Routine Screening for Chronic Human Immunodeficiency Virus Infection: Why Don’t the Guidelines Agree?

Roger Chou*

*Correspondence to Dr. Roger Chou, Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, 3181 SW Sam Jackson Park Road, Mail Code: BICC, Portland, OR 97239 (e-mail: chour@ohsu.edu).

Accepted for publication January 6, 2011.

Infection with human immunodeficiency virus remains a major public health problem in the United States. Prominent guidelines from the U.S. Preventive Services Task Force and the Centers for Disease Control and Prevention differ in their recommendations on whether and how to screen adults and adolescents not known to be at higher risk. These discrepancies have led to controversy and debate as well as confusion among clinicians. This article reviews principles of screening, explains specific issues related to screening for human immunodeficiency virus, reviews the discrepancies between the US Preventive Services Task Force and the Centers for Disease Control and Prevention guidelines and the methods used in each guideline, and describes potential reasons for the discrepancies. The case of screening for human immunodeficiency virus illustrates how discrepancies between guidelines may be related to different guideline development methods as well as the different perspectives of the guideline development groups.

Infection with human immunodeficiency virus (HIV) remains a major public health problem in the United States. In 2006, 1.1 million US persons 13 years of age or older were estimated to be living with HIV infection (1), with an annual incidence of about 56,000 new infections (2). The life expectancy of HIV-infected persons is now approaching that of uninfected persons, primarily because of the effectiveness and use of highly active antiretroviral therapy (HAART) (3, 4). Nonetheless, HIV infection remains incurable, and the vast majority of patients started on HAART require indefinite treatment, with its associated short- and long-term adverse effects (5, 6). Although less common in the HAART era, opportunistic infections still occur in patients with more advanced HIV infection (acquired immunodeficiency syndrome) and are linked with serious morbidity and mortality. Chronic HIV infection is associated with an average undiscounted lifetime cost of over $500,000 and an average discounted lifetime cost of more than $300,000 (both estimates in 2004 US dollars) (7).

About 21% of patients with HIV infection are thought to be unaware of their status (1). Screening could identify such persons before they develop symptoms or complications associated with chronic immune deficiency, which typically occur only after many years of infection (8). Identification of infected persons could allow earlier initiation of HAART therapy or prophylaxis for opportunistic infections, thus reducing the morbidity and mortality associated with advanced disease. Importantly, screening could also reduce the incidence of new HIV infections. Persons unaware of their HIV-positive status may unknowingly put uninfected individuals at risk because of risky sexual or intravenous drug use behaviors. Knowledge of positive HIV serostatus could reduce such behaviors, thus resulting in lower transmission rates (9). Even for persons who continue to engage in risky behaviors, earlier initiation of HAART therapy could decrease risk of transmission by reducing viral load—which predicts the degree of infectiousness—to lower or undetectable levels (10).

Clinicians look to screening guidelines to help guide their practice, and they receive recommendations from many different organizations, including professional societies, governmental agencies, health care payers, and others. In some
cases, screening recommendations from different groups largely agree with one another, and clinical decisions are relatively straightforward. For example, there is little debate regarding screening for high blood pressure. In other cases, screening recommendations disagree or conflict with one another in important ways. Recent examples include mammography screening of women 40–50 years of age (11), prostate cancer screening (12, 13), colon cancer screening (14), and screening for lipid disorders in children (15, 16). Such discrepancies are a source of confusion and frustration to clinicians, who would like to have consistent guidance about what to do.

In the case of HIV screening, prominent guidelines from the U.S. Preventive Services Task Force (USPSTF) (17) and the Centers for Disease Control and Prevention (CDC) (18) include some discordant recommendations, which have led to controversy and debate. The main area of discrepancy concerns whether to routinely perform HIV screening in persons not known to be at high risk and, to a lesser degree, how to implement screening in clinical practice.

The discrepancies between the USPSTF and CDC HIV screening guidelines provide an opportunity to explore how groups can examine essentially the same evidence about a screening intervention, yet reach different conclusions about what should be done. Understanding the reasons for the discrepancies can provide insight into how different methods for evaluating evidence and generating recommendations as well as different perspectives can affect the guideline development process.

PRINCIPLES OF SCREENING

If a patient asks a medical practitioner for help, the doctor does the best he can. He is not responsible for defects in medical knowledge. If, however, the practitioner initiates screening procedures he is in a very different situation. He should, in our view, have conclusive evidence that screening can alter the natural history of the disease in a significant proportion of those screened.


Screening has been defined as “the systematic application of a test or inquiry, to identify individuals at sufficient risk of a specific disorder to warrant further investigation or direct preventive action, amongst people who have not sought medical attention on account of symptoms of that disorder” (20, p. 12). The purpose of screening is to help identify a condition before it would typically be diagnosed based on the presence of clinical signs or symptoms.

