

# Testes de rastreio para alterações psiquiátricas em esclerose múltipla – Revisão sistemática com metanálise

*Screening tools for psychiatric disorders in multiple sclerosis – Systematic review and metanalysis*

Maren de Moraes e Silva<sup>1</sup>, Pilar Bueno Siqueira Mercer<sup>1</sup>, Maria Carolina Zavagna Witt<sup>1</sup>, Renata Guedes Ramina Pessoa<sup>1</sup>, Camila Poletto Viveiros<sup>2</sup>, Bruno Berardi Gazola<sup>3</sup>, Diana Herchonvicz de Oliveira<sup>4</sup>, Rafaela Ianisky<sup>4</sup>

## RESUMO

**Objetivo:** Realizar uma revisão sistemática com metanálise visando evidenciar as ferramentas de rastreio psiquiátrico mais adequadas na abordagem de pacientes adultos com esclerose múltipla avaliadas por estudos de acurácia diagnóstica. **Métodos:** As bases de dados Medline, SciELO, PubMed e Lilacs foram utilizadas para pesquisa de artigos referentes ao tema proposto. Para essa busca, foram utilizados os termos “multiple sclerosis psychiatric symptoms” e “multiple sclerosis psychiatric screening”, sendo incluídos artigos na língua portuguesa e inglesa publicados entre 2007 e 2017. A qualidade dos estudos incluídos foi avaliada utilizando o método QUADAS. Uma metanálise foi conduzida com o auxílio do programa RevMan 5.3. **Resultados:** Sete artigos foram selecionados para análise. A concordância dos revisores foi calculada com um *kappa* de 0,95. A maioria dos estudos selecionados realizou avaliação de ferramentas de rastreio para depressão, tendo apenas dois deles abordado a busca de transtornos de ansiedade. O rastreio de transtorno depressivo foi realizado a partir da avaliação de 11 instrumentos diferentes, enquanto o de ansiedade, por apenas dois. A maior parte dos testes analisados apresentou boa acurácia, e a ferramenta BAI foi a única com desempenho regular na análise da curva ROC (0,77 de área sob a curva). Os demais testes apresentaram desempenho bom ou excelente, cursando com mais de 83% na avaliação de área sob a curva. **Conclusões:** As ferramentas HADS (*Hospital Anxiety and Depression Scale*), CESD (*Center for Epidemiological Scale – Depression*) e PHQ-9 (*Patient Health Questionnaire-9*) parecem ser as mais indicadas para o rastreio psiquiátrico de pacientes com esclerose múltipla. **Registro PROSPERO:** CRD42017082741.

## Palavras-chave

Esclerose múltipla, depressão, transtornos de ansiedade.

## ABSTRACT

**Objective:** To conduct a systematic review with meta-analysis aimed to highlight the most appropriate psychiatric screening tools for clinical approach of multiple sclerosis adult patients evaluated by diagnostic accuracy studies. **Methods:** Medline, SciELO, PubMed e Lilacs databases were searched using the terms “multiple sclerosis psychiatric symptoms” and “multiple sclerosis psychiatric screening” for articles related to the proposed theme.

Recebido em  
21/7/2018  
Aprovado em  
2/7/2018

1 Hospital da Cruz Vermelha Filial do Paraná, Serviço de Neurologia, Curitiba, PR, Brasil.  
2 Faculdade Evangélica do Paraná (Fepar), Curitiba, PR, Brasil.  
3 Faculdades Pequeno Príncipe (FPP), Curitiba, PR, Brasil.  
4 Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, PR, Brasil.

DOI: 10.1590/0047-2085000000213

Endereço para correspondência: Maren de Moraes e Silva  
Av Vicente Machado, 1310  
80250-090 – Curitiba, PR, Brasil  
E-mail: marenms@live.com



Articles in Portuguese and English language published between 2007 and 2017 were included. The quality of included studies was assessed using the QUADAS method. A meta-analysis was conducted using RevMan 5.3. **Results:** Seven papers were selected for analysis. The reviewers' rating were in agreement with a 0.95 kappa. Most of the selected studies performed screening tools for depression, with only two of them addressing the search for anxiety disorders. Screening for depressive disorder was performed by the evaluation of 11 different instruments, while anxiety by only two. Most of the tests presented good accuracy – the BAI tool was the only one with a regular performance in the ROC analysis (0.77 AUC). The other tests presented with good or excellent performance, presenting more than 83% in the evaluation of AUC. **Conclusions:** The HADS (Hospital Anxiety and Depression Scale), CESD (Center for Epidemiological Scale – Depression) and PHQ-9 (Patient Health Questionnaire-9) tools seem to be the most suitable for the psychiatric screening of patients with multiple sclerosis. **PROSPERO registration:** CRD42017082741.

## Keywords

Multiple sclerosis, depression, anxiety disorders.

## INTRODUÇÃO

A intensa associação entre a esclerose múltipla (EM), doença desmielinizante neurodegenerativa incurável, e o acometimento por transtornos psiquiátricos já é reconhecida de longa data, chegando a ser verificada, em algumas séries, a ocorrência de transtorno depressivo maior em até 50% desses indivíduos<sup>1</sup>. Tais morbidades são associadas a piores desfechos em saúde, gerando impactos em qualidade de vida<sup>2</sup>, prejuízos laborais e sociais<sup>3</sup>, além de uma redução da aderência ao tratamento com suas consequências em termos de acometimento físico e cognitivo<sup>4</sup>. Essa associação é devida não apenas ao impacto psicológico das limitações geradas pela doença, mas também a alterações cerebrais estruturais e imunológicas, evidenciando uma etiologia multifatorial<sup>5</sup>. Diante disso e visando melhorar a qualidade da avaliação médica, são necessários o rastreamento e o tratamento precoce de morbidades psiquiátricas no contexto de uma abordagem global em saúde, objetivando prevenir o aumento da dependência do indivíduo e reduzir a possibilidade de suicídios<sup>6</sup>.

Há uma série de ferramentas propostas para rastreamento em psiquiatria para os mais diversos tipos de pacientes, entretanto muitas delas são pouco avaliadas e outras sequer são validadas para o uso específico em EM. Esse grupo constitui uma parcela distinta entre os pacientes neurológicos, trazendo desafios no que tange a uma adequada avaliação em psiquiatria: muitos dos sintomas da doença se sobrepõem àqueles da depressão, assim, causando vieses no momento da avaliação diagnóstica. Queixas de disfunção sexual, fadiga debilitante, retardo psicomotor e alterações de sono podem ser citadas como sintomas de sobreposição entre as duas patologias<sup>7</sup>.

