Clinical diagnosis of depression in primary care: a meta-analysis
Mitchell A J, Vaze A, Rao S

CRD summary
This review evaluated the accuracy of unassisted diagnoses of depression made by general practitioners (GPs). The authors concluded that GPs could rule out depression in most people who were not depressed, but the modest prevalence of depression in primary care meant that misidentifications outnumbered missed cases. The authors’ conclusions reflect the evidence presented and are overall likely to be reliable.

Authors' objectives
To assess the accuracy of unassisted diagnoses of depression made by general practitioners.

Searching
MEDLINE, PsycINFO, EMBASE and Scopus were searched from inception until 2009. Search terms were reported. A citation search was also performed on the Web of Knowledge.

Study selection
Eligible for inclusion in the review were studies in which general practitioners made an unassisted diagnosis of depression (i.e. without the use of severity scales, diagnostic instruments, education programmes or other organisational interventions). Outcomes had to be measured using interview-based diagnoses, a psychiatric expert diagnosis or validated-structured or semi-structured interview applied by a research interviewer. Included studies had to have a sample size of at least 50 cases. For studies assessing different types of depression, only data relating to major depression was used.

Studies were performed in USA, Canada, Europe and Australia. The mean patient age varied between studies (ranging from 15.6 years to 82.5 years). Gold standard reference tests included: composite international diagnostic interview (CIDI), diagnostic interview schedule (DIS), structural clinical interview for diagnostic and statistical manual (DSM) IIIR (SCID), Canberra interview for the elderly (CIE), schedules for clinical assessment in neuropsychiatry/present state examination (SCAN/PSE), schedule for affective disorders and schizophrenia, MINI internal neuropsychiatric interview, and the geriatric mental state examination. Case ascertainment was made through cumulative notes or electronic record, cross-sectional notes or electronic record, cumulative questionnaires, and contemporaneous prospective questionnaires.

The authors did not state how the papers were selected for the review, or how many reviewers performed the selection.

Assessment of study quality
Study quality was assessed using sample, data integrity, choice of reference criterion, method of case ascertainment, and duration of clinical assessment of depression. Quality was rated on a 5 point scale. Studies scoring one were of the highest quality (blind, sample size greater than 1000, drop-outs accounted for) and studies scoring five were of the lowest quality (unblinded or unclear whether blinded, sample size less than 100, other substantial methodological weaknesses).

Two authors assessed study quality. It was not reported whether this was performed independently or how any disagreements were resolved.

Data extraction
Data were extracted in order to calculate sensitivity, specificity, positive predictive value, negative predictive value and number-needed-to-screen.

Two reviewers independently extracted the data, which was then checked by a third reviewer.
Methods of synthesis
Summary estimates for sensitivity and specificity were calculated, along with corresponding 95% confidence intervals, using the random-effects model for all studies and for studies which examined the ability of general practitioners to accurately rule in a diagnosis of depressed people and rule out a diagnosis of non-depressed people.

Positive predictive values and negative predictive values were adjusted for variations in the prevalence of depression using a Bayesian curve of all post-test probabilities.

Bivariate meta-analysis was also performed and a summary Receiver Operator Characteristic curve was constructed.

Duration of assessment (cumulative versus versus cross-sectional), method of case ascertainment (case notes versus contemporaneous forms), country of origin, and patient ages were investigated as predictors of diagnostic accuracy.

Heterogeneity was assessed using the $I^2$ test.

Results of the review
Forty one studies were included in the review (50,371 patients). Four studies were assessed to have the highest quality and four studies were assessed to have the lowest quality. Nineteen studies examined the ability of general practitioners to accurately rule in a diagnosis of depressed people and rule out a diagnosis of non-depressed people. The overall prevalence of depression was 19.5% (95% confidence interval (CI): 15.7 to 23.7).

General practitioners correctly identified depression in 47.3% (95% CI: 41.7 to 53.01) of cases and recorded depression in their notes in 33.6% (95% CI: 22.4 to 45.7). However, there was evidence of statistically significant heterogeneity ($I^2=94\%$)

For rule-in and rule-out accuracy, sensitivity was 50.1% (95% CI: 41.3 to 59.0) and specificity was 81.3% (95% CI: 74.5 to 87.3). Again, there was evidence of statistically significant heterogeneity ($I^2=99.5\%$). The positive predictor value was 42.0% (95% CI: 39.6 to 44.3) and the negative predictor value was 85.8% (95% CI: 84.8 to 86.7).

Authors' conclusions
General practitioners could rule out depression in most people who were not depressed. However, the modest prevalence of depression in primary care meant that misidentifications outnumbered missed cases.

CRD commentary
The review addressed a clear research question and was supported by adequate inclusion criteria. The search strategy was adequate. However, there was no apparent attempt to search for unpublished studies, which means that relevant studies may have been missed. Furthermore, it was not reported how many reviewers were involved in the study selection process. This means that this process may have been subject to reviewer error or bias. The study quality assessment was adequate. Adequate details of the primary studies were provided and the synthesis methods were appropriate. The authors' conclusions reflect the evidence presented and are overall likely to be reliable.

Implications of the review for practice and research
Practice: The authors stated that repeated assessment by the general practitioner or other professional in a collaborative model with a case manager might help to reduce diagnostic errors and improve overall quality of care.

Research: The authors did not state any implications for research.

Funding
None.

Bibliographic details
Clinical Competence /statistics & numerical data; Depressive Disorder /diagnosis /epidemiology; Diagnostic Errors /methods /statistics & numerical data; Diagnostic and Statistical Manual of Mental Disorders; False Negative Reactions; False Positive Reactions; Family Practice /organization & administration; Humans; International Classification of Diseases; Interview, Psychological; Logistic Models; Multivariate Analysis; Prevalence; Primary Health Care /organization & administration; Psychiatric Status Rating Scales; Research Design; Sample Size; Sensitivity and Specificity

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.