

Оптимизация шкалы оценки депрессии Гамильтона на основе модели Раша

Ассанович М.А.

Ассанович Марат Алиевич

кандидат медицинских наук, доцент, заведующий кафедрой медицинской психологии и психотерапии, УО «Гродненский государственный медицинский университет», ул. Горького, 80, Гродно, 230013, Беларусь. Тел.: +375-152-75-66-73.

E-mail: 70malas@gmail.com

Аннотация. Цель исследования состояла в оптимизации шкалы оценки депрессии Гамильтона (HAM-D) с точки зрения конструктивной валидности диагностических пунктов. Методы исследования: исследование проводилось на основе политомического варианта модели Раша. Выборка испытуемых включала 550 пациентов с депрессией различной степени тяжести. Оценка конструктивной валидности пунктов проводилась после построения модели шкалы с помощью метрической системы Раша на основе значений индексов соответствия моделированных ответов наблюдаемым ответам испытуемых. Результаты: часть пунктов оригинальной шкалы HAM-D имеет неудовлетворительную конструктивную валидность. К ним относятся три пункта, содержащие симптомы нарушений сна, два пункта, включающие нарушения моторной активности, пункт с ипохондрическими симптомами и пункт «Критика к состоянию». В ходе анализа построены две оптимальные диагностические модели, включающие 12 и 10 диагностических пунктов. Обсуждение: обнаруженные психометрические проблемы HAM-D подтверждаются данными литературы. Симптом ипохондрии не является типичным симптомом депрессии, вследствие чего соответствующий пункт шкалы показал низкую конструктивную валидность. Структура пункта «Критика к состоянию» не обеспечивает адекватного поведения этого пункта в паттернах ответов испытуемых. Пункты, отражающие инсомнические нарушения, образуют выраженную интерференцию друг с другом, что обуславливает снижение их валидности. Пункты «Напряжение» и «Заторможенность» находятся в противоположных позициях и работают друг против друга, что определяет чрезмерно высокие значения индексов остатков. Вместе с тем шкала оценки депрессии Гамильтона содержит пункты, имеющие адекватную конструктивную валидность и приемлемые значения индексов качества. Удаление невалидных «шумящих» пунктов из модели позволяет оптимизировать структуру шкалы таким образом, что в ней будут находиться только валидные диагностические пункты. Нами получены две оптимальные диагностические модели. Одна включает 12 пунктов, вторая — 10 пунктов. Дальнейшие исследования должны быть направлены на оценку того, какая из этих моделей лучше согласуется с клиническими данными.

Ключевые слова: политомическая модель Раша; шкала оценки депрессии Гамильтона; измерительные модели; сепарационная статистика; индексы качества.

УДК 616-072.87(07)

Библиографическая ссылка по ГОСТ Р 7.0.5-2008

Ассанович М.А. Оптимизация шкалы оценки депрессии Гамильтона на основе модели Раша // Медицинская психология в России: электрон. науч. журн. – 2015. – N 2(31). – С. 7 [Электронный ресурс]. – URL: <http://mprj.ru> (дата обращения: чч.мм.гггг).

Поступила в редакцию: 20.03.2015 Прошла рецензирование: 03.04.2015 Опубликовано: 17.04.2015

Введение

Макс Гамильтон считал, что психометрика в клинической психиатрии должна занимать место научной дисциплины наряду с фармакологией и биохимией. Свои психометрические разработки Гамильтон проводил в клинике Университета Лидса в период 1957—1960 гг. Данный период был временем становления современной психофармакологии, начавшимся открытием антиманиакального эффекта лития и

антипсихотического эффекта хлорпромазина. Гамильтон четко видел необходимость в научно обоснованных коротких рейтинговых шкалах, которые могли бы использоваться в психофармакологических исследованиях для оценки эффекта новых препаратов. В 1959 г. Гамильтон разработал шкалу оценки тревоги, а в 1960 — шкалу оценки депрессии. В отличие от своего учителя Айзенка, который фокусировался на оценке нейротизма, Гамильтон, подобно Крепелину, концентрировался на оценке сугубо психиатрических симптомов тревоги и депрессии. Он полагал, что объективная оценка психопатологических проявлений является наилучшим способом получения клинического впечатления о пациенте. Цель применения шкал Гамильтона не состояла в постановке клинического диагноза, а заключалась только в оценке тяжести симптомов в течение недели. С целью исследования структуры и валидизации рейтинговых шкал Гамильтон применял спирменовский двухфакторный анализ. В процессе факторизаций шкалы депрессии он выявлял различное количество факторов в зависимости от гомогенности группы пациентов, оценки которых факторизировались. Фактически двухфакторный анализ оказался бесполезен в изучении структуры шкалы депрессии. В последующие 30 лет жизни, вплоть до своей смерти в 1989 году, Гамильтон полностью сосредоточился на исследовании созданной им шкалы депрессии, но так окончательно и не решил проблему ее структуры [7; 24].

Несмотря на незавершенный характер, шкала оценки депрессии Гамильтона (HAM-D) стала самой популярной объективной шкалой оценки выраженности депрессии в клинических исследованиях. Обзоры последних лет показывают, что методика обладает в целом удовлетворительными психометрическими характеристиками. Однако часть пунктов шкалы отличается конструктивной несостоятельностью, что сказывается на точности оценки тяжести депрессии. Отмечено, что пороговые критерии, разделяющие легкую степень депрессии и норму, не соответствуют критериальной валидности [23; 30]. Часть пунктов обнаруживает низкую конструктивную валидность. Из-за этого в последние годы HAM-D часто подвергается критическим нападкам [8; 23; 24; 30].

Цель настоящего исследования состояла в оптимизации шкалы оценки депрессии Гамильтона с точки зрения адекватной конструктивной валидности ее пунктов.

Методология исследования

Научное измерение в социальных науках основано на построении аддитивной равноинтервальной шкалы с единицами измерения, эквивалентными фиксированному количеству измеряемого психологического конструкта [5; 9; 17; 19; 20]. Сырые тестовые баллы не могут считаться единицами измерения, поскольку для них не установлены правила эквивалентности с уровнями выраженности конструкта. Последовательность сырых тестовых оценок также не может считаться измерительной шкалой, поскольку в ней отсутствует аддитивная структура, расстояния между баллами не являются равноинтервальными, а сами баллы не представляют собой единицы измерения [1; 4; 6; 10; 25].

Все виды нормализации, стандартизации и сглаживания, имеющие целью приведение сырых оценок к нормальному распределению, являются в большей мере искусственными и не имеют никакого отношения к измерению [1; 3; 4; 6; 9].

На сегодняшний день единственным подходом, который позволяет создать адекватную научным требованиям измерительную шкалу, является модель Раша [1; 7].

Модель Раша предложена датским математиком Г. Рашем в 1960 г. [29]. С тех пор она развилась в мощную систему вероятностных математических методов построения измерительных психодиагностических шкал [13; 22]. Концептуальное ядро данной модели основано на анализе каждого ответа испытуемых на вопрос каждого диагностического пункта методики [10; 11]. Базовое уравнение модели Раша описывает функциональную связь вероятности ключевого ответа испытуемого на вопрос пункта теста с уровнем выраженности у него измеряемого конструкта и трудностью вопроса пункта, на который дается ответ. С использованием этого базового уравнения в

качестве отправной точки была создана целая система математических итерационных методов построения линейных равноинтервальных измерительных шкал. Принципиальными достоинствами модели Раша по сравнению с другими подходами к созданию психометрических шкал являются возможность построения инвариантной диагностической шкальной модели и объективная оценка конструктивной валидности диагностических пунктов [10; 11; 13; 14; 19].

