

ORIGINAL RESEARCH

Potential Antidepressant Overtreatment Associated with Office Use of Brief Depression Symptom Measures

Anthony Jerant, MD, Richard L. Kravitz, MD, MSPH, Erik Fernandez y Garcia, MD, MPH, Mitchell D. Feldman, MD, MPhil, Camille Cipri, BS, Denyse Nishio, MD, Anca Knoepfler, MD, M. Kaleo Wooddell, MD, Victor Baquero, MD, and Peter Franks, MD

Background: Use of brief depression symptom measures for identifying or screening cases may help to address depression undertreatment, but whether it also leads to diagnosis and treatment of patients with few or no symptoms—a group unlikely to have major depression or benefit from antidepressants—is unknown. We examined the associations of use of a brief depression symptom measure with depression diagnosis and antidepressant recommendation and prescription among patients with few or no depression symptoms.

Methods: We conducted exploratory observational analyses of data from a randomized trial of depression engagement interventions conducted in primary care offices in California. Analyses focused on participants scoring <10 on a study-administered 9-item Patient Health Questionnaire (PHQ-9) (completed immediately before an office visit and not disclosed to the provider) with complete chart review data ($n = 595$). We reviewed visit notes for evidence of practice administration of a brief symptom measure (independent of the trial) and whether the provider (1) diagnosed depression or (2) recommended and/or prescribed an antidepressant.

Results: Among the 545 patients without a practice-administered measure, 57 (10.5%) had a visit diagnosis of depression; 9 (1.6%) were recommended and another 21 (3.8%) prescribed an antidepressant. Among the 50 patients (8.4% of total sample) with a practice-administered measure, 10 (20%) had a visit diagnosis of depression; 6 (12%) were recommended and another 6 (12%) prescribed an antidepressant. Adjusting for nesting within providers, trial intervention, stratification variables, and sample weighting, use of a brief symptom measure was associated with depression diagnosis (adjusted odds ratio, 3.2; 95% confidence interval, 1.1–9.2) and antidepressant recommendation and/or prescription (adjusted odds ratio, 3.80; 95% confidence interval, 1.0–13.9). Analyses using progressively lower PHQ-9 thresholds (<9 to <5) and examining antidepressant prescription alone yielded consistent findings. Analyses by practice-administered measure (PHQ-9 vs PHQ-2) indicated the study findings were largely associated with PHQ-9 use.

Conclusions: These exploratory findings suggest administration of brief depression symptom measures, particularly the PHQ-9, may be associated with depression diagnosis and antidepressant recommendation and prescription among patients unlikely to have major depression. If these findings are confirmed, researchers should investigate the balance of benefits and risks (eg, overdiagnosis of depression and overtreatment with antidepressants) associated with use of a brief symptom measure. (J Am Board Fam Med 2014;27:611–620.)

Keywords: Antidepressive Agents, Depression, Health Services Misuse

Underidentification of depression is prevalent and multifactorial in origin.^{1–3} Consequently, many practices use brief depression symptom measures,

such as the 2- or 9-item versions of the Patient Health Questionnaire (PHQ-2 or PHQ-9),^{4–6} to aid in identifying or screening cases. The US Pre-

This article was externally peer reviewed.

Submitted 21 January 2014; revised 18 April 2014; accepted 2 May 2014.

From the Department of Family and Community Medicine (AJ, PF), the Department of Internal Medicine,

Division of General Medicine (RLK), the Department of Pediatrics, Division of General Pediatrics (EFyG), and the Department of Internal Medicine, Division of General Medicine (MDF), University of California Davis School of Medicine, Sacramento; the University of California Davis

ventive Services Task Force (USPSTF) and others have endorsed the use of these measures in practices with appropriate diagnostic, treatment, and follow-up systems.^{7–10}

However, no studies have examined whether practice administration of a brief depression symptom measure also leads to antidepressant prescribing among people with few or no depression symptoms. Both the USPSTF and the Canadian Task Force on Preventive Health Care (CTFPHC) emphasize the need to examine this question.^{7,11} The issue is important to address, given that brief symptom measure validation studies using expert structured diagnostic interviews as the reference indicate that individuals with few or no depression symptoms are unlikely to benefit from antidepressants.^{4,6} In PHQ-9 validation studies, those with scores <10 largely comprised nondepressed people and those with minor depression.^{4,6} Such individuals may benefit from behavioral therapies,¹² but evidence from randomized controlled trials (RCTs) indicates most do not benefit from antidepressants.^{13–15} In other words, antidepressant treatment in such patients suggests potential overtreatment.

