The criterion validity of the Geriatric Depression Scale: a systematic review

CRD summary
The review assessed the diagnostic accuracy of the Geriatric Depression Scale (GDS-15 and GDS-30 versions). It concluded that the GDS is similar to the Centre for Epidemiological Studies Depression scale and better than the Yale-1-question screen. Limitations in the search, the reporting of the review and the meta-analytic methods mean that conclusions should be viewed with caution.

Authors' objectives
To assess the screening accuracy of both versions of the Geriatric Depression Scale (GDS-30 and GDS-15) and to compare the GDS with other screening instruments.

Searching
MEDLINE, EMBASE, CINAHL, PSYNDEX and the Cochrane Library were searched to September 2004. The search terms, which were reported, combined GDS with terms for diagnostic accuracy. The bibliographies of published reviews and retrieved papers were screened for additional studies.

Study selection
Study designs of evaluations included in the review
Observational and diagnostic accuracy studies were included. Studies with fewer than 10 depression cases, or where the absolute number of cases could not be calculated, were excluded.

Specific interventions included in the review
Studies assessing the accuracy of the GDS at any specified cut-off value were eligible for inclusion. Studies conducted in medical out-patient (including primary care) and in-patient settings, as well as in services for the elderly (e.g. nursing home) settings, were eligible for inclusion; studies of phone versions of the GDS were excluded. Reported cut-offs for the GDS-15 were between 3 and 10; reported cut-offs for the GDS-30 were between 7 and 14.

Reference standard test against which the new test was compared
The included studies were required to use an external reference standard (usually based on a psychiatric research interview) to confirm diagnosis. Other screening questionnaires, chart or routine diagnoses of treating physicians were not considered acceptable reference standards. Studies in which the reference standard was applied only in participants who were above the GDS cut-off were excluded. Most of the included studies used the American Psychiatric Association's DSM-III, DSM-III-R or DSM-IV criteria as the reference standard.

Participants included in the review
Studies of elderly participants were eligible for inclusion; studies of psychiatric patients were excluded. In the included studies, the mean age of the study participants ranged from 61 to 83 years and the prevalence of depression ranged from 6 to 52%.

Outcomes assessed in the review
Studies were included if they reported the sensitivity and specificity for any specified cut-off value of the GDS, or sufficient data to enable the calculation of these parameters.

How were decisions on the relevance of primary studies made?
Two reviewers independently read all retrieved abstracts, and any abstract classified as containing accuracy data (by either reviewer) was retrieved for full assessment. Two reviewers then independently assessed for inclusion all retrieved papers published in English, French or German (irrespective of the language of the GDS); any disagreements were resolved by discussion.

Assessment of study quality

Database of Abstracts of Reviews of Effects (DARE)
Produced by the Centre for Reviews and Dissemination
Copyright © 2018 University of York
The reviewers assessed whether the reference standard evaluation was conducted blind to the screening (GDS) result. The authors did not state how many reviewers were involved in this assessment.

Data extraction
The authors did not state how the data were extracted for the review, or how many reviewers performed the data extraction process. Data were extracted on: participant characteristics (including cognitive status); language, version and cut-off of GDS used; type of research interview used as the reference standard; sensitivity and specificity.

Methods of synthesis
How were the studies combined?
The mean sensitivity and specificity and 95% confidence intervals, weighted by sample size, were calculated for all screening instruments where data were available from three or more studies. The mean positive predictive value, negative predictive value and overall misclassification rate, weighted by sample size and prevalence of each study, were also calculated.

How were differences between studies investigated?
Differences between the studies were discussed in the text. No formal statistical assessment of between-study heterogeneity was reported. Separate analyses were also conducted, with studies grouped by setting.

Results of the review
Forty-two studies with a total of 6,314 participants were included in the review. The sample size of the included studies ranged from 40 to 715, and the prevalence of depression was between 5.6 and 51.6%.

