

СРАВНИТЕЛЬНАЯ ОЦЕНКА ПОРОГОВЫХ КРИТЕРИЕВ ВЫРАЖЕННОСТИ ДЕПРЕССИИ, РАЗРАБОТАННЫХ НА ОСНОВЕ КЛАССИЧЕСКОЙ ТЕОРИИ ТЕСТОВ И МОДЕЛИ РАША

М.А. Ассанович

*Гродненский государственный медицинский университет,
г. Гродно, Республика Беларусь*

Каждая используемая в психиатрии клиническая рейтинговая шкала должна иметь научно обоснованные критерии оценки выраженности измеряемого психопатологического конструкта. Цель описываемого в публикации исследования состояла в сравнительной оценке воспроизводимости и клинической применимости метрических критериев шкалы оценки тяжести депрессии Гамильтона (HRSD), разработанных в рамках двух подходов: по классической теории тестов (КТТ) и на основе модели Г. Раша (Georg Rasch). Для сравнительного анализа были сформированы три выборки протоколов исследования здоровых испытуемых и пациентов с депрессией, выполненных с помощью HRSD. Было выявлено, что критерии интерпретации, разработанные на основе классической теории тестов, отличаются низкой воспроизводимостью и не согласуются с клиническими данными. В отличие от критериев КТТ, разработанные на основе модели Раша метрические критерии отличаются высокой степенью воспроизводимости и эффективно согласуются с клиническими данными. Полученные результаты аргументированно объясняются концептуальными методологическими различиями обоих подходов в построении объективного измерения и определении стандартной ошибки измерения.

Ключевые слова: критерии интерпретации, шкала оценки депрессии Гамильтона, модель Раша, классическая теория тестов.

Введение

Каждая клиническая рейтинговая шкала, используемая в психиатрии, должна обладать надежными, валидными и точными критериями оценки выраженности измеряемого психопатологического конструкта. Такие критерии выводятся на основе разделения клинической шкалы на метрические уровни выраженности измеряемого конструкта. Каждый метрический уровень содержит суммарные оценки шкалы, которые со статистически заданной степенью вероятности эквивалентны определенной степени выраженности конструкта. Все оценки шкалы, входящие в один метрический уровень, имеют одинаковую интерпретацию, поскольку отражают один и тот же эквивалентный уровень конструкта. Таким образом, для интерпретации результатов психометрического исследования достаточно знать суммарные оценки, образующие границы метрических уровней шкалы (de Vet, Terwee, Mokkink, Knol, 2011).

В настоящее время в клинической психодиагностике существует несколько методологических подходов к разделению клинической шкалы на статистически значимые метрические уровни и определению критериев оценки результатов исследования. Наиболее точный и надежный подход, по мнению de Vet et al. (2011), основан на использовании стандартной ошибки измерения разности между оценками шкалы (SEM_{dif}). Если статистическая разница между суммарными оценками шкалы с заданной вероятностью превышает величину SEM_{dif} , то данные оценки относятся к разным метрическим уровням. В клинических и психометрических исследованиях для оценки статистической разницы задается вероятность, равная 95%. Тогда расстояние между оценками шкалы, относящимися к статистически разным метрическим уровням, должно превышать величину $1,96SEM_{dif}$, при этом число 1,96 представляет собой значение Z-оценки для 95%-ной плотности вероятно-

сти. Выделение в шкале метрических уровней проводится, начиная с минимальной (нулевой) суммарной оценки. Нулевая оценка представляет собой нижнюю границу первого метрического уровня, эквивалентного наименьшей степени выраженности измеряемого конструкта. Последовательно оценивается величина статистической разницы между нулевой и каждой последующей суммарной оценкой шкалы. Любая ближайшая суммарная оценка, имеющая разницу с нулевой оценкой, превышающую величину $1,96SEM_{dif}$, становится нижней границей следующего метрического уровня. Одновременно с этим предшествующая оценка образует верхнюю границу первого метрического уровня. Далее в качестве отправной точки используется суммарная оценка, эквивалентная нижней границе второго уровня, и цикл повторяется. Выявление метрических уровней в шкале проводится до максимальной суммарной оценки включительно. Оценки, находящиеся внутри одного метрического уровня, с 95%-ной вероятностью относятся к одной и той же степени выраженности измеряемого конструкта, которой эквивалентен данный уровень (Bech, Paykel, Sireling, Yiend, 2015; de Vet, Terwee, Mokkink, Knol, 2011).