Screening and early detection intuitively seem desirable (21), which has led to a tendency to assume that screening interventions are beneficial, or that doing something to detect a disease is inherently better than doing nothing. Unfortunately, knowing whether it makes sense to offer a screening intervention is not so straightforward. In 1975, Frame and Carlson (22) presented criteria that should be met for a disease to merit consideration for screening (Table 1). Similar criteria have been utilized by numerous groups evaluating screening interventions, including the World Health Organization (23), the Canadian Task Force on Preventive Health Care (24), the USPSTF (25), the UK National Screening Committee (26), and the Australian Government Department of Health and Aging (27). The disease should have an important societal burden based on its prevalence and effects on quality of life or mortality, there should be an asymptomatic phase in which early treatment is more effective than treatment delayed until the development of symptoms, and there should be acceptable screening tests as well as treatments. Ultimately, screening interventions that meet all these criteria must also be shown to be associated with benefits that clearly outweigh the harms.

A critical difference between screening and most other medical interventions is that it is performed on people not seeking help for a symptom. This difference has important ethical implications because every adverse outcome of screening is iatrogenic and could have been avoided. These include the effects of labeling a patient with a disease, the anxiety associated with the screening test, the harms of screening and subsequent interventions, and the burdens and costs associated with screening. Because the populations affected by screening are often large and the conditions to be detected usually uncommon, harms incurred by even a small percentage of those screened can affect a large number of people (25). In addition, since screening is performed on persons who do not have recognized symptoms or signs of the target condition, there is the real potential that the screening intervention and subsequent treatments could make them worse. One example is breast self-examination for breast cancer screening, which was associated with no improvement in breast-cancer-related mortality compared with no breast self-examination in randomized trials but increased the number of biopsies by 50%–100% (28). Therefore, the potential for doing greater harm than good must be taken seriously, and convincing evidence of benefits should be required before a screening intervention is recommended.

SPECIFIC ISSUES RELATED TO HIV SCREENING

Current USPSTF (17, 29) and CDC (18) guidelines largely agree with regard to many aspects related to HIV

---

Table 1. Six Criteria That Should Be Met to Justify Screening

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>The disease should have a significant effect on quantity or quality of life.</td>
</tr>
<tr>
<td>2</td>
<td>The incidence of the condition should be sufficient to justify the cost of screening.</td>
</tr>
<tr>
<td>3</td>
<td>The disease should have an asymptomatic period during which detection and treatment significantly reduce morbidity and/or mortality.</td>
</tr>
<tr>
<td>4</td>
<td>Tests should be available at reasonable cost to detect the condition during the asymptomatic period.</td>
</tr>
<tr>
<td>5</td>
<td>Acceptable methods of treatment should be available.</td>
</tr>
<tr>
<td>6</td>
<td>Treatment during the asymptomatic phase should yield a therapeutic result superior to that obtained by delaying treatment until symptoms appear.</td>
</tr>
</tbody>
</table>

* Source: Frame and Carlson (22).
screening. There is no debate about whether HIV infection meets criteria to warrant consideration for screening based on the burden of disease (prevalence of about 0.4% in the US population, with about 21% undiagnosed, and substantial morbidity and mortality) (1), the availability of accurate and acceptable screening tests (typically HIV antibody testing with confirmatory Western blot) (30–33), and the availability of effective and acceptable treatments (HAART and opportunistic infection prophylaxis) that can be started during a long, relatively asymptomatic latent period (34, 35).

Chronic HIV infection differs in some important aspects from screening for noninfectious cancers, cardiovascular disease, and other nongenetic conditions, where benefits and harms are largely limited to the individual being screened. Because HIV infection is an infectious disease, screening could also have an important impact on persons who are not screened themselves but who might be exposed by HIV-infected persons to risky behaviors. Therefore, evaluations of HIV screening must also consider the potential effects on risk of transmission. A challenge is that clinical studies would require extremely large samples to have enough power to adequately assess effects of HIV screening on transmission rates in low- or average-prevalence populations, given the low incidence in such settings.

HIV screening also differs from other screening interventions because of the persistent, strong stigma and discrimination associated with HIV infection (36). Historically, being diagnosed with HIV infection—or even agreeing to testing—has been associated with important negative social consequences. These factors led in part to HIV testing policies that required special informed consent and prevention counseling prior to testing (37, 38). The inconvenience of these requirements as well as the negative stigma associated with testing for HIV infection are potential barriers to screening. Furthermore, because the risk behaviors associated with HIV infection may also be viewed negatively, people can be reluctant to disclose them. Studies indicate that a relatively high proportion of patients with HIV infection reported no risk factors prior to diagnosis, even when asked about them (39). The difficulty in obtaining accurate information about risk factors poses challenges to traditional risk-based screening strategies, which are predicated on the ability to obtain reliable information in order to appropriately identify higher-risk persons for testing (40).

### DISCREPANCIES BETWEEN THE USPSTF AND CDC RECOMMENDATIONS

The USPSTF released updated recommendations on HIV screening in 2005 (Table 2) (17). Recommended was screening of all patients at higher risk of HIV infection (including those reporting risk factors and those in high-prevalence (>1%) or high-risk settings) and all pregnant women. These were grade A, or strong, recommendations, based on good evidence of substantial net benefits (that is, benefits relative to harms; refer to Tables 3 and 4 for USPSTF recommendation grades) (25). The USPSTF made no recommendation for or against routinely screening for HIV in nonpregnant adults and adolescents not known to be at higher risk of HIV infection, based on small estimated net benefits. This was a grade C recommendation.

