Published in final edited form as:

Psychiatr Serv. 2013 October ; 64(10): 1018–1025. doi:10.1176/appi.ps.201200449.

When Quality Indicators Undermine Quality: Bias in a Quality Indicator of Follow-up for Alcohol Misuse

Katharine A. Bradley, MD, MPH^{1,2,3,5,6}, Laura J. Chavez, MPH^{1,6}, Gwen T. Lapham, PhD, MPH, MSW^{1,2,6}, Emily C. Williams, PhD, MPH^{1,6}, Carol Achtmeyer, MN^{1,3,4}, Anna D. Rubinsky, PhD, MS^{1,6}, Eric J. Hawkins, PhD^{1,3,7}, Richard Saitz, MD, MPH^{8,9}, and Daniel R. Kivlahan, PhD^{1,3,7}

¹Health Services Research & Development (HSR&D) Northwest Center of Excellence, Seattle, WA

²Group Health Research Institute, Seattle, WA

³Center of Excellence in Substance Abuse Treatment and Education (CESATE), Seattle, WA

⁴General Medicine Service, Veterans Affairs (VA) Puget Sound Health Care System, Seattle, WA

⁵Departments of Medicine, Seattle, WA

⁶Health Services, Seattle, WA

⁷Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA

⁸Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, and Departments of Medicine, and Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA

⁹Boston Medical Center, Boston, MA

Abstract

Background—Valid quality indicators are needed to monitor and incentivize identification and management of mental health and substance use conditions (“behavioral conditions”). Because behavioral conditions are frequently under-identified, quality indicators often evaluate the proportion of patients who screen positive for a condition who have appropriate follow-up care documented. However, these “positive-screen-based” quality indicators of follow-up for behavioral conditions could be biased by differences in the denominator due to differential screening quality (“denominator bias”) and could reward identification of fewer patients with the behavioral condition(s) of interest.

Objective—To evaluate denominator bias in the performance of Veterans Health Administration (VA) networks on a quality indicator of follow-up for alcohol misuse that used patients with positive alcohol screens as the denominator.

Methods—The performance of 21 VA networks on a positive-screen-based quality indicator of follow-up for alcohol misuse was compared to the networks' performance on a population-based quality indicator (proportion of eligible patients who had alcohol misuse identified and follow-up documented) using medical record reviews (n=219,119).

Results—Results of the two quality indicators were inconsistent. For example, two networks performed similarly on the quality indicators (64.7%, 65.4%) even though one identified and

documented follow-up for almost twice as many patients (5,411 and 2,899 per 100,000 eligible, respectively). Networks that performed better on the positive-screen-based quality indicator identified fewer patients with alcohol misuse than networks that performed better on the population-based quality indicator (mean 4.1% vs 7.4% respectively).

Conclusion—A positive-screen-based quality indicator of follow-up for alcohol misuse preferentially rewarded networks that identified fewer patients with alcohol misuse.

Keywords

quality improvement; alcohol counseling; alcohol screening

Introduction

Over a quarter of the US population suffers from mental health or substance use (“behavioral”) condition(s) (1), but these conditions often remain unrecognized and untreated.(2-5) Identifying and offering evidence-based care to patients with behavioral conditions is therefore a major quality improvement priority for US health care.(2,6)

Improving the quality of behavioral health care requires valid quality indicators to measure and incentivize identification of and evidence-based follow-up for common behavioral conditions.(2,6-10) One commonly used approach to measuring and incentivizing behavioral care is to evaluate follow-up care provided to patients who screen positive on validated questionnaires.(9,11,12) We refer to these types of quality indicators of follow-up for behavioral conditions as “positive-screen-based” quality indicators.

One theoretic limitation of positive-screen-based quality indicators is that they might preferentially reward systems that identify fewer patients through screening. Table 1 shows how bias due to variation in the sensitivity of screening programs across health systems could undermine the validity of positive-screen-based quality indicators. Three hypothetical systems (A-C) with identical patient populations and identical true prevalence rates of a behavioral condition are modeled. Compared with systems “B” and “C,” system “A” has a more sensitive screening program, resulting in a two-fold higher prevalence of positive screens (10% vs. 5%). Therefore, although systems “A” and “B” have identical performance on a positive-screen-based quality indicator (50% of patients with positive screens have appropriate follow-up), system “A” identifies and offers follow-up to twice as many patients with the condition (5000 vs. 2500). Comparison of systems “A” and “C” demonstrates how “A”, with a more sensitive screening program, could perform worse on a positive-screen-based quality indicator (50% vs. 80%) despite identifying and offering follow-up care to more patients with the condition (5,000 vs. 4,000).

