Screening for Depression in Adults

TO THE EDITOR: Pignone and colleagues’ call for routine screening of general medical patients for depression (1) contrasts with past conclusions that the strength of evidence is insufficient for such a recommendation. The discrepancy is explained by both a shift in the question being addressed and the kind of evidence considered. Rather than evaluating screening in routine care, Pignone and colleagues examined whether routine screening combined with enhanced routine care improves outcomes. Fully 84% of the weight of evidence in their key meta-analysis is carried by three studies. Two (2, 3) yielded null results, so the positive effect is entirely carried by the study by Wells and colleagues (4). Screening was only a small part of this study, which recruited highly committed practices and provided training to staff, training materials, academic detailing, and additional staff. Also, a request was made that appointments be scheduled within 2 weeks for patients identified as depressed, and practices were provided with tracking lists and assistance in initiating and continuing treatment, as well as access to medication follow-up and brief, structured psychotherapy.

A primary reason that screening has not been shown to benefit patients is that under conditions of routine care, identification of depressed patients does not reliably improve outcomes (5). Wells and colleagues (4) and others have shown that short-term outcomes can be improved through the commitment of substantial resources and sustained effort. In the face of competing demands for resources, it would be difficult to introduce and maintain such modifications of routine care outside of a research project. The lack of appropriate reimbursement for such enhanced care is the bulk of the problem. Our concern is that Pignone and colleagues’ apparent endorsement of screening will distract from pursuit of a solution. Implemented by itself, screening serves only to commit more patients to ineffective care.

James C. Coyne, PhD
Steven C. Palmer, PhD
Patricia A. Sullivan, PhD
University of Pennsylvania Cancer Center
Philadelphia, PA 19104

References

TO THE EDITOR: The recommendations of the U.S. Preventive Services Task Force (USPSTF) (1) for screening for depression advise that a positive screening result should lead to a full diagnostic evaluation “to determine the presence or absence of specific depressive disorders, such as major depression or dysthymia.” The recommendations do not mention that a major depressive episode, as defined in the Diagnostic and Statistical Manual of Mental Disorders (2), can be a phase of a bipolar disorder or a major depressive disorder. This distinction is of great clinical significance.

Bipolar disorders, which include bipolar I, bipolar II, and cyclothymia, are relatively common mental disorders. They often pass undetected in general medical settings because patients rarely seek care when they are manic or hypomanic but are likely to do so when they are depressed. General physicians may not be familiar with mania or hypomania, and a history of mood elevations can easily be missed. In one family practice, for example, a careful assessment of consecutive adult patients presenting with anxiety or depressive symptoms revealed that approximately one quarter had a lifetime history of bipolar disorders (3). In the treatment of depression for patients who have bipolar disorders, there is a significant risk for complications if an antidepressant medication is prescribed without concurrent use of a mood stabilizer, such as lithium. Complications can include precipitation of mania or hypomania, mixed affective states, and rapid cycling between manic and depressive episodes (4). The USPSTF recommendations should thus specify the need to assess a history of mania or hypomania following a positive screening result for depression. The Mood Disorder Questionnaire (5), a brief screening measure for bipolar disorders, has been recently validated and can be used in general medical practice. Future studies on depression screening should make greater efforts to distinguish cases of bipolar disorders from those of major depressive disorders.

Amar K. Das, MD, PhD
Raz Gross, MD, MPH
Myrna M. Weissman, PhD
Columbia University
New York, NY 10032

References

IN RESPONSE: We agree with Coyne and colleagues that outcomes for depressed patients in primary care are better when screening is coupled with systematic efforts to improve initiation and continuation of effective therapy, as exemplified in the study by Wells and

© 2003 American College of Physicians

Downloaded from http://annals.org by guest on 12/27/2018
colleagues (1). The recommendation statement by the USPSTF (2) that accompanied our systematic review also explicitly noted that screening was more effective when coupled with systematic support. Rather than "distract[ing] from efforts to promote greater access to such services, we believe our evidence review and the USPSTF recommendation will further encourage payers and health systems to make the changes necessary to provide high-quality care for patients with depression. Existing evidence suggests that such care includes systematic efforts to identify patients with depression ("screening" or "case finding").

