



The influence of the administration method on scores of the 15-item Geriatric Depression Scale in old age

Margot W.M. de Waal^{a,*}, Gerda M. van der Weele^a, Roos C. van der Mast^b,
Willem J.J. Assendelft^a, Jacobijn Gussekloo^a

^a Department of Public Health and Primary Care (VO-P), Leiden University Medical Center, Leiden, The Netherlands

^b Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands

ARTICLE INFO

Article history:

Received 28 June 2010

Received in revised form 26 August 2011

Accepted 28 August 2011

Keywords:

Rating scales

Agreement

Depression

Psychometrics

ABSTRACT

Many rating scales can be self-administered or interviewer-administered, and the influence of administration method on scores is unclear. We aimed to study this influence on scores of the Geriatric Depression Scale (GDS-15), used as a screening instrument in general practice. In two general practices 376 registered patients aged 75 years and older were asked to participate. Exclusion criteria were dementia and current treatment for depression. The GDS-15 was administered twice within 1 month: self-administered by mail, and interviewer-administered during home visits. The sequence of administering the methods was different for the two practices. We analyzed differences in total and item GDS-scores. Of 141 subjects who participated (response rate 55%) 59 were men (42%). Mean age was 81.4 years (SD 4.8). When the GDS-15 was self-administered, 33 subjects (23.4%) left items unanswered. There were no items unanswered when the GDS-15 was interviewer-administered. On average the self-administered total GDS scores were 0.70 points higher than interviewer-administered scores (95% confidence interval = 0.41; 0.98), with a large range of variation in the scores (limits of agreement = −2.69 to 4.08). Item-item comparisons showed high percentages of agreement. Chance-corrected agreement (kappa) was moderate to fair, but three items showed only slight agreement (kappa values <0.21). In conclusion, compared to interviewer-administered scores, scores on the GDS-15 when self-administered were higher. The method of administration should be taken into account when interpreting scores.

© 2012 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Although depressive symptoms in old age have serious negative consequences and effective treatment is available, depressed older subjects are often not treated. Combined screening and treatment programs are being advocated to enhance recognition and to treat depressive symptoms in general practice more adequately (Pignone et al., 2002).

The Geriatric Depression Scale (GDS) is used frequently to screen for depressive symptoms in old age. It was originally developed as a 30-item self-rating scale with answers in simple yes/no format (Yesavage et al., 1982). The shortened 15-item version is considered to be more acceptable as a screening tool, given the shorter administration time (Sheikh and Yesavage, 1986). In a recent meta-analysis of the diagnostic validity and added value of the GDS in primary care,

the GDS-15 had adequate sensitivity and specificity and had good clinical utility as a screening test (Mitchell et al., 2010). The original instruction of Yesavage et al. was a combination of two methods of administration, stating that ‘patients who cannot complete the questionnaire unaided, have the questions read out to them’.

Little is written about the common practice of method of administration. In a literature search we found 12 studies in primary care in which the GDS-15 was used for screening purposes among persons aged 65 and over. In nine of these studies the GDS-15 was administered by interview (D'Ath et al., 1994; Iliffe et al., 1994; Noltorp et al., 1998; Whooley et al., 2000; Arthur et al., 2002; Freudenstein et al., 2002; Stek et al., 2004; Olivera et al., 2008; Weyerer et al., 2008) and in three studies by mail (Osborn et al., 2002; Harris et al., 2003; Licht-Strunk et al., 2005). Since several studies suggest that the method of administration influences the scores of scales (O'Neill et al., 1992; Geerlings et al., 1999; Smeeth et al., 2001), we questioned whether self-administration of the GDS-15 by mail would give comparable results as interviewer-administration when used to screen older subjects in general practice. Therefore we studied the influence of administering method on item and total scores of the GDS-15 among subjects aged 75 years and older in general practice.

* Corresponding author at: Leiden University Medical Center, Department of Public Health and Primary Care (VO-P), PO Box 9600, 2300 RC Leiden, The Netherlands. Tel.: +31 71 5268444; fax: +31 71 5268259.