The CDC released its updated recommendations on HIV screening in 2006 (Table 2) (18). It recommended routinely screening all patients aged 13–64 years regardless of assessed risk, unless the prevalence of undiagnosed HIV infection was documented to be less than 0.1% (the CDC did not grade its recommendations). The CDC also recommended opt-out HIV screening without requiring special consent or HIV prevention counseling, a departure from previous CDC guidelines. Opt-out screening means that patients are told that testing is routinely performed (41). Testing is not performed only if patients specifically refuse it. This is in contrast to traditional opt-in screening, in which the health care provider specifically asks the patient if he or she wants to be tested before ordering it. As noted earlier, many US jurisdictions required special opt-in consent procedures prior to testing for HIV infection, and guidelines also called for pretest HIV prevention counseling (37). The 2006 CDC guideline also recommended at least annual testing of high-risk persons. The USPSTF made no recommendations regarding frequency of screening or use of opt-out testing of nonpregnant persons, which were addressed as clinical considerations (17).

Following the release of the CDC guideline in 2006, the USPSTF commissioned an early update of the 2005 evidence review to evaluate whether new studies cited in the CDC guideline might change its recommendations (42). After reviewing the new evidence, the USPSTF confirmed the prior “C” recommendation for screening of nonpregnant

### Table 2. Summary of Human Immunodeficiency Virus Screening Guidelines

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High-risk persons and persons in high-risk settings</td>
<td>Screen (grade A recommendation)</td>
<td>Screen</td>
</tr>
<tr>
<td>Pregnant women</td>
<td>Screen (grade A recommendation)</td>
<td>Screen</td>
</tr>
<tr>
<td>Low- or average-risk persons</td>
<td>No recommendation for or against screening (grade C recommendation)</td>
<td>Opt-out screening without special informed consent; no pretest prevention counseling required</td>
</tr>
<tr>
<td>Screening implementation</td>
<td>No recommendation</td>
<td>Screen persons at high risk at least annually</td>
</tr>
<tr>
<td>Repeat screening</td>
<td>No recommendation</td>
<td></td>
</tr>
</tbody>
</table>
adults and adolescents not known to be at higher risk of HIV infection (43).

GUIDELINE DEVELOPMENT METHODS

Clinical practice guidelines are “systematically developed statements to assist practitioner and patient decisions about appropriate health care for specific clinical circumstances” (44, p. 38). Purposes of guidelines are to encourage the use of practices shown to be effective, to reduce the use of unproven or ineffective therapies, and to support the introduction of new knowledge into clinical practice (45). Guidelines are more likely to avoid biased or incorrect recommendations if they are based on a comprehensive, unbiased review of the evidence (46). It is also important for guideline developers to be transparent about how they went from evidence to recommendations. Groups can end up with similar recommendations even when the methods for developing them are quite different. Conversely, groups can publish discrepant recommendations even when the methods for developing them are similar.

These situations can occur because guideline developers vary in their interpretations of what constitutes convincing evidence. Published standards provide general guidance on how to determine whether evidence is convincing (47). However, even though such guidance is useful, it is difficult to define objective evidence thresholds. Some guideline developers may think that 2 well-conducted, moderately sized randomized trials are enough to estimate net benefit; others may think that additional research is needed. Therefore, guideline developers should be as transparent as possible about describing the evidence used to support their recommendations.

In addition, judgments regarding net benefits often require subjective interpretation. The evidence can tell us the likely benefits and the likely harms, burdens, and costs, but it does not directly tell us how to weigh all of these factors. When the evidence is clear that there are substantial benefits with minor harms, there is usually little room for subjectivity or disagreement. However, in the more common situations when the benefits are smaller, the harms more substantial, or the evidence less robust, subjective judgments are required, and groups are more likely to disagree. For example, groups making decisions on breast cancer screening of women 40–50 years of age must weigh a small reduction in risk of breast cancer–related death against a high number of false-positive mammograms and associated biopsies (28). For some women, even a small reduction in risk of breast cancer–related death is worth a large number of false-positive mammograms. Others would view these as significant trade-offs, with the harms largely offsetting the benefits.

It is critical for clinicians to know how guideline developers weighed the results for these very different outcomes to determine whether the conclusions are consistent with the

---

**Table 3.** Recommendation Grades from the U.S. Preventive Services Task Force^\text{a,b}\n
<table>
<thead>
<tr>
<th>Recommendation Grade</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>The USPSTF strongly recommends that clinicians routinely provide the service to eligible patients. (The USPSTF found good evidence that the service improves important health outcomes and concludes that benefits substantially outweigh harms.)</td>
</tr>
<tr>
<td>B</td>
<td>The USPSTF recommends that clinicians routinely provide the service to eligible patients. (The USPSTF found at least fair evidence that the service improves important health outcomes and concludes that benefits outweigh harms.)</td>
</tr>
<tr>
<td>C</td>
<td>The USPSTF makes no recommendation for or against routine provision of the service. (The USPSTF found at least fair evidence that the service can improve health outcomes but concludes that the balance of the benefits and harms is too close to justify a general recommendation.)</td>
</tr>
<tr>
<td>D</td>
<td>The USPSTF recommends against routinely providing the service to asymptomatic patients. (The USPSTF found at least fair evidence that the service is ineffective or that harms outweigh benefits.)</td>
</tr>
<tr>
<td>I</td>
<td>The USPSTF concludes that the evidence is insufficient to recommend for or against routinely providing the service. (Evidence that the service is effective is lacking, of poor quality, or conflicting and the balance of benefits and harms cannot be determined.)</td>
</tr>
</tbody>
</table>

Abbreviation: USPSTF, U.S. Preventive Services Task Force.

