IMPORTANCE Depression is common and associated with substantial burden. Suicide rates have increased over the past decade, and both suicide attempts and deaths have devastating effects on individuals and families.

OBJECTIVE To review the benefits and harms of screening and treatment for depression and suicide risk and the accuracy of instruments to detect these conditions among primary care patients.

DATA SOURCES MEDLINE, PsychINFO, Cochrane library through September 7, 2022; references of existing reviews; ongoing surveillance for relevant literature through November 25, 2022.

STUDY SELECTION English-language studies of screening or treatment compared with control conditions, or test accuracy of screening instruments (for depression, instruments were selected a priori; for suicide risk, all were included). Existing systematic reviews were used for treatment and test accuracy for depression.

DATA EXTRACTION AND SYNTHESIS One investigator abstracted data; a second checked accuracy. Two investigators independently rated study quality. Findings were synthesized qualitatively, including reporting of meta-analysis results from existing systematic reviews; meta-analyses were conducted on original research when evidence was sufficient.

MAIN OUTCOMES AND MEASURES Depression outcomes; suicidal ideation, attempts, and deaths; sensitivity and specificity of screening tools.

RESULTS For depression, 105 studies were included: 32 original studies (N=385 607) and 73 systematic reviews (including ≈2138 studies [N ≈9.8 million]). Depression screening interventions, many of which included additional components beyond screening, were associated with a lower prevalence of depression or clinically important depressive symptomatology after 6 to 12 months (pooled odds ratio, 0.60 [95% CI, 0.50–0.73]; reported in 8 randomized clinical trials [n=10 244]; I² = 0%). Several instruments demonstrated adequate test accuracy (eg, for the 9-item Patient Health Questionnaire at a cutoff of 10 or greater, the pooled sensitivity was 0.85 [95% CI, 0.79–0.89] and specificity was 0.85 [95% CI, 0.82–0.88]; reported in 47 studies [n = 11 234]). A large body of evidence supported benefits of psychological and pharmacologic treatment of depression. A pooled estimate from trials used for US Food and Drug Administration approval suggested a very small increase in the absolute risk of a suicide attempt with second-generation antidepressants (odds ratio, 1.53 [95% CI, 1.09–2.15]; n = 40 857; 0.7% of antidepressant users had a suicide attempt vs 0.3% of placebo users; median follow-up, 8 weeks). Twenty-seven studies (n = 24 826) addressed suicide risk. One randomized clinical trial (n=443) of a suicide risk screening intervention found no difference in suicidal ideation after 2 weeks between primary care patients who were and were not screened for suicide risk. Three studies of suicide risk test accuracy were included; none included replication of any instrument. The included suicide prevention studies generally did not demonstrate an improvement over usual care, which typically included specialty mental health treatment.

CONCLUSIONS AND RELEVANCE Evidence supported depression screening in primary care settings, including during pregnancy and postpartum. There are numerous important gaps in the evidence for suicide risk screening in primary care settings.

Depression is a common mental disorder in the US, with substantial economic costs. In 2019, an estimated 7.8% of US adults experienced at least 1 major depressive episode and 5.3% of adults experienced a major depressive episode with severe impairment in the past year. In 2019 in the US, 47,511 deaths were attributable to suicide. Suicide rates are increasing, with a 31% increase in suicide deaths in the US between 2001 and 2017.

In 2016, the US Preventive Services Task Force (USPSTF) recommended screening for depression in the general adult population, including pregnant and postpartum persons. The task force further stated that screening should be implemented with adequate systems in place to ensure accurate diagnosis, effective treatment, and appropriate follow-up (B recommendation). In 2014, the USPSTF concluded that the evidence was insufficient to assess the balance of benefits and harms associated with screening for suicide risk (I statement) in adolescents, adults, and older adults. This systematic review was conducted to support the USPSTF in updating its recommendations on depression and suicide risk screening.

Methods

Scope of Review

Figure 1 shows the analytic framework and key questions (KQs) that guided this review, which were developed in consultation with members of the USPSTF and covered screening for depression, anxiety, and suicide risk. The overall KQ (KQ1/1a) assessed the direct evidence on whether screening programs result in improved outcomes, while the other KQs assessed different parts of the indirect stream of evidence. There were no deviations from the original research plan. In the current publication, evidence on the benefits and harms of screening for and treatment of depression and suicide risk in adults and the accuracy of screening tools are discussed. Detailed methods and results are available in the full evidence review.

Data Sources and Searches

Ovid MEDLINE, the Cochrane Central Register of Controlled Clinical Trials, the Cochrane Database of Systematic Reviews, and PsycINFO were searched through September 7, 2022. For KQ1, KQ2, and KQ3, the search start dates were January of 2014 (depression) and January of 2012 (suicide risk) for original research (KQ1, KQ2, and KQ3, bridging from the previous USPSTF review) and January of 2015 for existing systematic reviews (KQ4 and KQ5, seeking existing systematic reviews published in the past 5 years). Detailed search strategies are listed in the eMethods in the Supplement. Searches were supplemented by reference lists of relevant reviews, including prior USPSTF reviews, and hand-searching the Cochrane Database of Systematic Reviews. Article alerts and targeted searches of journals to identify major studies published in the interim that may affect the conclusions or understanding of the evidence and the related USPSTF recommendation were used as part of ongoing surveillance. The last surveillance was conducted on November 25, 2022, and 1 existing systematic review was added to the review.

Study Selection

Two investigators independently reviewed titles, abstracts, and full-text articles using predefined eligibility criteria. For KQ1/1a and KQ3 (benefits and harms of screening), randomized clinical trials (RCTs) of adult primary care patients, including pregnant persons, investigating the benefits or harms of screening programs for depression or suicide risk were included. Studies were included that had unscreened control groups (KQ1) and in which the control group was also screened, but the screening results were not given to the participants’ primary care clinician (KQ1a). Included studies could have additional components beyond screening, such as referral support, training in diagnosis or management, and patient materials.

