

## Article

# Depression screening as a quality indicator

Daniel J Luchins MD

Chief of Mental Health Research, Jesse Brown VAMC and Professor of Clinical Psychiatry, University of Illinois at Chicago, USA

### ABSTRACT

**Objective** Although practice guidelines for depression screening are evidence based, with their development relying on reviews of controlled studies, their adaptation and use as quality indicators have not been subject to rigorous study. This paper will therefore review the evidence supporting this practice.

**Methods** A rational evaluation was carried out on both controlled studies and other sources of evidence related to the technical, clinical and policy assumptions underlying the use of depression screening guidelines as quality indicators.

**Results** 1) **Technical assumptions:** depression screening **could** be used as a quality indicator. Current information technology does not allow accurate determination of who would benefit from being screened, whether they actually were screened, or the optimal percentage that should be screened. 2) **Clinical assumptions:** depression screening **would** improve outcomes. The evidence

suggests that although depression screening might increase the diagnosis of depression, depressed patients so recognised tend to be less ill, less in need of treatment, or less likely to benefit from treatment, while screening, in the absence of other interventions, does not improve outcomes.

3) **Policy assumptions:** depression screening **should** be a focus of quality improvement. However, relative to other preventative measures, depression screening is a low priority. It does not meet usual cost-effectiveness criteria. There are more robust interventions for depression (i.e. collaborative care) that could be a focus of quality improvement efforts.

**Conclusion** Although routine depression screening may be an acceptable practice guideline, its use as a quality measure is not supported.

**Keywords:** depression, quality indicators, screening

## Introduction

Depression screening guidelines have been endorsed by the US Preventive Service Task Force (USPSTF),<sup>1,2</sup> the Canadian Task Force for Preventive Health Care<sup>3</sup> and the UK National Institute of Clinical Excellence (NICE),<sup>4</sup> and have been adopted as a quality indicator in many primary care settings.<sup>5</sup> Since such guidelines are based on controlled clinical trials, it has been assumed that quality indicators derived from these guidelines are 'evidence based' so that increased performance on the quality indicators should improve outcomes. But these controlled trials examined the impact of screening, not adherence to a quality indicator. Empirical studies of adherence

to quality indicators derived from screening guidelines, at least those for cancer, have questioned their benefits<sup>6-11</sup> while drawing attention to the distinctions between the purpose and function of a practice guideline and a quality indicator.<sup>7</sup> Guidelines are cognitive tools, like a mnemonic, that can assist physicians in their professional activities. They are used at the physician's discretion in order to improve their performance for whatever intrinsic and extrinsic rewards greater mastery of their profession provides.<sup>12</sup> Quality indicators, even if they paraphrase a guideline, are nevertheless rules. Their adoption requires an administrative not a clinician

decision, to which physicians adhere to meet some predetermined goal and achieve some external reward (e.g. a bonus, keeping your job). Since guidelines and quality indicators are different, the evidence that might support the adoption of a practice guideline does not necessarily support its adaptation into a quality indicator. Evaluating such a quality indicator requires more than studies of a screening instrument's sensitivity and specificity or a meta-analysis of controlled trials; instead it demands a realistic evaluation of the evidence related to the chain of assumptions underlying the decision to adopt the quality indicator.<sup>13</sup> With this in mind, this paper will focus on quality indicators derived from depression screening guidelines, examining assumptions in three domains: technical (can it be done?); clinical (what will happen if it is done?); and policy (should it be done?).

