


Beck Depression Inventory-II: Self-report or interview-based administrations show different results in older persons

Hana Stepankova Georgi,  Karolina Horakova Vlckova, Jiri Lukavsky, Miloslav Kopecek, and Martin Bares

National Institute of Mental Health, Klecany, Czech Republic

ABSTRACT

Beck Depression Inventory-II (BDI-II) is one of the most-used rating scales. It was developed as a tool administered either as a self-rating or interview-based, observer-rating scale.

Objective: The goal of this study is to compare BDI-II scores obtained with two standard methods of administration in community-based older persons.

Methods: BDI-II was administered at first in the self-rated version to a sample of 60 mentally healthy older persons (age 60–87 years). Afterward, the interview-based administration was performed.

Analyses: We compared the scores with nonparametric tests — Spearman's correlation coefficient and Wilcoxon Signed Ranks test. We also computed internal consistency.

Results: Self-rated BDI-II yielded significantly higher total score than interview ($p < 0.001$, $P = 88\%$). The correlation between total scores was moderate ($r_s = 0.46$, $p < 0.001$). Item analysis revealed a larger decrease (lower scores) in the somatic items in the interview-based version.

Conclusions: The two methods of administration result in different total score in healthy older persons. Therefore, interpretation of the scores should reflect the administration, which should be always specified in the studies.

Key words: Beck Depression Inventory, depression, psychogeriatrics, rating scales, research design and methodology

Introduction

Together with the development and production of antidepressants, regulation of the pharmacological industry, and increased prescription of medicines after the World War II (Healy, 1997), a need to assess the effectiveness of therapies has evolved. This need, or a demand, led to the introduction of many depression rating scales, among which Beck Depression Inventory (BDI) is one of the most applied (Demyttenaere and De Fruyt, 2003). BDI has been used for over half a century now, with the original version published in 1961 (Beck *et al.*, 1961). The paper also contained a full version of the questionnaire with 21 items and rating on a 4-point Likert-type scale (0–3 points per item), with a higher score for more severe symptoms. The administration was designed as a structured interview with questions

presented so that the patient's attitudes at the time of the interview would be elicited — an exact wording related to the period in mind was not specified (Beck *et al.*, 1961; Domino, 2006, p. 461). The authors warned that the method was to assess the intensity of depression but not to distinguish different diagnostic categories of mental disorders. BDI has gained immediate attention of clinicians and works on its psychometric properties in healthy and ill populations followed (Richter *et al.*, 1998). The abbreviated version, BDI-SF (Beck *et al.*, 1974), with 13 items has not found such popularity. Still, we can, for example, see it in Czech psychiatric literature (published in English) (Bares *et al.*, 2009; Kopecek *et al.*, 2007).

Shortly after the first revision, BDI-IA (Beck and Steer, 1993), the second revision was published: BDI-II (Beck *et al.*, 1996). BDI-II also consists of 21 items, but four differ from the original BDI. The BDI-IA items “weight loss”, “body image change”, “work difficulty,” and “somatic preoccupation” were substituted with “agitation”, “worthlessness”, “concentration difficulty,” and “loss of energy”.

Correspondence should be addressed to: Hana Stepankova Georgi, National Institute of Mental Health, Topolova 748, 250 67 Klecany, Czech Republic; Phone: +420 283 088 240; Email: hana.stepankova@nudz.cz. Received 12 Nov 2017; revision requested 21 Dec 2017; revised version received 23 May 2018; accepted 23 May 2018. First published online 9 October 2018.

The new items were aimed to correspond better to the diagnostic criteria for Major Depression in the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders DSM-IV (American Psychiatric Association, 1994), with three of those (except for “worthlessness”) also symptoms of anxiety disorders. The items scoring (0–3 points) remained the same, the time considered was extended to two weeks comparing to the original BDI. The severity of symptoms is interpreted as “minimal” (0–13), “mild” (14–19), “moderate” (20–28), and “severe” (29–63) (Beck *et al.*, 1996).

The BDI-II manual was also published in the Czech Republic with BDI-II introduced as a self-rated scale (Preiss and Vacíř, 1999), and has been used since both in clinical and research practice (e.g., Bezdicek *et al.*, 2014; Ptáček *et al.*, 2016).

Many psychometric analyses through the years rendered two to three underlying factors of BDI-II, cognitive, affective, and somatic factors. While there is not a consensus regarding the exact factor structure as the results are mixed, and each population renders different factor structure, its good psychometric qualities support its use in clinical and non-clinical samples (Huang and Chen, 2015; Manian *et al.*, 2013), including its use as a screening tool in older community-dwelling adults (Krell-Roesch *et al.*, 2018; Segal *et al.*, 2008).