В исследовании применялся политомический вариант модели Раша — модель частичного доверия, разработанная Мастерсом в 1982 году [18]. Модель разработана для построения и анализа шкал с пунктами, имеющими две категории ответов или более. Модель относится к семейству моделей Раша и содержит все полезные ее свойства: достаточность общей оценки по тесту для измерения уровня свойства, раздельное оценивание испытуемых и пунктов. По сути модель частичного доверия является простой адаптацией модели Раша для политомических пунктов. Категории ответа в каждом пункте представляются в обычной последовательности. Основное уравнение модели описывает функцию вероятности ответа на одну из категорий вопросов при определенном уровне измеряемого конструкта [18; 22].

HAM-D не содержит конкретных диагностических вопросов, привязанных к пунктам шкалы. Клиницисту самому предлагается формулировать вопросы, руководствуясь содержанием пунктов и собственным клиническим опытом. Естественно, это вносит существенный элемент субъективизма в диагностическое исследование, снижает валидность и надежность данных. С целью устранения указанного недостатка в 1988 г. Ж. Вильямс (Нью-Йоркский институт психиатрии) разработала «Структурированное интервью для шкалы депрессии Гамильтона» (SIGHD). Интервью состоит из 16 основных диагностических вопросов, соответствующих пунктам шкалы Гамильтона. 17-й пункт идентичен оригинальной HAM-D. К каждому основному вопросу прилагаются дополнительные с целью уточнения депрессивного симптома. Критерии оценки интервью соответствуют критериям оценки шкалы Гамильтона. Методика официально переведена на ряд европейских языков, в том числе и на русский. Обработка протокола исследования и оценка полученных данных соответствуют таковым в оригинальной шкале оценки депрессии Гамильтона [2]. Данное интервью использовалось нами в настоящем исследовании.

Выборка испытуемых для получения матрицы ответов составила 550 человек в возрасте от 23 до 54 лет. Среди них 328 женщин и 222 — мужчин. Все испытуемые являлись амбулаторными и стационарными пациентами психиатрического профиля с диагнозом аффективного расстройства. Исследование пациентов проводилось в течение первых трех-пяти дней после постановки клинического диагноза. Распределение испытуемых по критериальным группам представлено в таблице 1.

Таблица 1

Распределение испытуемых по нозологическим группам

Клинический диагноз	кол-во чел.
Легкий депрессивный эпизод	174
Умеренный депрессивный эпизод	173
Тяжелый депрессивный эпизод без психотических симптомов	125
Тяжелый депрессивный эпизод с психотическими симптомами	78

После вычисления линейных мер депрессии и трудностей пунктов шкалы Гамильтона проводилась оценка конструктивной валидности каждого пункта методики. Оценка конструктивной валидности пунктов осуществлялась по значениям среднеквадратичных остатков от разности между моделированными ответами на вопросы пунктов и фактически полученными при обследовании испытуемых [15; 16].

21]. В соответствии с правилами моделирования по Рашу оценивались два среднеквадратичных индекса валидности: невзвешенный (UMS) и взвешенный (WMS) [29]. Оба индекса по сути представляют собой критерий хи-квадрат, деленный на количество степеней свободы. Ожидаемое значение индексов, которое говорит о хорошем соответствии, равно 1. Значения меньше 1 свидетельствуют о чрезмерной предсказуемости ответов на вопрос, сформулированный в пункте, в контексте измеряемого конструкта. Значения больше 1 говорят о большом уровне «шума» и низком соответствии параметров пункта наблюдаемым данным. Эмпирически приемлемый диапазон оценок индексов валидности для диагностических интервью составляет 0,6–1,4 [27].

После построения объективной измерительной шкалы рассчитывалась психометрическая сепарационная статистика, включающая показатели надежности (сепарационный индекс) и дифференцирующей способности шкалы (индекс количества слоев) [12; 26; 28]. Пункты, не соответствующие установленным критериям индексов качества, удалялись, и диагностическая модель пересчитывалась вновь. Улучшение модели фиксировалось по оценкам индексов качества и количества слоев.

Результаты и их обсуждение

В результате применения алгоритма модели Раша было достигнуто схождение итерационного процесса на уровне критерия конвергенции 0,005.

Для каждого пункта шкалы Гамильтона рассчитаны индексы соответствия UMS и WMS.

Рассмотрим диагностическую состоятельность и конструктивную валидность пунктов шкалы Гамильтона после проведенного Раш-анализа. Значения индексов качества пунктов оригинальной шкалы приведены в таблице 2.

Таблица 2

Значения индексов качества пунктов оригинальной шкалы Гамильтона

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,58	0,59
2	Вина	0,64	0,63
3	Суицидальность	0,79	0,79
4	Трудности засыпания	1,26	1,25
5	Прерывистый сон	1,17	1,17
6	Раннее пробуждение	1,17	1,13
7	Работа, деятельность	0,54	0,56
8	Заторможенность	1,71	1,56
9	Напряжение	1,27	1,53
10	Психическая тревога	0,91	0,88
11	Соматическая тревога	0,85	0,77
12	Снижение аппетита	0,67	0,67
13	Общесоматические симптомы	1,03	1,03
14	Снижение либидо	0,48	0,44
15	Ипохондрия	1,23	1,16
16	Снижение веса	0,84	0,72
17	Критика к состоянию	1,77	1,80

Как следует из данной таблицы, 5 пунктов шкалы обнаруживают выходящие за допустимые рамки значения индексов качества. Это такие пункты, как «Депрессивное настроение», «Работа, деятельность», «Заторможенность», «Напряжение» и «Критика к состоянию». Данные пункты нарушают соответствие паттернов ответов испытуемых уровням выраженности депрессии. В результате снижается валидность и точность оценки тяжести депрессии по суммарному баллу шкалы. Фактически вышеперечисленные пункты не обладают надлежащей конструктивной валидностью в оригинальном формате HAM-D. Индекс количества слоев для данной модели составил 4,45, сепарационный индекс — 3,09. Это означает, что в исходном состоянии шкала способна дифференцировать 4 статистически значимых уровня выраженности депрессии. Анализ показателей в таблице 2 демонстрирует, что пункты 1, 7 и 14 имеют более низкие значения индексов. Это говорит о предсказуемости ответов испытуемых на вопросы, сформулированные в данных пунктах. При этом значения индексов для пунктов 1 («Депрессивное настроение») и 7 («Работа, деятельность») незначительно отклоняются от граничных значений, в то время как значение индекса для пункта 14 («Снижение либидо») снижено в значительной степени. Поведение этих пунктов в шкале легко объяснимо, поскольку все три пункта отражают типичные и ожидаемые симптомы депрессии. Большую предсказуемость пункта 14 («Снижение либидо») можно объяснить меньшим количеством категорий ответа в данном пункте по сравнению с другими пунктами.

