

# Does Meeting the HEDIS Substance Abuse Treatment Engagement Criterion Predict Patient Outcomes?

Alex HS Harris, PhD

Keith Humphreys, PhD

Thomas Bowe, PhD

Quyen Tiet, PhD

John W. Finney, PhD

## Abstract

*This study examines the patient-level associations between the Health Plan Employer Data and Information Set (HEDIS) substance use disorder (SUD) treatment engagement quality indicator and improvements in clinical outcomes. Administrative and survey data from 2,789 US Department of Veterans Affairs SUD patients were used to estimate the effects of meeting the HEDIS engagement criterion on improvements in Addiction Severity Index Alcohol, Drug, and Legal composite scores. Patients meeting the engagement indicator improved significantly more in all domains than patients who did not engage, and the relationship was stronger for alcohol and legal outcomes for patients seen in outpatient settings. The benefit accrued by those who engaged was statistically significant but clinically modest. These results add to the literature documenting the clinical benefits of treatment entry and engagement. Although these findings only indirectly support the use of the HEDIS engagement measure for its intended purpose—discriminating quality at the facility or system level—they confirm that the processes of care captured by the measure are associated with important patient outcomes.*

---

Address correspondence to Alex HS Harris, PhD, Center for Health Care Evaluation, Department of Veterans Affairs, Palo Alto Health Care System and Stanford University School of Medicine, 795 Willow Road (MPD-152), Menlo Park, CA 94025, USA. Phone: +1-650-4935000. Email: Alexander.Harris2@va.gov.

Keith Humphreys, PhD, Center for Health Care Evaluation, Department of Veterans Affairs, Palo Alto Health Care System and Stanford University School of Medicine, Menlo Park, CA, USA. Email: knh@stanford.edu

Thomas Bowe, PhD, Center for Health Care Evaluation, Department of Veterans Affairs, Palo Alto Health Care System and Stanford University School of Medicine, Menlo Park, CA, USA. Email: tbowe@va.gov

Quyen Tiet, PhD, Center for Health Care Evaluation, Department of Veterans Affairs, Palo Alto Health Care System and Stanford University School of Medicine, Menlo Park, CA, USA. Email: Quyen.Tiet@va.gov

John W. Finney, PhD, Center for Health Care Evaluation, Department of Veterans Affairs, Palo Alto Health Care System and Stanford University School of Medicine, Menlo Park, CA, USA. Email: John.Finney@va.gov

*Journal of Behavioral Health Services & Research*, 2008. © 2008 National Council for Community Behavioral Healthcare.

## Introduction

Quality indicators are used to evaluate health care systems by assessing the extent to which treatment conforms to established practice guidelines (process indicators) or other standards of quality, such as improvement in patients' symptoms (outcome indicators).<sup>1-3</sup> Some important indicators of quality, such as pretreatment and posttreatment assessments of symptoms, often are impractical to obtain. Quality indicators constructed from administrative data, such as those based on utilization and staffing data, are inexpensive and easy to generate but may be of low or unknown predictive validity. In this context, predictive validity refers to the association between a quality indicator and the more difficult to obtain gold standards of quality, such as patient outcomes.

It is thus highly preferable to identify process quality indicators, ideally calculable with readily available administrative data, that reliably predict subsequent outcomes. However, without knowing the extent to which such process quality measures are correlated with patient outcomes, it is impossible to rationally balance predictive validity and feasibility in choosing which indicators to use. Many widely adopted quality indicators are derived from administrative data, yet in many cases no or very weak direct validity evidence is available to support their use.<sup>4,5</sup> For example, many of the process measures of treatment quality adopted by the National Committee for Quality Assurance (NCQA) as Health Plan Employer Data and Information Set (HEDIS) measures and many of the Department of Veterans Affairs (VA) process quality measures are comprised of diagnostic, procedure, and/or utilization data from administrative sources,<sup>6,7</sup> yet the presumed links between many of these measures and patients' outcomes have never been evaluated.