One strategy to improve the validity of positive-screen-based quality indicators and avoid bias due to estimating quality indicators in different denominators (i.e. “denominator bias”) is to require use of a specific validated screening questionnaire and threshold to standardize the denominator.(13) This strategy is used by the Veterans Health Administration (VA) for alcohol misuse as well as depression and post-traumatic stress disorder (PTSD).(11,14) However, recent research has demonstrated that despite use of a uniform screening questionnaire and threshold for a positive screen, the sensitivity of alcohol screening programs may vary across VA networks,(15) likely due to differences in how screening is implemented in practice, e.g. non-verbatim interviews vs. paper.(16) Variation in the sensitivity of screening programs could undermine the validity of positive-screen-based quality indicators, but this has not been previously evaluated.

This study used a VA quality improvement dataset to determine whether variability in the prevalence of positive screens for alcohol misuse undermined the validity of a positive-screen-based quality indicator of follow-up for alcohol misuse (i.e. denominator bias). If such denominator bias existed in the VA despite high rates of screening with a uniform screening questionnaire and threshold, it would suggest that positive-screen-based quality indicators might unintentionally systematically reward health systems that identified fewer patients with alcohol misuse due to poorer quality alcohol screening programs. If this were true, positive-screen-based quality indicators for other behavioral conditions would need to be similarly evaluated.

Methods

Overview

Two quality indicators of follow-up for alcohol misuse were evaluated in a single sample of patients from each VA network. Both quality indicators were based on the same medical record reviews. The numerators of the two quality indicators were the same, but the denominators differed. The numerator was all patients in each network who had alcohol misuse identified and follow-up for alcohol misuse documented in the medical record. The denominator of one ("positive-screen-based") quality indicator included all patients who screened positive for alcohol misuse on VA's specified screen in a VA clinic. The denominator of the other ("population-based") quality indicator included all outpatients eligible for screening. First, each VA network was evaluated and its performance ranked on the two quality indicators. Second, convergent validity of the two quality indicators was assessed by calculating the difference in each network's rank on the two indicators. Third, denominator bias was evaluated by testing whether differences in rank were associated with the network prevalence of documented positive alcohol screens.

Data Sources and Sample

The VA Office of Analytics and Business Intelligence's (OABI's) External Peer Review Program (EPRP) conducts monthly standardized manual medical record reviews of stratified random samples of VA outpatients at all 139 facilities of the 21 VA networks. EPRP has assessed follow-up for alcohol misuse since 2006 (11) and EPRP data has high reliability. (17)

This study's sample included outpatients eligible for alcohol screening whose records were reviewed by EPRP from October 2007 (when follow-up for alcohol misuse was first required) through March 2010. Patients seen in VA clinics, including primary care and specialty medical, surgical and mental health clinics, were eligible for screening except for 0.003% with cognitive impairment or receiving hospice care (18). Each network is estimated to have provided care for 134,000-458,000 patients in 2008-2009. Because EPRP reviewed far fewer records, this study used data from 30 months to provide adequate sample sizes for network-level analyses (the level of accountability for VA performance measures).(11)

Measures

Alcohol Screening from EPRP Medical Record Reviews—The Alcohol Use Disorders Identification Test - Consumption (AUDIT-C), a validated screening questionnaire,(18) has been required for alcohol screening in the VA since 2006.(19) However, networks use variable approaches to implement AUDIT-C screening (e.g. interview, paper, etc.) which may account for differences in the quality of screening across networks.(15) AUDIT-C scores ≥ 5 were considered positive screens, consistent with the VA's quality indicator for follow-up for alcohol misuse.(11)

Follow-up for Alcohol Misuse from EPRP Medical Record Reviews—Patients who screened positive for alcohol misuse were considered to have been offered appropriate follow-up for the purposes of this study if EPRP abstractors found any documented alcohol-related advice or feedback, referral to addiction treatment, or discussion of referral within 30 days after a positive alcohol screen.(11)

Covariates—Patients' age, gender, and race were obtained from VA's National Patient Care Databases. An independent facility-level survey measure of the prevalence of alcohol misuse (AUDIT-C 5) was estimated from patient surveys based on the state where each facility was located using the VA's Survey of Healthcare Experiences of Patients (SHEP) for fiscal years 2007-2008 (20). SHEP was mailed monthly by OABI to a random sample of established outpatients with a recent visit (n 1,228-30,605 patients/state; response rate 54.5%).

Analyses

Descriptive Network Statistics—For each network, EPRP medical record review data were used to estimate the proportion of patients with documented screening for alcohol misuse, and the proportion of screened patients with positive screens (“screening prevalence of alcohol misuse”).