\(\text{a Source: Harris et al. (25).}\)

\(\text{b The recommendation language was revised after the human immunodeficiency virus screening guideline was published in May 2007 (85).}\)

---

**Table 4.** U.S. Preventive Services Task Force Recommendation Grid\(^{a,b}\)

<table>
<thead>
<tr>
<th>Quality of Evidence</th>
<th>Substantial</th>
<th>Moderate</th>
<th>Small</th>
<th>Zero/Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Fair</td>
<td>B</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>Poor</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(\text{a Source: Harris et al., 2001 (25).}\)

\(\text{b The recommendation grid was revised after the human immunodeficiency virus screening guideline was published in May 2007, with “quality of evidence” replaced by “level of certainty (high, moderate, or low)” (77).}\)

\(\text{c Refer to Table 3 for a description of the letter grades.}\)
values that they and their patients would place on them. Because judgments about the evidence can be influenced by financial, intellectual, or other conflicts of interest, it is also important for guideline developers to transparently describe any such issues and how they were managed (48).

**USPSTF methods**

The USPSTF is a panel of experts in primary care, prevention, and clinical epidemiology that issues recommendations for clinical preventive services (49). It receives support from the Agency for Healthcare Research and Quality, an agency of the US Department of Health and Human Services, but is an independent, nongovernmental body. The USPSTF process for developing guidelines is described in several publications (25, 50–53) and is regularly updated on the Agency for Healthcare Research and Quality Web site (54). A core principle of the USPSTF process is that all guidelines are built on systematic reviews that are based on a clearly defined scope and set of key questions, and the process uses explicit and transparent methods for identifying, selecting, and synthesizing the evidence (25). The systematic reviews are based on analytic frameworks that illustrate the key questions and the thought process used to evaluate the screening intervention (Figure 1).

As with many other screening interventions, direct evidence is not available showing that persons who undergo HIV screening experience better clinical outcomes compared with those who do not undergo screening (29). In these situations, the USPSTF analytic frameworks can be very useful for understanding the indirect chain of evidence necessary to determine whether a screening intervention is beneficial (25). For instance, in the case of HIV screening, it would not be sufficient to simply show that HIV screening accurately diagnoses HIV infection. It is also necessary to understand how many patients with screen-detected HIV infection would meet criteria for antiretroviral therapy or opportunistic infection prophylaxis, the degree to which treatment of those patients would result in improved clinical outcomes (reduced premature death, disability, or spread of disease), and the magnitude of harms associated with testing (such as false-positives, labeling, and psychological harms) and treatments. For each link in the indirect chain of evidence, the systematic review determines the quality of the evidence, based on the number, type, quality, and size of studies; the presence of important inconsistency between studies; and other factors (25). Gaps in any step of the chain of evidence decrease the certainty of estimates of the overall net benefit of a screening intervention, with important gaps in key areas rendering reliable estimates impossible.

In general, the USPSTF requires proof of improvements in patient-centered clinical outcomes rather than just intermediate outcomes in order to recommend a screening intervention. Intermediate outcomes are associated with

---

Figure 1. Analytic framework for screening for human immunodeficiency virus (HIV) in asymptomatic adolescents and adults (excluding pregnant women, dialysis patients, and transplant patients). Key questions (correspond to the circled numbers): 1) Does screening for HIV infection in asymptomatic adolescents and adults reduce premature death and disability or spread of disease? 2) Can clinical or demographic characteristics (including specific settings) identify subgroups of asymptomatic adolescents and adults at increased risk of HIV compared with the general population? 3) What are the test characteristics of HIV antibody test strategies? 4) What are the harms (including labeling and anxiety) associated with screening? Is screening acceptable to patients? 5) How many newly diagnosed HIV-positive patients meet criteria for antiretroviral treatment or prophylaxis for opportunistic infections? How many patients who meet criteria for interventions receive interventions? 6) What are the harms associated with the work-up for HIV infection? 7a) How effective are interventions (antiretroviral treatment, counseling on risky behaviors, immunizations, routine monitoring and follow-up, more frequent Papanicolaou testing, or prophylaxis for opportunistic infections) in improving clinical outcomes (mortality, functional status, quality of life, symptoms, opportunistic infections, or transmission rates)? 7b) In asymptomatic patients with HIV infection, does immediate antiretroviral treatment result in improvements in clinical outcomes compared with delayed treatment until symptomatic? 7c) How well do interventions reduce the rate of viremia, improve CD4 counts, or reduce risky behaviors? 8) What are the harms associated with antiretroviral therapy? 9) Have improvements in intermediate outcomes (CD4 counts, viremia, risky behaviors) been shown to reduce premature death and disability or spread of disease?
The assumptions and parameters used to develop an outcomes table can have an important effect on its estimates. A potential shortcoming of the USPSTF outcomes table on HIV screening is that most of its parameters and assumptions probably resulted in conservative estimates regarding potential benefits of screening (Table 6). For example, the outcomes table limited the time horizon to 3 years (based on the duration of trials of HAART), excluded interventions other than HAART (such as opportunistic infection prophylaxis and immunizations because of relatively small effects compared with those of HAART), could not estimate effects of screening on transmission risk, only assumed HAART for patients with CD4 counts of less than 0.200 cells/L (based on the guidelines and trials of HAART available at the time), and did not incorporate potential effects of screening on earlier diagnosis (no evidence). On the other hand, the assumption of 0.3% prevalence probably resulted in overestimation of the yield of screening, since the prevalence of undiagnosed HIV infection in the general population is closer to 0.1% (based on an overall prevalence of HIV infection of about 0.4%) (29).