For KQ2 (test accuracy), evidence on depression screening instruments was limited only to prespecified tools determined to be the most widely used or recommended screening tools for depression: Patient Health Questionnaire (PHQ), any version; Center for Epidemiologic Studies Depression Scale (CES-D); Edinburgh Postpartum Depression Scale (EPDS) for perinatal persons; and Geriatric Depression Scale (GDS) for older adults. These tools had been identified as a priori as being the most widely used or recommended, based on recommendations of professional societies and government entities, systematic reviews, implementation studies, and clinicians working in some large health systems. Existing systematic reviews were used to evaluate all instruments except the GDS, which was addressed using primary test accuracy studies because no relevant, relevant existing systematic reviews were found. Primary studies were also used to examine suicide risk screening, with no restrictions on specific tools.

For KQ4 and KQ5 (benefits and harms of treatment), RCTs of psychological, pharmacological, or combination interventions among people with elevated risk of suicide compared with control conditions (eg, placebo, usual care [including usual mental health specialty care], wait list, or attention control conditions) were included. For depression, existing systematic reviews were used to address KQ4 and KQ5, adapting a decision tool developed by Pollock et al to identify the most current and comprehensive evidence.

Data Extraction and Quality Assessment

Two independent investigators rated the quality of studies as “good,” “fair,” or “poor,” using predefined criteria for each study type, in accordance with the USPSTF methods (eTable 1 in the Supplement). Discrepancies between raters were resolved by discussion or consultation with the larger review team. Studies rated as “poor” quality due to critical methodological limitations were excluded.

Data from each included study were extracted into detailed forms using DistillerSR (Evidence Partners). One reviewer completed primary data abstraction, and a second reviewer checked all data for accuracy and completeness. Study inclusion criteria, population characteristics, intervention or screening tool details, comparators, and results for outcomes defined a priori were extracted.

Data Synthesis and Analysis

Findings were synthesized using text, tables, and figures; where possible, quantitative syntheses with meta-analysis were conducted of test accuracy and RCT findings. For meta-analysis of RCTs...
Figure 1. Analytic Framework: Depression and Suicide Risk Screening

Key questions

1. Do depression, anxiety, or suicide risk screening programs in primary care or comparable settings result in improved health outcomes in adults, including pregnant and postpartum persons?
   a. Does sending depression, anxiety, or suicide risk screening test results to providers (with or without additional care management supports) result in improved health outcomes?

2. Do instruments to screen for depression, anxiety, and/or high suicide risk accurately identify adults, including pregnant and postpartum persons, with depression, anxiety, and high suicide risk in primary care or comparable settings?

3. What are the harms associated with screening for depression, anxiety, or suicide risk in primary care or comparable settings in adults, including pregnant and postpartum persons?

4. Does treatment (psychotherapy or pharmacotherapy) of depression, anxiety, or high suicide risk result in improved health outcomes in adults, including pregnant and postpartum persons?

5. What are the harms of treatment (psychotherapy or pharmacotherapy) of depression, anxiety, or high suicide risk in adults, including pregnant and postpartum persons?

Evidence reviews for the US Preventive Services Task Force (USPSTF) use an analytic framework to visually display the key questions that the review will address to allow the USPSTF to evaluate the effectiveness and safety of a preventive service. The questions are depicted by linkages that relate interventions and outcomes. A dashed line indicates a health outcome that immediately follows an intermediate outcome. For more details see the USPSTF Procedure Manual.6

Results

Altogether, 105 publications were included for depression: 32 original studies (N=385,607) and 73 systematic reviews (including ≈2138 studies [N ≈ 9.8 million]); 27 studies (n = 24,826) addressed suicide risk (Figure 2).

Benefits of Screening

Key Question 1. Do depression or suicide risk screening programs in primary care or comparable settings result in improved health outcomes in adults, including pregnant and postpartum persons?

Key Question 1a. Does sending depression or suicide risk screening test results to providers (with or without additional care management supports) result in improved health outcomes?

Seventeen trials (reported in 28 publications) examined depression screening,15-30 including 1 that examined screening for...
Figure 2. Literature Search Flow Diagram: Depression and Suicide Risk Screening

Article summaries are included for KQ1, KQ2, KQ4, and KQ5. A summary for KQ3 is not included because the review incorporated and updated the evidence related to screening for and treatment of depression and suicide risk while adding evidence related to screening for and treatment of anxiety disorders and combination approaches that address more than 1 of these conditions. New primary evidence includes 3 studies for KQ1 evidence, 25 studies for KQ2 evidence, 0 studies for KQ3 evidence, 45 studies for KQ4 evidence, and 9 studies for KQ5 evidence. Existing systematic reviews for large, mature bodies of evidence were also included.

Reasons for Exclusion: Intervention: Study used an excluded intervention/screening approach. Population: Study was not conducted in an average-risk population. Setting: Study was not conducted in a country relevant to US practice. Aim: Study aim not relevant. Outcomes: Study did not have relevant outcomes or had incomplete outcomes. Comparator: Study included a comparator group that was not included. Quality: Study did not meet criteria for fair or good quality. Design: Study did not use an included design. Existing systematic review superseded: Existing systematic review was superseded by one that was more contemporary, comprehensive, or relevant.

Studies may appear in more than 1 key question (KQ).

© 2023 American Medical Association. All rights reserved.
existing systematic reviews covered various versions of the PHQ, 2- and 3-item Whooley screening questions, CES-D, and EPDS. For example, in individual patient data meta-analyses, the PHQ-9 correctly identified 85% of people with major depression and 85% of those without major depression, at the standard cutoff of 10 or greater, when compared with a semistructured interview reference standard (sensitivity, 0.85 [95% CI, 0.79-0.89]; specificity, 0.85 [95% CI, 0.82-0.87]; 47 studies [n = 11 234]) (Figure 3; eTable 7 in the Supplement). At the standard cutoff of 2 or greater and when compared with a semistructured interview, the PHQ-2 was more sensitive than the PHQ-9, correctly identifying 91% of people with major depression (sensitivity, 0.91 [95% CI, 0.88-0.94]; 48 studies [n = 11 703]). But specificity at that cutoff was lower, accurately identifying only 67% of people without depression (specificity, 0.67 [95% CI, 0.64-0.71]; 48 studies [n = 11 703]). The Whooley, CES-D, and EPDS demonstrated accuracy comparable with that of the PHQ-2.

The 14 primary studies all covered multiple versions of the GDS; the GDS-15 was the most common. The standard cutoff of 5 or greater had an acceptable balance of sensitivity and specificity. In the pooled analysis combining 7 studies, the GDS-15 accurately identified 94% of people with major depression (sensitivity, 0.94 [95% CI, 0.85-0.98]; I² = 85.7%) and 81% of those without (specificity, 0.81 [95% CI, 0.70-0.89]; I² = 99.2%) (eFigure 1 and eTable 8 in the Supplement).