Network Performance on the Two Quality Indicators—Two quality indicators were calculated for each VA network using patient-level data from medical record reviews. A network's positive-screen-based quality indicator of follow-up for alcohol misuse was:

$$\frac{\text{the number of patients with positive alcohol screens and appropriate follow-up documented}}{\text{all patients in the network with positive alcohol screens}}$$

A population-based quality indicator was selected as the comparator for the positive-screen-based quality indicator because a population-based measure is not biased by the specific definition of its denominator or by how the instrument used to determine the denominator is implemented clinically. A network's population-based quality indicator of follow-up for alcohol misuse was:

$$\frac{\text{the number of patients with positive alcohol screens and appropriate follow-up documented}}{\text{all patients in the network eligible for alcohol screening}}$$

Both quality indicators were expressed as percentages; the population-based quality indicator was also expressed as the number of patients who had alcohol misuse identified and appropriate follow-up documented per 100,000 eligible, to reflect the clinical implications of observed differences. Each network's relative performance (i.e. rank) on each quality indicator was then determined,(10) with “1” indicating best performance.

Assessment of Convergent Validity—Each network's difference in ranks on the two measures was calculated. In the absence of “gold standards” for quality indicators, convergent validity provides one indication of validity.(21)

Assessment of Denominator Bias—To evaluate whether networks that performed better on the positive-screen-based quality indicator were potentially biased by a lower screening prevalence of alcohol misuse documented in the medical record, networks were

divided into six groups based on differences in ranks on the two unadjusted quality indicators. Logistic regression was then used to estimate the adjusted screening prevalence of alcohol misuse across the six groups. Estimates were adjusted for demographics and the independent survey measure of alcohol misuse at each facility so that differences in the documented prevalence of positive alcohol screens across networks would not be biased by differences in patient demographics and/or drinking across geographically diverse facilities. Differences across groups were tested with post-estimation Wald tests.

Sensitivity Analyses: Adjusted Quality Indicators—Main analyses used unadjusted quality indicators.(22) Because differences in network performance on the quality indicators could reflect differences in demographics (23-26) or the true prevalence of alcohol misuse across networks, sensitivity analyses adjusted the two quality indicators for demographics and the independent facility-level survey measure of the prevalence of alcohol misuse, to determine if adjustment meaningfully altered findings.

Analyses were conducted in STATA 11.(27)

Results

Network Screening Characteristics

Across networks, 95.9-98.7% of eligible outpatients had screening with the AUDIT-C documented (Appendix A) and the screening prevalence of alcohol misuse varied two-fold (4.6-9.3%; Table 2).

Network Performance on the Two Quality Indicators

The positive-screen-based quality indicator of follow-up for alcohol misuse demonstrated marked variability across the networks: 46.3%-70.8% of patients who screened positive for alcohol misuse had appropriate follow-up documented. The population-based quality indicator demonstrated that 2.7%-5.4% of patients eligible for screening had alcohol misuse identified and appropriate follow-up documented (Table 2).

Convergent Validity of the Two Quality Indicators

Network performance on the two quality indicators was often inconsistent. For example, networks A and B had markedly different performance on the positive-screen-based quality indicator (46.3% and 70.8% respectively), but identified and documented follow-up for alcohol misuse in similar proportions of patients: 3.6% and 3.4% respectively on the population-based measure (Table 2). Conversely, networks P, J, E, Q and G had similar performance on the positive-screen-based quality indicator (54.2 -56.7%), but very different performance on the population-based measure (2.9-5.0%). Networks C, N, and R also had similar positive-screen-based quality indicators (64.7-65.7%) despite population-based measures ranging from 2.9% to 5.4%. Furthermore, these inconsistencies translate into large differences in the absolute number of patients with alcohol misuse identified and appropriately managed. For example, Networks C and R, with similar performance on the positive-screen-based quality indicator, differed by 2,512 patients with alcohol misuse identified and offered follow-up (2899 vs. 5411) per 100,000 eligible for screening.

Differences in each network's ranks on the two quality indicators ranged from +14 to -13 (Figure 1). Six of the 21 networks differed by over 7 ranks (lines, Figure 1).

Assessment of Denominator Bias

The mean adjusted screening prevalence of alcohol misuse based on medical record reviews differed significantly across the 6 groups of networks based on differences in ranks on the

two (unadjusted) quality indicators (Table 3). Networks that ranked over 7 ranks higher on the positive-screen-based quality indicator had a lower screening prevalence of alcohol misuse than networks that ranked over 7 ranks higher on the population-based quality indicator (4.1% vs. 7.4%; top, Table 3).