The question of whether screening without systematic support for treatment and follow-up produces any benefit has generated a great deal of controversy (3). Critics often cite the number of "negative" studies as evidence of lack of efficacy, but many of the screening trials were too small to exclude important benefits of screening. Meta-analysis of the trials identified in our review, excluding the trials by Wells and Katzelnick and their colleagues (1, 4), shows that screening without such support produces an absolute reduction of 6 percentage points (95% CI, 0 to 12 percentage points) in the proportion of patients depressed at 6 months (Figure). This effect is smaller than the effect seen in the overall meta-analysis, which found a reduction of 9 percentage points (CI, 4 to 14 percentage points). These findings are consistent with our conclusion that screening without systematic support is probably more effective than usual care but that screening with systematic support is better than screening without it.

Michael Pignone, MD, MPH
Bradley Gaynes, MD, MPH
Kathleen Lohr, PhD
Research Triangle Institute–University of North Carolina
Evidence-based Practice Center
Chapel Hill, NC 27514

Jerry Rushton, MD, MPH
University of Michigan
Ann Arbor, MI 48109

Cynthia Mulrow, MD, MSc
University of Texas Health Science Center
San Antonio, TX 78284

References

Screening for Breast Cancer

TO THE EDITOR: We write in response to Humphrey and colleagues’ article on breast cancer screening (1). The annual rate of new, treated cases of breast cancer has more than tripled since mammography screening began in the United States, an increase exclusive to women who have been screened.

Four random autopsy studies (2–5), using Kramer’s exhaustive microscopic detection technique, predict that 16 million women with undiagnosed breast cancer are present in the current population of the United States. Of those cancers, 88% are in situ and 12% are

Table

<table>
<thead>
<tr>
<th>Study (Reference)</th>
<th>Treatment Group, n/n</th>
<th>Control Group, n/n</th>
<th>RD (95% CI Random)</th>
<th>Weight, %</th>
<th>RD (95% CI Random)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Williams et al. (5)</td>
<td>56/153</td>
<td>30/65</td>
<td>16.0</td>
<td>-0.10 (-0.24 to 0.05)</td>
<td></td>
</tr>
<tr>
<td>Whooley et al. (6)</td>
<td>41/97</td>
<td>54/109</td>
<td>17.7</td>
<td>-0.07 (-0.21 to 0.06)</td>
<td></td>
</tr>
<tr>
<td>Callahan et al. (7, 8)</td>
<td>87/100</td>
<td>66/75</td>
<td>31.1</td>
<td>-0.01 (-0.11 to 0.09)</td>
<td></td>
</tr>
<tr>
<td>Zung and King (9)</td>
<td>7/21</td>
<td>13/20</td>
<td>4.1</td>
<td>-0.32 (-0.61 to -0.03)</td>
<td></td>
</tr>
<tr>
<td>Lewis et al. (10)</td>
<td>111/161</td>
<td>117/157</td>
<td>31.1</td>
<td>-0.06 (-0.15 to 0.04)</td>
<td></td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>302/532</td>
<td>280/426</td>
<td>100.0</td>
<td>-0.06 (-0.12 to 0.00)</td>
<td></td>
</tr>
</tbody>
</table>

For the overall effect, P = 0.04. RD = risk difference.
invasive. Most women with breast cancer do not benefit from treatment, since it would take 400 years for the 16 million U.S. women with undiagnosed breast cancer to die of the disease at the current mortality rate of 40 000 women per year.

Medicine does not employ a less sensitive or less specific test than screening mammography. In 2001, 1.5 million of 30 million U.S. mammograms were interpreted as false positive for breast cancer, promising every woman following screening guidelines two negative breast biopsies in her natural lifetime. In 2001, when two thirds of U.S. women with 10 million cases of breast cancer were screened, 250 000 cases of breast cancer were detected.

There is no justification for the current practice of treating every breast cancer detected and no sense to detection until the few breast cancers that threaten women can be distinguished from the majority, which do not. Breast cancer treatments are not harmless.

A woman who undergoes screening is more than five times more likely to become a patient with breast cancer than a woman who does not. It is misleading for survivors of unnecessary breast cancer treatments to claim in public advertisements to be survivors of breast cancer. Most women with breast cancer survive without treatment.

Data from the National Center for Health Statistics show that 28 830 women died of breast cancer in 1970 (mortality rate, 27 per 100 000 women) and 41 737 died of breast cancer in 1997 (mortality rate, 30.6 per 100 000 women). Four hundred fifty million screening mammograms in the United States over 30 years did not reduce breast cancer deaths but pointlessly increased treatments.

