

Funcionamento Diferencial do Item no Inventário de Depressão Beck

Differential Item Functioning in the Beck Depression Inventory

Stela Maris de Jesus Castro^I, Mariana Cúri^{II}, Vanessa Bielefeldt Leotti Torman^I, João Riboldi^I

RESUMO: *Introdução:* Diversos estudos mostram o Funcionamento Diferencial do Item (DIF) em itens do Inventário de Depressão Beck (BDI), ao compararem homens e mulheres. A presença de um grande número de itens com DIF no BDI é uma severa ameaça à validade da medida da intensidade de sintomas depressivos obtida pela Teoria da Resposta ao Item (TRI) e às conclusões baseadas nos escores derivados dos itens com e sem DIF. *Objetivo:* Os objetivos deste estudo foram identificar esses itens do BDI, ajustar o modelo de TRI para itens constrangedores (modelo 2), o qual acomoda itens com a presença de DIF, e comparar esses resultados com os do ajuste do modelo logístico de dois parâmetros tradicional da TRI (modelo 1). *Métodos:* Os resultados obtidos com ambos os modelos foram comparados. *Resultados:* Os itens que apresentaram DIF foram: tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fadigabilidade e perda da libido. Os resultados do ajuste dos dois modelos são similares quanto à discriminação, gravidade (à exceção dos itens com DIF) e no cálculo de escores para os indivíduos. Apesar disso, o modelo 2 é vantajoso, pois mostra as diferenças em gravidade do sintoma depressivo para os grupos avaliados, trazendo, dessa forma, mais informação ao pesquisador sobre a população estudada. *Conclusão:* Esse modelo, que tem um alcance mais amplo em termos de população-alvo, pode ser uma ótima alternativa na identificação e acompanhamento de indivíduos com potencial depressivo.

Palavras-chave: Teoria da Resposta ao Item. Funcionamento Diferencial do Item. Intensidade de Sintomas Depressivos. Inventário de Depressão Beck. Traço latente. Modelo TRI para itens constrangedores.

^IUniversidade Federal do Rio Grande do Sul – Porto Alegre (RS), Brasil.

^{II}Universidade de São Paulo – São Carlos (SP), Brasil.

Autor correspondente: Stela Maris de Jesus Castro. Rua João Mendes Ouriques, 650, Ipanema, CEP: 91760-450, Porto Alegre, RS, Brasil. E-mail: stela.castro@ufrgs.br

Conflito de interesses: nada a declarar – **Fonte de financiamento:** nenhuma.

ABSTRACT: Introduction: There are several studies showing the presence of Differential Item Functioning (DIF) in some items of the Beck Depression Inventory (BDI), when comparing men and women. The presence of a large number of items with DIF in BDI is a severe threat to the validity of measurement of the intensity of depressive symptoms obtained by Item Response Theory (IRT) and to the conclusions based on the scores derived from the items with or without DIF. **Objective:** The objectives of this study were to identify these items from the BDI, adjust the IRT model for embarrassing items (model 2), which accommodates items with the presence of DIF, and compare these results with the fit of the traditional two-parameter logistic IRT model (model 1). **Methods:** The results obtained with the both models were compared. **Results:** Items with DIF were: sadness, feeling of failure, dissatisfaction, guilty, punishment, crying, fatigability and loss of libido. The results of the adjustment of the two models are similar in discrimination, gravity (except for items with DIF), and in the calculation of scores for individuals. Nevertheless, model 2 is beneficial because it shows the differences in gravity of depressive symptoms for groups evaluated, thus providing more information to the researcher on the study population. **Conclusion:** This model, which has a broader scope in terms of target population, may be a good alternative to the identification and follow-up of individuals with potential depression.

Keywords: Item Response Theory. Differential Item Functioning. Intensity of Depressive Symptoms. Beck Depression Inventory. Latent trait. IRT Model for embarrassing items.

INTRODUÇÃO

Um traço latente é uma variável que não pode ser observada diretamente. Na tentativa de medi-lo, faz-se necessária a utilização de um instrumento composto por itens que presumivelmente o refletem. Estabelecer equivalência de medidas entre grupos que diferem em características como educação, gênero e raça, por exemplo, é importante em avaliação de saúde mental, para que esses grupos possam ser comparados em termos de suas medidas dos traços de interesse, tais como intensidade de sintomas depressivos, funcionamento físico ou satisfação com cuidado, por exemplo¹. Sendo assim, antes de comparar grupos de respondentes (em relação a idade ou gênero, por exemplo) em termos do traço latente a ser medido, deve-se estar confiante de que os itens que compreendem a medida operam equivalentemente entre os diferentes grupos¹. Em outras palavras, existe a possibilidade de que itens, em especial em medidas psicológicas e/ou psiquiátricas, funcionem diferentemente ou de forma enviesada para diferentes grupos de respondentes². Se um item tem uma função de resposta diferente para dois grupos, esse item é dito viesado³.

