Brief report

Rasch analysis of the Beck Depression Inventory-II in stroke survivors: A cross-sectional study

Anners Lerdal [a,b,*], Anders Kottorp [c], Caryl L. Gay [a,d,e], Ellen K. Grov [b,f], Kathryn A. Lee [d]

[a] Lovisenberg Diakonale Hospital, Lovisenberggata 17, NO-0440 Oslo, Norway
[b] Department of Nursing Science, Institute of Health and Society, Faculty of Medicine, Oslo, Norway
[c] Division of Occupational Therapy, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden
[d] Department of Family Health Care Nursing, University of California, San Francisco, CA, USA
[e] Lovisenberg Diakonale University College, Oslo, Norway
[f] Førde University College, Førde, Norway

ABSTRACT

Background: The Beck Depression Inventory-II (BDI-II) is often used to assess depressive symptoms among stroke patients, but more evidence is needed regarding its psychometric properties in this population. The purpose of this study was to assess the BDI-II's psychometric properties using a Rasch model application in a sample of patients 6 months after a first clinical stroke.
Methods: Data were collected prospectively from patient medical records and from questionnaires (with assistance if needed) as a part of a longitudinal study of poststroke fatigue. Data from the 6-month follow-up were used in this analysis. The sample consisted of 106 patients with first-ever stroke recruited from two Norwegian hospitals between 2007 and 2008. Depressive symptoms were measured with the BDI-II. Rasch analysis was used to assess the BDI-II's psychometric properties in this sample.
Results: Five BDI-II items did not demonstrate acceptable goodness-of-fit to the Rasch model: items 10 (crying), 16 (changes in sleep), 17 (irritability), 18 (changes in appetite), and 21 (loss of interest in sex). If these 5 items were removed, the resulting 16-item version not only had fewer items, it also had better internal scale validity, person-response validity, and person-separation reliability than the original 21-item version in this sample of stroke survivors.
Limitations: The study did not include a clinical evaluation of depression.
Conclusion: A 16-item version of the BDI-II, omitting items 10, 16, 17, 18 and 21, may be more appropriate than the original 21-item BDI-II for use as a unidimensional measure of depression in patients following first-ever stroke.

1. Introduction

An estimated 33% of stroke survivors experience clinical depression (Hackett et al., 2005), and post-stroke depression (PSD) is associated with both reduced quality of life (Gainotti et al., 2001; Sturm et al., 2004) and increased mortality (Williams et al., 2004). Thus, it is an important clinical and research goal to develop accurate assessments of depressive symptoms (Forkmann et al., 2013) in order to elucidate the etiology, evaluate interventions and identify patients needing treatment.

Many assessments of depression severity use summative rating scales addressing a variety of symptoms and compare individual sum scores with standard cut-offs. However, such approaches assume that all items contribute equally to the same underlying construct. Assumptions regarding unidimensionality and interval-level data are rarely confirmed empirically and are routinely violated when applied to ordinal data (Kottorp, 2003). Such applications may also lead to results with misleading clinical interpretations (Fisher, 1993; Merbitz et al., 1989; Wright and Linacre, 1989), as depression may be either overestimated or underestimated depending on the specific items endorsed.

To address these concerns, Rasch models are increasingly used for psychometric evaluation of both new and existing instruments (Tesio, 2003). Rasch analysis can be used to examine whether items measure a unidimensional construct, a crucial aspect of both classical and modern measurement statistics. Items that demonstrate poor fit to the Rasch model might be considered for deletion

* Corresponding author. Tel.: +47 2322 5000; fax: +47 2322 5023.
E-mail addresses: anners.lerdal@medisin.uio.no,
anners.lerdal@lds.no (A. Lerdal), anders.kottorp@ki.se (A. Kottorp),
caryl.gay@ucsf.edu (C.L. Gay), e.k.grov@medisin.uio.no (E.K. Grov),
kathryn.lee@nursing.ucsf.edu (K.A. Lee).

to improve the instrument (Kendel et al., 2010). If Rasch model assertions are supported by empirical data, individual raw scores from empirically validated items can be converted into valid equal-interval measures to be used for further analysis (e.g., comparison to standard cut-offs). Rasch models can also identify differential item functioning (DIF), which occurs when peoples' ratings on particular items differ systematically according to specific characteristics (e.g. socio-demographic or clinical factors) relative to their ratings on other items. Prior studies have reported contradictory findings regarding PSD's relationship with socio-

Table 1
Overview of the analytic process using a Rasch model approach.

