаев специалисты используют другие заключения, такие как, например, «наиболее вероятно, заболевание имеется» или «наиболее вероятно, заболевание отсутствует». Подобные нюансы в принятии врачебных заключений отражают другие характеристики информативности – отношения правдоподобия [10, с. 219].

Отношение правдоподобия результатов диагностического теста (likelihood ratio, LR) – вероятность того, что результат диагностического теста будет ожидаться у пациента с заболеванием по сравнению с таким же результатом у пациента без заболевания. Иначе говоря, тест показывает, во сколько раз выше (ниже) вероятность получить данный результат теста от больных, нежели от здоровых.

Вычисление показателей отношения правдоподобия для положительного результата (LR+) и отрицательного результата (LR–) производится на основе табличных данных (см. табл. 1) или значений чувствительности и специфичности теста, по формулам:

$$LR+ = [A/(A+C)]/[B/(B+D)] = Se/(1-Sp);$$

$$LR- = [C/(A+C)]/[D/(B+D)] = (1-Se)/Sp.$$

Если отношение правдоподобия положительного результата теста равно 1, то вероятность положительного результата теста у больного такая же, как и вероятность положительного результата теста у здорового. Если этот показатель, например, равен 5, то вероятность положительного теста у больного в 5 раз выше, чем вероятность того же результата теста у здорового.

В науке широко используют прогностические модели, рассчитанные с помощью номограмм – графических двумерных инструментов, которые позволяют определить приблизительное значение вероятности результата. Например, по номограмме Фегана [29] на основе показателей претестовой вероятности и отношения правдоподобия можно рассчитать посттестовую вероятность болезни (рис. 2). Так, если показатель вероятности в предварительном тестировании (pre-test probabili-

ty) лежит между 30 и 70%, то показатель отношения правдоподобия теста будет очень высоким. Очень низкий показатель теста, полученный из отношения правдоподобия (ниже 0,1), практически исключает вероятность того, что пациент имеет заболевание.

Показатель отношения правдоподобия используют для оценки применимости диагностического теста. Преимущество этого показателя состоит в том, что он помогает выйти за рамки грубой классификации результатов теста (норма–патология), с которой сталкиваются пользователи, если описывают точность теста только в терминах чувствительности и специфичности при единственной точке разделения.

ROC-анализ

Для того чтобы оценить диагностическую эффективность метода с учетом последствий ложных решений, используют метод составления характеристик

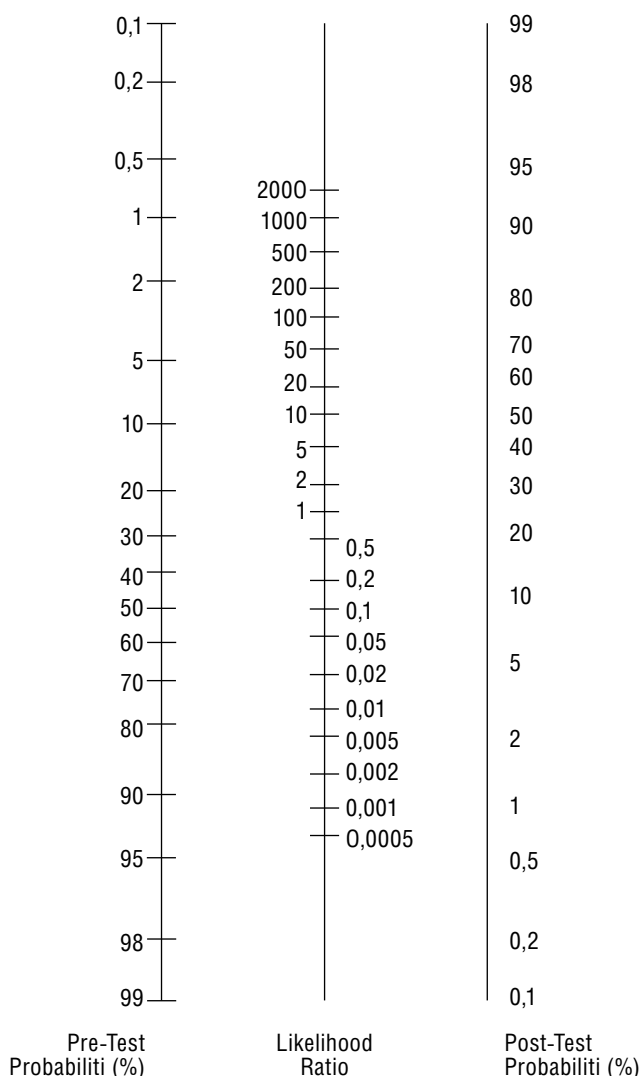


Рис. 2. Номограмма отношения правдоподобия (Likelihood Ratio Nomogram) [34]

