Journal of Substance Abuse Treatment 53 (2015) 16–21

Contents lists available at ScienceDirect

Journal of Substance Abuse Treatment

Examining the Specification Validity of the HEDIS Quality Measures for Substance Use Disorders ☆

Alex H.S. Harris, Ph.D. [a,*], Laura Ellerbe, M.S. [a,1], Tyler E. Phelps, M.L.I.S., M.S. [a,2], John W. Finney, Ph.D. [a,3], Thomas Bowe, Ph.D. [a,4], Shalini Gupta, M.S. [a,5], Steven M. Asch, M.D., M.P.H. [a,6], Keith Humphreys, Ph.D. [a,7], Jodie Trafton, Ph.D. [b,8]

[a] Center for Innovation to Implementation, VA Palo Alto Health Care System and Stanford University School of Medicine, 795 Willow Road (MPD152), Menlo Park, CA 94025
[b] Program Evaluation and Resource Center, VA Palo Alto Health Care System and Stanford University School of Medicine, 795 Willow Road (MPD152), Menlo Park, CA 94025

ARTICLE INFO

Article history:
Received 8 July 2014
Received in revised form 30 December 2014
Accepted 6 January 2015

Keywords:
Quality measurement
Specification validity
Electronic medical records
Substance-related disorders
Veterans Health Administration

ABSTRACT

Accurate operationalization is a major challenge in developing quality measures for substance use disorder treatment. *Specification validity* is a term used to describe whether a quality measure is operationalized such that it captures the intended care processes and patients. This study assessed the specification validity of the 2009 Healthcare Effectiveness Data and Information Set (HEDIS®) substance use disorder initiation and engagement measures by examining whether encounters assumed to include relevant treatment have corroborating evidence in the clinical progress notes. The positive predictive values were excellent (>90%) for residential and outpatient records selected from addiction treatment programs but more modest for records generated in non-addiction settings, and were highly variable across facilities. Stakeholders using these measures to compare care quality should be mindful of the clinical composition of the data and determine if similar validation work has been conducted on the systems being evaluated.

Published by Elsevier Inc.

Quality measures are increasingly used in behavioral health care for reimbursement and for monitoring conformance with clinical policies, often without fully understanding the limitations of the measures. As with any clinical test, the validity of quality measures depends on how the measures are operationalized and the quality and specificity of the underlying data. If the quality measures are valid for the intended purposes, their use can significantly improve the health care of patients (Chassin, Loeb, Schmaltz, & Wachter, 2010). However, unintended consequences can arise if the quality measures are based on weak evidence, are poorly operationalized, or in the case of process measures, are not associated with better patient outcomes (Chassin et al., 2010; Wachter, Flanders, Fee, & Pronovost, 2008).

In many areas of health care, one of the major challenges in quality measure development is accurately specifying the targeted patients, contexts, and processes using available and reliable data elements. In mental health and substance use disorder treatment, practices for diagnosing particular conditions or disorders may vary both across and within healthcare systems. Furthermore, the Common Procedural Terminology (CPT®) and other available administrative and billing codes related to these treatments are often much less specific than would be ideal for operationalizing evidence-based treatments. For example, psychotherapies vary substantially in the quality and strength of their supporting scientific evidence; only the most effective are included in definitions and measures of quality. However, administrative codes for individual psychotherapy (such as 90804–90819) do not make any distinctions about the type of psychotherapy provided. It can be similarly difficult to determine from available data whether addiction treatment was provided at all, let alone whether it was an evidence-based form.

Ideally, a process quality measure has specification validity, meaning that the strategy for operationalizing the targeted processes, patients, and contexts does so with acceptable accuracy. One indicator of accuracy is a measure's positive predictive value (PPV), which can be defined as the probability of finding any corroborating evidence of treatment that meets the quality indicator's definition (true positives/measured positives). Quality measure developers often make substantial efforts to operationalize the targeted processes but do not take the

☆ This study was funded by the VA Health Services Research and Development Service (Grants Nos. SUS 99-015; IIR 10-370-2; RCS-14-232). None of the authors has any conflict of interest related to this work. The views expressed herein are not necessarily those of the Department of Veterans Affairs.