Breast cancer screening does not save lives, it creates patients with cancer. Overtreatment of breast cancer is the rule, not the exception.

John K. Whiteford, MD
Verona, PA 15147

John Ryan Whiteford, MD
University of Pittsburgh
Pittsburgh, PA 15260

References

TO THE EDITOR: Humphrey and colleagues (1) noted that misclassification of cause of death could change the direction of the results of the trials they examined. However, they argued that the reduction in rates of advanced breast cancer in the Two-County Study (which are not related to judgments on cause of death) is similar to the reduction in breast cancer mortality. The problem with this reasoning is that assessment of cause of death is unreliable and biased in favor of screening (2). This was recently confirmed. The 2002 overview of the Swedish trials (3) reported only a 10% reduction in breast cancer mortality for the Östergotland part of the Two-County Study, whereas the authors of the Two-County Study recently reported a 24% reduction (Table) (4).

In its recommendations (5), the U.S. Preventive Services Task Force lists under harms only false-positive test results and unnecessary anxiety, biopsies, and cost. I find it pretty euphemistic to talk about unnecessary “biopsies” when, in fact, there is good evidence from the randomized trials that screening increases both tumorectomies and mastectomies because of overdiagnosis and overtreatment (2). Tumorectomy is usually followed by radiation therapy, which is likely to reduce overall survival because of damage to the heart and vessels (2). These findings and the lack of an effect on all-cancer mortality (including breast cancer) support our conclusion that “the currently available reliable evidence does not show a survival benefit of mass screening for breast cancer” (2).

Informed consent for women contemplating whether to attend a screening program is an illusion as long as the screening advocates and their organizations don’t honestly convey the uncertainty about the survival benefit and continue to omit information on major harms. Women could become better informed by consulting a consumer organization, such as the National Breast Cancer Coalition (www.stopbreastcancer.org).

Peter C. Gøtzsche, MD
Righospitalet Nordic Cochrane Centre
DK-2100 Copenhagen Ø, Denmark

Disclaimer: The views expressed are those of the author and are not necessarily the views of or the official policy of the Cochrane Collaboration.

References

Table. Breast Cancer Mortality for the Östergotland Part of the Swedish Two-County Study, as Reported in a Recent Overview of the Swedish Mammography Screening Trials and a Recent Update of the Two-County Study*

<table>
<thead>
<tr>
<th>Study, Year (Reference)</th>
<th>Breast Cancer Deaths, n</th>
<th>Time in Person-Years of Follow-up (in Thousands)</th>
<th>Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Invited Group</td>
<td>Control Group</td>
<td></td>
</tr>
<tr>
<td>Nystrom et al., 2002 (3)</td>
<td>177</td>
<td>190</td>
<td>569</td>
</tr>
<tr>
<td>Tabar et al., 2000 (4)</td>
<td>167</td>
<td>213</td>
<td>660</td>
</tr>
</tbody>
</table>

* Age group, 40 to 74 years; evaluation model.
IN RESPONSE: We agree with Drs. Whiteford and Whiteford that mammography detects many non–life-threatening malignancies and that, fortunately, it misses many more microscopic ones. Mammography is far from an ideal screening test. However, for clinically important cancer, the sensitivity and specificity of mammography compare favorably with those of other screening tests. Drs. Whiteford and Whiteford also claim that the incidence of breast cancer has tripled because of wider use of mammography and that the death rate from breast cancer has risen sharply. Actually, between 1973 and 1999, a period in which mammography rates increased substantially, the annual incidence of invasive breast cancer increased from 98.5 to 139.1 per 100,000 women (1). Over this period, age-adjusted breast cancer deaths decreased from 32.3 to 27 per 100,000 women.

We agree with Dr. Götzsche that classification of cause of death can be biased, but misclassification is not necessarily biased in favor of screening. Misclassification could as easily bias findings against a real mammography benefit by virtue of “sticky diagnosis” bias (2). In any case, rates of advanced breast cancer are not affected by classification of the cause of death, which is why we cite them as an independent measure of the impact of screening. We also agree with Dr. Götzsche and Drs. Whiteford and Whiteford that a full assessment of the harms of mammography should include the predictable harms of treatment, including the risk for undergoing lumpectomies and mastectomies because of overdiagnosis and overtreatment. This issue was discussed in our Appendix, in presentations to the U.S. Preventive Services Task Force, and in the full systematic evidence review on which our Annals article was based (3).

