

## СРАВНИТЕЛЬНАЯ ОЦЕНКА ПОРОГОВЫХ КРИТЕРИЕВ ШКАЛЫ ДЕПРЕССИИ ГАМИЛЬТОНА, РАЗРАБОТАННЫХ НА ОСНОВЕ ROC-АНАЛИЗА И МОДЕЛИ РАША

**М.А. Ассанович**

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

Интерпретируемость психометрической шкалы основывается на наличии определенных пороговых критериев, которые позволяют разделить все оценки шкалы по определенным уровням выраженности измеряемого конструкта. Цель описываемого в публикации исследования состояла в сравнительной оценке воспроизводимости и диагностической эффективности пороговых критериев шкалы оценки тяжести депрессии Гамильтона (HRSD), разработанных в рамках двух подходов: ROC-анализа и 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 пороговые ROC-критерии отличаются низкой воспроизводимостью и характеризуются высокой точностью только в случае их использования на той же выборке данных, на которой они разрабатывались. В отличие от ROC-критериев, разработанные на основе модели Раша критерии отличаются высокой степенью воспроизводимости и эффективно согласуются с клиническими данными независимо от особенностей выборки. Полученные результаты обоснованы концептуальными и методологическими различиями обоих подходов в построении пороговых критериев.

*Ключевые слова: пороговые критерии, шкала оценки депрессии Гамильтона, модель Раша, ROC-анализ.*

Интерпретируемость психометрической шкалы основывается на наличии определенных пороговых критериев, которые позволяют разделить все оценки шкалы по уровням выраженности измеряемого конструкта. Пороговые критерии должны объективно отражать уровни выраженности конструкта независимо от особенностей популяций испытуемых и согласовываться с клиническими данными (de Vet et al., 2011).

В настоящее время в клинической психометрике применяется несколько способов определения пороговых критериев. В одном из наших исследований (Ассанович, 2017) была проведена сравнительная оценка воспроизводимости и согласованности с клиническими данными пороговых критериев, разработанных на основе классической теории тестов (КТТ) и модели Раша. Анализ полученных в данном исследовании результатов показал, что пороговые критерии, рассчитанные на основе КТТ, характеризуются низкой воспроиз-

водимостью при исследовании различных выборок испытуемых и не согласуются с клиническими данными. В то же время пороговые критерии, разработанные на основе модели Раша, не зависят от особенностей популяции и соответствуют клиническим данным (Ассанович, 2017).

В настоящее время в клинической психодиагностике для определения пороговых критериев достаточно часто используется ROC-анализ (Hajian-Tilaki, 2013). Основная цель использования данного метода состоит в оценке внешней валидности и диагностической эффективности психодиагностических шкал. Кроме этого, ROC-анализ позволяет также определить отсекающие оценки, обладающие максимальной чувствительностью и специфичностью, которые устанавливаются как пороговые критерии (de Vet et al., 2011). Выборки данных для проведения ROC-анализа формируются с помощью внешнего клинического критерия. В качестве такового

могут использоваться оценки экспертов или данные по другим методикам, имеющим статус «золотого стандарта». Например, с помощью экспертных оценок клиницистов можно сформировать три группы пациентов с легкой, умеренной и значительной выраженностью проявлений измеряемого конструкта. Далее, проведя ROC-анализ для каждой пары смежных групп, определяются оценки, дифференцирующие три степени выраженности конструкта. Эти оценки и будут являться пороговыми критериями интерпретации. Для данных конкретных выборок ROC-анализ предлагает наилучшую дифференцирующую оценку. Однако остается неясной степень объективности (воспроизводимости) пороговых критериев, установленных таким методом. Неизвестно, как степень воспроизводимости и надежности таких критериев связана с их диагностической эффективностью.

Цель настоящего исследования состояла в оценке воспроизводимости и диагностической эффективности пороговых критериев шкалы оценки депрессии Гамильтона, разработанных на основе модели Раша, в сравнении с пороговыми критериями, определенными с помощью ROC-анализа.

### Методология исследования

Материал настоящей статьи представляет собой продолжение исследования, результаты которого нами представлены в работе (Ассанович, 2017). В качестве психометрической шкалы использовалась шкала оценки депрессии Гамильтона (HDRS, Hamilton, 1960). Первичные данные включали протоколы исследования с помощью HRDS пациентов с легкой, умеренной и тяжелой степенями депрессии, а также здоровых испытуемых. Протоколы были сгруппированы в три выборки данных. Подробная характеристика испытуе-

мых и выборок представлена в исследовании (Ассанович, 2017).