Таблица 3. Классификация качества модели в зависимости от площади под характеристической кривой [13]

Интервал значений площади под характеристической кривой	Качество модели
0,91–1,0	Отличное
0,81–0,90	Очень хорошее
0,71–0,80	Хорошее
0,60–0,70	Среднее
Менее 0,60	Неудовлетворительное

тической кривой (Receiver Operating Characteristic Curve, ROC-curve). Действия, выполняемые для ее построения, называют ROC-анализом [31, 39]. Поскольку ROC-кривая отражает взаимозависимость ложноположительных и истинно-положительных результатов, такой подход позволяет наглядно сопоставить диагностическую эффективность различных методов при обнаружении одного и того же патологического процесса. Для этого, например, формируют группу врачей-специалистов, которые с помощью двух методов проводят диагностику пациентов с верифицированным диагнозом. Результаты, имеющие количественный характер или измеренные в порядковой шкале измерений, ранжируют [20]. Полученные таким образом данные сводят в таблицы и вычисляют показатели специфичности и чувствительности, а затем по ним строят ROC-кривые и вычисляют под ними площадь.

Площадь под ROC-кривой (area under ROC-curve, AUC) – это количественное значение ROC-анализа. Чем выше показатель AUC, тем качественнее классификатор, при этом значение 0,5 соответствует непригодности выбранного метода классификации. В литературе [13] приводится экспертная шкала для значений AUC, по которой можно судить о качестве модели (табл. 3).

Надо сказать, что процесс расчетов трудоемкий, но с помощью современных компьютерных технологий он существенно облегчается.

Пример. В гипотетически сформированной группе из 50 участников исследования проведена оценка 5-балльного диагностического теста на наличие признака в сравнении с результатами стандартного критерия диагностики. Задачи: определить чувствительность и специфичность диагностического метода, рассчитать оптимальную точку распределения «больной/здоровый», а также установить качество диагностической модели, определяемой по площади под ROC-кривой. Для расчетов применили компьютерную программу MedCalc (Ver.12.7.8; MedCalc Software, 2014; Belgium).

Заключение о диагностической эффективности метода получено на основе результатов анализа данных (рис. 3–5): в группе у 25 из 50 пациентов выявлен признак болезни. В точке разделения «больной/здоровый», соответствующей 2 баллам (на рис. 3 отмечено звездочкой), определены оптимальные значения чувствительности и специфичности – 88,0 и 80,0%, соответственно. Площадь под ROC-кривой составляет 0,892 (95% доверительный интервал от 0,772 до 0,962) – показатель, характеризующий хорошее качество модели диагностического теста (см. табл. 3). Исходя из этого тест можно использовать для диагностики изучаемого признака. Превышение значений тестирования более 2 баллов будет свидетельствовать о наличии признака болезни.

Ошибки в интерпретации

Систематические ошибки интерпретации результатов – распространенное явление. Заключение по данным инструментальных исследований часто

POSITIVE GROUP Признак = 1 Sample size = 25					
NEGATIVE GROUP Признак = 0 Sample size = 25					
Disease prevalence unknown.					
Area under the ROC curve = 0,892					
Standard error = 0,048					
95% Confidence interval = 0,772 to 0,962					
Criterion	Sens.	(95% C.I.)	Spec.	(95% C.I.)	+LR -LR
>=1	100,0	(86,2–100,0)	0,0	(0,0–13,8)	1,00 0,00
> 1	96,0	(79,6–99,3)	44,0	(24,4–65,1)	1,71 0,09
> 2 *	88,0	(68,8–97,3)	80,0	(59,3–93,1)	4,40 0,15
> 3	76,0	(54,9–90,6)	92,0	(73,9–98,8)	9,50 0,26
> 4	40,0	(21,2–61,3)	96,0	(79,6–99,3)	10,00 0,62
> 5	0,0	(0,0–13,8)	100,0	(86,2–100,0)	1,00 0,00