Na literatura sobre Teoria da Resposta ao Item (TRI), o termo viés do item tem sido essencialmente substituído pela expressão Funcionamento Diferencial do Item (DIF, do inglês *Differential Item Functioning*). O DIF ocorre quando a probabilidade de determinada resposta a um item do instrumento não tem a mesma relação com o traço latente em dois

ou mais grupos de respondentes, ou seja, quando a probabilidade de se escolher como resposta uma categoria de um item não depende apenas do traço latente do indivíduo, mas também do fato de ele pertencer a um determinado grupo (por exemplo, a probabilidade de escolher uma categoria de resposta é diferente entre homens e mulheres com o mesmo nível do traço latente). Mais concretamente, o DIF ocorre quando um item apresenta uma Curva Característica do Item (CCI) diferente para cada grupo ou, equivalentemente, quando qualquer parâmetro do item difere entre os grupos. Se um item está livre desse viés, as respostas para esse item serão relacionadas somente com o nível do traço latente que o item está procurando medir. Se o item apresenta viés, as respostas ao mesmo serão relacionadas a algum outro fator além do nível do traço latente.

Muitos instrumentos de medida, em especial na área da psiquiatria, apresentam itens que podem funcionar de maneira diferente nos diversos grupos. Entre eles, pode-se citar o Inventário de Depressão Beck (BDI, do inglês *Beck Depression Inventory*), um instrumento que estima o traço latente Intensidade de Sintomas Depressivos. Alguns estudos relatam a presença de itens com DIF no BDI em relação ao gênero^{4,6}. A diferença entre as distribuições das respostas de homens e mulheres se deram em itens referentes a choro, sentimento de fracasso, punição, perda da libido, insatisfação, culpa e fadigabilidade.

A presença de um grande número de itens com DIF no BDI é uma severa ameaça à validade da medida da intensidade de sintomas depressivos obtida pela TRI e às conclusões baseadas nos escores derivados dos itens com e sem DIF. Uma possível solução para esse problema poderia ser a eliminação dos mesmos do instrumento de medida. No entanto, isso poderia comprometer a medida do traço latente, pois os itens provavelmente contêm informação relevante à medida do mesmo, já que o BDI foi construído para abranger todos os sintomas depressivos observáveis⁷. O uso de um modelo que permita a manutenção de todos os itens no instrumento e, ao mesmo tempo, contemple as diferenças entre os grupos trabalhados vem a ser uma ótima alternativa para a análise de dados oriundos do BDI.

O Modelo TRI para itens constrangedores, proposto por Cúri et al.⁸, insere-se dentro dessa perspectiva, pois preserva tais características. Sendo assim, este estudo teve por objetivos identificar itens do BDI que apresentem DIF para gênero, isto é, que apresentem viés comparando-se homens e mulheres, por meio da análise diferencial do item, ajustar o modelo para itens constrangedores para a amostra de sujeitos considerada e comparar esses resultados com os do ajuste do modelo logístico de dois parâmetros tradicional da TRI.

MÉTODOS

AMOSTRA

Os indivíduos são provenientes de um estudo transversal conduzido para realizar a adaptação, a normatização e a validação para o português das Escalas Beck, em um estudo conduzido pela Dra. Jurema Alcides Cunha e publicado em 2001⁹.

A escala do BDI, originalmente com 4 pontos, para os objetivos do presente trabalho, foi dicotomizada de modo que a resposta assume o valor 1 ($X_{ij} = 1$) quando o indivíduo j relata apresentar o sintoma descrito no item i (isto é, escolhe uma das categorias com escores 1, 2 ou 3 do determinado item) e 0 ($X_{ij} = 0$) caso não apresente aquele sintoma.

MODELOS CONSIDERADOS

Foram adotados dois modelos de TRI para variáveis dicotômicas (aqui sendo do tipo ausência ou presença de sintoma depressivo).

Modelo logístico unidimensional de 2 parâmetros (Modelo 1)

Este é um modelo de TRI para resposta dicotômica, apropriado para medidas nas quais os itens não discriminam igualmente os níveis do traço latente^{2,10}. O modelo de 2 parâmetros prediz a probabilidade de o indivíduo j apresentar o sintoma medido no item i , condicionado a sua intensidade de sintomas depressivos, isto é, $P(X_{ij} = 1 \mid \theta_j, \zeta_i)$, como segue:

$$P(X_{ij} = 1 \mid \theta_j, \zeta_i) = \frac{1}{1 + e^{-a_i(\theta_j, \zeta_i)}} \quad (1)$$

onde: $i = 1, \dots, 21$ itens, $j = 1, \dots, n$ indivíduos, $\zeta_j = (a_j, b_j)^t$, θ_j é a intensidade de sintomas depressivos (traço latente) do indivíduo j (parâmetro de indivíduo); b_i é o parâmetro de gravidade (posição) do item i e representa a gravidade do sintoma depressivo descrito pelo item i (quando $\theta_j = b_i$, a probabilidade de presença do sintoma i é 0,5); a_i é o parâmetro de discriminação (ou inclinação) do item i .