Despite trial evidence of little benefit, antidepressants are frequently prescribed to patients with few or no depression symptoms,^{16,17} resulting in unnecessary costs and potential detrimental effects (eg, labeling, medication toxicity).¹⁸ Discussions about antidepressants with such patients may also burden office visits,¹⁹ distracting from more salient issues.^{20,21} In contrast with the USPSTF, the CTFPHC recommended against using brief symptom measures, in part because of concerns about antidepressant overtreatment, but it noted a lack of RCTs examining how brief measures contribute both to appropriate depression treatment and overtreatment.¹¹

We examined these issues in exploratory observational analyses of data from an RCT.²² We focused on the subgroup of participants with few or

no depression symptoms, defined by a score of <10 on a study-administered PHQ-9.^{4,6} We examined whether a practice-administered brief depression symptom measure was associated with (1) increased diagnosis of depression and (2) increased recommendation and/or prescription of antidepressants. Providers were not informed about the study-administered PHQ-9 or the results. Depression diagnosis and antidepressant recommendation and/or prescription were considered in tandem to help gauge whether antidepressants were intended to address depression or other conditions (eg, insomnia, chronic pain). We examined the composite outcome of antidepressant recommendation and/or prescription because both components reflect clinicians' assessments of treatment need; prescriptions also reflect patient preferences.^{23–25} We ascertained use of brief symptom measures, diagnosis of depression, recommendation and prescription of antidepressants from visit medical records.

Given evidence that case finding and screening to identify other conditions lead to overdiagnosis and overtreatment,^{26,27} we hypothesized that use of a brief depression symptom measure would be associated with both increased depression diagnosis and increased antidepressant recommendation and prescription. While a PHQ-9 score <10 helps identify patients unlikely to have major depression, any cut point results in misclassification; lower cut points have increasing specificity in ruling out major depression.^{4,6} Thus, we also examined depression diagnosis and antidepressant recommendation and/or prescription associated with use of a brief symptom measure at progressively lower study-administered PHQ-9 thresholds (<9, <8, <7, <6, <5). Finally, we examined how the associations of use of a brief symptom measure with the study outcomes varied by the specific measure used.

Methods

We used data from an RCT examining the effects of 2 different in-office, previsit patient engagement and activation interventions on depression care and outcomes. The RCT procedures, interventions, and outcomes have been reported elsewhere and included patients across the spectrum of PHQ-9 scores.^{22,28} In previous PHQ-9 validation work, among individuals with a score of <10, only 5 (1.0%) had major depression based on expert diagnostic interview.^{4,6} Thus, the sample for these anal-

Center for Healthcare Policy and Research, Sacramento (AJ, RLK, CC, PF); and the University of California Davis Health System, Sacramento (DN, AK, MKW, VB).

Funding: The work was supported by grant 1R01MH079387 (to RLK) from the National Institute of Mental Health.

Conflict of interest: none declared.

Corresponding author: Anthony Jerant, MD, Department of Family and Community Medicine, University of California Davis School of Medicine, 4860 Y Street, Suite 2300, Sacramento, CA 95817 (E-mail: ajerant@ucdavis.edu).

yses included only RCT participants with a PHQ-9 score <10 and complete visit record review data.

Sample Recruitment and Enrollment

We recruited providers from primary care offices affiliated with the University of California, San Francisco; the San Francisco Veterans Affairs Medical Center; the University of California, Davis, Ambulatory Care Center; the University of California, Davis, Primary Care Network; the Northern California (Sacramento) Veterans Affairs Health System; Kaiser Permanente, Sacramento; and Sutter Medical Group, Sacramento. We obtained ethics approval from the institutional review boards at all performance sites. Clinicians were told the study was an RCT of interventions designed to improve communication about common physical and mental health symptoms. Although not blinded to their patients' participation, clinicians were not alerted to patients' group assignments.

Study eligibility criteria were age 25 to 70 years, able to read and understand English and use a touchscreen notebook computer, and not currently taking antidepressants (except low-dose tricyclic agents reported for pain or sleep). At all but one study office, eligibility screening was conducted by telephoning patients already scheduled (at their own initiative, not because of trial participation) to make a primary care visit in the next 1 to 2 weeks. Patients were told that the study was aimed at improving care for people with common symptoms including sleep problems, depression, and chronic pain. Patients with scores of ≥ 5 on the PHQ-8²⁹ (used in lieu of the PHQ-9 for initial telephone eligibility screening) were oversampled. A study research assistant met with eligible patients, who provided preliminary verbal phone consent 1 hour before their scheduled provider visit. At a single University of California, San Francisco, office, research assistants approached patients in waiting rooms without prior phone screening. In-office written informed consent was obtained from all participants.

Following informed consent, patients completed a preintervention computerized questionnaire containing the study measures. Next, a software program randomly assigned them to receive one of the study interventions. After receiving their intervention assignment from the computer, patients attended their scheduled visit. We provided patients

\$20 to \$35 for completing the in-office study procedures.