Рис. 3. Экранная форма вывода обобщенных результатов расчета в программе MedCalc

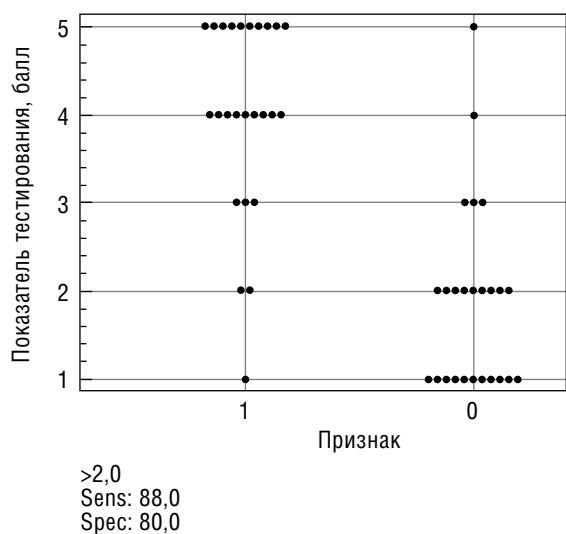


Рис. 4. Экранная форма результата расчета точки распределения «больной/здоровый» с помощью программы MedCalc

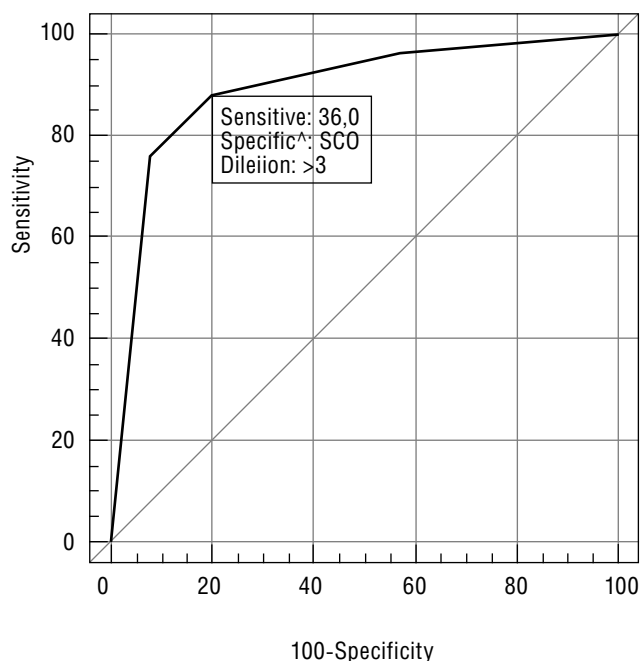


Рис. 5. Экранная форма вывода ROC-кривой диагностического теста

во многом субъективны и подвержены влиянию дополнительной клинической информации. В связи с этим применяют проверку на сходимость результатов клинического и инструментального методов исследования дважды: первый раз – не имея клинической информации, второй раз – располагая ею. Интерпретация результатов под влиянием клинической информации приводит к повышению соответствия между исследуемым тестами и стандартом диагностики, т.е. тест кажется более информативным, чем на самом деле [36].

Чаще всего случайная ошибка при оценке чувствительности и специфичности обусловлена тем, что информативность диагностической методики изучается в малых группах, поэтому результаты могут искажаться вследствие случайных вариаций данных, извлеченных из популяции с тестируемым заболеванием. Чтобы минимизировать ошибки такого рода, необходимо проводить исследования на достаточном количестве участников (предварительно рассчитанный объем выборки), формировать выборку в соответствии с условиями репрезентативности, в статистических расчетах значений в малых выборках применять корректирующие методы (например, введение поправки Йетса).

Заключение

Признавая эффективность психодиагностических тестов, вслед за Л.Н. Собчик [18] подчеркнем, что выводы об особенностях личности больного человека и о его психическом состоянии должны делаться на основании совокупности всех имею-

щихся о нем сведений. Недопустимо формирование диагноза психического расстройства только на основе психодиагностических данных. Диагностику состояния психического здоровья необходимо вести с позиций комплексного подхода, включающего не только признаки болезни и формальные показатели диагностических исследований, но и биографические данные, сведения о его профессиональном, социально-экономическом и семейном статусах, о той ситуации и жизненном пространстве, в которых он находится, о тех обстоятельствах (на момент обследования), которые могут оказать значительное влияние на результаты психометрического тестирования. Необходимо видеть пациента, оценить его общефизиологические и физиогномические особенности, наладить с ним контакт и развеять те опасения, которые может вызвать у него процедура обследования. Соблюдение этих условий позволит провести исследование и относиться с доверием к полученным результатам.

