Screening for Intimate Partner Violence: An Evaluation of the USPSTF Systematic Evidence Review and Recommendation

Mary Lang-Furr

A Master's thesis submitted to the faculty of the University of North Carolina at Chapel Hill
In partial fulfillment of the requirements for the degree of Master of Public Health in the Public Health Leadership Program.

Chapel Hill
2005

Russell Harris
Samantha Meltzer-Brody, MD

11-4-05 Date
Abstract

Background: The United States Preventive Services Task Force (USPSTF) issued an “I” recommendation in 2004 for screening for intimate partner violence (IPV); it found insufficient evidence to recommend in favor of or against screening. This recommendation has been criticized. It is inconsistent with the recommendations of the American Medical Association and the American College of Obstetrics and Gynecology, both of which recommend that clinicians routinely inquire directly about a history of abuse. The purpose of this paper is to evaluate whether the “I” recommendation is consistent with the evidence that exists for IPV screening and whether the criteria used by the USPSTF to evaluate IPV screening were comparable to the criteria used for other health topics.

Methods: I used one set of criteria to evaluate the USPSTF’s systematic evidence review (SER) on screening for IPV and to compare it to the SER’s on screening for depression and alcohol misuse. Another set of criteria was used to evaluate the USPSTF recommendation for IPV screening and to compare it to other USPSTF recommendations.

Results: There were important differences in the SER’s that may have resulted in adverse effects being given more weight in determining the benefit-harm ratio in the IPV SER than they were given in the other two SER’s. However, there was significantly less evidence available on IPV screening than for depression or alcohol misuse screening.
Conclusion: Evidence for screening for IPV may not have been assessed in a comparable way to that for depression and alcohol misuse, but due to a shortage of evidence on IPV screening, the “I” recommendation would likely stand if disparities in the synthesis of the evidence were remedied.

Background

USPSTF Recommendations for Screening for Intimate Partner Violence

The United States Preventive Services Task Force (USPSTF) systematic evidence review (SER) for screening for intimate partner violence (IPV) “identified no studies that directly addressed the effectiveness of screening in a health care setting in reducing harm from family and intimate partner violence or the adverse effects of screening and interventions.”

Screening instruments were validated, but none were evaluated on violence or health outcomes. Authors of the SER stated a need for studies of the effectiveness of treatment programs for survivors of violence and perpetrators to demonstrate that identification and intervention does or does not lead to improved health outcomes, including reduced violence, improved quality of life, mental health, social support, self-esteem, and productivity.

The USPSTF gave screening for IPV an “I” recommendation, which means they found insufficient evidence to recommend for or against the routine screening of women for intimate partner violence by clinicians. Because of gaps in the
evidence, the USPSTF could not determine the balance between the benefits and harms of screening for intimate partner violence.

One Critique of the USPSTF’s Recommendation for Screening for IPV

Dr. Linda Chamberlain is an epidemiologist specializing in the health effects of domestic violence who works as a consultant for the Family Violence Prevention Fund (FVPF). She wrote an article for FVPF that detailed several criticisms of the USPSTF guidelines for IPV. She felt IPV should be evaluated as its own topic, rather than being evaluated in the same review with child and elder abuse. She also said there should be a separate recommendation on screening pregnant women, like USPSTF did for screening for alcohol and tobacco use. Instead, the authors dismissed the limited evidence that exists for screening and intervention for IPV during pregnancy, rather than lending weight to findings in this high-risk group.

Chamberlain asserted that evaluating IPV as a screening practice forces the topic into a medical model that requires a greater burden of evidence than if it had been evaluated as a behavioral assessment and counseling practice, which would be a more appropriate framework for the evaluation of ways to address IPV. The authors excluded trauma studies in order to emphasize screening asymptomatic patients, but the USPSTF’s previous recommendations emphasized injury-related health risks of IPV in the primary care setting.
Another problem Chamberlain identified was that authors speculated on potential adverse effects of screening based on studies that were not germane to screening for IPV. In fact these studies did not meet the inclusion criteria for the SER. She asserted that emphasizing theoretical adverse effects of screening without considering potential benefits resulted in misleading analysis. The USPSTF did not acknowledge the adverse effects of doing nothing to address IPV in the primary care setting. The adverse effects of not asking women about IPV should be given as much consideration as the adverse effects of screening. Finally, Chamberlain stated that other screening services that have been given a favorable recommendation by the USPSTF have gaps in the evidence similar to those that exist for IPV screening.

**USPSTF Recommendations for Screening for Depression and Alcohol Misuse**

The USPSTF systematic evidence review for screening for depression found that the “rate of detection and diagnosis of depression, based mainly on chart reviews or the completion of a study-specific form” increased from 37% to 47% in six studies that examined the effect of providing feedback of depression screening results to primary care providers. Among studies with interpretable clinical outcomes, screening was associated with 9% absolute reduction in proportion of patients with persistent depression. The USPSTF gave screening for depression a “B” recommendation, which means that they recommend that clinicians routinely screen for depression because there is at least fair evidence that the benefits of screening outweigh the harms of screening. Specifically, the USPSTF
recommends screening adults for depression in clinical practices that have systems in place to assure accurate diagnosis, effective treatment, and follow up.

The evidence review for screening for alcohol misuse found that for good-quality brief, multi-contact behavioral counseling interventions, absolute risk reduction of risky and harmful alcohol use by primary care patients was 10.5% (NNT=10). Mean drinks per week were reduced 2.9-8.7 compared to controls. There was not sufficient evidence to assess the effects of interventions on morbidity and mortality. The USPSTF gave screening for alcohol misuse a “B” recommendation. The Task Force recommended in favor of screening and behavioral counseling interventions to reduce alcohol misuse by adults, including pregnant women, in primary care settings.

Key Questions

- Was evidence for screening for IPV assessed by the USPSTF-designated Evidence-based Practice Center (EPC) in a fair way compared to other problems assessed by the USPSTF and its EPC’s?
- Is the recommendation for IPV screening consistent with other USPSTF recommendations?

Methods
To answer my key questions I used two sets of criteria, one for evaluating SER’s and one for evaluating the USPSTF recommendations. The criteria for evaluating evidence reviews are as follows:

1. Is there a focused question, including specific key questions, defined outcomes, and defined harms?

2. Was there an adequate search of the literature for evidence including appropriate sources of data, appropriate inclusion and exclusion criteria decided upon prior to the literature search, reasonable methods for determining article eligibility, and uniform methods of article review?

3. Was the evaluation of quality of individual studies standardized according to USPSTF criteria?

4. Was the assessment of overall quality of the evidence standardized according to USPSTF criteria?

5. Was the assessment of magnitude of net benefit of screening reasonable based on the quality of available evidence?

To evaluate the USPSTF recommendations on screening for IPV, depression, and alcohol misuse, I used the following criteria:

1. Was the recommendation consistent with the overall quality of evidence for the benefits of screening?

2. Was the recommendation consistent with the overall quality of the evidence for harms of screening?

3. Was the recommendation consistent with the net magnitude of benefit found?
4. Was the recommendation consistent with the net magnitude of harm found?
I compared and contrasted the three different sets of recommendations issued by the USPSTF based on these questions.

Results

*Key Questions on Screening for IPV*

All three SER's included specific key questions, but there were some important differences in what they asked. The EPC that conducted the SER on screening for IPV asked the following key questions:

1. How well does screening identify current harm or risk for harm from IPV?
2. What are the adverse effects of screening?
3. How well do interventions reduce harm from IPV?
4. What are the adverse effects of interventions?
5. Does screening for IPV reduce harm, premature death, and disability?