Levando em consideração tal importância nos desfechos em saúde, a presente revisão sistemática tem como objetivo evidenciar, na literatura médica vigente, as ferramentas de rastreamento psiquiátrico mais adequadas na abordagem de pacientes adultos com EM, avaliadas por estudos de acurácia diagnóstica.

## MÉTODOS

O presente artigo trata-se de uma revisão sistemática com metanálise, que consta registrada na base PROSPERO, com código de cadastro CRD42017082741.

### Busca na literatura médica

Em sua realização, inicialmente foi feito um breve rastreamento na literatura para identificar possíveis palavras-chave que se adequariam da melhor forma a guiar o início da revisão.

As bases de dados Medline, SciELO, PubMed e Lilacs foram utilizadas para pesquisa de artigos referentes à avaliação de acurácia diagnóstica de testes de rastreamento para alterações psiquiátricas em pacientes com EM. Para essa busca, foram utilizados os termos “*multiple sclerosis psychiatric symptoms*” e “*multiple sclerosis psychiatric screening*”. Utilizamos como critérios de inclusão artigos na língua portuguesa e inglesa publicados entre janeiro de 2007 e novembro de 2017 conduzidos em população adulta com diagnóstico de EM e que realizassem validação de ferramentas diagnósticas para transtornos psiquiátricos. Foram incluídos estudos observacionais transversais e longitudinais, retrospectivos e prospectivos, e excluídos editoriais, *guidelines*, cartas e revisões. Foram excluídos também estudos que não trouxessem uma ferramenta padrão-ouro para validação dos resultados das ferramentas propostas. A busca foi finalizada em 27 de novembro de 2017.

### Seleção de estudos

A busca e a seleção de artigos foram realizadas por cinco pesquisadores independentes, de forma cegada. Em uma primeira etapa, revisamos títulos e resumos como forma de determinar se os estudos se encaixavam nos critérios de inclusão. Concomitantemente, ocorreu uma avaliação com base na estratégia “PICO”<sup>8</sup> para um rastreamento inicial dos estudos potencialmente elegíveis. Tal ferramenta consiste em um acrônimo, que se refere aos seguintes critérios: “P” para “participantes” (pacientes adultos com EM), “I” para intervenção (instrumentos de rastreamento para alterações psiquiátricas), “C” para compa-

ração (ferramenta padrão-ouro) e “O” para “outcomes” ou, em português, “desfechos” (propriedades de acurácia). Após, revisamos o texto completo dos artigos, excluindo aqueles que não se encaixaram nos critérios de nossa pesquisa. Depois, todo o material foi revisado por todos os integrantes da equipe de pesquisa, e discordâncias foram resolvidas por consenso. Por fim, foi utilizada uma medida de concordância, *kappa*, para avaliar o nível de acordo para elegibilidade dos estudos.

## Análise dos dados

Dados referentes ao poder diagnóstico das ferramentas avaliadas pelos estudos selecionados foram extraídos dos próprios artigos e, quando não presentes em tais trabalhos, foram calculados. Todos os dados gerados foram expressos em tabelas ou gráficos. Uma metanálise foi conduzida com o auxílio do programa RevMan 5.3 para a realização de cálculos e a geração de gráficos *forest plot*. Foi calculado também um intervalo de confiança de 95% para os dados de sensibilidade e especificidade das ferramentas de rastreo abordadas. A avaliação da homogeneidade foi realizada por meio da análise visual dos gráficos, uma vez que o tradicional meio estatístico  $I^2$  não é recomendado para essa análise em sensibilidade e especificidade, por se tratarem de medidas univariadas<sup>9</sup>.

## Análise do rigor metodológico

A qualidade dos estudos incluídos foi avaliada utilizando o método QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*)<sup>10</sup>, que se presta como ferramenta específica de avaliação de trabalhos de acurácia diagnóstica. Domínios relacionados a seleção da amostra, qualidade e descrição metodológica e fluxo dos participantes nos estudos foram acessados, classificando-se os artigos selecionados de acordo com a quantidade de respostas afirmativas obtidas em referência aos 14 questionamentos avaliados na ferramenta. Não há consenso em termos de padronização da nota final pretendida, porém utilizamos a mesma classificação adotada por inúmeros outros estudos: a avaliação ficou acordada em “Excelente” para estudos com pontuação de 11 a 14, “Bom” para a pontuação entre 9 e 10, “Adequado” para a pontuação de 6 a 8, “Fraco” para a pontuação de 4 a 5 e “Inadequado” para a pontuação entre 0 e 3.

## RESULTADOS

### Seleção dos estudos

A busca na literatura identificou um total de 4.413 artigos. Dos estudos encontrados, 4.389 foram excluídos na etapa de leitura de título e resumo. Dos restantes, sete foram excluídos após aplicação dos critérios PICO e, por fim, após acesso ao texto completo, sete artigos foram selecionados para análise. Tal fluxo, justificado, pode ser visto na figura 1. A concordância dos revisores foi calculada com *kappa* de 0,95.

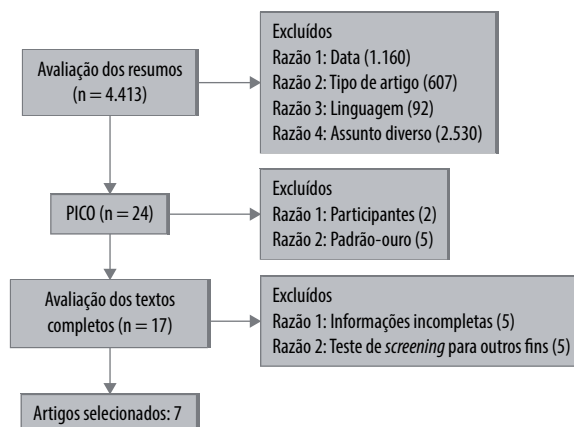


Figura 1. Fluxograma de seleção e exclusão dos estudos.

## Características dos estudos revisados

A maioria dos estudos foi publicada entre 2012 e 2015, sendo estudos recentes. Apenas um entre os selecionados foi publicado em 2009, sendo o mais antigo<sup>11</sup>. As populações utilizadas em tais registros são pertencentes à América do Norte e à Europa, não tendo sido selecionados estudos de pacientes provenientes de outros locais. As características gerais dos artigos selecionados podem ser verificadas na tabela 1.