Modelo TRI para itens constrangedores (Modelo 2)

Este modelo para itens dicotômicos, proposto por Cúri et al.⁸, possibilita diferenciar a gravidade da presença do sintoma depressivo entre indivíduos constrangidos e não constrangidos por um determinado item de forma a apresentarem comportamentos diferenciados frente à respectiva CCI. A probabilidade de o indivíduo j apresentar ou não o sintoma medido no item i ($X_{ij} = 1$ ou 0, respectivamente) e constranger-se ou não pelo item i ($C_{ij} = 1$ ou 0, respectivamente) é:

$$P(X_{ij} = x_{ij}, C_{ij} = c_{ij} \mid \theta_j, \zeta_i) = P(X_{ij} = x_{ij} \mid C_{ij} = c_{ij}, \theta_j, \zeta_i) \times P(C_{ij} = c_{ij} \mid \theta_j, \zeta_i) \quad (2)$$

$$= [(P_{ij}^*)^{x_{ij}} (1 - P_{ij}^*)^{(1-x_{ij})} \zeta_i^{c_{ij}}] \times [P_{ij}^{x_{ij}} (1 - P_{ij})^{(1-x_{ij})} (1 - \zeta_i)^{(1-c_{ij})}]$$

onde:

$\zeta_i = (a_{ij}, b_{1j}, b_{2j}, \gamma_i, \delta_i)^t$; $P_{ij}^* = \frac{\gamma_i}{1 + e^{-a_i(\theta_j, b_{2i})}} = P(X_{ij} = 1 \mid C_{ij} = 1, \theta_j, \zeta_j)$; $P_{ij} = \frac{1}{1 + e^{-a_i(\theta_j, b_{1i})}} = P(X_{ij} = 1 \mid C_{ij} = 0, \theta_j, \zeta_j)$; θ_j é a intensidade de sintomas depressivos (traço latente) do indivíduo j (parâmetro de indivíduo); b_{1i} é o parâmetro de gravidade do item i para indivíduos não constrangidos, denominados a partir de agora como grupo com comportamento padrão (mulheres); b_{2i} é o parâmetro de gravidade do item i para indivíduos constrangidos, denominados a partir de agora como grupo com comportamento diferenciado (homens); a_i é o parâmetro de discriminação (ou inclinação) do item i ; γ_i é a probabilidade de o indivíduo do grupo com comportamento diferenciado afirmar que tem o sintoma depressivo, isto é, probabilidade de o indivíduo constrangido dizer que tem o sintoma dado que realmente o apresenta (repare que, no grupo não constrangido, assume-se que essa probabilidade é 1); δ_i é a probabilidade de um indivíduo apresentar comportamento diferenciado com relação ao sintoma i . Neste estudo, se assumirá que a classificação dos indivíduos como constrangidos ou não constrangidos será dada pelo sexo, ou seja, $C_{ij} = 1$, se homem, ou 0, se mulher.

Esse modelo, além do parâmetro de discriminação do item, comum aos demais modelos de TRI, estima outros parâmetros que contemplam o funcionamento diferencial daqueles itens que apresentam DIF. Para esses itens, os grupos são comparáveis entre si, mas não se pode fazer isso olhando para os parâmetros de gravidade. Os parâmetros b_{1i} e b_{2i} expressam probabilidades diferentes de o indivíduo apresentar o sintoma. A comparação correta entre gravidades deve ser feita entre b_{1i} e $\theta_{0,5}^*$.

Repare que b_{1i} , como no modelo logístico de 2 parâmetros, pode ser interpretado como a intensidade dos sintomas depressivos de um indivíduo com comportamento padrão, tal que a probabilidade de presença do sintoma i é igual a 0,5 (quando $\theta_j = b_i$, $P_{ij} = 0,5$). Por outro lado, para indivíduos com comportamento diferenciado, quando $\theta_j = b_{2i}$, $P_{ij}^* = \gamma_i/2$. Por essa razão, não faz sentido comparar b_{1i} e b_{2i} . Neste trabalho, a interpretação da gravidade do sintoma avaliado por um item com DIF será feita pela comparação das intensidades de sintomas depressivos de indivíduos de cada um dos 2 grupos para os quais a probabilidade de presença do sintoma é igual a 0,5. No grupo com comportamento padrão, equivale a $\theta_j = b_i$ e, no grupo com comportamento diferenciado, $\theta_j = -(1/a_i) \ln[(\gamma_i - 0,5) / 0,5] + b_{2i}$ (cuja estimativa é denotada por $\theta_{0,5}^*$).

ESTRATÉGIA DE ANÁLISE

A análise de funcionamento diferencial do item foi realizada por meio da técnica conhecida como Razão de Verossimilhanças em Teoria da Resposta ao Item¹¹ (IRTLR, do inglês *Item Response Theory Log-Likelihood Ratio*, versão 2.0b), utilizando o *software* IRTLRI, desenvolvido por Dave Thissen e disponibilizado em sua *homepage*¹². Esse procedimento parte da definição de Frederic Lord sobre DIF (então chamado de viés do item) e usa o teste da razão de verossimilhança como um teste de significância para a hipótese de nulidade de que os parâmetros de uma função de resposta de um item não diferem entre grupos — um

resultado significativo aponta a detecção de DIF. Como, para modelos de TRI paramétricos, o grupo de parâmetros do item é isomorfo (tem a mesma forma) com a função de resposta do item, o exame das diferenças entre os parâmetros do item é equivalente ao exame da diferença nas funções de resposta do item. O *software* IRTLRDIF tem implementado dois dos modelos de TRI mais comumente usados: o modelo logístico de 3 parâmetros e o modelo politômico de resposta gradual de Samejima¹³. O modelo logístico de 2 parâmetros (utilizado para a identificação de itens com DIF) é um caso especial de ambos os modelos anteriores e, nesse *software*, está implementado como um modelo de resposta gradual com duas categorias de resposta. Em função do tamanho da amostra, o nível de significância adotado para a identificação dos itens com DIF foi de 1%.