Steps	Psychometric property	Statistical approach and criteria	Original BDI-II (21 items) results	Reduced BDI-II (16 items) results[a]
1	**Rating scale functioning**: Does the rating scale function consistently across items?	• Average measures for each step category on each item should advance monotonically • z-values < 2.0 in outfit mean square ($MnSq$) values for step category calibrations Linacre (2002)	• All items met criteria • The outfit $MnSq$ z-values ranged from 0.80 to 1.68	
2	**Internal scale validity**: How well do the actual item responses match the expected responses from the Rasch model?	Item goodness-of-fit statistics (Smith et al., 1998)[b] • Infit $MnSq$ values ≤ 1.2 or • Infit $MnSq$ values ≤ 1.4 and z-value < 2.0	• 5 items (21, 16, 17, 10, 18) failed to meet criteria[c]. • See Table 2 for infit $MnSq$ statistics	• All items met criteria
3	**Internal scale validity**: Is the scale unidimensional?	Principal component analysis of the residuals (Linacre, 2011) • $\geq 50\%$ of total variance explained by first component (depression) • Any additional component explains $< 5\%$ of the remaining variance of residuals after removing first component	• First component explained 38.0% of total variance • Second component explained 7.3% of total variance	• First component explained 51.8% of total variance • Second component explained 6.8% of total variance
4	**Person-response validity**: How well do the actual individual responses match the expected responses from the Rasch model?	Person goodness-of-fit statistics • Person goodness-of-fit infit $MnSq$ values < 1.5 associated with a z-value < 2.0 (Kottorp, 2003) • $\leq 5\%$ of sample fails to demonstrate acceptable goodness-of-fit values (Kottorp, 2003)	• 5.6% of sample failed to demonstrate acceptable goodness-of-fit values	• 1.9% of sample failed to demonstrate acceptable goodness-of-fit values
5	**Person-separation reliability**: Can the scale distinguish at least 3 distinct groups of depression in the sample tested?	Person-separation index (Fisher, 1992) • ≥ 2.0	• 1.99	• 2.06
6	**Internal consistency**: Are item responses consistent with each other?	• Cronbach alpha coefficient • > 0.8 (Fisher, 1992)	• 0.90	• 0.91
7	**Differential item functioning (DIF)**: Are item difficulty calibrations stable in relation to demographic and clinical variables?	• Mantel–Haenszel statistic • $p < 0.01$ with Bonferroni correction (Mantel, 1963)	• Item 15 (*loss of energy*) had DIF in relation to lesion type (infarct vs hemorrhage) • People with hemorrhage had relatively higher scores on item 15 as compared to people with infarct (item DIF contrast=13.42)	• Item 15 (*loss of energy*) had DIF in relation to age ($<$ or ≥ 60 years old) • People < 60 years old had relatively higher scores on item 15 as compared to people ≥ 60 years old (item DIF contrast=13.74)
8	**Differential test functioning (DTF)**: How consistent are the scores from the modified and original BDI-II versions?	• z-score differences < 1.96 • Pearson correlation $r > 0.80$ and $p < 0.05$	• 2 measures (1.9%) had z-scores exceeding ± 1.96 • $r=0.981$, $p < 0.01$	

[a] After initial evaluation of the original 21-item BDI-II, a stepwise process was used whereby items failing to meet criteria were removed one at a time, and only those meeting criteria in earlier steps advanced to subsequent steps. If more than one item failed to meet a criterion at a given step, the item with the worst fit was removed and the step was repeated with the remaining items. Through this process, a 16-item version omitting items 10, 16, 17, 18 and 21 was idenitified as meeting almost all criteria and is included in the last column.

[b] We focus on infit rather than outfit statistics because they are usually considered more informative in analyses of goodness-of-fit, as they focus on the degree of fit in the most typical observations in the data, and are therefore less sensitive to extreme scores endorsed by only a few participants (McNamara, 1996).

[c] The five misfitting items did not demonstrate misfit in the first iteration, some emerged in subsequent iterations; items are listed in the order of removal and the $MnSq$ values shown reflect the iteration prior to the item's removal as provided in Table 2.

demographic factors (Kouwenhoven et al., 2011; Aben et al., 2002), and thus, the possibility of DIF warrants consideration.