Пункты 8, 9 и 17 имеют значительную долю случайной дисперсии, что отражается в повышенных значениях индексов качества. Что касается пунктов 8 («Заторможенность») и 9 («Напряжение»), то их поведение объяснимо, поскольку эти два пункта находятся в контрпозиции по отношению друг к другу. Это означает, что если испытуемые одному из этих пунктов дают высокие оценки, то второму, как правило, — низкие. Депрессивный пациент не может одновременно находиться в напряжении и заторможенности. Часть пациентов имеет симптомы напряжения, другая часть — симптомы заторможенности. Пункт 17 («Критика к состоянию») отличается наиболее высокими значениями индексов соответствия, что говорит о низкой конструктивной валидности этого пункта в рамках всей шкалы. Таким образом, результаты первичного анализа шкалы Гамильтона с помощью модели Раша в русскоязычной популяции подтверждают данные зарубежных исследований о диагностической несостоятельности отдельных пунктов шкалы.

Следующий этап настоящего исследования был направлен на оптимизацию шкалы HAM-D с точки зрения конструктивной валидности пунктов. Из диагностической модели прежде всего были удалены пункты, имевшие повышенные значения индексов UMS и WMS, в частности пункты 8, 9, 17. Далее повторно был выполнен итерационный анализ с помощью модели частичного доверия. Оценки остатков различий между ожидаемыми и наблюдаемыми ответами на вопросы пунктов, т.е. значения индексов качества, представлены в таблице 3.

Значения индексов качества пунктов диагностической модели HAM-D № 2

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,68	0,68
2	Вина	0,77	0,75
3	Суицидальность	0,88	0,86
4	Трудности засыпания	1,48	1,56
5	Прерывистый сон	1,37	1,49
6	Раннее пробуждение	1,39	1,38
7	Работа, деятельность	0,61	0,62
10	Психическая тревога	0,96	0,98
11	Соматическая тревога	0,89	0,84
12	Снижение аппетита	0,73	0,72
13	Общесоматические симптомы	1,17	1,18
14	Снижение либидо	0,52	0,47
15	Ипохондрия	1,44	1,44
16	Снижение веса	0,94	0,84

Анализ данных в таблице 3 показывает, что значения индексов для пунктов 1 и 7 «подтянулись» к приемлемому диапазону от 0,6 до 1,4. Пункт 14 также повысил значения индексов качества, однако они еще остаются ниже допустимой границы. Обращают на себя внимание новые пункты, значения которых выбились на уровень выше верхней границы допустимого диапазона. Это пункты, касающиеся симптомов нарушений сна: п. 4 («Трудности засыпания») и п. 5 («Прерывистый сон»). Также выбился из диапазона пункт 15 («Ипохондрические симптомы»). Невалидное поведение п. 4 и п. 5 можно объяснить тем, что нарушения сна представляют собой довольно частое явление при депрессии любой степени тяжести. Форма представления этих симптомов в шкале Гамильтона нечетко связывает тяжесть нарушений сна с тяжестью депрессии. Довольно выраженные инсомнические симптомы могут беспокоить пациентов с легкой и умеренной степенью выраженности депрессии. Что касается девиантного поведения п. 15 («Ипохондрия»), то это вполне ожидаемо. Ипохондрические проявления актуальны для тяжелых степеней депрессии, когда у пациента отмечаются психотические симптомы ипохондрического бреда. При легкой и средневыраженной депрессии имеющиеся ипохондрические симптомы отражают, скорее, не депрессию, а наличие коморбидного ипохондрического расстройства. Показатели сепарационной статистики модели Раша улучшили свои значения. В частности, сепарационный индекс составил 3,27, индекс количества слоев — 4,69.

Следующий этап анализа состоял в исключении невалидных пунктов из шкалы и пересчете диагностической модели. Показатели качества пунктов после исключения девиантных симптомов представлены в таблице 4.

Таблица 4

Значения индексов качества пунктов диагностической модели HAM-D № 3

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,76	0,77
2	Вина	0,89	0,86
3	Суицидальность	1,04	1
6	Раннее пробуждение	1,58	1,62
7	Работа, деятельность	0,65	0,67
10	Психическая тревога	1,06	1,09
11	Соматическая тревога	1,02	0,98
12	Снижение аппетита	0,81	0,79
13	Общесоматические симптомы	1,3	1,34
14	Снижение либидо	0,58	0,5
16	Снижение веса	1,09	1,03

Анализ значений индексов WMS и UMS в таблице 4 показывает, что только два пункта имеют отклоняющиеся значения данных индексов: п. 6 («Раннее пробуждение») и п. 14 («Снижение либидо»). При этом п. 6 характеризуется аномально высокими значениями WMS и UMS, что говорит о его слабом участии в измерении тяжести депрессии. Напротив, пункт 14 по-прежнему отличается предсказуемым поведением, которое отражается в сниженных значениях индексов. Показатели сепарационной статистики для этой модели характеризуются улучшенными значениями по сравнению с предыдущей. Сепарационный индекс составил 3,41, индекс количества слоев — 4,88. Это означает, что число статистически дифференцируемых степеней тяжести депрессии приближается к 5. С учетом большой величины остатка в пункте 6 было принято решение о его исключении из шкалы и последующем пересчете показателей модели. Значения индексов WMS и UMS следующей диагностической модели представлены в таблице 5.

Таблица 5

Значения индексов качества пунктов диагностической модели HAM-D № 4

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,78	0,79
2	Вина	0,91	0,88
3	Суицидальность	1,09	1,04
7	Работа, деятельность	0,68	0,69
10	Психическая тревога	1,15	1,20
11	Соматическая тревога	1,11	1,08
12	Снижение аппетита	0,86	0,84
13	Общесоматические симптомы	1,38	1,43
14	Снижение либидо	0,62	0,53
16	Снижение веса	1,19	1,15

Диагностическая модель № 4 характеризуется практически адекватными характеристиками. Значения индексов двух пунктов (п. 13 и п. 14) отличаются незначительными отклонениями от допустимого диапазона: от 0,6 до 1,4. Сепарационный индекс равен 3,51, индекс количества слоев составил 5,02, т.е. шкала HAM-D в форме диагностической модели № 4 может дифференцировать 5 статистически значимых уровней депрессии. Стоит ли удалять из модели пункты 13 и 14? Их отклонения в значениях UMS и WMS носят не столь значительный характер, как у других пунктов. Тем не менее они выходят за рамки установленного диапазона. На следующем этапе анализа мы удалили эти пункты и пересчитали параметры модели. В результате пересчета все оставшиеся 8 пунктов характеризовались приемлемыми значениями индексов. Однако удаление указанных пунктов ухудшило показатели сепарационной статистики. Сепарационный индекс уменьшил значение до 3,19, а индекс количества слоев — до 4,59. Это означает, что удаление пунктов с незначительными отклонениями значений индексов остатков при существенном снижении числа диагностических пунктов в шкале снижает надежность шкалы и ухудшает ее дифференцирующие свойства. Поэтому удаление данных пунктов, в целом имеющих адекватную конструктивную валидность, нежелательно.