Once the systematic review has been completed, the USPSTF considers its findings and develops and votes on the actual recommendations based on a formal set of rules (25). The rules for grading recommendations (Table 3) are based on both the estimated net benefit of the screening intervention and the certainty in the estimates (based on the quality of the supporting evidence) (Table 4). The recommendation grade can help clinicians and patients prioritize screening interventions, given the limited resources and time available to address preventive health. The USPSTF generally recommends routine implementation of grade A and B recommendations, based on the principle that proven interventions with larger benefits should receive the highest priority. For routine HIV screening, given the relatively large number of patients the USPSTF estimated would need to be screened to prevent one case of clinical progression or death in low-prevalence populations, the net benefit was

**Table 5. Outcomes of Counseling and One-Time Screening for HIV Infection After 3 Years in a Hypothetical Cohort of Asymptomatic Adolescents and Adults**

<table>
<thead>
<tr>
<th>Screening Results</th>
<th>Prevalence 0.3%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persons screened, no.</td>
<td>10,000</td>
</tr>
<tr>
<td>Persons identified as HIV-positive, no.</td>
<td>30</td>
</tr>
<tr>
<td>Patients receiving test results, no.</td>
<td>24–28</td>
</tr>
<tr>
<td>Partners identified as HIV-positive, no.</td>
<td>2–6</td>
</tr>
<tr>
<td>Total HIV-positive patients identified, no.</td>
<td>26–34</td>
</tr>
<tr>
<td>Patients with CD4 cell count &lt;0.200 × 10^9 cells/L, no.</td>
<td>3–15</td>
</tr>
<tr>
<td>Cases of clinical progression or death prevented over 3 years with highly active antiretroviral therapy, no.</td>
<td>0.7–8.2</td>
</tr>
<tr>
<td>No. needed to screen to prevent 1 clinical progression or death over 3 years</td>
<td>1,210–13,800</td>
</tr>
<tr>
<td>No. needed to screen to prevent 1 horizontal transmission over 3 years</td>
<td>Unable to calculate</td>
</tr>
<tr>
<td>No. needed to screen to cause 1 cardiovascular or cerebrovascular event over 3 years</td>
<td>16,900–1,580,500</td>
</tr>
</tbody>
</table>

Abbreviation: HIV, human immunodeficiency virus.

* Source: Chou et al., 2005 (29, 87).
In the case of HIV screening, no USPSTF members reported significant potential conflicts. In some cases, USPSTF members with significant conflicts are not permitted to participate in discussions or voting, whether they were identified and considered at all. Failure to use explicit and transparent methods does not necessarily mean the recommendations are wrong. However, it does make it difficult or impossible to determine how much certainty to have in them and how strongly to prioritize them.

The American College of Physicians (59) independently reviewed the USPSTF and the CDC HIV screening guidelines by using the Appraisal of Guidelines Research and Evaluation (AGREE) instrument (57). In the 7 domains related to rigor of development (each scored on a 1–4 scale), the USPSTF screening guideline received 25 points (maximum, 28) compared with 18 for the CDC guideline (59). The main areas in which the USPSTF guideline scored substantially higher than the CDC guideline were in use of systematic methods to search for evidence (mean score, 3.5 vs. 2.3), clear description of criteria for selecting evidence (mean score, 4.0 vs. 1.5), clear description of methods used to formulate the recommendations (mean score, 4.0 vs. 2.5), and explicit link between the recommendations and the supporting evidence (mean score, 4.0 vs. 2.0).

### CDC methods

The CDC is one of the major operating components of the US Department of Health and Human Services. It has a broad mandate to protect the health of Americans through health promotion; prevention of disease, injury, and disability; and preparedness for new health threats, often operating from a public health perspective (56).

The CDC did not describe the methods used to develop its HIV screening guidelines (18). Although supporting evidence for the CDC recommendations was described, no formal methods for identifying, selecting, grading, and synthesizing the evidence were provided; no process for reporting and managing conflicts of interest was described; the guideline did not estimate the magnitude of net benefit; and the recommendations were not graded. Such methods fall short of published standards for guideline development (57, 58) because it is difficult to understand how the CDC conceptualized the important issues related to HIV screening, whether all of the evidence was identified, how the evidence was synthesized, how the expected benefits and harms were weighed, whether conflicts of interest might have affected those judgments, and how important research gaps might have influenced the recommendations—or whether they were identified and considered at all. Failure