E-mail address: m.w.m.de-waal@lumc.nl (M.W.M. de Waal).

2. Methods

2.1. Subjects

In two general practices in The Netherlands, in the cities Leiden and Katwijk, registered patients aged 75 years and over were asked to participate. General practitioners (GPs) excluded patients with current treatment for depression, severe cognitive dysfunction (diagnosis of dementia or Alzheimer disease or clinically known Mini Mental State Examination (MMSE) scores <19 points), loss of partner or child within the last 3 months, a life expectancy of less than 3 months, or patients who do not speak Dutch. All participants were visited at their own home, during which all exclusion criteria were checked. For this study, we further excluded all participants with MMSE scores below 24 points to minimize the influence of cognitive dysfunction (Korner et al., 2007; Lach et al., 2010).

2.2. Measurement of depressive symptoms

To screen for the presence of depressive symptoms, the 15-item Geriatric Depression Scale (GDS-15) was used (Sheikh and Yesavage, 1986). The answers are in a yes/no response format. The total depression score ranges from 0 to 15 points, with higher scores indicating more depressive symptoms. In this study a score of 5 points or higher was considered as clinically relevant (D'Ath et al., 1994).

The GDS-15 was administered twice: self-administered by mail, and interviewer-administered during home visits. The sequence of administering methods was different for the two practices. In the first general practice, subjects were invited by mail to complete and return the GDS, with one postal reminder after 2 weeks. After the self-administered GDS-15 was returned by mail, subjects were contacted for a home visit in which trained interviewers administered the GDS-15 a second time. The interviewers were kept blind to the scores on the self-administered GDS. In the interviews all questions were read out to the participant, and on request some additional explanation was given, e.g. as to time frame ('last month' as stated in the introduction of both written and interviewer versions) or reference group ('of same age', not stated in the introduction). In the second general practice, subjects were invited by letter to participate, with one postal reminder after 2 weeks. When the response card was returned by mail, subjects were contacted for a home visit in which interviewers administered the GDS-15. Two weeks after this home visit, subjects were asked by mail to complete and return the GDS-15 with one postal reminder after 2 weeks.

There were five well-trained interviewers, three of them performed the majority of interviews in both practices (for practice A: 41%, 18% and 27%; and for practice B: 31%, 16% and 53%), and two others only performed interviews in the first practice (resp. 1% and 13%).

2.3. Further measurements

Cognitive functioning was measured using the Mini Mental State Examination (MMSE) during the home visit mentioned above (Folstein et al., 1975). Scores range from 0 to 30 points, with lower scores indicating increasing cognitive impairment. A score below 24 points indicates cognitive impairment (Kempen et al., 1995). Finally, questions were added about education, income and living situation.

2.4. Statistical analyses

To compute the GDS-15 total score, first we interpreted all missing items as 'not-depressed' (0 points). We chose to do so because the GDS-15 total score generally has a low positive predictive value, which would even be lower by interpreting missing items as 'depressed'. Secondly, we computed the GDS-15 total scores using prorating of scores to check whether this would make a difference: for each missing item the average score of completed items per individual was imputed and was added to the total score of completed items (<http://www.stanford.edu/~yesavage/GDS.html>; visited November 14th 2010). This imputation assumes that items are 'missing at random'.

We calculated Cronbach's alpha as a measure of internal coherence of the questionnaire (Bland and Altman, 1997).

For visual judgment of agreement of the two administering methods, differences in scores were plotted using the Bland–Altman method (Bland and Altman, 1986). In this plot the x-axis represents the average between the first and second GDS-15 total score, and the y-axis represents the mean difference in scores for the whole sample with 95% limits of agreement (mean difference \pm 1.96 standard deviation of the mean difference). Since we did not randomly assign sequence of administration (and interviewer) to subjects, we corrected for an uneven distribution of patient characteristics by stratified analysis of difference in scores on age, education, cognitive functioning (24–26 versus >26), sequence of administration, time between measurements, and interviewer.