Возникает другой вопрос: можно ли улучшить психометрические свойства рассмотренной выше диагностической модели № 4 путем добавления новых пунктов? Дело в том, что нами были удалены две группы симптомов, которые можно было бы объединить в два новых диагностических пункта. Первая группа включает симптомы напряжения (п. 8) и заторможенности (п. 9). Вторая группа включает нарушения сна (п. 4—6). В отдельности каждый из этих пунктов показал высокие значения разницы между ожидаемыми и наблюдаемыми ответами. Во многом это было обусловлено их дискоординированным поведением в шкале. Однако можно попытаться устранить дискоординированность пунктов путем их объединения в каждой группе в один общий симптом по принципу «или-или». Группу пунктов 8 и 9 мы объединяем в один пункт «Психомоторные нарушения». Данный пункт оценивается по максимальной шкале оценки ответа на любой вопрос из пунктов 8—9. Точно так же мы объединяем симптомы нарушений сна (п. 4—6) в один диагностический пункт «Нарушения сна», оцениваемый по максимальной оценке любого из пунктов 4—6. Сначала в модель № 4 был добавлен пункт «Психомоторные нарушения». Показатели полученной диагностической модели № 5 представлены в таблице 6.

Таблица 6

Значения индексов качества пунктов диагностической модели HAM-D № 5

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,75	0,77
2	Вина	0,87	0,84
3	Суицидальность	1,11	1,10
7	Работа, деятельность	0,69	0,70
8	Психомоторные нарушения	0,85	0,86
9	Психическая тревога	1,22	1,26
10	Соматическая тревога	1,21	1,18
12	Снижение аппетита	0,87	0,85
13	Общесоматические симптомы	1,38	1,45
14	Снижение либидо	0,62	0,52
16	Снижение веса	1,20	1,11

Анализ данных в таблице 6 показывает, что характеристики этой диагностической модели незначительно отличаются от характеристик предыдущей модели № 4. Индексы UMS и WMS для пункта «Психомоторные нарушения» (п. 8) находятся недалеко от

единицы, что говорит об адекватном поведении этого пункта в шкале. Показатели сепарационной статистики улучшили свои значения. Сепарационный индекс составил 3,76, индекс количества слоев — 5,36. На следующем этапе в модель был добавлен новый объединенный пункт «Нарушения сна». Значения индексов остатков для новой модели № 6 показаны в таблице 7.

Таблица 7

Значения индексов качества пунктов диагностической модели HAM-D № 6

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,75	0,77
2	Вина	0,87	0,85
3	Суицидальность	1,12	1,13
4	Нарушения сна	0,92	0,91
5	Работа, деятельность	0,69	0,71
6	Психомоторные нарушения	0,87	0,88
7	Психическая тревога	1,22	1,29
8	Соматическая тревога	1,21	1,21
9	Снижение аппетита	0,88	0,86
10	Общесоматические симптомы	1,39	1,47
11	Снижение либидо	0,61	0,52
12	Снижение веса	1,21	1,15

Как следует из таблицы 7, пункт «Нарушения сна» (п. 4) довольно удачно занял свое место в шкале, поскольку значения индексов качества близки к единице, т.е. к идеальному значению. Несколько ухудшились значения индексов UMS и WMS для пунктов «Общесоматические симптомы» и «Снижение либидо» (п. 10, 11). Показатели сепарационной статистики между тем повысили свои значения. Сепарационный индекс составил 3,92, индекс количества слоев — 5,56. Далее была предпринята попытка очередного удаления слегка «шумящих» пунктов 10 и 11 из модели с последующим пересчетом параметров. Значения индексов остатков получившейся модели № 7 представлены в таблице 8.

Таблица 8

Значения индексов качества пунктов диагностической модели HAM-D № 7

№ п.	Название пункта	Значения индексов качества	
		WMS	UMS
1	Депрессивное настроение	0,77	0,78
2	Вина	0,90	0,89
3	Суицидальность	1,13	1,20
4	Нарушения сна	0,94	0,94
5	Работа, деятельность	0,75	0,77
6	Психомоторные нарушения	0,90	0,91
7	Психическая тревога	1,19	1,30
8	Соматическая тревога	1,15	1,17
9	Снижение аппетита	0,88	0,86
12	Снижение веса	1,18	1,09

Как следует из анализа данных в таблице 8, все 10 пунктов имеют допустимые значения индексов WMS и UMS. В отношении величин остатков модель № 7 имеет оптимальные характеристики качества. Однако значения сепарационной статистики, отражающие надежность шкалы и ее дифференцирующие свойства, после удаления пунктов 10 и 11 несколько ухудшили свои показатели. Сепарационный индекс равен 3,66, индекс количества слоев — 5,21. Тем не менее данная модель имеет право на жизнь и нуждается в дальнейшем рассмотрении.

Заключение

Таким образом, в результате проведенного исследования было выявлено, что оригинальная шкала оценки депрессии Гамильтона содержит ряд пунктов, обладающих неудовлетворительной конструктивной валидностью. Конструктивная валидность оценивалась на основе значений показателей соответствия ожидаемых ответов, моделированных с помощью метрической системы Раша, наблюдаемым ответам, полученным при исследовании группы пациентов с депрессией различной степени тяжести. Пункты, имеющие большие величины остатков от разницы между ответами, не вписываются в рамки диагностической направленности шкалы. В шкале Гамильтона такими пунктами явились отдельные симптомы нарушений сна, психомоторных нарушений, ипохондрические симптомы и критика к состоянию. Ипохондрические симптомы и пункт «Критика к состоянию» были исключены из анализа. Три пункта, отражающие нарушения сна, и два пункта, относящиеся к психомоторным симптомам, были переформированы в объединенные два пункта. В результате были получены две диагностические модели (модели № 6 и № 7), обладающие приемлемыми характеристиками индексов качества и сепарационной статистики. Выделение этих двух моделей определяет задачи следующего исследования, которое должно быть направлено на определение в этих моделях критериев оценки тяжести депрессии и оценку того, как каждая из них согласуется с внешними клиническими критериями.

Литература

1. Ассанович М.А. Клиническая психодиагностика. – Минск: «Беларусь», 2012. – 344 с.
2. Ассанович М.А. Клиническая психодиагностика. Специализированные методики и опросники: учебное пособие. – Гродно: ГрГМУ, 2013. – 520 с.
3. Ассанович М.А. Инвариантность психометрических моделей // Журн. Гродн. гос. мед. универ. – 2014. – № 46. – С. 47–50.
4. Ассанович М.А. Проблема научного измерения в психодиагностике // Журн. Гродн. гос. мед. универ. – 2014. – № 45. – С. 9–14.
5. Andrich D. Rasch models for measurement. – Sage university, 1988. – 95 p.
6. Angoff W.H. Scales, norms, and equivalent scores. – Princeton, NJ: Educational Testing Service, 1984. – 144 p.
7. Bech P. Clinical psychometrics. – Wiley-Blackwell, 2012. – 211 p.
8. Bech P., Allerup P., Larsen E.R. The Hamilton Depression Scale (HAM-D) and the Montgomery-Asberg Depression Scale (MADRS). A psychometric re-analysis of the European genome-based therapeutic drugs for depression study using Rasch analysis // Psychiatry Res. – 2014. – Vol. 217, № 3. – P. 226–232.
9. Bond T.G. Applying the rasch model: fundamental measurement in the human science. – LEA, 2001. – 255 p.
10. Crocker L. Introduction to classical and modern test theory. – CENGAGE Learning, 2008. – 527 p.
11. Emberston S.E. Item Response Theory for psychologists. – LEA, 2000. – 371 p.