После разделения шкалы на метрические уровни их следует подвергнуть клинической валидации путем сопоставления с клиническими внешними критериями. После такой клинической валидации метрических критериев в большинстве случаев клинически ориентированная психометрическая шкала включает четыре клинко-метрических уровня (Bech, 2012): клинически незначимой выраженности (первый уровень); уровень легкой степени выраженности (второй уровень); уровень умеренной степени выраженности (третий уровень) и уровень значительной степени выраженности (четвертый метрический уровень).

В психометрике существуют два варианта расчета величины SEM_{dif} . Первый вариант базируется на методологии классической теории тестов (КТТ) (Crocker, Algina, 2008). Второй основан на методологии модели Дж. Раша, которая в последние годы активно применяется для разработки и анализа клинических рейтинговых шкал (Bond, Fox, 2007).

Различия между этими двумя вариантами носят концептуальный характер. Для более глубокого понимания целесообразно кратко рассмотреть методологию расчета SEM_{dif} .

Поскольку психометрическая шкала направлена на измерение латентного конструкта, полученные в результате этого данные всегда включают в себя случайную ошибку. Как известно, стандартное отклонение случайной ошибки получило название стандартной ошибки измерения (SEM):

$$SEM = \sqrt{\sigma_{error}^2}, \quad (1)$$

где σ_{error}^2 – дисперсия случайной ошибки (Crocker, Algina, 2008).

Пусть необходимо определить стандартную ошибку разницы между любыми двумя шкальными оценками X_i и X_j , т.е. ($SEM_{dif}_{X_j-X_i}$). Поскольку SEM, согласно (1), представляет собой стандартное отклонение случайной ошибки шкальной оценки, то $SEM_{dif}_{X_j-X_i}$ является стандартным отклонением случайной ошибки разности между шкальными оценками X_i и X_j :

$$SEM_{dif}_{X_j-X_i} = \sqrt{\sigma_{error(X_j-X_i)}^2}, \quad (2)$$

где $\sigma_{error(X_j-X_i)}^2$ – дисперсия случайной ошибки разности оценок X_i и X_j .

Как известно, дисперсия разности двух переменных равна сумме дисперсий этих двух переменных. Следовательно, формула расчета дисперсии ошибки разности оценок X_i и X_j будет иметь вид:

$$\sigma_{error(X_j-X_i)}^2 = \sigma_{error(X_i)}^2 + \sigma_{error(X_j)}^2. \quad (3)$$

Отсюда, значение $SEM_{dif}_{X_j-X_i}$ можно определить по следующей формуле:

$$\begin{aligned} SEM_{X_j-X_i} &= \sqrt{\sigma_{error(X_j-X_i)}^2} = \\ &= \sqrt{\sigma_{error(X_i)}^2 + \sigma_{error(X_j)}^2} = \\ &= \sqrt{SEM_{X_i}^2 + SEM_{X_j}^2}. \end{aligned} \quad (4)$$

Таким образом, стандартная ошибка разности двух оценок равна квадратному корню из суммы квадратов стандартных ошибок измерения каждой из этих оценок. Следовательно, для расчета стандартной ошибки разности между оценками (SEM_{dif}) достаточно знать значения SEM – стандартных ошибок измерения этих оценок (de Vet, Terwee, Mokkink, Knol, 2011).

В классической теории тестов значение SEM рассчитывается по следующей формуле:

$$SEM = \sigma_X \sqrt{(1-r)}, \quad (5)$$

где σ_X – стандартное отклонение распределения оценок в популяции; r – коэффициент надежности (Bond, Fox, 2007).