Three studies were included that screened for suicidal ideation (eTable 9 in the Supplement).57-59 Most screening instruments reported sensitivity and specificity above 0.80 for at least 1 reported cutoff (eTable 10 in the Supplement). However, there was no replication of any instrument, and 2 of the 3 studies included only 3 individuals57 and 12 individuals59 with suicidal ideation or at very high risk according to the reference standards. The study with the most events was limited to older adults.58

Table 1. Summary of Meta-analysis Results for Depression Outcomes Among Depression Screening Studies (KQ1)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No. studies (No. analyzed)</th>
<th>Pooled result (95% CI)</th>
<th>I², %</th>
<th>τ² Range of effectsb</th>
<th>Median (IQR) effectsb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalence (met criteria for depression or score above cutoff)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td>8 (10 244)</td>
<td>OR, 0.60 (0.50-0.73)</td>
<td>0</td>
<td>0</td>
<td>0.20 to 1.11 ARD, −9.1 to 1.4</td>
</tr>
<tr>
<td>General</td>
<td>1 (218)</td>
<td>OR, 0.67 (0.37-1.21)</td>
<td>NA</td>
<td>NA</td>
<td>0.67 ARD, −9</td>
</tr>
<tr>
<td>Older</td>
<td>1 (206)</td>
<td>OR, 0.70 (0.38-1.26)</td>
<td>NA</td>
<td>NA</td>
<td>0.70 to 0.80 ARD, −8 to −5</td>
</tr>
<tr>
<td>Postpartum</td>
<td>5 (9202)</td>
<td>OR, 0.54 (0.40-0.73)</td>
<td>25.6</td>
<td>.02</td>
<td>0.30 to 1.11 ARD, −9.1 to 1.4</td>
</tr>
<tr>
<td>Pregnant</td>
<td>1 (618)</td>
<td>OR, 0.80 (0.48-1.35)</td>
<td>NA</td>
<td>NA</td>
<td>0.80 ARD, −2</td>
</tr>
<tr>
<td>Remission (did not meet criteria for depression or score below cutoff, among those with symptoms at baseline)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td>8 (2302)</td>
<td>OR, 1.58 (1.23-2.02)</td>
<td>0</td>
<td>0</td>
<td>0.81 to 4.81 ARD, −18 to 33.8</td>
</tr>
<tr>
<td>General</td>
<td>3 (1396)</td>
<td>OR, 1.52 (1.41-1.63)</td>
<td>0</td>
<td>0</td>
<td>0.81 to 4.05 ARD, −5 to 33</td>
</tr>
<tr>
<td>Older</td>
<td>2 (259)</td>
<td>OR, 0.97 (0.21-4.41)</td>
<td>0</td>
<td>0</td>
<td>0.83 to 2.49 ARD, −18 to 5</td>
</tr>
<tr>
<td>Postpartum</td>
<td>2 (562)</td>
<td>OR, 1.83 (0.27-12.27)</td>
<td>0</td>
<td>0</td>
<td>1.67 to 2.34 ARD, 11.7 to 19</td>
</tr>
<tr>
<td>Pregnant</td>
<td>1 (85)</td>
<td>OR, 4.81 (1.81-12.80)</td>
<td>NA</td>
<td>NA</td>
<td>4.81 ARD, 33.8</td>
</tr>
<tr>
<td>Combined reduced depressionc,d</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td>16 (8448)</td>
<td>OR, 1.63 (1.37-1.95)</td>
<td>0.5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>5 (1675)</td>
<td>OR, 1.53 (1.38-1.70)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Older</td>
<td>4 (675)</td>
<td>OR, 1.00 (0.56-1.78)</td>
<td>15.2</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Postpartum</td>
<td>6 (6013)</td>
<td>OR, 1.98 (1.60-2.43)</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Pregnant</td>
<td>1 (85)</td>
<td>OR, 4.81 (1.81-12.80)</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Symptom severity (change in depression symptom scores)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All studies</td>
<td>9 (5543)</td>
<td>Mean difference in change, −1.0 (−2.3 to 0.3)</td>
<td>74.4</td>
<td>1.1</td>
<td>−8.2 to 2.6</td>
</tr>
<tr>
<td>All studies</td>
<td>6 (3790)</td>
<td>SMD, −.09 (−0.36 to 0.18)</td>
<td>79.6</td>
<td>0.04</td>
<td>NR</td>
</tr>
</tbody>
</table>

Abbreviations: ARD, absolute risk difference; NA, not applicable; NR, not reported; OR, odds ratio; SMD, standardized mean difference.

a Effect based on restricted maximum likelihood model with the Knapp-Hartung adjustment for small samples.
b Range of effects for all study groups, subgroup analyses, and time points, ie, not limited to records in the meta-analysis.
c As available, selected depression remission or scoring below a cutoff first, depression prevalence or scoring above a cutoff (reversed) second, and depression response third.
d Range of effects and median (IQR) effects not shown because reversal of results from some studies creates misleading ARD values.
Figure 3. Test Accuracy of the PHQ, CES-D, Whooley, and EPDS From Published Systematic Evidence Reviews (Key Question 2)