Measures

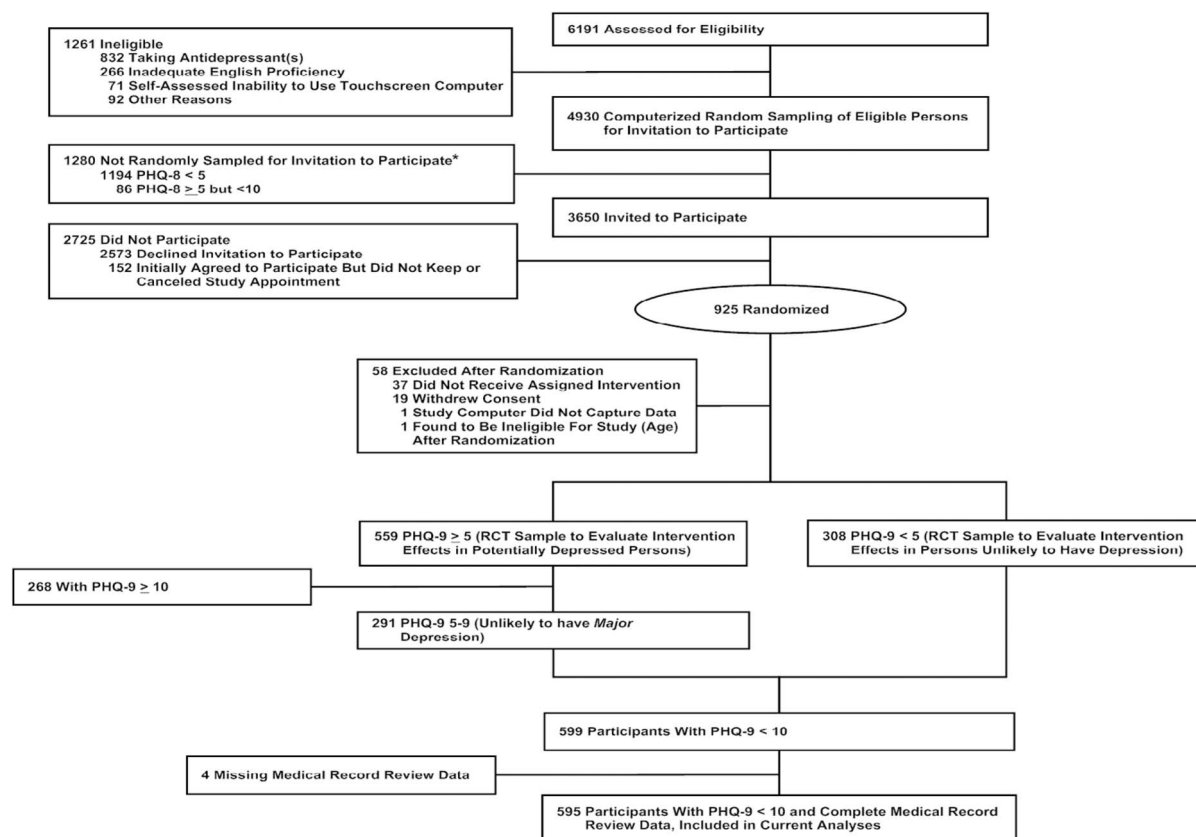
Depression symptoms were measured for all trial participants in the preintervention computerized questionnaire using the PHQ-9, administered within 20 minutes before the provider visit. PHQ-9 items assess how often respondents have experienced various symptoms (eg, feeling down, sleep problems, thoughts of harming oneself) in the preceding 2 weeks (sum of 9 items, each scored from 0 [not at all] to 3 [nearly every day]; scores can range from 0 to 27). Study providers and office staff were not notified that participants completed a PHQ-9 as part of the trial, nor were they given the results. In the previsit questionnaire, patients also self-reported their age in years, sex, race (white, black/African American, Asian, Native Hawaiian/other, Pacific Islander, or American Indian/Alaska native), and ethnicity (Hispanic vs not). They also completed the Medical Outcomes Study 12-item Short Form health status measure, which yields scores for both a Physical Component Summary and Mental Component Summary (range, 0–100; higher scores equate to better health).³⁰

Trained abstractors subsequently reviewed medical records for evidence (yes vs no) of whether the provider listed depression in the study visit note (ie, not the "master" problem list). The abstractors also reviewed visit notes for evidence (yes vs no) of practice (eg, provider or office staff) administration of a brief depression symptom measure (ie, not the study-administered PHQ-9, which was undisclosed to study providers).²⁸ If relevant, abstractors documented which specific measure was administered. Record abstractors also ascertained evidence of provider recommendation and/or prescription of an antidepressant during the study visit.²⁸ Coding options were antidepressant recommended (but not prescribed), antidepressant prescribed, or neither.