Differences in item scores were analyzed with percentage of agreement (unadjusted agreement) and kappa. Kappa takes into account the agreement occurring by chance, thus representing a measure of agreement beyond chance. Kappa varies from -1 to $+1$, and agreement is considered poor when kappa is less than 0.00, slight 0.00–0.21, fair 0.21–0.40, moderate 0.41–0.60, substantial 0.61–0.80, and almost perfect 0.81–1.00 (Landis and Koch, 1977).

3. Results

3.1. Study population

In general practice A, 69 (32%) out of 218 enlisted older subjects were excluded and 82 out of 149 remaining subjects participated (response rate 55%). In general practice B, 46 (29%) out of 158 enlisted older subjects were excluded, and 59 out of 112 remaining subjects participated (response rate 53%). The mean age of all participants was 81.4 years (S.D. 4.8), 59 (42%) were male, and 82 (58%) lived alone. Low cognitive functioning according to an MMSE-score below 27 points was found in 23 (16%) subjects. Comparing both general practices, subjects in general practice A were older and had a lower level of education (see Table 1). The time between administering the two GDS-15 scales was at average 30 days (95% CI = 27–33). For 45% of the subjects time between both tests was within 21 days: in practice A this was 28% and for practice B this was 68%.

3.2. Influence of administration method on GDS scores

When the GDS-15 was self-administered, 33 subjects (23.4%) left some items unanswered, of which four subjects (2.5%) left five or more items unanswered. There were no items unanswered when the GDS-15 was interviewer-administered.

The internal coherence (Cronbach's alpha) was 0.77 for the self-administered GDS-15 and 0.69 for the interviewer-administered GDS-15. Table 2 shows that the mean difference in total GDS-15 score was 0.70 points (95% CI (0.41; 0.98)), i.e. total GDS-15 scores were on average 0.70 points higher when the GDS was self-administered compared to interviewer-administered. Using a GDS-15 cut-off score of ≥ 5 points, 16 subjects (11.3%, 95% CI (6.1; 16.7)) were considered to have clinically relevant depressive symptoms on the self-administered GDS-15 and five subjects (3.5%, 95% CI (0.0; 7.0)) on the interviewer-administered GDS. Prorating of scores, instead of interpreting missing items as 'not-depressed', gave almost similar results: median GDS total score was 1.0 (IQR 0; 2.5) when self-administered, the mean difference between self-administered and interview-administered was 0.76 (CI 95% 0.47; 1.05), and GDS total score ≥ 5 for 12.1% of subjects ($n = 17$).

The Bland–Altman plot (Fig. 1) shows a large range of agreement in individual scores (S.D. = 1.73; 95% limits of agreement (-2.69 ; 4.09)). The mean difference lies above the null line, indicating a bias towards a higher score on self-administered GDS-15 total score. Eight subjects could be seen as 'outliers', as they fell outside the limits of agreement. Their characteristics did not differ from the characteristics of the total group; seven of these eight subjects had MMSE scores above 26. Stratified analyses on age, education, cognitive functioning, sequence of administration, time between measurements, and interviewer showed that GDS-15 total scores were higher in the self-administered version in all defined subgroups. (See Table 3.)

Table 4 shows for self-administered GDS per item the percentage of missing answers, the percentage of depressive answers per

Table 1
Sociodemographics and cognitive functioning of all study subjects ($n = 141$).

	Practice A		Practice B		P-value ^a
	Self/int.-adm.		Int./self.-adm.		
	(n = 82)		(n = 59)		
Sociodemographics					
Age: ≥80 years	54	67%	25	49%	0.006
Gender: male	32	39%	27	46%	
Education: basic schooling only	15	22%	23	56%	0.006
Income: social security only	11	16%	7	13%	
Living situation: alone	47	58%	27	50%	
Cognitive functioning:					
MMSE scores 24–26	11	13%	12	20%	
MMSE median score (IQR)	29 (28–30)		28 (27–29)		

^a Chi-square test.

Table 2Characteristics of the GDS-15 among all study subjects ($n = 141$).