A number of screening instruments have been used to assess depression among stroke patients (Aben et al., 2002; Turner et al., 2012; Kang et al., 2013). Although the Beck Depression Inventory-II (BDI-II) (Beck et al., 1996) is one of the more widely used instruments (Haghgoo et al., 2013; Weaver et al., 2013) and has acceptable validity for depression screening for among stroke patients (Aben et al., 2002), it has not been evaluated using Rasch analysis in this specific population. When the BDI-II was evaluated in a broader population of patients with neurological disease (Siegert et al., 2010), participant responses demonstrated good overall fit to the Rasch model when three items were deleted. However, it is unclear whether these findings apply to stroke patients specifically. Thus, the aim of this study was to assess the BDI-II's psychometric properties in a sample of first-ever stroke patients using a Rasch model approach.

2. Methods

2.1. Sample and procedures

The study was approved by the Regional Medical Research Ethics Committee of Health East of Norway, the Norwegian Data Inspectorate, and the two hospitals from which patients were recruited. In 2007–2008, patients with first-ever stroke admitted to one of two hospitals in Eastern Norway were asked to participate in a previously described longitudinal fatigue study (Lerdal et al., 2011). Included patients had first-ever clinical presentation of stroke (World Health Organization, 2007), were 18 years or older, had sufficient cognitive and communication skills to participate, and provided written informed consent. Data were collected from medical records and validated questionnaires; 6-month follow-up data were used for this analysis to minimize the influence of somatic symptoms (Kang et al., 2013).

2.2. Instrument

The BDI-II (Beck et al., 1996) was used to measure depressive symptom severity, with interviewer assistance when needed. It consists of 21 groups of four or seven statements, and patients select the statements that best reflect how they felt for the past 2 weeks. Scores range 0–63, with higher scores indicating more severe symptoms.

2.3. Statistical analysis

A Rasch, partial-credit model application was used to analyze BDI-II raw scores using WINSTEPS Rasch analysis software, version 3.69.1.16 (Linacre, 2010). The step-wise analytic approach and criteria for evaluating each psychometric property are summarized in Table 1 and have been previously described (Lerdal and Kottorp, 2011; Linacre, 2002, 2011; Mantel, 1963, Smith et al., 1998).

3. Results

3.1. Participants

Of the 193 patients screened for eligibility, 115 completed a baseline assessment, and the 106 with 6-month follow-up data were included in the analysis. The 9 participants who did not complete the 6-month assessment did not differ from those who did with respect to baseline depressive symptoms. The final sample was 40% male, 70% married/cohabitating, 25% had paid employment, and 29% had at least 12 years of education. Mean age was 68 ± 13 years (range 29–91), and 80% were 60 years of age or older. Based on MRI or CT, stroke type was classified as infarct (63%), hemorrhage (8%), or unknown (29%). Of the 67 with valid data on stroke location, 48% were right hemisphere, 46% were left, and 6% were bilateral. Mean BDI score was 10.4 ± 8.3 (range 0–52).

3.2. Analysis of the BDI-II's psychometric properties

Results for each step of the Rasch analysis are presented in Table 1. All 21 items met criteria for acceptable rating scale functioning. Five items failed to demonstrate acceptable item goodness-of-fit (see Tables 1 and 2), while the remaining 16 items demonstrated acceptable item goodness-of-fit (see Table 2). For subsequent analytic steps, we evaluated psychometric properties of both the original 21-item BDI-II and a 16-item version omitting

Table 2
Summary of iterations and item difficulty calibrations (including fit statistics) and item-total correlations (point-biserial correlations).

Iteration	Item #		Measure	SE	Infit MnSq	Item-total correlation
1	#21	Loss of interest in sex[a]	39.91	1.16	1.72	0.33
2	#16	Changes in sleeping patterns[a]	46.29	1.71	1.56	0.24
3	#17	Irritability[a]	45.11	1.52	1.48	0.40
4	#10	Crying[a]	51.95	1.97	1.54	0.30
5	#18	Changes in appetite[a]	50.22	1.88	1.29	0.47
6	#1	Sadness	59.25	2.47	0.86	0.61
Less frequent symptoms	#9	Suicidal thoughts	57.66	2.93	0.74	0.61
	#5	Guilty feelings	57.51	2.27	1.10	0.51
	#12	Loss of interest	56.43	2.09	1.18	0.50
	#3	Past failure	56.22	2.04	1.00	0.58
	#6	Punishment feelings	52.39	2.29	0.50	0.67
	#8	Self-criticalness	51.84	1.99	0.91	0.64
	#7	Self-dislike	50.30	1.95	0.76	0.73
	#11	Agitation	49.71	2.22	1.16	0.44
	#14	Worthlessness	49.71	2.01	1.05	0.60
	#2	Pessimism	49.43	1.93	0.84	0.68
	#13	Indecisiveness	46.70	1.85	0.83	0.71
	#19	Concentration difficulty	46.41	2.07	1.16	0.49
	#4	Loss of pleasure	45.67	1.75	1.20	0.53
	#20	Tiredness or fatigue	36.36	1.78	1.00	0.61
More frequent symptoms	#15	Loss of energy	34.42	1.99	1.13	0.54