<table>
<thead>
<tr>
<th>Source</th>
<th>Reference standard</th>
<th>No. of studies</th>
<th>No. of participants</th>
<th>Cutoff</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHQ-9 Linear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negeri et al,55 2021</td>
<td>Fully structured diagnostic interview</td>
<td>20</td>
<td>17167</td>
<td>≥10</td>
<td>0.64 (0.53-0.74)</td>
<td>0.88 (0.83-0.92)</td>
</tr>
<tr>
<td>MINI</td>
<td></td>
<td>33</td>
<td>16102</td>
<td>≥10</td>
<td>0.74 (0.67-0.79)</td>
<td>0.89 (0.86-0.91)</td>
</tr>
<tr>
<td>Semistructured diagnostic interview</td>
<td></td>
<td>47</td>
<td>11234</td>
<td>≥10</td>
<td>0.85 (0.79-0.89)</td>
<td>0.85 (0.82-0.87)</td>
</tr>
<tr>
<td>Wang et al,52 2021 (prenatal)</td>
<td>Fully/semistructured diagnostic interview</td>
<td>3</td>
<td>1465</td>
<td>≥10</td>
<td>0.82 (0.74-0.91)</td>
<td>0.77 (0.68-0.87)</td>
</tr>
<tr>
<td>PHQ-9 Algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>He et al,48 2020</td>
<td>Fully structured diagnostic interview</td>
<td>13</td>
<td>7577</td>
<td>Original</td>
<td>0.35 (0.26-0.46)</td>
<td>0.95 (0.93-0.97)</td>
</tr>
<tr>
<td>MINI</td>
<td></td>
<td>14</td>
<td>2780</td>
<td>Original</td>
<td>0.51 (0.49-0.53)</td>
<td>0.97 (0.96-0.98)</td>
</tr>
<tr>
<td>Semistructured diagnostic interview</td>
<td></td>
<td>27</td>
<td>6331</td>
<td>Original</td>
<td>0.57 (0.49-0.64)</td>
<td>0.95 (0.94-0.97)</td>
</tr>
<tr>
<td>PHQ-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al,53 2020</td>
<td>Fully structured diagnostic interview</td>
<td>13</td>
<td>7596</td>
<td>≥10</td>
<td>0.63 (0.52-0.72)</td>
<td>0.86 (0.81-0.90)</td>
</tr>
<tr>
<td>MINI</td>
<td></td>
<td>14</td>
<td>2784</td>
<td>≥10</td>
<td>0.72 (0.63-0.79)</td>
<td>0.88 (0.84-0.91)</td>
</tr>
<tr>
<td>Semistructured diagnostic interview</td>
<td></td>
<td>27</td>
<td>6362</td>
<td>≥10</td>
<td>0.88 (0.80-0.90)</td>
<td>0.86 (0.83-0.89)</td>
</tr>
<tr>
<td>PHQ-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harel et al,54 2022</td>
<td>Fully structured diagnostic interview</td>
<td>15</td>
<td>12109</td>
<td>≥4</td>
<td>0.68 (0.56-0.78)</td>
<td>0.85 (0.78-0.90)</td>
</tr>
<tr>
<td>MINI</td>
<td></td>
<td>31</td>
<td>14870</td>
<td>≥4</td>
<td>0.80 (0.73-0.85)</td>
<td>0.83 (0.80-0.86)</td>
</tr>
<tr>
<td>Semistructured diagnostic interview</td>
<td></td>
<td>29</td>
<td>7719</td>
<td>≥4</td>
<td>0.88 (0.81-0.93)</td>
<td>0.79 (0.74-0.83)</td>
</tr>
<tr>
<td>PHQ-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levis et al,50 2020</td>
<td>Fully structured diagnostic interview</td>
<td>20</td>
<td>17319</td>
<td>≥2</td>
<td>0.82 (0.75-0.87)</td>
<td>0.71 (0.63-0.77)</td>
</tr>
<tr>
<td>MINI</td>
<td></td>
<td>32</td>
<td>15296</td>
<td>≥2</td>
<td>0.89 (0.84-0.92)</td>
<td>0.68 (0.64-0.73)</td>
</tr>
<tr>
<td>Semistructured diagnostic interview</td>
<td></td>
<td>48</td>
<td>11703</td>
<td>≥2</td>
<td>0.91 (0.88-0.94)</td>
<td>0.67 (0.64-0.71)</td>
</tr>
<tr>
<td>PHQ-2 + PHQ-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levis et al,50 2020</td>
<td>Semistructured diagnostic interview</td>
<td>44</td>
<td>10627</td>
<td>≥2 (PHQ-2), ≥10 (PHQ-9)</td>
<td>0.82 (0.76-0.86)</td>
<td>0.87 (0.84-0.89)</td>
</tr>
<tr>
<td>CES-D</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vilagut et al,51 2016</td>
<td>Standardized diagnostic interview</td>
<td>22</td>
<td>NR</td>
<td>≥16</td>
<td>0.87 (0.82-0.91)</td>
<td>0.70 (0.65-0.75)</td>
</tr>
<tr>
<td>Whooley</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bosanquet et al,47 2015</td>
<td>Diagnostic interview</td>
<td>10</td>
<td>4618</td>
<td>NA</td>
<td>0.95 (0.88-0.97)</td>
<td>0.65 (0.56-0.74)</td>
</tr>
<tr>
<td>Smith et al,56 2022 (prenatal)</td>
<td>Diagnostic interview</td>
<td>5</td>
<td>1402</td>
<td>NA</td>
<td>0.95 (0.81-0.99)</td>
<td>0.60 (0.44-0.74)</td>
</tr>
<tr>
<td>EPDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levis et al,46 2020</td>
<td>Fully structured diagnostic interview</td>
<td>4</td>
<td>3188</td>
<td>≥11</td>
<td>0.90 (0.58-0.98)</td>
<td>0.83 (0.62-0.94)</td>
</tr>
<tr>
<td>MINI</td>
<td></td>
<td>18</td>
<td>3302</td>
<td>≥11</td>
<td>0.82 (0.71-0.89)</td>
<td>0.84 (0.79-0.89)</td>
</tr>
<tr>
<td>Semistructured diagnostic interview</td>
<td></td>
<td>36</td>
<td>9066</td>
<td>≥11</td>
<td>0.81 (0.75-0.87)</td>
<td>0.88 (0.85-0.91)</td>
</tr>
</tbody>
</table>

CES-D indicates Center for Epidemiologic Studies Depression scale; EPDS, Edinburgh Postnatal Depression Scale; MINI, Mini International Neuropsychiatric Interview; NA, not applicable; NR, not reported; PHQ, Patient Health Questionnaire.
Harms of Screening

**Key Question 3.** What are the harms associated with screening for depression or suicide risk in primary care or comparable settings in adults, including pregnant and postpartum persons?

Only 1 depression screening study reported on harms. That study, conducted in Hong Kong among postpartum patients, reported that there were no adverse events in either group. Across all depression screening studies included for KQ1, there was no pattern of effects indicating that screening might paradoxically worsen any outcomes the interventions were aiming to benefit.

For harms of suicide screening, the same short-term study (n = 443) that was included for KQ1 was the only evidence included for assessing the harms of suicide risk screening (ie, a possible increase in suicidal ideation) and indicated no differences between groups (eTable 4 in the Supplement).