Because treatment options, radiation doses, and practice styles have changed substantially since the trials we examined were done, we did not see value in focusing on evidence of overtreatment and overdiagnosis. If treatment and surgeons have become more conservative over time, the harms of mammography might be less severe today than they were in the 1970s and 1980s. If cancer (particularly ductal carcinoma in situ) is treated too aggressively today, the harms might be greater. Unfortunately, the data were not suitable for assessing the actual harms of treatment in current practice, so we settled for informing the Task Force of these uncertainties. Better data are needed to assess the impact of overdiagnosis and overtreatment in today’s practice settings.

Linda L. Humphrey, MD, MPH
Mark Helfand, MD, MS
Benjamin K.S. Chan, MS
Oregon Health & Science University
Portland Veterans Affairs Medical Center
Portland, OR 97239

References

The Mammography Dilemma

TO THE EDITOR: The editorial by Dr. Steven Goodman (1) on the U.S. Preventive Task Force (USPSTF) recommendations for breast cancer screening (2) made important observations on the difficulties in crafting evidence-based recommendations regarding mammography. The members of the USPSTF share Goodman’s belief that no amount of rigor in critiquing evidence will extinguish the mammography debate or completely eliminate the subjectivity of grading evidence, but we take exception to three points.

First, Goodman asserted that the USPSTF was inattentive to the harms of mammography. The full USPSTF recommendation and rationale statement, only part of which was published in Annals, more thoroughly discusses the potential harms of screening (3). The USPSTF considers the harms of screening in all of its assessments, highlighting the anxiety, labeling effects, false-positive results, and complications induced by the cascade of tests and the treatments that they propagate. Our mammography report discussed the potential for mammography to cause anxiety, unnecessary biopsies, and the detection and treatment of cancer of uncertain clinical significance (for example, ductal carcinoma in situ). In addition, our recommendation noted the balance between benefits and harms that becomes steadily more favorable as women get older. Ductal carcinoma in situ accounts for a large and growing proportion of screening-detected cancer. To the extent that treatment of ductal carcinoma in situ constitutes overtreatment, screening mammography would be expected to induce unnecessary surgery, although we did not find good evidence to confirm this suspicion. The best evidence of harms relates to false-positive mammography and biopsy results.

Second, although Goodman made important epistemological observations about the inability of research to define truth, we dispute the implication that appraisals based on randomized trials involving 400,000 women offer no clearer insights to the truth than an expert’s guesses based on personal opinion and experience. Evidence-based medicine does not expect trial data to eliminate uncertainty, only to reduce it. The most important attribute of USPSTF recommendations is that the rationale is transparent. The letter grade ‘B’ that the USPSTF assigned to the mammography recommendation and the supporting narrative make explicit our concerns about the strength of the evidence.

Finally, Goodman faulted the USPSTF for not recalculating trial results to capture the uncertainty introduced by design flaws. Although we agree in principle with this Bayesian goal, in our view limitations in current data and statistical methods would make such adjustments less transparent and of uncertain validity. We therefore used language rather than numbers to express our concerns about the quality of evidence.

In the face of these uncertainties and difficult tradeoffs, the USPSTF concluded that there was fair evidence that mammography was effective for women 40 years of age and older, but that clinicians should inform women about the potential benefits and harms of screening and the limitations of the test as they apply to women of
their age. Women should then consider their own values and preferences in deciding whether to be screened.

Alfred O. Berg, MD, MPH
Janet Allan, PhD, RN, CS
Steven Woolf, MD, MPH
U.S. Preventive Services Task Force
Rockville, MD 20852

References

TO THE EDITOR: Goodman’s review of the mammography controversy (1) is precisely what is required to avoid the self-referential gold standard status of randomized, controlled trials in the “new paradigm” of evidence-based medicine. Any gold standard must be incorporated into a systematic web of belief about clinical evidence that is not limited to outcomes of randomized, controlled trials but must consider the basic principles of medicine. As Goodman rightly pointed to the failure of the logical positivists to establish an indubitable standard of proof by describing strict conditions of verifiability, so evidence-based medicine will fail unless we all try to understand the relation of empirical knowledge to a system of medical evidence. This empirical evidence does not and cannot stand alone but must, as Sackett has insisted, be used judiciously. This judicious use of empirical information is and always will be the challenge to physicians.