## Ferramentas analisadas

A maioria dos estudos selecionados realizou avaliação de ferramentas de rastreo para depressão, tendo apenas dois deles abordado o rastreo de transtornos de ansiedade (Tabela 1). Não foram encontrados estudos elegíveis a respeito de outros tipos de morbidades psiquiátricas na população pretendida.

O rastreo de transtorno depressivo foi realizado a partir da avaliação de 11 instrumentos diferentes, a saber: PHQ-9 (*Patient Health Questionnaire-9*), PHQ-2 (*Patient Health Questionnaire-2*), BDI (*Beck Depression Inventory*), BDI-FS (*Beck Depression Inventory – Fast Screen*), BDI-II (*Beck Depression Inventory II*), CESD (*Center for Epidemiological Scale – Depression*), Neuro-QOL (*Neurological Quality of Life*), IDS-SR30 (*Inventory of Depressive Symptomatology – Self-report*), HADS (*Hospital Anxiety and Depression Scale*), MSDRS (*Multiple Sclerosis Depression Rating Scale*) e HDRS (*Hamilton Depression Rating Scale*). A ansiedade foi avaliada pelas seguintes duas escalas: BAI (*Beck Anxiety Inventory*) e HADS (*Hospital Anxiety and Depression Scale*). A distribuição dessas escalas e características dos estudos selecionados pode ser observada na tabela 1.

O padrão-ouro utilizado pela maioria das escalas foi o SCID (*Structured Clinical Interview for DSM-IV*), seguido pelo DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*) e o MINI (*Mini International Neuropsychiatric Interview*). Tanto o SCID quanto o MINI são entrevistas estruturadas baseadas nos critérios diagnósticos do DSM.

## Qualidade metodológica

Entre sete artigos classificados pela ferramenta de qualidade QUADAS, cinco foram avaliados como “excelentes”<sup>11-15</sup> e dois como “bons”<sup>16,17</sup>. Não houve artigos avaliados como “adequados”, “fracos” ou “inadequados”.

Todos os artigos selecionados utilizaram critérios padrão-ouro como base de comparação, os quais foram aplicados de forma homogênea a todos os participantes, além de descreverem de forma adequada todos os critérios de inclusão e apresentarem execução metodológica detalhada.

## Avaliação do poder dos testes

A maior parte dos testes analisados apresentou boa acurácia, e a ferramenta BAI foi a única com desempenho regular na análise da curva ROC (0,77 de área sob a curva). Os demais testes, apresentaram desempenho bom ou excelente, curvando com mais de 83% na avaliação de área sob a curva (Tabela 2).

O mais alto valor preditivo positivo (96%) foi obtido na avaliação da escala HDRS, no estudo de Raimo *et al.*<sup>15</sup>, enquanto o melhor valor preditivo negativo (98%) pode ser

**Tabela 1.** Características dos estudos selecionados

Autor (data)	Morbidade	País	Escalas
Amtmann <i>et al.</i> (2015)	Depressão	Estados Unidos	PHQ-9, BDI-FS, Neuro-QOL, PHQ-2, CESD
Fischer <i>et al.</i> (2015)	Depressão	Alemanha	BDI, IDS-SR30
Honarmand e Feinstein (2009)	Depressão e Ansiedade	Canadá	HADS
Patten <i>et al.</i> (2015)	Depressão	Canadá	PHQ-9, PHQ-2, CESD, HADS-D
Quaranta <i>et al.</i> (2012)	Depressão	Itália	MSDRS, BDI
Raimo <i>et al.</i> (2015)	Depressão	Itália	HDRS
Watson <i>et al.</i> (2014)	Ansiedade/Depressão	Inglaterra	BAI, BDI-II, HADS-A, HADS-D

PHQ: Patient Health Questionnaire; BDI-FS: Beck Depression Inventory – Fast Screen; Neuro-QOL: Neurological Quality of Life; CESD: Center for Epidemiologic Studies Depression Scale; BDI: Beck Depression Inventory; IDS-SR30: 30-item self-rated Inventory of Depressive Symptomatology; HADS-A: Hospital Anxiety and Depression Scale – Anxiety; HADS-D: Hospital Anxiety and Depression Scale – Depression; MSDRS: Multiple Sclerosis Depression Rating Scale; BAI: Beck Anxiety Inventory; BDI-II: Beck Depression Inventory II; HDRS: Hamilton Depression Rating Scale.

**Tabela 2.** Estudos de acurácia e resultados

Autor	Amostra	Padrão-ouro	Instrumento	ASC	Sens	Esp	VPP	VPN
<b>Depressão</b>								
Amtmann <i>et al.</i> (2015) (a)	164	SCID	PHQ-9	0,89	0,94	0,61	0,59	0,95
Amtmann <i>et al.</i> (2015) (b)	164	SCID	PHQ-2	ND	0,71	0,81	0,60	0,87
Amtmann <i>et al.</i> (2015) (c)	164	SCID	BDI-FS	0,86	0,81	0,78	0,60	0,91
Amtmann <i>et al.</i> (2015) (d)	164	SCID	CESD-20	0,89	0,94	0,69	0,55	0,97
Amtmann <i>et al.</i> (2015) (e)	164	SCID	CESD-10	0,89	0,83	0,83	0,67	0,93
Amtmann <i>et al.</i> (2015) (f)	164	SCID	Neuro-QOL	0,88	0,87	0,70	0,54	0,93
Fischer <i>et al.</i> (2015) (a)	31	MINI	IDS-SR30	0,86	0,80	0,76	0,89	0,62
Fischer <i>et al.</i> (2015) (b)	31	MINI	BDI	0,85	0,87	0,70	0,87	0,70
Honarmand e Feinstein (2009)	180	SCID	HADS-D	0,94	0,90	0,87	0,57	0,97
Patten <i>et al.</i> (2015) (a)	152	SCID	PHQ-9	0,95	0,86	0,91	0,60	0,98
Patten <i>et al.</i> (2015) (b)	152	SCID	PHQ-2**	0,94	0,80	0,93	0,64	0,97
Patten <i>et al.</i> (2015) (c)	152	SCID	HADS-D	0,92	0,85	0,82	0,42	0,97
Patten <i>et al.</i> (2015) (d)	152	SCID	CESD-20	0,94	0,95	0,73	0,34	0,99
Quaranta <i>et al.</i> (2012) (a)	94	DSM-IV	MSDRS***	0,93	0,61	0,98	0,83	0,94
Quaranta <i>et al.</i> (2012) (b)	94	DSM-IV	BDI	0,83	0,86	0,83	0,46	0,91
Raimo <i>et al.</i> (2015)	70	DSM-IV	HDRS	0,99	0,97	0,97	0,96	0,98
Watson <i>et al.</i> (2014) (a)	34	DSM-IV	HADS-D	0,88	0,77	0,67	0,59	0,82
Watson <i>et al.</i> (2014) (b)	34	DSM-IV	BDI-II	0,89	0,85	0,67	0,61	0,88
<b>Ansiedade</b>								
Honarmand e Feinstein (2009)	140	SCID	HADS-A	0,91	0,88	0,81	0,53	0,97
Watson <i>et al.</i> (2014) (c)	34	DSM-IV	BAI****	0,77	0,70	0,79	0,58	0,86
Watson <i>et al.</i> (2014) (d)	34	DSM-IV	HADS-A*****	0,94	0,90	0,71	0,56	0,94