Определение пороговых критериев интерпретации с помощью ROC-анализа (ROC-критериев) проводилось в следующей последовательности отдельно для каждой выборки данных.

1. Проведение ROC-анализа оценок испытуемых в парах групп, смежных по выраженности депрессии: здоровые – легкая степень депрессии, легкая степень депрессии – умеренная степень депрессии, умеренная степень депрессии – тяжелая степень депрессии.

2. Построение таблиц показателей чувствительности и специфичности для каждой оценки.

3. Определение в каждой паре групп пороговых критериев, обладающих максимальной чувствительностью и специфичностью. Пороговые критерии представляли собой оценки шкалы, соответствовавшие максимальному значению критерия Юдена (J Yuden), определяемого по формуле:  $J = \text{чувствительность} + \text{специфичность} - 1$  (de Vet et al., 2011).

В качестве пороговых критериев, установленных на основе метрической системы Раша (Раш-критериев), были использованы критерии, рассчитанные на первом этапе исследования и показавшие полную воспроизводимость во всех трех выборках (Ассанович, 2017). Эти критерии представлены в табл. 1.

Сравнительная оценка диагностической эффективности пороговых критериев осуществлялась путем оценки степени согласованности классификации испытуемых по степени тяжести депрессии, выполненной на основе различных пороговых критериев HRSD, с одной стороны, с классификацией, построенной на основе клинических критериев МКБ-10, с другой. Оценка степени соответствия класси-

Таблица 1

Пороговые критерии HRDS, установленные на основе модели Раша (Раш-критерии)

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	

фикаций проводилась с помощью статистических критериев кси-квадрат, показателей величины эффекта Коэна (Cohen, 1992). Сравнительная оценка диагностической эффективности пороговых критериев проводилась в двух выборках данных: первой (выборка 1) и третьей (выборка 3). Оцениванию подвергались пороговые критерии, разработанные с помощью ROC-анализа и метрической системы Раша, а также оригинальные критерии HRSD. Сравнительный анализ диагностической эффективности проводился с помощью отношения шансов (ОШ), статистического критерия Z и показателя абсолютного снижения риска (Fritz, Morris, Richler, 2012). Оригинальные пороговые критерии HRSD включали два набора. Первый состоял из классических, традиционно устоявшихся пороговых оценок (классические критерии), второй набор включал пороговые оценки (APA-критерии, рекомендуемые Американской психиатрической ассоциацией, Zimmerman et al., 2013). Пороговые критерии обоих наборов представлены в табл. 2.

## Результаты исследования

### Оценка воспроизводимости пороговых критериев, установленных с помощью ROC-анализа

В табл. 3 представлены значения чувствительности (Se) и специфичности (Sp) для каждой оценки шкалы HRSD и дифференцирующие критерии в трех выборках испытуемых по результатам ROC-анализа групп «здоровые – легкая степень депрессии».

Анализ данных в табл. 3 обнаруживает, что ROC-критерии разделения групп здоровых и пациентов с легкой степенью депрессии в трех выборках данных имеют различные значения. В выборке 1 оценка, соответствующая максимальному значению индекса J Юдена, равна 5. NFrz оценка соответствует нижней границе уровня легкой степени тяжести депрессии. Это означает, что диапазон оценок, отражающий отсутствие депрессии, составил от 0 до 4 баллов.

В выборке 2 критерием разделения групп является оценка, равная 10 баллам, соответст-

Таблица 2

Пороговые критерии оригинальной шкалы оценки тяжести депрессии Гамильтона

Степень тяжести депрессии	Классические критерии, баллы	APA-критерии, баллы
Депрессия отсутствует	0–6	0–7
Легкая депрессия	7–17	8–13
Умеренная депрессия	18–24	14–18
Тяжелая депрессия	25 и более баллов	19 и более баллов

Таблица 3

Критерии разделения групп «здоровые – легкая степень депрессии» в трех выборках испытуемых

