

Selective Cutoff Reporting in Diagnostic Accuracy Studies of the PHQ-9 and EPDS Depression Screening Tools

Dipika Neupane^{1,2}, Brooke Levis¹⁻³, Parash Mani Bhandari^{1,2}, Brett D Thombs^{1,2}, Andrea Benedetti^{2,4}, and the DEPRESSION Screening Data (DEPRESSD) Collaboration

¹*Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Québec, Canada*

²*Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Québec, Canada*

³*Centre for Prognosis Research, School of Primary, Community and Social Care, Keele University, Staffordshire, UK*

⁴*Respiratory Epidemiology and Clinical Research Unit, McGill University Health Centre, Montréal, Québec, Canada*

Background: Selectively reporting accuracy results from only well-performing cutoffs in studies of diagnostic or screening tests may result in biased estimates when synthesized. Extent of bias may differ depending on the availability of a well-defined standard cutoff.

Aims: We compared bias in accuracy estimates and cutoff reporting patterns for the Patient Health Questionnaire-9 (PHQ-9; well-defined standard cutoff ≥ 10) and Edinburgh Postnatal Depression Scale (EPDS; no standard cutoff, common cutoffs ≥ 10 to ≥ 13).

Methods: We analyzed subsets of datasets from two separate individual participant data meta-analyses (IPDMAs) on PHQ-9 and EPDS accuracy. Separately, for the PHQ-9 and EPDS, we used bivariate random effects meta-analysis to compare accuracy estimates based on published cutoffs only versus all cutoffs from all studies. We also compared the number of published cutoffs below and above the standard or common cutoffs in relation to study-specific “optimal” cutoffs.

Results: For the PHQ-9 (30 studies, $N = 11,773$), published results underestimated sensitivity compared to results for all cutoffs for cutoffs below ≥ 10 (median difference: -0.06) and overestimated for cutoffs above ≥ 10 (median difference: 0.07). EPDS (19 studies, $N = 3,637$) sensitivity estimates were similar for cutoffs below ≥ 10 (median difference: 0.01) but higher for published cutoffs above ≥ 13 (median difference: 0.14). Mean cutoff of all cutoffs reported among PHQ-9 studies with optimal cutoffs below ≥ 10 was 8.8 compared to 11.8 for studies with optimal cutoffs above ≥ 10 . 18 of 19 EPDS studies had optimal cutoffs below ≥ 13 ; those below ≥ 10 did not report more cutoffs below ≥ 10 (mean cutoff: 9.9), but those with above ≥ 10 reported more above ≥ 10 (mean cutoff: 11.8).

Conclusion: Selective cutoff reporting and resulting bias in accuracy estimates were more pronounced for the PHQ-9 than EPDS. Researchers evaluating diagnostic accuracy of screening tools should report results for all relevant cutoffs.

Keywords

diagnostic test accuracy, individual participant data meta-analysis, meta-analysis, selective cutoff reporting, publication bias