The few efforts to directly evaluate the link between process performance criteria and outcomes across diverse areas of medicine have yielded mixed results. Most commonly, nonsignificant or very weak associations between processes measures and subsequent outcomes are found,<sup>5,8-10</sup> although some positive studies exist.<sup>11, 12</sup> Currently unknown is why some process measures have the expected relationship with outcomes, whereas most that have been directly evaluated do not. Aspects of the underlying condition, the maturity of the clinical evidence, or the quality measure development process itself may be related to the predictive validity of process quality measures. Regardless, the use of quality measures with unknown or poor predictive validity runs the risk of incentivizing poor or incomplete care. Therefore, before candidate process quality measures are adopted, the relationship between the measures and clinical outcomes should be verified.

To measure the quality of substance use disorder (SUD) treatment services, the Washington Circle, an organization supported by the US Center for Substance Abuse Treatment, developed the Initiation and Engagement quality measures.<sup>2,13</sup> These were adopted subsequently by the NCQA into HEDIS measures. Because HEDIS is the most widely used set of quality measures in the managed health care industry in the US, NCQA's adoption of these measures means that many health care systems are now tracking them.

"Initiation" refers to the percentage of identified SUD patients with at least a prior 60-day SUD service-free period that has either a residential SUD admission or an SUD-related outpatient visit *and* an additional SUD-related visit within 14 days. "Engagement" refers to the percentage of identified SUD patients that receives two additional SUD-related visits within 30 days of initiation. For those initiating as inpatients or residential patients, the two additional visits for engagement must occur within 30 days of discharge. These measures are constructed from Common Procedural Terminology (CPT), diagnosis-related group (DRG), and ICD-9 codes and are designed to be calculable by any health care system.

Although the initiation and engagement indicators have been widely adopted, evidence linking them to other quality measures, especially patient outcomes, is minimal. Only one prior study has examined the association between the HEDIS SUD measures and subsequent patient-level outcomes. When examining SUD treatment data in a sample of adult clients treated in publicly

funded Oklahoma substance abuse outpatient programs, engagement, but not initiation, was associated with lower likelihood of subsequent arrest and incarceration but not charges of driving under the influence.<sup>14</sup> The present study examines the patient-level association between meeting the engagement criteria and subsequent improvements in alcohol, drug, and legal symptoms. This study focuses primarily on engagement and not initiation because the vast majority of our eligible sample initiated (i.e., 100% of inpatients, 86% of residential patients, 78% of outpatients, 90% of intensive outpatients, 85% of domiciliary patients, and 86% of methadone patients), resulting in very little between-patient variability. Nevertheless, we provide findings from some exploratory analyses of the relationship between initiation and patient improvement. The current study extends our knowledge in this area by using a nationwide sample of VA patients seen in both inpatient–residential and outpatient settings and examining the effect of engagement in more domains of functioning.

## Methods

### Sample selection

For a VA-funded Outcomes Monitoring Project,<sup>15</sup> up to 50 new patients were randomly selected from each program in a randomly selected and representative sample of VA SUD treatment programs. Baseline self-administered Addiction Severity Index (ASI)<sup>16,17</sup> data were collected on 5,723 patients who entered treatment in one of 118 VA SUD treatment programs located at 73 VA facilities. The programs were of six types: inpatient, domiciliary, residential, intensive outpatient, outpatient, and methadone. Programs that indicated they provided specialized services to SUD patients who had serious comorbid psychiatric disorders (dual diagnosis) were included in the sampling frame, but psychiatric treatment programs that provided specialized services for comorbid SUD patients were not. In order to be eligible to meet the engagement criterion, patients must have had a *qualifying visit*, defined as an SUD-related health care encounter (specified by CPT and primary or secondary ICD-9-CM codes) that was preceded by at least 60 days without an SUD-related encounter. From the sample of 5,723 patients, this study focused on the 2,789 who had a qualifying visit and were thus eligible to engage, within the 14 days preceding the baseline assessment.

Although the 14-day aspect of this sample definition is somewhat arbitrary, it was chosen to include people who started a new episode of care just prior to enrolling in the Outcomes Monitoring Project and to exclude people who started care far in advance of the baseline assessment. For example, a patient who had a qualifying outpatient visit 10 days prior to an inpatient admission during which he or she were enrolled in and completed the baseline assessment for the Outcomes Monitoring Project was included. A patient with a qualifying inpatient admission 30 days prior to the Outcomes Monitoring Project baseline assessment was excluded. As discussed more fully in the “Results” section, sensitivity analyses were conducted to assess the impact of the sample definition on the major findings.