12. Fisher W. Reliability, separation, strata statistics // Rasch Measurement Transactions. – 1992. – Vol. 6, № 3. – P. 238.
13. Green K.E., Franton C.G. Survey development and validation with the Rasch model. – Charleston, 2002. – 42 p.
14. Hambleton R.K., Swaminathan H., Rogers H.J. Fundamentals of Item Response Theory. – Sage Publication, 1991. – 174 p.
15. Linacre J.M. What fo Infit and Outfit Mean-Square and Standardized mean? // Rasch Measurement transactions. – 2002. – Vol. 16, № 2. – P. 878.
16. Linacre J.M. Rasch power analysis: size vs. significance: standardized chi-square fit statistic // Rasch Measurement transactions. – 2003. – Vol. 17, № 1. – P. 918.
17. Massof R.W. The measurement of vision disability // Optometry and vision science. – 2002. – Vol. 79, № 8. – P. 516–552.
18. Masters G.N. A Rasch model for partial credit scoring // Psychometrika. – 1982. – Vol. 47.2. – P. 149–174.
19. Measurement in medicine / H. de Vet, C. Terwee, L. Mokkink [et al.]. – Cambridge: Cambridge University Press, 2011. – 338 p.
20. Michell J. Measurement in psychology: critical history of a methodological concept. – Cambridge: Cambridge University Press, 1999. – 265 p.
21. Onder I. An investigation of goodness of model data fit // Hacettepe Universitesi Egitim Fakultesi Degrisi. – 2007. – Vol. 32. – P. 210–220.
22. Reeve B.B. An introduction to modern measurement theory. – NCI, 2001. – 67 p.
23. Romera I., Perez V., Menchon J.M. Optimal cutoff point of the Hamilton Rating Sacale for Depression according to normal levels of social and occupational functioning // Psych. Res. – 2011. – Vol. 186, № 1. – P. 133–137.
24. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? / R.M. Bagby, A.G. Ryder, D.R. Schuller [et al.] // American Journal of Psychiatry. – 2004. – Vol. 161. – P. 2163–2177.
25. Wright B.D. Thinking with raw scores // Rasch Measurement Transactions. – 1993. – Vol. 7, № 2. – P. 299–300.
26. Wright B.D. Realiability and separation // Rasch Measurement Transactions. – 1996. – Vol. 9, № 4. – P. 472.
27. Wright B., Linacre J.M. Reasonable mean-square fit values // Rasch Measurement Transactions. – 1994. – Vol. 8, № 3. – P. 370.
28. Wright B.D., Masters G.N. Number of person or item strata // Rasch Measurement transactions. – 2002. – Vol. 16, № 3. – P. 888.
29. Wright B., Stone M. Measurement essentials. – Wilmington: Wide Range, 1999. – 221 p.
30. Zimmerman M., Posternak M.A., Chelminski I. Is the cutoff to define remission on the Hamilton Rating Scale for Depression too high? // J. Nerv. Ment. Dis. – 2005. – Vol. 193, № 3. – P. 170–175.

Optimization of the Hamilton Depression Rating Scale using Rasch model

Assanovich M.A.

Assanovich Marat Alievich

Candidate of Psychological Sciences, Department of Clinical Psychology and Psychotherapy; Grodno State Medical University, ul. Gorkogo, 80, Grodno, 230013, Belarus. Phone: +375-152-75-66-73.

E-mail: 70malas@gmail.com

Abstract. The aim of study was to optimize the Hamilton Depression Rating Scale (HAM-D) in terms of diagnostic items construct validity. Research methods: The study was conducted using polytomous Rasch model. The sample consisted of 550 patients suffering from depression of varying severity levels. Assessment of the item's construct validity have held after the scale building using Rasch metric system on the base of fit-indexes values. Results: Some items of the original HAM-D have had unsatisfactory construct validity. These includes three items containing symptoms of sleep disorders, two items, including violations of motor activity, hypochondriac symptom's item and item "Insight". During analysis there were built two optimal diagnostic models, including 12 and 10 diagnostic items. Discussion: psychometric problems HAM-D are supported by the literature. The hypochondriasis is not a typical symptom of depression, resulting in the corresponding item scale showed low construct validity. The structure of the item "Insight" does not provide adequate behavior of this item in responses patterns. Items reflecting insomnia form a major interference with each other, resulting in decrease of their validity. Items "Agitation" and "Retardation" are in opposite positions and work against each other, which determines the extremely high values of the fit-indices. However, Hamilton Depression Rating Scale contains items with adequate construct validity and acceptable values of quality indices. Removal of non-valid "noisy" items from the model allows to optimize the structure of the scale so that it will contain only valid diagnostic items. We obtained two optimal diagnostic models. The one includes 12 points, the second — 10 points. Further research should focus on the assessment which models is better agreement with the clinical data.

Keywords: polytomous Rasch model; Hamilton Depression Rating Scale; measuring models; separation statistics and indices of quality.

Bibliographic reference

Assanovich M.A. Optimization of the Hamilton Depression Rating Scale using Rasc model. *Med. psihol. Ross.*, 2015, no. 2(31), p. 7 [in Russian, in English]. Available at: <http://mprj.ru>

Received: March 20, 2015

Accepted: April 3, 2015

Publisher: April 17, 2015

Introduction

Max Hamilton believed that psychometrics in clinical psychiatry should take place as scientific discipline along with pharmacology and biochemistry. His psychometric elaborations Hamilton made at the clinic of the University of Leeds in 1957—1960. This period was a time modern psychopharmacology formation, the beginning of open antimanic effect of lithium and antipsychotic effect of chlorpromazine. Hamilton clearly saw the need for science-based brief rating scales, which could be used in psychopharmacological studies to assess the effect of new drugs. In 1959, Hamilton developed a scale for assessing anxiety, and in 1960 — rating scale for depression. Unlike his teacher Eysenck, who focused on the assessment of neuroticism, Hamilton, like Kraepelin, concentrated on the assessment of purely psychiatric symptoms of anxiety and depression. He believed that an objective assessment of psychopathology is the best way to get the clinical impression of the patient. The purpose of the Hamilton scales was not to make clinical diagnosis, but it was only in the evaluation of the severity of symptoms during a week. In order to study the structure and validation of rating scales Gamilton applied Spirmen's two-factor analysis. During factorizations of depression scale he identifies the different number of factors, depending on the homogeneity of the group of patients whose scale scores evaluated to factor. In fact, two-factor analysis proved useless in the study of the depression scale structure. The next 30 years of his life until his death in 1989 Hamilton was fully focused on the research the depression scale he created, but not completely solved the problem of its structure [7; 24].

Despite it's incomplete structure the Hamilton Depression Rating Scale (HAM-D) has become the most popular objective rating scale of depression in clinical studies. Recent reviews show that the scale has a generally satisfactory psychometric characteristics. However, some items are characterized by structural failure, which affects on the accuracy depression severity assessing. It is noted that the threshold criteria for differentiation a light

degree of depression and the norm don't conform to criterion validity of [23; 30]. Some of the items reveals a low construct validity. Because of this, in recent years, HAM-D is often subjected to critical attacks [8; 23; 24; 30].

The purpose of this study was to optimize the Hamilton Depression Rating Scale in terms of adequate construct validity of its items.

