

Diagnosis Accuracy of Baptista Depression Scale: Adult and Screening Versions

André Pereira Gonçalves* , Makilim Nunes Baptista , Anna Elisa de Villemor-Amaral , & Lucas de Francisco Carvalho 

Universidade São Francisco, Campinas, SP, Brasil

ABSTRACT – This study aimed to investigate the diagnostic accuracy of the Baptista Depression Scale adult and screening versions (EBADEP-A and EBADEP-screening) for discrimination of people with a major depressive episode (MDE) or major depressive disorder (MDD) and to compare discriminative capacities. Participants were 187 people, 52 patients, and 135 non-clinical individuals. Results indicated sensitivity equal to 0.92 and specificity equal to 0.88 for EBADEP-A (cut-off = 54), and 0.96 and 0.82, respectively, for EBADEP-screening (cut-off = 17). Findings suggest equivalence between the two versions of EBADEP to discriminate people with MDE or MDD. Results indicate that the two versions can provide a correct diagnostic indicative for MDE and MDD.

KEYWORDS: psychopathology, mental health, psychometric properties, cutoff

Acurácia Diagnóstica da Escala Baptista de Depressão: Versões Adulto e Triagem

RESUMO – O objetivo da presente investigação foi verificar a acurácia diagnóstica da Escala Baptista de Depressão adulto e triagem (EBADEP-A e EBADEP- triagem) para discriminação de pessoas com episódio depressivo maior (EDM) e/ou transtorno depressivo maior (TDM), e comparar suas capacidades discriminativas. Participaram 187 pessoas, 52 pacientes e 135 do grupo não clínico. Os resultados indicaram sensibilidade igual a 0,92 e especificidade igual a 0,88 para a EBADEP-A (ponto de corte igual a 54), e 0,96 e 0,82, respectivamente, para a EBADEP-triagem (ponto de corte igual a 17). Os achados sugerem equivalência entre as duas versões da EBADEP para discriminação de pessoas com EDM e/ou TDM. Os resultados indicam que as duas versões podem fornecer um indicativo diagnóstico correto para EDM e TDM.

PALAVRAS-CHAVE: psicopatologia, saúde mental, propriedades psicométricas, ponto de corte

Depressive disorders present the highest prevalence compared to other psychiatric disorders. According to the World Health Organization (WHO, 2017), 300 million people suffer from a depressive episode worldwide, representing 5% of the population. Besides, depressive disorders are related to disability and withdrawal from professional activities, representing about 12% of total leave from work, impacting the individual's life and the organizational environment.

The etiology of depressive disorders is still complex to identify, arising from several factors, biological (e.g., genetics, hormonal dysregulation), psychological (e.g., self-

criticism, personality traits), and social (e.g., relationships, social acceptance; Fried & Neste, 2015). The combination of these factors is responsible for the onset, development, and remission of depressive symptoms (Durisko et al., 2015; Fried & Neste, 2015; Lichtenberg & Belmaker, 2010).

Within the category of depressive disorders, the Major Depressive Episode (MDE) represents the most classic condition of this group, and the Major Depressive Disorder (MDD). MDD can reach 16.2% of people worldwide (Kessler et al., 2003). For MDD configuration, recurrent depressive episodes must occur, that is, at least two depressive episodes. Depressive episodes correspond to the

* E-mail: andregoncalvespsi@gmail.com

■ Submetido: 02/04/2019; Revisado: 06/09/2019; Aceito: 08/12/2020.

presence of diagnostic criteria for MDE and/or MDD for at least two weeks. In the case of a first or single episode, MDE is characterized (APA, 2014).

Detection of MDE and MDD can be conducted following the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA, 2014) and the International Classification of Diseases and Health-Related Problems (ICD-10; WHO, 1996). In DSM-5, the individual must present at least one of the two symptoms considered central, namely, anhedonia or depressed mood. It is also necessary to manifest at least four more symptoms, which may be: loss or marked gain in weight, insomnia or hypersomnia, fatigue, agitation or psychomotor retardation, feeling of uselessness or excessive or inadequate guilt, decreased ability to think or concentrate and recurring thoughts of death. Besides, symptoms must be present for at least two intermittent weeks (APA, 2014).