* Corresponding author. Tel.: +1 650 493 5000x23423; fax: +1 650 617 2736.
E-mail addresses: Alexander.Harris2@va.gov (A.H.S. Harris), Laura.Ellerbe@va.gov (L. Ellerbe), Tyler.Phelps@va.gov (T.E. Phelps), John.Finney@va.gov (J.W. Finney), Thomas.Bowe@va.gov (T. Bowe), Shalini.Gupta2@va.gov (S. Gupta), Steven.Asch@va.gov (S.M. Asch), knh@stanford.edu (K. Humphreys), Jodie.Trafton@va.gov (J. Trafton).
[1] Tel.: +1 650 493 5000x27357; fax: +1 650 617 2736.
[2] Tel.: +1 650 493 5000x27553; fax: +1 650 617 2736.
[3] Tel.: +1 650 493 5000x22848; fax: +1 650 617 2736.
[4] Tel.: +1 650 493 5000x22666; fax: +1 650 617 2736.
[5] Tel.: +1 650 493 5000x25562; fax: +1 650 617 2736.
[6] Tel.: +1 650 493 5000x23476; fax: +1 650 617 2736.
[7] Tel.: +1 650 493 5000x22814; fax: +1 650 617 2736.
[8] Tel.: +1 650 493 5000x22474; fax: +1 650 617 2736.

http://dx.doi.org/10.1016/j.jsat.2015.01.002
0740-5472/Published by Elsevier Inc.

important step of assessing whether their judgments and decisions were acceptably correct.

The selection of specification strategies, which range from labor-intensive methods (e.g., chart review or direct observation of care) to the use of preexisting and easily accessible administrative data, faces tradeoffs among feasibility, cost, and accuracy. Although strategies based on administrative data are usually inexpensive and feasible, they are usually of unknown validity. Without knowing the accuracy of a particular specification, quality measure developers may find it impossible to determine and make transparent whether that strategy appropriately balances specification validity with feasibility and cost.

For example, as the most widely used set of quality measures in the U.S. managed healthcare industry, the Healthcare Effectiveness Data and Information Set (HEDIS®) operationalizes quality measures across many domains ("HEDIS®, 2013," 2013). Two of these measures, initiation and engagement, are intended to assess early involvement in addiction treatment. The specification strategy used in these measures is based on commonly available diagnostic and procedure codes (see Appendix A). This strategy is feasible and inexpensive, but only one evaluation of the specification validity has been undertaken (Harris, Reeder, Ellerbe, & Bowe, 2011).

In that study, Harris and colleagues estimated the overall and facility variation in positive predictive value of records meeting the HEDIS specification criteria for substance use disorder treatment using data from the U.S. Veterans Health Administration (VHA). To make this assessment, the investigators calculated the probability of finding any evidence of addiction treatment in the clinicians' progress notes for records that met the HEDIS criteria. For records selected from substance use disorder treatment programs, the positive predictive values for inpatient and outpatient records were excellent (99% and 92%, respectively). However, of outpatient encounters with a qualifying diagnosis/procedure code combination outside of addiction treatment clinics (e.g., in primary care or mental health clinics), only 63% had chart evidence of relevant treatment. Within non-addiction inpatient units, 46% of sampled qualifying records had chart evidence of addiction treatment. Given that the initiation and engagement measures are designed to capture treatment both within and outside of specialty addiction programs, and that most of the qualifying records are generated outside of addiction programs, the modest positive predictive values for records generated outside of addiction treatment programs make interpretation of the resulting performance data difficult.

VHA is comprised of over 140 major medical centers (with additional affiliated community-based outpatient clinics) organized into 21 regional Veterans Integrated Service Networks (VISNs). Perhaps more alarming than the modest overall positive predictive values for records generated outside of addiction settings was the variation in positive predictive values among the 21 networks (36%–85% in outpatient records and 18%–68% for inpatient records). Such variation in specification validity makes use and interpretation of the performance data even more complex and challenging.