Research Methodology

The scientific measurement in the social sciences is based on the construction of the additive equal-interval scale with units equivalent to a fixed quantity of measured psychological construct. Raw test scores can not be considered as units of measurement, because they do not set the rules of equivalence with the construct levels. The sequence of the raw scores can not be considered as a measuring scale, since it lacks the additive structure, the distance between the scores are not equal-interval and scores themselves are not the units of measurement [5; 9; 17; 19; 20].

All kinds of normalization, standardization and smoothing, with the aim of adduction the raw estimates to the normal distribution are increasingly artificial and have no relation to the measurement [1; 3; 4; 6; 9].

The present day only approach that allows to create an adequate scientific standards measuring scale is the Rasch model [1; 7].

Rasch model was proposed by Danish mathematician G. Rasch in 1960 [29]. Since then, it has been evolved into a powerful system of probabilistic mathematical methods for constructing psychometric scales [13; 22]. The conceptual core of this model is based on an analysis of each answer to each test item [10; 11]. The basic Rasch model equation describes the functional relationship between probability of key response on the test item and the severity level of the construct had measured and the item difficulty to which the answer is given. Using as a starting point this basic equation there was developed the system of mathematical iterative methods for linear measuring equal-interval scales constructuion. The principal advantages of the Rasch model comparing with other psychometric approaches are: the possibility of constructing an invariant diagnostic scaling models and objective evaluation of the diagnostic items construct validity [10; 11; 13; 14; 19].

In the study we used polytomous Rasch model — partial credit model developed Masters in 1982 [18]. The model has been developed for the design and analysis of measuring scales with items that have two or more response categories. The model belongs to the family of Rasch model and contains all its useful properties: sufficiency of test score for construct measurement, separate assessment of persons and items. In fact the partial model is a simple adaptation of the Rasch model for polytomous items. The response categories for each item are represented in the usual sequence. The model basic equation describes the probability function of response for one of the categories at a particular level of construct [18; 22].

HAM-D does not contain any specific diagnostic questions linked to the diagnostic items. The clinician is offered to formulate questions, guided by the content of the items and one's own clinical experience. Naturally, this makes a significant element of subjectivity in the diagnostic assessment, reduces the validity and reliability of the data. In order to eliminate this drawback, in 1988 J. Williams (New York Institute of Mental Health) has developed a "Structured interview for the Hamilton Depression Rating Scale» (SIGH-D). The interview consists of 16 major diagnostic questions relevant to the items of the original Hamilton scale. The 17th item is identical to HAM-D's original one. Each main question has additional ones to clarify the depressive symptoms. Diagnostic criteria of interview meets original criteria of Hamilton scale. SIGH-D officially translated into several European languages, including Russian. Treatment protocol research and evaluation of the data corresponds to that of the original Hamilton Depression Rating Scale [2]. This interview we used in our study.

The sample for the response matrix consisted of 550 persons aged 23–54. Among them there were 328 women and 222 men. All subjects were psychiatric inpatients and outpatients having diagnosis of affective disorder. The study was conducted during the first three – five days after the clinical diagnosis. The distribution of patients according to group criteria is presented in Table 1.

Table 1

The distribution of patients by nosological groups

Clinical diagnosis	number of persons
Mild depressive episode	174
Moderate depressive episode	173
Severe depressive episode without of psychotic symptoms	125
Severe depressive episode with psychotic symptoms	78

Next after calculating the linear measures of depression and item difficulties construct validity of each item was evaluated. Evaluation of the items construct validity was carried using mean squares residuals values of difference between the predicted and actually obtained responses [15; 16; 21]. According to the rules of Rasch modeling we have assessed two standard indices: unweighted (UMS) and weighted (WMS) [29]. Both indices are essentially a chi-square test, divided by the number of degrees of freedom. The expected value of the index, which shows good agreement, equal to 1. Values less than 1 indicate over-predictable responses to the item in the context of the measured construct. Values greater than 1 indicate high noise level and low item agreement with the observed data. Empirically acceptable range of estimates for the index validity for diagnostic interviews is following [0,6 1,4] [27].

Next after building an objective measurement scale we have calculated separation statistics, including reliability index (separation index) and index of differentiating ability (number of strata) [12; 26; 28]. Items that do not meet the quality indices criteria were deleted and then diagnostic model was recalculated again. Model improving was determined by estimating fit-index values and number of strata.

Results and Discussion

As a result of applying the Rasch model algorithm we have achieved convergence of the iteration process at the level of the convergence criteria equal to 0.005. For each item of Hamilton scale we have calculated fit-indices UMS and WMS.

Let consider diagnostic consistency and construct validity of the scale items after conducted Rasch analysis. The fit-index values of the original scale items are shown in Table 2.

Table 2

The fit-index values of the Hamilton original scale items

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.58	0.59
2	Guilt	0.64	0.63
3	Suicide	0.79	0.79
4	Insomnia early	1.26	1.25

5	Insomnia middle	1.17	1.17
6	Insomnia late	1.17	1.13
7	Work and activities	0.54	0.56
8	Retardation	1.71	1.56
9	Agitation	1.27	1.53
10	Anxiety psychic	0.91	0.88
11	Anxiety somatic	0.85	0.77
12	Gastrointestinal symptoms	0.67	0.67
13	Somatic symptoms general	1.03	1.03
14	Genital symptoms	0.48	0.44
15	Hypochondriasis	1.23	1.16
16	Loss of weight	0.84	0.72
17	Insight	1.77	1.80

As follows from the table, 5 points scale find out values beyond the permissible range of fit-indices. These are such items as "Depressed mood", "Work and activities", "Retardation", "Agitation" and "Insight". These points may alter the response patterns agreement to levels of depression severity. This reduces the validity and accuracy of the assessment of the depression severity using scale total score. In fact, the above items do not have adequate construct validity in its original format of HAM-D. The number of strata index for this model was 4.45, the separation index — 3.09. This means that initially the scale is able to differentiate 4 statistically significant levels of depression severity. Analysis of the values in Table 2 shows that items 1, 7 and 14 have lower fit-index values, indicating the excessive predictability of the patients responses on these items. The fit-values for items 1 ("Depressed mood") and 7 ("Work and activities") slightly deviate from the limit values, while the index value for item 14 ("Genital symptoms") is reduced large. The behavior of these items in the scale is explained easily, since all three ones reflect the typical and expected symptoms of depression. Greater predictability of item 14 ("Genital symptoms") can be explained by a smaller number of response categories that has this item in comparison with others.

Items 8, 9 and 17 have a significant proportion of random variation that is reflected in the increased values of quality indices. With regard to item 8 ("Retardation") and 9 ("Agitation"), their behavior is understandable, since these two points are counterclaims against each other. This means that if the patient would receive high score at one of these points, in the second, as a rule — he would receive a low one. Depressive patients can not be simultaneously in agitation and retardation. Some patients have symptoms of agitation, the other part — the symptoms of retardation. Item 17 ("Insight") is characterized by the highest values of the fit-index indicating that shows the lowest construct validity throughout the scale. Thus, the results of the primary analysis of HAM-D using Rasch model in the Russian-speaking population confirm the data of foreign studies about diagnostic failure of individual scale items.

The next phase of this study was aimed at optimizing the HAM-D scale in terms of items construct validity. Primarily there have been deleted items with very high UMS and WMS values from the diagnostic model, in particular, items 8, 9, 17. Next, iterative re-analysis was performed using the partial credit model. Fit-index values are shown in Table 3.