Расчитанное таким образом значение SEM одинаково для всех возможных оценок шкалы. Так как все оценки шкалы имеют одинаковые стандартные ошибки измерения, формулу (4) для расчета стандартной ошибки разницы между оценками X_i и X_j можно преобразовать следующим образом:

$$\begin{aligned} SEM_{dif_{X_j-X_i}} &= \sqrt{SEM_{X_i}^2 + SEM_{X_j}^2} = \\ &= \sqrt{2SEM^2} = \sqrt{2}SEM. \end{aligned} \quad (6)$$

Тогда, по de Vet et al. (2011), величина статистической разницы между оценками шкалы равна значению выражения $1,96\sqrt{2}SEM$.

Второй вариант расчета значений SEM_{dif} основан на методологии предложенной датским математиком Георгом Рашем в 1960 г. модели, получившей его имя. В настоящее время эта модель развилась в самостоятельную систему вероятностных математических методов построения психометрических шкал¹. Концептуальное ядро модели Раша основано на анализе каждого ответа испытуемых на каждый диагностический пункт методики. Базовое уравнение модели (так называемая однопараметрическая логистическая функция Раша) описывает функциональную связь вероятности ключевого ответа испытуемого на пункт теста с уровнем выраженности у него измеряемого конструкта и трудностью пункта, на который дается ответ:

$$\frac{P(\beta)}{1 - P(\beta)} = e^{(\beta - \delta_i)}, \quad (7)$$

где β – уровень измеряемого конструкта у испытуемого; δ_i – трудность пункта i ; $P(\beta)$ – вероятность ключевого ответа при уровне конструкта β ; e – экспоненциальная функция, равная 2,718.

Используя в качестве отправной точки это базовое уравнение, было создана целая система математических итерационных методов построения линейных равноинтервальных измерительных шкал. Использование такой

¹ В зарубежной литературе по этому вопросу приводится термин Rasch Measurement (RM), а в русскоязычных публикациях в последние годы используется термин «метрическая система Георга Раша» (прим. ред.).

метрической системы Раша позволяет трансформировать суммарные оценки шкалы в эквивалентные меры выраженности конструкта, представленные в логитах². В результате такой трансформации образуется равноинтервальная метрическая шкала (Bond, Fox, 2007). В отличие от построенной по принципам КТТ системы, в модели Раша для каждой шкальной оценки, выраженной в логитах, рассчитывается индивидуальная ошибка измерения. Расчет такого рода ошибок проводится в соответствии с итерационным математическим алгоритмом модели (Bond, Fox, 2007; Wright, Stone, 1999). В связи с этим формула расчета SEM_{dif} в модели Раша имеет классический вид (см. формулу (5)). Отсюда, по Wright, Stone (1999), величина статистической разницы между оценками шкалы в формате модели Раша имеет вид:

$$1,96\sqrt{SEM_{X_i}^2 + SEM_{X_j}^2}. \quad (8)$$

Учитывая наличие двух версий расчета величины статистической разницы, в процессе разработки критериев интерпретации клинической рейтинговой шкалы встает проблема подбора наиболее адекватного варианта. Адекватная методика расчета SEM_{dif} должна соответствовать требованиям объективного, воспроизводимого и валидного определения метрических уровней в клинической шкале.

Цель настоящего исследования состояла в сравнительной оценке воспроизводимости и клинической применимости метрических критериев клинической рейтинговой шкалы, разработанных в рамках классической теории тестов и на основе модели Раша.

Материал и методы исследования

В качестве анализируемой психометрической шкалы была выбрана шкала оценки тяжести депрессии Гамильтона (HRSD, Hamilton, 1960). Эта шкала впервые была опубликована в Англии в 1960 г., затем переведена на все европейские языки и в настоящее время является одной из наиболее широ-

² Логит – единица измерения уровней выраженности конструкта и трудностей тестовых заданий в рамках логистических моделей. В частном случае – если разность между упомянутыми двумя понятиями составляет 1 логит, то вероятность ключевого ответа конкретным испытуемым j -го пункта теста равна 0,73. Математически логит представляет собой логарифм шанса ключевого ответа на пункт теста (прим. авт.).

ко используемых методик психометрической оценки тяжести депрессии (McDowell, 2006).

Шкала HRSD предназначена для оценки выраженности депрессивного состояния у пациентов с уже установленным диагнозом депрессии. Перечень симптомов в шкале достаточно ограничен и не исчерпывает все проявления депрессии. Структура HRSD включает 21 диагностический пункт, который оценивается по трехбалльной (от 0 до 2) или пятибалльной (от 0 до 5) категориям (de Vet, Terwee, Mokkink, Knol, 2011).