К сожалению, бывает и так, что профессиональные психодиагностические инструменты, совершенно не защищенные от неконтролируемого распространения, попадают в руки горе-энтузиастов или просто предприимчивых людей. Этому способствуют декларативность многих регламентирующих документов, отсутствие реакции разработчиков тестов на нелегитимное использование интеллектуальной собственности, неразработанность этических аспектов этой деятельности. Такое положение дел приводит к тому, что валидность известных и достаточно эффективных диагностических ме-



тодик резко снижается [27], они фактически выводятся из профессионального оборота. Широкое распространение компьютерной самодиагностики приводит пациентов не только к заблуждениям, но и зачастую наносит значительный ущерб их психике [4, с. 119]. В Россию хлынул поток тестов из-за рубежа. Однако установлено, что редко какой из них соответствует изложенным выше требованиям [4, с. 124].

Проблема совершенствования диагностики психических расстройств открыта и актуальна. Для

поисков в этой области требуются высокий уровень развития научно-исследовательских технологий и опора на дальнейшее изучение сложных психических процессов [18]. В связи с интенсификацией применения психометрических методик в психиатрии возникает ряд актуальных вопросов: отбор и разработка высококачественных диагностических инструментов, наличие головной научно-методической организации, специализация и усовершенствование обеспечивающих специалистов и др.

Литература

1. Ардашкин И.Б., Дубинина И.А. Психодиагностика: Учеб. пособие. – Томск: Изд-во ТПУ, 2001. – 100 с.
2. Блейхер В.М., Крук И.В. Патопсихологическая диагностика. – Киев: Здоров'я, 1986. – 280 с.
3. Булыгина В.Г., Макурина А.П., Жумагалиева М.Ю. и др. Проблема создания новых экспериментальных методов для решения прикладных задач в патопсихологии и судебно-психиатрической практике // Психология и право [Электронный ресурс]. – 2011. – № 4. – URL: http://psyjournals.ru/files/49293/psyandlaw_2011_4_Buligina.pdf (дата обращения 10.03.2014).
4. Бурлачук Л.Ф. Психодиагностика: Учебник для вузов. – СПб.: Питер, 2006. – 351 с.
5. Власов В.В. Как читать медицинские статьи: часть 2. Исследования, посвященные методам диагностики // Междунар. журн. мед. практики [Электронный ресурс]. – 1997. – № 2. – URL: <http://www.mediasphera.aha.ru/mjmr/97/2/2-97-1.htm> (дата обращения 18.01.2014).
6. Власов В.В. Эпидемиология: Учебное пособие для вузов. – 2-е изд., испр. – М.: ГЭОТАР-Медиа, 2005. – 464 с.
7. Гурович И.Я., Любов Е.Б., Чапурин С.А. и др. Бремя депрессивных расстройств в отечественных психиатрических службах // Журн. неврол. и психиатр. им. С.С. Корсакова. – 2010. – Т. 110, № 3. – С. 77–82.
8. Жариков Н.М., Морозов Г.В., Хритинин Д.Ф. Судебная психиатрия: Учебник для вузов / Под общ. ред. акад. РАМН Г.В. Морозова. – 3-е изд., перераб. и доп. – М.: Норма, 2004. – 528 с.
9. Ключников С.А., Смирнов О.Р., Захарова Е.Ю. Случай болезни Ниманна–Пика типа С // Неврология, нейропсихиатрия, психосоматика. – 2013. – № 4. – С. 43–48.
10. Королюк И.П. Медицинская информатика: Учебник. – 2-е изд., перераб. и доп. – Самара: Офорт; ГБОУ ВПО СамГМУ, 2012. – 244 с.
11. Курсаков А.А. К вопросу об эволюции и перспективах развития психометрического метода: опыт применения клинического интервью для оценки депрессии и смежных синдромов // Психиатрия и психофармакотер. – 2013. – Т. 15, № 3. – С. 78–81.
12. Ланг Т.А., Сесик М. Как описывать статистику в медицине. Аннотированное руководство для авторов, редакторов и рецензентов / Пер. с англ. под ред. В.П. Леонова. – М.: Практическая медицина, 2011. – 480 с.
13. Логистическая регрессия и ROC-анализ – математический аппарат // BaseGroup Labs: технологии анализа данных [Электронный ресурс]. – URL: <http://www.basegroup.ru/library/analysis/regression/logistic> (дата обращения 18.01.2014).
14. Мосолов С.Н. Шкалы психометрической оценки симптоматики шизофрении и концепция позитивных и негативных расстройств. – М.: Новый цвет, 2001. – 238 с.
15. Мосолов С.Н., Костюкова Е.Г. Современные подходы к диагностике и фармакотерапии биполярного аффективного расстройства // Трудный пациент [Электронный ресурс]. – 2005. – № 12. – URL: <http://t-pacient.ru/articles/6566/> (дата обращения 10.03.2014).
16. Носс И.Н. Психодиагностика: Учебник. – 2-е изд. – М.: Юрайт, 2014. – 500 с.
17. Носс И.Н., Васина Н.В. Введение в практику психологического исследования: Учебное пособие. – М.: Изд-во Института психотерапии, 2004. – 348 с.
18. Собчик Л.Н. Психодиагностика в медицине: практическое руководство. – М.: БОРГЕС, 2007. – 416 с.
19. Трущелёв С.А. Глобальное бремя психических болезней // Здравоохран. Рос. Федерации. – 2009. – № 4. – С. 33–39.
20. Трущелёв С.А. Общие вопросы методологии статистического анализа: типы данных и алгоритмы подбора методов // Рос. психиатр. журн. – 2014. – № 1. – С. 68–73.
21. Трущелёв С.А. Решение задач описательной статистики средствами пакета анализа Microsoft Excel // Рос. психиатр. журн. – 2013. – № 2. – С. 38–42.
22. Altman D.G., Bland J.M. Statistics Notes: Diagnostic tests 1: sensitivity and specificity // BMJ. – 1994. – Vol. 308. – P. 1552. – URL: <http://www.bmj.com/content/308/6943/1552.pdf%2Bhtml>
23. Baldessarini R.J., Finkelstein S., Arana G.W. The predictive power of diagnostic tests and the effect of prevalence of illness // Arch. Gen. Psychiatry. – 1983. – Vol. 40(5). – P. 569–573.
24. Bech P. Applied psychometrics in clinical psychiatry: the pharmacopsychometric triangle // Acta Psychiatr. Scand. – 2009. – Vol. 120. – P. 400–409.
25. Bech P. Clinical psychometrics. – Oxford: Wiley-Blackwell, 2012. – 200 p.
26. Bech P. Rating scales for psychopathology, health status and quality of life. – Berlin: Springer, 1993. – 520 p.