	Study subjects
GDS total score	
Self-administered#, median (IQR)	1.0 (0; 2)
Interviewer-administered, median (IQR)	0.0 (0; 2)
Difference between self-administered and interviewer-administered GDS total score	
Mean difference (± 1.96 S.E. = 95% CI)	0.70 (0.41; 0.98)
Mean difference (± 1.96 S.D. = limits of agreement) ^a	0.70 (−2.69; 4.08)
GDS total score ≥ 5	
Self-administered ^b $n = 16$	11.3% (6.1; 16.7)
Interviewer-administered $n = 5$	3.5% (0.0; 7.0)

IQR = interquartile range.

95% CI = 95% confidence intervals (mean ± 1.96 S.E.).^a According to Bland and Altman (1986).^b To compute the total score, missing items were interpreted as 'not-depressed'.

administration method and agreement between administration methods per item. On items 2, 9 and 10, subjects gave far more often a depressive answer when the GDS was self-administered than when it was interviewer-administered; e.g. on item 10 'Do you feel you have more problems with memory than most other people?', 30 subjects (21.3%) gave a depressive answer when self-administered, but only 10 subjects (7.1%) when interviewer-administered. Item-item comparisons showed high unadjusted percentages of agreement. Kappa values showed that chance-corrected agreement was moderate to fair, but items 1, 3, and 12 showed poorest agreement (values < 0.21). Kappa is dependent on observed prevalence rates of marginal totals per item, and therefore kappa values inherently vary across the items. Consequently, for these items, which had a high agreement, one should interpret the value of kappa with caution, because of the so-called high agreement but low kappa paradoxes (Feinstein and Cicchetti, 1990).

4. Discussion

We questioned whether self-administration of the GDS-15 by mail would give comparable results as interviewer-administration when

Table 3

Stratified analyses for differences in GDS-15 total scores between self-administered and interviewer-administered.

	N	Difference in total score between self-administered and interviewer-administered Mean difference (CI 95%)
Age group		
75–79 years	62	0.39 (0.001; 0.71)
80 years and over	79	0.94 (0.52; 1.35)
Gender		
Male	59	0.78 (0.43; 1.13)
Female	82	0.63 (0.20; 1.07)
Education		
Basic level (max 6 years)	38	1.18 (0.55; 1.82)
More than basic level	103	0.51 (0.20; 0.83)
Cognitive functioning		
MMSE scores 24 to 26	23	1.48 (0.72; 2.24)
MMSE scores 27 or higher	118	0.54 (0.23; 0.85)
Interviewers		
AC	51	0.55 (0.13; 0.97)
CM	1	3.00
EH	24	0.71 (−0.30; 1.72)
IM	55	0.60 (0.20; 1.00)
PT	10	1.70 (0.19; 3.21)
Order of administration		
1st self, 2nd interviewer (practice B)	82	0.80 (0.40; 1.21)
1st interviewer, 2nd self (practice A)	59	0.54 (0.14; 0.94)
Time between test 1 and test 2		
Within 3 weeks	78	0.50 (0.10; 0.90)
Longer than 3 weeks	63	0.94 (0.52; 1.36)

used to screen subjects aged 75 years and over in general practice. On average total depression scores were 0.70 points higher when the GDS-15 was self-administered than when interviewer-administered, with a large variation between subjects (limits of agreement (−2.69;