**Benefits of Treatment**

**Key Question 4.** Does treatment of depression or high suicide risk result in improved health outcomes in adults, including pregnant and postpartum persons?

Thirty-nine existing systematic reviews (reported in 41 publications) of treatment for depression were included; 30 addressed psychological treatment (eTable 11 in the Supplement), and 10 addressed pharmacologic treatment (eTable 12 in the Supplement). Psychological treatment improved depression symptom severity (Figure 2 in the Supplement). This was the case both in broad analyses that included a wide range of populations and specific interventions and in analyses of some important specific populations, including older adults, perinatal populations, and primary care patients. For example, the broadest analysis, which included any type of psychological treatment compared with any kind of control condition, measuring the depression outcome immediately after treatment (typically 2 to 6 months after baseline), had a standardized mean difference (SMD) of –0.72 (95% CI, –0.78 to –0.67; 385 studies [N not reported but estimated at ≈3 000]), suggesting a moderate to large effect size. When limited to studies in primary care patients, the effect was smaller but statistically significant (SMD, –0.42 [95% CI, –0.56 to –0.29]; 59 studies [N not reported]). Remission and response to treatment (ie, a prespecified level of symptom reduction, such as a certain number of points or a percentage decline relative to baseline score) were more sparsely reported.

Data were limited for populations who were socially or economically disadvantaged or in specific racial or ethnic groups; however, the limited evidence supported benefits of psychological treatment in these populations as well. For example, an analysis of 5 trials among people described as having low socioeconomic status found reduced depressive symptoms at up to 12 weeks after baseline (SMD, –0.66 [95% CI, –0.92 to –0.41]; 5 studies [n = 424]), and a separate analysis found no differences in effect size between studies limited to race or ethnic “minority” populations vs not limited to these populations.

For antidepressant medications, pooled effects consistently demonstrated small but statistically significant reductions in depressive symptom severity (eFigure 3 in the Supplement) as well as increased rates of remission (eFigure 4 in the Supplement) and response to treatment (eFigure 5 in the Supplement) in the short term (typically 8 weeks). For example, fluoxetine, which had the largest body of evidence with 117 studies, was associated with a small reduction in symptom severity (SMD, –0.23 [95% CI, –0.28 to –0.19]), a 46% increase in the odds of remission (OR, 1.46 [95% CI, 1.34-1.60]) and a 52% increase in the odds of treatment response (OR, 1.52 [95% CI, 1.40-1.66]; number of studies, number of individuals, and P² not reported). Little information was available on the long-term impact of antidepressants in the synthesized literature, and information was absent or extremely limited on the benefits of pharmacologic treatment among socially or economic disadvantaged or specific racial or ethnic groups. The existing systematic reviews did not report on baseline symptom levels, making it impossible to determine whether the samples in the pharmacotherapy trials were comparable to those in the trials of psychological treatment.

Twenty-three RCTs (reported in 36 articles [n = 22 632]) of suicide prevention among people at increased risk of suicide were included (eTable 13 in the Supplement). One study examined the impact of a pharmacologic intervention (lithium), and the remaining studies examined psychological treatment interventions, along with usual mental health care. The impact of the interventions on suicide deaths could not be determined, with only 1 death reported in the included studies. One large (n = 18 882) good-quality multisite trial conducted in US integrated care settings tested 2 suicide prevention interventions among adults with an elevated risk for suicide based on item 9 of the PHQ-9. That study found that, compared with usual care, a care management intervention had no impact on the rate of suicide attempts (hazard ratio, 1.07 [95% CI, 0.84 to 1.37]; P = .52), and a low-intensity online skills training intervention was associated with an increased rate of suicide attempts (hazard ratio, 1.29 [95% CI, 1.02 to 1.64]; P = .02). Most other studies reported 5 or fewer suicide attempts per study group, and the pooled effect was not statistically significant (OR, 0.94 [95% CI, 0.73-1.22]; 12 RCTs [n = 14 573]; I² = 11.2%, including only the care management group of the large trial [Figure 4]). Usual mental health care was the most common control group and was in some cases enhanced or optimized, so most of the included studies could be considered comparative effectiveness studies.

Harms of Treatment

**Key Question 5.** What are the harms of treatment of depression or high suicide risk (psychotherapy or pharmacotherapy) in adults, including pregnant and postpartum persons?

Four existing systematic reviews addressing harms of psychological interventions (eTable 14 in the Supplement) were included. Psychological interventions did not increase the risk of harm, as measured by deterioration of depressive symptoms (eTable 15 in the Supplement).

For pharmacologic treatment, 1 cohort study and 22 existing systematic reviews (eTable 16 in the Supplement) were included. There was clear evidence that persons receiving antidepressants were at a higher risk of dropout because of adverse events (eFigure 6 in the Supplement). There was also evidence of an increased risk of serious adverse events with use of selective serotonin reuptake inhibitors (OR, 1.39 [95% CI, 1.12-1.72]; 44 RCTs [n = 13 198]; I² = 0%) (eFigure 7 in the Supplement). The absolute risk of serious adverse events appears to be relatively low, however, and evidence for specific serious adverse events was very limited. There were too few suicide deaths to determine the association between antidepressant use and suicide death, but both RCT and...
observational evidence supported a small absolute increase in risk of suicide attempts with second-generation antidepressant use among adults up to age 65 years (Figure 8 in the Supplement). For example, a review of US Food and Drug Administration regulatory data indicated a 53% increase in the odds of a suicide attempt at post-treatment evaluation with the use of second-generation antidepressants (OR, 1.53 [95% CI, 1.09-2.15]; n = 41,861); 0.7% of antidepressants seekers had a suicide attempt vs 0.3% of placebo users.160 Evidence on other outcomes was limited and generally included only observational evidence, including harms of pharmacotherapy in pregnant or postpartum people.

Two of the included RCTs of suicide prevention treatment reported on harms.106,132 There were no differences between groups at follow-up on an instrument designed to assess the perceived level of coercion experienced by service users during hospital admission.106,132 There was no pattern of effect in the studies included for KQ4 to indicate paradoxic harms of treatment. The study of lithium found a higher rate of nonserious adverse events (75.7% with lithium, 69.0% with placebo; P value not reported), and a slightly higher rate of serious adverse events (38.8% with lithium, 34.1% with placebo; P value not reported) but no difference in withdrawals due to adverse events (1.2% with lithium, 1.5% with placebo; P value not reported).132

**Discussion**

Direct evidence indicated that screening programs improved depression outcomes. In addition, robust indirect evidence exists that screening tools feasible to administer in primary care settings have reasonable accuracy and that treatment is effective (Table 2). The direct evidence is more equivocal than the indirect evidence, being based on a smaller number of studies and having fewer statistically significant findings. The presence of additional program components beyond screening in many of the depression screening studies made it difficult to isolate the specific effects of screening alone in these studies.