Для подсчета общего показателя по шкале учитываются только первые 17 пунктов. Оставшиеся 4 симптома используются для детализации характера депрессии. Все диагностические пункты оценивают состояние пациента в течение последних нескольких дней. В клинике часто используется сокращенная 17-пунктовая версия шкалы, лишённая последних четырех пунктов, поскольку они не учитываются при подсчете (McDowell, 2006).

Первичные данные формировались на основании результатов, включенных в протоколы исследования с помощью HRSD пациентов с легкой, умеренной и тяжелой степенями депрессии, а также здоровых испытуемых.

Пациенты обследовались при амбулаторном лечении (с легким депрессивным эпизодом) и стационарном лечении (с умеренным и тяжелым депрессивным эпизодом) в Гродненском областном клиническом центре «Психиатрия-Наркология» и в Гродненской городской клинической больнице № 3 (Республика Беларусь). Психодиагностическое обследование проводилось в течение первых 5 дней после вынесения клинического диагноза. Для целей сравнительного анализа были сформированы три выборки первичных данных. Выборку № 1 составили данные исследования группы здоровых лиц, пациентов с клинически верифицированным диагнозом депрессии различной степени тяжести численностью 551 человек, в том числе 231 мужчина и 320 женщин, все в возрасте 23–54 года. Выборка № 2 являлась рандомизированно сформированной частью выборки № 1 в количестве 331 наблюдений (60 % всего объема выборки № 1), в том числе 138 мужчин и 193 женщины в возрасте 24–52 года. Выборка № 3 состояла из данных исследования вновь сформированной группы здоровых лиц и пациентов с депрессией в количестве 333 человек в возрасте 24–53 года, в том числе 145 мужчин и 189 женщин.

Метрические уровни рассчитывались по данным исследования каждой выборки отдельно. Технология расчета метрических уровней в рамках КТТ включала расчет значений показателя дисперсии, коэффициента надежности, которые использовались для последующего расчета стандартной ошибки измерения по формуле (5). Величина статистической разницы между оценками шкалы рассчитывалась по формуле (7). Технология расчета метрических уровней в рамках модели Раша включала:

- вероятностный анализ ответов испытуемых на пункты HRSD;
- итерационное оценивание мер выраженности депрессии и параметров трудности пунктов;
- построение равноинтервальной шкалы логитов, эквивалентной шкале сырых баллов;
- расчет индивидуальной ошибки измерения для каждой меры депрессии, выраженной в логитах.

Величина статистической разницы между оценками шкалы рассчитывалась по формуле (8).

После определения метрических уровней в рамках обоих методологических подходов проводилась оценка согласованности метрических уровней с клиническими степенями тяжести депрессии, определенными по критериями МКБ-10. Клиническая верификация проводилась путем статистического анализа таблиц сопряженности клинических и метрических критериев. При этом строки таблицы включали частоты клинических критериев выраженности конструкта по МКБ-10, столбцы таблицы содержали частоты метрических критериев. В ячейках на пересечении строк и столбцов локализовались совпадающие частоты. Статистический анализ заключался в определении максимальной частоты каждого клинического критерия, которая совпадала с максимальной частотой определенного метрического критерия. Метрический критерий, совпавший по максимальной частоте с определенным внешним клиническим критерием, приобретал свойство клинической интерпретируемости, соответствующий данному внешнему критерию. Если внешнему клиническому критерию по максимальным частотам соответствовали два и более метрических критерия, то они объединялись в один общий критерий, имеющий два метрических уровня. В качестве статистических критериев опреде-

Психодиагностика

ления значимой максимальной частоты использовался точный метод Фишера (ТМФ).

Первоначально оценивались метрические уровни, разработанные в рамках КТТ, а затем – метрические уровни, рассчитанные на основе модели Раша.

Результаты

В табл. 1 представлены значения показателей дисперсии, коэффициента надежности, стандартной ошибки измерения (*SEM*) и статистической разницы между оценками вида $1,96\sqrt{2SEM}$, рассчитанные по материалам шкалы оценки депрессии Гамильтона во всех трех выборках с использованием технологии классической теории тестов.