Both methods of administration of scales, self-rating, and observer-rating are used in inventories for depression in older population. In older persons, their cognitive status is one of the factors that should play a role in a clinician’s decision of the scale used. Cognitively impaired older patients should be assessed with observer-rated scales, because the patient may lack an understanding and insight into their problems (Ganguli *et al.*, 2006). For the cognitively healthy, the self-rated scales are used (Sharp and Lipsky, 2002). The BDI-II manual allows for two ways of administration: Self-Report or an Interview.

While BDI was originally proposed as an observer-rated scale based on an interview with a trained administrator (not necessarily a clinician), and BDI-II as administered in either way (self-rated or interview), they are now widely presented and used as self-rated instruments (Brown *et al.*, 2012; Hagen, 2007; Joe *et al.*, 2008; Segal *et al.*, 2008). Regularly, the little training to administer the self-rated methods is presented as one of its advantages (Smarr and Keefer, 2011). BDI-II is used both in clinical practice and in research to assess level of depressivity in patients and also in healthy persons. Also, it is commonly used to document that a study sample was depression-free (Bezdicek *et al.*, 2014; Krell-Roesch *et al.*, 2018; Rahe *et al.*, 2015; Stepankova *et al.*, 2012).

This study was inspired by our cognitive training study experience (Stepankova *et al.*, 2012), and oral comments and inquiries of the healthy older persons related to several items of the BDI-II, especially to the last one (“sexual interest”) during their filling in the self-report BDI-II form. It became apparent that older persons tended to compare the past two weeks (as written in the original standard instructions) with their prime or youth. Reflecting that, we designed this study to find whether there would be a consistent picture and what the effects would show in the scores.

Methods

Participants

We recruited 86 adult community-dwelling persons of normal cognition (Mini-Mental State Examination MMSE 28–30) (Folstein *et al.*, 2001) who had participated in the previous cognitive training study approved by the institutional ethics committee. All participants signed informed consent forms.

Procedure

BDI-II was administered as a self-report questionnaire in accordance with the manual (Beck *et al.*, 1996; Preiss and Vacíř, 1999) (BDI-II-SR). The forms were checked for completeness. Then, after a short break of several minutes, we administered the BDI-II in a structured interview (BDI-II-IB). The original instruction read from the form was emphasized: to consider the past two weeks, including today. This was a vocal cue from the administrator. The participants made the scores on the paper-form such as in the first administration method. When participants hesitated about the assessed period, the instruction was repeated. If necessary, and directly asked, it was answered not to compare the current status for example with their youth. The order of administration methods was stable. In an opposite order, the participants would have been biased as to spontaneity in the Self-Reports. In most studies, self-rating is used without prior exposure to the tool. Thus, the administration in this study reflected the real-life assessment situation.

The same reviewers administered both versions, thus they were not blinded to the scores. They were not personally interested in the results — it was not important for them whether the scores in both administration methods differed or not; they did not co-author this manuscript; and the assessed persons were not patients of the institute or the administrators. In that way we tried to avoid a possible administrator bias.

Statistical methods and data analyses

For comparison of these two methods of administrations, we computed total scores in BDI-II-IB and BDI-II-SR. Because the distribution of total scores was not normal, we applied the nonparametric tests. We analyzed these total scores using Spearman's correlation coefficient and using Wilcoxon Signed Ranks test for two dependent samples. Using Wilcoxon Signed Ranks test we also tested all of items to present the source of the total score difference. Thus, we used Bonferroni correction for multiple comparisons. The effect size was computed using an analogy of probabilistic index for dependent samples. The best equivalent is to report percentages of higher scores as this method retains the clarity of $P(X > Y)$ reporting (Acion *et al.*, 2006). Thus, we report percentages of cases when scores from the Interview are lower than the scores from the Self-Report Using Wilcoxon Signed Ranks test. We also tested the hypothesis that the answers to each item did not differ despite using the different method of administration. We calculated mean difference between the scores to present the source of the total score difference. We also computed internal consistency using Cronbach's alpha. All statistical procedures were done using SPSS 16.0.

Results

Sample

The recruited persons were scrutinized with an objective of a homogenous sample as for mental health. After we applied exclusion criteria of no psychiatric history and no current or previous psychiatric medication, the convenience sample consisted of 60 mentally healthy older persons (Table 1).