In addition to the criteria reported in the manuals for the diagnosis of MDE and MDD, symptoms perceived by clinicians are presented by literature as relevant for the correct diagnosis of MDE and MDD (Goldberg, 2011; Lichtenberg & Belmaker, 2010). Specifically, the diagnostic criteria do not include some symptoms observed in clinical practice, like anguish, anxiety, social instability, irritability, hopelessness, and hypochondria (Beard et al., 2016; Beck et al., 1997; Fried et al., 2016). These characteristics make MDE and MDD heterogeneous diagnostic categories (Goldberg, 2011).

Some techniques and tools are employed to assist the professional in diagnosing MDE and MDD, such as interviews, observations, and scales. Concerning scales, these are generally based on the manuals' diagnostic criteria for the development of items. However, some scales have also been designed considering the symptoms observed by clinicians, trying to encompass all the characteristics that configure MDE and MDD, and assist in the correct diagnosis of these conditions (Baptista, 2012; Beck et al., 1997; Fried, 2016; Fried et al., 2016; Lichtenberg & Belmaker 2010).

The verification of a scale's ability to assist in diagnosing a pathology is typically conducted in diagnostic accuracy studies. Diagnostic accuracy concerns the capability that an instrument, technique, or methodology has to discriminate groups, that is, how much an assessment tool is capable of identifying people with the condition and also discarding diagnosis for people without the condition (Bossuyt & Leeflang, 2008; Faraone & Tsuang, 1994; Parshall, 2013). These studies have clinical contributions, as they provide information that grant professionals greater security in decision making (Glasser, 2014).

The main diagnostic accuracy indicators are sensitivity, specificity, positive predictive value, negative predictive value, probability rates, and effectiveness test. Sensitivity refers to how well an instrument can correctly select a person with the pathology; specificity concerns how much a tool can correctly select a person who does not have the pathology; the positive predictive value refers to the probability that a

positive result will be truly positive based on the prevalence of the investigated pathology; the negative predictive value refers to the likelihood of a negative result being truly negative based on the prevalence of the investigated pathology; the probability rates correspond to the probability of a person actually having the pathology; effectiveness rate refers to the instrument's total ability to discriminate between individuals with pathology and individuals without pathology. Through these indicators, one can conclude about the instruments' diagnostic capacity (Glasser, 2014; Lalkhen & McCluskey, 2008; Van Stralen et al., 2009).

Studies verified the diagnostic accuracy of instruments that assess depressive symptoms. For instance, Santos et al. (2013), in a study using the Patient Health Questionnaire-9 (PHQ-9), found a sensitivity of 77.5% and specificity of 86.7%, in a sample of 447 participants from the general population, 40 of which were diagnosed with MDD by the Mini International Neuropsychiatric Interview (MINI). Another study, using the Beck Depression Inventory-II (BDI-II), found sensitivity equal to 70% and specificity equal to 87% with a sample of 242 individuals from the general population from which possible individuals diagnosed with depression were extracted through the Structured Clinical Interview for the DSM-IV (Gomes-Oliveira et al., 2012).

In Brazil, the Baptista Depression Scale (EBADEP-A; Baptista, 2012) was created to track depressive symptoms in adults. EBADEP-A is a self-report scale composed of 45 items in pairs of antagonistic sentences at each end of the item to be answered in a four-point Likert format. Studies carried out using EBADEP-A observed evidence for its use and good psychometric properties (Baptista, 2012; Baptista et al., 2012; Bighetti et al., 2014; Carneiro & Baptista, 2012; Souza et al., 2015). In addition to EBADEP-A, a screening version was developed, with 15 items on a 4-point Likert scale (Baptista & Carvalho, 2018). Shorter scales can provide benefits, such as lower spending and faster assessments, and are useful tools mainly in public health services and primary care (Bolsoni & Zuardi, 2015).