### I Technical assumptions: adherence to depression screening guidelines could be used as a quality indicator

Practice guidelines are clinical tools; a clinician determines whether they are applicable to any particular patient using all available information. Quality indicators are administrative rules; an administrator decides that they are applicable to all patients in a particular category unless excluded by an algorithm that can only utilise data available to administrators. The patients to whom a guideline may be clinically applicable are not necessarily the same as those to whom it is applied administratively. Since the extent of this difference may vary amongst the patient populations of different clinicians, sites or programmes, performance on this guideline may not provide a comparable measure of their quality. Without studying the clinical applicability of a quality indicator in a particular clinical setting, the extent of bias this introduces will remain unknown. This problem could be reduced if clinicians had the option of noting that a screening guideline is not applicable to a particular case, but administrators are loath to offer such an option for fear that its abuse could undermine the value of their quality indicator.<sup>14</sup>

#### 1 You can determine how to screen

The ideal screen would take very little (if any) of the clinician's time yet would strongly affect their clinical decision making. Unfortunately, these two variables tend to be inversely related. Patient self-screening requires the least clinician time but may produce little impact on physician behaviour. This

phenomenon was well illustrated with the Primary Care Evaluation of Mental Disorder (PRIME-MD).<sup>15</sup> This involved a two-step process: 1) patients completed a 26-item self-administered questionnaire, and 2) if this was positive, clinicians completed an evaluation guide. Since this required a prohibitive amount of clinician time (mean = 8.4 minutes), a completely self-administered version was subsequently developed (PRIME-MD Patient Health Questionnaire, or PHQ)<sup>16</sup> which required far less clinician time (less than three minutes in 85% of cases) but it had almost no effect on clinician behaviour. Of the 74 patients with depression newly identified on the PHQ, only 22% received follow-up visits, 10% were prescribed an antidepressant and 5% were given a mental health referral. As the authors conclude, 'Although the PHQ is clearly more efficient for clinicians to use than the original PRIME-MD, our study indicates that it may also be easier to ignore'.<sup>16</sup>

An alternative approach is to have the screening carried out by other personnel (nurse, physician assistant, etc.) with the result then transmitted to the physician. But having staff provide physicians with diagnostic information is known to have only a modest effect on their behaviour. A meta-analysis examining controlled studies in which clinicians were provided with the results of positive depression screening tests found no effect on their prescription of antidepressant medications.<sup>17</sup> Although it can be used by clinic administrators to produce high levels of guideline adherence,<sup>10</sup> this approach risks dissociating the quality indicator from clinical outcomes.

#### 2 You can determine who to screen

Using cancer screening guidelines as quality indicators often mandates screening individuals who could not benefit. This may not only be futile<sup>7-11</sup> but can also have negative consequences.<sup>6</sup> To avoid screening individuals unlikely to benefit, some depression screening guidelines (i.e. NICE)<sup>4</sup> emphasise screening only high-risk groups. Since several 'red flags' are known to raise the suspicion of depression, in theory such patients could be identified through an electronic medical record and their rates of screening assessed.<sup>18</sup> But such a quality indicator would require a computer algorithm that could successfully identify at-risk patients. Some risk factors might be identifiable: utilisation of healthcare, persistent physical symptoms, or chronic cardiac, cancer and CNS disorders.<sup>19</sup> But other, perhaps stronger, indicators such as being a 'difficult' patient<sup>20</sup> or undergoing a stressful life experience<sup>21</sup> are not currently readily identifiable through an electronic record. Perhaps it is for this reason that there have

been no clinical trials of this high-risk screening strategy.<sup>5</sup>

Finally, avoiding needless screening also requires that those diagnosed and in treatment for depression should be excluded. Unfortunately, available information technology cannot reliably identify such patients. A study done in the Veterans Administration (VA),<sup>22</sup> which is a leader in electronic medical records, found that a computerised review of electronic data misclassified as undiagnosed and untreated over one-third of the patients whose written charts showed otherwise.