*Key Questions on Screening for Depression*

The EPC that conducted the SER on screening for depression asked three key questions that were similar to questions 1, 3, and 5 asked by the IPV group:

1. What is the accuracy of screening instruments for depression in primary care populations?
2. Is treatment of depression in primary care patients effective in improving outcomes?
3. Is screening more effective than usual care in identifying patients with depression, facilitating treatment of patients with depression, and improving outcomes?

**Key Questions on Screening for Alcohol Misuse**

The key questions asked in the SER for screening and behavioral interventions for alcohol misuse were as follows:

1. Is there direct evidence that behavioral counseling interventions to reduce risky or harmful alcohol use reduce morbidity and/or mortality?
2. What methods were used to identify target populations for behavioral counseling interventions in primary care?
3. What are adverse effects associated with alcohol use screening and screening-related assessment?
4. Does behavioral counseling intervention in primary care reduce risky or harmful alcohol use in the targeted subgroup? What are the essential elements of efficacious interventions?
5. Are there other positive outcomes from behavioral counseling interventions to reduce risky/harmful alcohol use?
6. What are adverse effects associated with behavioral counseling interventions for risky/harmful alcohol use?
7. What health care system influences are present in effective screening and screening-related assessments and interventions to reduce risky/harmful alcohol use or its outcomes?
Comparison of Key Questions

The difference in the sets of key questions for screening for IPV and screening for depression may have resulted in a differential assessment of harms between the IPV and depression SER’s and, consequently, the recommendations that followed each. The IPV group gave more weight to adverse effects of screening and interventions by making each of them a key question. The depression group cited a lack of evidence about adverse effects as a problem in the Results section of the SER, but the examination of adverse effects was not given the status of “key question.”

As with the IPV paper, the EPC evaluating evidence on screening and behavioral interventions for alcohol misuse designated two key questions for asking about adverse effects, those associated with screening and those associated with behavioral counseling interventions.

The papers were inconsistent in the way they dealt with the evaluation of screening instruments. The IPV and depression papers both had key questions that specifically asked about the accuracy of screening instruments; whereas the alcohol misuse paper instead asked, “What methods were used to identify target populations for behavioral counseling interventions in primary care?” without examining studies on the accuracy of screening instruments. This discrepancy
may have created a lesser burden of evidence for the alcohol misuse review than for the other two reviews.

The papers also had different approaches to answering the question of how interventions impacted outcomes. First, both the IPV and depression review articles asked one key question about how interventions affected outcomes in general. The alcohol misuse paper asked three different key questions about outcomes, which included morbidity, mortality, risky or harmful alcohol use, and “other positive outcomes.”

Finally, both the IPV and depression groups searched for evidence about the effects of screening on outcomes, but the IPV group sought data about harm, premature death, and disability, none of which were found. The depression group was seeking data on the effect of screening on rate of diagnosis, rate of treatment, and non-specific “outcomes” compared to usual care, which it was able to find. My own search of the literature for studies on the effect of screening for IPV on rate of detection of violence, rate of treatment, questionnaire scores, and other outcomes revealed one article that would have been included if the authors had allowed these as outcomes. It was a systematic review article that concluded that there is insufficient evidence of benefit to recommend in favor of implementing screening programs. The alcohol misuse SER did not include a question about the effects of screening on outcomes, another discrepancy between its key questions and those of the other two SER’s.
Defined Outcomes

IPV screening outcomes included:

1. Patient-oriented outcomes like indicators of physical abuse, neglect, emotional abuse, sexual abuse, and related health outcomes if reported such as depression.
2. How well a screening instrument compared to a standard.
3. Outcomes for evaluating studies on interventions: scores on Severity of Violence Against Women Scale, which includes questions about frequency of threats and actual physical and sexual violence.

Depression screening outcomes included:

1. Patient oriented outcomes like recovery from or improvement in depression after certain period of time and mean months of depression in 1 year
2. Sensitivity, specificity, ±LR of screening instruments.
3. Intermediate outcomes like improvement in SCL-20 score, % reduction in depressive severity score, rate of diagnosis, rate of treatment, % with >12 point decrease on SDS at 1 month, % with HAM-D<10 at 6 months, Mean GHQ at 6 months, Zung scale score, % depressed at 3 months DSM-IIIR criteria, % with <1 DSM-IIIR criteria symptoms, % depressed at 6 months, % depressed at 12 months, % depressed at 24 months, mean months of depression in 1 year, % with >12 point decrease on SDS at 1 month, % with HAM-D<10 at 6 months.
Alcohol misuse screening and behavioral counseling interventions outcomes included:

1. Patient-oriented outcomes like reduced morbidity and mortality and “other positive outcomes.”

2. 3 primary alcohol use outcome categories:
   1. average consumption (mean drinks per week)
   2. binge use (proportion bingeing)
   3. safe/moderate/recommended use (based on attaining each study’s recommended limits on average consumption and/or binge use)

The disparity in defined outcomes reflected the disparity in key questions, with the alcohol misuse group neglecting to evaluate any outcome reflecting the accuracy of screening instruments. All three defined some outcomes quantitatively and left room for more open-ended outcomes to emerge.

Harms of Screening

The same amount of evidence was found on the adverse effects of screening for all three problems. However, the assessment of potential adverse effects, which in no case was based on any evidence, was different in the three SER’s. The only potential adverse effect considered in the alcohol misuse SER was the potential for offending patients. The depression group discussed a longer list of potential adverse effects: false-positive results, adverse effects of treatment, adverse effects and costs of treatment for false-positives, and potential adverse effects of
labeling. The IPV group listed some potential adverse effects that were similar to those listed for the depression group like false-negatives discouraging inhibiting identification of those at risk, and false-positives leading to "inappropriate labeling and punitive attitudes." The IPV group also enumerated potential adverse effects that were more severe than those mentioned for depression and alcohol misuse screening without any additional evidence to warrant them. These potential consequences included psychological distress, escalation of abuse and family tension, loss of personal residence and financial resources, erosion of established family structure, loss of autonomy for the victim, lost time from work, and homicide.

Data Sources
Sources of data were comparable for the IPV, depression, and alcohol misuse SER's. The IPV group searched MEDLINE, Cochrane Register, PsychINFO, CINAHL, Health & Psychosocial Instruments, reference lists of systematic reviews, and experts. The depression EPC used MEDLINE, Cochrane database on depression, neurosis, and anxiety disorders, and bibliographies of SR's, original articles, Guide to Clinical Preventive Services, and the AHCPR 1993 Clinical Practice Guideline on Depression. The authors of the SER on screening and behavioral interventions for alcohol misuse used the Cochrane Database of Systematic Reviews and Database of Research Effectiveness, bibliographies of SR's, unpublished studies located through expert contacts and referrals, 1996
Inclusion and Exclusion Criteria

The inclusion and exclusion criteria for each SER are described in Table 1. The criteria were similar for all three. One difference for the SER for IPV was that it included studies that were conducted in primary care, ob-gyn, and emergency room settings, excluding studies of patients presenting with trauma. The other two SER’s included studies done in primary care and community settings but excluded those that were conducted in emergency departments. The IPV SER seemed to be more inclusive than the other two SER’s.