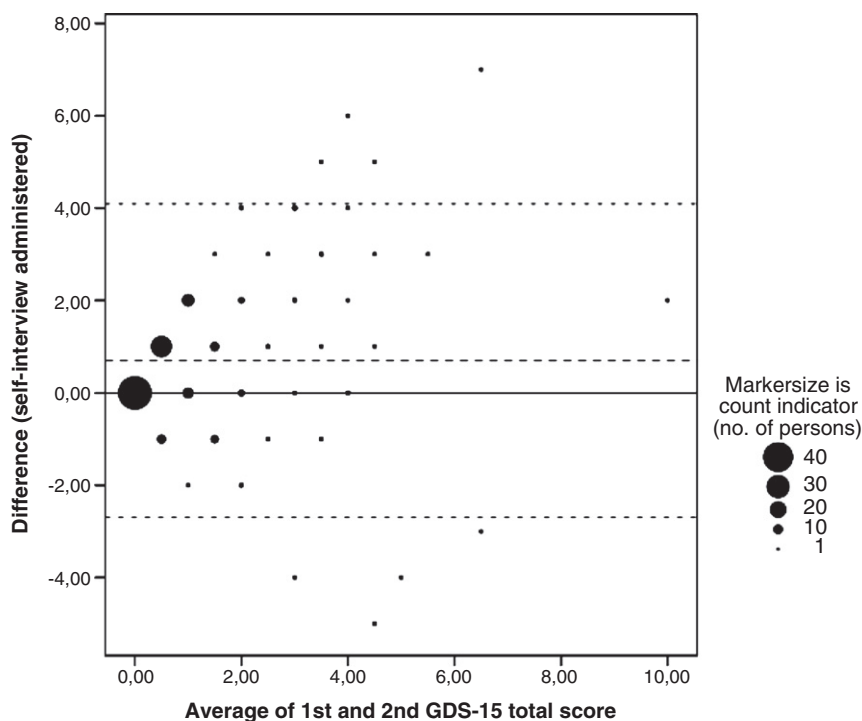


Fig. 1. Plot of mean of self-administered and interviewer-administered GDS-15 total score (x-axis) and difference between self-administered and interviewer-administered total score (y-axis). Dotted lines are limits of agreement, i.e. 95% confidence interval of mean difference in scores (± 1.96 S.D.).

Table 4

Answers on each GDS-15 item self-administered or interviewer-administered ($n = 141$), with unadjusted percentage of agreement and kappa. Items are ordered by kappa. Noteworthy numbers are highlighted in bold.

	Self-administered		Interviewer-administered	Item–item agreement	
	No answer (%)	Depressive answer (%)	Depressive answer (%)	Unadjusted percentage agreement	Kappa ^a
14 Do you feel that your situation is hopeless?	2	4	5	96	0.48
13 Do you feel full of energy?	7	27	22	79	0.45
11 Do you think it is wonderful to be alive now?	1	2	4	96	0.43
8 Do you often feel helpless [hopeless]?	4	4	2	96	0.43
6 Are you afraid that something bad is going to happen to you?	6	22	18	81	0.41
4 Do you often get bored?	2	1	2	98	0.39
5 Are you in good spirits most of the time?	3	4	1	96	0.28
10 Do you feel you have more problems with memory than most?	7	21	7	82	0.27
2 Have you dropped many of your activities and interests [last month]?	1	16	4	87	0.25
9 Do you prefer to stay at home, rather than going out and doing new things?	7	45	15	65	0.23
15 Do you think that most people are better off than you are?	4	8	4	92	0.21
7 Do you feel happy most of the time?	4	6	5	91	0.21
3 Do you feel that your life is empty?	1	6	8	89	0.16
12 Do you feel pretty worthless the way you are now?	2	9	4	89	0.07
1 Are you basically satisfied with your life?	3	3	7	91	–0.03

^a Kappa: poor <0.00, slight 0.00–0.21, fair 0.21–0.40, moderate 0.41–0.60.

4.09)). On some items subjects gave a depressive answer far more often when self-administered than when interviewer-administered. Apart from three items, item–item comparisons showed fair to moderate agreement.

Our study on the GDS-15 is in concordance with other studies on depression suggesting that self-administered measures give higher scores compared to interviewer-administered measures. Among patients of a medical geriatric unit, the GDS-30 was administered twice within 5 days, once self-administered and once staff-administered in a random sequence showing that total scores on the GDS-30 were on average 2 points higher when self-administered (O'Neill et al., 1992). Another study using the Center for Epidemiologic Studies Depression (CES-D) Scale also found higher scores when the CES-D was self-administered compared to interviewer-administered, both in lower (<70 years) as in higher age groups (over 70 years) (Geerlings et al., 1999).