Administration methods

Internal consistency (Cronbach's alpha) was 0.77 for the Interview and 0.79 for the Self-Report. The BDI-II-IB mean was 3.22 points (SD 3.77), the BDI-II-SR mean was 9.67 points (SD 6.27), thus the mean difference was 6.45 points. Figure 1 shows a histogram of total scores in both administration methods. The distribution of total scores varied and was normal in the Self-Report ($Z = 0.740$, $p = 0.645$, skewness 1.476, kurtosis 5.670), but for the Interview the distribution was not normal and skewed toward lower values ($Z = 1.519$, $p = 0.020$, skewness 1.430, kurtosis 1.869). In the Interviews, 20 persons had a total score of 0; in the Self-Reports, two persons had a total score of 0 (Figure 1). All participants gave either equal or lower scores in all items in the Interview than in the Self-Report. Based on the

Table 1. Demographic data of the sample ($n = 60$)

GENDER	13 MEN (22%)
Education	Lower secondary – 3% Complete secondary – 57% Tertiary – 40%
Age	60–87 years; $M = 67.9$ years; $SD = 6.6$

recommended standard cut-off values (Beck *et al.*, 1996), the Interview resulted in 98% of persons reporting minimal depression and 2% reporting mild depression, while Self-Report showed 78% minimally depressed, 20% mildly depressed, and 2% severely depressed. Scrutinizing the results, we found that in Self-Reports 64% of participants gave the lowest scores to all the items (0 points), while in Interview 87%. Using Wilcoxon test, we found the most stable answers — i.e., answers remained the same both in Self-Report and Interview and the differences between them were not significant (Table 2) — were to nine items: Sadness, Past Failure, Punishment Feelings, Self-Dislike, Suicidal Thoughts or Wishes, Crying, Agitation, Loss of Interest, and Irritability.

Spearman's rank correlation of total scores of both methods of administration was $r_s = 0.46$ ($p < 0.001$). A Wilcoxon Signed-Ranks test indicated that the Self-Report (Median = 9) total score was statistically significantly higher than the Interview-Based total score (Median = 2, $z = -6.3$, $p < 0.001$).

Discussion

The goal of this study was to compare two ways of an administration — Self-Report and Interview — of a standard method for a detection of depressive symptomatology, BDI-II, and their effect on the scores, in cognitively and mentally healthy older persons. BDI-II is often used as a measure documenting mood status in research (Wang and Gorenstein, 2013).

Because our subjects were cognitively healthy, non-depressed older persons, the level of reported depressive symptomatology was not expected to be high. Results confirmed this expectation with the majority of the sample reporting minimal to no depression (SR: 78%, IB 98%). Both methods of administration showed similar good internal consistency, SR $\alpha = 0.79$ and IB $\alpha = 0.77$. Our results showed that both method's total scores correlated moderately, while Self-Report gave significantly higher total scores than the Interview. Nobody scored higher in the Interview than in the Self-Report. The largest differences in individual item scores were

Table 2. Descriptive statistics for the 21 BDI-II items in both methods of administration, Interview and Self-Report

ITEM	INTERVIEW (N = 60)		SELF-REPORT (N = 60)		DIFFERENCE		
	ITEM MEAN	ITEM SD	ITEM MEAN	ITEM SD	MEAN DIFFERENCE	P (IB < SR)	95% CI
1. Sadness	0.05	0.22	0.07	0.25	0.03	3%	(-0.01, 0.08)
2. Pessimism	0.27	0.48	0.47	0.62	0.23*	19%	(0.09, 0.38)
3. Past failure	0.30	0.62	0.49	0.70	0.22	17%	(0.08, 0.35)
4. Loss of satisfaction	0.17	0.38	0.39	0.52	0.23*	22%	(0.11, 0.35)
5. Guilty feelings	0.28	0.52	0.51	0.59	0.23*	22%	(0.11, 0.35)
6. Punishment feelings	0.13	0.57	0.24	0.70	0.10	7%	(-0.01, 0.21)
7. Self-dislike	0.03	0.10	0.08	0.20	0.07	7%	(0.00, 0.13)
8. Self-criticism	0.13	0.43	0.61	0.77	0.47*	35%	(0.28, 0.65)
9. Suicidal thoughts or wishes	0.03	0.18	0.12	0.38	0.08	7%	(0.00, 0.17)
10. Crying	0.08	0.33	0.20	0.58	0.12	7%	(0.00, 0.24)
11. Agitation	0.18	0.39	0.24	0.43	0.07	7%	(0.00, 0.13)
12. Loss of interest	0.07	0.25	0.12	0.33	0.08	7%	(0.00, 0.17)
13. Indecision	0.12	0.32	0.61	0.70	0.50*	42%	(0.32, 0.68)
14. Feelings of worthlessness	0.12	0.32	0.31	0.57	0.20*	17%	(0.07, 0.33)
15. Loss of energy	0.15	0.36	0.78	0.70	0.63*	50%	(0.44, 0.82)
16. Changes in sleeping pattern	0.42	0.79	0.98	0.86	0.52*	40%	(0.32, 0.71)
17. Irritability	0.07	0.25	0.17	0.50	0.10	7%	(-0.01, 0.21)
18. Changes in appetite	0.12	0.37	0.54	0.95	0.45*	24%	(0.21, 0.69)
19. Concentration	0.12	0.37	0.56	0.58	0.45*	38%	(0.28, 0.62)
20. Tiredness	0.30	0.53	0.76	0.70	0.45*	36%	(0.27, 0.63)
21. Loss of interest in sex	0.08	0.42	1.32	1.24	1.22*	58%	(0.89, 1.54)
Total score	3.22	3.77	9.67	6.27	6.45	88%	(5.07, 7.83)