The SER for screening for depression stated that inclusion and exclusion criteria were agreed upon prior to performing a search of the literature. Neither the authors of the IPV SER nor the authors of the SER on screening for alcohol misuse stated whether such criteria were established prior to searching for evidence.

The method for determining eligibility for inclusion in the SER was not specified by the IPV authors. The depression group stated that two reviewers independently reviewed titles and abstracts of articles identified by literature searches and excluded ones which they agreed did not meet eligibility criteria. When they disagreed, all authors reviewed full articles and reached consensus.
about inclusion. For the SER of screening and behavioral interventions for alcohol misuse, one investigator reviewed 4,331 titles and abstracts, and a second investigator reviewed a random 35% of titles and abstracts for agreement. Approximately 95% agreement was found, and no articles that met review inclusion criteria were coded differently by the two reviewers. Seventeen studies met setting and quality criteria, but one did not have study results in time for the SER. Twelve of the sixteen reviewed studies addressed non-pregnant adults and were included in the SER.

*Methods of Article Review*

All three SER’s had similar methods for review articles, though the alcohol misuse group described their methods in more detail. The IPV group abstracted study design, number of participants, setting, length and type of interventions, length of follow-up, outcomes, methods of outcome measurement, and study duration of each study. The depression group abstracted study design and outcomes data from articles on screening accuracy, screening outcomes, and treatment. These data were used to construct evidence tables. For alcohol misuse, one primary reviewer abstracted relevant information for many more categories than were included in the other two SER’s. All data in evidence tables were checked by a second reviewer and articles excluded for quality or relevance were listed in Table 10 of the Appendix. Abstracted data included author, year published, type of trial, setting, definition of a standard drink used in the study, total number randomized to the study, proportion of female and non-white
participants, baseline alcohol consumption, proportion that was alcohol dependent, and proportion that was motivated, help-seeking, or thought they had a problem with alcohol. The alcohol misuse SER was the most explicit about outcomes that were abstracted, including mean drinks per week or reduction in mean drinks per week, percentage of participants without binge drinking (usually defined as more than four drinks per occasion), and percentage of participants achieving recommended drinking levels or patterns as defined by the individual studies.

Methods of Evaluating Quality of Individual Studies

The methods for evaluating the quality of individual studies were similar for all three SER's. In the IPV SER, two reviewers independently rated each study's quality using criteria specific to different study designs developed by the USPSTF. Each study was categorized as good, fair, or poor. When reviewers disagreed, a final score was reached "through consensus." In the depression SER, internal and external validity for each article were rated separately using criteria developed by the USPSTF. The criteria for determining internal validity were described and were exactly the same as those used by the IPV authors. In the SER on screening and interventions for alcohol misuse, the quality of the articles was graded using the same USPSTF criteria, supplemented by guidelines on evaluating study randomization, attrition, and intention-to treat analyses from the Cochrane Drug and Alcohol Group. The final quality rating for each study was assigned "by consensus" of the investigator team.
Evaluation of Quality of Individual Studies

To answer the question of how accurate screening instruments are, the IPV SER found four good-quality studies, seven fair-quality studies, and three poor-quality studies, as defined by the USPSTF. The depression SER found many more studies addressing the accuracy of instruments designed to screen for depression: twenty-three good-quality studies and eighteen fair-quality studies. The alcohol misuse SER did not seek evidence on the accuracy of screening instruments, and none was included in the final evidence review.

The IPV SER found two fair-quality studies on the effectiveness of interventions for IPV while the authors of the depression review found eight good-quality studies and eleven fair-quality studies on the effectiveness of either pharmacotherapy or psychotherapy at treating depression. The alcohol misuse SER included seven good-quality studies and five fair-quality; this group excluded poor-quality studies from eligibility for the SER. There was a large disparity in the amount of evidence available for the three different SER’s and a notable lack of any good-quality studies on the effectiveness of interventions designed to improve outcomes for IPV.

No studies of any quality were found that addressed the question of adverse effects due to either screening or interventions for any of these health problems.
There was no difference in the evidence for adverse effects among the SER’s since none was found for any of the three topics covered in this review.

The authors of the IPV SER sought studies on the effects of screening on harm, premature death, and disability, but none was available that met the eligibility criteria for the review. In evaluating the clinical outcomes of screening, the authors of the depression SER asked the question about outcomes in terms of the rate of diagnosis of depression, the rate of treatment, and any other outcomes of screening. They found eight good-quality studies and five fair-quality studies that evaluated the effect of screening on diagnosis, care, and outcomes. If the authors of the IPV SER had expanded the outcomes beyond premature death, disability and “harm,” more similar to the outcomes evaluated by the authors of the depression study, they may have found more than the two outcomes studies that were ultimately included. A quick search of MEDLINE for “outcomes of screening” and “intimate partner violence” revealed at least one more systematic review that appears to meet the SER’s inclusion and exclusion criteria. However, that systematic review concluded that there was insufficient evidence to recommend for universal screening. The authors of the review of screening and behavioral interventions for alcohol misuse did not seek or find evidence of the effects of screening on outcomes.

Methods of Evaluating Overall Quality of Evidence
The methods for evaluating the overall quality of the evidence were not discussed by the authors of the IPV SER. However, a summary-of-evidence table included an assessment of internal and external validity, but not coherence (number of studies, homogeneity of those studies, precision of findings, and direction of results). The authors of the depression SER described their methods for evaluating the overall quality of the evidence as follows: they rated the aggregate internal validity, external validity, and coherence for each of the key questions. These three ratings were combined to rate the overall quality of evidence. The SER for screening and behavioral interventions for alcohol misuse described a process identical to the one used in the SER for screening for depression.

_Evaluation of the Overall Quality of Evidence for Screening for IPV_

The IPV screening review rated the overall quality of evidence separately for the accuracy of screening instruments and for the efficacy of interventions. The overall quality of evidence for the accuracy of screening instruments was assessed as “poor-good” for non-randomized controlled trials. Screening instruments were found to have fair-good correlation with longer instruments, but the authors were concerned that there were no studies of instruments that followed women longitudinally. The external validity was not given a rating, but the authors noted that several instruments were tested using a variety of settings and populations. The overall quality of evidence for the efficacy of interventions was assessed as “fair” for one randomized controlled trial and one controlled trial without randomization. External validity was a problem for these trials because the study
populations were small and included only pregnant women. No evidence was found on the effects of screening on clinical outcomes, nor on adverse effects of screening or interventions.

*Evaluation of the Overall Quality of Evidence for Screening for Depression*

The SER for screening for depression evaluated the overall quality of the evidence on accuracy of screening instrument and efficacy of treatments as well the effects of screening on diagnosis, care, and outcomes. The investigators found good overall internal and external validity and good coherence for the studies on the accuracy of instruments designed to screen for depression in adults. They found good overall internal and external validity and good coherence for the studies on the effectiveness of pharmacotherapy in adults. They found fair overall internal validity and good overall external validity and coherence for the effectiveness of psychotherapy in adults. The investigators assessed the studies on the effects of screening on diagnosis, care, and outcomes as having good internal and external validity, but only fair-poor coherence. No evidence was found on the topic of adverse effects of depression screening and interventions.