27. Berlin J., Rennie D. Measuring the quality of trials: the quality of quality scales // JAMA. – 1999. – Vol. 282. – P. 1083–1085.
28. Buwalda V.J., Nugter M.A., van de Brug S.Y. et al. ROM in adult psychiatry: an evaluation of measurement instruments // Tijdschr. Psychiatr. – 2012. – Vol. 54(2). – P. 147–152.
29. Fagan T.J. Nomogram for Bayes theorem // N. Engl. J. Med. – 1975. – Vol. 293. – P. 257.
30. Feighner J.P., Robins E., Guze S.B. et al. Diagnostic criteria for use in psychiatric research // Arch. Gen. Psychiatry. – 1972. – Vol. 26(1). – P. 57–63.
31. Gardner I.A., Greiner M. Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests // Vet. Clin. Pathol. – 2006. – Vol. 35. – P. 8–17.
32. Kagee A., Tsai A.C., Lund C. et al. Screening for common mental disorders in low resource settings: Reasons for caution and a way forward // Int. Health. – 2013. – Vol. 5 (1). – P. 11–14.
33. Killeen A.A. The Effect of Disease Prevalence on the Predictive Value of Diagnostic Tests // Ipathology [Electronic Resource]. – URL: <http://www.ipathology.com/ipathology/effect-of-disease-prevalence.html> (Accessed 18.01.2014).
34. Likelihood Ratios. – Centre for Evidence Based Medicine [Electronic Resource]. – 2013. – URL: <http://www.cebm.net/index.aspx?o=1043> (Accessed 18.01.2014).
35. Mackinnon A. A spreadsheet for the calculation of comprehensive statistics for the assessment of diagnostic tests and inter-rater agreement // Comput. Biol. Med. – 2000. – Vol. 30 (3). – P. 127–134.
36. Mallett S., Halligan S., Thompson M. et al. Interpreting diagnostic accuracy studies for patient care // BMJ. – 2012. – Vol. 345. – P. e3999.
37. Mosolov S.N., Ushkalova A.V., Kostukova E.G. et al. Validation of the Russian version of the hypomania checklist (HCL-32) for the detection of bipolar II disorder in patients with a current diagnosis of recurrent depression // J. Affect. Disord. – 2013. – Vol. 155. – P. 90–95.
38. Phillips J. Detecting diagnostic error in psychiatry // Diagnosis. – 2014. – Vol. 1 (1). – P. 75–78.
39. Zweig M.H., Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine // Clin. Chem. – 1993. – Vol. 39. – P. 561–577.