Item Mean – arithmetic mean of the scores; SD – Standard Deviation; 95% CI – confidence interval; Mean difference: Self-report – Interview.

*p < 0.0023 (We used Bonferroni correction for multiple comparisons $0.05/21 = 0.0023$); P(IB < SR) - probability that the scores from Interview are lower than from Self-Report; corresponding P(IB > SR) = 0% for each item.

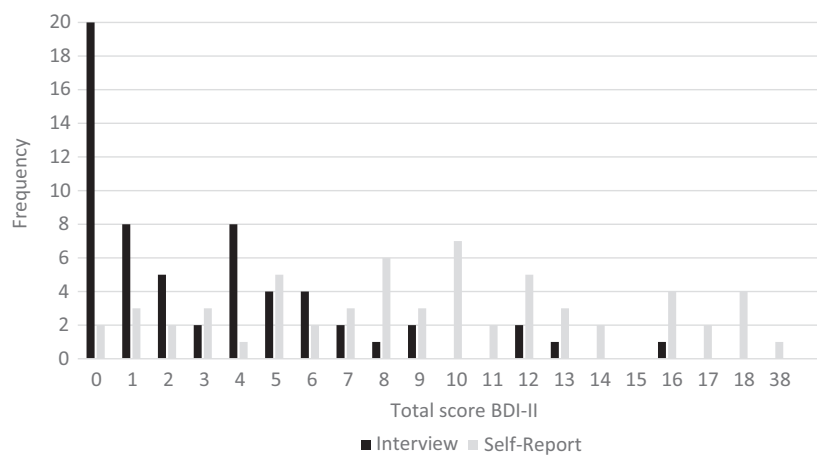


Figure 1. Frequency distribution of total scores in both administration methods.

found in Item 21: Loss of Interest in Sex, and in Item 15: Loss of Energy.

The process of reviewing this article led to several suggestions of sources for such a large difference in the mean total scores of about 6 points. Among them could be an order effect, comments made by

the administrators, the social situation of the interview, and so on (as discussed below), and the source explicitly mentioned by the participants. From those comments of the participants, we conclude a probable source: The differences in the scores were due to a possible explanation of items wording.

Most items include “as ever,” “than usual,” and “used to be.” These can be understood in various ways, as inquiries of our subjects showed. The drop of scores in the Interview resulted from such inquiries, after the older persons asked whether they are to compare themselves to (1) their prime years, youth; (2) other people, i.e., what is “normal” (which is, of course, very subjective); or (3) what is usual for them now in old age. When inquiring, the standard instruction was repeated to them with an emphasis on the last two weeks including today. In case they needed more information and wanted to talk more on the subject of the item, they were advised to consider their usual, normal healthy state in recent years, older age, and not compare to their youth. This became apparent in the items Loss of Interest in Sex, Loss of Energy, Indecisiveness, Changes in Sleeping Pattern, Changes in Appetite, Concentration Difficulty, and Tiredness or Fatigue. These items correspond to changes in normal aging: certain decline in the somatic-vegetative area such as in appetite (Hetherington, 1998), sleep (Ohayon *et al.*, 2004), and libido (Taylor and Gosney, 2011), and also in higher-order brain system and cognitive functions (Bishop *et al.*, 2010; Deary *et al.*, 2009), which may be perceived as reduction in mental energy (Glisky, 2007). Therefore, if an older person compares his/her own level in those areas with his/her youth, a decrease needs to be expected; it is less, lower, or worse than it “used to be,” while it is perfectly normal, non-pathologic in older age. And not just “normal”; these issues may not be perceived negatively, or hampering. They may be perceived as a given reality, not hindering well-being in older age. This is, of course, a matter of individual approach to aging, and also a matter of gravity, a scope of such changes.