Two studies were conducted aiming to verify the diagnostic accuracy of EBADEP-A. The first study encompassed 1,676 individuals, observing sensitivity equal to 77.5% and specificity equal to 87.5 (Baptista, 2012); and another study with 40 depressive patients (MDE or MDD) and 40 people from the general population and without a diagnosis, observing sensitivity equal to 97.5% and specificity equal to 90.5% (Baptista & Carvalho, 2018). In this same study (Baptista & Carvalho, 2018), the short version of EBADEP-A (EBADEP-screening) was also used, which obtained 95% sensitivity and 87.5% specificity. Screening tests should have greater sensitivity in comparison to specificity, making it possible to identify all individuals with the condition (Germans et al., 2012).

Despite satisfactory findings with EBADEP-A and EBADEP-screening, studies aiming to verify these scales'

diagnostic accuracy are still scarce. Considering that MDE and MDD have a high prevalence in the population, that the correct diagnosis can improve a patient's prognosis, and that EBADEP-A and EBADEP-screening can be useful tools to diagnose these psychopathologies, our study aimed to verify the diagnostic accuracy of EBADEP-A and EBADEP-screening in the discrimination of people with MDD and MDE. Besides, we sought to verify which of the two versions of EBADEP is the most discriminative. In order to achieve the proposed objectives, we replicated the procedures used by Baptista and Carvalho (2018), although

in a larger sample. We created four hypotheses: H1) the clinical group should present higher means in EBADEP-A and EBADEP-screening (Baptista, 2012); H2) EBADEP-A and EBADEP-screening should show an area under the curve (ROC curve) > 0.80, considered good (Bewick et al., 2004; Baptista & Carvalho, 2018 Hosmer & Lemeshow, 2000); H3) EBADEP-A and EBADEP-screening should have greater sensitivity and less specificity to select a greater number of cases (Germans et al., 2012); H4) EBADEP-A should have greater discriminative power compared to EBADEP-screening (Baptista & Carvalho, 2018).

METHOD

Participants

Our sample was composed of 187 individuals divided into clinical and non-clinical groups. The clinical group consisted of 52 participants, mostly female (92.3%) aged between 18 and 62 years old ($M = 35.71$ $SD = 11.54$), who were being treated for MDE or MDD in a Psychosocial Care Center (CAPS) and the Family Health Strategy (ESF). As inclusion criteria, the patient should present a diagnosis of MDE or MDD by a psychiatrist. Therefore, the gold standard for the clinical sample was the diagnosis by a psychiatrist.

The non-clinical group was composed of 135 individuals, college students, the majority (72.2%) was female, aged between 18 and 51 ($M = 23.8$ and $SD = 6.49$). Participants in this group reported never having undergone psychiatric and psychological treatment, not using medication for any psychopathology, and had no symptoms of MDE or MDD. The Beck Depression Inventory (BDI) was used to confirm this psychopathology absence in the group.

Instruments

Baptista Depression Scale Adult Version (EBADEP-A; Baptista, 2012)

This instrument was designed and standardized in Brazil to track symptoms of depression in psychiatric and non-psychiatric samples. The scale was developed based on the DSM-IV-TR, ICD-10, the Cognitive Model, and the Behavioral Theory. This one-dimension scale consists of 45 items disposed on a semantic-differential format, containing two contrasting statements each. A rating scale is used to measure if the participant agrees more with the first statement, the second, or with both at the same magnitude. Studies have revealed good psychometric properties for the scale (Baptista, 2012; Baptista & Carvalho, 2018; Bighetti et al., 2014).

Baptista Depression Scale - Screening Version (EBADEP- Screening)

EBADEP-screening was developed based on the adult version of EBADEP (EBADEP-A; Baptista, 2012). This scale aims to assess the symptoms of depression and is composed of 15 items selected based on the main symptoms of psychiatric manuals, such as sad mood, anhedonia, guilt, fatigue, concentration, suicidal ideation, and sleep. In the study conducted by Baptista e Carvalho (2018), the EBADEP-screening was able to discriminate 40 patients diagnosed with depression by SCID-I from 40 people without depression, with sensitivity equal to 95.0 and specificity of 87.5. In this study, the internal consistency reliability of EBADEP-screening was equal to 0.78.

