

There may be many more people with depression (with high or low Patient Health Questionnaire (PHQ) scores) who were unidentified because a SCID was not applied to the 780 and the screen depends wholly on the PHQ-9 on a single application. At a typical (medium-risk) prevalence rate of 20% there would be around 156 cases of depression in the group of 780, but in a high-risk group where the prevalence may be 35% this would mean around 273 true cases. Just relying on a single application of the PHQ-9 on its own (either by algorithm or recommended cut-off) is probably insufficient. Assuming (like the authors) that the sensitivity of the PHQ-9 is a generous 0.88,² then there might be 33 missed cases in a high-risk sample. However, a meta-analysis from Wittkamp *et al*³ found a pooled sensitivity of 0.77 and Gilbody *et al*⁴ found 0.81 (both in primary care), which would translate into 52–63 missed cases. Of course there is offset by the issue of false positives which should also be examined in a screening implementation study. However, this remains a speculation without the SCID data from the parent 780 sample which is not reported (but perhaps available to the authors?).

In summary, I suggest this is a genuinely useful paper about the hazards of screening implementation but it is not really about screening success, for which a screening randomised controlled trial or pre-post screen design is needed. A simple guide to interpretation of screening studies can be downloaded from www.psych-oncology.info/education.htm.

- 1 Baas KD, Wittkampf KA, van Weert HC, Lucassen P, Huyser J, van den Hoogen H, et al. Screening for depression in high-risk groups: prospective cohort study in general practice. *Br J Psychiatry* 2009; **194**: 399–403.
- 2 Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**: 606–13.
- 3 Wittkampf KA, Naeije L, Schene AH, Huyser J, van Weert HC. Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. *Gen Hosp Psychiatry* 2007; **29**: 388–95.
- 4 Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007; **22**: 1596–602.

Alex J. Mitchell, Leicester General Hospital, Leicester LE5 4PW, UK.
Email: alex.mitchell@leicspart.nhs.uk

doi: 10.1192/bjp.195.5.462a

Author's reply: Dr Mitchell raises two interesting issues: (1) the way we calculated results of the screening procedure, in particular the number needed to screen to treat one additional patient; and (2) the PHQ-9 as a screening instrument.

As Dr Mitchell mentions, our study was indeed a screening implementation study. We wanted to learn whether screening in high-risk groups could detect a substantial number of so far not detected and treatable patients with depression. For this we conducted a pragmatic study and determined the gain of a stepped care screening (and treatment) programme in real

practice, with real doctors and real patients. A GP who wants to screen his patients can read what to expect. Failures (refusals, no-shows, misclassifications) are inherent to such conditions and should be incorporated in the calculations.

We defined our target population and included patients ($n=2005$) from the GPs' medical files and surgery. Our screening cascade showed 17 new patients that could be treated for depression. Perhaps this calculation is a bit optimistic because treatment was directly available without costs, which is not always the case.

Dr Mitchell makes a comparison with a drug trial. Unfortunately, we do not think this comparison makes the interpretation of our data easier. We consider as our screen the PHQ and use the SCID as the reference diagnostic standard. So the 780 patients who returned the PHQ and gave informed consent form the screened population. From there we count downwards to the number of detected cases and upwards to the number needed to be invited for the screening to be able to screen those 780 patients.

There can be discussion about the way we corrected for patients that did not adhere to the programme. We presented each step (with number of people who refused, did not attend and the reasons therefore), so that readers can make their own judgements, as Dr Mitchell has done. However, we disagree with his interpretation. If we use his analogy of a drug trial, then an intention-to-treat analysis is the best analysis. That means that non-adherence should be incorporated in the number needed to treat (or screen). Starting the analysis with the number of patients that completed the SCID (as Dr Mitchell does) provides the GPs with the number of patients they have to see, after a pre-screen with the PHQ-9.

It is correct that the PHQ-9 misses some cases, although not as many as Dr Mitchell supposes. We used the PHQ-9 in the screening mode (a cut off score of 10) and not in the diagnostic mode (algorithm). Sensitivity of the screening mode is 0.93, not 0.77.¹ However, a GP who uses the PHQ-9 will follow the results of the screening and invite patients with a positive score for clinical evaluation, thus also missing patients with a score below threshold. We unfortunately do not have SCID data of those who scored negative on the PHQ.

- 1 Wittkampf K, van Ravesteijn H, Baas K, van de Hoogen H, Schene A, Bindels P, et al. The accuracy of the Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009; **31**: 451–9.

Henk van Weert, Department of General Practice, Academic Medical Center, University of Amsterdam; **Aart H. Schene**, Department of Psychiatry, Academic Medical Center (Room PA1-132), Meibergdreef 5 1105 AZ Amsterdam, The Netherlands. Email: a.h.schene@amc.uva.nl

doi: 10.1192/bjp.195.5.463