Выборка 1 (N=264)			Выборка 2 (N=156)			Выборка 3 (N=196)		
Оценка	Se	Sp	Оценка	Se	Sp	Оценка	Se	Sp
0	1,000	0,000	1	1,000	0,000	0	1,000	0,000
1	1,000	0,073	2	1,000	0,115	1	1,000	0,075
2	1,000	0,167	3	1,000	0,246	2	1,000	0,151
3	1,000	0,292	4	1,000	0,361	3	1,000	0,280
4	1,000	0,385	5	0,926	0,508	4	1,000	0,312
<b>5*</b>	<b>0,940*</b>	<b>0,521</b>	6	0,883	0,574	5	0,911	0,387
6	0,886	0,573	7	0,777	0,656	6	0,832	0,710
7	0,766	0,667	8	0,521	0,820	7	0,683	0,828
8	0,443	0,833	<b>10*</b>	<b>0,479*</b>	<b>1,000*</b>	8	0,644	0,935
10	0,407	1,000	11	0,319	1,000	<b>9*</b>	<b>0,604*</b>	<b>0,978*</b>
11	0,263	1,000	12	0,266	1,000	10	0,564	1,000
12	0,222	1,000	13	0,213	1,000	11	0,436	1,000
13	0,186	1,000	14	0,170	1,000	12	0,416	1,000
14	0,162	1,000	15	0,149	1,000	13	0,287	1,000
15	0,138	1,000	16	0,117	1,000	14	0,198	1,000
16	0,114	1,000	17	0,043	1,000	15	0,069	1,000
17	0,048	1,000	–	–	–	–	–	–

\* Обозначена оценка, соответствующая максимальному значению критерию Юдена ( $1 - Se + Sp$ ).

## Психодиагностика

вующая нижней границе уровня легкой степени депрессии. Таким образом, диапазон оценок, свидетельствующих об отсутствии депрессии, представлен интервалом от 1 до 8 баллов.

Отметим, что ни один испытуемый выборки 2 не имеет оценки, равной 9. Поэтому с помощью ROC-анализа оказалось невозможным оценить дифференцирующие свойства такой оценки. Фактически она попадает в зону диагностической неопределенности. Следовательно, верхней границей диапазона оценок, соответствующего отсутствию депрессии, является оценка, равная 8, а нижней границей уровня умеренной степени тяжести служит оценка, равная 10. В выборке 3 дифференцирующий критерий равен 9 баллам, а границы диапазона, соответствующего отсутствию депрессии, равны 0 и 8 баллам.

В табл. 4 представлены значения чувствительности (Se) и специфичности (Sp) для каждой оценки шкалы HRDS в трех выборках испытуемых по результатам ROC-анализа групп «легкая степень – умеренная степень тяжести депрессии».

Анализ данных в табл. 4 обнаруживает различные критерии дифференцирования групп пациентов с легкой и умеренной депрессией в трех выборках. В выборке 1 оцен-

ка, соответствующая максимальному значению индекса J Юдена, равна 11. Таким образом, диапазон оценок, соответствующий легкой степени депрессии, составил от 5 (табл. 4) до 11 баллов. В выборке 2 критерием разделения групп является значение оценки, равное 13. Отсюда диапазон оценок, свидетельствующих о наличии легкой степени депрессии, составляет от 10 (табл. 4) до 13 баллов. В выборке 3 дифференцирующий критерий равен 14 баллам, а границы диапазона, соответствующего легкой степени депрессии, равны 9 и 14 баллам.

В табл. 5 представлены значения чувствительности (Se) и специфичности (Sp) для каждой оценки шкалы HRDS в трех выборках испытуемых по результатам ROC-анализа групп «умеренная степень – тяжелая степень депрессии».

Как и в предыдущих случаях, данные в табл. 5 позволяют обнаружить различные критерии дифференцирования групп пациентов с умеренной и тяжелой депрессией во всех трех выборках. В выборке 1 оценка, соответствующая максимальному значению индекса J Юдена, равна 19 (т. е. диапазон оценок, соответствующий средней степени тяжести депрессии, составил от 11 до 19 баллов). В выборке 2 критерием разделения групп яв-

Таблица 4  
Критерии разделения групп «легкая степень – умеренная степень депрессии» в трех выборках испытуемых