Таблица 1
Значения дисперсии (σ), коэффициента надежности α Кронбаха, *SEM* и статистической разницы между оценками, рассчитанные по технологии классической теории тестов в трех выборках

Выборка	Значения показателей			
	σ	α Кронбаха	<i>SEM</i>	Статистическая разница
1	7,83	0,88	2,74	7,59
2	8,43	0,86	3,12	8,65
3	8,65	0,84	3,46	9,59

Как следует из данных в табл. 1, во всех трех выборках обнаружилось различные значения *SEM* и статистических различий между

оценками по шкале HRSD, что создает нестабильные условия для определения метрических критериев выраженности депрессии в данной шкале.

В табл. 2 представлены границы уровней выраженности депрессии, рассчитанные на основе значений статистической разницы между оценками в трех выборках.

Анализ данных в табл. 2 показывает, что во всех трех выборках метрические критерии выраженности депрессии имеют различные значения. При этом, чем выше уровень выраженности депрессии, тем больше расхождения в метрических критериях. Нестабильность критериев оценки выраженности депрессии обуславливает различия в интерпретации полученных данных при исследовании с помощью HRSD различных популяций испытуемых.

В табл. 3 приведены сопряженные частоты метрических и клинических уровней выраженности депрессии в трех выборках испытуемых. В таблицах также отмечены максимальные частоты в столбцах и строках, рассчитанные с помощью точного метода Фишера (ТМФ). Статистически значимые ($p < 0,05$) максимальные частоты отмечены жирным шрифтом и звездочкой. Анализ максимальных частот в табл. 3 позволяет заключить, что во всех трех выборках первый метрический уровень статистически соответствует двум кли-

Таблица 2
Метрические критерии выраженности депрессии в шкале HRSD, рассчитанные на основе классической теории тестов в трех выборках данных

Метрический уровень депрессии	Значения метрических критериев в выборках, отличающихся величиной статистической разницы между оценками		
	Выборка № 1	Выборка № 2	Выборка № 3
	1	0	0
2	8	9	10
	9	10	11
3	17	19	21
	18	20	22
4	26	29	32
	27	30	33
5	35	39	43
	36	40	44
6	44	49	52
	45	50	–
	52	52	–

* Значения статистической разницы округлялись до целого числа, поскольку метрические критерии в рамках КТТ рассчитываются в баллах, являющихся целыми числами. Истинные значения составили: для выборки № 1 – $(1,96\sqrt{2SEM} = 7,59 \approx 8)$, для выборки № 2 – $(1,96\sqrt{2SEM} = 8,65 \approx 9)$, для выборки № 3 – $(1,96\sqrt{2SEM} = 9,59 \approx 10)$.

Таблица 3

Таблица сопряженности частот¹ метрических уровней (рассчитанных на основе КТТ) и клинических степеней выраженности депрессии

Клинические степени тяжести депрессии ²	Метрические уровни (выборка № 1)					Метрические уровни (выборка № 2)				Метрические уровни (выборка № 3)			
	1	2	3	4	5	1	2	3	4	1	2	3	4
Первая	*97	0	0	0	0	*62	0	0	0	*94	0	0	0
Вторая	*99	68	0	0	0	*49	45	0	0	*57	44	0	0
Третья	0	*135	*41	0	0	0	*93	16	0	5	*79	18	0
Четвертая	0	0	*48	*61	2	0	1	*45	*20	0	0	*31	*2

¹ Здесь и далее: результаты оценки максимальных частот в выборках № 1–3 установлены с помощью ТМФ.

² Здесь и далее: использованным формулировкам иерархии уровней депрессии соответствуют: первая – «здоровые, (депрессия отсутствует)»; вторая – «депрессивный эпизод легкой степени», третья – «депрессивный эпизод умеренной степени», четвертая – «депрессивный эпизод тяжелой степени».

ническим степеням тяжести депрессии: отсутствию депрессии и легкой степени депрессии. Отсюда следует, что система метрических критериев, рассчитанных на основе классической теории тестов, неэффективно согласуется с клиническими данными и с трудом интерпретируется. Первый метрический уровень может интерпретироваться как отсутствие депрессии, либо как легкая степень депрессии.