### Patient baseline data

In most cases, a self-administered questionnaire focusing on ASI composite score items was completed by patients upon enrollment into the Outcomes Monitoring Project at the beginning of an episode of VA SUD treatment (the other patients were interviewed by clinical staff). Like the interview-based ASI, the self-administered version results in a 0.00 to 1.00 (less to more severe) composite score that reflects symptoms–functioning over the past 30 days in each of seven domains (i.e., alcohol, drug, psychiatric, medical, legal, employment, and family–social). The present study focused on the associations between engagement and change in the alcohol, drug,

and legal composites and utilized baseline values of the other composites and demographic characteristics as case-mix adjustment variables.

Self-administered ASI composite scores have been shown to have adequate reliability and high correspondence with standard interview-derived ASI composite scores.<sup>16,17</sup> The composite scores reported here differ from the standard interview-based ASI composite scores in several ways. Two items in the legal domain on illegal activities and income from such activities were not assessed by the self-administered ASI. In the drug use domain, the self-report composites were based on asking whether or not each of the nine illicit drugs had been used in the past 30 days, whereas the standard interview-based ASI drug composite items ask on how many days in the last 30 days each drug had been used. In addition to the ASI data, patients' VA medical records were queried to gather information regarding their gender, race, age, marital status, and homeless status.

## **Engagement**

To identify qualifying visits and to determine if each patient in the eligible sample satisfied the HEDIS engagement criteria, data used to calculate engagement (i.e., DRG categories, ICD-9-CM diagnosis codes, CPT codes, and ICD-9 procedure codes) were extracted from the VA National Patient Care Database. Technical specifications for the engagement indicator are available at <http://www.washingtoncircle.org> and <http://www.ncqa.org/>.

## **Missing data**

Of the 2,789 patients with baseline ASI data, 1,820 (65.3%) provided follow-up ASI data, an average of 7.3 (SD=2.4) months later. There is a rich but little-known literature on inference in the presence of missing data and particularly in the context of missing follow-up data in longitudinal studies.<sup>18–22</sup> To address the missing follow-up data and potential attrition bias as recommended by this literature, a multiple imputation strategy was used and supplemented by sensitivity analyses.<sup>21–23</sup>

For these analyses, all baseline characteristics and ASI items were used to impute missing ASI items at follow-up. Although patients with complete data differed on several baseline characteristics from those with missing follow-up data (they were younger, less likely to be homeless, and had lower baseline ASI drug and employment and higher ASI psychiatric and legal composite scores), the imputation model adjusts for these differences when making the imputations. Five datasets were imputed with draws from the theoretically complete data distribution using the Multivariate Imputation by Chained Equations program implemented in the R statistical language.<sup>24,25</sup> All analyses were conducted on the imputed datasets and combined into the final statistics and inferences presented here.

## **Primary outcome and analysis strategy**

Our primary indices of patient outcomes were changes from intake to follow-up on ASI alcohol, drug, and legal composite scores, with positive scores indicating improvement. For several reasons, the use of improvement scores as outcomes was favored over using follow-up scores with baseline scores as covariates, most importantly because the change score analyses met the fundamental assumption of mixed-effect regression regarding the distribution of errors whereas a follow-up score analyses would have severely violated this assumption. Improvement scores were outcomes in unadjusted mixed-effects regression models with engagement (yes–no) as the predictor. Adjusted models additionally included baseline values of the ASI composite scores (except the dependent variable) and demographic characteristics as covariates. A random effect for SUD treatment program was included in all models to address potential nonindependence of patient observations.

## Results

The baseline characteristics of the entire sample of patients and by subsequent engagement status are presented in Table 1. The sample was mostly male with an average age of 47.7 years and prominent medical and employment problems. Patients' self-reported race–ethnic identity was 63% Caucasian, 22% African American, 2.3% Hispanic, and 12.7% in other less frequent classifications (e.g., Native American, Pacific Islander) or not reported. The patients who subsequently engaged in treatment were more likely to be African American and homeless, and less likely to be Caucasian than those who did not engage. Furthermore, those who engaged had more severe baseline ASI alcohol, psychiatric, and legal composite scores than nonengagers.