Beck Depression Inventory (BDI; Beck et al., 1961)

The BDI is a self-report scale composed of 21 items, with a 4-point Likert scale. The cutoffs are the following: 0 to 11 = minimum, 12 to 19 = mild, 20 to 35 = moderate, and 36 to 63 = severe. Studies indicate good psychometric qualities of this scale (Beck et al., 1988; Paranhos et al., 2010; Wang & Gorenstein, 2013). We employed the cutoff indicated by Kendall et al. (1987), suggesting that individuals with scores above 20 points should be classified as mild depressive.

Procedures

Our research was received by the University of São Francisco ethics committee and approved. Data collection with the group of depressives was performed with people designated by the referred treatment centers who had a confirmed diagnosis of depression. Participants were informed about the data collection procedures. The applications, including EBADEP-A, were carried out in an appropriate service room, individually.

The non-clinical group was composed of people without a diagnosis of depression and a self-reported psychiatric history. This group responded to EBADEP-A and BDI. BDI was used as a gold standard to verify the absence of depressive conditions in this group. The exclusion criterion was presenting indicators of depression in the BDI.

We highlight that EBADEP-screening scores were calculated later, extracting the items that make up this version of EBADEP-A.

Data Analysis

We organized data in an electronic spreadsheet and later exported it to the statistical program SPSS version 21 and

the statistical program Medcalc. The t-test was employed to compare the mean score of depressive and non-clinical groups, and the Cohen's d (Cohen, 1992) indicator was used to verify the differences' magnitude. We used the ROC (Receiver Operating Characteristic) curve analysis to investigate the best cutoff and calculate the sensitivity, specificity, true predictive value, negative predictive value, positive probability rates, negative probability rate, and the efficiency test of the scales. In order to calculate the positive and negative predictive values, we used the formula proposed by Streiner (2003) for samples without prevalence were used. Finally, to verify the equivalence of EBADEP-A and EBADEP-screening, the intraclass correlation coefficient (ICC) and binary logistic regression were calculated.

RESULTS

Table 1 presents the t-test results to compare the means between the groups in EBADEP-A and EBADEP-screening.

The clinical group showed a higher mean than the non-clinical group in the total score of EBADEP-A and EBADEP-screening. We observed big difference between the groups. In Figure 1 and Table 2, we present the ROC curve results, including diagnostic accuracy indicators.

The area under the ROC curve was 0.96 (CI = 0.925 and 0.985; $p = 0.001$) for EBADEP-A and 0.95 (CI = 0.90 and 0.98; $p = 0.001$) for EBADEP-screening.

The best cutoff was 54, with better sensitivity indicators than the cutoff 59, as suggested by the EBADEP-A manual. Screening scales should be highly sensitive, indicating that the instrument has little chance of false negative.

We use Streiner's formula (2003) for samples without prevalence information to calculate the positive and negative predictive values. The values were 0.91 and 0.92, respectively, indicating that an individual with a positive result on EBADEP-A has 91% chance of having depression, while a person with a negative result on EBADEP-A has 92% chance of not having depression. The positive and negative probability rates were calculated. The values were 7.90 and 0.09, respectively, showing moderate ability to identify who has MDE or MDD and high ability to identify who does not have these conditions. The effectiveness test showed that EBADEP-A correctly identified 89% of the total cases (people with and without a diagnosis).

Table 1

Mean comparison between depressive and non-clinical groups.

Scales	Groups	M	SD	p (d)
EBADEP-A	-	28,70	19,07	< 0,001 (2,55)
	+	80,54	23,30	
EBADEP-screening	-	10,30	7,34	<0,001 (2,56)
	+	27,94	6,70	

Note. - = non-clinical group; + = clinical group.

Table 2

Diagnostic accuracy indicators for EBADEP-A and EBADEP-screening

	CO	Ss	Sp	+PV	-PV	+PR	-PR	ET
EBADEP-A	54	0,92	0,88	0,91	0,92	7,90	0,09	0,89
	59	0,84	0,91	0,86	0,91	9,66	0,17	0,89
EBADEP-screening	17	0,96	0,82	0,96	0,81	5,45	0,05	0,87

Note. CO: cutoff; Ss = sensitivity; Sp = specificity; +PV = positive predictive value; -PV = negative predictive value; +PR = positive probability rates; -PR = negative probability rates; TE = effectiveness test.