Because of administrators' reluctance to allow a 'does not apply' category, as well as difficulties in identifying high-risk patients, excluding those already depressed, or otherwise focusing screening on those most likely to benefit, when depression screening guidelines are turned into quality indicators they are generally applied to all patients, regardless of whether a given individual would stand to benefit. This makes it likely that very high levels of adherence do not necessarily reflect exemplary practice but rather the screening of individuals regardless of whether this is clinically appropriate.<sup>23</sup>

### 3 You can determine who was screened

Current automated methods to review electronic records do not allow an accurate determination of whether a patient met a quality indicator. In a study of coronary artery disease patients, the authors<sup>24</sup> manually re-examined the records of patients that an automated electronic review determined had not achieved one of seven quality measures. Depending on the indicator, they found that 15–81% were actually not failures. These inconsistencies are particularly problematic when measures are compared between settings. A field test<sup>25</sup> of a HEDIS colorectal cancer screening measure in five different health plans, measured adherence using 1) administrative data, 2) a hybrid of administrative data and medical record data and 3) patient survey data only. The relative ranking varied according to the data used; one plan ranked first on administrative data, second on hybrid data and fourth on survey data.

The significance of these issues has been ignored in studies that have used depression screening as a quality indicator. A VA study which was published as 'best practice' reported 97% adherence to annual depression screening.<sup>26</sup> But despite the VA's exemplary computerised medical record the finding relied on computerised review **plus** the author's manual review of all encounter forms and clinician notes. Attempts to match this benchmark, relying only on computerised review without comparable labour-intensive methods, would be expected to fail.

### 4 You can determine a target screening rate

As has been pointed out,<sup>7</sup> there are no empirical studies to justify target rates for adherence to any screening measure. Instead they are derived by two alternative methods. The first uses the adherence rates or target goals of other health systems. But this could only be valid if the systems being compared shared similar clinical characteristics (patient demographics, patient severity, etc.) and data elements. The second method uses a continuous quality improvement model. For example, for some performance indicators the VA resets the target goal to the previous year's screening rate of the 20% of VA networks who achieve the highest scores on this measure.<sup>10</sup> At best this suggests that such a rate is possible, not that it is desirable.

There is indirect evidence that adherence to mental health screening guidelines might be inversely related to clinical outcomes. Two VA studies, one<sup>27</sup> examining a facility's performance on alcohol screening guidelines (depression screening was not examined) and the other<sup>15</sup> surveying primary care physicians' use of all computerised clinical reminders (including screening for mental health conditions), both found academic affiliation related to lower rates of adherence. Since academic affiliation is usually associated with better quality care<sup>28</sup> this is contrary to the assumption that increasing adherence to screening guidelines is necessary to promote the best care.

## II Clinical assumptions: adherence to depression screening quality indicators will improve outcomes

Clinical guidelines are based on a systematic review of clinical trials; for example, the USPTF's 2002 recommendation<sup>1</sup> relied on a meta-analysis of seven controlled studies showing depression screening to be associated with a 13% reduction in relative risk for persistent depression. However, several of the studies included in this review utilised depression screening in conjunction with other quality improvement/care management strategies. A 2005 Cochrane review<sup>17</sup> that excluded such studies concluded that depression screenings alone did not improve outcomes. This is consistent with the USPTF 2009 update<sup>2</sup> that concluded, 'Depression screening programs without substantial staff-assisted depression care supports are unlikely to improve depression outcomes'. Although such findings might still provide sufficient grounds for recommending these guidelines to clinicians, they do not necessarily justify the administrative decision to use them as a quality indicator. If anything, they would suggest that administrators

need to focus their attention on these additional supports and not on a procedure (depression screening) which is ineffective in their absence.

Why screening by itself may not affect outcomes will be discussed below.

### *1 Unrecognised cases of depression would benefit from treatment*

Implicit in the use of depression screening is the assumption that the patients so recognised would generally benefit from treatment. The evidence for this is at best equivocal. Many patients identified by screening techniques have transient symptoms (possibly related to life events) not true depressive disorders, and do not require treatment. For example, two weeks after primary care patients were screened for depression,<sup>29</sup> most who had screened positive (52%) did not meet the criteria for a mood disturbance (either major depression or dysthymia). Furthermore, those who meet criteria for depression but were previously unrecognised by their primary care physician tend to be less severely ill<sup>30-32</sup> compared to those who are recognised and therefore over time (i.e. a year), despite receiving no intervention, do no worse than those who are recognised.<sup>30-34</sup> This is consistent with a recent meta-analysis suggesting that antidepressant treatment may be no more effective than a placebo in just such milder cases.<sup>35</sup> On the other hand, long-term follow-up (to four years) of patients who had not been in treatment for depression, but who screened positive, detected high rates of depression and overall poor mental health outcomes.<sup>36</sup> So, no doubt some unrecognised cases of depression would benefit from treatment. But this tendency for unrecognised cases to be less severe, less in need of treatment and less likely to benefit from antidepressant treatment does raise the question of whether this subpopulation should be the focus of quality improvement efforts.