O ajuste dos modelos de TRI logístico de 2 parâmetros (modelo 1) e para itens constrangedores (modelo 2) foi realizado por meio de rotinas elaboradas no WinBUGS, versão 1.4.3¹⁴. As rotinas referentes a ambos os modelos utilizam um método bayesiano de estimação de parâmetros por meio de simulação de Monte Carlo via Cadeia de Markov (MCMC).

Este estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Sul (UFRGS), na reunião n° 37, ata n° 117, de 30 de outubro de 2008.

RESULTADOS

As características demográficas da amostra podem ser encontradas no artigo Teoria da resposta ao item aplicada ao Inventário de Depressão Beck¹⁵, onde o modelo de Resposta Gradual de Samejima foi ajustado a esses dados. É importante ressaltar que os indivíduos da amostra estão divididos quase que igualmente entre homens e mulheres, com pequena vantagem para as últimas.

Os itens que apresentaram DIF, de acordo com a técnica da Razão de Verossimilhança, foram: tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fatigabilidade e perda da libido. Os resultados do ajuste do modelo 1 encontram-se na Tabela 1 e os resultados do ajuste do modelo 2, considerando-se os 8 itens que apresentaram DIF e o grupo masculino como o grupo com os indivíduos constrangidos pelos mesmos, na Tabela 2.

As estimativas dos parâmetros de discriminação nos modelos 1 e 2 (Tabelas 1 e 2, respectivamente) indicam que praticamente todos os itens podem ser considerados apropriados quanto a essa característica ($a_i > 1^{8,10}$), à exceção de perda de peso e autoacusações. Os itens com maior poder de discriminação são os relacionados a sintomas de sentimento de fracasso e insatisfações.

A partir das estimativas de gravidade (b_i) dos sintomas depressivos (Tabela 1), observa-se que os sintomas autoacusações e irritabilidade são os menos graves e os sintomas perda de peso e ideias suicidas, os mais graves. É interessante ressaltar que perda de peso é o sintoma depressivo mais grave e, ao mesmo tempo, o que menos discrimina a população ($\hat{a}_{19}=0,93$). No entanto, o sintoma ideias suicidas é o segundo mais grave ($\hat{b}_9=1,20$) e discrimina bem a população quanto ao nível de intensidade de sintomas depressivos ($\hat{a}_9=1,71$).

Tabela 1. Média e desvio padrão da distribuição *a posteriori* dos parâmetros do modelo logístico de 2 parâmetros (modelo 1).

Item	a_i (DP)	b_i (DP)
1 Tristeza	2,38 (0,09)	0,16 (0,02)
2 Pessimismo	2,41 (0,10)	0,76 (0,03)
3 Sentimento de fracasso	2,90 (0,12)	0,71 (0,03)
4 Insatisfações	2,79 (0,11)	0,10 (0,02)
5 Culpa	1,98 (0,08)	0,57 (0,03)
6 Punição	1,48 (0,06)	0,66 (0,04)
7 Autoaversão	2,50 (0,10)	0,51 (0,02)
8 Autoacusações	0,97 (0,05)	-1,19 (0,06)
9 Ideias Suicidas	1,71 (0,08)	1,20 (0,04)
10 Choro	1,69 (0,06)	0,36 (0,03)
11 Irritabilidade	1,07 (0,05)	-0,65 (0,04)
12 Retraimento Social	1,51 (0,06)	0,71 (0,03)
13 Indecisão	1,99 (0,08)	0,13 (0,03)
14 Mudança na autoimagem	1,72 (0,07)	0,52 (0,03)
15 Dificuldade de trabalhar	1,94 (0,07)	0,30 (0,03)
16 Insônia	1,44 (0,06)	0,06 (0,03)
17 Fatigabilidade	1,41 (0,06)	-0,25 (0,03)
18 Perda de apetite	1,22 (0,06)	0,96 (0,04)
19 Perda de peso	0,93 (0,06)	2,00 (0,11)
20 Preocupações somáticas	1,19 (0,05)	0,35 (0,04)
21 Perda da libido	1,43 (0,06)	0,74 (0,04)

a_i : parâmetro de discriminação do item i ; b_i : parâmetro de gravidade do item i ; DP: desvio padrão.

Quanto à gravidade dos sintomas, os resultados são os mesmos do modelo 1 para todos os itens do BDI que não apresentam DIF. As diferenças ocorrem apenas nos oito itens restantes. Observa-se que a presença de culpa tem maior probabilidade de ser observada em níveis mais altos de intensidade de sintomas depressivos ($\hat{b}_{1,5}=0,58$) nas mulheres e mais baixos nos homens ($\theta_{0,5}^* = 0,53$). Isso se inverte para os sintomas tristeza, sentimento de fracasso, insatisfações, punição, perda da libido, choro e fatigabilidade. Por exemplo, a presença de perda da libido tem maior probabilidade de ser observada em níveis mais baixos de intensidade de sintomas depressivos ($\hat{b}_{1,21}=0,48$)

Tabela 2. Média e desvio padrão da distribuição *a posteriori* dos parâmetros do modelo para itens constrangedores (modelo 2).