### Association between engagement and improvement in ASI alcohol scores

ASI alcohol improvement had a mean (SD) of 0.20 (0.30) and a range from  $-0.94$  to  $0.93$ . The unstandardized mixed-effects regression coefficients for the model predicting improvement in ASI alcohol score are presented in Table 2. In the unadjusted model, patients who did not engage (represented by the intercept) improved 0.18 points (95% CI 0.15 to 0.20). Patients who met the engagement criteria improved on average an additional 0.03 points (95% CI 0.00 to 0.05). In the model adjusting for clinical and demographic baseline characteristics, patients who did not engage improved 0.01 points (95% CI 0.05 to 0.14); those who met the engagement criteria improved on average an additional 0.04 points (95% CI 0.01 to 0.06).

To test the robustness of our results in various contexts and using different specifications, several sensitivity–exploratory analyses were conducted. First, the analyses were repeated separately in outpatients and inpatient–residential patients. Outpatients included individuals treated in programs classified as outpatient, intensive outpatient, or methadone. Inpatients included individuals treated in programs classified as inpatient, residential, or SUD domiciliary. Patients were classified by the

**Table 1**

Baseline characteristics by subsequent engagement status for 2,789 veterans assessed within 14 days of beginning a new SUD care episode

Characteristic	Total ( <i>N</i> =2,789)	Engaged ( <i>N</i> =1,463)	Not engaged ( <i>N</i> =1,326)
Male	92.1%	91.5%	92.9%
African American**	22.0%	24.3%	19.5%
Hispanic	2.3%	2.7%	1.7%
Caucasian**	63.0%	60.3%	66.4%
Married	18.8%	18.0%	19.6%
Homeless**	6.2%	7.0%	5.0%
Age	47.7 (8.9)	47.6 (8.7)	47.8 (9.1)
ASI alcohol*	0.39 (0.30)	0.40 (0.31)	0.38 (0.30)
ASI drug	0.18 (0.18)	0.19 (0.18)	0.18 (0.18)
ASI psychiatric*	0.36 (0.26)	0.37 (0.26)	0.35 (0.26)
ASI medical	0.44 (0.37)	0.44 (0.36)	0.45 (0.37)
ASI legal **	0.19 (0.30)	0.22 (0.32)	0.17 (0.29)
ASI employment	0.68 (0.29)	0.68 (0.29)	0.69 (0.29)
ASI family–social	0.26 (0.24)	0.26 (0.24)	0.26 (0.24)

Means (SD) unless otherwise identified as percentages

\* $p < 0.05$  difference between engaged and not engaged patients is significant; \*\* $p < 0.01$  difference between engaged and not engaged patients is significant

**Table 2**

Engagement and patient baseline characteristics predicting improvement in ASI alcohol composite at 7-month follow-up

<b>Mixed-effects regression coefficients (95% confidence intervals)</b>						
	<b>All patients (N=2,789)</b>		<b>Inpatients only (N=1,251)</b>		<b>Outpatients only (N=1,536)</b>	
	<b>Unadjusted</b>	<b>Adjusted</b>	<b>Unadjusted</b>	<b>Adjusted</b>	<b>Unadjusted</b>	<b>Adjusted</b>
Intercept	0.23 (0.20, 0.25)**	0.14 (0.10, 0.19)**	0.27 (0.24, 0.31)**	0.14 (0.07, 0.22)*	0.20 (0.16, 0.22)**	0.13 (0.08, 0.19)**
Met engagement criteria	0.03 (0.00, 0.05)*	0.03 (0.01, 0.06)*	0.03 (-0.01, 0.06)	0.02 (-0.02, 0.05)	0.05 (0.02, 0.08)*	0.05 (0.02, 0.09)*
Age*10		0.00 (-0.01, 0.02)		0.02 (-0.00, 0.04)		0.00 (-0.02, 0.02)
Male		0.10 (.06, 0.14)**		0.12 (.05, 0.19)**		0.08 (0.02, 0.13)*
African American		-0.04 (-0.07, -0.01)*		-0.04 (-0.08, 0.01)		-0.05 (-0.08, -0.01)*
Married		-0.01 (-0.04, 0.02)		-0.00 (-0.05, 0.05)		-0.01 (-0.05, 0.02)
Homeless		0.00 (-0.04, 0.05)		-0.04 (-0.12, 0.04)		0.04 (-0.02, 0.10)
Baseline ASI drug		-0.01 (-0.08, 0.06)		-0.04 (-0.14, 0.06)		0.02 (-0.09, 0.10)
Baseline ASI psychiatric		0.14 (.09, 0.19)**		0.18 (.11, 0.27)**		0.11 (0.05, 0.19)**
Baseline ASI legal		-0.05 (-0.08, -0.01)*		-0.05 (-0.10, 0.01)		-0.04 (-0.09, 0.00)
Baseline ASI medical		0.00 (-0.03, 0.03)		0.01 (-0.04, 0.06)		0.00 (-0.04, 0.04)
Baseline ASI employment		-0.05 (-0.09, 0.01)		-0.09 (-0.15, 0.02)*		-0.03 (-0.08, 0.02)
Baseline ASI family-social		0.09 (0.04, 0.14)**		0.04 (-0.04, 0.12)		0.13 (0.07, 0.21)**