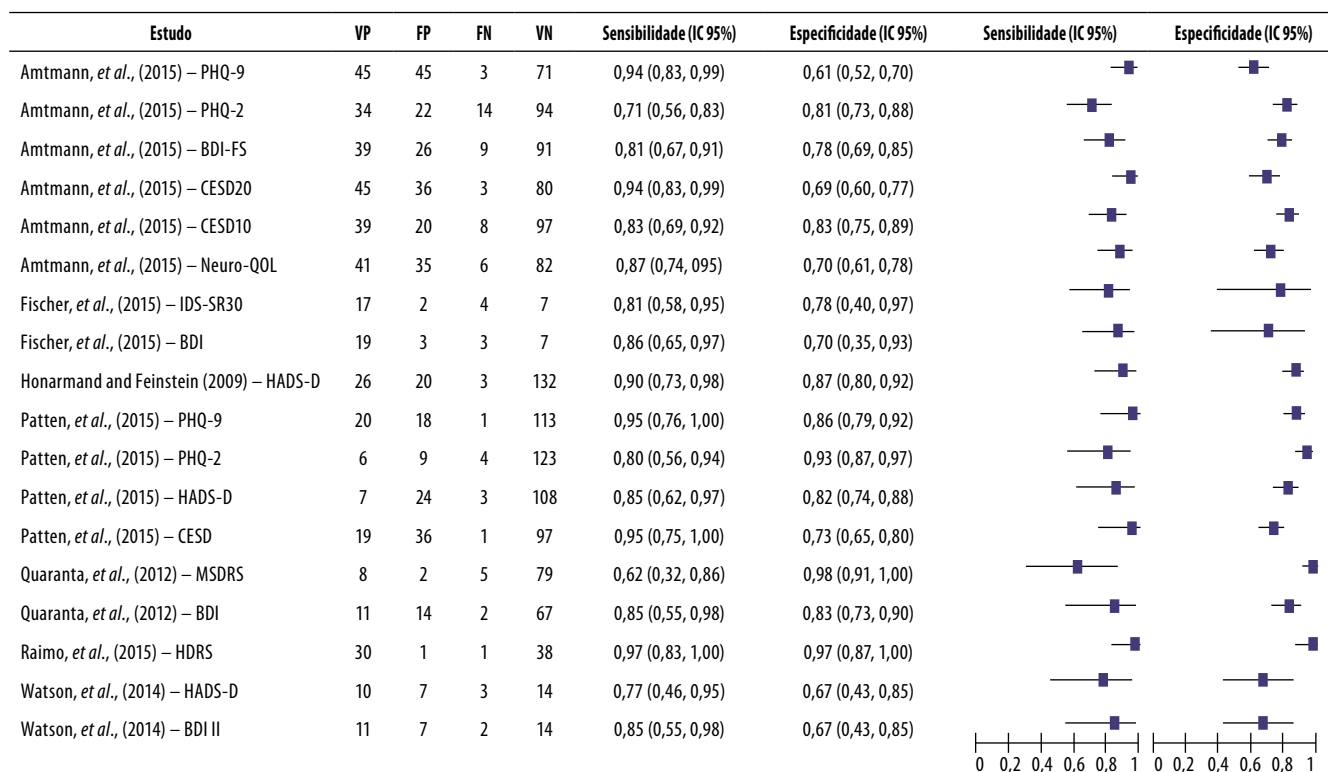
ASC: área sob a curva; Sens: sensibilidade; Esp: especificidade; VPP: valor preditivo positivo; VPN: valor preditivo negativo; SCID: Structured Clinical Interview for DSM-IV; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders; HDRS: Hamilton Depression Rating Scale; MSDRS: Multiple Sclerosis Depression Rating Scale; BDI-II: Beck Depression Inventory II; PHQ: Patient Health Questionnaire; BDI-FS: Beck Depression Inventory – Fast Screen; CESD: Center for Epidemiologic Studies Depression Scale; Neuro-QOL: Neurological Quality of Life; IDS-SR30: 30-item self-rated Inventory of Depressive Symptomatology; MINI: Mini International Neuropsychiatric Interview; HADS-D: Hospital Anxiety and Depression Scale – Depression; HADS-A: Hospital Anxiety and Depression Scale – Anxiety; BAI: Beck Anxiety Inventory; ND: não disponível.

verificado na avaliação das escalas PHQ-9, por Patten *et al.*<sup>13</sup>, e HDRS, por Raimo *et al.*<sup>15</sup>. Este último apresentou também desempenho excelente em todos os demais quesitos. Esses e outros valores de sensibilidade e especificidade podem ser vistos de maneira comparativa na figura 2.

Poucas opções de estudos se prestam a avaliar o rastreamento de ansiedade<sup>11,17</sup>. Entre eles, pode-se observar homogeneidade (Figura 3) no desempenho da escala HADS, subescala para ansiedade, apresentando *performance* melhor que o BAI. A sensibilidade de 70% dela não é satisfatória, sequer sua especificidade (79%).