Item	a_i (DP)	b_{1i} (DP)	b_{2i} (DP)	$\theta_{0,5}^*$	γ_i (DP)	δ_i (DP)
1 Tristeza	2,37 (0,12)	0,07 (0,03)	0,16 (0,07)	0,28	0,88 (0,04)	0,45 (0,008)
2 Pessimismo	2,34 (0,10)	0,76 (0,03)	–	–	–	–
3 Sentimento de fracasso	2,94 (0,15)	0,68 (0,03)	0,70 (0,06)	0,76	0,92 (0,06)	0,45 (0,007)
4 Insatisfações	2,87 (0,14)	0,04 (0,03)	0,06 (0,05)	0,14	0,90 (0,04)	0,45 (0,007)
5 Culpa	2,06 (0,09)	0,58 (0,03)	0,50 (0,05)	0,53	0,97 (0,03)	0,45 (0,007)
6 Punição	1,53 (0,07)	0,65 (0,04)	0,54 (0,10)	0,67	0,91 (0,06)	0,45 (0,007)
7 Autoaversão	2,53 (0,10)	0,51 (0,02)	–	–	–	–
8 Autoacusações	0,99 (0,05)	-1,17 (0,06)	–	–	–	–
9 Ideias suicidas	1,71 (0,08)	1,20 (0,04)	–	–	–	–
10 Choro	1,62 (0,07)	0,29 (0,04)	0,42 (0,07)	0,49	0,95 (0,04)	0,45 (0,008)
11 Irritabilidade	1,08 (0,05)	-0,65 (0,04)	–	–	–	–
12 Retraimento social	1,51 (0,06)	0,71 (0,03)	–	–	–	–
13 Indecisão	1,99 (0,08)	0,13 (0,03)	–	–	–	–
14 Mudança na autoimagem	1,71 (0,07)	0,52 (0,03)	–	–	–	–
15 Dificuldade de trabalhar	1,91 (0,07)	0,30 (0,03)	–	–	–	–
16 Insônia	1,43 (0,06)	0,06 (0,03)	–	–	–	–
17 Fatigabilidade	1,38 (0,06)	-0,33 (0,05)	-0,25 (0,07)	-0,19	0,96 (0,03)	0,45 (0,007)
18 Perda de apetite	1,21 (0,06)	0,96 (0,05)	–	–	–	–
19 Perda de peso	0,93 (0,06)	2,00 (0,10)	–	–	–	–
20 Preocupações somáticas	1,18 (0,05)	0,35 (0,03)	–	–	–	–
21 Perda da libido	1,19 (0,06)	0,48 (0,04)	1,39 (0,11)	1,50	0,94 (0,05)	0,45 (0,007)

a_i : parâmetro de discriminação do item i ; b_{1i} : parâmetro de gravidade do item i para indivíduos do grupo feminino; DP: desvio padrão; b_{2i} : parâmetro de gravidade do item i para indivíduos do grupo masculino; $\theta_{0,5}^*$: nível de intensidade de sintomas depressivos de um indivíduo do grupo com comportamento diferenciado, em que a probabilidade de presença do sintoma é 0,5; γ_i : probabilidade de um indivíduo do grupo masculino dizer que tem o sintoma depressivo, isto é, probabilidade de presença de sintoma em homens com um alto nível de intensidade de sintomas depressivos; δ_i : probabilidade de um indivíduo apresentar comportamento diferencial com relação ao sintoma i .

nas mulheres e mais altos nos homens ($\theta_{0,5}^* = 1,50$). Ainda como um resultado do modelo 2, estima-se que a probabilidade de um homem com alta intensidade de sintomas depressivos expressar a presença dos sintomas relativos a tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fatigabilidade e perda da libido é maior ou igual a 88% ($\hat{\gamma}_i \geq 0,88$). A Figura 1 mostra as CCI's produzidas pelos modelos

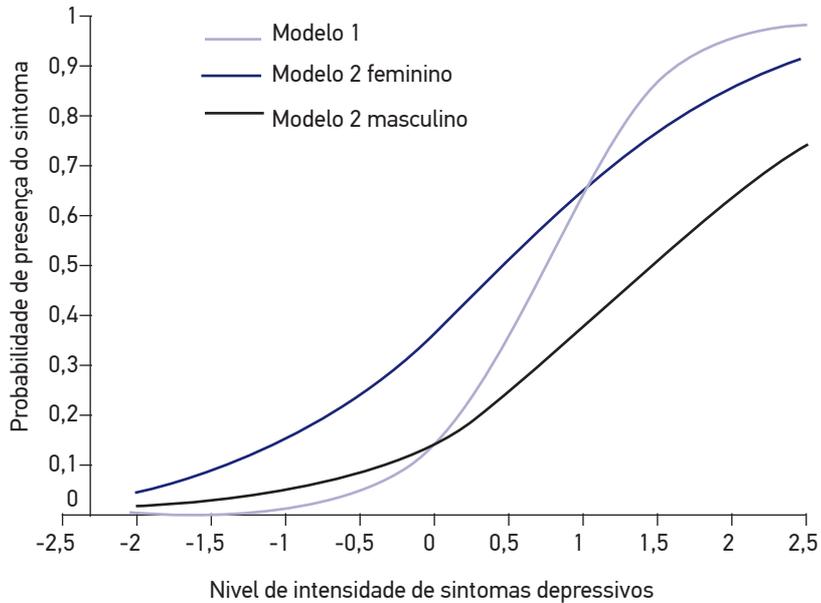


Figura 1. Curva Característica do Item (CCI) para o sintoma perda da libido (item 21) segundo o modelo logístico de 2 parâmetros (1) e o modelo para Itens Constrangedores (2) para os sexos feminino e masculino.