### *2 Screening will increase recognition of depression*

Even if on average they are less ill, a proportion of depressed patients are not recognised by their primary care physicians. In cross-sectional studies this may approach one in every two cases,<sup>37</sup> with longitudinal follow-up suggesting that over time (i.e. three years) about one in seven remains undiagnosed.<sup>38</sup> There is an assumption that if the depression screens from all patients were shared with their clinicians, the recognition of depression would increase but this is only very weakly supported by the evidence. In their 2005 Cochrane review, Gilbody *et al*<sup>17</sup> included 11 controlled studies in which

clinicians were randomly presented with the results of depression screens on patients whom they did not recognise as depressed and their charts were subsequently reviewed for any entry about depression. They noted that the intervention had a borderline positive impact on the diagnosis of depression (RR = 1.27, 95% CI = 1.02 to 1.59), with high heterogeneity ( $I^2 = 69\%$ ) that was due to a more robust effect in studies in which notification and randomisation were restricted to patients with high scores (RR = 1.67, 95% CI = 0.89 to 3.16), but was negligible in increasing diagnosis in unselected patients (RR = 1.03, 95% CI = 0.85 to 1.24). The authors hypothesise that this reflects the clinicians' implicit use of Bayesian reasoning. Since the prevalence of depression in an unselected sample is relatively low, the post-screen probability of depression for any positive screen (as opposed to only those with high scores) will be less than 50%, wrong more often than it is right. The meta-analysis suggests that when clinicians are provided with the results of all positive depression screens, they tend to ignore the results.

### *3 Recognition will increase the use of appropriate interventions*

The meta-analysis by Gilbody *et al*<sup>17</sup> of studies in which clinicians were randomly presented with the results of a depression screen also examined whether those patients subsequently received interventions for depression. Notification had a borderline significant effect on increasing 'any intervention' (RR = 1.30, 95% CI = 0.97-1.76), but the heterogeneity was large ( $I^2 = 81\%$ ), with studies that randomised only high-risk patients tending to show a larger effect size than those that randomised unselected patients. When the analysis focused on the prescription of antidepressants, there was no difference between the intervention and the control group (RR = 1.20, 95% CI = 0.87-1.66). This suggests that when provided with the results of depression screens, even on selected cases, clinicians might be minimally inclined to do something, but not necessarily to prescribe antidepressants.

### *4 Recognition will improve outcomes*

Since depression screening identifies the less severely ill cases, while having a minimal effect on increasing either the diagnosis or treatment of depression, it is not surprising that in the absence of other interventions, depression screening may not be effective. When Gilbody *et al*<sup>17</sup> restricted their attention to studies in which depression screening was tested without any additional enhancement of care, there was no indication of any impact of screening on depression outcomes (standardised mean difference

= -0.02, 95% CI = -0.25–0.20, with low study heterogeneity  $I^2 = 31\%$ ). This does not support the assumption that clinical outcomes would necessarily benefit should quality improvement efforts improve compliance with depression screening guidelines.

### III Policy assumptions: should depression screening play a role in improving mental health services?

Even if depression screening guidelines could be made into accurate quality indicators, and even if adherence to these guidelines would improve clinical outcomes, there remains the question of what role such measures should play in efforts to enhance mental healthcare. As will be shown, the conclusion that adherence to depression screening guidelines should be adopted as a quality measure rests on several policy assumptions. This section will examine the evidence related to these assumptions.