Both the direct and indirect evidence on screening for suicide risk was extremely limited, and the indirect evidence indicated that implementation of some interventions that are feasible for widespread use in health care settings may either have no impact on suicide attempts or paradoxically increase the risk of a suicide attempt. However, the treatment evidence was predominantly compared with usual specialty mental health care, making it difficult to understand the absolute treatment effects. Unlike the previous review, the current review did not include treatment studies in persons seeking treatment in urgent or emergency settings, due to their low applicability to screening in primary care settings; however, the conclusions of the current review are consistent with those from the previous review.

**Screening for Depression**

The direct evidence for the benefits of screening for depression was very similar to that in the previous review, with only 2 new studies added.28,31 Trials in general primary care populations and in perinatal populations in particular demonstrate increased rates of depression remission or falling below a specified symptom severity level after 6 to 12 months. The evidence in older adult populations was more limited and did not show a clear benefit but also had some important limitations. Chiefly, only 4 studies examined screening in older adults, and only 1 used a depression measure specifically designed for older adults.2,3 This is important because some somatic and sleep-related symptoms of depression are common in older adults without depression. There is ample evidence that screening instruments can identify people with major depressive disorder with reasonable accuracy, and cutoffs could be optimized for specific local settings and populations. Further, a large cohort study showed that disparities in screening rates between Black and other English-speaking primary care patients were eliminated after implementation of a routine screening.
Table 2. Summary of Evidence: Depression and Suicide Risk Screening

<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of studies (No. randomized)</th>
<th>Summary of findings</th>
<th>Consistency and precision</th>
<th>Other limitations</th>
<th>Strength of evidence</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ1: Benefits of screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>14 RCTs, 3 CCTs (n = 18,437)</td>
<td>Evidence supported the benefits of screening for depression; eg, at 6 mo postbaseline or 6 mo postpartum (or the closest follow-up time point to 6 mo): Prevalence of depression or clinically important symptomatology: OR, 0.60 (95% CI, 0.50 to 0.73); 8 studies (n = 10,244); I² = 0%. Remission or falling below a specific level of depression symptomatology: OR, 1.58 (95% CI, 1.23 to 2.02); 8 studies (n = 2,302); I² = 0%. However, no clear benefit in symptom severity measures was found (pooled mean difference in change, −1.0 [95% CI, −2.3 to 0.3]); 9 studies (n = 5,543); I² = 74.4%</td>
<td>Reasonably consistent, reasonably precise</td>
<td>Few studies with unscreened control groups and limited capacity for conducting such studies as screening for depression becomes the standard of care; heterogeneity in interventions and limited evidence on screening without further practice supports</td>
<td>Moderate for benefit</td>
<td>Most studies either conducted outside the US or, among US-based studies, published &gt;15 years ago Applicability to current US health care systems unclear</td>
</tr>
<tr>
<td>Suicide risk</td>
<td>1 RCT (n = 443)</td>
<td>Among primary care patients who screened positive for depression, there was 1 suicide attempt after 2 weeks; there were no group differences on any of 3 items measuring suicidal ideation</td>
<td>Consistency NA, imprecise</td>
<td>Single study, very short-term follow-up, limited to people who screened positive for depression</td>
<td>Insufficient</td>
<td>Conducted in the UK and limited to people with symptoms of depression</td>
</tr>
<tr>
<td>KQ2: Accuracy of screening tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>10 Existing systematic reviews (~196 studies [n ~ 75,000]) 14 Test accuracy studies (n = 8,819)</td>
<td>Adequate sensitivity and specificity for the PHQ Linear, PHQ-8, PHQ-2, Whooley questions, CES-D, EPDS, and GDS</td>
<td>Consistent, precise</td>
<td>Most of the existing systematic reviews were not restricted to primary care populations</td>
<td>High</td>
<td>Most of the studies were not conducted in the US</td>
</tr>
<tr>
<td>Suicide risk</td>
<td>3 Test accuracy studies (n = 17,511)</td>
<td>GDS-15, GDS-Suicide Ideation, and the SDDS-PC had adequate test accuracy to detect suicidal ideation</td>
<td>Consistency NA, precision NA</td>
<td>Not replicated in more than 1 study</td>
<td>Insufficient</td>
<td>All studies conducted in the US, 2 in primary care</td>
</tr>
<tr>
<td>KQ3: Harms of screening</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>Directly assessed harms: 1 (n = 642) Indirectly used to infer harms: 14 RCTs, 3 CCTs (n = 18,437)</td>
<td>One study reported no adverse events in either group Studies included for KQ1 did not show a pattern of results indicating harmful impact</td>
<td>Consistent, imprecise</td>
<td>Adverse events rarely directly assessed</td>
<td>Moderate for little to no harm</td>
<td>Most studies either conducted outside the US or, among US-based studies, published &gt;15 years ago Applicability to current US health care systems unclear</td>
</tr>
<tr>
<td>Suicide risk</td>
<td>1 RCT (n = 443)</td>
<td>Two of 3 suicidal ideation items indicated a possible higher risk with screening; however, the findings were inconclusive due to lack of statistical significance and very wide confidence intervals</td>
<td>Consistency NA, imprecise</td>
<td>Single study, very short-term follow-up, limited to people who screened positive for depression</td>
<td>Insufficient</td>
<td>Conducted in the UK and limited to people with symptoms of depression</td>
</tr>
</tbody>
</table>
### Table 2. Summary of Evidence: Depression and Suicide Risk Screening (continued)