1 e 2 para o item 21, relativo à perda da libido. Aqui, fica evidente a vantagem do uso do modelo 2 em relação ao modelo 1, pois as diferenças no comportamento de um item com DIF, em relação à sua gravidade, são mostradas claramente para os dois grupos comparados.

Os níveis de intensidade de sintomas depressivos estimados sob os modelos de TRI estão na mesma escala da gravidade do sintoma estimada para cada item do BDI; logo, são comparáveis. Os percentis 95 dos níveis de intensidade de sintomas depressivos são 1,598 e 1,593 para os modelos 1 e 2, respectivamente. Dos 201 indivíduos com intensidade de sintomas depressivos acima do percentil 95 para cada modelo, 194 são classificados de forma igual pelos dois modelos. As características (Tabela 3) desse grupo mostram que quase 80% são oriundos do grupo psiquiátrico, aproximadamente 68% são mulheres, a maioria (mais de 58%) não tem um companheiro(a) e tem, em média, 37 anos de idade.

As estimativas de intensidade de sintomas depressivos obtidas segundo os modelos 1 e 2 apresentam alta associação, com coeficiente de correlação igual a 0,99.

Tabela 3. Descrição dos indivíduos com nível alto de sintomas depressivos, estimado como um valor acima do percentil 95.

Variável	Psiquiátrica	Clínica	População geral	Total
Modelo 1	n = 157	n = 18	n = 26	n = 201
Idade (n = 200)				
Média	39,18	34,67	29,76	37,60
Desvio padrão	12,5	13,26	13,83	13,08
Mínimo	15	18	18	15
Máximo	75	64	67	75
Escolaridade (n = 194)				
Menos de 5 anos	54	9	12	38,1%
Ensino Fundamental Completo	43	3	5	25,9%
Ensino Médio Completo	43	5	5	26,9%
Ensino Superior Completo	11	0	4	7,7%
Estado Civil (n = 199)				
Solteiro	45	8	19	36,2%
Casado	74	8	1	41,7%
Desquitado, divorciado ou viúvo	36	2	6	22,1%
Sexo (n = 201)				
Masculino	36	11	16	31,3%
Feminino	121	7	10	68,7 %
Modelo 2	n = 155	n = 17	n = 29	n = 201
Idade (n = 200)				
Média	39,03	34,12	28,64	37,15
Desvio padrão	12,46	13,45	13,48	13,16
Mínimo	15	18	15	15
Máximo	75	64	67	75
Escolaridade (n = 194)				
Menos de 5 anos	53	8	13	37,6%
Ensino Fundamental Completo	43	3	6	26,4%
Ensino Médio Completo	43	5	5	26,9%
Ensino Superior Completo	10	0	4	7,1%
Estado Civil (n = 199)				
Solteiro	47	8	22	38,7%
Casado	72	7	1	40,2%
Desquitado, divorciado ou viúvo	34	2	6	21,1%
Sexo (n = 201)				
Masculino	38	10	18	32,8%
Feminino	117	7	11	67,2 %

$\hat{\theta}$: traço latente (nível de sintomas depressivos) estimado a partir dos dados da amostra.
 $\hat{\theta}=1,598$ para o modelo 1 e $\hat{\theta}=1,593$ para o modelo 2.

DISCUSSÃO

Utilizou-se o modelo logístico de dois parâmetros (modelo 1) visando à comparação com o modelo para itens constrangedores (modelo 2) porque ambos incluem parâmetros de discriminação e gravidade do sintoma depressivo. Outros estudos já utilizaram o modelo logístico de 2 parâmetros a dados psiquiátricos: Schaeffer¹⁶, em 1988, ajustou esse modelo a respostas para 11 sintomas de depressão para os quais existiam 4 categorias de resposta (“nunca”, “uma vez até agora”, “com relativa frequência” e “muitas vezes”) e Kessler et al.¹⁷ aplicaram-no na construção de 2 escalas (uma com 10 itens e outra com 6) sobre saúde mental.