Analyses

Stata version 13.0 (Stata Corp., College Station, TX) was used. The primary analyses used logistic regression, which was implemented via generalized estimating equations to account for nesting of patients within clinicians, to model the associations of use of a brief depression symptom measure by the practice (yes/no; the key independent variable) with the following dependent variables in the entire

Figure 1. Flow of subjects through the study. *These subjects were excluded by a computerized random sampling program, based on low scores on the telephone-based 8-item Patient Health Questionnaire (PHQ-8) (used in lieu of the 9-item Patient Health Questionnaire [PHQ-9] for initial telephone eligibility screening), because the aim of the parent trial was to oversample patients with higher levels of depression symptoms. The PHQ-9 was administered to all participants in the prestudy visit questionnaire (but was not disclosed to study providers). RCT, randomized controlled trial.



study sample: (1) depression diagnosis during the study visit (yes/no); and (2) recommendation and/or prescription of an antidepressant during the visit (yes/no). Additional analyses examined these associations in patient subsamples defined by lower cut points on the study-administered PHQ-9: <9, <8, <7, <6, and <5. These analyses were conducted because specificity in ruling out major depression increases at lower cut points. In a final analysis, use of a brief symptom measure was categorized as either PHQ-2, PHQ-9 (the 2 measures used by study practices; see Results), or none to examine their adjusted relationships with the study outcomes. All analyses were adjusted for trial intervention group and stratification variables: practice setting (academic vs nonacademic) and patient sex and race/ethnicity (non-Hispanic white or other). Regressions were weighted to adjust for the oversampling in the RCT of people with more depres-

sion symptoms, yielding estimates applicable to unselected primary care samples.

Results

Figure 1 shows the flow of participants through the parent RCT and the derivation of the current analytic sample comprising participants with a PHQ-9 score <10 and complete visit record review data ($n = 595$). In the current sample, the mean number of patients enrolled per provider was 4.8 (range, 1–11). Table 1 shows the characteristics of the analytic sample by practice use of a brief depression symptom measure (used vs not) and overall. Medical records revealed practice administration of a brief depression symptom measure (ie, not the study-administered PHQ-9) for 50 patients (8%). Men ($P < .001$) and patients in Veterans Administration and health maintenance organiza-

Table 1. Characteristics of the Study Sample by Depression Screener Status and Overall

| Characteristics | Practice or Provider Use of a Brief Depression Symptom Measure at the Study Visit* | | Total |
|---------------------------------|--|-------------|-------------|
| | Did Not Use | Used | |
| Total | 545 (91.6) | 50 (8.4) | 595 (100.0) |
| Categorical variables | | | |
| Patient sex | | | |
| Female | 316 (58.0) | 11 (22.0) | 327 (55.0) |
| Male | 229 (42.0) | 39 (78.0) | 268 (45.0) |
| Race/ethnicity | | | |
| Non-Hispanic white | 328 (60.2) | 30 (60.0) | 358 (60.2) |
| Non-Hispanic black | 88 (16.1) | 5 (10.0) | 93 (15.6) |
| Hispanic (any race) | 65 (11.9) | 9 (18.0) | 74 (12.4) |
| Non-Hispanic, other race | 64 (11.7) | 6 (12.0) | 70 (11.8) |
| Education | | | |
| College graduate | 291 (53.8) | 24 (48.0) | 315 (53.3) |
| Some college | 175 (32.3) | 13 (26.0) | 188 (31.8) |
| High school graduate | 65 (12.0) | 12 (24.0) | 77 (13.0) |
| <High school | 10 (1.8) | 1 (2.0) | 11 (1.9) |
| Practice setting | | | |
| Multispecialty group | 269 (49.4) | 5 (10.0) | 274 (46.1) |
| Academic | 179 (32.8) | 12 (24.0) | 191 (32.1) |
| Veterans Administration | 60 (11.0) | 25 (50.0) | 85 (14.3) |
| Health maintenance organization | 37 (6.8) | 8 (16.0) | 45 (7.6) |
| Trial intervention group | | | |
| Intervention 1 | 192 (35.2) | 19 (38.0) | 211 (35.5) |
| Intervention 2 | 176 (32.3) | 13 (26.0) | 189 (31.8) |
| Attention control video | 177 (32.5) | 18 (36.0) | 195 (32.8) |
| Continuous variables | | | |
| Age, mean (SD) | 52.4 (11.7) | 53.5 (12.5) | 52.5 (11.8) |
| Health status | | | |
| Mental (MCS-12), mean (SD) | 51.2 (9.2) | 48.5 (8.0) | 51.0 (9.2) |
| Physical (PCS-12), mean (SD) | 43.7 (12.8) | 44.1 (10.5) | 43.7 (12.6) |

Data are n (%) unless otherwise indicated.

*Based on a review of the study visit note in the medical record.

MCS-12, 12-item Short Form Mental Component Summary score; PCS-12, 12-item Short Form Physical Component Summary score; SD, standard deviation.

tion settings ($P < .001$) were overrepresented in the group with practice-administered measures. There was evidence of practice administration of a brief symptom measure to at least one study patient for 33 of the 135 study providers (24.4%; median 30%; interquartile range, 20% to 41.7%).