There were no significant relationships for engagement in models with alcohol abstinence or change in days of intoxication as the outcome or when limiting the sample to patients with an alcohol use disorder diagnosis at baseline. Time between the beginning of the episode and baseline assessment, ranging from 1 to 14 days, was not a significant covariate in these models. Intraclass correlations for alcohol change scores were 0.095, 0.115, and 0.066 for all programs, inpatient-residential programs, and outpatient programs, respectively. All numeric covariates were mean-centered.

\* $p < 0.05$ ; \*\* $p < 0.01$

type of program in which they completed their baseline assessments. As presented in Table 2, the effect of engagement was only significant in the sample receiving care in outpatient settings. In the adjusted outpatient model, patients who did not engage improved 0.09 points (95% CI 0.03 to 0.14) and patients who met the engagement criteria improved on average 0.06 points more (95% CI 0.02 to 0.09).

Analyses also were conducted with more concrete and clinically transparent outcomes, such as change in days of drinking to intoxication and alcohol abstinence at follow-up. Additionally, these and the main analyses were rerun limiting the sample to patients with an alcohol use disorder diagnosis at baseline. Although the direction and magnitude of the engagement effects mirrored those in the primary analyses, the confidence intervals included zero, perhaps reflecting the reduced reliability of the single-item outcomes and the reduction in sample size.

### **Association between engagement and improvement in ASI drug scores**

Improvement on the ASI drug composite had a mean (SD) of 0.09 (0.16) and ranged from -0.52 to 0.91. The unstandardized mixed-effects regression coefficients for the model predicting improvement in ASI drug score are presented in Table 3. In the unadjusted model, patients who did not engage (represented by the intercept) improved 0.08 points (95% CI 0.06 to 0.09). Patients who met the engagement criteria improved on average 0.02 additional points (95% CI 0.00 to 0.03). In the model adjusting for clinical and demographic baseline characteristics, patients who did not engage improved 0.06 points (95% CI 0.03 to 0.08). Patients who met the engagement criteria, on average, improved 0.02 points more (95% CI 0.00 to 0.03).

To examine the sensitivity of our findings, the analyses were repeated separately in outpatients and inpatient-residential patients (presented in Table 3) and additional analyses used drug abstinence at follow-up as the outcome. Although the direction and magnitude of the engagement effects mirrored those in the primary analyses, the reduced reliability of the outcomes and the reduction in sample size increased the width of the confidence intervals in the adjusted models to cover zero.

### **Association between engagement and improvement in ASI legal scores**

ASI legal improvement had a mean (SD) of 0.05 (0.31) and a range from -1.00 to 1.00. The unstandardized mixed-effects regression coefficients for the models predicting improvement in ASI legal scores are presented in Table 4. In the unadjusted model, patients who did not engage improved 0.03 points (95% CI 0.01 to 0.06). Patients who engaged, on average, improved 0.04 points more (95% CI 0.01 to 0.06). In the model adjusting for clinical and demographic baseline characteristics, patients who did not engage were functioning better at follow-up by an average of 0.035 points (95% CI -0.13 to 0.08). Patients who met the engagement criteria, on average, improved 0.043 points more (95% CI 0.02 to 0.07).