Sensitivity Analyses

Adjustment of the two quality indicators did not meaningfully change any findings (Appendices B-D).

Discussion

This study demonstrates important limitations of quality indicators of follow-up care for alcohol misuse that use patients with positive alcohol screens as the denominator. First, network performance on the positive-screen-based quality indicator did not reflect the proportion of patients who had alcohol misuse identified and appropriate follow-up documented. Moreover, the magnitude of the observed inconsistencies was clinically meaningful. For example, two networks performed almost identically on the positive-screen-based quality indicator (64.7% and 65.4%) despite one identifying and offering appropriate follow-up for alcohol misuse to almost twice as many patients (5,411 vs. 2,899) per 100,000 eligible for screening. Given that some VA networks screen over 450,000 patients a year, two networks with comparable sizes and performance on a positive-screen-based quality indicator could differ by over 11,000 patients identified and offered care for alcohol misuse each year. Moreover, results suggest that the positive-screen-based quality indicator was biased by insensitive screening programs: the better that networks performed on the positive-screen-based quality indicator compared to the population-based quality indicator, the lower their screening prevalence of alcohol misuse (i.e. the less likely they were to identify alcohol misuse by screening).

Alcohol screening and brief counseling interventions have been deemed the third highest prevention priority for US adults (28,29) among practices recommended by the US Preventive Services Task Force.(30) Positive-screen-based quality indicators of follow-up for alcohol misuse have been put forth by The Joint Commission (TJC),(9) as well as the National Business Coalition on Health (NBCH) to increase alcohol screening and follow-up.(12) Our results demonstrate potential problems with these quality indicators. Additionally, whereas the VA has required use of a common alcohol screening questionnaire and threshold to standardize the denominator of its positive-screen-based quality indicator, TJC and the NBCH have not specified standard alcohol screening questionnaires or thresholds.(9,12) This will likely result in even greater variability in the screening prevalence of alcohol misuse which could further bias positive-screen-based quality indicators.(23-26)

These findings also call into question other behavioral quality indicators. Positive-screen-based quality indicators are increasingly used for depression and other behavioral conditions.(31,32) These measures, developed from clinical guidelines and expert opinion,(13) are often paired with measures to incentivize behavioral screening because under-identification is one of the greatest barriers to quality behavioral health care.(2) However, no previous study to our knowledge has evaluated whether positive-screen-based quality indicators for follow-up on behavioral conditions preferentially reward health systems that identify *fewer* patients with the condition of interest, despite known limitations of other quality indicators based on clinical guidelines (33-35). Furthermore, this bias could also impact “diagnosis-based” behavioral accountability measures that use patients diagnosed with behavioral conditions as the denominator,(35) such as the Healthplan Effectiveness Data and Information Set (HEDIS) alcohol or other drug (AOD) measures used by The National Committee for Quality Assurance (NCQA).(36)

This study suggests that alternatives to positive-screen-based quality indicators for behavioral health conditions are needed. The American Medical Association Physician Consortium for Performance Improvement has proposed a population-based quality indicator, similar to that used in this study,(37) that incentivizes identification as well as appropriate follow-up of alcohol misuse. However, population-based quality indicators can seem counterintuitive to clinicians because follow-up is evaluated in all patients irrespective of their need (i.e. in patients with positive and negative screens). Further, population-based quality indicators could be biased due to differences in clinical samples. Therefore, although adjustment did not meaningfully change results in this study, population-based quality indicators may need to be case-mix adjusted. Moreover, all measures that rely on provider documentation for the numerator could be biased by electronic medical records that encourage identical documentation of follow-up irrespective of care provided.

Patient report of appropriate care for alcohol misuse on surveys that include standardized alcohol screening is likely the optimal quality indicator for follow-up of alcohol misuse.(38) Mailed patient surveys are used to assess smoking cessation counseling, and Medicare is planning to use surveys to assess other preventive counseling.(39) Alcohol-related advice is a key component of evidence-based brief alcohol counseling,(40) and the VA has screened for alcohol misuse and measured receipt of alcohol-related advice on patient surveys since 2004.(41) This survey administers the AUDIT-C in a standard fashion, and then asks about alcohol-related advice. Standardized screening on a mailed survey avoids differences in screening methods across systems and patient survey measures are not biased by variability in provider documentation.(38)

This study has several limitations. First, both quality indicators relied on medical record reviews of clinical documentation of appropriate follow-up; there was no external “gold” standard for alcohol-related discussions. The quality of documented alcohol-related discussions is unknown, especially when documentation of follow-up is rewarded, as in the VA since 2007 (11). In addition, this study compared performance at the network level and used data from a 30 month period to improve the precision of estimates,(42) obscuring possible variability across facilities and time. Further research is needed to explore other factors that bias quality measurement, particularly the severity of identified alcohol misuse and the prevalence of identified alcohol use disorders.(23-26) Finally, the generalizability of these findings from the VA to other health systems is unknown, although such systems are increasingly implementing screening with the AUDIT-C (11,13,18,41,43-46) and recent incentives for EHRs (47-50) and Centers for Medicare & Medicaid Services reimbursement for annual alcohol screening (51) will likely increase implementation and monitoring of alcohol screening and follow-up.