[a] Item excluded from next iteration for failing to meet criteria [28]. Item #18 also had a z-value < 2.0.

the 5 misfitting items. Principal components analysis indicated that only the 16-item version met the criterion of $\geq 50\%$ of the variance explained by the first dimension, and both versions failed the criterion of $< 5\%$ of the variance explained by the second dimension. The 16-item version demonstrated better person goodness-of-fit than the original 21-item BDI-II, which failed to meet the criterion of $\leq 5\%$ of respondents having poor fit. Despite having fewer items, the 16-item version also demonstrated better person-separation reliability than the original BDI-II. Cronbach alpha coefficients for both BDI-II versions were acceptable. Of the original 21 items, item 15 (*loss of energy*) demonstrated DIF in relation to lesion type; patients with hemorrhage were more likely to experience this symptom compared to those with infarct. None of the other original 21 items showed significant DIF in relation to age, gender, cohabitant status, education, or stroke location. In the 16-item version, however, item 15 again demonstrated DIF but now in relation to age, with it being relatively easier to endorse among older patients compared with younger patients. Finally, comparison of the individual measures from the 21-item and 16-item versions indicated that the two versions generated similar measures for most respondents, and the measures were highly correlated.

4. Discussion

The main finding from this Rasch-analysis of the BDI-II in a sample of first-ever stroke patients was that five items failed to demonstrate acceptable goodness-of-fit to the Rasch model and might be considered for omission to improve the instrument's psychometric properties with this population. A 16-item version excluding these items demonstrates unidimensionality and seems to generate a more internally valid estimate of depression severity than the original 21-item BDI-II, and therefore may be a more appropriate scale for patients following stroke. Furthermore, 3 of the 5 items that did not show acceptable item goodness-of-fit consist of 4 response statements and 2 consist of 7 statements, so excluding these 5 items reduces the instrument length by 26 response statements. Reducing response burden is particularly beneficial for patients who tire easily.

Somewhat consistent with our findings, a prior Rasch analysis (Siegert et al., 2010) examining patients with different neurological diseases also recommended excluding three of the five BDI-II items (16, 18, and 21) excluded in the current study. Items 16 and 18 were modified from the original BDI to include 7 response statements capturing both increases and decreases in sleep and appetite (Beck et al., 1996). Since increases and decreases are coded similarly, higher values reflect only the amount of change in sleep or appetite, not the direction of change. However, the poor fit of these items across two samples may indicate a lack of empirical support for this revised coding. Only our study excluded items 10 and 17, but it is unclear whether the differences between the two studies are related to chance or to clinical diagnosis. Together, these studies provide evidence that items 16, 18 and 21 do not contribute to a unidimensional measure of depression among patients with neurological conditions or the more specific population of stroke patients. Larger studies with more homogeneous neurological samples may reveal whether there are systematic explanations for the other misfitting items (e.g., diagnostic profiles) (Kottorp et al., 2003).

The presence of DIF in the *loss of energy* item in both the 21- and 16-item versions, albeit regarding different patient characteristics, suggests that *loss of energy* may be experienced differently across various groups relative to other BDI-II items. It is also possible that the DIF occurred by chance and thus has minimal impact on the BDI-II's validity across lesion types and age groups.