<table>
<thead>
<tr>
<th>Assumption</th>
<th>Expected Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited time horizon to 3 years</td>
<td>Underestimates benefits of screening</td>
</tr>
<tr>
<td>Excluded interventions other than HAART</td>
<td>Underestimates benefits of screening</td>
</tr>
<tr>
<td>Assumed HAART for only those patients with CD4 cell count &lt; 0.200 ( \times 10^3 ) cells/L</td>
<td>Underestimates benefits of screening</td>
</tr>
<tr>
<td>Unable to estimate effects of screening on transmission risk</td>
<td>Underestimates benefits of screening</td>
</tr>
<tr>
<td>Unable to estimate effects of screening on earlier diagnosis</td>
<td>Probably underestimates benefits of screening</td>
</tr>
<tr>
<td>Prevalence of undiagnosed infection 0.3%</td>
<td>Overestimates benefits of screening (prevalence of undiagnosed infection in the general population 0.1%)</td>
</tr>
<tr>
<td>Included serious cardiovascular harms only</td>
<td>Underestimates harms of screening</td>
</tr>
</tbody>
</table>

Abbreviation: HAART, highly active antiretroviral therapy.

* Source: Chou et al., 2005 (29, 86).
the modest decrease in risky behaviors in the small number of persons testing HIV-positive would reduce the overall transmission risk, since any beneficial effects could be offset by increased risky behaviors in the much larger proportion of persons testing HIV-negative. These uncertainties were not mentioned in the CDC guideline.

Both the USPSTF and the CDC considered evidence on effects of HIV screening on reducing risky behaviors. Each group reviewed a meta-analysis finding that patients aware of their HIV status reduced the rate of unprotected intercourse with serodiscordant partners by 68% compared with those unaware of their status (64).

However, the 2 groups differed substantially in how they assessed the quality of the meta-analysis. The USPSTF review rated the meta-analysis of fair quality because it did not rank the quality of the included studies or report reasons for excluding studies (42). Upon examination of the individual studies included in the meta-analysis, the USPSTF review found that all had important methodological shortcomings, such as high rates of loss to follow-up, no clear inception cohort, or unclear blinding status of investigators to the testing status of subjects (65–67). The meta-analysis also included unpublished data from 4 studies, and it provided no detail about their methods to enable judgments about their quality (64). Other factors also complicated interpretation of the results. Statistical heterogeneity was not reported for the pooled estimate, and all of the studies included in the meta-analysis were conducted at a time when extensive counseling prior to HIV testing was recommended. It is therefore difficult to determine to what degree the reductions in risky behaviors were related to the screening test itself or to the counseling interventions. In addition, reductions in risky behaviors were all based on patient self-reports, which are difficult to verify. Because reductions in risky behaviors are a socially desirable response, patients might be reluctant to report a failure to change behaviors or overstate any reductions (68). Finally, the USPSTF noted that surveys found high rates of continued risky behaviors in HIV-infected persons (69), and epidemiologic studies showed that rates of risky behaviors and sexually transmitted disease were on the rise among certain highly tested populations of men who have sex with men (70, 71). On the basis of all of these findings, the USPSTF concluded that it could not estimate the effect of reductions in risky behaviors on transmission rates.

The CDC accepted the estimates from the meta-analysis (64) described above on decreased risky behaviors following screening. Because the CDC did not assess study quality, it is not possible to know whether the methodological shortcomings in the meta-analysis or the individual studies included might have affected interpretation of the estimates. The CDC also accepted results of a modeling study (from the same author of the meta-analysis) that estimated a 31% decline in new sexually transmitted HIV infections per year (from 32,000 to 22,150) if all HIV-positive patients unaware of their status became aware of it (9). The USPSTF concluded that even if the estimated reduction in risky behaviors was accurate, the reliability of the model was questionable because the formula used did not capture a number of factors (such as the type of risky behaviors, number of risky behavior episodes, number of sexual partners, effects of viral load, use of antiretroviral therapy, presence of other sexually transmitted diseases, CD4 count, and time since diagnosis) (72) that affect the likelihood of HIV transmission.

The model also assumed that all patients who are HIV-positive would be tested and would reduce risky behaviors to the same degree. However, patients who are at highest risk of transmitting infection may not be equally likely to be tested or to reduce risky behaviors compared with those at lower risk of transmitting infection. The model did not consider the potential effects of testing in those with negative results, who, in some studies, had higher rates of subsequent sexually transmitted diseases (63).

**Interpretation of cost-effectiveness and yield of screening**

Both the CDC and the USPSTF reviewed 2 cost-effectiveness analyses (73, 74). In the first analysis, the incremental cost-effectiveness of one-time screening versus no screening was less than $50,000 per quality-adjusted life-year (QALY) when the HIV prevalence was 0.5% (the prevalence of undiagnosed HIV infection in the United States is about 0.1%), excluding potential transmission benefits (74). After potential beneficial effects on secondary transmission were incorporated, the incremental cost-effectiveness remained less than $50,000 per QALY at very low HIV prevalence (0.05%). However, cost-effectiveness in low-prevalence settings was sensitive to estimates of beneficial effects of screening on transmission. The other study (73), which did not directly incorporate secondary transmission benefits into its model, estimated an incremental cost-effectiveness of one-time screening compared with no screening of $113,000 per QALY when the prevalence of undiagnosed HIV infection was 0.1%. This study did not directly incorporate secondary transmission benefits, but it estimated that screening 100,000 persons in the general population would prevent 10 of the 780–1,060 expected secondary transmissions.