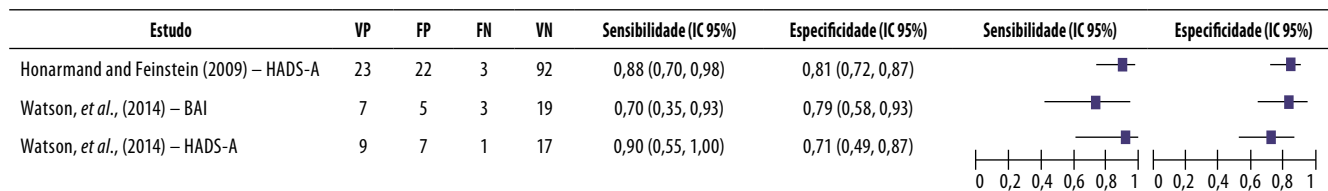
Todas as variantes da escala BDI cursaram com boa sensibilidade e especificidade não muito atrás, sendo a escala BDI-II, avaliada por Watson *et al.*<sup>17</sup>, a com pior desempenho. Em uma análise conjunta da escala BDI por meio do gráfico *forest plot*, é possível perceber homogeneidade nos resultados (Figura 4), com apenas a especificidade apresentada por Fisher *et al.*<sup>16</sup> fugindo do esperado.

Em relação à ferramenta PHQ-9, foi verificada alta sensibilidade, concordante entre os estudos, ante especificidade algo heterogênea (Figura 5). As notas de corte foram avaliadas entre tradicional e ideal, mantendo-se o padrão gráfico.



**Figura 2.** Sensibilidade e especificidade das ferramentas para depressão.

VP: verdadeiros-positivos; FP: falsos-positivos; FN: falsos-negativos; VN: verdadeiros-negativos; IC: intervalo de confiança; PHQ: Patient Health Questionnaire; BDI-FS: Beck Depression Inventory – Fast Screen; CESD: Center for Epidemiologic Studies Depression Scale; Neuro-QOL: Neurological Quality of Life; IDS-SR30: 30-item self-rated Inventory of Depressive Symptomatology; BDI: Beck Depression Inventory; HADS-D: Hospital Anxiety and Depression Scale – Depression; MSDRS: Multiple Sclerosis Depression Rating Scale; HDRS: Hamilton Depression Rating Scale; BDI-II: Beck Depression Inventory II.



**Figura 3.** Sensibilidade e especificidade das ferramentas para ansiedade.

VP: verdadeiros-positivos; FP: falsos-positivos; FN: falsos-negativos; VN: verdadeiros-negativos; IC: intervalo de confiança; HADS-A: Hospital Anxiety and Depression Scale – Anxiety; BAI: Beck Anxiety Inventory.

É possível perceber desempenho homogêneo também entre os estudos que avaliam a escala HADS (Figura 6), com desempenho ligeiramente pior no estudo de Watson *et al.*<sup>17</sup>. Apresentou boa aplicabilidade tanto no quesito depressão quanto ansiedade.

Já o CESD apresentou excelente desempenho em termos de sensibilidade nos dois estudos nos quais foi avaliado, com *performance* bastante homogênea entre eles (Figura 7). Alta sensibilidade, com especificidade regular. As notas de corte foram ajustadas para o ideal à população de cada estudo.

#### BDI

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Fischer, <i>et al.</i> , (2015)	19	3	3	7	0,86 (0,65, 0,97)	0,70 (0,35, 0,93)		
Quaranta, <i>et al.</i> , (2012)	11	14	2	67	0,85 (0,55, 0,98)	0,83 (0,73, 0,90)		
Valor sumário	30	17	5	74	0,86 (0,70, 0,95)	0,81 (0,72, 0,89)		

#### BDI-II

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Watson, <i>et al.</i> , (2014)	11	7	2	14	0,85 (0,55, 0,98)	0,67 (0,43, 0,85)		

#### BDI-FS

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Amtmann, <i>et al.</i> , (2015)	39	26	9	91	0,81 (0,67, 0,91)	0,78 (0,69, 0,85)		

**Figura 4.** Sumário ferramenta BDI (*Beck Depression Inventory*).

BDI: *Beck Depression Inventory*; VP: verdadeiros-positivos; FP: falsos-positivos; FN: falsos-negativos; VN: verdadeiros-negativos; Sens: sensibilidade; Espec: especificidade; IC: intervalo de confiança; BDI-II: *Beck Depression Inventory II*; BDI-FS: *Beck Depression Inventory – Fast Screen*.

#### PHQ-9

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Amtmann, <i>et al.</i> , (2015)	45	45	3	71	0,94 (0,83, 0,99)	0,61 (0,52, 0,70)		
Patten, <i>et al.</i> , (2015)	20	18	1	113	0,95 (0,76, 1,00)	0,86 (0,79, 0,92)		
Valor sumário	65	63	4	184	0,94 (0,86, 0,98)	0,74 (0,69, 0,80)		

**Figura 5.** Sumário ferramenta PHQ-9 (*Patient Health Questionnaire 9*).

PHQ-9: *Patient Health Questionnaire 9*; VP: verdadeiros-positivos; FP: falsos-positivos; FN: falsos-negativos; VN: verdadeiros-negativos; Sens: sensibilidade; Espec: especificidade; IC: intervalo de confiança.

#### HADS-D

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Honarmand and Feinsten (2009)	26	20	3	132	0,90 (0,73, 0,98)	0,87 (0,80, 0,92)		
Patten, <i>et al.</i> , (2015)	18	24	3	107	0,86 (0,64, 0,97)	0,82 (0,74, 0,88)		
Watson, <i>et al.</i> , (2014)	10	7	3	14	0,77 (0,46, 0,95)	0,67 (0,43, 0,85)		
Valor sumário	54	51	9	253	0,86 (0,75, 0,93)	0,83 (0,79, 0,87)		

#### HADS-A

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Honarmand and Feinsten (2009)	23	22	3	92	0,88 (0,70, 0,98)	0,81 (0,72, 0,87)		
Watson, <i>et al.</i> , (2014)	9	7	1	17	0,90 (0,55, 1,00)	0,71 (0,49, 0,87)		
Valor sumário	32	29	4	109	0,89 (0,74, 0,97)	0,79 (0,71, 0,85)		

**Figura 6.** Sumário ferramenta HADS (*Hospital Anxiety and Depression Scale*).

HADS-D: *Hospital Anxiety and Depression Scale – Depression*; VP: verdadeiros-positivos; FP: falsos-positivos; FN: falsos-negativos; VN: verdadeiros-negativos; Sens: sensibilidade; Espec: especificidade; IC: intervalo de confiança; HADS-A: *Hospital Anxiety and Depression Scale – Anxiety*.