The findings of this study need to be considered in light of its limitations. Findings are limited to patients 6 months following first-ever stroke, and additional studies are needed to evaluate the BDI-II's psychometric properties among patients at different stages of recovery, with recurrent stroke, and using more detailed stroke diagnostics. Given the lack of unidimensionality in the original BDI-II, the cut-off scores used to indicate depression severity may warrant further evaluation among stroke patients. Because this study did not include a clinical evaluation for depression, it remains unclear whether the current cut-off scores were clinically useful in this sample of stroke survivors and what adjustments, if any, would be needed for use with the 16-item version. The study is also limited by the relatively small sample size for this type of analysis, particularly for evaluating DIF. Replication in larger samples is warranted. Furthermore, the influence of stroke severity on the BDI-II's psychometric properties was not evaluated in this study, but warrants consideration in future research given recent studies indicating stroke severity's association with post-stroke depression (Hackett and Anderson, 2005; De Ryck et al., 2013) and influence on a screening instrument's psychometric properties, particularly one containing somatic items such as the BDI (Kang et al., 2013). Finally, optimizing some of the psychometric properties of the BDI-II based on the Rasch model may result in the exclusion of clinically valuable information from the instrument. Because screening instruments are typically designed to generate a single measure of an underlying unidimensional construct, we recommend that future studies consider excluding the five items that posed threats to the BDI-II's validity and reliability as a measure of depression among stroke survivors. However, this suggestion does not mean that these items are not relevant to the diagnosis of depression, as these symptoms likely still have considerable clinical relevance and thus warrant systematic assessment as part of the clinical evaluation process.

5. Conclusion

Assessment of psychometric properties in a variety of samples is important for demonstrating an instrument's validity, especially when comparing findings across studies and for a scale originally developed with psychiatric patients (Beck et al., 1988) and now often used with somatic patients. The findings of this explorative study and those of Siegert et al. (2010) raise concern about the original 21-item BDI-II for depression screening among stroke survivors. Given the questionable validity of the summed scores, use of the 21-item BDI-II may lead to over- or under-estimation of depression among people with stroke or other neurological conditions. A 16-item version that omits items 10, 16, 17, 18 and 21 may be more appropriate than the original version as a unidimensional measure of depression among stroke patients and would likely reduce respondent burden. Omission of these five items might also increase the BDI-II's ability to distinguish between patients with low, medium, and high levels of depression, although further research is needed to determine appropriate cut-off scores for the 16-item version.

Conflict of interest

None of the authors has a conflict of interest to disclose.

Acknowledgments

References

Aben, I., Verhey, F., Lousberg, R., Lodder, J., Honig, A., 2002. Validity of the beck depression inventory, hospital anxiety and depression scale, SCL-90, and hamilton depression rating scale as screening instruments for depression in stroke patients. Psychosomatics 43, 386–393.

Beck, A.T., Steer, R.A., Garbin, M.G., 1988. Psychometric properties of the Beck depression inventory: twenty-five years of evaluation. Clin. Psychol. Rev. 8, 77–100.

Beck, A.T., Steer, R.A., Brown, G.K., 1996. BDI-II, Beck Depression Inventory Manual. Psychological Corporation, San Antonio, TX

De Ryck, A., Brouns, R., Fransen, E., Geurden, M., Van, G.G., Wilssens, I., De Ceulaer, L., Marien, P., De Deyn, P.P., Engelborghs, S., 2013. A prospective study on the prevalence and risk factors of poststroke depression. Cerebrovasc. Dis. Extra 3, 1–13.

Fisher, A.G., 1993. The assessment of IADL motor skills: an application of many-faceted Rasch analysis. Am. J. Occup. Ther. 47, 319–329.

Fisher, W., 1992. Reliability statistics. Rasch Meas. Trans. 6, 238.

Forkmann, T., Gauggel, S., Spangenberg, L., Brahler, E., Glaesmer, H., 2013. Dimensional assessment of depressive severity in the elderly general population: psychometric evaluation of the PHQ-9 using Rasch Analysis. J. Affect. Disord. 148, 323–330.

Gainotti, G., Antonucci, G., Marra, C., Paolucci, S., 2001. Relation between depression after stroke, antidepressant therapy, and functional recovery. J. Neurol. Neurosurg. Psychiatry 71, 258–261.

Hackett, M.L., Anderson, C.S., 2005. Predictors of depression after stroke: a systematic review of observational studies. Stroke 36, 2296–2301.

Hackett, M.L., Yapa, C., Parag, V., Anderson, C.S., 2005. Frequency of depression after stroke: a systematic review of observational studies. Stroke 36, 1330–1340.

Kang, H.J., Stewart, R., Kim, J.M., Jang, J.E., Kim, S.Y., Bae, K.Y., Kim, S.W., Shin, I.S., Park, M.S., Cho, K.H., Yoon, J.S., 2013. Comparative validity of depression assessment scales for screening poststroke depression. J. Affect. Disord. 147, 186–191.