#### 1 Depression screening is a priority among preventive services

The relative importance of depression screening compared to other preventive services was evaluated in a systematic review of all 25 recommended by the USPSTF.<sup>39</sup> For each, it calculated the clinically preventable burden (CPB) and the cost-effectiveness, ranked these on a five-point scale (five being best) and then added these to compute a total score. Depression screening in adults scored four, with 17 services ranked above it and only 4 below it. When attention is focused only on screening services it was outscored by colorectal screening,<sup>8</sup> hypertension screening,<sup>8</sup> vision screening in adults,<sup>8</sup> cervical cancer screening,<sup>7</sup> breast cancer screening,<sup>6</sup> chlamydia screening,<sup>6</sup> vision screening in children<sup>6</sup> and obesity screening.<sup>5</sup> Relative even to other screening measures, depression screening is not a priority.

#### 2 Depression screening is cost effective

The cost-utility of depression screening in primary care was modelled by Valenstein *et al.*<sup>40</sup> Annual screening produced a cost-utility ratio of over \$225 000/quality adjusted life year (QALY), while \$50 000/QALY is often used as a benchmark for cost-effectiveness. Of course, the utility of screening depends on the quality of care that can be provided once depression is identified. But even were annual screening to be carried out in an environment with optimal care for depression, a collaborative care model (see below), the cost-utility ratio would still be well over the \$50 000 benchmark.<sup>41</sup> For annual

screening to be cost-effective by comparison with no screening (<\$50 000/QALY), screening would have to cost \$3.00 or less, serve a population with a prevalence of 13% or more, lead to treatment for 80% of diagnosed patients and achieve remission in 85% of treated patients.<sup>40</sup> It is highly unlikely that all these criteria could be met.

#### 3 Screening is the most effective intervention to improve clinical outcomes in depression

Unlike screening in primary care settings, which has not been demonstrated to improve outcomes for depression,<sup>17</sup> there is one intervention, collaborative care, that has been convincingly shown to do so. Collaborative care involves three elements: a case manager, a primary care physician and access to specialist input. It has been subject to multiple randomised trials; a meta-analysis<sup>42</sup> of 35 studies found a positive effect on outcomes at six months (standardised mean difference (SMD) = 0.25, 95% CI = 0.18–0.32), and analysis of 11 longer-term studies identified statistically significant improvement at 18 months (SMD = 0.15, 95% CI = 0.03–0.46) and similar trends for up to five years. Cost-effectiveness models based on clinical data are favourable,<sup>43</sup> with benefits at a cost of \$10 000 to \$35 000/QALY, well within the usual \$50 000 benchmark. The meta-analysis also showed that studies with low fidelity to the collaborative care model had poorer outcomes, while regular supervision with a specialist and/or case manager with a mental health background was associated with better outcomes. These findings raise the possibility that a setting's adherence to quality indicators that assessed fidelity to the collaborative care model, rather than its clinicians' adherence to depression screening guidelines, might lead to improvement in clinical outcomes for depression and therefore might be a better focus for administrative attention.

## Conclusion

This paper examines whether adherence to depression screening guidelines in primary care should be a quality indicator, by examining relevant assumptions in three domains – the technical 'how', the clinical 'why' and the policy 'should'. It found little empirical support for any of these assumptions, but on the contrary found much that contradicted them. That despite this body of evidence a wide range of health systems continue to use routine depression screening as a quality indicator is consistent with the conclusion that 'it can be virtually

impossible for anyone who looks at a (*quality*) measure to determine how accurate it is. An absolutely terrible measure will still produce a result, which for all intents and purposes will look just as authentic as the result produced by an accurate measure'.<sup>44</sup> On such indicators, outstanding performance may be more a measure of looking good than doing good.