## DISCUSSÃO

### Depressão

A maioria dos artigos que abordam o rastreo de transtornos psiquiátricos se presta a buscar apenas sintomas depressivos, outros poucos procuram verificar sintomas de ansiedade e nenhum dos selecionados para esta revisão teve como objetivo a busca de outras morbidades psiquiátricas, tais como psicoses.

O foco principal dessas ferramentas é realizar um rastreo inicial, sendo o paciente, então, encaminhado para um diagnóstico especializado. Tais ferramentas, para serem consideradas ideais, devem ter alta confiabilidade e validade na distinção entre casos e não casos de uma condição, traduzidas por alta sensibilidade e acurácia, com o menor número possível de falsos-negativos<sup>18</sup>. Levando em consideração este último quesito, tanto PHQ-9, quanto CESD tiveram bom desempenho em pacientes com EM, apresentando alta sensibilidade (Tabela 2). O Questionário de Saúde do Paciente-9 (PHQ-9) é validado para língua portuguesa e está disponível gratuitamente para a utilização – em sua avaliação, entrevistados são solicitados a avaliar com que frequência apresentaram cada sintoma depressivo nas últimas duas semanas em uma escala de quatro pontos<sup>19</sup> e nove itens, embasados nos critérios do DSM-IV<sup>20</sup>. Há também uma variante, o PHQ-2 (Questionário de Saúde do Paciente-2), desenvolvido objetivando um rastreo ainda mais rápido<sup>21</sup> e consistindo nas duas primeiras questões do PHQ-9. Nos dois estudos que o avaliaram, a escala foi aplicada juntamente com o PHQ-9<sup>12,13</sup> e, por serem variáveis binárias, não se pôde comparar a curva ROC. O PHQ-2, por si só, já não apresenta sensibilidade tão adequada, porém mostra-se interessante em termos de especificidade e alto valor de verdadeiros-negativos. Além disso, é importante que seja ressaltado o fato de que os resultados de PHQ-9 foram analisados com ponto de corte tradicional de 10, valor esse que, pelas observações de Amtmann *et al.*<sup>12</sup>, deveria ser ajustado para 12, visando à melhor acurácia em paciente com EM. Diante dos achados de Patten *et al.*<sup>13</sup>, esse corte parece, de fato, melhorar os parâmetros, porém a heterogeneidade (Figura 5) em termos de especificidade entre ambos os estudos fica mantida independentemente do ponto de corte escolhido, o que pode ser explicado pela

diferença no padrão das populações estudadas em termos de EDSS (*Expanded Disability Status Scale*), e isso pode levar a um fator de confusão por sintomas sobrepostos em pacientes mais debilitados. Uma baixa especificidade pode subestimar o número de pacientes acometidos por depressão, sobremaneira.

Em relação ao BDI (*Beck Depression Inventory*), há a possibilidade de verificar essa escala em outras duas versões além da original: uma versão de aplicabilidade rápida, denominada BDI-FS (*Beck Depression Inventory – Fast Screen*), e a versão revisada em 1996<sup>22</sup>, BDI-II (*Beck Depression Inventory II*). Essa última, apesar de ter sido revisada e modificada com base na original e ser validada para língua portuguesa<sup>23</sup>, ainda é pouco usada em trabalhos de acurácia nessa população, tendo apenas a BDI original estudos para comparação (Figura 4). Assim como o PHQ-9, o BDI também avalia sintomas nas últimas duas semanas<sup>24</sup>, no entanto sua versão de avaliação rápida (BDI-FS) não inclui sintomas somáticos, o que, de certa forma, traz vantagens na avaliação de paciente com EM, eliminando alguns dos possíveis vieses de sobreposição de sintomas<sup>25</sup>. Todas as versões citadas cursaram com boa sensibilidade e especificidade mediana, sendo a do BDI-II avaliada por Watson *et al.* a de piora acurácia diagnóstica<sup>17</sup>. Entre as avaliações de BDI, os resultados foram homogêneos (Figura 4). Em relação ao BDI-II, é sugerido, inclusive, ajuste nos pontos de corte, elevando-os para melhor desempenho nessa população de pacientes<sup>26</sup>. Isso pode se dever ao fato de, com a progressão das limitações da EM, mais sintomas se sobrepõem ao diagnóstico de depressão, necessitando de mudança nos valores de corte para melhor resultado. Um possível empecilho, em termos de padrão de população brasileira, pode ser o fato de o BDI ter acurácia influenciada pelo nível educacional – fato que pode ser relevante, também, em todas as outras escalas autoadministráveis aqui citadas<sup>14</sup>.

Alternativa a ser considerada é a ferramenta CESD, também validada para falantes da língua portuguesa. Foi obtido excelente desempenho em termos de sensibilidade em ambos os estudos que com ela trabalharam<sup>12,13</sup>, sendo verificada uma distribuição bastante homogênea se utilizado o ponto de corte tradicional para a avaliação de Patten *et al.*<sup>13</sup> e outra, quatro pontos acima para Amtmann *et al.*<sup>12</sup> (Figura 7), apresentando alta sensibilidade e especificidade

#### CESD

Estudo	VP	FP	FN	VN	Sensibilidade (IC 95%)	Especificidade (IC 95%)	Sensibilidade (IC 95%)	Especificidade (IC 95%)
Amtmann, <i>et al.</i> , (2015)	45	36	3	80	0,94 (0,83, 0,99)	0,69 (0,60, 0,77)	■	■
Patten, <i>et al.</i> , (2015)	19	36	1	97	0,95 (0,75, 1,00)	0,73 (0,65, 0,80)	■	■
Valor sumário	64	72	4	177	0,94 (0,86, 0,98)	0,71 (0,65, 0,77)	■	■

**Figura 7.** Sumário ferramenta CESD (*Center for Epidemiologic Studies Depression Scale*).

CESD: *Center for Epidemiologic Studies Depression Scale*; VP: verdadeiros-positivos; FP: falsos-positivos; FN: falsos-negativos; VN: verdadeiros-negativos; Sens: sensibilidade; Espec: especificidade; IC: intervalo de confiança.