Kendel, F., Wirtz, M., Dunkel, A., Lehmkuhl, E., Hetzer, R., Regitz-Zagrosek, V., 2010. Screening for depression: Rasch analysis of the dimensional structure of the PHQ-9 and the HADS-D. J. Affect. Disord. 122, 241–246.

Kottorp, A., 2003. Occupation-Based Evaluation and Intervention. Validity of the Assessment of Motor and Process Skills when Used with Persons with Mental Retardation (Umeå university medical dissertation). Umeå University, Umeå, Sweden.

Kottorp, A., Bernspang, B., Fisher, A.G., 2003. Validity of a performance assessment of activities of daily living for people with developmental disabilities. J. Intell. Disabil. Res. 47, 597–605.

Kouwenhoven, S.E., Kirkevold, M., Engedal, K., Kim, H.S., 2011. Depression in acute stroke: prevalence, dominant symptoms and associated factors. A systematic literature review. Disabil. Rehabil. 33, 539–556.

Lerdal, A., Bakken, L.N., Rasmussen, E.F., Beiermann, C., Ryen, S., Pynten, S., Drefvelin, A.S., Dahl, A.M., Rognstad, G., Finset, A., Lee, K.A., Kim, H.S., 2011. Physical impairment, depressive symptoms and pre-stroke fatigue are related to fatigue in the acute phase after stroke. Disabil. Rehabil. 33, 334–342.

Lerdal, A., Kottorp, A., 2011. Psychometric properties of the Fatigue Severity Scale – Rasch analyses of individual responses in a Norwegian stroke cohort. Int. J. Nurs. Stud. 48, 1258–1265.

Linacre, J.M., 2002. Optimizing rating scale category effectiveness. J. Appl. Meas. 3, 85–106.

Linacre, J.M., 2010. Winsteps – Rasch Model Computer Program. Version 3.69.1.16. ⟨www.winsteps.com⟩ Retrieved June 2010.

Linacre, J.M., 2011. A User's Guide to Winstep. Ministep Rasch-Model Computer Programs. Program Manual 3.73.0.

Mantel, N., 1963. Chi-square tests with one degree of freedom: extensions of the Mantel Haenszel procedure. J. Am. Stat. Assoc. 58, 690–700.

McNamara, T., 1996. Measuring Second Language Performance. Addison Wesley Longman, London

Merbitz, C., Morris, J., Grip, J.C., 1989. Ordinal scales and foundations of misinference. Arch. Phys. Med. Rehabil. 70, 308–312.

Siegert, R.J., Tennant, A., Turner-Stokes, L., 2010. Rasch analysis of the Beck Depression Inventory-II in a neurological rehabilitation sample. Disabil. Rehabil. 32, 8–17.

Smith, R.M., Schumacker, R.E., Bush, M.J., 1998. Using item mean squares to evaluate fit to the Rasch model. J. Outcome Meas. 2, 66–78.

Sturm, J.W., Donnan, G.A., Dewey, H.M., Macdonell, R.A., Gilligan, A.K., Srikanth, V., Thrift, A.G., 2004. Quality of life after stroke: the North East Melbourne Stroke Incidence Study (NEMESIS). Stroke 35, 2340–2345.

Tesio, L., 2003. Measuring behaviours and perceptions: Rasch analysis as a tool for rehabilitation research. J. Rehabil. Med. 35, 105–115.

Turner, A., Hambridge, J., White, J., Carter, G., Clover, K., Nelson, L., Hackett, M., 2012. Depression screening in stroke: a comparison of alternative measures with the structured diagnostic interview for the diagnostic and statistical manual of mental disorders, fourth edition (major depressive episode) as criterion standard. Stroke 43, 1000–1005.

Williams, L.S., Ghose, S.S., Swindle, R.W., 2004. Depression and other mental health diagnoses increase mortality risk after ischemic stroke. Am. J. Psychiatry 161, 1090–1095.

World Health Organization, 2007. International Classification of Diseases (ICD). apps.who.int/classifications/apps/icd/icd10online Version for 2007 [Downloaded 13.04.2011].

Wright, B.D., Linacre, J.M., 1989. Observations are always ordinal; measurements, however, must be interval. Arch. Phys. Med. Rehabil. 70, 857–860.