Table 2 shows the proportion of study patients overall and the proportion of patients with a practice-administered depression symptom measure and who had evidence in the medical record of an antidepressant recommendation and/or prescription at various cut points of the study-administered PHQ-9. Across the PHQ-9 score categories, an

antidepressant was recommended and/or prescribed in 4% to 7% of visits overall versus in 22% to 30% of visits with a practice-administered brief depression symptom measure (Table 2). Among the 545 patients (91.6% of the total sample) without record evidence of a practice-administered symptom measure, 57 (10.5%) had a visit diagnosis of depression; 9 (1.6%) were recommended and another 21 (3.8%) prescribed an antidepressant. Among the 50 patients (8.4% of the total sample) with evidence of a practice-administered measure, 10 (20%) had a visit diagnosis of depression; 6 (12%) were recommended and another 6 (12%)

Table 2. Number and Percentage of Study Visits with Depression Diagnosis and with Antidepressant Recommendation and/or Prescription at Various Study 9-Item Patient Health Questionnaire (PHQ-9) Score Thresholds

| PHQ-9 Score | All Visits | | | Visits with Practice or Provider Use of a Brief Depression Symptom Measure | | |
|-------------|------------|-----------------------|--|--|-----------------------|--|
| | Visits (n) | Depression Diagnosis* | Antidepressant Recommendation and/or Prescription* | Visits (n) | Depression Diagnosis* | Antidepressant Recommendation and/or Prescription* |
| <10 | 595 | 67 (11.3) | 42 (7.1) | 50 | 10 (20.0) | 12 (24.0) |
| <9 | 538 | 56 (10.4) | 32 (6.0) | 43 | 10 (23.3) | 10 (23.3) |
| <8 | 486 | 45 (9.3) | 31 (6.4) | 36 | 8 (22.2) | 9 (25.0) |
| <7 | 421 | 35 (8.3) | 21 (5.0) | 28 | 6 (21.4) | 8 (28.6) |
| <6 | 369 | 29 (7.9) | 19 (5.2) | 23 | 5 (21.7) | 7 (30.4) |
| <5 | 304 | 20 (6.6) | 11 (3.6) | 18 | 4 (22.2) | 4 (22.0) |

Data are n (%) unless otherwise indicated. The PHQ-9 was administered to all participants in the prestudy visit questionnaire (but was not disclosed to study providers).

*Based on a review of the study visit note in the medical record.

prescribed an antidepressant. Accounting for nesting of patients within providers and trial sample weighting, practice administration of a brief symptom measure was associated with increased odds of antidepressant recommendation and/or prescription (adjusted odds ratio [AOR] 4.4; 95% confidence interval [CI], 2.3–8.6) and with increased odds of antidepressant prescription alone (AOR, 3.8; 95% CI, 1.6–9.1).

In multivariable analyses, the odds of depression diagnosis, antidepressant recommendation and/or prescription, and antidepressant prescription alone were higher among those with versus without a practice-administered brief depression symptom

measure (Table 3). For all outcomes, analyses at progressively lower PHQ-9 score thresholds (<9 to <5) yielded similar findings (Table 3).

Among the 50 participants completing a practice-administered brief measure, 26 completed the PHQ-2 and 23 completed the PHQ-9 (1 had missing data). Among those completing the PHQ-9, 2 had scores >10 (10 in one, 11 in the other). In a logistic regression analysis excluding these 2 participants and the participant with missing data for the specific brief measure, the likelihood of antidepressant recommendation and/or prescription was increased for practice administration of the PHQ-9 (AOR, 10.0; 95% CI, 1.8–55.4) but not the PHQ-2 (AOR, 1.1; 95% CI,

Table 3. Adjusted Odds of Depression Diagnosis, Antidepressant Recommendation, and/or Prescription and Antidepressant Prescription Alone Associated with Use of Brief Depression Symptom Measure

| PHQ-9 Score | Depression Diagnosis* | | Antidepressant Recommendation and/or Prescription* | | Antidepressant Prescription Alone* | |
|-------------|-----------------------|------|--|------|------------------------------------|-----|
| | AOR (95% CI) | P | AOR (95% CI) | P | AOR (95% CI) | P |
| <10 | 3.2 (1.1–9.2) | .02 | 3.8 (1.0–13.9) | .04 | 4.1 (1.1–16.2) | .04 |
| <9 | 4.4 (1.4–13.1) | .008 | 4.8 (1.1–20.2) | .04 | 6.0 (1.2–30.9) | .03 |
| <8 | 4.7 (1.5–15.1) | .01 | 4.4 (1.0–19.7) | .05 | 6.3 (1.2–33.3) | .03 |
| <7 | 6.1 (1.6–22.7) | .007 | 7.6 (1.8–32.2) | .006 | 7.0 (1.3–37.6) | .02 |
| <6 | 5.5 (1.4–22.3) | .02 | 6.9 (1.4–35.1) | .02 | 7.1 (1.3–38.6) | .02 |
| <5 | 11.1 (2.6–47.6) | .001 | 155.2 (0.7–32,540.6) | .06 | 1.7 (0.1–24.8) | .69 |

The PHQ-9 was administered to all participants in the prestudy visit questionnaire (but was not disclosed to study providers).