The analyses were repeated separately in outpatient and inpatient-residential patient samples. As presented in Table 4, the effect of engagement was only significant in the sample receiving care in outpatient settings. In the adjusted outpatient model, patients who did not engage improved 0.04 points (95% CI 0.00 to 0.07) and patients who met the engagement criteria improved an average of 0.04 points more (95% CI 0.00 to 0.09).

### **Translating effects on change in ASI into a clinically meaningful metric**

ASI scores (and changes in them) are opaque in terms of clinical meaning. Although there was no effect of engagement on clinically transparent outcomes, such as change in days of intoxication, the relationships between the ASI composite scores and these metrics can be used to translate the ASI effects into more clinically meaningful terms. To do this, the association between the ASI

**Table 3**

Engagement and patient baseline characteristics predicting improvement in ASI drug composite at 7-month follow-up

<b>Mixed-effects regression coefficients (95% confidence intervals)</b>						
	<b>All patients (N=2,789)</b>		<b>Inpatients only (N=1,251)</b>		<b>Outpatients only (N=1,536)</b>	
	<b>Unadjusted</b>	<b>Adjusted</b>	<b>Unadjusted</b>	<b>Adjusted</b>	<b>Unadjusted</b>	<b>Adjusted</b>
Intercept	0.10 (0.09, 0.11)**	0.08 (0.05, 0.10)*	0.12 (0.10, 0.14)**	0.10 (0.06, 0.14)**	0.08 (0.07, 0.10)**	0.06 (0.03, 0.08)*
Met engagement criteria	0.02 (0.00, 0.03)*	0.01 (0.00, 0.03)*	0.02 (0.00, 0.05)*	0.02 (-0.00, 0.04)	0.02 (0.00, 0.03)*	0.02 (-0.00, 0.03)
Age*10		-0.01 (-0.04, 0.02)*		-0.03 (-0.04, -0.02)*		-0.03 (-0.04, -0.02)*
Male		0.02 (-0.00, 0.04)		0.01 (-0.03, 0.05)		0.02 (-0.01, 0.04)
African American		0.05 (0.04, 0.07)**		0.05 (0.03, 0.08)**		0.05 (0.04, 0.07)**
Married		-0.01 (-0.03, 0.00)		-0.01 (-0.04, 0.02)		-0.01 (-0.03, 0.00)
Homeless		0.02 (-0.00, 0.05)		0.03 (-0.01, 0.07)		0.02 (-0.01, 0.05)
Baseline ASI alcohol		-0.00 (-0.03, 0.02)		-0.00 (-0.03, 0.03)		-0.00 (-0.04, 0.02)
Baseline ASI psychiatric		0.05 (0.02, 0.08)*		0.07 (0.03, 0.12)**		0.03 (-0.00, 0.07)
Baseline ASI legal		0.00 (-0.02, 0.03)		0.01 (-0.03, 0.05)		0.00 (-0.03, 0.05)
Baseline ASI medical		0.02 (0.00, 0.04)*		0.02 (-0.01, 0.05)		0.02 (-0.02, 0.03)
Baseline ASI employment		-0.03 (-0.05, -0.01)*		-0.03 (-0.07, 0.00)		-0.03 (-0.05, -0.00)*
Baseline ASI family-social		0.10 (0.07, 0.14)**		0.09 (0.05, 0.14)**		0.11 (0.07, 0.15)**

There were no significant relationships for engagement in models with drug abstinence as the outcome or when limiting the sample to patients with a drug use disorder diagnosis at baseline. Intraclass correlations for drug change scores were 0.069, 0.066, and 0.046 for all programs, inpatient-residential programs, and outpatient programs, respectively. All numeric covariates were mean-centered.

\* $p < 0.05$ ; \*\* $p < 0.01$

**Table 4**

Engagement and patient baseline characteristics predicting improvement in ASI legal composite at 7-month follow-up