Os achados em relação ao modelo 1, quanto à presença de DIF em oito itens do BDI, mostram que homens e mulheres com o mesmo nível de intensidade de sintomas depressivos responderam de modo diferente aos itens tristeza, sentimento de fracasso, insatisfações, culpa, punição, choro, fadigabilidade e perda da libido. Diversos estudos^{4-6,18-24} corroboram esses achados; entretanto, o funcionamento diferencial (DIF) do item choro em relação ao gênero é o que aparece na maioria deles. Boa parte dos estudos que mostram a diferença de gênero em relação ao choro salienta que mulheres tendem a chorar mais que homens^{5,6,21}. Isso pode ser mais um reflexo da bem conhecida tendência das mulheres chorarem mais prontamente e mais intensamente que os homens em uma variedade de situações angustiantes do que um indicador de uma diferença de gênero na prevalência da depressão¹⁸. Essa constatação sugere que o choro em resposta à angústia é, em grande medida, determinado pelo gênero; logo, homens e mulheres com o mesmo nível de intensidade de sintomas depressivos provavelmente não responderão ao item choro da mesma forma, o que se confirma neste estudo. Originalmente, a escala do BDI tem quatro categorias, sendo que, em especial no item sobre choro, a categoria de ordem mais alta afirma que o indivíduo perdeu a capacidade de chorar, mesmo que tenha vontade, enquanto as três primeiras categorias determinam um crescente na quantidade de vezes que se costuma chorar. De todos os homens que receberam o valor 1 na escala dicotomizada, mais da metade respondeu a categoria 3, o mesmo ocorrendo quando se observa apenas os homens que pertencem ao grupo dos 5% maiores valores estimados para nível de intensidade de sintomas depressivos, mostrando que são sérios candidatos a um diagnóstico positivo para depressão. Essa perda da capacidade de chorar masculina também aparece no estudo de Hammen e Padesk⁴, no qual se trabalhou com o BDI na escala original.

Comparando-se os resultados encontrados para os modelos 1 e 2 em relação à discriminação dos sintomas depressivos medidos pelos itens, pode-se perceber que, considerando-se itens com valores de $a_i \geq 1$ ^{8,10} como tendo razoável discriminação, os mesmos 19 itens nos 2 modelos estão nessa categoria, com exceção apenas de perda de peso e autoacusações. No estudo de Cúri et al.⁸, no qual foi ajustado um modelo logístico de três parâmetros a dados do BDI, apenas perda de apetite obteve estimativa para discriminação abaixo desse ponto de corte, porém o sintoma perda de peso está muito próximo dessa região. Já entre os sintomas que mais discriminam, sentimento de fracasso e insatisfação, são os que aparecem

nos modelos 1, 2 e no ajustado por Cúri et al.⁸, mostrando que estes são sintomas importantes na discriminação da população quanto à intensidade de sintomas depressivos.

Um resultado que aparece no modelo 2 é a maior gravidade do sintoma perda da libido para os homens do que para as mulheres, pois a maior probabilidade de presença do mesmo ocorre em níveis de intensidade de sintomas depressivos mais altos para homens do que para as mulheres. A importância da perda da libido para os homens aparece em diversos estudos. Em um ensaio clínico randomizado sobre os efeitos sexuais (tais como melhora na perda da libido e na disfunção erétil) da reposição de testosterona em homens com diagnóstico de depressão maior²⁵, os autores pretendiam verificar se o tratamento seria eficaz nessa população da mesma forma que o é na população geral. No entanto, a reposição de testosterona não surtiu o efeito já conhecido, indicando que talvez o impedimento tenha sido a condição de depressão da população-alvo.

Os grupos formados pelos 5% de indivíduos com maiores estimativas de intensidade de sintomas depressivos (traço latente sendo medido), obtidas a partir dos modelos 1 e 2, evidenciam a superioridade feminina no grupo psiquiátrico, pois mais de 75% desses grupos é formado por mulheres. Esses dados são consistentes com a evidência de que a depressão é de duas a três vezes mais comum em mulheres adolescentes e adultas do que em homens adolescentes e adultos²⁶, pois essas mulheres têm os níveis de intensidade de sintomas depressivos mais altos, sendo fortes candidatas a terem um diagnóstico positivo para depressão.

É importante ressaltar que os modelos 1 e 2 rastreiam praticamente os mesmos indivíduos como pertencentes a esses grupos com as maiores estimativas de intensidade de sintomas depressivos. Dos 201, apenas 7 mulheres e 7 homens apresentam classificações discordantes, sendo que o modelo 1 rastreia mais mulheres como acima do percentil 95 e o modelo 2 rastreia mais homens como acima do seu respectivo percentil 95. Essas diferenças parecem ocorrer devido ao fato de que os níveis de intensidade de sintomas depressivos estimados para esses indivíduos estão no limite de seu respectivo valor do percentil 95.

CONCLUSÃO

Foram ajustados dois modelos de TRI aos dados dicotomizados do BDI: o modelo logístico de 2 parâmetros (modelo 1) e o modelo de TRI para itens constrangedores (modelo 2), o qual acomoda a presença de itens que apresentam DIF.

Os resultados encontrados entre os modelos 1 e 2 são bastante similares, especialmente no caso das estimativas de intensidade de sintomas depressivos para cada indivíduo, comprovado pela alta correlação entre os escores de TRI. Apesar disso, o modelo 2 ainda é vantajoso, pois mostra as diferenças em gravidade do sintoma depressivo para os grupos avaliados, trazendo, dessa forma, mais informação ao pesquisador sobre a população estudada. A utilização de um modelo de alcance mais amplo em termos de população-alvo pode ser uma alternativa

bastante útil também na área clínica, onde a existência de modelos validados pode contribuir para a identificação de indivíduos com potencial depressivo.

Uma limitação deste trabalho é a de que ele consiste de uma comparação empírica, fazendo-se necessário um estudo mais amplo, valendo-se, por exemplo, de dados simulados.

Ainda, como Cúri et al.⁸ comentam, faz-se necessária a extensão do modelo 2 a itens com resposta ordinal, pois, assim como o BDI, inúmeros instrumentos de medidas psiquiátricas contemplam itens de resposta ordinal, e a sua transformação em itens dicotômicos (do tipo ausência ou presença, por exemplo) não aproveita totalmente a informação disponível, podendo produzir resultados com alguma inconsistência.