В табл. 4 представлены границы метрических уровней шкалы HRDS, рассчитанные на основе модели Раша, с представлением рассчитанной в логитах эквивалентной для каждой оценки по шкале меры выраженности депрессии. При этом каждая такая мера имеет свою индивидуальную ошибку измерения

(SE). Определение метрических критериев проводилось по специальной формуле, построенной на основе модели Раша.

Как следует из анализа данных в табл. 4, метрические критерии, выраженные в логитах, в трех выборках данных также имеют различные значения. Особенно явно это видно при сравнении метрических критериев выборки № 1 и выборки № 2. Однако, если сравнить эквивалентные логитам оценки в баллах, то во всех трех выборках метрические критерии имеют одинаковые значения, за исключением границы между 7-м и 8-м уровнями выраженности депрессии. Отсюда следует, что метрические критерии, рассчитанные на основе модели Раша, обладают значительно большей

Таблица 4

Метрические критерии выраженности депрессии в шкале HRDS, рассчитанные на основе модели Раша в трех выборках данных

Уровень выраженности депрессии	Выборка № 1			Выборка № 2			Выборка № 3		
	Логиты	SE	Баллы	Логиты	SE	Баллы	Логиты	SE	Баллы
1	-4,3	1,54	0	-5,08	1,53	0	-5,17	1,54	0
	-1,38	0,39	5	-2,18	0,37	5	-2,24	0,40	5
2	-1,17	0,37	6	-1,98	0,36	6	-2,03	0,38	6
	-0,32	0,28	12	-1,12	0,28	12	-1,17	0,29	12
3	-0,21	0,27	13	-1,01	0,27	13	-1,06	0,28	13
	0,42	0,24	20	-0,38	0,24	20	-0,41	0,24	20
4	0,5	0,24	21	-0,30	0,23	21	-0,33	0,24	21
	1,09	0,23	29	0,29	0,23	29	0,28	0,24	29
5	1,16	0,23	30	0,36	0,23	30	0,35	0,24	30
	1,81	0,26	38	1,00	0,27	38	1,05	0,27	38
6	1,91	0,27	39	1,10	0,28	39	1,15	0,28	39
	2,68	0,34	45	1,87	0,35	45	2,00	0,36	45
7	2,86	0,37	46	2,05	0,38	46	2,20	0,39	46
	4,09	0,61	50	2,88	0,51	49	3,53	0,53	50
8	4,82	0,86	51	3,49	0,62	50	4,29	0,63	51
	6,06	1,55	52	5,27	1,55	52	5,54	1,55	52

Психодиагностика

стабильностью и воспроизводимостью по сравнению с критериями, рассчитанными на основе КТТ. Расхождения в баллах на 7-м и 8-м уровнях в двух выборках можно объяснить тем, что метрические критерии на этих уровнях соответствуют таким высоким оценкам, которые практически не встречаются в популяциях пациентов. Вместе с тем, модель Раша строит полноразмерную шкалу, включающую в себя, в том числе, и оценки, которые могут не встречаться в популяции, но являются при этом психометрически обоснованными. Такие маргинальные оценки теоретически моделируются на шкале с помощью математического итерационного алгоритма модели. В связи с этим теоретически ожидаемые метрические критерии могут не всегда соответствовать наблюдаемым данным. Однако такие расхождения в модели Раша редки и касаются таких уровней выраженности конструкта, которые не определяются в реальности, но допускаются конструкцией шкалы. На всех остальных уровнях выраженности депрессии в трех выборках данных проявляется полное соответствие метрических критериев, выраженных в баллах.

В табл. 5 приведены установленные для трех выборок сопряженные частоты метрических уровней, рассчитанных на основе модели

Раша, и клинических степеней выраженности депрессии. В таблицах также отмечены максимальные значения частот в столбцах и строках, рассчитанные с помощью точного метода Фишера (ТМФ). Статистически значимые ($p < 0,05$) максимальные частоты отмечены полужирным шрифтом и знаком «звездочка» (*).