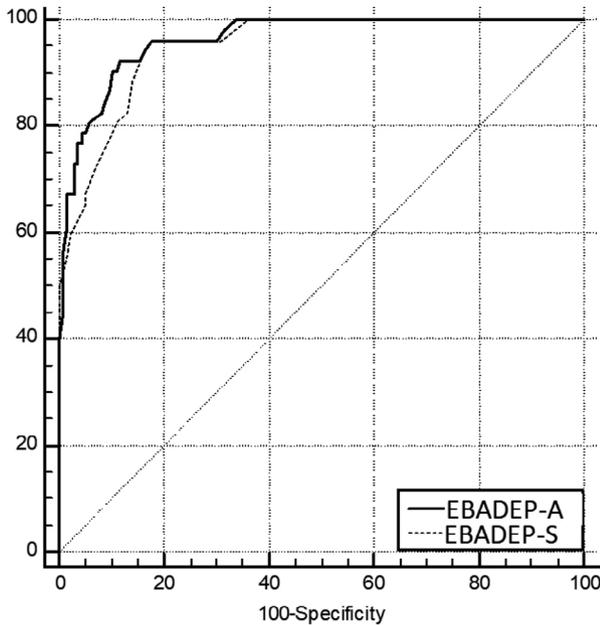


Figure 1. AUC of EBADEP-A and EBADEP-screening

Note. $AUC_{\text{ebadep-a}} = 0,96$ (IC = 0,925 - 0,985; $p = 0,001$); $AUC_{\text{ebadep-screening}} = 0,95$ (IC = 0,90 - 0,98; $p = 0,001$)

The results indicated that, for EBADEP-screening, the cutoff 17 is the most appropriate. As in EBADEP-A, the screening version also showed higher sensitivity values. The values suggest that 96% of individuals positive for depression on EBADP-screening have a diagnosis (MDE or MDD), and 81% of people negative on EBADEP-screening did not have a diagnosis for the same conditions. The positive probability rate was 5.45, and the negative probability rate was 0.05, indicating moderate ability to identify who has MDE or MDD and a high ability to identify individuals without these conditions. The effectiveness test showed that the EBADEP-screening correctly identified 87% of the cases.

The cutoff of 59 for EBADEP-A correctly identified 45 patients with MDE or MDD and 123 non-patients. In contrast, the cutoff 54 indicated by our study's results, correctly identified 47 patients and 125 non-patients. EBADEP-screening was able to identify 50 patients and 110 non-patients correctly.

We employed the ICC and logistic regression analysis to investigate the equivalence between EBADEP-A and EBADEP-screening. The ICC was 0.88 ($p < 0.01$), and in logistic regression, the $r^2_{\text{nagelkerke}}$ was 0.70 for EBADEP-A, and $r^2_{\text{nagelkerke}}$ was 0.63 for EBADEP-screening.

DISCUSSION

Diagnostic accuracy indicators provide information to the professional, which helps decide whether a person is diagnosed with a particular condition. Besides, diagnostic accuracy information may indicate a correct prognosis for a case (Glasser, 2014). The present study aimed to verify the diagnostic accuracy of EBADEP-A and EBADEP-screening in the discrimination of people with and without MDE and MDD and verify the equivalence between the two versions of the scale. Our hypotheses were partially confirmed, with the clinical group presenting higher means in both versions of EBADEP (H1). The AUC was above 0.90 (H2), in addition to both versions of the scale showing greater sensitivity in comparison to specificity (H3). The hypothesis that EBADEP-A would have a better discriminative capacity than screening (H4) was refuted.

The depressive group presented higher mean than the non-clinical group in EBADEP-A and EBADEP-screening, with differences of strong magnitude. Similar results with EBADEP-A previously indicated this scale's capacity to identify people with MDE or MDD (Baptista & Gomes, 2011). The findings with EBADEP-screening also exhibited the ability to separate the groups, even slightly higher than EBADEP-A, which may indicate that the screening version provides an assessment close to the full version. Bolsoni and Zuardi (2015) emphasize the usefulness of brief scales mainly for primary health care use. The full version of the scale can generate more diagnostic information and be recommended for use in contexts where the professional

is available for further investigation, while the screening version can be used in contexts of faster assessments such as screening in public health services.