Thirty-two of the included studies investigated the GDS used in its original language (English), while the remaining 10 using translated versions.

In 26 studies, researchers assessing the reference standard were blinded to the results of the GDS.

Diagnostic accuracy of the GDS-15.
Not including modified Mandarin and Cantonese versions, the sensitivity ranged from 0.600 to 0.940 and the specificity from 0.570 to 0.870. For all GDS-15 studies, the weighted mean sensitivity was 0.805 and the weighted mean specificity 0.750 (21 studies).

Diagnostic accuracy of the GDS-30.
The sensitivity ranged from 0.340 to 1.000 and the specificity from 0.629 to 0.964. For all GDS-30 studies, the weighted mean sensitivity was 0.753 and the weighted mean specificity 0.770 (33 studies).

Comparison of the GDS-15 and GDS-30.
Using studies reporting data for the GDS-15 and GDS-30 in identical samples (9 studies), sensitivity was significantly higher for the GDS-30 than for the GDS-15 while specificity was significantly higher for the GDS-15 than for the GDS-30. The mean overall misclassification rate was lower for the GDS-15.

Comparison of the GDS-15 and GDS-30 with other screening instruments.
Direct comparisons of the Centre for Epidemiological Studies Depression scale (CES-D) with the GDS-15 (4 studies) showed a significantly lower mean sensitivity and a significantly higher mean specificity for the CES-D; results were similar for comparisons with the GDS-30 (6 studies). Direct comparisons of the Yale-1-question screen with the GDS-15 (4 studies) and GDS-30 (3 studies) showed significantly lower mean sensitivity and specificity for the Yale-1-question screen in both cases.
Authors' conclusions
The accuracy of both versions of the GDS is similar to that of the CES-D for diagnosing depression and significantly better than that of the Yale-1-question screen. Methodological limitations of the primary studies impede the generalisability of the results of the meta-analyses.

CRD commentary
The review addressed a clearly stated question, defined by appropriate inclusion criteria. The search strategy covered a range of relevant sources, but the use of methodological search terms to identify diagnostic accuracy studies (as described) is likely to result in studies being missed; this is due to inconsistency in the reporting of such studies and in their indexing in bibliographic databases. Post-search restrictions to studies reported in English, French or German might have resulted in further loss of data. Appropriate measures to reduce the potential for error and bias in the study selection process were reported. However, it is unclear whether similar measures were applied to the data extraction process. The assessment of the methodological quality of included studies was limited to a single criterion and the results of this assessment were not reported for individual studies. Similarly, although key study characteristics were reported for individual included studies, numerical results were not (illustration on a forest plot only). This omission makes interpretation of the results of the review very difficult.

The relevance of pooled estimates is doubtful, as no formal testing of between-study heterogeneity or threshold effect (the effect on diagnostic performance or variation in cut-off) was reported; it was explicitly stated in the text that studies using a range of cut-off values were pooled and that visual inspection of the forest plots presented indicates the presence of significant between-study heterogeneity. The authors focused their conclusions upon the findings of the small number of studies which directly compared the GDS with other screening tools. These conclusions should be viewed with caution given the limitations outlined.

Implications of the review for practice and research
Practice: The authors stated that the GDS is not useful for the diagnosis of depression in persons with marked cognitive impairment. The accuracy of the GDS in non-psychiatric hospital in-patients is sufficiently high to consider its use.

Research: The authors stated that further research is needed to clarify the role of the GDS in nursing home residents and out-patients.

Funding
Pfizer Corporation, Austria; Wyeth-Lederle, Austria.

Bibliographic details

PubMedID
17087788

DOI
10.1111/j.1600-0447.2006.00888.x

Other publications of related interest

Indexing Status
Subject indexing assigned by NLM

MeSH
Record Status
This is a critical abstract of a systematic review that meets the criteria for inclusion on DARE. Each critical abstract contains a brief summary of the review methods, results and conclusions followed by a detailed critical assessment on the reliability of the review and the conclusions drawn.