The different answers in both administrations could not be due to low understanding of the items in the phase of Self-Report as our sample was quite highly educated, even comparing to the Czech general older population (Ceský statistický úřad [Czech Statistical Office], 2011, Chap. 21). Our sample had 40% of persons with completed tertiary education (in general Czech population: 10%), 57% with complete secondary education (general Czech: 28%).

A fixed order (Self-Report – Interview) was used so that we compared the usual setting (participant receives a questionnaire with written instructions, fills it, and hands over) with another, less-usual one of an interview where there is an interaction with the administrator and a chance to discuss the unclear items. If we rotated the order, participants who would do the Interview first would already be briefed as to the meaning of the items and their answers would reflect that. Because the number of potential

participants in the study was not high, we decided to follow the fixed routine. Rotating the order would bring additional information in the picture. We can hypothesize that if persons undergoing the “IB – SR” order had close to equal scores in both administration methods, and very close to the scores of the Interviews of the participants “SR-IB,” it would support our view the stand alone written instruction may be unclear to the older healthy persons.

Also, the social context of the Interview may play its role. One can imagine that older people may be reluctant to give high scores, e.g., to Item 21 (I lost interest in sex completely) in an interview. Our experience was not such. The participants discussed that item most often, emphasizing that they were very much active in their youth so now it is almost nothing compared to those times. It is possible that persons who are on the introvert side, shy, doubt their understanding, or interacting with an administrator who they perceive as intimidating or personally unpleasant may not ask for the explanation. Then their scores could remain unchanged. On the other hand, the participants had the paper forms to mark the scores; it was not the administrator who wrote the scores. Thus, we tried to secure a comfortable situation for participants to be as honest as possible in their answers.

The elapsed time between the two administrations may have played a role. It could be that a longer time would cause participants to forget how they scored in the first administration, and they would be unbiased by their first response. We were not able to invite them for another session, so we can only deduce we would have to also check for their life situation at the second assessment, whether it generally remained unchanged, and their mood and health status as well. The design we applied, with the two consecutive assessments, lead to lower scores even if we assumed our healthy participants remembered their first scores in the Self-Report. We also observed that nobody gave a higher score in any item in the Interview, i.e., did not report graver depressive symptoms.

While it is not possible to make a definite conclusion, our project suggests that comparisons with their youth may be the natural answer to the BDI-II items in older, healthy persons. The somatic and non-affective symptoms that are known to be often present in older, depressed persons (Wilkowska-Chmielewska *et al.*, 2013) and may disguise the mental disorder need to be addressed with carefully chosen wording especially in a self-report measure, so that such symptoms are not overlooked if present but do not show while actually absent. For example, a popular instrument for screening, Geriatric Depression Scale and its abbreviated form GDS-15 (Sheikh and Yesavage, 1986), avoids falsely elevated

scores in somatic issues by their omission, focusing on the psychological aspects of depression.

Our results brought us to ask: What are we really interested to know when subjecting a healthy older person to such a scale where phrases such as “as ever,” “than usual,” and “used to be” are used? Are we after a change occurring between two distinct points in a recent period of time, e.g., now versus two weeks ago? Or rather, what period of time or a state of overall fitness of body and mind should the person consider as the benchmark, what he or she should compare the status in the last two weeks to? Is the benchmark his/her best physical and mental status in youth? Is the benchmark his/her usual state in the past years, this stage of life, maybe when he/she was fit according to his/her opinion? Thus, the main message of this contribution may be a suggestion for the instructions to be more specified in this so that the meaning is clear to both raters and assessed subjects, and consequently, results are more reliable. As the improved reliability of scales used in clinical trials becomes a sought-after feature (Williams and Kobak, 2008), a consensus and better understanding of the relevant time periods to be compared should be reached. Such clarification and consensus will improve comparability of studies as well as diagnostic procedures. Furthermore, it was shown that non-cognitive parameters may have a predictive value in preclinical stages of dementia, which should be researched further (Masters *et al.* 2015), and only reliable and valid data will enable correct research outcomes.