Выборка 1 (N=343)			Выборка 2 (N=203)			Выборка 3 (N=203)		
Оценка	Se	Sp	Оценка	Se	Sp	Оценка	Se	Sp
4	1,000	0,000	4	1,000	0,000	4	1,000	0,000
5	1,000	0,054	5	1,000	0,065	5	1,000	0,080
6	1,000	0,108	6	1,000	0,108	6	1,000	0,160
7	1,000	0,229	7	1,000	0,215	7	1,000	0,310
8	1,000	0,554	8	1,000	0,473	8	1,000	0,350
10	1,000	0,590	10	1,000	0,516	9	1,000	0,390
<b>11*</b>	<b>0,926*</b>	<b>0,735*</b>	11	0,917	0,677	10	1,000	0,430
12	0,881	0,777	12	0,862	0,731	11	0,951	0,560
13	0,847	0,813	<b>13*</b>	<b>0,826*</b>	<b>0,785*</b>	12	0,941	0,580
14	0,761	0,837	14	0,734	0,828	13	0,912	0,710
15	0,625	0,861	15	0,569	0,849	<b>14*</b>	<b>0,824*</b>	<b>0,800*</b>
16	0,432	0,886	16	0,385	0,882	15	0,598	0,930
17	0,295	0,952	17	0,266	0,957	16	0,343	1,000
18	0,233	1,000	18	0,220	1,000	17	0,255	1,000
19	0,193	1,000	19	0,174	1,000	18	0,225	1,000
20	0,142	1,000	20	0,147	1,000	19	0,157	1,000
21	0,114	1,000	21	0,110	1,000	20	0,108	1,000
22	0,063	1,000	22	0,055	1,000	21	0,069	1,000
23	0,040	1,000	23	0,028	1,000	22	0,029	1,000
24	0,011	1,000	–	–	–	23	0,010	1,000
–	–	–	–	–	–	–	–	–

ляется оценка в 20 баллов, а диапазон оценок, отражающих наличие средней степени тяжести депрессии, составил от 13 до 10 баллов. В выборке 3 дифференцирующий критерий равен 22 баллам. Обращает внимание отсутствие в данной выборке испытуемых, набравших по шкале HRSD суммы в 20 баллов и 21 балл. Такие оценки остаются в зоне диагностической неопределенности. Отсюда границы диапазона, соответствующего средней степени депрессии, равны 14 и 19 баллам, а

верхняя граница уровня тяжелой степени депрессии составляет 22 балла.

На основе проведенного анализа во всех трех выборках были определены критерии интерпретации выраженности депрессии по шкале HRSD, которые представлены в табл. 6.

Таким образом, во всех трех выборках пороговые ROC-критерии имеют различные значения. Из этого следует, что пороговые критерии, установленные с помощью ROC-

Таблица 5  
Критерии разделения групп «умеренная степень – тяжелая степень депрессии» в трех выборках испытуемых

Выборка 1 (N=288)			Выборка 2 (N=175)			Выборка 3 (N=136)		
Оценка	Se	Sp	Оценка	Se	Sp	Оценка	Se	Sp
10	1,000	0,000	10	1,000	0,000	10	1,000	0,000
11	1,000	0,069	11	1,000	0,074	11	1,000	0,040
12	1,000	0,114	12	1,000	0,130	12	1,000	0,050
13	1,000	0,149	13	1,000	0,167	13	1,000	0,079
14	1,000	0,234	14	1,000	0,259	14	1,000	0,168
15	1,000	0,371	15	1,000	0,426	15	1,000	0,396
16	1,000	0,566	16	1,000	0,611	16	1,000	0,653
17	1,000	0,703	17	1,000	0,731	17	1,000	0,743
18	1,000	0,766	18	1,000	0,778	18	1,000	0,772
<b>19*</b>	<b>1,000*</b>	<b>0,806*</b>	19	1,000	0,815	19	1,000	0,802
20	0,938	0,857	<b>20*</b>	<b>0,985*</b>	<b>0,852*</b>	<b>22*</b>	<b>1,000*</b>	<b>0,822*</b>
21	0,902	0,886	21	0,939	0,889	23	0,794	0,901
22	0,821	0,937	22	0,788	0,944	24	0,647	1,000
23	0,795	0,960	23	0,758	0,972	25	0,529	1,000
24	0,759	0,989	24	0,712	1,000	26	0,471	1,000
25	0,732	1,000	25	0,697	1,000	27	0,412	1,000
26	0,705	1,000	26	0,515	1,000	28	0,353	1,000
27	0,563	1,000	27	0,424	1,000	29	0,294	1,000
28	0,429	1,000	28	0,348	1,000	30	0,235	1,000
29	0,366	1,000	29	0,303	1,000	31	0,176	1,000
30	0,304	1,000	30	0,197	1,000	32	0,147	1,000
31	0,170	1,000	31	0,136	1,000	33	0,088	1,000
32	0,125	1,000	32	0,106	1,000	34	0,059	1,000
33	0,098	1,000	33	0,076	1,000	35	0,029	1,000
34	0,063	1,000	34	0,045	1,000	-	-	-
35	0,036	1,000	35	0,015	1,000	-	-	-
36	0,018	1,000	36	1,000	0,000	-	-	-
37	0,009	1,000	37	1,000	0,074	-	-	-

\* Обозначена оценка, соответствующая максимальному значению критерию Юдена (1- Se + Sp).