Nevertheless, these findings regarding “1st generation” quality indicators for follow-up care of alcohol misuse can inform development of evidence-based “2nd generation” measures. While several national organizations have developed accountability measures for follow-up of alcohol misuse,(9,12,37) others such as the National Quality Forum and NCQA have not, in part due to lack of information on the optimal approach to measuring the quality of appropriate follow-up care. This study evaluated the convergent validity between positive-screen-based and population-based quality indicators, an essential step in improving quality measurement for behavioral conditions.(21) Findings suggest that positive-screen-based quality indicators systematically favor health systems with insensitive alcohol screening programs, undermining efforts to improve identification of alcohol misuse. Other positive-screen-based behavioral quality indicators may have similar limitations. Because under-recognition of behavioral conditions is a critical barrier to quality care,(2) positive-screen-based quality indicators for other behavioral conditions should be evaluated in future research.

In conclusion, valid measures of the quality of care will be essential for improving the recognition of and follow-up care for common behavioral conditions, such as alcohol misuse.(2,6) This study suggests that positive-screen-based quality indicators derived from medical record reviews of provider documentation—like those used by VA and The Joint Commission—should be avoided, due to bias favoring systems with screening programs that identify fewer patients with alcohol misuse (i.e. “denominator bias”) even when a uniform screen and screening threshold are used across all systems.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was supported by VA HSR&D Substance Use Disorders Quality Enhancement Research Initiative (SUD QUERI #SUB98-000), VA Health Services Research and Development (HSR&D), NIAAA R21 AA020894-01A1 and Group Health Research Institute (GHRI). Funders had no role in the conduct or reporting of research. Data were provided via a data use agreement with the VA Office of Analytics and Business Intelligence (OABI; formerly the VA Office of Quality and Performance), which had no role in design or analyses, but reviewed the manuscript before submission to ensure accurate use and reporting of data.

References

1. Kessler RC, Chiu WT, Demler O, et al. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*. 2005; 62:617–627. [PubMed: 15939839]
2. Institute of Medicine: Improving the Quality of Health Care for Mental and Substance-Use Conditions: Quality Chasm Series. Washington D.C.: National Academies Press; 2006.
3. Substance Abuse Mental Health Services Administration. (2010, Table B22) State Estimates of Substance Use from the 2007-2008 National Surveys on Drug Use and Health. Retrieved 2011 Oct 3, from <http://www.oas.samhsa.gov/2k8state/AppB.htm#TabB-9>
4. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *Journal of the American Medical Association*. 2004; 291:2581–2590. [PubMed: 15173149]
5. McGlynn EA, Asch SM, Adams JL, et al. The quality of health care delivered to adults in the United States. *New England Journal of Medicine*. 2003; 348:2635–2645. [PubMed: 12826639]
6. Pincus HA, Spaeth-Rublee B, Watkins KE. Analysis & commentary: the case for measuring quality in mental health and substance abuse care. *Health Affairs (Millwood)*. 2011; 30:730–736.
7. Chassin MR, Loeb JM, Schmaltz SP, et al. Accountability measures--using measurement to promote quality improvement. *New England Journal of Medicine*. 2010; 363:683–688. [PubMed: 20573915]
8. Kilbourne AM, Greenwald DE, Hermann RC, et al. Financial incentives and accountability for integrated medical care in Department of Veterans Affairs mental health programs. *Psychiatric Services*. 2010; 61:38–44. [PubMed: 20044416]
9. The Joint Commission. Tobacco and Alcohol Measures. 2011 Jan 21. Retrieved 2011 Oct 3, from http://www.jointcommission.org/tobacco_and_alcohol_measures/
10. Humphreys K, McLellan AT. A policy-oriented review of strategies for improving the outcomes of services for substance use disorder patients. *Addiction*. 2011; 106:2058–2066. [PubMed: 21631620]
11. Lapham GT, Achtmeyer CE, Williams EC, et al. Increased documented brief alcohol interventions with a performance measure and electronic decision support. *Medical Care*. 2012; 50:179–187. [PubMed: 20881876]
12. National Business Coalition on Health. About eValue8. 2009. Retrieved 2011 Oct 3, from <http://evalue8.org/eValue8>