The technical specifications for the HEDIS Initiation and Engagement measures have undergone substantial revision since the aforementioned study, most notably with the addition of Healthcare Common Procedure Coding System (HCPCS) codes. HCPCS includes much more specific codes (e.g., H0005—Alcohol and/or drug services; group counseling by clinician) than CPT and can be used by clinicians who are not licensed independent practitioners ("HEDIS®, 2009," 2009). Theoretically, this should improve the specification validity of the HEDIS measures. Therefore, the purpose of the present study was to evaluate the updated specifications and to assess if the problems identified in our previous study have been ameliorated.

1. Methods

We estimated the positive predictive value of the updated 2009 HEDIS specification strategy for identifying substance use disorder care in a health encounter. In this study, positive predictive value was defined as the proportion of records that met the HEDIS specifications that could be verified as true positives through chart review. Due to limited resources and the labor intensive nature of chart review, we did not pursue other measures of performance (e.g., sensitivity, specificity) that would have required the sampling of large numbers of records that did not meet the HEDIS specifications. Although chart notes are not a perfect record of the clinical encounter, they provide a much more logistically feasible means to check if relevant treatment occurred compared to direct observation or interviews. First, we retrieved over 5 million records from outpatient and inpatient/residential encounters in fiscal years (FY) 2008 and 2009 from the VHA National Patient Care Databases that met the specifications of the 2009 HEDIS measures (2009) and were therefore presumed to involve addiction treatment. Records meeting the specifications, or considered "HEDIS-qualified," contained both a substance use disorder diagnosis and a mental health procedure code from among those listed in Appendix A. VHA-specific treatment specialty codes (i.e., clinic stop and bed section codes) were used to assess the specification validity of the initiation and engagement measures by type: in outpatient versus inpatient/residential and substance abuse specialty versus non-specialty settings. Thus, 700 records from each of four HEDIS-qualified record types were randomly sampled: outpatient care in an addiction specialty setting, outpatient care in a general setting, inpatient/residential care in an addiction specialty setting, and inpatient/residential care in a general setting ($n = 2800$).

The research team conducted a chart review of these records to determine if the progress note from each encounter contained any mention of addiction treatment, as is presumed by the logic of the measures. In order to do so, VistAWeb, an intranet Web application of the Veterans Health Information Systems and Technology Architecture (VistA) was used to extract the corresponding clinical progress notes on the date of the qualifying encounter. Progress notes were extracted from 2800 records, entered into a secure Microsoft Access database, and then de-identified.

The chart review team developed a rating procedure to conduct a qualitative content analysis of the notes; we established guidelines for determining if there was sufficient documentation of addiction treatment to assume that substance use disorder was addressed in a mental health assessment, outpatient encounter, inpatient stay, or detoxification. Two reviewers determined if there was documentation of relevant care first by identifying phrases commonly used to describe addiction treatment, and then examining the context in which the words were used. For example, phrases or acronyms such as AUDIT-C, Addiction Severity Index (ASI), or Substance Abuse Treatment Program often appeared in the progress notes. However, the reviewers also considered the context, focusing on the degree of patient interaction and provider follow-up, in order to prevent misclassification. We broadly defined 'treatment' for the purposes of coding as at least an exchange between patient and provider(s) regarding substance use without judgment on the quality or depth of the care provided. For instance, if a progress note included "An alcohol screening test (AUDIT-C) was positive (score = 9)," which suggests drinking behavior at risky levels, but did not mention any provider response or intervention (e.g., "The patient was advised to abstain from alcohol use"), then the record was coded as not having sufficient documentation of addiction treatment. Likewise, if a progress note mentioned that a "rally pack" (i.e., an intravenous nutritional support including thiamine, folic acid, and magnesium) had been given to an intoxicated patient, but there was no documentation of further clinical action (e.g., Clinical Institute Withdrawal Assessment given; detox meds or treatment program consult ordered), then the record was coded as not having enough documentation of addiction treatment.

However, some records documented provider response but no interaction between patient and provider; these were differentiated with a "partial care" code. For example, if a progress note from an outpatient