*Based on a review of the study visit note in the medical record. Values are adjusted for nesting of patients within providers; trial intervention group; trial stratification variables (practice setting [academic versus nonacademic], patient sex, and patient race/ethnicity category [non-Hispanic white or all others]); and trial sample weighting (ie, oversampling of people with higher levels of depressive symptoms [PHQ-9 score >5]).

AOR, adjusted odds ratio; CI, confidence interval; PHQ-9, 9-item patient health questionnaire.

0.3–4.0). The adjusted marginal probability of antidepressant recommendation and/or prescription was 0.03 for patients with no practice-administered measure, 0.04 for those completing a PHQ-2, and 0.23 for those completing a PHQ-9. Similarly, the likelihood of a depression diagnosis was increased for practice administration of the PHQ-9 (AOR, 5.3; 95% CI, 1.4–20.6) but not the PHQ-2 (AOR, 1.7; 95% CI, 0.4–8.1). The adjusted marginal probability of a depression diagnosis was 0.08 for patients with no practice-administered measure, 0.13 for those completing a PHQ-2, and 0.31 for those completing a PHQ-9.

Discussion

In exploratory analyses of data from an RCT focused on participants with relatively few or no depression symptoms (defined by a PHQ-9 score <10), practice administration of a brief depression symptom measure during office visits was associated with hypothesized increases in depression diagnosis and antidepressant recommendation and/or prescription. Additional analyses suggested these findings were associated primarily with the PHQ-9 (not the PHQ-2).

Caution is required when interpreting our findings, given their preliminary nature and the limitations of the study design. The parent RCT did not include an expert diagnostic interview for depression. It is possible that some patients had clinical depression. However, PHQ-9 validation studies using an expert diagnostic interview reference standard indicate only about 1% of people scoring <10 have major depression,^{4,6} the only form of depression consistently shown in RCTs to benefit from antidepressants.^{13–15} Also, we lacked information regarding the indication(s) for antidepressant recommendation and prescription. Thus we examined whether practice use of a brief measure was associated with the diagnosis of depression in the visit note to help gauge the likelihood that antidepressants were recommended or prescribed for depression versus other indications. That use of a brief depression symptom measure had parallel associations with both depression diagnosis and antidepressant recommendation and prescription (Table 3) suggests that a depression diagnosis was associated with these antidepressant recommendations and prescriptions.

In turn, our findings tentatively suggest practice administration of brief depression symptom mea-

sures may be associated with overdiagnosis of depression and overtreatment with antidepressants, both prevalent in primary care.^{16,17} Our exploratory findings suggest the need for more definitive studies. The wide confidence intervals around the point estimates for the study outcomes stem from the relatively few participants with these events and with record evidence of use of a brief symptom measure, signaling further need for caution in interpreting the findings. Still, use of a brief symptom measure was associated with increased depression diagnosis and antidepressant recommendation and prescription in appropriately weighted analyses adjusted for potential confounders and in patient subsamples defined using progressively lower PHQ-9 cut points (with a correspondingly decreasing likelihood of major depression).^{4,6} Thus one might expect similar odds of depression diagnosis and antidepressant recommendation and prescription associated with use of a brief symptom measure in unselected primary care samples.

While the observational nature of our analyses precludes causal inference, we consider 2 plausible hypotheses regarding possible mechanisms. One is that use of a brief symptom measure is simply a marker for providers or practices that place emphasis on identifying and treating depression. Prior work suggests considerable variation among providers in this regard.³¹ Highly engaged providers might be more likely to use brief symptom measures for case finding or screening. They might also be inclined to choose more comprehensive brief measures like the PHQ-9 over briefer options like the PHQ-2, potentially contributing to the differences in findings for these measures. At the same time, such providers might have a lower threshold for treating with antidepressants because of their greater interest in and sensitivity to depression cues, even for patients not meeting criteria for major depression.

A second possibility is that use of brief symptom measures could foster overtreatment with antidepressants by nonspecifically heightening provider consideration of major depression, a difficult diagnosis for many providers.³ Such prompting could occur directly (eg, provider is handed a completed brief measure), indirectly through patient activation (eg, patient asks provider questions about depression after working through a brief measure), or by both routes. Such nonspecific prompting might be triggered more by the detailed PHQ-9 than the

brief PHQ-2. However, both heightened provider engagement and nonspecific prompting associated with use of a brief symptom measure could lead to more appropriate treatment of patients with major depression. Studies designed to explore the balance of risks and benefits associated with the commonly used brief measures are needed.