13. Bradley KA, Williams EC, Achtmeyer CE, et al. Measuring performance of brief alcohol counseling in medical settings: a review of the options and lessons from the Veterans Affairs (VA) health care system. *Substance Abuse*. 2007; 28:133–149. [PubMed: 18077309]
14. Oslin DW, Ross J, Sayers S, et al. Screening, assessment, and management of depression in VA primary care clinics. The Behavioral Health Laboratory. *Journal of General Internal Medicine*. 2006; 21:46–50. [PubMed: 16423122]
15. Bradley KA, Lapham GT, Hawkins EJ, et al. Quality concerns with routine alcohol screening in VA clinical settings. *Journal of General Internal Medicine*. 2011; 26:299–306. [PubMed: 20859699]
16. Williams, EC.; Achtmeyer, CE.; Grossbard, JR., et al. Barriers and facilitators to implementing alcohol screening with a clinical reminder in the VA: A qualitative study. Research Society on Alcoholism; Atlanta, GA: 2011.
17. Jha AK, Perlin JB, Kizer KW, et al. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. *New England Journal of Medicine*. 2003; 348:2218–2227. [PubMed: 12773650]
18. Bradley KA, Kivlahan DR, Williams EC. Brief approaches to alcohol screening: practical alternatives for primary care. *Journal of General Internal Medicine*. 2009; 24:881–883. [PubMed: 19495888]
19. Bradley KA, DeBenedetti AF, Volk RJ, et al. AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcoholism, Clinical and Experimental Research*. 2007; 31:1208–1217.
20. Wright SM, Craig T, Campbell S, et al. Patient satisfaction of female and male users of Veterans Health Administration services. *Journal of General Internal Medicine*. 2006; 21(3):S26–32. [PubMed: 16637941]
21. Pronovost PJ, Lilford R. Analysis & commentary: A road map for improving the performance of performance measures. *Health Affairs (Millwood)*. 2011; 30:569–573.
22. Damman OC, Stubbe JH, Hendriks M, et al. Using multilevel modeling to assess case-mix adjusters in consumer experience surveys in health care. *Medical Care*. 2009; 47:496–503. [PubMed: 19238105]
23. Buchsbaum DG, Buchanan RG, Lawton MJ, et al. A program of screening and prompting improves short-term physician counseling of dependent and nondependent harmful drinkers. *Archives of Internal Medicine*. 1993; 153:1573–1577. [PubMed: 8323420]
24. Buchsbaum DG, Buchanan RG, Poses RM, et al. Physician detection of drinking problems in patients attending a general medicine practice. *Journal of General Internal Medicine*. 1992; 7:517–521. [PubMed: 1403208]
25. Burman ML, Kivlahan DR, Buchbinder MB, et al. Alcohol-related advice for Veterans Affairs primary care patients: Who gets it? Who gives it? *Journal of Studies on Alcohol*. 2004; 65:621–630. [PubMed: 15536772]
26. Volk RJ, Steinbauer JR, Cantor SB. Patient factors influencing variation in the use of preventive interventions for alcohol abuse by primary care physicians. *Journal of Studies on Alcohol*. 1996; 57:203–209. [PubMed: 8683970]
27. StataCorp. Stata Statistical Software: Release 11. College Station, TX: StataCorp LP; 2009. in Edited by
28. Solberg LI, Maciosek MV, Edwards NM. Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. *American Journal of Preventive Medicine*. 2008; 34:143–152. [PubMed: 18201645]
29. Maciosek MV, Coffield AB, Edwards NM, et al. Priorities among effective clinical preventive services results of a systematic review and analysis. *American Journal of Preventive Medicine*. 2006; 31:52–61. [PubMed: 16777543]
30. Jonas DE, Garbutt JC, Amick HR, et al. Behavioral Counseling After Screening for Alcohol Misuse in Primary Care: A Systematic Review and Meta-analysis for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*. 2012
31. Desai MM, Rosenheck RA, Craig TJ. Case-finding for depression among medical outpatients in the Veterans Health Administration. *Medical Care*. 2006; 44:175–181. [PubMed: 16434917]