Таблица 6  
Пороговые критерии HRSD, установленные с помощью ROC-анализа в трех выборках испытуемых

Интерпретация	Критерии		
	Выборка 1	Выборка 2	Выборка 3
Депрессия отсутствует	0–4	1–8	0–8
Зона неопределенности	–	9	–
Легкая степень тяжести депрессии	5–10	10–12	9–13
Умеренная (средняя) степень тяжести депрессии	11–18	13–19	14–19
Зона неопределенности	–	–	20–21
Тяжелая степень выраженности депрессии	19 и более	20 и более	22 и более

анализа, в отличие от критериев, созданных на основе модели Раша, характеризуются низкой воспроизводимостью. Стабильность ROC-критериев зависит от особенностей распределения оценок в выборке испытуемых.

### *Сравнительная оценка диагностической эффективности пороговых критериев HRSD*

В табл. 7 приведены значения показателей согласованности с клиническими данными четырех видов пороговых критериев HRSD: классических, АРА-критериев, ROC-критериев и Раш-критериев. При этом ROC-критерии были разделены на два типа: прямые и перекрестные. Прямые ROC-критерии оценивались на тех же выборках данных, на которых они и разрабатывались. Перекрестные критерии были образованы путем замены ROC-критериев выборки 1 критериями выборки 3 и, наоборот, критерии, разработанные на выборке 1 оценивались на выборке 3.

Анализ данных в табл. 7 показывает, что почти все пороговые критерии обнаруживают в той или иной степени рассогласование с клиническими критериями. Только классификация на основе Раш-критериев в выборке 1

достоверно не отличается от клинической классификации ( $X^2=7,80$ ;  $p=0,05$ ). Классические критерии и перекрестные ROC-критерии обнаруживают самые низкие уровни согласования с клиническими данными в обеих выборках. Степень рассогласования этих критериев с клиническими критериями достигает 40 %, что отражается в низких значениях показателя величины эффекта  $w$  (0,07–0,14). Лучше всего с клиническими критериями согласуются прямые ROC-критерии и критерии, разработанные на основе модели Раша. Промежуточное положение занимают АРА-критерии, которые в выборке 1 показали низкий уровень согласования ( $g=0,12$ ), а выборке 3 – умеренный ( $g=0,20$ ).

В табл. 8 представлены результаты сравнения различных пороговых критериев друг с другом. Как показывают результаты сравнения, максимальные эффекты улучшения согласованности с клиническими данными демонстрируют прямые ROC-критерии и Раш-критерии. Величина отношения шансов для данных критериев в сравнении с классическими критериями превышает 2. Абсолютный прирост эффекта в отношении снижения риска рассогласования с клиническими критериями небольшой – от 15 до 17 %.

Таблица 7  
Показатели согласованности с клиническими критериями различных пороговых критериев HRSD

Пороговые критерии HRSD	Выборки данных N1 = 551 N3 = 333	Оценка рассогласования с клиническими критериями $X^2$ , $p$	Величина эффекта рассогласования $X^2$ $w$ Cohen's	Уровень рассогласования с клиническими критериями, %	Величина эффекта совпадения классификаций $g$ Cohen's
Классические	1	$X^2=169,29^{***}$	0,55 <sup>1</sup>	42,9	0,07 <sup>A</sup>
	3	$X^2=88,70^{***}$	0,52 <sup>1</sup>	43,3	0,07 <sup>A</sup>
АРА	1	$X^2=108,11^{***}$	0,44 <sup>2</sup>	38,4	0,12 <sup>A</sup>
	3	$X^2=32,04^{***}$	0,31 <sup>2</sup>	29,7	0,20 <sup>B</sup>
ROC (прямые)	1	$X^2=23,88^{***}$	0,21 <sup>3</sup>	25,5	0,25 <sup>C</sup>
	3	$X^2=43,53^{***}$	0,36 <sup>2</sup>	29,7	0,20 <sup>B</sup>
ROC (перекрестные)	1	$X^2=147,29^{***}$	0,52 <sup>1</sup>	36,1	0,14 <sup>A</sup>
	3	$X^2=40,97^{***}$	0,35 <sup>2</sup>	40,9	0,09 <sup>A</sup>
Критерии Раша	1	$X^2=7,80^*$	0,12 <sup>1</sup>	27,0	0,23 <sup>B</sup>
	3	$X^2=11,90$	0,19 <sup>1</sup>	30,3	0,20 <sup>B</sup>

Примечания. 1. \* – различия достоверны на уровне  $p<0,05$ ; \*\* – различия достоверны на уровне  $p<0,01$ ; \*\*\* – различия достоверны на уровне  $p<0,0001$ .