The two versions of EBADEP presented an excellent result in the AUC (Bewick et al., 2004; Hajian-Tilaki, 2013; Hosmer & Lemeshow, 2000; Obuchowski, 2000), which refers to the instrument's general accuracy capacity (Glasser, 2014), being slightly higher in EBADEP-A. These data indicate both versions have good general capacity to identify people with MDE or MDD. The cutoffs suggested in this study (54 for EBADEP-A and 17 for EBADEP-screening) were shown to correctly identify the largest number of cases of MDE and MDD.

As for the ability to identify true positive and negative cases, the findings indicate that EBADEP-A and EBADEP-screening present superior results when compared with other tools such as BDI-I (Cunha, 2016), BDI-II (Gomes et al., 2012), and PHQ-9 (Santos et al., 2013). The positive and negative predictive values were calculated based on Streiner (2003) for samples with no prevalence and indicated that the results of EBADEP-A and EBADEP-screening provide correct information for diagnostic decision-making. Both versions showed good capacity to identify people with or without depression, demonstrating the usefulness of these scales in the clinical context.

Our findings indicate that the two versions of EBADEP are equivalent. Moreover, the results suggest that the screening version, with a smaller number of items, has a

capacity similar to the full version of EBADEP-A. This finding is in line with what was observed by Baptista and Carvalho (2018), who also found equivalence between the versions. The results of the diagnostic accuracy indicators, in general, also suggest the use of the screening version as this version obtained similar results to the full version, using fewer items and requiring less administration time.

Our findings indicate that the two versions of EBADEP provide essential information to assist clinicians in making

decisions on the presence of MDE or MDD, specifically for screening purposes. However, the results must be weighed in light of the main methodological limitations of the study. The restricted sample size may have decreased the variability of the observed responses; the unequal number of men and women may have impacted the results; we did not administer other tests to assess MDE and MDD, making it impossible to compare the versions of EBADEP with external measures.