32. Kilbourne AM, Keyser D, Pincus HA. Challenges and opportunities in measuring the quality of mental health care. *Canadian Journal of Psychiatry*. 2010; 55:549–557.
33. Walter LC, Davidowitz NP, Heineken PA, et al. Pitfalls of converting practice guidelines into quality measures: lessons learned from a VA performance measure. *Journal of the American Medical Association*. 2004; 291:2466–2470. [PubMed: 15161897]
34. Landon BE, O'Malley AJ, Keegan T. Can choice of the sample population affect perceived performance: implications for performance assessment. *Journal of General Internal Medicine*. 2010; 25:104–109. [PubMed: 19936847]
35. Garnick DW, Horgan CM, Chalk M. Performance measures for alcohol and other drug services. *Alcohol Research & Health*. 2006; 29:19–26. [PubMed: 16767849]
36. National Committee for Quality Assurance. HEDIS 2011 Technical Specifications. Vol. 2. Washington, D.C.: 2011.
37. Physician Consortium for Performance Improvement® (PCPI). Preventive care and screening: percentage of patients aged 18 years and older who were screened for unhealthy alcohol use at least once during the two-year measurement period using a systematic screening method and who received brief counseling if identified as an unhealthy alcohol user. The National Quality Measures Clearinghouse™ (NQMC). 2008. Retrieved 2011 Oct 3, from <http://www.qualitymeasures.ahrq.gov/content.aspx?id=13366=alcohol>
38. Bradley KA, Johnson ML, Williams EC. Commentary on Nilsen et al.: The importance of asking patients-the potential value of patient report of brief interventions. *Addiction*. 2011; 106:1757–1759. [PubMed: 21917038]
39. HEDIS. CAHPS Health Plan Survey 4.0H Adult Questionnaire Commercial. 2008. Retrieved 2011 Oct 3, from http://www.cchri.org/programs/CAHPS%20PDFs/CAHPS_4_0H_Adult_Commercial_survey.pdf
40. Whitlock EP, Polen MR, Green CA, et al. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*. 2004; 140:557–568. [PubMed: 15068985]
41. Bradley KA, Williams EC, Achtmeyer CE, et al. Implementation of evidence-based alcohol screening in the Veterans Health Administration. *American Journal of Managed Care*. 2006; 12:597–606. [PubMed: 17026414]
42. Keenan PS, Cleary PD, O'Malley AJ, et al. Geographic area variations in the Medicare health plan era. *Medical Care*. 2010; 48:260–266. [PubMed: 20182269]
43. Seale JP, Shellenberger S, Tillery WK, et al. Implementing alcohol screening and intervention in a family medicine residency clinic. *Substance Abuse*. 2005; 26:23–31. [PubMed: 16492660]
44. Seale JP, Shellenberger S, Boltri JM, et al. Effects of screening and brief intervention training on resident and faculty alcohol intervention behaviours: a pre- post-intervention assessment. *BMC Family Practice*. 2005; 6:46. [PubMed: 16271146]
45. Rose HL, Miller PM, Nemeth LS, et al. Alcohol screening and brief counseling in a primary care hypertensive population: a quality improvement intervention. *Addiction*. 2008; 103:1271–1280. [PubMed: 18422825]
46. Bradley, KA.; Williams, EC. Implementation of Screening and Brief Intervention in Clinical Settings Using Quality Improvement Principles, in *Principles of Addiction Medicine*. Fiellin, D.; Miller, S.; Saitz, R.; Reis, R., editors. Lippincott Williams & Wilkins; 2009. p. 288-292.
47. Blumenthal D. Launching HITECH. *New England Journal of Medicine*. 2010; 362:382–385. [PubMed: 20042745]
48. Blumenthal D, Tavenner M. The “meaningful use” regulation for electronic health records. *New England Journal of Medicine*. 2010; 363:501–504. [PubMed: 20647183]
49. Ahern DK, Woods SS, Lightowler MC, et al. Promise of and potential for patient-facing technologies to enable meaningful use. *American Journal of Preventive Medicine*. 2011; 40:S162–172. [PubMed: 21521591]
50. Druss BG, Mauer BJ. Health care reform and care at the behavioral health--primary care interface. *Psychiatric Services*. 2010; 61:1087–1092. [PubMed: 21041346]
51. Walker, EP. Medicare to cover alcohol, depression screening. 2011 Oct 17. Retrieved 2012 Feb 1, from <http://www.medpagetoday.com/PrimaryCare/PreventiveCare/29085>

Figure 1.
 Lower numbered ranks reflect higher performance (1 = highest performance).