The CDC and the USPSTF differed in how they interpreted the cost-effectiveness analyses. The CDC stated that both studies demonstrated cost-effectiveness of screening in low-prevalence settings. By contrast, the USPSTF viewed the results as somewhat inconsistent, since the estimate in the second study was more than twice as high as that in the first study and over the $100,000 per QALY threshold often (although somewhat arbitrarily) used to distinguish cost-effective from non-cost-effective interventions (75). It is unclear how the CDC interpreted the different cost-effectiveness estimates. In the study that found screening cost-effective in low-prevalence populations (74), results were sensitive to estimates on effects on transmission risk, an area previously characterized by the USPSTF as marked by substantial uncertainty. Finally, although the USPSTF found that both studies met criteria for good-quality cost-effectiveness analyses, neither directly answered the question of whether routine universal screening is more cost-effective than targeted screening in low-prevalence settings. Rather, each of the studies compared screening with no screening. Thus, the analyses do not necessarily reflect the comparison relevant to clinical
practice, where targeted screening has long been recommended even in low-prevalence populations.

**Potential effects of screening on earlier diagnosis**

Many patients with HIV infection meet criteria for acquired immunodeficiency syndrome at the time of diagnosis or soon afterward (76). Based on the usual course of chronic HIV infection, this means that patients have been infected—and therefore have posed an unknowing transmission risk—for many years before diagnosis. In addition, late initiation of HAART is associated with increased morbidity and mortality (77–80). The CDC guideline (18) suggests that universal screening could result in earlier diagnosis but does not cite any studies demonstrating stage shift as a result of screening. On the basis of the lack of evidence, the USPSTF found that it was unable to estimate effects of screening on stage at diagnosis.

**Extrapolation of evidence from high-risk and prenatal settings**

The acceptability and uptake of opt-out testing has been evaluated primarily in prenatal settings (41). The CDC guideline appeared to extrapolate data showing high acceptability of opt-out testing from prenatal settings to nonpregnant settings (18). The USPSTF methods allow for extrapolation of data from one population or setting to another (25). In this case, however, the USPSTF made no recommendation about use of opt-out testing in nonpregnant persons. Because pregnant women concerned about potential risks to a fetus might make testing decisions that are different than those of nonpregnant persons without such considerations because of the high value placed on preventing neonatal transmission, the USPSTF felt that it was not clear how well evidence could be extrapolated.

**Approach to potential barriers to testing**

Potential barriers to HIV screening include the stigma associated with HIV disease and testing for it, the unwillingness of patients to disclose risky behaviors, the lack of local prevalence data to guide prevalence-based testing, and clinician burdens associated with traditional requirements for special informed consent procedures and prevention counseling prior to testing (18). The CDC recommendations to routinize and streamline screening by removing specialized informed consent and prevention counseling prior to screening could overcome such barriers, but no studies showing improvement in uptake of screening after implementation of the approaches recommended by the CDC were cited. The USPSTF guideline did not comment on these issues because of the lack of evidence.

**Frequency of testing**

The USPSTF did not make recommendations on the frequency of screening because no studies compared the effects of different screening intervals on clinical outcomes. The CDC recommended at least annual testing of high-risk persons (18), although this recommendation seems to be inconsistent with the cost-effectiveness analyses discussed earlier, which found incremental cost-effectiveness ratios of well over $100,000 per QALY for annual screening versus one-time screening even of high-risk persons (73, 74).

**CONCLUSIONS**

When the evidence is strong about a screening intervention and the benefits clearly and substantially outweigh the harms, there is usually little disagreement about what to do. When the evidence is weaker and the balance between benefits and harms is closer, disagreements are more likely, in part because subjective judgments about what constitutes sufficient evidence and how to weigh benefits relative to harms assume greater relative importance. In such cases, it is critical for guideline developers to use systematic and transparent methods so readers can determine which recommendations are based on solid evidence of clear benefits and which ones may be more highly influenced by the values and perspectives of the guideline development group.

In the case of HIV screening, the CDC guideline fell short on current standards for developing guidelines (57). This does not necessarily mean that the CDC recommendations were off base, and it is possible that the CDC would have reached the same recommendations if they had used the same process as the USPSTF did. The problem is that it is also possible that the recommendations may have differed if factors such as study quality and degree of certainty in estimates of benefits had been formally assessed. Some aspects of the CDC recommendations seem to be based on flawed evidence (reduced transmission), plausible but unproven assumptions about likely benefits (earlier diagnosis), or extrapolation of evidence from higher-risk or prenatal populations to low-risk nonpregnant populations (use of opt-out testing), with little acknowledgment or discussion of how these factors were considered during the guideline development process.

Conversely, although the USPSTF guideline was more adherent to standards for developing guidelines (59), that does not necessarily mean that its recommendations were correct. Even when methods are more systematic, explicit, and transparent, many areas still require subjective interpretation when it comes to synthesizing evidence and determining net benefits. For example, the USPSTF has traditionally emphasized the importance of clinical studies while downplaying studies that evaluate intermediate outcomes, modeling studies, and cost-effectiveness analyses. In the case of HIV screening, such studies might be the best available evidence, since studies directly showing decreased transmission after screening are very difficult to carry out. The focus on reliable data from clinical studies also resulted in an outcomes table based on mostly conservative parameters and assumptions, which could have resulted in significant underestimates of the benefits of HIV screening (Table 6).