2. Уровень согласованности (по значениям эффекта рассогласования  $X^2$   $w$  Cohen's): <sup>1</sup> – низкий (<0,3); <sup>2</sup> – средний (0,3–0,5); <sup>3</sup> – высокий – (>0,5).

3. Уровень совпадения классификаций (по значениям коэффициента  $g$  Cohen's): <sup>A</sup> – низкий (<0,15); <sup>B</sup> – средний (0,15–0,25); <sup>C</sup> – высокий – (>0,25).

Таблица 8

Результаты сравнения пороговых критериев HRSD друг с другом по степени согласованности  
с клиническими данными

Сравниваемые критерии	Выборки данных	ОШ	Z, p	Абсолютное снижение риска рассогласования с клиническими критериями, %
ROC (прямые) – классические	1	2,18	5,98 p<0,0001	17,4
	3	1,80	3,63 p<0,001	13,6
ROC (прямые) – APA	1	1,82	4,56 p<0,0001	12,9
	3	1,00	0 p=1,00	0
ROC (перекрестные) – классические	1	1,33	2,28 p<0,05	6,8
	3	1,10	0,63 p=0,53	2,4
ROC (перекрестные) – APA	1	1,11	0,81 p=0,42	2,3
	3	1,63	2,99 p<0,01	0
ROC (прямые) – ROC (перекрестные)	1	1,64	3,77 p<0,001	10,6
	3	1,63	2,99 p<0,001	11,2
Раш – традиционные	1	2,02	5,46 p<0,0001	15,9
	3	1,75	3,44 p<0,001	13,0
Раш – APA	1	1,68	4,03 p<0,0001	11,4
	3	1,03	0,17 p=0,87	0
ROC (прямые) – Раш	1	1,08	0,55 p=0,58	1,5
	3	1,03	0,17 p=0,87	0,6
Раш – ROC (перекрестные)	1	1,53	3,23 p<0,01	9,1
	3	1,59	2,83 p<0,01	10,6

Сравнительный анализ не выявил достоверных различий между прямыми ROC-критериями и Раш-критериями, показывающих одинаковую эффективность. Однако сравнение Раш-критериев с перекрестными ROC-критериями показало, что пороговые Раш-критерии достоверно лучше согласуются с клиническими данными, чем ROC-критерии в случае применения последних к выборке данных, которая не использовалась для разработки этих критериев. Следует отметить, что перекрестные ROC-критерии обнаружили низкую эффективность при сравнении не только с Раш-

критериями, но и оригинальными критериями HRSD, а также прямыми ROC-критериями.

#### Обсуждение результатов и заключение

Результаты проведенного исследования позволяют сделать несколько обоснованных выводов.

ROC-анализ представляет собой метод доказательной оценки диагностической эффективности и критериальной валидности. Как правило, ROC-кривая практически не зависит от особенностей выборки. В плане оценки критериальной валидности ROC-анализ зарекомендовал себя как объективный

метод, дающий воспроизводимые кривые диагностической эффективности (Fawcett, 2006). Однако результаты настоящего исследования показывают, что данный метод нельзя использовать для получения пороговых оценок, объективно разделяющих группы испытуемых по клиническому критерию. Пороговые критерии, определяемые с помощью ROC-анализа, зависят от распределения оценок в выборке испытуемых и в полной мере относятся только к той выборке, данные исследования которой использовались для проведения анализа. На других выборках данных непосредственно не относящиеся к ним ROC-критерии могут характеризоваться низкой точностью и невысокой диагностической эффективностью. Пороговые ROC-критерии наилучшим образом воспроизводят классификацию испытуемых по внешнему критерию только на той выборке данных, на которой проводился ROC-анализ. Помимо низкой воспроизводимости, разработка пороговых критериев с помощью ROC-анализа не учитывает ошибку измерения и полностью зависима от заранее проведенного разделения испытуемых на основе внешнего критерия. При этом остается неизвестной величина ошибки, с которой применялся внешний критерий. ROC-критерии лишь позволяют найти наилучшие точки разделения испытуемых на основе заранее определенного внешнего (клинического) критерия на данной конкретной выборке испытуемых. Переносить эти разделяющие оценки на другие выборки нельзя вследствие их низкой воспроизводимости.