The CTFPHC expressed concerns about overtreatment with antidepressants in their 2013 statement recommending against use of brief depression symptom measures.¹¹ Nonetheless, both the CTFPHC and the USPSTF acknowledged that no studies have examined the associations of the use of such measures with overtreatment with antidepressants.^{7,11} Our findings call for RCTs designed and powered to address the associations of commonly used brief symptom measures with potential risks (eg, overtreatment with antidepressants leading to unnecessary adverse medication effects) and benefits (eg, increased recognition and treatment of depression, reduced functional limitations, fewer suicides) and to explore causal mechanisms.

In the absence of such RCTs, the USPSTF's qualified endorsement of office use of brief depression symptom measures was predicated on the presumption of benefits and did not consider possible risks.^{7,10} The USPSTF acknowledged extrapolating from RCTs in which routine use of brief depression symptom measures was bundled with immediate enrollment of patients with major depression into collaborative care.⁷ These RCTs demonstrated improved outcomes for patients with major depression but apart from collaborative care did not examine the independent effects of use of a brief depression symptom measure. Confirmation in future studies of our findings suggesting potential overdiagnosis of depression and overtreatment with antidepressants associated with use of a brief symptom measure would suggest the need to re-evaluate the USPSTF recommendation.

Future confirmation of our findings also would suggest the need to move beyond the use of simple brief depression symptom measures for case finding or screening to developing and studying the impact of novel tools that could help to better match depression treatment to patient need. Individually tailored patient activation computer programs that incorporate brief depression symptom measures represent one example.²² Rather than simply informing patients of their symptom measure scores, such programs provide individualized information

and motivational messages.³² In the context of significant depression symptoms, the programs provide messages to activate the patient to discuss the symptoms with a provider and to enhance receptiveness to offers of treatment. Conversely, in the context of few or no depression symptoms, tailored programs inform patients that they are not likely to be depressed and not likely to benefit from treatment, potentially buffering against subsequent provider recommendation of an antidepressant. Initial RCT evidence indicates the promise of such tailored programs.²² Nonetheless, larger trials, powered to detect small but potentially clinically significant increases in diagnosis of depression and treatment with antidepressants in patients with few or no symptoms, are needed to determine whether the programs offer advantages over isolated use of brief symptom measures.

Beyond the study limitations noted previously, the parent RCT involved interventions to improve depression recognition and treatment, raising the possibility of a Hawthorne effect influencing our findings. Participating providers were not informed that depression was the focus of the study.^{22,28} Nonetheless, some may have been prompted to consider and treat depression during participation (eg, after encountering patients activated by the interventions to discuss depression), potentially increasing their tendency to use brief measures and/or prescribe antidepressants. Also, we lacked information regarding the specifics of provider use of brief symptom measures, such as the indication (eg, case finding vs screening) and interpretation (eg, cut points used to define clinical depression). Despite these limitations, given there is little evidence supporting benefits of antidepressants among patients lacking major depression,^{13–15} the findings have potential clinical relevance regardless of the reasons for and approaches to using the brief symptom measures. If our key findings are replicated by others, studies designed specifically to address mechanisms will be required to tease out the “causes” of the observed association.

The accuracy of the medical record information we used is unknown, with uncertain net impact on the findings. Other ascertainment methods also have drawbacks. For example, audio or video recording of visits may alter patient and/or provider behaviors. The generalizability of the findings is also uncertain, particularly to categories of individuals ineligible for the parent RCT (eg, non-English

speaking people, individuals with sensorimotor impairments precluding use of a touchscreen computer). Replicating our findings in studies with samples that include such people and that use other methods of ascertaining use of brief depression symptom measures, antidepressant recommendation and treatment, and their indications and applications will be useful.

Conclusion

In exploratory observational analyses of RCT data from patients unlikely to have major depression, and therefore unlikely to benefit from antidepressants, use of a brief depression symptom measure during an office visit was associated with increased depression diagnosis and increased antidepressant recommendation and/or prescription. Analyses examining the specific brief measure used (PHQ-2 vs PHQ-9) suggested these associations were primarily attributable to use of the PHQ-9. Further studies are needed to confirm and explore the mechanisms of these findings and to investigate the balance of benefits and risks associated with the use of brief depression symptom measures.

The authors are grateful to the following individuals, who coordinated or facilitated recruitment and participation of patients in the study: Christina Slee, MPH, Julia Huerta, MPH, and Dustin Gottfeld, BS (University of California Davis); Sarah Olson, BA, Ana Fernandez-Lamothe, Jeff Kohlwe, MD, and Seth Berkowitz, MD (University of California San Francisco). The authors also are indebted to all the physicians, offices, and patients who participated.