*Evaluation of the Overall Quality of Evidence for Screening and Behavioral Interventions for Alcohol Misuse*

The authors of the SER on screening and behavioral interventions for alcohol misuse evaluated the overall quality of evidence available for each of the seven key questions they asked. They found fair overall evidence that behavioral
counseling interventions to reduce harmful alcohol use actually reduce morbidity and/or mortality. They found fair overall evidence about which methods were useful for identifying target populations for behavioral counseling interventions in primary care. They found poor overall evidence on adverse effects associated with alcohol use screening and screening-related assessment. They rated evidence on whether behavioral counseling interventions in primary care reduces risky or harmful alcohol use as poor-fair quality overall. They found poor-fair overall evidence for other positive outcomes from behavioral counseling interventions. There was poor overall evidence on the adverse effects associated with behavioral counseling interventions. Finally, they evaluated evidence on “health care system influences” that are present in effective screening and screening-related assessments and interventions as being of fair overall quality.

**Direction and Magnitude of the Evidence for IPV Screening and Interventions**

Screening instruments for IPV were found to have fair-good correlation with longer instruments. Two studies of interventions “suggest benefit using self-reported outcomes,” but the magnitude of increased detection was not reported by the SER. Both used the Severity of Violence Against Women Scale (SVAWS) to evaluate the screening instruments. One of these studies found that one intervention consisting of “unlimited access to counseling services of a ...professional counselor with expertise in domestic violence,” and the services of a “mentor mother” who “offered support, education, referral, and assistance in accessing community resources through personal visits and telephone contacts”
had significantly lower ($p<0.05$) physical violence scores on the SVAWS two months following initiation of the intervention than those who received only professional counseling (adjusted means of 34.7 and 39.5, respectively). However, the physical violence score for those who received professional counseling and a “mentor mother” were not significantly lower than for those who received only a brief intervention consisting of “a wallet-sized resource card that included phone numbers of local agencies to assist with domestic violence” and “information about planning for personal safety” (adjusted mean of 38.2). No evidence was found for the effects of screening on clinical outcomes nor for the adverse effects of screening or interventions.

The second study included in the IPV SER on the effects of interventions on outcomes found differences between the intervention group whose participants received a “one-to-one interview conducted by …nurses who had received a minimum of 4 hours training with the investigator” along with “information on applying for legal protection orders and filing criminal charges, and community resource phone numbers such as the shelter, hot lines, and law enforcement” and the comparison group whose participants received only a “wallet-sized card with information on community resources for abuse, including the crisis hot line, the shelter, and law enforcement.” The authors found higher mean SVAWS scores for threats and physical violence in the comparison group than in the intervention group (33.4 vs. 27.3 for threats; 35.9 vs. 33.1 for actual violence) six months following the initiation of intervention. However, the comparison group started
out with higher mean scores than the intervention group at the time of entry into the study (46.4 vs. 39.1 for threats; 52.0 vs. 47.9 for actual violence), and it is unclear from the study which of these differences is statistically significant.

Direction and Magnitude of the Evidence for Depression Screening and Interventions

Screening instruments for depression were found to improve the rate of detection of depression by 10-47%. The use of pharmacologic interventions resulted in a 10-30% absolute risk reduction for main outcomes, including HAM-D score reduction and recovery from depression. Investigators reported that psychotherapeutic interventions appeared as effective as antidepressant interventions. No evidence was found on the adverse effects of screening or interventions. The effects of screening on clinical outcomes were mixed in different studies. Two small older trials found large improvements in depression. Two larger good quality trials found moderate improvements (9%) in recovery from depression. Four other studies found small or no improvements in clinical outcomes for those who had been screened.

Direction and Magnitude of the Evidence for Alcohol Misuse Screening and Interventions

Studies found by the investigators that addressed the question of whether interventions for alcohol misuse reduce morbidity or mortality produced mixed results: in two of four studies examining problem scores, those in all groups
improved with no differences between intervention and control groups at follow-up. Two other studies showed no changes from baseline to follow-up within or between groups. With other outcomes, studies generally found no improvement or similar improvements in interventions and controls over the duration of the trials. Of the five trials that examined health care utilization, only one found reduced self-reported hospital days at twelve months. In one good-quality trial there was a trend toward reduced all-cause mortality in interventions compared to controls (three vs. seven deaths; p>0.1). In a second study a brief single-contact intervention had no long-term effects on morbidity, mortality, or alcohol consumption at ten-year follow-up. A third study found that men who received an intervention had a significantly lower total mortality rate (24/100,000 person-years) than non-invited controls (30/100,000 person-years) (p<0.02) and had significantly reduced alcohol-related mortality after 3 to 21 years.

Various methods were used to identify target populations for behavioral counseling interventions in primary care, including screening (identifying patients with probable risky/harmful alcohol use) and screening-related assessment (confirming screening results and distinguishing patients suitable for brief interventions from those needing specialty care referral). Screening usually involved self-administered questionnaires or brief interviews to assess average quantity or frequency and binge use.
The investigators found that screening and interventions integrated into primary care reduce alcohol consumption and improve use patterns: four good-quality trials reported that weekly drinking was reduced 13-34% compared to controls, resulting in 2.9-8.7 fewer mean drinks per week at follow-up in intervention group compared to control group. One good-quality trial did not find significant change in average use. All five good-quality trials found significant effects on recommended or sage alcohol use, resulting in 10-19% more intervention participants than controls reporting recommended or safe drinking patterns. Two of four good-quality trials reported significantly reduced binge drinking.

The investigators found that there are often health care system influences present in effective screening and screening-related assessments and interventions to reduce risky/harmful alcohol use or its outcomes. In all 12 trials, additional staff or systems support were required to provide screening and assessment services and sometimes intervention support. In nearly every study, research staff conducted the screening and assessment outside the routine care encounter. Most of these processes took more than 30 minutes. Provider training sessions ranged from fifteen minutes to 2.5 hours. Only three studies used incentives for participating providers or patients. Besides usual care physicians, studies also used research staff or non-physician health care staff to deliver some or all of the intervention. Research staff often performed important support functions like prompting the provider and supplying intervention materials to the chart.
No studies were found that addressed what adverse effects were associated with alcohol use screening or assessment. No studies were found that addressed the question of what adverse effects were associated with behavioral counseling interventions for risky or harmful alcohol use. No other positive outcomes from behavioral counseling interventions emerged. “Null findings may reflect lack of an effect, lack of appropriate measures, and/or reduced power for secondary analyses.”

**IPV Recommendations**

Next I examined the consistency of each recommendation statement with the overall quality of the evidence and the net benefit found for each key question in its corresponding SER. For the question of accuracy of IPV screening instruments the SER found four good quality studies and two fair studies that all showed good correlation (R=0.69-0.85) between screening instruments and longer questionnaires. It also found three studies rated “poor” because they used an interview as the reference standard, but the interview was not defined. In each of these studies the screening instruments yielded a higher rate of detection of IPV (14% vs. 0.4%, 85% vs. 59%, and 41% vs. 14%). In the Discussion section of the Recommendation Statement, the USPSTF wrote, “Newer brief instruments designed to identify women who are victims of IPV in primary care settings compare well with lengthier, previously validated instruments,” but in the Summary section of the same paper, they wrote, “The USPSTF found no existing
studies that determine the accuracy of screening tools for identifying...IPV among...women in the general population.”

To answer the question of how effective interventions are for IPV, the SER found two fair quality studies that indicated interventions improved self-reported outcomes. One study compared three interventions (card with resources vs. unlimited access to counselor in clinic vs. counseling plus “mentor mother” in community).⁹ According to the USPSTF’s SER, “abuse decreased significantly in all groups with no significant differences between the groups.” However, a careful reading of the original paper revealed that the counseling plus “mentor mother” group had significantly lower physical violence scores on the SVAWS two months following initiation of intervention than those who received only counseling.