Пороговые критерии, разработанные на основе модели Раша, лишены недостатков ROC-критериев и при этом не уступают последним в точности. Раш-критерии рассчитываются на максимально правдоподобной равноинтервальной шкале со стабильными параметрами. Одним из таких параметров, имеющим непосредственное отношение к определению пороговых критериев, является индивидуальная ошибка измерения для каждой шкальной оценки. Следует отметить, что параметры шкалы Раша не зависят от особенностей популяции, что и обуславливает высокую воспроизводимость пороговых критериев. В нашем исследовании (Ассанович, 2017) пороговые критерии, рассчитанные на основе модели Раша, показали полную воспроизводимость на трех выборках, в отличие от критериев, созданных на основе классической

теории тестов и ROC-критериев, описанных в настоящей работе. Раш-критерии обладают высокой точностью, не уступающей ROC-критериям в случае оценки эффективности последних на той же выборке данных, на которой они разрабатывались.

В отличие от ROC-критериев, технология определения критериев на основе модели Раша фактически не зависит от распределения оценок в выборке испытуемых, поскольку учитывает ошибки измерения оценок моделированной шкалы, а не наблюдаемых ответов испытуемых.

Классические пороговые критерии шкалы оценки депрессии Гамильтона (HRSD) отличаются низкой точностью. Критерии, рекомендованные APA, обладают более высокой точностью, которая, вместе с тем, характеризуется выраженной нестабильностью в разных популяциях пациентов. Установленные на основе модели Раша критерии, как и прямые ROC-критерии, имеют значительно более высокую точность, по сравнению с оригинальными критериями HRSD. Однако установленные нами критерии все-таки обнаруживают достаточно высокий уровень расхождения с клиническими данными, достигающий 30 %. Такой уровень расхождения может быть обусловлен психометрическими проблемами HRSD, обнаруженными как в зарубежных исследованиях (Bagby et al., 2004), так и в одном из наших исследований (Ассанович, 2015). Вышеизложенное диктует необходимость проведения существенной психометрической модификации HRSD и разработки новых, более точных пороговых критериев.

### Литература

1. Ассанович, М.А. Оптимизация шкалы оценки депрессии Гамильтона на основе модели Раша / М.А. Ассанович // *Медицинская психология в России*. – 2015. – 2(31). – С. 7. – <http://mprj.ru> [Accessed March 30, 2017].
2. 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.
3. Bagby, R.M. *The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight?* / R.M. Bagby, A.G. Ryder, D.R. Schuller, M.B. Marshall // *American Jour-*

*nal of Psychiatry*. – 2004. – 161(12). – P. 2163–2177. – [http:// psychiatryonline.org/doi/abs/10.1176/appi.ajp.161.12.2163](http://psychiatryonline.org/doi/abs/10.1176/appi.ajp.161.12.2163).

4. Cohen, J. *Quantitative Methods in Psychology* / J. Cohen // *Psychological Bulletin*. – 1992. – 112(1). – P. 155–159.

5. Fawcett, T. *An introduction to ROC analysis* / T. Fawcett // *Pattern Recognition Letters*. – 2006. – 27(8). – P. 861–874.

6. Fritz, C.O. *Effect size estimates: Current use, calculations, and interpretation* / C.O. Fritz, P.E. Morris, J.J. Richler // *Journal of Experimental Psychology: General*. – 2012. – 141(1). – P. 2–18. – <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0024338>.

7. Hajian-Tilaki, K. *Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation* / K. Hajian-Tilaki //

*Caspian Journal of Internal Medicine*. – 2013. – 4(2). – P. 627–635.

8. Hamilton, M. *A rating scale for depression* / M. Hamilton // *Journal of neurology*. – 1960. – 23. – P. 56–62. – Available at: /Users/Kayleigh/Desktop/Papers/ReadCube/Hamilton-1960- *Journal of neurology.pdf*.

9. De Vet, H.C.W. *Measurement in Medicine. A Practical Guide* / H.C.W. de Vet, C.B. Terwee, L.B. Mokkink, D.L. Knol. – Cambridge University Press, 2011. – Available at: [www.cambridge.org](http://www.cambridge.org) [Accessed March 30, 2017].