REFERÊNCIAS

1. Teresi JA, Fleishman JA. Differential item functioning and health assessment. *Qual Life Res* 2007; 16(Suppl 1): 33-42.
2. Embretson SE, Reise SP. *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates; 2000.
3. Lord F. *Applications of item response theory to practical testing problems*. Hillsdale: Routledge; 1980.
4. Hammen CL, Padesky CA. Sex differences in the expression of depressive responses on the Beck Depression Inventory. *J Abnorm Psychol* 1977; 86(6): 609-14.
5. Santor D, Ramsay J, Zuroff D. Nonparametric item analyses of the Beck Depression Inventory: evaluating gender item bias and response option weights. *Psychol Assess* 1994; 6: 255-70.
6. Salokangas RK, Vaahtera K, Pacriev S, Sohlman B, Lehtinen V. Gender differences in depressive symptoms. An artefact caused by measurement instruments? *J Affect Disord* 2002; 68(2-3): 215-20.
7. Beck AT, Steer RA. *Beck Depression Inventory*. Manual. San Antonio, TX: Psychological Corporation; 1993.
8. Cúri M, Singer JM, Andrade DF. A model for psychiatric questionnaires with embarrassing items. *Stat Methods Med Res* 2001; 20(5): 451-70.
9. Cunha JA. *Manual da versão em português das Escalas Beck*. São Paulo: Casa do Psicólogo; 2001.
10. Andrade DF, Tavares HR, Valle RC. Teoria da Resposta ao Item: conceitos e aplicações. In: *Anais do 14º SINAPE*; 2000 jul 28; Caxambu (MG).
11. Teresi JA, Ocepek-Welikson K, Kleinman M, Cook KF, Crane PK, Gibbons LE, et al. Evaluating measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): applications (with illustrations) to measure of physical functioning ability and general distress. *Qual Life Res* 2007; 16(Suppl 1): 43-68.
12. Thissen D. Dave Thissen's Front Page. Disponível em www.unc.edu/~dthissen/dl.html. (Acessado em 26 de julho de 2008).
13. Samejima F. *Estimation of latent ability using a response pattern of graded scores*. Madison (WI): Psychometric Society; 1969.
14. Lunn DJ, Thomas A, Best N, Spiegelhalter D. Winbugs – a Bayesian modeling framework: concepts, structure, and extensibility. *Stat Comput* 2000; 10: 325-37.
15. Castro SMJ, Trentini C, Riboldi J. Teoria da resposta ao item aplicada ao Inventário de Depressão Beck. *Rev Bras Epidemiol* 2010; 13(3): 487-501.
16. Schaeffer NC. An application of item response theory to the measurement of depression. *Sociol Methodol* 1988; 18: 271-307.
17. Kessler RC, Andrews G, Colpe LJ, Hiripi E, Mroczek DK, Normand SLT, et al. Short screening scales to monitor population prevalence and trends in non-specific psychological distress. *Psychological Medicine* 2002; 32: 959-76.
18. Romans SE, Tyas J, Cohen MM, Silverstone T. Gender differences in the symptoms of major depressive disorder. *J Nerv Ment Dis* 2007; 195(11): 905-11.
19. Stommel M, Given BA, Given CW, Kalaian HA, Schulz R, McCorkle R. Gender bias in the measurement properties of the Center for Epidemiologic Studies Depression Scale (CES-D). *Psychiatry Res* 1993; 49(3): 239-50.
20. Wilhelm K, Parker G, Asghari A. Sex differences in the experience of depressed mood state over fifteen years. *Soc Psychiatry Psychiatr Epidemiol* 1998; 33(1): 16-20.
21. Carter JD, Joyce PR, Mulder RT, Luty SE, McKenzie J. Gender differences in the presentation of depressed outpatients: a comparison of descriptive variables. *J Affect Disord* 2000; 61(1-2): 59-67.

22. Gelin M, Zumbo B. Differential item functioning results may change depending on how an item is scored: an illustration with the Center for Epidemiologic Studies Depression Scale. *Educ Psychol Meas* 2003; 63: 65-74.
23. Wenzel A, Steer RA, Beck AT. Are there any gender differences in frequency of self-reported somatic symptoms of depression? *J Affect Disord* 2005; 89(1-3): 177-81.
24. Angst J, Gamma A, Gastpar M, Lepine JP, Mendlewicz J, Tylee A; Depression Research in European Society Study. Gender differences in depression. Epidemiological findings from the European DEPRES I and II studies. *Eur Arch Psychiatry Clin Neurosci* 2002; 252(5): 201-9.
25. Seidman SN, Roose SP. The sexual effects of testosterone replacement in depressed men: randomized, placebo-controlled clinical trial. *J Sex Marital Ther* 2006; 32(3): 267-73.
26. Beyer JL, Nash J, Shelton R, Loosen PT. Transtorno depressivo maior. In: Jorge MR. Manual diagnóstico e estatístico de transtornos mentais. 4ª edição. Porto Alegre: Artmed; 2000. p. 288-324.

Recebido em: 13/12/2012

Versão final apresentada em: 29/03/2013

Aceito em: 05/06/2013