The other study on IPV outcomes compared two interventions (three counseling sessions vs. card with resources). The IPV SER stated that less violence occurred in the intervention group at six and twelve months using the SVAWS (p=0.052) and using the Index of Spouse Abuse (ISA) (p=0.007).¹⁰ The Discussion section of Recommendation Statement said, “both studies...showed a trend (not statistically significant) in women reporting decreased violence after brief counseling or outreach interventions.” In fact, both studies produced results that showed a statistically significant difference between different types of interventions. The first study showed that the support of a “mentor mother” was
associated with significantly lower physical violence scores on the SVAWS; the second showed significantly less violence using the ISA instrument and an arguably significant difference using the SVAWS. However, given the fair quality of the studies and the fact that in the second study participants were not assigned to intervention groups randomly, the Summary of Recommendation accurately stated, “The USPSTF found limited evidence as to whether interventions reduce harm to women.”

On the question of effect of screening on harm, premature death, and disability for IPV, the SER found no evidence on the effect of screening on harm, premature death, and disability for IPV. In the Discussion section of the Recommendation Statement the USPSTF wrote, “No studies were found that directly addressed the impact of screening on reducing harmful outcomes.” The Summary of Recommendation states, “The USPSTF found no direct evidence that screening for family and intimate partner violence leads to decreased disability or premature death.”

The USPSTF Recommendation Statement for IPV reflected the findings in the SER on the adverse effects of IPV screening, but it treats the same amount of evidence differently in the IPV Recommendation Statement than in either the depression or the alcohol misuse Recommendation Statement. The SER found no evidence on the adverse effects of IPV screening. The USPSTF stated in the Discussion section of its Recommendation statement, “No studies have directly
addressed the harms of screening and intervention for...IPV. False-positive test results, most common in low-risk populations, may compromise the clinician-patient relationship. Additional possible harms of screening may include loss of contact with established support systems, psychological distress, and an escalation of abuse. However, none of these potential harms has been studied.” In the Summary it was reiterated that “no studies have directly addressed the harms of screening and interventions for...IPV. As a result, the USPSTF could not determine the balance between the benefits and harms of screening for...IPV.”

**Depression Recommendations**

Regarding the question of accuracy of depression screening instruments, the SER found “good” overall quality of evidence that screening improved rate of detection by 10-47%. In the Recommendation Statement, the USPSTF stated that “most instruments have relatively good sensitivity but only fair specificity.”

“Assuming...a prevalence of major depression of 5-10% in primary care settings, about 24-40% of patients who screen positive will have major depression.” This is consistent with the Summary of Recommendations, which stated “The USPSTF found good evidence that screening improves the accurate identification of depressed patients in primary care settings.”

The SER for depression found good internal and external validity and coherence for studies that showed moderate benefits for pharmacotherapy. It found fair internal validity and good external validity and coherence for studies that showed
moderate benefits for psychotherapy. Treatment resulted in about 10-30% ARR for main outcomes for both pharmacotherapy and psychotherapy. In the Recommendation Statement the USPSTF wrote, “Antidepressant medications for major depression...are clearly more effective than placebo,” and “Psychosocial and psychotherapeutic interventions are probably as effective as antidepressant medications for major depression.” The Summary of Recommendations stated, “The USPSTF found good evidence...that treatment of depressed adults identified in primary care settings decreases clinical morbidity.”

The question on the effects of screening for depression asked about diagnosis, care, and outcomes for depression, rather than harm, premature death, and disability like in the IPV paper. The SER characterized the studies evaluating the effect of screening on patient-oriented outcomes as having good internal and external validity and fair-poor coherence. The effects of screening on outcomes were mixed. In the Recommendation Statement the USPSTF described the trials that demonstrated positive effects of screening and those that showed no effect. The Summary of Recommendations states, “Trials that have directly evaluated the effect of screening on clinical outcomes have shown mixed results.” The Recommendation statement was consistent with the quality, direction, and magnitude of the evidence reported in the SER.

Like the IPV SER, the depression SER found no evidence on the adverse effects of depression screening. The USPSTF stated in the Discussion section of its
Recommendation statement, “The potential harms of screening include false-positive screening results, the inconvenience of further diagnostic work-up, the adverse effects and costs of treatment for patients who are incorrectly identified as being depressed, and potential adverse effects of labeling. None of the research reviewed provided useful empirical data regarding these potential adverse effects.” In the Summary the authors wrote, “The USPSTF concluded the benefits of screening are likely to outweigh any potential harms” without any mention of the lack of evidence on harms.

*Alcohol Misuse Recommendations*

The authors of the SER on screening and behavioral interventions for alcohol misuse discovered various methods for identifying target populations for behavioral counseling interventions in primary care: screening (identifying patients with probable risky/harmful alcohol use) and screening-related assessment (confirming screening results and distinguishing patients suitable for brief interventions from those needing specialty care referral). Screening usually involved self-administered questionnaires or brief interviews to assess average quantity or frequency and binge use. Unlike IPV and depression, the authors did not determine the extent to which screening improved the rate of detection of risky or harmful alcohol use. In fact, accuracy of screening instruments was not a question posed by the authors of the review, who instead asked, “What methods were used to identify risky/harmful drinkers for behavioral counseling interventions in primary care?” Because of this discrepancy, it seemed
misleading for the USPSTF to write in the Recommendation Statement, “The USPSTF found good evidence that screening in primary care settings can accurately identify patients whose levels or patterns of alcohol consumption...place them at risk for increased morbidity and mortality...”

The USPSTF recommendation for screening and behavioral interventions for alcohol misuse was consistent with the SER findings. The SER concluded that screening and interventions integrated into primary care reduce alcohol consumption and improve use patterns: four good-quality trials reported that weekly drinking was reduced 13-34% compared to controls, resulting in 2.9-8.7 fewer mean drinks per week at follow-up in intervention group compared to control group. One good-quality trial did not find a significant change in average use. All five good-quality trials found significant effects on recommended or safe alcohol use, resulting in 10-19% more intervention participants than controls reporting recommended or safe drinking patterns. Two of four good-quality trials reported significantly reduced binge drinking. Correspondingly, the USPSTF reported in the Recommendation Statement, “good evidence that brief behavioral counseling interventions with follow-up produce small to moderate reductions in alcohol consumption that are sustained over 6- to 12-month periods or longer.”

Unlike the authors of the SER’s on IPV and depression, the authors of the SER for alcohol misuse did not include in their list of key questions one about the effects of screening on health outcomes. Instead they asked, “Is there direct
evidence that behavioral counseling interventions to reduce risky or harmful alcohol use reduce morbidity and/or mortality?” Eligible studies produced mixed results, which are described above. In their Recommendation Statement the Task Force wrote, “The USPSTF found some evidence that interventions lead to positive health outcomes 4 or more years post-intervention, but found limited evidence that screening and behavioral counseling reduce alcohol-related morbidity.” The Recommendation Statement is consistent with the quality of the evidence and the magnitude of benefit found.