10. Zimmerman, M. *Severity classification on the Hamilton Depression Rating Scale* / M. Zimmerman, J.H. Martinez, D. Young et al. // *Journal of affective disorders*. – 2013. – 150(2). – P. 384–388. Available at: <http://www.science-direct.com/science/article/pii/S0165032713003017>.

**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 3 апреля 2017 г.

DOI: 10.14529/psy170202

## COMPARATIVE EVALUATION OF CUT-OFF CRITERIA OF THE HAMILTON RATING SCALE FOR DEPRESSION DEVELOPED USING ROC-ANALYSIS AND RASCH MODEL

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

Interpretability of the psychometric scale is based on the presence of certain threshold criteria that allow to divide all scores into levels of the construct measured. The purpose of this study was to conduct the comparative evaluation of the reproducibility and diagnostic efficiency of cut-off criteria of The Hamilton Rating Scale for Depression (HRSD) developed using the ROC-analysis and 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 cut-off ROC-criteria had low reproducibility and were characterized by high accuracy only if they were used in the same sample of data which was used for their developing. Unlike ROC-criteria, cut-off criteria developed on the base of Rasch model characterized by a high degree of reproducibility and were effectively consistent with clinical data regardless of the sample characteristics. These results were explained by the conceptual and methodological differences of both approaches in defining of cut-off criteria.

*Keywords:* cut-off criteria, Hamilton Rating Scale for Depression, Rasch model, ROC-analysis.

### References

1. Assanovich M.A. *Optimizatsiya shkaly otsenki depressii Gamil'tona na osnove modeli Rasha. Meditsinskaya psikhologiya v Rossii* [Optimization of the Scale of Assessing Hamilton's Depression on the Basis of the Rush Model. Medical Psychology in Russia], 2015, no. 2(31), p. 7. Available at: <http://mprj.ru> (accessed 30.03.2017).
2. 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
3. Bagby R.M. et al. The Hamilton Depression Rating Scale: Has the Gold Standard Become a Lead Weight? *American Journal of Psychiatry*, 2004, vol. 161(12), pp. 2163–2177. Available at: <http://psychiatryonline.org/doi/abs/10.1176/appi.ajp.161.12.2163>. DOI: <https://doi.org/10.1176/appi.ajp.161.12.2163>.
4. Cohen J. Quantitative Methods in Psychology. *Psychological Bulletin*, 1992, vol. 112(1), pp. 155–159. DOI: <https://doi.org/10.1037/0033-2909.112.1.155>.
5. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*, 2006, vol. 27(8), pp. 861–874. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.
6. Fritz C.O., Morris, P.E. & Richler, J.J. Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 2012, vol. 141(1), pp. 2–18. Available at: <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0024338>. DOI: <https://doi.org/10.1037/a0024338>.
7. Hajian-Tilaki K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 2013, vol. 4(2), pp. 627–635.
8. Hamilton M. A rating scale for depression. *Journal of neurology*, 1960, pp. 56–62. Available at: </Users/Kayleigh/Desktop/Papers/ReadCube/Hamilton-1960-Journal of neurology.pdf>.
9. De Vet H.C.W. et al. *Measurement in Medicine A Practical Guide*, Cambridge University Press, 2011. Available at: [www.cambridge.org](http://www.cambridge.org) [Accessed March 30, 2017]. DOI: <https://doi.org/10.1017/CBO9780511996214>.
10. Zimmerman M. et al. Severity classification on the Hamilton Depression Rating Scale. *Journal of affective disorders*, 2013, vol. 150(2), pp. 384–8. Available at: <http://www.sciencedirect.com/science/article/pii/S0165032713003017>. DOI: <http://dx.doi.org/10.1016/j.jad.2013.04.028>.

*Received 3 April 2017*

---

### ОБРАЗЕЦ ЦИТИРОВАНИЯ

Ассанович, М.А. Сравнительная оценка пороговых критериев шкалы депрессии Гамильтона, разработанных на основе ROC-анализа и модели Раша / М.А. Ассанович // Вестник ЮУрГУ. Серия «Психология». – 2017. – Т. 10, № 2. – С. 15–24. DOI: 10.14529/psy170202

### FOR CITATION

Assanovich M.A. Comparative Evaluation of Cut-off Criteria of the Hamilton Rating Scale for Depression Developed Using ROC-Analysis and Rasch Model. *Bulletin of the South Ural State University. Ser. Psychology*. 2017, vol. 10, no. 2, pp. 15–24. (in Russ.). DOI: 10.14529/psy170202