## REFERENCES

- 1 Pignone MP, Gaynes BN, Rushton JL *et al.* Screening for depression in adults: a summary of the evidence for the US Preventive Services Task Force. *Annals of Internal Medicine* 2002;136:765–76.
- 2 O'Connor EA, Whitlock EP, Beil TL and Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Annals of Internal Medicine* 2009;151:793–803.
- 3 MacMillan HL, Patterson CJS and Wathen CN. Screening for depression in primary care: recommendation statement from the Canadian Task Force on Preventive Health Care. *Canadian Medical Association Journal* 2005;172:33–5.
- 4 National Institute for Clinical Excellence. *Depression: core interventions in the management of depression in primary and secondary care*. London: HMSO, 2004.
- 5 Gilbody S, Sheldon T and Wessely S. Should we screen for depression? *BMJ* 2006;332:1027–30.
- 6 Walter LC, Eng C and Covinsky KE. Screening mammography for frail older women: what are the burdens. *Journal of General Internal Medicine* 2004;16:779–84.
- 7 Walter LC, Davidowitz NP, Heineken PA *et al.* Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. *Journal of the American Medical Association* 2004;291:2466–70.
- 8 Walter LC, Bertenthal D, Linquist K and Konety BR. PSA screening among elderly men with limited life expectancies. *Journal of the American Medical Association* 2006;296:2336–42.
- 9 Walter LC, Lindquist K, Nugent S *et al.* Impact of age and comorbidity on colorectal cancer screening among older veterans. *Annals of Internal Medicine* 2009;150:465–73.
- 10 Heineken PA, Wenell C, Kerlikowske K and Walters LC. Problems with using women's cancer screening rates to measure performance. *Federal Practitioner* 2009;26:17–20.
- 11 Sultan S, Conway J, Edelman D *et al.* Colorectal cancer screening in young patients with poor health and severe comorbidity. *Archives of Internal Medicine* 2006;166:2209–14.
- 12 MacInyre A. *After Virtue: a study in moral theory* (2e). Notre Dame, Indiana: University of Notre Dame, 1984.
- 13 Pawson R and Tilley N. *Realistic Evaluation*. London: Sage, 1997.
- 14 Fung CH, Tsai JS, Lulejian A *et al.* An evaluation of the Veterans Health Administration's clinical reminder system: a national survey of generalists. *Journal of General Internal Medicine* 2007;23:392–8.
- 15 Spitzer RL, Williams JBW, Kroenke K *et al.* Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *Journal of the American Medical Association* 1994;271:1749–56.
- 16 Spitzer RL, Kroenke K, Williams JBW *et al.* Validation and utility of a self-report version of the PRIME-MD: the PHQ primary care study. *Journal of the American Medical Association* 1999;282:1737–44.
- 17 Gilbody S, Sheldon T and House A. Screening and case-finding instruments for depression: a meta-analysis. *Canadian Medical Association Journal* 2008;179:997–1003.
- 18 Kroenke K. Depression screening is not enough. *Annals of Internal Medicine* 2001;134:418–20.
- 19 Kroenke K, Spitzer RL, Williams JB *et al.* Physical symptoms in primary care: predictors of psychiatric disorders and functional impairment. *Archives of Family Medicine* 1994;3:774–9.
- 20 Hahn SR, Kroenke K, Spitzer RL *et al.* The difficult patient: prevalence, psychopathology and functional impairment. *Journal of General Internal Medicine* 1996;11:1–8.
- 21 Saltini A, Mazzi MA, Del Piccolo I and Zimmerman C. Decisional strategies for attribution of emotional distress in primary care. *Psychological Medicine* 2004;34:729–39.
- 22 Kramer TL, Owen RR, Cannon D *et al.* How well do automated performance measures assess guideline implementation for new-onset depression in the Veterans Health Administration? *Joint Commission Journal on Quality and Patient Safety* 2003;29:479–89.
- 23 Luchins DJ. Improving care, improving performance or just improving numbers. *Psychiatric Services* 2009;59:1328–30.
- 24 Persell SD, Wright JM, Thompson JA *et al.* Assessing the validity of national quality measures for coronary artery disease using an electronic record. *Archives of Internal Medicine* 2006;166:2272–7.
- 25 Schneider EC, Nadel MR, Zaslavsky AM and McGlynn EA. Assessment of scientific soundness of clinical performance measures. *Archives of Internal Medicine* 2008;168:876–82.
- 26 Kirkcaldy RD and Tynes LL. Depression screening in a VA primary care clinic. *Psychiatric Services* 2006;57:1694–6.
- 27 Desai MM, Rosenheck RA and Craig TJ. Screening for alcohol use disorders among medical outpatients: the influence of individual and facility characteristics. *American Journal of Psychiatry* 2005;162:1521–6.
- 28 Rosenthal GE, Harper DE, Quinn LM and Cooper GS. Severity adjusted mortality and length of stay in teaching and non-teaching hospitals: results of a regional study. *Journal of the American Medical Association* 1997;278:485–90.
- 29 Coyne JC and Schwenk TL. The relationship of distress to mood disturbance in primary care and