REFERÊNCIAS

- American Psychiatric Association. (2014). *DSM-5: Manual Diagnóstico e Estatístico de Transtornos Mentais* [Diagnostic and Statistical Manual of Mental Disorders]. Artmed Editora.
- Baptista, M. N. (2012). *Manual técnico da Escala Baptista de Depressão em Adultos (EBADEP-A)* [Baptist Adult Depression Scale Technical Manual]. Vetor.
- Baptista, M. N., & Carvalho, L. D. F. (2018). Diagnostic Accuracy of a Brazilian Depression Self-Report Measure (EBADEP): Original and Short Versions. *Avaliação Psicológica*, 17(4), 484-492. <http://dx.doi.org/10.15689/ap.2018.1704.8.08>
- Baptista, M. N., & Gomes, J. O. (2011). Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: evidências de validade de construto e de critério [Escala Baptista de Depressão (Versão Adulto) - EBADEP-A: Construct and criterion Validity Evidences]. *Psico-USF*, 16(2), 151-161. <https://doi.org/10.1590/S1413-82712011000200004>
- Baptista, M. N., Cardoso, H. F., & Gomes, J. O. (2012). Escala Baptista de Depressão (Versão Adulto) - EBADEP-A: Validade Convergente e Estabilidade Temporal [Escala Baptista de Depressão (Versão Adulto) – EBADEP-A: Convergent Validity and Temporal Stability]. *Psico-USF*, 17(3), 407-416. <https://doi.org/10.1590/S1413-82712012000300007>
- Beard, C., Millner, A. J., Forgeard, M. J. C., Fried, E. I., Hsu, K. J., Treadway, M. T., Leonard, C. V., Kertz, S. J., & Björgvinsson, T. (2016). Network Analysis of Depression and Anxiety Symptom Relationships in a Psychiatric Sample. *Psychological Medicine*, 46(16), 3359-3369. <https://doi.org/10.1017/S0033291716002300>
- Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1997). *Terapia Cognitiva da Depressão* [Cognitive Therapy for Depression]. Artes Médicas.
- Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric Properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8(1), 77-100. [https://doi.org/10.1016/0272-7358\(88\)90050-5](https://doi.org/10.1016/0272-7358(88)90050-5)
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An Inventory for Measuring Depression. *Archives of General Psychiatry*, 4(6), 561-571. <https://doi.org/10.1001/archpsyc.1961.01710120031004>
- Bewick, V., Cheek, L., & Ball, J. (2004). Statistic Review 13: Receiver Operating Characteristic Curves. *Critical Care*, 8(6), 508-512. <https://doi.org/10.1186/cc3000>
- Bighetti, C. A., Alves, G. A. S., & Baptista, M. N. (2014). Escala Baptista de Depressão (EBADEP-A): Evidências de Validade com o Big Five [Baptista Depression Scale (EBADEP-A): Evidence for Validity with the Big Five]. *Avaliação Psicológica: Interamerican Journal of Psychological Assessment*, 13(1), 29-36.
- Bolsoni, L. M., & Zuardi, A. W. (2015). Estudos Psicométricos de Instrumentos Breves de Rastreamento para Múltiplos Transtornos Mentais [Psychometric Studies of Brief Screening Tools for Multiple Mental Disorders]. *Jornal Brasileiro de Psiquiatria*, 64(1), 63-69. <https://doi.org/10.1590/0047-2085000000058>
- Bossuyt P. M & Leflang M. M. (2008). Developing Criteria for Including Studies. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. The Cochrane Collaboration.
- Carneiro, A. M., & Baptista, M. N. (2012). Saúde Geral e Sintomas Depressivos em Universitários [General Health and Depressive Symptoms in Undergraduate students]. *Salud & Sociedad*, 3(2), 166-178. <https://doi.org/10.22199/S07187475.2012.0002.00004>
- Cunha, J.A. (2016) *Manual da Versão em Português das Escalas Beck* [Manual of the Portuguese Version of the Beck Depression Inventory]. Casa do Psicólogo.
- Durisko, Z., Mulsant, B. H., & Andrews, P. W. (2015). An Adaptationist Perspective on the Etiology of Depression. *Journal of Affective Disorders*, 1(172), 315-323. <https://doi.org/10.1016/j.jad.2014.09.032>
- Farone, S. V., & Tsuang, M. T. (1994). Measuring Diagnostic Accuracy in the Absence of a “Gold Standard”. *The American Journal of Psychiatry*, 151(5), 650. <https://doi.org/10.1176/ajp.151.5.650>
- Fried, E. I. (2016). Are More Responsive Depression Scales Really Superior Depression Scales? *Journal of Clinical Epidemiology*, 77, 4-6. <https://doi.org/10.1016/j.jclinepi.2016.05.004>
- Fried, E. I., & Nesse, R. M. (2015). Depression Is Not a Consistent Syndrome: An Investigation of Unique Symptom Patterns in the STAR*D study. *Journal of Affective Disorders*, 172, 96–102. <https://doi.org/10.1016/j.jad.2014.10.010>
- Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., & Borsboom, D. (2016). What Are ‘Good’ Depression Symptoms? Comparing the Centrality of DSM and Non-DSM Symptoms of Depression in a Network Analysis. *Journal of Affective Disorders*, 189, 314-320. <https://doi.org/10.1016/j.jad.2015.09.005>
- Germans, S., Van Heck, G., L., & Hodiament, P. P. G. (2012). Results of the Search for Personality Disorder Screening Tools: Clinical Implications. *Journal of Clinical Psychiatry*, 73(2), 165-173. <https://doi.org/10.4088/JCP.11m07067>
- Glasser, S. P. (2014). Research Methodology for Studies of Diagnostic Tests. In S. P. Glasser (Ed.). *Essentials of Clinical Research* (pp. 313-326). Springer.
- Goldberg, D. (2011). The Heterogeneity of “Major Depression”. *World Psychiatry*, 10(3), 226-228. <https://doi.org/10.1002/j.2051-5545.2011.tb00061.x>
- Gomes-Oliveira, M.H., Gorenstein, C., Lotufo-Neto, F., Andrade, L.H. & Wang, Y.P. (2012). Validation of the Brazilian Portuguese Version of the Beck Depression Inventory-II in a Community Sample. *Revista Brasileira de Psiquiatria*, 34, 389-394. <https://doi.org/10.1187/5047-89722056321477>
- Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*, 4(2), 627-635.

- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). John Wiley.
- Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and Recommendations Regarding Use of the Beck Depression Inventory. *Cognitive Therapy and Research*, *11*(3), 289-299. <https://doi.org/10.1007/BF01186280>
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A. J., Walters, E. E., & Wang, P. S. (2003). The Epidemiology of Major Depressive Disorder Results from the National Comorbidity Survey Replication. *Journal of the American Medical Association*, *289*(23), 3095-3105. <https://doi.org/10.1001/jama.289.23.3095>
- Lalkhen, A. G., & McCluskey, A. (2008). Statistics V: Introduction to Clinical Trials and Systematic Reviews. *Continuing Education in Anaesthesia, Critical Care & Pain*, *8*(4), 143-146. <https://doi.org/10.1093/bjaceaccp/mkn023>
- Lichtenberg, P., & Belmaker, R. H. (2010). Subtyping Major Depressive Disorder. *Psychotherapy and Psychosomatics*, *79*(3), 131-135. <https://doi.org/10.1159/00028695>
- Obuchowski, N. A. (2000). Sample Size Tables for Receiver Operating Characteristic Studies. *American Journal of Roentgenology*, *175*(3), 603-608. <https://doi.org/10.2214/ajr.175.3.1750603>
- Paranhos, M. E., Argimon, I. I. L., & Werlang, B. S. G. (2010). Propriedades Psicométricas do Inventário de Depressão de Beck-II (BDI-II) em Adolescentes [Psychometric Properties of the Beck Depression Inventory-II (BDI-II) in Adolescents]. *Avaliação Psicológica*, *9*(3), 383-392.
- Parshall, M. B. (2013). Unpacking the 2x2 table. *Heart & Lung: The Journal of Acute and Critical Care*, *42*(3), 221-226. <https://doi.org/10.1016/j.hrtlng.2013.01.006>
- Santos, I. S., Tavares, B. F., Munhoz, T. N., Almeida, L. S. P. D., Silva, N. T. B. D., Tams, B. D., Patella, A. M., & Matijasevich, A. (2013). Sensibilidade e Especificidade do Patient Health Questionnaire-9 (PHQ-9) entre Adultos da População Geral [Sensitivity and Specificity of the Patient Health Questionnaire-9 (PHQ-9) Among Adults from the General Population]. *Cadernos de Saúde Pública*, *29*, 1533-1543.
- Souza, M.S, Baptista, M.N, & Alves, G.A.S. (2015). Estudos Psicométricos Preliminares da Escala Baptista de Depressão para Adultos [Preliminary Psychometric Studies of Baptista's Depression Scale for Adults]. *Estudos de Psicologia (Campinas)*, *32*(3), 357-370. <https://dx.doi.org/10.1590/0103-166X2015000300002>
- Streiner, D. L. (2003). Being Inconsistent About Consistency: When Coefficient Alpha Does and Doesn't Matter. *Journal of Personality Assessment*, *80*(3), 217-222. https://doi.org/10.1207/S15327752JPA8003_01
- Van Stralen, K. J., Stel, V. S., Reitsma, J. B., Dekker, F. W., Zoccali, C., & Jager, K. J. (2009). Diagnostic Methods I: Sensitivity, Specificity, and Other Measures of Accuracy. *Kidney International*, *75*(12), 1257-1263. <https://doi.org/10.1038/ki.2009.92>
- Wang, Y. P., & Gorenstein, C. (2013). Psychometric Properties of the Beck Depression Inventory-II: A Comprehensive Review. *Brazilian Journal of Psychiatry*, *35*(4), 416-431. <https://doi.org/10.1590/1516-4446-2012-1048>
- World Health Organization (1996). *Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde: CID-10 Décima revisão*. [Trad. do Centro Colaborador da OMS para a Classificação de Doenças em Português]. 3 ed. EDUSP.
- World Health Organization. (2017). *Depression*. Recuperado de <http://www.who.int/mediacentre/factsheets/fs369/en/>