Because of the inherent gray areas in the guideline development process, understanding the perspectives of the
guideline developers is especially important when there are discrepancies. The CDC approached the issue of HIV screening largely from the perspective that the unchanged incidence of HIV infections was an urgent public health problem, that standard targeted screening approaches had failed, and that significant changes were required to make screening more effective (18). A 2003 CDC initiative described a key strategic goal of making HIV testing a routine part of medical care on the same voluntary basis as other diagnostic and screening tests (81). Guideline development groups that view the clinical situation as dire and have expressed a goal to revamp screening approaches may be more likely to make plausible assumptions about potential benefits based on weaker or nontraditional types of evidence. In addition, because authors of the CDC guideline have conducted research on HIV infection and were authors of some of the studies cited in the guideline, it is possible that intellectual conflicts of interest might have influenced judgments about the strength of the evidence. Recently, efforts to limit the impact of potentially conflicted experts in the guideline development process have been proposed and implemented by some groups (48).

The USPSTF, on the other hand, approached HIV screening much like it approached any other preventive services; that is, decisions regarding any preventive service should be based on whether there is convincing evidence of clear clinical benefit. Requiring proof of benefits carries the advantage of consistency and may reduce arbitrariness. It may also result in a tendency to be more reactive than innovative, particularly if few allowances for flexibility are permitted in situations characterized by more clinical urgency, which may not come with the luxury of waiting 10 or 20 years for the needed research. On the other hand, the role of guideline development groups may not be to innovate themselves but rather to promote uptake of proven interventions (innovative or otherwise).

So what should clinicians do in the face of discrepant guidelines? It would be appropriate for those who want to prioritize preventive services based on proven benefits to adopt the USPSTF recommendation and individualize decisions about HIV screening of low-risk patients, after offering grade A and B preventive services. For clinicians who place a high value on potentially reducing HIV incidence and are more willing to accept plausible but unproven benefits, it might be reasonable to adopt the CDC recommendations. This approach was followed by the American College of Physicians, which largely endorsed the CDC recommendations despite finding more methodological shortcomings (59).

As is usually the case, new evidence may help resolve the discrepancies between HIV screening guidelines. Additional evidence regarding benefits of earlier versus later initiation of HAART has led to changes in practice that will impact estimates of benefits from screening, since more patients are now being treated at earlier stages of disease (82). Several studies have evaluated implementation of routine HIV screening in low-prevalence settings, with results perhaps less impressive than hoped. In a large study performed in a low-prevalence emergency department setting, opt-out routine testing resulted in a modest increase in the number of newly diagnosed HIV infections but was much less efficient than testing based on the presence of symptoms or risk factors (10 of 6,702, or 0.15% compared with 5 of 231, or 2.2%) (83). In addition, patients were diagnosed with advanced HIV infection with either screening strategy (mean CD4 cell count of 0.069 vs. 0.013 cells/µL). Another study showed high rates of false-positives (84%) following rapid HIV testing (prior to confirmatory testing), illustrating some of the challenges with applying even a highly accurate test in low-prevalence settings (84). As yet, no studies are known to have demonstrated reduced HIV transmission or other clinical benefits with routine screening in low-risk patients.

Future updates of the HIV screening guidelines will provide opportunities to incorporate new evidence and to reassess the methods and decisions used in previous guidelines. The USPSTF will begin its updating process in 2011 and might reconsider when and how to use modeling and cost-effectiveness studies and when it might be appropriate to consider more liberal assumptions when estimating net benefits. It is likely that the USPSTF will continue to focus on the question of who to screen and leave details regarding use of opt-out screening and provision of prevention counseling to the “clinical considerations” section of the recommendations, unless strong evidence is published on the impact of their implementation. The CDC should institute methods that meet current standards for developing guidelines, so it is clear when recommendations are supported by strong evidence and when they are based on less solid ground. Adoption of more formal methods by the CDC might result in different characterizations of the reliability of the evidence and estimates of potential benefits.

Even if the methods of the 2 groups converge, discrepancies in the recommendations may remain unresolved. As much as we would like all guidelines to be based strictly on the evidence, it is impossible to remove the human element. If this was not the case, we would not need guidelines and guideline panels at all, and we could base all of our clinical decisions on systematic reviews alone. In the case of HIV screening, potential users of guidelines should recognize that guideline development groups that view themselves as screening innovators and agents of change are likely to act on the same evidence quite differently from those who mainly view themselves as consistent, impartial interpreters of evidence.

ACKNOWLEDGMENTS

Author affiliations: Department of Medicine, Oregon Health and Science University, Portland, Oregon (Roger Chou); and Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, Oregon (Roger Chou).

Conflict of interest: No funding was received for this manuscript. The author was the lead investigator on the evidence reviews (funded by the Agency for Healthcare Research and Quality) used to inform the USPSTF recommendations. He was not a member of the USPSTF and does not.

Epidemiol Rev 2011;33:7–19
speak for or represent the views of the Agency or the USPSTF. The author is also currently a member of the American College of Physicians Clinical Guidelines Committee, although he was not at the time that it reviewed the CDC and USPSTF HIV screening recommendations.

REFERENCES