Table 3

The fit-index values of the HAM-D diagnostic model № 2

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.68	0.68
2	Guilt	0.77	0.75
3	Suicide	0.88	0.86
4	Insomnia early	1.48	1.56
5	Insomnia middle	1.37	1.49
6	Insomnia late	1.39	1.38
7	Work and activities	0.61	0.62
10	Anxiety psychic	0.96	0.98
11	Anxiety somatic	0.89	0.84
12	Gastrointestinal symptoms	0.73	0.72
13	Somatic symptoms general	1.17	1.18
14	Genital symptoms	0.52	0.47
15	Hypochondriasis	1.44	1.44
16	Loss of weight	0.94	0.84

Analysis of the data in Table 3 shows that the fit-index values for items 1 and 7 "caught up" to an acceptable range from 0.6 to 1.4. Item 14 also have raised the values of fit-indices, but they still remain below the limits. Emphasis is placed on new items, the values of which had escaped the upper limit of the acceptable range. These items concern on sleep symptoms: item 4 ("Insomnia early") and item 5 ("Insomnia middle"). Item 15 ("Hypochondriasis") also have pulled out the acceptable range. Invalid behaviour of items 4 and 5 can be explained by the fact that sleep disorders are a fairly frequent occurrence in depression of any severity. The presentation of these symptoms in the Hamilton scale unclearly links the severity of sleep disorders, depending on the severity of depression. Quite pronounced insomnia symptoms may disturb patients with mild to moderate degrees of depression. As for the deviant behavior of item 15 ("Hypochondriasis"), it is to be expected. Hypochondriacal symptoms are relevant for severe depression when patients have a psychotic hypochondriacal delusions. In mild and moderately severe depression available hypochondriacal symptoms reflect not so much depression but the presence of comorbid hypochondriasis. Separation statistics of Rasch model have improved their values. In particular, the separation index is 3.27, the index of the number of layers — 4.69.

The next phase of the analysis had consisted in the exclusion of non-valid items from the scale and recalculation of diagnostic model. Quality item indicators after eliminating deviant symptoms are presented in Table 4.

Table 4

The fit-index values of the HAM-D diagnostic model № 3

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.76	0.77
2	Guilt	0.89	0.86
3	Suicide	1.04	1

6	Insomnia early	1.58	1.62
7	Work and activities	0.65	0.67
10	Anxiety psychic	1.06	1.09
11	Anxiety somatic	1.02	0.98
12	Gastrointestinal symptoms	0.81	0.79
13	Somatic symptoms general	1.3	1.34
14	Genital symptoms	0.58	0.5
16	Loss of weight	1.09	1.03

Analysis of the WMS index and UMS indices in Table 4 shows that only two items are outliers of data: i. 6 ("Insomnia early") and i. 14 ("Genital symptoms"). At the same i. 6 is characterized by abnormally high values of WMS and UMS, which indicates its low participation in measuring the severity of depression. Conversely, item 14 is still characterized by predictable behavior, which is reflected in the reduced values of the fit-indices. Separation statistics of the model are characterized by improved values in comparison with the previous one. The separation index is 3.41, the index of the number of layers — 4.88. This means that the number of statistically differentiate degrees of severity of depression is approaching to 5. Given the high value of the residue in item 6, it was decided to remove this item from the scale with the subsequent recalculation of the model. The WMS and UMS values of the next diagnostic model are presented in Table 5.

Table 5

The fit-index values of the HAM-D diagnostic model № 4

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.78	0.79
2	Guilt	0.91	0.88
3	Suicide	1.09	1.04
7	Work and activities	0.68	0.69
10	Anxiety psychic	1.15	1.20
11	Anxiety somatic	1.11	1.08
12	Gastrointestinal symptoms	0.86	0.84
13	Somatic symptoms general	1.38	1.43
14	Genital symptoms	0.62	0.53
16	Loss of weight	1.19	1.15

Diagnostic model № 4 has practically adequate properties. The fit-index values of the two items (i. 13 and 14) characterized by insignificant deviations from the acceptable range from 0.6 to 1.4. The separation index is 3.51, the index of the number of strata is 5.02, i.e., HAM-D scale in the form of diagnostic model №4 can differentiate 5 statistically significant levels of depression. Whether it is necessary to remove from the model 13 and 14? These deviations in the UMS and WMS are not as significant as in the nature of other items. However, they go beyond the acceptable range. In the next stage of the analysis we have removed these items and recalculate the model parameters. After recalculation of all of the remaining 8 items characterized by suitable index values. However, the removal of these

items had worsened separating statistics. The separation index value decreased to 3.19, while the index of the number of layers — up to 4.59. This means that the removal of points with minor deviations of the fit-index values at a significantly reduced number of diagnostic items in the scale reduces the reliability of the scale and degrades its differentiating properties. Therefore, the removal of these items generally having adequate construct validity is undesirable.

Another question arises: is it possible to improve the psychometric properties of diagnostic model number 4 discussed above by the addition of new items? The fact that we had removed two groups of symptoms, which could be combined into two new diagnostic items. The first group includes the symptoms of agitation (i. 8) and retardation (i. 9). The second group includes sleep disturbances (i. 4–6). Separately, each of these items has showed high values of the difference between expected and observed responses. This was largely due to their incoordinate behavior in scale. However, you can try to correct incoordinate items by combining them in each group in a common symptom using principle of "either-or". Group of items 8 and 9, we have combined into a single item "Psychomotor impairments." This item is measured at the maximum score of the answer to any of items 8 or 9. In the same way, we have combined the symptoms of sleep disorders (i. 4–6) in one diagnostic item "sleep disorders", estimated at a maximum score on any of the items 4 to 6. First, we have added the item "psychomotor impairments" in the model № 4. Characteristics of resulting diagnostic model № 5 are shown in Table 6.

Table 6

The fit-index values of the HAM-D diagnostic model № 5

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.75	0.77
2	Guilt	0.87	0.84
3	Suicide	1.11	1.10
7	Work and activities	0.69	0.70
8	Psychomotor impairments	0.85	0.86
9	Anxiety psychic	1.22	1.26
10	Anxiety somatic	1.21	1.18
12	Gastrointestinal symptoms	0.87	0.85
13	Somatic symptoms general	1.38	1.45
14	Genital symptoms	0.62	0.52
16	Loss of weight	1.20	1.11

Analysis of the data in Table 6 shows that the characteristics of this diagnostic model are slightly different from those of the previous model № 4. Indexes UMS and WMS for item "psychomotor impairments" (i. 8) are close to 1, indicating that adequate behavior of this item. Separation statistics have improved their values. The separation index is 3.76, the index of the number of layers — 5.36. The next step there was added the conjoint item "sleep disorders" in the new model. The fit-index values of the new model number 6 are shown in Table 7.