Limitations

One of the limitations of our study is the gender ratio with prevailing women, which therefore did not allow for a gender-related analysis. This could be clarified in more focused follow-up research. Another limitation is the lack of counterbalancing of the order of administrations, Interview – Self-Report, which would enrich the data, and could support our view upon the understanding of the items by older healthy people. The number of participants in the study did not allow for the rotating and consecutive analyses though. The time elapsed between the two administrations could be also considered a limitation. It would be advisable to allow for a longer period to pass, and distract the persons with other tasks so that their attention would be led away from the questions and their answers. It is questionable how long the period should be, if the administrations should be done in one session with tasks between, or to allow for longer time to pass in a span of days or weeks. If a longer time passed, it could be discussed whether the status was the same or the personal situation could objectively change. The scheme of our study allowed

only for the described procedure. Some could suggest that a golden standard for an assessment of depression syndromes should have been used (a standardized clinical interview using the Diagnostic and Statistical Manual of Mental Disorders 4th edition (DSM-IV) or the International Statistical Classification of Diseases and Related Health Problems 10th Revision criteria (World Health Organization, 2016)). However, our sample included only healthy persons; therefore, we concluded the golden standard administration would not render useful data. Also, data from older persons suffering from clinical depression are not available for such a comparison in this study. It is possible they would yield different results.

Conclusion

It is highly recommendable to carefully choose the administration methods used in the assessment of older persons including healthy ones, and bear in mind the self-reports may produce different results than interviews due to a different understanding of the instructions and natural way of relating to the questions about one’s past. Thus, one cut score may not be adequate for both administration methods. Our study suggests that healthy older persons differ in their understanding of standard instructions that ask about their past. Patients may understand they are supposed to compare their current (e.g., past two weeks including today) state with when they felt well, were healthy, or remitted. Healthy older subjects may be confused regarding the benchmark they should use for comparison.

Conflict of interest

None.

Descriptions of authors’ roles

HSG designed the study and drafted the manuscript. KHV and JL were in charge of statistical analyses. MK and MB assisted in writing and revising the manuscript.

Acknowledgements

This work was supported by the Grant Agency of the Czech Republic (“Impact of settlement size on cognition in older age,” grant Nr. 17-14829S), and by the project “Sustainability for the National Institute of Mental Health,” under grant number LO1611, with a financial support from the Ministry

of Education, Youth and Sports of the Czech Republic.