The SER on screening and behavioral interventions for alcohol misuse reported that no studies were found that addressed adverse effects associated with alcohol use screening or assessment or with behavioral counseling interventions for alcohol use. The USPSTF mentioned in the Discussion section of the Recommendation Statement that little evidence was found to address the question of the magnitude of harms of screening and behavioral counseling interventions. In contrast to IPV Recommendation Statement, the lack of harms was not stated in the Summary of Recommendations, which, parallel to the depression recommendation, stated, “The USPSTF concluded that the benefits of behavioral counseling interventions to reduce alcohol misuse by adults outweigh any potential harms.”

Conclusions
It is likely that evidence for screening for IPV was not assessed in a fair way compared to other problems addressed by the USPSTF. The USPSTF Recommendation for screening for IPV was not consistent with recommendations for comparable health problems. However, it is unlikely that the recommendation for screening for IPV would be different if disparities in assessment and synthesis were addressed.

The largest discrepancy among the three sets of guidelines was the different speculation and emphasis placed on adverse effects of screening and interventions. It is unclear why a lack of evidence on adverse effects caused the USPSTF to declare in its Recommendation Statement on screening for IPV that it was unable to determine the balance between benefits and the harms, but the same lack of evidence for screening for depression and screening and behavioral interventions for alcohol misuse did not prevent the USPSTF from issuing a B recommendation and stating that there is at least fair evidence that the benefits of screening outweigh harms for those two health problems.

Nonetheless, it is clear that there is much less evidence available on the effects of screening and interventions for IPV than for depression and for alcohol misuse. The USPSTF assessment that there is insufficient evidence to recommend in favor or against universal screening for IPV is probably correct. Since there appear to be several good quality studies demonstrating the usefulness of screening
instruments at identifying women at-risk for IPV, what is needed is evidence on whether screening leads to interventions that improve outcomes for such women.

Future Research

To answer the question of whether the benefits of universal screening for IPV outweigh the risks, a randomized controlled trial should be conducted in several primary care settings like family practice offices in rural and urban settings as well as a hospital-based emergency room. Women in each setting should be randomized to screening for IPV or to usual care. It should be determined ahead of time the sample size that would be required to detect a significant difference between the two comparison groups. Participants should be randomized to either the screening group or to the usual care group by a reliable method like using a random number table or computer-generated random numbers. Following randomization, investigators should ensure that the groups have comparable demographic characteristics and do not differ in a way that would contribute to or mask differences between the measured outcomes for each of the two groups.

Patients who are randomized to the screening group should be screened using a validated screening instrument like the HITS scale, which asks patients how often their partner physically hurt, insulted, threatened, and screamed at them. Within 72 hours of randomization, patients should be assessed using a “gold standard” for measuring intimate partner violence. The most widely used standard is the Conflict Tactics Scale (CTS), but the authors of the SER on Screening for
Family and Intimate Partner Violence note that, “the Conflict Tactics Scales may not have undergone sufficient testing of its validity to qualify as a gold standard” (cite SER). The person conducting the assessment using the CTS or other standard should be blinded to the screening results.

Following screening, those participants who are identified as having an increased risk of experiencing abuse should receive and intervention consisting of, at a minimum, education about intimate partner violence and referral to a program or agency that provides support for victims of IPV. Interventions taken by clinicians following the screening vs. usual care stage should be assessed by reviewing the medical records of participants. Patients should be classified as having been treated for IPV if the medical record documented counseling for or education about IPV and a referral to a mental health provider or community agency that provides support for victims of IPV. Patient and physician satisfaction should be assessed using an instrument such as Likert scales.

Patients in the usual care and screening groups should be assessed at three months, six months, one year, two years, five years, and ten years following the beginning of the study. One instrument that has been used to measure outcomes of interventions for IPV in the past is the Severity of Violence Against Women Scale (SVAWS). It measures threats of physical violence, actual physical violence, and sexual abuse. Other outcomes that ought to be measured include homicide, suicide, all-cause mortality, ER visits, spontaneous abortion, quality of
life, and psychiatric disorders like depression and anxiety. These outcomes should be measured in a way that they would be assessed equivalently whether they were positively or negatively influenced by screening.

Only a randomized controlled trial would provide evidence that would answer the question of whether the risks of screening outweigh the benefits. The more women that can be followed for longer periods of time, the more information will be gained. Consultation with a member of the University of North Carolina-Chapel Hill’s Institutional Review Board revealed that this design would likely be approved. She anticipated that the biggest challenge being the ability to ensure confidentiality for the participants. Especially concerning would be how to contact them for follow-up so that their partners would not be alerted to their status as a study participant. Solutions to this problem include using only contact information to which the participants consent, not leaving messages for the participants, not sending letters or e-mail messages, clarifying times when it is safe to call, blocking features like caller ID, and always allowing the participant to control when and how she is contacted.

Limitations

One limitation of this review is my personal experience as an advocate for survivors of domestic and sexual violence since 1998. I have made every effort to maintain an unbiased attitude in this examination of USPSTF guidelines but must acknowledge my frustration with the lack of choices and resources available to
the many women, men, and children who are abused by their own family members. To address my own bias I have depended on my advisor and other readers to bring to my attention statements and conclusions that are or appear to be lacking in objectivity.
References


3. Chamberlain, L. The USPSTF recommendation on intimate partner violence: what we can learn from it and what we can do about it. Family Violence Prevention and Health Practice. Vol. 01, January 2005