Table 1
Example of Denominator Bias in Quality Indicators of Follow-up Care for Behavioral Conditions that Use Patients with Positive Screens as the Denominator (“Positive-screen-based” Quality Indicators)^{*}

	Number of Patients With Behavioral Condition [*]	Screening Prevalence of Behavioral Condition [*]	Number of Patients with Positive Screen [*]	Percent with Documented Appropriate Follow-up [*]	N and % Offered Appropriate Follow-up for Condition [§]
A	13,000	10%	10,000	50%	5000 5.0%
B	13,000	5%	5,000	50%	2500 2.5%
C	13,000	5%	5,000	80%	4000 4.0%

^{*} Three identical healthcare systems, each with 100,000 identical patients, 13% with a behavioral condition, demonstrating how variation in clinical screening programs could bias positive-screen-based quality indicators of follow-up for behavioral health conditions. Numbers and percentages in bold are assumptions: true prevalence 13%, the prevalence of positive screens (“screening prevalence”) are 5% or 10%, and 50% or 80% have appropriate follow-up documented; un-bolded numbers and percentages are calculated based on assumptions.

[§] The proportion of all eligible patients offered appropriate follow-up is a “population-based” quality indicator.

Table 2
Each Network's Screening Prevalence of Alcohol Misuse Documented in the Medical Record based on Medical Record Reviews* and Each Network's Performance on the Two Types of Quality Indicators of Follow-up for Alcohol Misuse

Network	Screening Prevalence of Alcohol Misuse*			Quality Indicators of Follow-up for Alcohol Misuse		
	%	95% CI	Positive-screen-based [§]	%	95% CI	Population-based [‡]
C	4.6	4.1-5.0	64.7	59.7-69.7	2.9	2.5-3.3
S	4.8	4.2-5.3	63.0	56.9-69.2	2.9	2.5-3.4
B	4.9	4.5-5.2	70.8	67.2-74.3	3.4	3.1-3.7
G	5.5	5.0-5.9	54.2	50.0-58.3	2.9	2.5-3.2
L	5.5	5.1-5.9	49.6	45.9-53.4	2.7	2.4-2.9
P	6.0	5.6-6.4	56.7	53.4-59.9	3.3	3.0-3.6
U	6.5	6.1-6.9	59.0	55.6-62.4	3.8	3.4-4.1
M	6.6	6.2-7.0	63.2	60.1-66.3	4.1	3.8-4.5
I	6.8	6.2-7.3	51.6	47.6-55.5	3.4	3.0-3.7
O	6.9	6.4-7.4	53.3	49.5-57.0	3.6	3.3-4.0
N	7.0	6.4-7.7	65.7	61.2-70.1	4.5	4.0-5.0
F	7.1	6.6-7.6	62.5	59.0-65.9	4.3	3.9-4.7
T	7.2	6.7-7.7	48.3	44.7-51.9	3.4	3.0-3.7
J	7.3	6.8-7.8	55.3	51.9-58.6	4.0	3.6-4.3
K	7.4	6.9-8.0	58.1	54.3-62.0	4.2	3.8-4.6
H	7.5	6.9-8.0	69.5	65.8-73.2	5.1	4.6-5.6
E	7.8	7.3-8.3	55.5	52.2-58.8	4.2	3.9-4.6
A	7.9	7.3-8.4	46.3	42.7-49.9	3.6	3.2-3.9
R	8.4	7.8-9.0	65.4	61.7-69.0	5.4	4.9-5.9
D	9.1	8.6-9.6	48.4	45.3-51.4	4.3	3.9-4.7
Q	9.3	8.8-9.9	54.8	51.7-57.9	5.0	4.6-5.4

Table 2 Legend: Networks were labeled (A-U) by letters in non-numeric order consistent with a previous report,(15) and ordered based on the screening prevalence of alcohol misuse.

* The proportion of patients who screened positive for alcohol misuse on the AUDIT-C (5 points) based on External Peer Review Program (EPRP) medical record reviews.

[§] Positive-screen-based quality indicator of follow-up for alcohol misuse: the proportion of patients with alcohol misuse (AUDIT-C ≥ 5) who had follow-up documented based on EPRP medical record reviews.

[†] Population-based quality indicator of follow-up for alcohol misuse: the proportion and number per 100,000 of patients eligible for screening who had alcohol misuse identified (AUDIT-C ≥ 5) and follow-up documented based on EPRP medical record reviews.

Table 3

Differences in the Adjusted Screening Prevalence of Alcohol Misuse Documented in Medical Records Across Groups of Networks Based on Differences in Ranks on the Two Types of Quality Indicators of Follow-up for Alcohol Misuse.

Six Groups of Networks Based on Differences in Rank on the Two Unadjusted Quality Indicators of Follow-up for Alcohol Misuse													

Legend Table 3. Positive values of differences in rank indicate higher performance on the positive-screen-based quality indicator; negative values indicate higher performance on population-based quality indicator.

* The proportion of patients who screened positive for alcohol misuse on the AUDIT-C (5 points) based on EPRP medical record reviews adjusted for age, gender, race and an independent survey measure the facility prevalence of alcohol misuse.