Table 7

The fit-index values of the HAM-D diagnostic model № 6

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.75	0.77
2	Guilt	0.87	0.85
3	Suicide	1.12	1.13
4	Sleep disorders	0.92	0.91
5	Work and activities	0.69	0.71
6	Psychomotor impairments	0.87	0.88
7	Anxiety psychic	1.22	1.29
8	Anxiety somatic	1.21	1.21
9	Gastrointestinal symptoms	0.88	0.86
10	Somatic symptoms general	1.39	1.47
11	Genital symptoms	0.61	0.52
12	Loss of weight	1.21	1.15

As shown in Table 7, item "Sleep disorders" (i. 4) has taken its place quite successfully in the scale, since the values of quality indices are close to 1, i.e., the ideal values. Such items as "General somatic symptoms" and "Genital symptoms" (i. 10, 11) have slightly worsened their UMS and WMS values. Separation statistics, meanwhile, increased their values. The separation index was 3.92, the index of the number of strata — 5.56. Further, another attempt was made to remove a bit "noisy" items 10 and 11 from the model with the subsequent recalculation of parameters. The fit-index values of resulting model № 7 are presented in Table 8.

Table 8

The fit-index values of the HAM-D diagnostic model № 7

№ i.	Title of the item	Fit-index values	
		WMS	UMS
1	Depressed mood	0.77	0.78
2	Guilt	0.90	0.89
3	Suicide	1.13	1.20
4	Sleep disorders	0.94	0.94
5	Work and activities	0.75	0.77
6	Psychomotor impairments	0.90	0.91
7	Anxiety psychic	1.19	1.30
8	Anxiety somatic	1.15	1.17
9	Gastrointestinal symptoms	0.88	0.86
12	Loss of weight	1.18	1.09

As follows from the analysis of the data in Table 8, all 10 points have valid indices of WMS and UMS. With respect to residual values model number 7 has the best quality characteristics. However, the separation statistics, reflecting the reliability of the scale and its differentiating properties deteriorate their performance after removal of items 10 and 11.

The separation index is 3.66, the index of the number of strata — 5.21. However, this model has a "right to life" and needs further consideration.

Conclusion

Thus, during the study we have revealed that the original Hamilton Depression Rating Scale contains a number of items that have poor construct validity. Construct validity was evaluated based on the values of the indicators of the agreement of expected answers, modeled using the Rasch metric system, to observed responses obtained in the study of group of patients with depression of varying severity. Items that have a large amount of residuals on the difference between the answers, do not fit into the framework of diagnostic orientation of the scale. In the Hamilton scale such items were some symptoms of sleep disorders, psychomotor impairments, hypochondriacal symptoms and insight. Hypochondriacal symptoms, and the "insight" were excluded from the analysis. Three items, reflecting sleep disorder, and two items related to psychomotor symptoms were reformed in the combined two points. As a result, the formation of two diagnostic model (model number 6 and № 7), have acceptable performance indexes of quality and separation statistics. Development of these two models defines the tasks of the next study, which should be aimed at the evaluation of the criteria for assessing of depression severity and an assessment of how each of them is consistent with external clinical criteria.

References

1. Assanovich M.A. *Klinicheskaya psikhodiagnostika* [Clinical psychodiagnosics]. Minsk, "Belarus" Publ., 2012. 344 p.
2. Assanovich M.A. *Klinicheskaya psikhodiagnostika. Spetsializirovannye metodiki i oprosniki: uchebnoe posobie* [Clinical psychodiagnosics. Specialized techniques and questionnaires: tutorial]. Grodno, GrGMU Publ., 2013. 520 p.
3. Assanovich M.A. Invariantnost' psikhometricheskikh modelei [Invariance of psychometric models]. *Zhurn. Grodn. gos. med. univer.*, 2014, no. 46, pp. 47–50.
4. Assanovich M.A. Problema nauchnogo izmereniya v psikhodiagnostike [Problem of scientific measurement in psychodiagnosics]. *Zhurn. Grodn. gos. med. univer.*, 2014, no. 45, pp. 9–14.
5. Andrich D. *Rasch models for measurement*. Sage university, 1988. 95 p.
6. Angoff W.H. *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service, 1984. 144 p.
7. Bech P. *Clinical psychometrics*. Wiley-Blackwell, 2012. 211 p.
8. Bech P., Allerup P., Larsen E.R. The Hamilton Depression Scale (HAM-D) and the Montgomery-Asberg Depression Scale (MADRS). A psychometric re-analysis of the European genome-based therapeutic drugs for depression study using Rasch analysis. *Psychiatry Res.*, 2014, vol. 217, no. 3, pp. 226–232.
9. Bond T.G. *Applying the rasch model: fundamental measurement in the human science*. LEA, 2001. 255 p.
10. Crocker L. *Introduction to classical and modern test theory*. CENGAGE Learning, 2008. 527 p.
11. Emberston S.E. *Item Response Theory for psychologists*. LEA, 2000. 371 p.
12. Fisher W. Reliability, separation, strata statistics. *Rasch Measurement Transactions*, 1992, vol. 6, no. 3, p. 238.
13. Green K.E., Franton C.G. *Survey development and validation with the Rasch model*. Charleston, 2002. 42 p.
14. Hambleton R.K., Swaminathan H., Rogers H.J. *Fundamentals of Item Response Theory*. Sage Publication, 1991. 174 p.
15. Linacre J.M. What fo Infit and Outfit Mean-Square and Standardized mean? *Rasch Measurement transactions*, 2002, vol. 16, no. 2, p. 878.

16. Linacre J.M. Rasch power analysis: size vs. significance: standardized chi-square fit statistic. *Rasch Measurement transactions*, 2003, vol. 17, no. 1, p. 918.
17. Massof R.W. The measurement of vision disability. *Optometry and vision science*, 2002, vol. 79, no. 8, pp. 516–552.
18. Masters G.N. A Rasch model for partial credit scoring. *Psychometrika*, 1982, vol. 47, pp. 149–174.
19. de Vet H., Terwee C., Mokkink L., Knol D. *Measurement in medicine*. Cambridge, Cambridge University Press, 2011. 338 p.
20. Michell J. *Measurement in psychology: critical history of a methodological concept*. Cambridge, Cambridge University Press, 1999. 265 p.
21. Onder I. An investigation of goodness of model data fit. *Hacettepe Universitesi Egitim Fakultesi Degrisi*, 2007, vol. 32, pp. 210–220.
22. Reeve B.B. *An introduction to modern measurement theory*. NCI, 2001. 67 p.
23. Romera I., Perez V., Menchon J.M. Optimal cutoff point of the Hamilton Rating Scale for Depression according to normal levels of social and occupational functioning. *Psych. Res.*, 2011, vol. 186, no. 1, pp. 133–137.
24. Bagby R.M., Ryder A.G., Schuller D.R., Marshall M.B. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *American Journal of Psychiatry*, 2004, vol. 161, pp. 2163–2177.
25. Wright B.D. Thinking with raw scores. *Rasch Measurement Transactions*, 1993, vol. 7, no. 2, pp. 299–300.
26. Wright B.D. Reliability and separation. *Rasch Measurement Transactions*, 1996, vol. 9, no. 4, p. 472.
27. Wright B., Linacre J.M. Reasonable mean-square fit values. *Rasch Measurement Transactions*, 1994, vol. 8, no. 3, p. 370.
28. Wright B.D., Masters G.N. Number of person or item strata. *Rasch Measurement transactions*, 2002, vol. 16, no. 3, p. 888.
29. Wright B., Stone M. *Measurement essentials*. Wilmington, Wide Range, 1999. 221 p.
30. Zimmerman M., Posternak M.A., Chelminski I. Is the cutoff to define remission on the Hamilton Rating Scale for Depression too high? *J. Nerv. Ment. Dis.*, 2005, vol. 193, no. 3, pp. 170–175.