regular, evidenciando mais falsos-positivos que verdadeiros-positivos. Entretanto, pensando na utilização apenas como rastreio, tal fato pode ser de alguma maneira relevante, ainda diante da baixa frequência de falsos-negativos – exigindo-se, logicamente, a aplicação de avaliação diagnóstica posterior. A discrepância entre os pontos de corte se deve, possivelmente, a diferenças no padrão de população, sendo uma com mais gravidade que a comparada. Essa ferramenta apresenta duas possibilidades de aplicação: CESD-20 e CESD-10. A versão completa do CESD possui 20 itens e os entrevistados são solicitados a avaliarem a frequência de sentimentos positivos e negativos apresentados na semana prévia à avaliação<sup>27</sup>, enquanto o CESD-10 inclui metade dos itens da escala original<sup>28</sup>. Apenas o estudo de Amtmam *et al.* avaliou a versão enxuta da escala, a qual foi aplicada com a completa<sup>12</sup>. É proposto, também por Amtmann *et al.*<sup>12</sup>, aumento do ponto de corte ideal, com aumento da especificidade do teste, mantendo a homogeneidade com o estudo de Patten *et al.*<sup>13</sup> (Figura 7). No entanto, a aplicação do CESD encontra prejuízo em termos de ser uma escala longa, que exige disponibilidade de tempo<sup>13</sup>.

A escala HADS, validada em português<sup>29</sup>, aparenta ter bom desempenho, a despeito do alto número de falsos-negativos, sendo a avaliação homogênea entre os estudos avaliados (Figura 6). Tal escala tem aplicabilidade tanto para ansiedade como depressão, apresentando duas divisões, o que se apresenta como vantagem em termos de praticidade. A primeira será citada em momento propício. Parece cumprir sua proposta de ferramenta de rastreio, com frequência baixa de falsos-negativos. Validada e gratuita em português, a escala respeita a visão de que a avaliação de transtornos psiquiátricos pode encontrar dificuldades em relação a pacientes neurológicos e, portanto, não segue à risca os critérios do DSM-IV<sup>30</sup>, sendo desenhada para evitar confundidores e sintomas sobrepostos<sup>24</sup>. Contudo, um de seus itens pode gerar confusão em termos de sobreposição: o que cursa a respeito de lentificação, que pode gerar dúvidas em pacientes com EM no momento do preenchimento, referindo sintomas não psicológicos.

Há também outras escalas, menos estudadas na população de EM, que não tiveram aqui critérios de comparação, tais como IDS-SR30, HDRS, MSDRS e NeuroQOL. O IDS-SR30, meio novo de avaliação para depressão, apesar de não específico para EM, apresentou parâmetros ótimos quando da avaliação da acurácia (Tabela 2), podendo ser uma ferramenta promissora na avaliação desses pacientes e necessitando, portanto, de novos estudos para corroborar esse desempenho. Apresenta características dignas de nota, como cobrir os critérios do DSM-IV e oferecer versões paralelas paciente-avaliador<sup>16</sup>. No entanto, ainda não é disponível para o público brasileiro. Outra opção com bom desempenho é o HDRS (*Hamilton Depression Rating Scale*), que possui validação para português. Essa é uma das ferramentas semiestruturadas

mais frequentemente utilizadas para verificação de severidade e modificação dos sintomas depressivos ao longo da evolução clínica<sup>15</sup>, porém ainda pouco explorada para EM em específico. De acordo com achados de Raimo *et al.*<sup>15</sup> nessa população, a escala apresentou excelentes valores de acurácia diagnóstica (Tabela 2).

Já uma alternativa mais específica é o MSDRS, escala desenvolvida por Quaranta *et al.*<sup>14</sup> especialmente para pacientes com EM, apresentando a maior especificidade entre os testes, ante menor sensibilidade. Visto que objetiva rastreio, é de suma importância que o teste seja sensível. Diante disso, talvez sejam necessários mais estudos em relação a algum papel dessa ferramenta no acompanhamento do transtorno já diagnosticado. Em termos de ferramentas específicas para pacientes neurológicos, pode ser citada a NeuroQOL<sup>31</sup>, também pouco avaliada e não validada para falantes da língua portuguesa, porém com aparente bom desempenho em termos de acurácia, necessitando de maiores estudos para EM.

## Transtornos de ansiedade

O número de publicações envolvendo rastreio de sintomas ansiosos em EM é reduzido, a despeito da alta frequência dessa morbidade nesses pacientes<sup>32</sup>.

Foi verificado que a subescala para ansiedade da ferramenta HADS teve melhor desempenho que a BAI (Figura 3). A sensibilidade de 70% da BAI não é muito satisfatória, o teste tampouco compensa em especificidade (0,79), e foi, ainda, único com acurácia regular (0,77) entre todos os testes avaliados nesta revisão, os quais apresentaram desempenho de área sob a curva maior que 83% (Tabela 2). Parece superestimar a presença de transtorno de ansiedade nesse extrato de pacientes, uma vez que foca excessivamente em sintomas físicos<sup>17</sup>. Além disso, essa escala ainda não é validada para o português brasileiro, o que limita sua aplicabilidade em nosso país. Diante desses achados, não parece ser a ferramenta mais indicada para a avaliação em EM.

Já em avaliação da ferramenta HADS, foi verificada homogeneidade entre os estudos relatados para ansiedade (Figura 6), com boa sensibilidade. É importante ter em mente que até dois terços dos pacientes com transtornos de ansiedade possuem de forma concomitante depressão maior<sup>32</sup>, o que torna o HADS uma opção adequada para a prática clínica, por sua versatilidade.

## Limitações e contribuições

Esta revisão se limita no que diz respeito ao risco que se deve, primariamente, a um viés na seleção, inerente a revisões sistemáticas, devido à restrição de idiomas e datas, com risco de perda de estudos elegíveis. Entretanto, tem força no sentido de a qualidade dos estudos utilizados ter sido avaliada com base em protocolos reconhecidos<sup>10</sup>, em termos de trazer contribuições de boa evidência na avaliação de pacientes com EM.



## CONCLUSÃO

Com base na literatura atual, as ferramentas HADS, CESD e PHQ-9 parecem ser as mais indicadas para o rastreo psiquiátrico de pacientes com esclerose múltipla, sendo indicada avaliação psiquiátrica especializada posterior para confirmação diagnóstica. A escala HDRS parece ter resultado possivelmente promissor, necessitando de novas avaliações, assim como ferramentas mais específicas no manejo psiquiátrico de pacientes com esclerose múltipla, as quais ainda possuem estudos escassos, em especial aquelas desenhadas para avaliação exclusiva de transtorno de ansiedade.