Table 1. Systematic Evidence Review Criteria

<table>
<thead>
<tr>
<th></th>
<th>Focused question: key questions</th>
<th>Focused question: defined outcomes</th>
<th>Focused question: defined harms</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPV</td>
<td>1. Accuracy of screening instruments?</td>
<td>1. Patient-oriented: Harm, premature death, disability (no studies were found)</td>
<td>Speculative: false-negatives, false-positives, psychological distress, escalation of abuse and family tension, loss of personal residence and financial resources, erosion of established family structure, loss of autonomy for the victim, lost time from work, and homicide. No evidence found on potential adverse effects of either screening or interventions for IPV.</td>
</tr>
<tr>
<td></td>
<td>2. Adverse effects of screening?</td>
<td>2. Studies on instruments: how well instrument in question correlated with previously validated instrument (standard instrument varied)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Adverse effects of interventions?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5. Does screening reduce harm, premature death, and disability?</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Does treatment improve outcomes?</td>
<td>2. Recovery from depression, improvement in scores</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Does screening improve diagnosis, treatment, and outcomes better than usual care?</td>
<td>3. Rates of diagnosis and treatment, mean months of depression in 1 year, % with improvement on standard questionnaires, % depressed after 3 months, 6 months, 12 months, and 24 months.</td>
<td></td>
</tr>
<tr>
<td>Alcohol misuse</td>
<td>Focused question: key questions</td>
<td>Focused question: defined outcomes</td>
<td>Focused question: defined harms</td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------</td>
<td>-----------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>1. Is there direct evidence that behavioral counseling interventions to reduce risky or harmful alcohol use reduce morbidity and/or mortality?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. What methods were used to identify target populations for behavioral counseling interventions in primary care?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. What are adverse effects associated with alcohol use screening and screening-related assessment?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Does behavioral counseling intervention in primary care reduce risky or harmful alcohol use in the targeted subgroup?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a. What are the essential elements of efficacious interventions?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Are there other positive outcomes from behavioral counseling interventions to reduce risky/harmful alcohol use?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. What are adverse effects associated with behavioral counseling interventions for risky/harmful alcohol use?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. What health care system influences are present in effective screening and screening-related assessments and interventions to reduce risky/harmful alcohol use or its outcomes?</td>
<td>1. Reduced morbidity and mortality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. 3 primary alcohol use outcome categories:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. average consumption (mean drinks per week)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. binge use (proportion bingeing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. safe/moderate use (based on attaining each study’s recommended limits on average consumption and/or binge us)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Other positive outcomes</td>
<td>Speculative: potential for offending patients. No evidence found on potential adverse effects for either screening or interventions for alcohol problems.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence search: sources of data</td>
<td>Evidence search: inclusion criteria</td>
<td>Evidence search: exclusion criteria</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>IPV</strong></td>
<td>MEDLINE, Cochrane Register, PsychINFO, CINAHL, Health &amp; Psychosocial Instruments, reference lists of SR's, and experts</td>
<td>English, applicable to U.S. clinical practice, describes abuse of women, conducted in primary care, ob-gyn, or ER settings, includes a clinician in process of assessment or intervention. For assessment: evaluates assessment procedure, clearly describes sample, instrument outcomes, and data collection. For intervention: measures effectiveness of intervention compared to comparison groups</td>
<td>Studies on patients presenting with trauma, interventions designed to educate health care professionals or increase screening rates, mandatory reporting laws, accuracy of physician diagnosis and reporting, physician factors related to reporting, and descriptions of programs</td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td>MEDLINE, Cochrane database on depression, neurosis, and anxiety disorders, and bibliographies of SR's, original articles, Guide to Clinical Preventive Services, and the AHCPR 1993 Clinical Practice Guideline on Depression</td>
<td>English, humans, original data only. For assessment: 1994-1999, includes “criterion standard”, measures sensitivity and specificity, conducted in primary care or community setting. For effectiveness of pharmacotherapy: 1994-1999, RCT’s, conducted in primary care or community settings. For effectiveness of psychotherapy: 1966-1999, same as above</td>
<td>Conducted in hospitals or psychiatry clinics, non-published studies, non-English, animal studies, letters, editorials, non-systematic reviews</td>
</tr>
<tr>
<td>Evidence search: sources of data</td>
<td>Evidence search: inclusion criteria</td>
<td>Evidence search: exclusion criteria</td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol misuse</strong></td>
<td><strong>Key questions 1,4,5,7:</strong></td>
<td>Emergency department or hospital</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cochrane Database of Systematic</td>
<td>setting, non-human, non-English</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reviews and Database of Research</td>
<td>abstract, study design (case</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Effectiveness, bibliographies of</td>
<td>reports or series, letters to</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SR's, unpublished studies located</td>
<td>editor or editorials,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>through expert contacts and</td>
<td>observational studies),</td>
<td></td>
</tr>
<tr>
<td></td>
<td>referrals, 1996 Guide to Clinical</td>
<td>population (age &lt; 12 years,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>preventive Services, MEDLINE,</td>
<td>dependent drinkers; comorbid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cochrane Controlled Clinical</td>
<td>treatment populations),</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trials, PsychInfo, HealthSTAR, and</td>
<td>setting (specialty treatment,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CINAHL from 1994-April 2002</td>
<td>behavioral health, community or</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Key questions 3 and 6:</strong></td>
<td>school without clinic/health care</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MEDLINE and PsychInfo 1994-April</td>
<td>personnel, non-comparable cultural</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2002</td>
<td>setting), intervention (no</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>behavioral counseling intervention,</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>intervention delivered in</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>non-replicable ways), poor quality</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>according to USPSTF criteria;</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>studies on pregnant women and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>adolescents.</td>
<td></td>
</tr>
<tr>
<td>IPV</td>
<td>No</td>
<td>Not specified</td>
<td>Abstracted study design, number of participants, setting, length and type of interventions, length of follow-up, outcomes, methods of outcome measurement, and study duration of each study.</td>
</tr>
<tr>
<td>-----</td>
<td>----</td>
<td>--------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Depress</td>
<td>Yes</td>
<td>2 reviewers independently reviewed titles and abstracts of articles identified by literature searches and excluded ones which they agreed did not meet eligibility criteria. When they disagreed, all authors reviewed full articles and reached consensus about inclusion.</td>
<td>Abstracted study design and outcomes data from articles on screening accuracy, screening outcomes, and treatment. These data were used to construct evidence tables.</td>
</tr>
<tr>
<td>Evidence search: I/E Criteria decided on prior to lit. search?</td>
<td>Evidence search: Method for determining eligibility</td>
<td>Evidence search: Methods of article review</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Alcohol misuse</td>
<td>No</td>
<td>One primary reviewer abstracted relevant information using data-abstraction forms. All data in evidence tables were checked by a second reviewer and articles excluded for quality or relevance appear in Table 10 of the Appendix. Abstracted author, year published, type of trial, setting, and definition of a standard drink used in the study. Abstracted total number randomized to the study, proportion of female and non-white participants, baseline alcohol consumption, proportion that was alcohol dependent, and proportion that was motivated, help-seeking, or thought they had a problem with alcohol. Alcohol use outcomes measured: 1. mean drinks per week or reduction in mean drinks per week. 2. percentage of participants without binge drinking (usually defined as &gt;4 drinks per occasion). 3. percentage of participants achieving recommended drinking levels or patterns (as defined by the study).</td>
<td></td>
</tr>
<tr>
<td>One investigator reviewed 4331 non-duplicative titles and abstracts, and a second investigator reviewed a random 35% of titles/abstracts for concordance. Approximately 95% agreement was found, and non articles that met review inclusion criteria were discrepantly coded by he 2 reviewers. 17 studies met setting and quality criteria, but 1 did not have study results in time for the SER. 12 of the 16 reviewed studies addressed non-pregnant adults and were included in the SER</td>
<td>One primary reviewer abstracted relevant information using data-abstraction forms. All data in evidence tables were checked by a second reviewer and articles excluded for quality or relevance appear in Table 10 of the Appendix. Abstracted author, year published, type of trial, setting, and definition of a standard drink used in the study. Abstracted total number randomized to the study, proportion of female and non-white participants, baseline alcohol consumption, proportion that was alcohol dependent, and proportion that was motivated, help-seeking, or thought they had a problem with alcohol. Alcohol use outcomes measured: 1. mean drinks per week or reduction in mean drinks per week. 2. percentage of participants without binge drinking (usually defined as &gt;4 drinks per occasion). 3. percentage of participants achieving recommended drinking levels or patterns (as defined by the study).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methods for evaluating quality of individual studies</td>
<td>Evaluation of quality of individual studies</td>
<td>Methods for evaluating overall quality of the evidence</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------------</td>
<td>--------------------------------------------</td>
<td>-----------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>IPV</td>
<td>Accuracy of screening instruments: 4 “good” quality, 7 “fair,” 3 “poor.” All are non-randomized controlled trials. Adverse effects of screening: no evidence available Effectiveness of interventions: 2 studies rated “fair.” 1 RCT and 1 non-randomized controlled trial Adverse effects of interventions: no evidence available Effect of screening on harm, premature death, and disability: no evidence available</td>
<td>Not discussed. Summary of evidence table included internal and external validity, but not coherence.</td>
<td></td>
</tr>
<tr>
<td>Depression</td>
<td>Accuracy of screening instruments: 23 “good,” 18 “fair” Effectiveness of interventions: Pharmacotherapy: 8 “good,” 11 “fair” Effect of screening on diagnosis, care, and outcomes: 8 “good,” 5 “fair”</td>
<td>Rated the aggregate internal validity, external validity, and coherence (number of studies, homogeneity of those studies, precision of findings, and direction of results) for each of the key questions. These 3 ratings were combined to rate the overall quality of evidence.</td>
<td></td>
</tr>
</tbody>
</table>

Two reviewers independently rated each study’s quality using criteria specific to different study designs developed by the USPSTF. Each study was categorized as good, fair, or poor. When reviewers disagreed, a final score was reached “through consensus.”