References

- Acion, L., Peterson, J. J., Temple, S. and Arndt, S.** (2006). Probabilistic index: an intuitive non-parametric approach to measuring the size of treatment effects. *Statistics in Medicine*, 25, 591–602. doi: [10.1002/sim.2256](https://doi.org/10.1002/sim.2256).
- American Psychiatric Association** (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4th ed.). Washington, DC: American Psychiatric Association.
- Bares, M. et al.** (2009). Low frequency (1-Hz), right prefrontal repetitive transcranial magnetic stimulation (rTMS) compared with venlafaxine ER in the treatment of resistant depression: a double-blind, single-centre, randomized study. *Journal of Affective Disorders*, 118, 94–100. doi: [10.1016/j.jad.2009.01.032](https://doi.org/10.1016/j.jad.2009.01.032).
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J. and Erbaugh, J.** (1961). An inventory for measuring depression. *Archives of General Psychiatry*, 4, 561–571. doi: [10.1001/archpsyc.1961.01710120031004](https://doi.org/10.1001/archpsyc.1961.01710120031004).
- Beck, A. T., Rial, W. Y. and Rickels, K.** (1974). Short form of depression inventory: cross-validation. *Psychological Reports*, 34, 1184–1186.
- Beck, A. T. and Steer, R. A.** (1993). *Manual for the Beck Depression Inventory*. San Antonio, TX: Psychological Corporation.
- Beck, A. T., Steer, R. A. and Brown, G. K.** (1996). *Manual for Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.
- Bezdicek, O. et al.** (2014). Czech version of Rey auditory verbal learning test: normative data. *Aging, Neuropsychology, and Cognition*, 21, 693–721. doi: [10.1080/13825585.2013.865699](https://doi.org/10.1080/13825585.2013.865699).
- Bishop, N. A., Lu, T. and Yankner, B. A.** (2010). Neural mechanisms of ageing and cognitive decline. *Nature*, 464, 529–535. doi: [10.1038/nature08983](https://doi.org/10.1038/nature08983).
- Brown, M., Kaplan, C. and Jason, L.** (2012). Factor analysis of the Beck Depression Inventory-II with patients with chronic fatigue syndrome. *Journal of Health Psychology*, 17, 799–808. doi: [10.1177/1359105311424470](https://doi.org/10.1177/1359105311424470).
- Ceský statistický úřad [Czech Statistical Office]** (2011). Senioři v ČR 2010 [Seniors in the Czech Republic in 2010]; available at: http://www.czso.cz/csu/2010edicniplan.nsf/publ/1417-10-nepravidelne_2010; last accessed 10 November 2017.
- Deary, I. J. et al.** (2009). Age-associated cognitive decline. *British Medical Bulletin*, 92, 135–152. doi: [10.1093/bmb/ldp033](https://doi.org/10.1093/bmb/ldp033).
- Demyttenaere, K. and De Fruyt, J.** (2003). Getting what you ask for: on the selectivity of depression rating scales. *Psychotherapy and Psychosomatics*, 72, 61–70. doi: [10.1159/000068690](https://doi.org/10.1159/000068690).
- Domino, G.** (2006). *Psychological Testing: An Introduction* (2nd ed.). Cambridge, NY: Cambridge University Press.
- American Psychiatric Association** (1994). *DSM-IV*. Washington, DC: American Psychiatric Association.
- Folstein, M. F. et al.** (2001). *Mini-Mental State Examination User's Guide*. Odessa, FL: Psychological Assessment Resources.
- Ganguli, M. et al.** (2006). Discrepancies in information provided to primary care physicians by patients with and without dementia: the steel valley seniors survey. *The American Journal of Geriatric Psychiatry*, 14, 446–455. doi: [10.1097/01.JGP.0000199340.17808.77](https://doi.org/10.1097/01.JGP.0000199340.17808.77).
- Glisky, E. L.** (2007). Changes in cognitive function in human aging. In: D. R. Riddle (Ed). *Brain Aging: Models, Methods, and Mechanisms*. Boca Raton, FL: CRC Press (Chap. 1); Available at: <http://www.ncbi.nlm.nih.gov/books/NBK3885/>; accessed 10 November 2017.
- Hagen, B.** (2007). Measuring melancholy: a critique of the Beck Depression Inventory and its use in mental health nursing. *International Journal of Mental Health Nursing*, 16, 108–115. doi: [10.1111/j.1447-0349.2007.00453.x](https://doi.org/10.1111/j.1447-0349.2007.00453.x).
- Healy, D.** (1997). *The Antidepressant Era*. Cambridge, MA: Harvard University Press.
- Hetherington, M. M.** (1998). Taste and appetite regulation in the elderly. *Proceedings of the Nutrition Society*, 57, 625–631. doi: [10.1079/PNS19980091](https://doi.org/10.1079/PNS19980091).
- Huang, C. and Chen, J.-H.** (2015). Meta-analysis of the factor structures of the Beck Depression Inventory-II. *Assessment*, 22, 459–472. doi: [10.1177/1073191114548873](https://doi.org/10.1177/1073191114548873).
- Joe, S., Woolley, M. E., Brown, G. K., Ghahramanlou-Holloway, M., and Beck, A. T.** (2008). Psychometric properties of the Beck Depression Inventory-II in low-income, African American suicide attempters. *Journal of Personality Assessment*, 90, 521–523. doi: [10.1080/00223890802248919](https://doi.org/10.1080/00223890802248919).
- Kopecek, M., Cerná, L., Sulak, J., Raszka, M., Bares, M. and Seifertová, D.** (2007). Depressed patients' perception of the efficacy of electroconvulsive therapy and venlafaxine therapy. *NeuroEndocrinology Letters*, 28, 889–894.
- Krell-Roesch, J. et al.** (2018). Depressive and anxiety symptoms and cortical amyloid deposition among cognitively normal elderly persons: The Mayo Clinic study of aging. *International Psychogeriatrics*, 30, 245–251. doi: [10.1017/S1041610217002368](https://doi.org/10.1017/S1041610217002368).
- Manian, N., Schmidt, E., Bornstein, M. H. and Martinez, P.** (2013). Factor structure and clinical utility of BDI-II factor scores in postpartum women. *Journal of Affective Disorders*, 149, 259–268. doi: [10.1016/j.jad.2013.01.039](https://doi.org/10.1016/j.jad.2013.01.039).
- Masters, M. C., Morris, J. C. and Roe, C. M.** (2015). 'Noncognitive' symptoms of early Alzheimer disease: a longitudinal analysis. *Neurology*, 84, 617–622. doi: [10.1212/WNL.0000000000001238](https://doi.org/10.1212/WNL.0000000000001238).
- Ohayon, M. M., Carskadon, M. A., Guilleminault, C. and Vitiello, M. V.** (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep*, 27, 1255–1273. doi: [10.1093/sleep/27.7.1255](https://doi.org/10.1093/sleep/27.7.1255).
- Preiss, M. and Vacíř, K.** (1999). *Beckova sebepoznovací škála depresivity pro dospělé BDI-II. Příručka. [Beck self-report depression inventory for adults. BDI-II. Manual]*. Brno: Psychodiagnostika s r.o.
- Ptáček, R., Raboch, J., Vnuková, M., Hlinka, J. and Anders, M.** (2016). Beckova škála deprese BDI-II - Standardizace a využití v praxi [Beck Depression Inventory BDI-II-Standardization and practical use]. *Ceská a Slovenská Psychiatrie*, 112, 270–274.
- Rahe, J., Petrelli, A., Kaesberg, S., Fink, G. R., Kessler, J. and Kalbe, E.** (2015). Effects of cognitive training with