Why would self-administered scores be higher (i.e. more depressive) than interviewer-administered scores? These findings could suggest that older subjects need help filling out questionnaires, and that this need for help is not restricted to the oldest old. This is affirmed by our finding that many older subjects (23.4%) left items unanswered when the GDS-15 was self-administered. But what kind of help is needed? Perhaps, it is help in general, such as explaining procedures and solving misunderstandings, which cannot be given by mail. On the other hand, the written GDS-15 could be improved. Given the yes–no answering categories, subjects are unable to score nuances, which may lead to unanswered items. Subjects may have difficulty interpreting some specific items: e.g. Segulin and Deponte (2007) suggested rephrasing some items to make them less 'philosophical' and more concrete. Additional written information may be needed concerning the reference group ('of same age') and time frame ('last month'). Another improvement might be to shorten the GDS-15 by removing problematic items. Several shortened versions have been proposed by evaluating item correlations with depression (e.g. D'Ath et al., 1994) or item suitability (e.g. Jongenelis et al., 2007). Unfortunately, in the GDS-10 and GDS-8 different items were removed and we identified again other items as problematic, except item 9 which is commonly considered as problematic. The GDS-10 and GDS-8 both still include items 1 and 3, which had a low kappa in our analyses (Jongenelis et al., 2007). Other reasons for systematic discrepancies between the two ways of administration may be rating by proxy (e.g. through visual problems, cognitive problems, language skills, illiteracy or lack of motivation), or socially desirable responding. Some authors suggest that sensitive

questions are answered more truthfully when self-administered, since the presence of the interviewer might influence scores towards socially desirable answers, that is, not being depressed (De Leeuw, 2005). If this is true, this would favor a self-administration method. For some populations (such as the oldest old) one might want to include a cognitive test. This is more feasible during an interview, although this can be costly. In a preceding pilot study 12% of elderly respondents mentioned that the postal GDS-15 was filled out by others. We assume that this might lead to somewhat higher but valid scores, as was found in studies developing an informant-version of the GDS (Nitcher et al., 1993; Brown and Schinka, 2005). An alternative, to overcome missing items and rule out filling out by proxy, is to use the GDS by telephone (Burke et al., 1995). This might still induce socially desirable answers though.

To our knowledge, this study is the first to assess the difference between self-administered and interviewer-administered GDS-15 scores among subjects aged 75 years and older in the general population. Many GDS validation studies have included patients aged 60 and older, and we showed that the method of administration had influence on scores among the oldest old. It could be seen as a limitation that the sequence of administration to each subject was not randomly assigned. And we did not always succeed in keeping the time between the two tests to a preferred minimum. However, stratified analyses consequently showed higher scores when the GDS-15 was self-administered. Therefore, we expect that scores will be systematically higher when the GDS is self-administered. Given these limitations, however, we cannot exactly quantify the difference. The interviewers did not restrict themselves to 'read out loud' the GDS, as is often advised. To our opinion, however, it was desirable that they gave synonyms or examples to explain items to enable subjects to choose between 'yes' and 'no'. We advise other researchers to administer the GDS in the same manner.

What are the implications of our findings for research and practice? We consider a mean difference of 0.5 to 1 point on a total of 15 points could be clinically relevant, especially when a strict cut-off point is used to act on. Screening by mail is less costly and may have fewer barriers in revealing sensitive information (De Leeuw, 2005). On the other hand, interviews give no missing answers and help can be offered when questions are not clear to the participant or are misunderstood, resulting in more accurate answers. It is advisable to register whether persons received help. Differences between both administration methods varied a lot between individual older subjects, but we could not identify a specific subgroup for which

screening by mail is not applicable or reliable. For epidemiological studies the extra costs of interviewing all subjects may be worthwhile to get more valid data. Perhaps, administration of the GDS by telephone is another option (Burke et al., 1995). In clinical practice the costs may not weigh up to the benefits, e.g. in a combined screening and treatment program initiated by the GP (Van der Weele et al., 2011). In this case a two-step design, firstly mail and secondly an interview among screen positives, can be chosen.

To conclude, our study indicates that the method of administering the GDS should be carefully weighed and reported. It should be taken into account when interpreting scores, e.g. when comparing studies or choosing a cut-off point.

Acknowledgments

This study was funded by a grant from the Netherlands Organisation for Health Research and Development (ZonMw, grant number: 945-07-502). We thank Annemarie Kraaijpoel for her preliminary work.