Internal and external validity for each article were rated separately using criteria developed by the USPSTF. The criteria for determining internal validity were described and were exactly the same as those used by the IPV authors.
<table>
<thead>
<tr>
<th>Alcohol misuse</th>
<th>Methods for evaluating quality of individual studies</th>
<th>Evaluation of quality of individual studies</th>
<th>Methods for evaluating overall quality of the evidence</th>
</tr>
</thead>
</table>
| Quality of the articles was graded using the USPSTF criteria, supplemented by guidelines on evaluating study randomization, attrition, and intention-to treat analyses from the Cochrane Drug and Alcohol Group. The final quality rating for each study was assigned by consensus of the investigator team. | *KQ1:* 2 “good” and 2 “fair” quality studies  
*KQ2:* 7 “good” and 4 “fair” quality RCT’s; 1 “fair” quality CCT of relevant primary care populations  
*KQ3:* No studies found  
*KQ4:* 7 “good” and 4 “fair” quality RCT’s; 1 “fair” quality CCT of relevant primary care populations (same studies used to answer KQ2)  
*KQ4a:* Same studies examined as in KQ4, but less information available about the elements of effective interventions.  
*KQ5:* 3 studies of “limited usefulness” found that addressed cost-effectiveness  
*KQ6:* No studies found  
*KQ7:* “little information reported about contextual factors” | “We used the UPSTF approach to grade the overall quality of evidence for each key question.” (As described above for the depression paper) |
<table>
<thead>
<tr>
<th></th>
<th>Overall assessment of quality of evidence</th>
<th>Evaluation of magnitude of net benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>IPV</strong></td>
<td><em>Accuracy of screening instruments</em>: Internal validity: Poor-good. External validity: Several brief instruments tested using a variety of settings and populations. <em>Effectiveness of interventions</em>: Internal validity: Fair External validity: Studied in small populations of pregnant women only.</td>
<td>Screening instruments: “fair-good correlation” with longer instruments. 2 intervention trials “suggest benefit using self-reported outcomes.” Both used the SVSWS; magnitude not reported. No evidence for adverse effects or effect of screening on clinical outcomes.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td><em>Accuracy of screening instruments (for adults)</em>: Good internal/external validity and coherence <em>Effectiveness of treatments</em>: Good internal/external validity and coherence for pharmacotherapy. Fair internal validity and good external validity and coherence for psychotherapy. <em>Effect of screening on diagnosis, care, and outcomes</em>: Good internal/external validity and fair-poor coherence.</td>
<td>Screening improved rate of detection by 10-47%. Treatment resulted in 10-30% ARR for main outcome. No evidence for adverse effects. Effects of screening on outcomes were mixed. 2 found large improvements. 2 found moderate (9%) recovery. 4 found small or no improvements.</td>
</tr>
</tbody>
</table>

---

49
<table>
<thead>
<tr>
<th>Alcohol misuse</th>
<th>Overall assessment of quality of evidence</th>
<th>Evaluation of magnitude of net benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>KQ1</em>: Fair overall evidence</td>
<td><em>KQ1</em>: Mixed results: in 2 of 4 studies examining problem scores, those in all groups improved with no differences between intervention and control groups at follow-up. Other 2 studies showed no changes from baseline to follow-up within or between groups. With other outcomes, studies generally found no improvement or similar improvements in interventions and controls over the duration of the trials. Of the 5 trials that examined health care utilization only 1 found reduced self-reported hospital days at 12 months. In one good-quality trial there was a trend toward reduced all-cause mortality in interventions compared to controls (3 vs. 7 deaths; p&gt;0.1). In a 2nd study a brief single-contact intervention had no long-term effects on morbidity, mortality, or alcohol consumption at 10-year follow-up. A 3rd study men had significantly lower total mortality rate (24/100,000 person-years) than non-invited controls (30/100,000 person-years) (p&lt;0.02) and had significantly reduced alcohol-related mortality after 3 to 21 years.</td>
</tr>
<tr>
<td></td>
<td><em>KQ2</em>: Fair overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ3</em>: Poor overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ4</em>: Good overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ4a</em>: Poor-to-fair overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ5</em>: Poor-to-fair overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ6</em>: Poor overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ7</em>: Fair overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ8</em>: Good overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ9</em>: Poor overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ10</em>: Fair overall evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ11</em>: Mixed results: in 2 of 4 studies examining problem scores, those in all groups improved with no differences between intervention and control groups at follow-up. Other 2 studies showed no changes from baseline to follow-up within or between groups. With other outcomes, studies generally found no improvement or similar improvements in interventions and controls over the duration of the trials. Of the 5 trials that examined health care utilization only 1 found reduced self-reported hospital days at 12 months. In one good-quality trial there was a trend toward reduced all-cause mortality in interventions compared to controls (3 vs. 7 deaths; p&gt;0.1). In a 2nd study a brief single-contact intervention had no long-term effects on morbidity, mortality, or alcohol consumption at 10-year follow-up. A 3rd study men had significantly lower total mortality rate (24/100,000 person-years) than non-invited controls (30/100,000 person-years) (p&lt;0.02) and had significantly reduced alcohol-related mortality after 3 to 21 years.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ2</em>: Various methods were used: screening (identifying patients with probable risky/harmful alcohol use) and screening-related assessment (confirming screening results and distinguishing patients suitable for brief interventions from those needing specialty care referral). Screening usually involved self-administered questionnaires or brief interviews to assess average quantity or frequency and binge use.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ3</em>: No studies found that addressed adverse effects associated with alcohol use screening or assessment or with behavioral counseling interventions for alcohol use.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>KQ4</em>: Screening and interventions integrated into primary care reduce alcohol consumption and improve use patterns: 4 good-quality trials reported that weekly drinking was reduced 13-34% compared to</td>
<td></td>
</tr>
</tbody>
</table>
controls, resulting in 2.9-8.7 fewer mean drinks per week at follow-up in intervention group compared to control group. 1 good-quality trial did not find significant change in average use. All 5 good-quality trials found significant effects on recommended or safe alcohol use, resulting in 10-19% more intervention participants than controls reporting recommended or safe drinking patterns. 2 of 4 good-quality trials reported significantly reduced binge drinking.

KQ4a: Not enough information available about the elements that make an intervention effective.

KQ5: "Null findings may reflect lack of an effect, lack of appropriate measures, and/or reduced power for secondary analyses."

KQ6: No studies address this question.

KQ7: In all 12 trials, additional staff or systems support were required to provide screening and assessment services and sometimes intervention support. In nearly every study, research staff conducted the screening and assessment outside the routine care encounter. Most of these processes took more than 30 minutes. Provider training sessions ranged from 15 min to 2.5 hours. Only 3 studies used incentives for participating providers or patients. Besides usual care physicians, studies also used research staff or non-physician health care staff to deliver some or all of the intervention. Research staff often performed important support functions like prompting the provider and supplying intervention materials to the chart.