- additional physical activity compared to pure cognitive training in healthy older adults. *Clinical Interventions in Aging*, 2015, 297–310. doi: [10.2147/CIA.S74071](https://doi.org/10.2147/CIA.S74071).
- Richter, P., Werner, J., Heerlein, A., Kraus, A. and Sauer, H.** (1998). On the validity of the Beck Depression Inventory. A review. *Psychopathology*, 31, 160–168. doi: [10.1159/000066239](https://doi.org/10.1159/000066239).
- Segal, D. L., Coolidge, F. L., Cahill, B. S. and O'Riley, A. A.** (2008). Psychometric properties of the Beck Depression Inventory II (BDI-II) among community-dwelling older adults. *Behavior Modification*, 32, 3–20. doi: [10.1177/0145445507303833](https://doi.org/10.1177/0145445507303833).
- Sharp, L. K. and Lipsky, M. S.** (2002). Screening for depression across the lifespan: a review of measures for use in primary care settings. *American Family Physician*, 66, 1001–1008.
- Sheikh, J. I. and Yesavage, J. A.** (1986). Geriatric depression scale (GDS): recent evidence and development of a shorter version. *Clinical Gerontologist*, 5, 165–173. doi: [10.1300/J018v05n01_09](https://doi.org/10.1300/J018v05n01_09).
- Smarr, K. L. and Keefer, A. L.** (2011). Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care & Research*, 63, S454–S466. doi: [10.1002/acr.20556](https://doi.org/10.1002/acr.20556).
- Stepankova, H., Lukavsky, J., Kopecek, M., Steinova, D. and Ripova, D.** (2012). Modification of subjective cognitive outcomes in older persons through memory training. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, 25, 117–125. doi: [10.1024/1662-9647/a000061](https://doi.org/10.1024/1662-9647/a000061).
- Taylor, A. and Gosney, M. A.** (2011). Sexuality in older age: essential considerations for healthcare professionals. *Age and Ageing*, 40, 538–543. doi: [10.1093/ageing/afr049](https://doi.org/10.1093/ageing/afr049).
- Wang, Y. and Gorenstein, C.** (2013). Assessment of depression in medical patients: A systematic review of the utility of the Beck Depression Inventory-II. *Clinics*, 68, 1274–1287. doi: [10.6061/clinics/2013\(09\)15](https://doi.org/10.6061/clinics/2013(09)15).
- Williams, J. B. W. and Kobak, K. A.** (2008). Development and reliability of a structured interview guide for the Montgomery-Asberg depression rating scale (SIGMA). *The British Journal of Psychiatry*, 192, 52–58. doi: [10.1192/bjp.bp.106.032532](https://doi.org/10.1192/bjp.bp.106.032532).
- Wilkowska-Chmielewska, J., Szelenberger, W. and Wojnar, M.** (2013). Age-dependent symptomatology of depression in hospitalized patients and its implications for DSM-5. *Journal of Affective Disorders*, 150, 142–145. doi: [10.1016/j.jad.2012.12.012](https://doi.org/10.1016/j.jad.2012.12.012).
- World Health Organization** (2016). *ICD-10. International Statistical Classification of Diseases and Related Health Problems, 10 Revision* (5th ed.). Geneva: World Health Organization.