References

- Arthur, A.J., Jagger, C., Lindesay, J., Matthews, R.J., 2002. Evaluating a mental health assessment for older people with depressive symptoms in general practice: a randomised controlled trial. *The British Journal of General Practice* 52, 202–207.
- Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310.
- Bland, J.M., Altman, D.G., 1997. Cronbach's alpha. *British Medical Journal* 314, 572.
- Brown, L.M., Schinka, J.A., 2005. Development and initial validation of a 15-item informant version of the Geriatric Depression Scale. *International Journal of Geriatric Psychiatry* 20, 911–918.
- Burke, W.J., Roccaforte, W.H., Wengel, S.P., Conley, D.M., Potter, J.F., 1995. The reliability and validity of the Geriatric Depression Rating Scale administered by telephone. *Journal of American Geriatrics Society* 43, 674–679.
- D'Ath, P., Katona, P., Mullan, E., Evans, S., Katona, C., 1994. Screening, detection and management of depression in elderly primary care attenders. I: the acceptability and performance of the 15 item Geriatric Depression Scale (GDS15) and the development of short versions. *Family Practice* 11, 260–266.
- De Leeuw, E.D., 2005. To mix or not to mix data collection modes in surveys. *Journal of Official Statistics* 21, 233–255.
- Feinstein, A.R., Cicchetti, D.V., 1990. High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology* 43, 543–549.
- Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 12, 189–198.
- Freudenstein, U., Arthur, A., Matthews, R., Jagger, C., 2002. Can routine information improve case finding of depression among 65 to 74 year olds in primary care? *Family Practice* 19, 520–522.
- Geerlings, S.W., Beekman, A.T.F., Deeg, D.J., van Tilburg, W., Smit, J.H., 1999. The Center for Epidemiologic Studies Depression Scale (CES-D) in a mixed-mode repeated measurements design: sex and age effects in older adults. *International Journal of Methods in Psychiatric Research* 8, 102–109.
- Harris, T., Cook, D.G., Victor, C., Rink, E., Mann, A.H., Shah, S., DeWilde, S., Beighton, C., 2003. Predictors of depressive symptoms in older people—a survey of two general practice populations. *Age and Ageing* 32, 510–518.
- Iliffe, S., Mitchley, S., Gould, M., Haines, A., 1994. Evaluation of the use of brief screening instruments for dementia, depression and problem drinking among elderly people in general practice. *The British Journal of General Practice* 44, 503–507.
- Jongenelis, K., Gerritsen, D.L., Pot, A.M., Beekman, A.T., Eisses, A.M., Kluiters, H., Ribbe, M.W., 2007. Construction and validation of a patient- and user-friendly nursing home version of the Geriatric Depression Scale. *International Journal of Geriatric Psychiatry* 22, 837–842.
- Kempen, G.I., Brilman, E.I., Ormel, J., 1995. The Mini Mental Status Examination. Normative data and a comparison of a 12-item and 20-item version in a sample survey of community-based elderly. *Tijdschrift voor Gerontologie en Geriatrie* 26, 163–172.
- Korner, A., Lauritzen, L., Abelskov, K., Gulmann, N.C., Brodersen, A.M., Wedervang-Jensen, T., Marie, K.K., 2007. Rating scales for depression in the elderly: external and internal validity. *The Journal of Clinical Psychiatry* 68, 384–389.
- Lach, H.W., Chang, Y.P., Edwards, D., 2010. Can older adults with dementia accurately report depression using brief forms? Reliability and validity of the Geriatric Depression Scale. *Journal of Gerontological Nursing* 36, 30–37.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33, 159–174.
- Licht-Strunk, E., van der Kooij, K.G., van Schaik, D.J., van Marwijk, H.W., van Hout, H.P., de Haan, M., Beekman, A.T., 2005. Prevalence of depression in older patients consulting their general practitioner in The Netherlands. *International Journal of Geriatric Psychiatry* 20, 1013–1019.