References

1. Rost K, Nutting P, Smith J, Coyne JC, Cooper-Patrick L, Rubenstein L. The role of competing demands in the treatment provided primary care patients with major depression. *Arch Fam Med* 2000; 9:150–4.
2. Bell RA, Franks P, Duberstein PR, et al. Suffering in silence: reasons for not disclosing depression in primary care. *Ann Fam Med* 2011;9:439–46.
3. Mitchell AJ, Rao S, Vaze A. Do primary care physicians have particular difficulty identifying late-life depression? A meta-analysis stratified by age. *Psychother Psychosom* 2010;79:285–94.
4. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
5. Lowe B, Kroenke K, Grafe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res* 2005;58:163–71.
6. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. *Psychiatr Ann* 2002;32:509–21.
7. Screening for depression in adults: U.S. preventive services task force recommendation statement. *Ann Intern Med* 2009;151:784–92.
8. Katz PS. Observer extra: depression. An internist's guide to screening, diagnosing and treating clinical depression. Philadelphia: American College of Physicians; 2006. Available from: <http://www.acpinternist.org/archives/2006/12/depression.pdf>. Accessed January 13, 2014.
9. Mitchell J, Trangle M, Degnan B, et al. Adult depression in primary care. 16th ed. Bloomington, MN: Institute for Clinical Systems Improvement; 2013. Available from: <https://www.icsi.org/asset/fnhdm3/Depr-Interactive0512b.pdf>. Accessed January 13, 2014.
10. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med* 2009;151:793–803.
11. Joffres M, Jaramillo A, Dickinson J, et al. Recommendations on screening for depression in adults. *CMAJ* 2013;185:775–82.
12. Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol* 2008;76:909–22.
13. Barbui C, Cipriani A, Patel V, Ayuso-Mateos JL, van Ommeren M. Efficacy of antidepressants and benzodiazepines in minor depression: systematic review and meta-analysis. *Br J Psychiatry* 2011;198:11–6, sup 1.
14. Hegerl U, Schonknecht P, Mergl R. Are antidepressants useful in the treatment of minor depression: a critical update of the current literature. *Curr Opin Psychiatry* 2012;25:1–6.
15. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 2010;303:47–53.
16. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;374:609–19.
17. Mojtabai R. Clinician-identified depression in community settings: concordance with structured-interview diagnoses. *Psychother Psychosom* 2013;82: 161–9.
18. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med* 2011;155:772–85.
19. Schmitt MR, Miller MJ, Harrison DL, Touchet BK. Relationship of depression screening and physician office visit duration in a national sample. *Psychiatr Serv* 2010;61:1126–31.
20. Jaen CR, Stange KC, Nutting PA. Competing de-

- mands of primary care: a model for the delivery of clinical preventive services. *J Fam Pract* 1994;38:166–71.
21. Parchman ML, Pugh JA, Romero RL, Bowers KW. Competing demands or clinical inertia: the case of elevated glycosylated hemoglobin. *Ann Fam Med* 2007;5:196–201.
22. Kravitz RL, Franks P, Feldman MD, et al. Targeted and tailored patient engagement programs for recognition and initial treatment of depression in primary care: a randomized trial. *JAMA* 2013;310:1818–28.
23. Aikens JE, Nease DE Jr, Klinkman MS. Explaining patients' beliefs about the necessity and harmfulness of antidepressants. *Ann Fam Med* 2008;6:23–9.
24. Johnson MD, Meredith LS, Hickey SC, Wells KB. Influence of patient preference and primary care clinician proclivity for watchful waiting on receipt of depression treatment. *Gen Hosp Psychiatry* 2006;28:379–86.
25. van Schaik DJ, Klijn AF, van Hout HP, et al. Patients' preferences in the treatment of depressive disorder in primary care. *Gen Hosp Psychiatry* 2004;26:184–9.
26. Esserman LJ, Thompson IM, Reid B. Overdiagnosis and overtreatment in cancer: an opportunity for improvement. *JAMA* 2013;310:797–8.
27. Lauer MS. Pseudodisease, the next great epidemic in coronary atherosclerosis?: comment on “Impact of coronary computed tomographic angiography results on patient and physician behavior in a low-risk population”. *Arch Intern Med* 2011;171:1268–9.
28. Tancredi DJ, Slee CK, Jerant A, et al. Targeted versus tailored multimedia patient engagement to enhance depression recognition and treatment in primary care: randomized controlled trial protocol for the AMEP2 study. *BMC Health Serv Res* 2013;13:141.
29. Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J Affect Disord* 2009;114:163–73.
30. Ware J Jr, Kosinski M, Keller SD. A 12-item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34:220–33.
31. Mojtabai R. Diagnosing depression and prescribing antidepressants by primary care physicians: the impact of practice style variations. *Ment Health Serv Res* 2002;4:109–18.
32. Jerant A, Sohler N, Fiscella K, Franks B, Franks P. Tailored interactive multimedia computer programs to reduce health disparities: opportunities and challenges. *Patient Educ Couns* 2011;85:323–30.