<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of studies (No. randomized)</th>
<th>Summary of findings</th>
<th>Consistency and precision</th>
<th>Other limitations</th>
<th>Strength of evidence</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depression</td>
<td>30 Existing systematic reviews of psychological treatment (≈346 RCTs [n = 45 078]) 10 Existing systematic reviews of pharmacologic treatment (≈522 studies [n = 116 477])</td>
<td>Psychological treatment improved depression and other health outcomes such as anxiety symptoms, hopelessness, quality of life, and functioning  The broadest analysis indicated a moderate to large effect on depression (SMD, −0.72 [95% CI, −0.78 to −0.67]; 385 studies [N not reported but estimated at ≈33 000])  The effect was smaller when limited to studies in primary care patients but was clearly statistically significant (SMD, −0.42 [95% CI, −0.56 to −0.29]; 59 studies [N not reported])  Antidepressant medications consistently demonstrated increased rates of remission and response to treatment and small but statistically significant reductions in depressive symptom severity; eg, fluoxetine, which had the largest body of evidence with 117 studies, was associated with a small reduction in symptom severity (SMD, −0.23 [95% CI, −0.28 to −0.19]), a 46% increase in the odds of remission (OR, 1.46 [95% CI, 1.34 to 1.60]), and a 52% increase in the odds of treatment response (OR, 1.52 [95% CI, 1.40 to 1.66]).</td>
<td>Consistent, precise</td>
<td>Most existing systematic reviews examined posttreatment outcomes with little information on longer-term follow-up  There was evidence of publication or reporting bias; however, effects were still statistically significant after adjusting for these biases  Evidence for benefit in a priori populations of interest was limited in the synthesized literature, particularly on the effect of antidepressant medications</td>
<td>High for benefit</td>
<td>Studies recruited from a wide range of community, online, and clinic sources; wide range of countries; effect sizes in subgroup analyses limited to primary care settings tended to be smaller than in broad-based analyses</td>
</tr>
<tr>
<td>Suicide risk</td>
<td>23 RCTs (n = 22 632)</td>
<td>A very large (n = 18 882) study conducted in 4 US health care systems found that 2 separate suicide prevention interventions were associated with either no impact on suicide attempts (HR, 1.07 [97.5% CI, 0.84 to 1.37]) for a care management intervention or an increased risk of suicide attempts (HR, 1.29 [97.5% CI, 1.02 to 1.64]) for a low-intensity online intervention)  Most other studies had very few participants with suicide attempts, and the pooled effect was not statistically significant (OR, 0.94 [95% CI, 0.73 to 1.22]).</td>
<td>Inconsistent, imprecise</td>
<td>Control groups were typically usual specialty mental health care (enhanced or optimized in some cases) and so may be considered comparative effectiveness studies; some trials had primary aims of broad self-harm reduction (ie, not focused on self-harm with suicidal intent)  Suicide death: insufficient  Suicide attempts: moderate that some interventions are associated with no benefit or increased risk of harm compared with usual mental health care  Suicidal ideation, depression, other mental health: low for small to no benefit compared with usual specialty mental health care</td>
<td>Moderate for harm</td>
<td>Studies recruited from a wide range of community, online, and clinic sources; wide range of countries; effect sizes in subgroup analyses limited to primary care settings tended to be smaller than in broad-based analyses</td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of studies (No. randomized)</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>KQ5: Harms of treatment</td>
<td></td>
<td>In 3 existing systematic reviews, deterioration rates were either lower with psychological interventions or did not differ statistically from control groups. A separate review among older adults reported that none of the 14 included trials reported safety data. Pharmacologic treatment was associated with a higher risk of dropout due to adverse events with all agents examined and a higher risk of serious adverse events with SSRI use (OR, 1.39 [95% CI 1.12 to 1.72]; 44 RCTs [N not reported]; I² = 0%). There were too few suicide deaths to determine the association between antidepressant use and suicide death, but both RCT and observational evidence supported an increased risk of suicide attempts with second-generation antidepressant use among adults up to age 65 y (OR, 1.53 [95% CI, 1.09 to 2.15]; N = 40,857; 0.7% of antidepressants users vs 0.3% of placebo users). Other outcomes were largely limited to observational evidence.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Condition</th>
<th>No. of studies (No. randomized)</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suicide risk</td>
<td>Directly assessed harms: 2 RCTs (n = 607) Indirectly used to infer harms: 15 RCTs (n = 1994)</td>
<td>Two studies reported on harms. There were no differences between groups at follow-up on an instrument designed to assess the perceived level of coercion experienced by service users during hospital admission. The study of lithium found a higher rate of nonserious adverse events (75.7% with lithium, 69% with placebo; P value not reported) and a slightly higher rates of serious adverse events (38.8% with lithium, 34.1% with placebo; P value not reported) but no difference in withdrawals due to adverse events (1.2% with lithium, 1.5% with placebo; P value not reported). There was no pattern of effect in the studies included for KQ4 to indicate paradoxical harms of treatment.</td>
</tr>
</tbody>
</table>
program in a large health system,\textsuperscript{167} suggesting that routine depression screening may promote equitable mental health outcomes across racial and ethnic groups.

In contrast to the findings of the current review, the Canadian Task Force on Preventive Health Care (CTFPHC) does not recommend screening for depression, based on the lack of direct evidence on the benefits and harms of routinely screening asymptomatic adults.\textsuperscript{162} The review of the CTFPHC recommendation was based on only included studies in which the screening intervention was a normal part of care and that had an unscreened comparison group. Thus, only 3 of the 17 screening studies included in the current review could have met the Canadian task force’s inclusion criteria and were published at the time the CTFPHC review was conducted, and only 1 of these was explicitly listed as examined\textsuperscript{17} and excluded from the review. The study excluded from the CTFPHC review did not integrate screening into the normal care process; instead, screening was undertaken by study staff. In addition, the CTFPHC stated that it “had concerns about the potential harms of screening (e.g. false positive, unnecessary treatment, labelling and stigma) and appropriate use of limited resources.”

**Screening for Suicide Risk**

Suicide prevention efforts have the potential to save many US lives, and according to the Centers for Disease Control and Prevention, success in preventing suicide is most likely if addressed at multiple levels and in multiple sectors.\textsuperscript{163} While there is likely an important role for health care settings, only 1 trial reporting direct evidence on suicide risk screening among primary care patients was found, and it was limited to patients who had screened positive for depression.\textsuperscript{32} The findings were inconclusive. This review was scoped to include evidence on screening in broad populations (not only those who screen positive for depression), but no such evidence was found. In addition, there was minimal evidence on the test performance of suicide risk screening instruments; no instrument was addressed in more than 1 study. However, since many depression screening instruments include a suicide risk question, suicide risk screening often occurs in the context of real-world depression screening.