<b>Mixed-effects regression coefficients (95% confidence intervals)</b>						
	<b>All patients (N=2,789)</b>		<b>Inpatients only (N=1,251)</b>		<b>Outpatients only (N=1,536)</b>	
	<b>Unadjusted</b>	<b>Adjusted</b>	<b>Unadjusted</b>	<b>Adjusted</b>	<b>Unadjusted</b>	<b>Adjusted</b>
Intercept	0.03 (0.00, 0.05)*	0.03 (-0.01, 0.08)	0.01 (-0.02, 0.05)	0.07 (-0.00, 0.15)	0.04 (0.00, 0.07)*	0.02 (-0.05, 0.09)
Met engagement criteria	0.04 (0.01, 0.06)*	0.04(0.02, 0.07)**	0.02 (-0.02, 0.05)	0.02 (-0.02, 0.06)	0.04 (0.00, 0.09)*	0.05 (0.01, 0.09)**
Age* 10		-0.00 (-0.02, 0.01)		-0.01 (-0.03, 0.01)		0.00 (-0.02, 0.02)
Male		-0.01 (-0.05, 0.04)		-0.06 (-0.12, 0.02)		0.02 (-0.04, 0.08)
African American		-0.03 (-0.06, 0.00)		-0.05 (-0.09, 0.00)		-0.02 (-0.06, 0.02)
Married		0.04 (0.00, 0.06)*		0.06 (0.01, 0.10)*		0.02 (-0.02, 0.06)
Homeless		-0.05 (-0.10, 0.00)		-0.06 (-0.13, 0.01)		-0.04 (-0.12, 0.03)
Baseline ASI alcohol		-0.04 (-0.09, -0.01)*		-0.02 (-0.08, 0.03)		-0.06 (-0.12, 0.01)
Baseline ASI drug		-0.08 (-0.15, 0.00)		-0.06 (-0.15, 0.04)		-0.07 (-0.19, 0.06)
Baseline ASI psychiatric		-0.03 (-0.02, 0.03)		-0.07 (-0.15, 0.01)		-0.01 (-0.08, 0.08)
Baseline ASI medical		0.01 (-0.02, 0.05)		0.02 (-0.02, 0.08)		-0.01 (-0.05, 0.05)
Baseline ASI employment		0.04 (-0.00, 0.09)		0.09 (0.03, 0.15)**		0.01 (-0.05, 0.07)
Baseline ASI family-social		0.05 (-0.01, 0.13)		0.04 (-0.02, 0.13)		0.06 (-0.01, 0.17)

Intraclass correlations for legal change scores were 0.041, 0.036, and 0.015 for all programs, inpatient-residential programs, and outpatient programs, respectively. All numeric covariates were mean-centered.

\* $p < 0.05$ ; \*\* $p < 0.01$

composite change scores and the component item(s) were estimated (e.g., predicting improvement in days of intoxication with ASI alcohol change scores). These equations were then used to estimate the average value of the clinically meaningful metric at the mean of the ASI change score and again when adding the effect of engagement. For example, based on our models and the relationship between change in the ASI alcohol composite and change in days of alcohol intoxication, one would expect the average effect of engagement on the alcohol composite (0.025 points) to be associated with 0.62 fewer days of alcohol intoxication.

The average effect of engagement on the drug composite (0.015 points) is estimated to increase the predicted probability of drug abstinence by 0.003 (from 0.775 to 0.778). The average effect of engagement on the legal composite (0.036 points) is estimated to decrease the predicted probability of being incarcerated in the previous 30 days by 0.002 (from 0.048 to 0.046). These estimates emphasize that, although the associations between engagement and ASI improvement scores are statistically significant, they are clinically modest.

### **Exploratory analyses on the effects of initiation on outcomes**

Because the vast majority of our sample met the HEDIS initiation criterion (see “Introduction”), this study focused on the association between meeting the engagement criterion and outcomes. However, exploratory analyses on the association between meeting the initiation criterion and outcomes were conducted. In the full sample, initiation had a small but significant relationship to improvement in both the alcohol and drug ASI composites, but the effects became nonsignificant with case-mix adjustment. No effect of initiation was found when the sample was stratified by treatment setting or with improvement in the legal ASI composite as the outcome. These largely null results should not be given much weight because of the limited variability on initiation in our sample.

### **Additional sensitivity analyses**

The extent to which our results were robust to choice of sample definition, method for addressing missing data, and parametric assumptions of statistical procedures was examined. For reasons previously outlined, only patients who had a qualifying visit within 14 days of completing the baseline assessment were included in the primary sample. The analyses were repeated with two variations of this definition: (1) changing the sample to include patients who had a qualifying visit within 7 days of the baseline assessment ( $N=2,022$ ); (2) changing the sample to include patients who completed the baseline on the same day as a qualifying visit ( $N=1,049$ ). The point estimates for the effects of interest were virtually identical to the main results, although the confidence intervals were wider and subsumed zero reflecting the reduced sample sizes.