Анализ данных в табл. 5 показывает, что метрические уровни, рассчитанные на основе модели Раша, эффективно согласуются с клиническими данными во всех трех выборках. В табл. 6 приведены эквивалентные соответствия метрических уровней и клинических степеней тяжести депрессии, определенных по критериям МКБ-10.

Анализ соответствий, представленных в табл. 6, выявляет, что тяжелой степени депрессии в шкале Гамильтона соответствует пять метрических уровней. Вместе с тем, анализ частот в табл. 5 свидетельствует о том, что в выборках испытуемых фактически наблюдается не более 5 метрических уровней депрессии. Отсюда следует, что три последних уровня соответствуют таким высоким оценкам, которые в настоящее время не встречаются даже у пациентов с тяжелой депрессией. Однако, учитывая наличие установ-

Таблица 5
Таблица сопряженности частот метрических уровней (рассчитанных на основе модели Раша) и клинических степеней выраженности депрессии

Клинические степени тяжести депрессии	Метрические уровни (выборка № 1)					Метрические уровни (выборка № 2)					Метрические уровни (выборка № 3)				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Первая	*56	41	0	0	0	*36	26	0	0	0	*67	27	0	0	0
Вторая	19	*117	31	0	0	11	*63	20	0	0	17	*55	29	0	0
Третья	0	27	*129	20	0	0	19	*78	12	0	0	9	*75	18	0
Четвертая	0	0	11	*67	*34	0	0	4	*42	*20	0	0	0	*26	*7

Таблица 6
Эквивалентные статистически значимые соответствия метрических уровней шкалы HRDS, установленных на основе модели Раша, и клинических степеней тяжести депрессии (по МКБ-10)

Метрический уровень (модель Раша)	Диапазон оценок (баллы)	Клиническая степень тяжести депрессии (МКБ-10)
1	0–5	Отсутствуют признаки депрессии
2	6–12	Депрессия легкой степени тяжести
3	13–20	Депрессия средней степени тяжести
4	21–29	Депрессия тяжелой степени тяжести
5	30–38	
6	39–45	
7	46–50	
8	51–52	

ленных эквивалентных связей трех последних уровней с оценками шкалы, такие уровни логично отнести к тяжелой степени депрессии.

Обсуждение результатов и выводы

В результате проведенного исследования было выявлено, что критерии оценки тяжести депрессии в шкале Гамильтона, разработанные на основе классической теории тестов, носят нестабильный характер и не согласуются с клиническими данными. Методология определения метрических критериев в формате КТТ имеет существенные недостатки, которые в значительной степени снижают валидность и надежность установленных критериев интерпретации. Представляется целесообразным рассмотреть эти методологические недостатки подробнее.

Первый недостаток связан с общей методологией конструирования психометрических шкал в классической теории тестов. Шкалы, созданные на ее основе, оказываются зависимыми от особенностей популяции. Это означает, что значения параметров «дисперсия оценок» и «коэффициент надежности» будут меняться при исследовании различных выборок. Оба параметра, согласно формуле (5), участвуют в расчете стандартной ошибки измерения (*SEM*). Исходя из этого, значение статистической разницы между оценками, рассчитанное в процессе метрической калибровки шкалы на основе данных исследования одной выборки, может быть непригодным для использования при исследовании другой выборки испытуемых.

Второй методологический недостаток заключается в том, что все оценки шкалы имеют одинаковое значение стандартной ошибки измерения (*SEM*). В классической теории тестов значение *SEM* рассчитывается на основе стандартного отклонения оценок от средней и значения коэффициента надежности. Оба параметра имеют отношение ко всей шкале и нивелируют индивидуальные особенности каждой оценки на шкале. Психометрическая шкала не может обладать одинаковой измерительной точностью во всем диапазоне выраженности конструкта вследствие неравномерного распределения диагностических пунктов в соответствии со степенью их трудности. Точное установление критериев интерпретации требует правильного определения величин стандартных ошибок измерения оценок,

которые образуют границы метрических уровней выраженности конструкта. Если значения ошибок измерения определены неверно, то критерии интерпретации не будут соответствовать клинической реальности.