Donald E. Stanley, MD
Nobleboro, ME 04555

Reference

IN RESPONSE: I appreciate the opportunity to clarify some points made in my editorial on the mammography controversy. I will address Drs. Berg, Allan, and Woolf’s points in reverse order. My mention of alternative approaches to representing uncertainty was not a criticism of the USPSTF methods; indeed, the USPSTF has been a leader in developing methods to translate evidence into policy recommendations. But the USPSTF has changed its methods over time (1), and I assume that methodologic progress will continue. My suggestion—not a “faulting”—was aimed at moving the USPSTF and the field toward different means of representing uncertainty.

Regarding Berg and colleagues’ second point, I did not state that the empirical evidence from 400 000 women is worth little more than “an expert’s guesses.” I said that “the justification for why studies are included or excluded from the evidence base can rest on competing claims of methodologic authority that look little different from the traditional claims of medical authority that proponents of evidence-based medicine have criticized.” The USPSTF documents stated only that the study flaws introduced biases that were “unlikely” or “inadequate” to account for the observed effects. These are expressions of opinion, expert though they may be. I do not criticize such opinions, but I do take issue with claims or implications that they are absent when evidence-based methods are used.

The first issue Berg and colleagues raise is the most critical and may be an area in which we will remain in disagreement. I stated that their discussion of increased surgical rates was oblique. By oblique I meant not presented in a way that woman and their doctors could use in decision making, which requires knowing the actual surgical risks. Neither the Web version of their report nor the published document provide the actual surgical risks (reported by Olsen and Gotzsche [2] as an increase of approximately 4 surgeries per 1000 screened women), nor do they explain why they are absent. Their letter and the published USPSTF rationale (3, 4) appear to imply that an elevation in surgery rates is acceptable if it produces a mortality benefit. While it is true that we would expect surgical rates to be higher in a mammographically screened group, the central question for physicians and for each woman (particularly those younger than 50) is whether the (possible) mortality benefit is worth these surgical risks, which exist in addition to risk for extra biopsies and anxiety due to false-positive results. Without numbers, that is hard to judge.

Steven N. Goodman, MD, MHS, PhD
Johns Hopkins School of Medicine
Baltimore, MD 21205

References

CORRECTION

Correction: Factors That Affect Accuracy of Screening Mammography

In an article on accuracy of screening mammography (1), Table 3 contained errors. The values in the rows labeled “Yes” and “No” under “Current use of HRT” should be reversed in the columns titled “Cases of Cancer,” “Sensitivity,” and “Specificity.” Values for participants currently using hormone replacement therapy and those who were not, respectively, should have been 904 cases of cancer versus 1319 cases, 72.7% versus 76.6% for sensitivity, and 91.7% versus 92.6% for specificity.

Reference
Quality Grand Rounds

TO THE EDITOR: As chief residents in an internal medicine program, we recognize that residency is a time when addressing medical mistakes is especially important. Failing to recognize errors misses the opportunity to cultivate a vigilance that is essential in medicine. Being called to task on an error, however, can be an ego-crushing experience. We agree with the editorial accompanying the first article in the Annals Quality Grand Rounds series (1) that morbidity and mortality conferences or department-wide quality assurance rounds are often inadequate forums for learning from medical errors. In addition, we think such conferences inhibit group participation.

We recently held a workshop at our institution to help residents think through a case of a patient who died after decisions that were retrospectively criticized. The participants were the four chief residents and 32 members of the housestaff from all 3 years. A summary of the case was read aloud, and the housestaff were separated into four small groups in which a chief resident facilitated discussion. The groups tried to identify where the problems occurred in the patient’s care and what could have been done differently. The groups then merged to summarize the discussions that took place, focusing on errors in clinical thinking, system failures, and cultural factors within the residency that may have played a role in the outcome.

This forum promoted the collegiality and honesty necessary for effective criticism, and the small-group format helped facilitate widespread participation. Furthermore, reconvening at the conclusion of the session was a good exercise in group thinking and problem solving. These are critical skills for the doctor-in-training to develop to help avoid the breakdown in communication exemplified by Chassin and Becher’s report, “The Wrong Patient” (2).

David Hirsch, MD
Peter Gliatto, MD
Yasmin Meah, MD
Mount Sinai Hospital
New York, NY 10029

References