## CONTRIBUIÇÕES INDIVIDUAIS

**Maren de Moraes e Silva** – Seleção dos artigos, análise crítica, estatística e redação do manuscrito.

**Pilar Bueno Siqueira Mercer** – Revisão do manuscrito.

**Maria Carolina Zavagna Witt** – Revisão do manuscrito.

**Renata Guedes Ramina Pessoa** – Revisão do manuscrito.

**Camila Poletto Viveiros** – Busca e seleção dos artigos, redação.

**Bruno Berardi Gazola** – Busca e seleção dos artigos, redação.

**Diana Herchovicz de Oliveira** – Busca e seleção dos artigos, redação.

**Rafaela Ianisky** – Busca e seleção dos artigos, redação.

## CONFLITOS DE INTERESSE

Declaramos não ter conflitos de interesse na publicação deste artigo.

## REFERÊNCIAS

- Patten SB, Beck CA, Williams JV, Barbui C, Metz LM. Major depression in multiple sclerosis: a population-based perspective. *Neurology*. 2003;61(11):1524-7.
- D'Alisa S, Miscio G, Baudo S, Simone A, Tesio L, Mauro A. Depression is the main determinant of quality of life in multiple sclerosis: a classification-regression (CART) study. *Disabil Rehabil*. 2006;28(5):307-14.
- Patten SB, Williams JV, Lavorato DH, Koch M, Metz LM. Depression as a predictor of occupational transition in a multiple sclerosis cohort. *Funct Neurol*. 2013;28(4):275-80.
- Tarrant M, Oleen-Burkey M, Castelli-Haley J, Lage MJ. The impact of comorbid depression on adherence to therapy for multiple sclerosis. *Mult Scler Int*. 2011;20(11):2713-21.
- Feinstein A, Magalhaes S, Richard JF, Audet B, Moore C. The link between multiple sclerosis and depression. *Nat Rev Neurol*. 2014;10(9):507-17.
- Feinstein A. Multiple sclerosis, depression, and suicide. *BMJ*. 1997;315:691-2.
- Goodkin DE. The natural history of multiple sclerosis. In: Rudick RA, Goodkin GE (Eds.). *Treatment of multiple sclerosis: trial design, results, and future perspectives*. New York: Springer-Verlag; 1992. p. 17-46.
- Santos CMC, Pimenta CAM, Nobre MRC. A estratégia PICO para a construção da pergunta de pesquisa e busca de evidências. *Rev Latino-Am Enfermagem*. 2007 ;15(3):508-11.
- Takwoingi Y, Riley RD, Deeks JJ. Meta-analysis of diagnostic accuracy studies in mental health. 2015;18(4):103-9.
- Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
- Honarmand K, Feinstein A. Validation of the Hospital Anxiety and Depression Scale for use with multiple sclerosis patients. *Mult Scler*. 2009;15(12):1518-24.
- Amtmann D, Bamer AM, Johnson KL, Ehde DM, Beier ML, Elzea JL, et al. A comparison of multiple patient reported outcome measures in identifying major depressive disorder in people with multiple sclerosis. *J Psychosom Res*. 2015;79(6):550-7.
- Patten SB, Burton JM, Fiest KM, Wiebe S, Bulloch AG, Koch M, et al. Validity of four screening scales for major depression in MS. *Mult Scler*. 2015;21(8):1064-71.
- Quaranta D, Marra C, Zinno M, Patanella AK, Messina MJ, Piccinini C, et al. Presentation and validation of the multiple sclerosis depression rating scale: a test specifically devised to investigate affective disorders in multiple sclerosis patients. *Clin Neuropsychol*. 2012;26(4):571-87.
- Raimo S, Trojano L, Spitaleri D, Petretta V, Grossi D, Santangelo G. Psychometric properties of the Hamilton Depression Rating Scale in multiple sclerosis. *Qual Life Res*. 2015;24(8):1973-80.
- Fischer A, Fischer M, Nicholls RA, Lau S, Poettgen J, Patas K, et al. Diagnostic accuracy for major depression in multiple sclerosis using self-report questionnaires. *Brain Behav*. 2015;5(9):e00365.
- Watson TM, Ford E, Worthington E, Lincoln NB. Validation of mood measures for people with multiple sclerosis. *Int J MS Care*. 2014;16(2):105-9.
- Zhu W, Zeng N, Wang N. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations. *Northeast SAS User Group proceedings, Section of Health Care and Life Sciences, Baltimore, Maryland, 14-17 November 2010, 1-9*.
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-13.
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Text Revision. Washington, DC: American Psychiatric Association; 2000.
- Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care*. 2003;41(11):1284-92.
- Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess*. 1996;67(3):588-97.
- Gomes-Oliveira MH, Gorenstein C, Lotufo Neto F, Andrade LH, Wang YP. Validation of the Brazilian Portuguese version of the Beck Depression Inventory-II in a community sample. *Rev Bras Psiquiatr*. 2012;34(4):389-94.
- Beck AT, Steer RA, Brown GK. *BDI – Fast Screen for Medical Patients: Manual*. San Antonio, TX: Pearson; 2000.
- Benedict RH, Fishman I, McClellan MM, Bakshi R, Weinstock-Guttman B. Validity of the Beck Depression Inventory-Fast Screen in multiple sclerosis. *Mult Scler*. 2003;9(4):393-6.
- Sullivan MJ, Weinschenker B, Mikail S, Bishop SR. Screening for major depression in the early stages of multiple sclerosis. *Can J Neurol Sci*. 1995;22(3):228-31.
- Radloff LS. The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas*. 1977;1(3):385-401.
- Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med*. 1994;10(2):77-84.
- Botega NJ, Bio MR, Zomignani MA, Garcia Jr. C, Pereira WAB. Transtornos do humor em enfermagem de clínica médica e validação de escala de medida (HAD) de ansiedade e depressão. *Rev Saúde Pública*. 1995;29(5):359-63.
- Minden SL, Schiffer RB. Affective disorders in multiple sclerosis. Review and recommendations for clinical research. *Arch Neurol*. 1990;47(1):98-104.
- Gershon RC, Lai JS, Bode R, Choi S, Moy C, Bleck T, et al. Neuro-QOL: quality of life item banks for adults with neurological disorders: item development and calibrations based upon clinical and general population testing. *Qual Life Res*. 2012;21(3):475-86.
- Korostil M, Feinstein A. Anxiety disorders and their clinical correlates in multiple sclerosis patients. *Mult Scler*. 2007;13(1):67-72.