- Mitchell, A.J., Bird, V., Rizzo, M., Meader, N., 2010. Diagnostic validity and added value of the Geriatric Depression Scale for depression in primary care: a meta-analysis of GDS30 and GDS15. *Journal of Affective Disorders* 125, 10–17.
- Nitcher, R.L., Burke, W.J., Roccaforte, M.D., Wengel, S.P., 1993. A collateral source version of the Geriatric Depression Rating Scale. *The American Journal of Geriatric Psychiatry* 1, 143–152.
- Noltorp, S., Gottfries, C.G., Norgaard, N., 1998. Simple steps to diagnosis at primary care centres. *International Clinical Psychopharmacology* 13 (Suppl 5), S31–S34.
- Olivera, J., Benabarre, S., Lorente, T., Rodriguez, M., Pelegrin, C., Calvo, J.M., Leris, J.M., Idanez, D., Arnal, S., 2008. Prevalence of psychiatric symptoms and mental disorders detected in primary care in an elderly Spanish population. The PSICOTARD Study: preliminary findings. *International Journal of Geriatric Psychiatry* 23, 915.
- O'Neill, D., Rice, I., Blake, P., Walsh, J.B., Coakley, D., 1992. The Geriatric Depression Scale: rater-administered or self-administered? *International Journal of Geriatric Psychiatry* 7, 511–515.
- Osborn, D.P., Fletcher, A.E., Smeeth, L., Stirling, S., Nunes, M., Breeze, E., Siu-Woon, N.E., Bulpitt, C.J., Jones, D., Tulloch, A., 2002. Geriatric Depression Scale Scores in a representative sample of 14,545 people aged 75 and over in the United Kingdom: results from the MRC Trial of Assessment and Management of Older People in the Community. *International Journal of Geriatric Psychiatry* 17, 375–382.
- Pignone, M.P., Gaynes, B.N., Rushton, J.L., Burchell, C.M., Orleans, C.T., Mulrow, C.D., Lohr, K.N., 2002. Screening for depression in adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 136, 765–776.
- Segulin, N., Deponte, A., 2007. The evaluation of depression in the elderly: a modification of the Geriatric Depression Scale (GDS). *Archives of Gerontology and Geriatrics* 44, 105–112.
- Sheikh, J.I., Yesavage, J.A., 1986. Geriatric Depression Scale (GDS): recent findings and development of a shorter version. In: Brink, T.L. (Ed.), *Clinical gerontology: a guide to assessment and intervention*. Howarth Press, New York.
- Smeeth, L., Fletcher, A.E., Stirling, S., Nunes, M., Breeze, E., Ng, E., Bulpitt, C.J., Jones, D., 2001. Randomised comparison of three methods of administering a screening questionnaire to elderly people: findings from the MRC trial of the assessment and management of older people in the community. *British Medical Journal* 323, 1403–1407.
- Stek, M.L., Gussekloo, J., Beekman, A.T., van Tilburg, W., Westendorp, R.G., 2004. Prevalence, correlates and recognition of depression in the oldest old: the Leiden 85-plus study. *Journal of Affective Disorders* 78, 193–200.
- Van der Weele, G.M., de Waal, M.W., van den Hout, W.B., van der Mast, R.C., de Craen, A.J., Assendelft, W.J., Gussekloo, J., 2011. Yield and costs of direct and stepped screening for depressive symptoms in subjects aged 75 years and over in general practice. *International Journal of Geriatric Psychiatry* 26, 229–238.
- Weyerer, S., Eifflaender-Gorfer, S., Kohler, L., Jessen, F., Maier, W., Fuchs, A., Pentzek, M., Katuszkiewicz, H., Bachmann, C., Angermeyer, M.C., Lippa, M., Wiese, B., Mosch, E., Bickel, H., 2008. Prevalence and risk factors for depression in non-demented primary care attenders aged 75 years and older. *Journal of Affective Disorders* 111, 153–163.
- Whooley, M.A., Stone, B., Soghikian, K., 2000. Randomized trial of case-finding for depression in elderly primary care patients. *Journal of General Internal Medicine* 15, 293–300.
- Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., Leirer, V.O., 1982. Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research* 17, 37–49.