Studies without control groups (and therefore excluded from the current review) have indicated that asking adults about suicidality in mental health settings does not increase suicidality.\textsuperscript{160,164,165} Similarly, a randomized trial among adults with borderline personality disorder comparing frequent and repeated mental health assessment (5 times per day initially, then daily, then weekly), with or without items assessing suicidal ideation, found no increase in suicidal thoughts or behaviors with suicide-related screening compared with mental health screening without suicide-related items.\textsuperscript{166} Some health care systems have implemented suicide risk screening in primary care settings, without reports of harms. These include the US Department of Veterans Affairs system, which recommends using the PHQ-9 (which includes a suicide-related item),\textsuperscript{167} and the Chickasaw Nation Departments of Health and Family Services,\textsuperscript{168} which recommends administering the full PHQ-9 to people who screen positive on the PHQ-2.

Qualitative patient interviews among people who screened positive for depression in primary care settings indicated that being asked about suicidal thoughts felt appropriate and valuable, given the context of the positive depression screen result.\textsuperscript{169} One theme that emerged, however, was difficulty answering the PHQ-9 item about suicide (“Thoughts that you would be better off dead, or thoughts of hurting yourself in some way”), since some felt that while they thought about suicide or wishing they were dead, they felt strongly that they would never attempt suicide. Other themes included weighing the hope for help against fears of negative consequences (eg, loss of autonomy or feelings of shame), the importance of a trusting relationship with the clinician, and the value of the clinician’s willingness to listen without judgement.

Other potential harms of screening for suicide risk in primary care settings have been observed or postulated. For example, there are documented cases in which insurers have denied medical coverage for health care associated with suicide attempts,\textsuperscript{170,171} and having a positive suicide risk screen result may increase the risk that some types of injuries could be interpreted as suicide attempts. Similarly, life insurance payouts could potentially be affected by findings of increased suicide risk in medical records, since most policies do not pay out for suicide deaths in the first 2 years of coverage.\textsuperscript{172,173} Thus, a screening result in the medical record indicating an elevated risk of suicide could result in serious financial implications for people who struggle with mental health issues and their families.

**Mental Health Equity Across Racial and Ethnic Groups**

We found minimal information on the effects of mental health screening in traditionally underserved patient groups, including Black, Latino, and Native American adults. Evaluating the effectiveness of depression and suicide risk screening in these groups is particularly important because of the high burden of depression in these communities and because they may be at elevated risk of misdiagnosis or barriers to treatment. Racism and discriminatory policies in the US have adversely affected the mental and economic well-being of these communities.\textsuperscript{174-181} The health care system has contributed to these inequities through bias in diagnosis, even if inadvertent, and by tolerating differential barriers to receiving appropriate treatment. For example, compared with White patients, misdiagnosis of mental health conditions appears to be more common in Black and Hispanic/Latino patients,\textsuperscript{182-184} who are also less likely to receive mental health services than Asian American or White patients.\textsuperscript{185,186} The cost of treatment and lack of insurance are among the main barriers to receiving mental health services,\textsuperscript{187} and in the US these barriers tend to have a greater impact on Black persons and other racial and ethnic groups than on White persons, since racism and structural policies in the US have contributed to large inequities in wealth.\textsuperscript{178}

**Limitations**

This review had several limitations. First, it excluded studies in narrow populations with findings that were not widely applicable to screening in primary care settings. For example, we did not include studies limited to persons with physical or developmental disabilities or to people with medical or other mental health comorbidities such as heart disease, cancer, substance use disorders, bipolar disorder, or posttraumatic stress disorder. Second, similarly, the screening instruments selected for review may not apply to some important groups of patients, such as those with low literacy, low health literacy, limited verbal language, or patients who do not speak English.
Third, there remains uncertainty about the benefits of depression screening in older adults, and studies are needed that report outcomes using instruments specifically designed for older adults, and both short-term (<6 months) and long-term (>2 years) outcomes.

Fourth, there are also limitations to understanding of the direct impact of screening relative to other depression management supports. As depression screening becomes the standard of care, this is increasingly difficult to study. Nevertheless, rigorous examinations of implementation programs are needed that report the percentage of patients being screened, referred, and treated. These examinations should also report patient health outcomes, such as depression symptoms and quality of life, prior to program implementation and in control clinics.

Fifth, in addition, more research is needed to understand the impact of depression screening and the most appropriate screening tools to use among Asian American, Black, Hispanic/Latino, and Native American/Alaska Native communities. Native American/Alaska Native communities were not represented in the included studies, despite disproportionately high depression prevalence. Similarly, more information is needed on screening in other underrepresented groups such as gender-nonconforming, immigrant, and non-English-speaking patients. Research is also needed on whether implicit bias among primary care clinicians is associated with lower likelihood of screening some patients or the likelihood of appropriate diagnosis and treatment.

The evidence base to support broad suicide screening in primary care settings is extremely limited. Foundational research is needed in primary care populations, including determining which tools should be used and how screening should be implemented. For example, research needs to examine what training is needed and for whom, what system-level supports are needed, and how to minimize the risk of harms, such as feeling judged or stigmatized, feeling that a cry for help was ignored, or experiencing unnecessary loss of autonomy. The National Institute of Mental Health has called for research examining the use of the Zero Suicide approach, which could also help support a USPSTF screening recommendation if conducted in primary care settings.

Conclusions

Evidence supported depression screening in primary care settings, including during pregnancy and postpartum. There are numerous important gaps in the evidence for suicide risk screening in primary care settings.


49. Levis B, Negeri Z, Sun Y, Benedetti A, Thombs BD. DEPRESSION Screening Data
Clinical Review & Education

US Preventive Services Task Force

USPSTF Review: Depression and Suicide Risk Screening


Depression screening to detect major depression among pregnant and postpartum women: systematic review and meta-analysis of individual participant data. BMJ. 2020;371:m4022. doi:10.1136/bmj.m4022


Effect of offering care management or online healthcare setting to prevent suicide in an integrated suicidal behavior. A randomized controlled trial. JAMA. 2022;327(9):630-638. doi:10.1001/jama.2022.0423


159. Gumsugol SB, Schickling BM, Vignato JA, Santillan DA, Santillan MK. Selective serotonin