- psychiatric populations. *Journal of Consulting and Clinical Psychology* 1997;65:161–8.
- 30 Ormel J, Koeter MWJ, van den Brink W and van de Willige G. Recognition, management, and course of anxiety and depression in general practice. *Archives of General Psychiatry* 1991;48:700–6.
- 31 Simon GE and Von Korff M. Recognition, management and outcomes of depression in primary care. *Archives of Family Medicine* 1995;4:95–6.
- 32 Simon GE, Goldberg D, Tiemans BG and Ustun TB. Outcome of recognized and unrecognized depression in an international primary care study. *General Hospital Psychiatry* 1999;21:97–105.
- 33 Coyne JC, Klinkman MS, Gallo SM and Shwenk TL. Short-term outcomes of detected and undetected depressed primary care patients and depressed psychiatric patients. *General Hospital Psychiatry* 1997; 19:333–43.
- 34 Dowrich C and Buchan I. Twelve month outcome of depression in general practice: does detection or disclosure make a difference? *BMJ* 1995;311:1274–6.
- 35 Fournier JC, DeRubeis RJ, Hollon SD *et al.* Antidepressant drug effects and depression severity: a patient-level meta-analysis. *Journal of the American Medical Association* 2010;303:47–53.
- 36 Weissman MM, Neria Y, Gameroff MC *et al.* Positive screens for psychiatric disorders in primary care: a long term follow-up of patients who were not in treatment. *Psychiatric Services* 2010;61:151–9.
- 37 Kessler D, Lloyd K, Lewis G and Gray DP. Cross sectional study of symptom attribution and recognition of depression and anxiety in primary care. *BMJ* 1999;318:536–40.
- 38 Kessler D, Bennewith O, Lewis G and Sharp D. Detection of depression and anxiety in primary care: follow up study. *BMJ* 2002;325:1016–17.
- 39 Maciosek MV, Coffield AB, Edwards NM *et al.* Priorities among effective clinical preventative services: results of a systematic review and analysis. *American Journal of Preventive Medicine* 2006;31:52–61.
- 40 Valenstein M, Vijan S and Zeber JE. The cost–utility of screening for depression in primary care. *Annals of Internal Medicine* 2001;134:345–60
- 41 Zeber J, Vijan S and Valenstein M. *The Cost Utility of Depression Screening: Usual Care Versus Collaborative Treatment*. Abstracts from the Academy for Health Services Research and Health Policy Meeting 2002; 19:14.
- 42 Gilbody S, Bower P, Fletcher J *et al.* Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. *Archives of Internal Medicine* 2006;166:2314–21.
- 43 Schoenbaum M, Unutzer J, Sherbourne C *et al.* Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *Journal of the American Medical Association* 2001;286:1325–30.
- 44 Eddy DM. Performance measurement: problems and solutions. *Health Affairs* 1998;17:7–25.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

Daniel J Luchins, Chief of Mental Health Research, Jesse Brown VAMC, 820 South Damen Ave, Chicago, Illinois 60612, USA. Email: [daniel.luchins@va.gov](mailto:daniel.luchins@va.gov)

Accepted June 2010