В отличие от критериев КТТ метрические критерии, разработанные на основе модели Раша, отличаются высокой стабильностью и эффективно согласуются с клиническими данными. Это обусловлено некоторыми особенностями метрической системы Раша, в частности, независимостью от параметров распределения конструкта в популяции и возможностью определения индивидуальной ошибки измерения для каждой оценки шкалы.

Таким образом, применение модели Раша в психометрической разработке клинических психометрических шкал позволяет не только построить равноинтервальную аддитивную измерительную шкалу, но и выделить в шкале статистически значимые метрические уровни выраженности конструкта, эффективно согласующиеся с клиническими данными.

Литература/References

1. Bech P. *Clinical psychometrics*. Chichester, West Sussex: Wiley-Blackwell, 2012. 202 p. DOI: 10.1002/9781118511800.
2. Bech P., Paykel E., Sireling L., Yiend J. Rating scales in general practice depression: Psychometric analyses of the Clinical Interview for Depression and the Hamilton Rating Scale. *Journal of Affective Disorders*. 2015, vol. 171, pp. 68–73. DOI: <http://dx.doi.org/10.1016/j.jad.2014.09.013>
3. Bond T.G., Fox C.M. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. LEA, 2007. 356 p.
4. Crocker L., Algina J. *Introduction to Classical and Modern Test Theory*. Cengage Learning, 2008. 541 p.
5. Hamilton M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatr.* 1960, pp. 56–62.
6. McDowell I. *Measuring Health: A Guide to Rating Scales and Questionnaires*. New York: Oxford University Press, 2006. 748 p. DOI: 10.1093/acprof:oso/9780195165678.001.0001.
7. De Vet H.C.W., Terwee C.B., Mokkink L.B., Knol D.K. *Measurement in Medicine*. Cambridge University Press, 2011. 350 p. DOI: <https://doi.org/10.1017/CBO9780511996214>.
8. Wright B.D., Stone M.H. *Measurement essentials*. Wilmington, Delaware: Wide Range, Inc., 1999. 205 p.

Ассанович Марат Алиевич, кандидат медицинских наук, доцент, заведующий кафедрой медицинской психологии и психотерапии, Гродненский государственный медицинский университет (г. Гродно, Республика Беларусь), 70malas@gmail.com

Поступила в редакцию 22 февраля 2017 г.

DOI: 10.14529/psy170103

COMPARATIVE EVALUATION OF CUT-OFF CRITERIA OF DEPRESSION SEVERITY DEVELOPED ON THE BASIS OF CLASSICAL TEST THEORY AND RASCH MODEL

*M.A. Assanovich, 70malas@gmail.com
Grodno State Medical University, Grodno, Republic of Belarus*

Each clinical rating scale used in psychiatry should have evidence-based criteria for assessing the severity of psychopathological constructs measured. The purpose of this study was to conduct the comparative evaluation of the reproducibility and clinical applicability of metric criteria of The Hamilton Rating Scale for Depression (HRSD) developed using the classical test theory (CTT) and developed on the basis of Rasch model. Three samples of research protocols obtained using HRSD included healthy subjects and patients with depression were established for comparative study. As a result of the study, it was found that interpretation criteria developed using the classical test theory had low reproducibility and were not applicable for clinical data. Unlike CTT-criteria metric criteria developed on the base of Rasch model characterized by a high degree of reproducibility and were effectively consistent with clinical data. These results were explained by the conceptual methodological differences of both psychometric approaches in construction of objective measurement and determination of the standard error of measurement.

Keywords: interpretation criteria, Hamilton Rating Scale for Depression, Rasch model, classic test theory.

Received 22 February 2017

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Ассанович, М.А. Сравнительная оценка пороговых критериев выраженности депрессии, разработанных на основе классической теории тестов и модели Раша / М.А. Ассанович // Вестник ЮУрГУ. Серия «Психология». – 2017. – Т. 10, № 1. – С. 26–34. DOI: 10.14529/psy170103

FOR CITATION

Assanovich M.A. Comparative Evaluation of Cut-Off Criteria of Depression Severity Developed on the Basis of Classical Test Theory and Rasch Model. *Bulletin of the South Ural State University. Ser. Psychology*. 2017, vol. 10, no. 1, pp. 26–34. (in Russ.). DOI: 10.14529/psy170103