The analyses using different strategies to address missing follow-up data were also repeated. Although it is plausible that the missing at random (MAR) assumption of the multiple imputation procedure was adequately satisfied and evidence exists that this strategy’s performance is often good even when the MAR assumption is not fully satisfied, the data were reanalyzed using two different missing data methods that have other limitations but do not require the MAR assumption. Specifically, pattern mixture modeling<sup>18,19</sup> and a selection model were used.<sup>26</sup> Both of these analyses yielded substantively identical results to our main analyses.

## **Conclusions**

### **Scope of inference**

An important group of patients was not included in the Outcomes Monitoring Project or by proxy in the current study: patients identified with SUDs who did not seek treatment. The rate of engagement in

the current sample of patients entering treatment was over 50%, whereas the national rate of engagement in the VA health care system is about 9%. This difference highlights an important limitation to the scope of inference in our study: These results apply to VA patients who enter new episodes of SUD treatment. They do not address the relationship of engagement to outcomes relative to outcomes of patients who have never sought treatment or those who are ineligible because they are involved in treatment often enough not to have a 60-day service-free period. The extent to which engagement is associated with outcomes in the population of all identified VA SUD patients is unknown.

### **Importance of distinguishing patient-level and facility-level effects**

In evaluating the association between process-of-care indicators and subsequent outcome quality measures, it is critical both conceptually and statistically to distinguish patient-level and facility-level effects.<sup>27</sup> A recent study examined the extent to which VA medical centers' performance on the initiation and engagement measures were associated with the effectiveness of their specialty substance abuse treatment services after adjusting for patient case-mix. Higher initiation rates were not associated with facility-level average improvement in ASI alcohol composite scores but were very modestly associated with greater improvements on ASI drug composite scores. Facility rates of engagement were not associated with facility averages in patient improvement.<sup>9</sup>

The assumption that relationships observed for groups should necessarily hold for individuals (or visa versa) is so common that it has a name: the "ecological fallacy."<sup>28</sup> In the context of the present study, it may seem incongruous to some that the association between a quality indicator and outcomes may be significant at the patient level but not at the facility level. However, the sample criteria, sample sizes, hypotheses, and scopes of inference are very different at the two levels. In patient-level analyses, such as those presented in the current study, the sample consists of individuals who are eligible to meet the criteria and are then tracked. As noted above, for patient-level analyses such as these, the sampling frame is all identified SUD patients with a 60-day service-free period, not all SUD patients. Furthermore, individual-level constructs often are very different and have different confounders, once aggregated to the facility level.<sup>29</sup>

Contrast the person-level analysis described above by considering rates of engagement as characteristics of facilities. Do facilities with higher rates of engagement produce better average outcomes for *all* their patients with SUDs? Put differently, should administrators, clinical managers, and consumers assume that facilities having higher engagement rates also produce better average outcomes for all patients, whether or not they meet (or are eligible to meet) the engagement criteria? This question differs conceptually and statistically from the patient-level question. Facility rates of engagement are calculated with all identified SUD patients with a 60-day service-free period but often are assumed to gauge the quality of treatment for all patients.

Individual-level constructs often have very different meaning and correlates once aggregated to the facility level. Individual-level correlations are often used to define and operationalize quality indicators, yet quality indicators are used to measure facility-level performance. The underlying assumption that individual-level associations will hold at the facility level needs to be checked, as done recently for Medicare's Hospital Compare performance measures.<sup>30</sup> Although the constituent processes of care of these measures had been linked to clinical outcomes, including reduced mortality at the individual level, when combined and aggregated to the facility level, the relationships with outcomes were much weaker or nonexistent. Therefore, when asking patient-level questions and recommending provider practices, one needs to consult the patient-level evidence, such as the results of the current study. However, one of the primary purposes of quality indicators is to distinguish facility-level quality.<sup>4,31</sup> Performance measures often are designed to both guide the care of individual patients and discriminate between high and low performing facilities. Our results provide some support for the former reason, but other analyses<sup>9</sup> do not support the latter use of the engagement measure.

## Outpatient and inpatient–residential treatment

In the adjusted models, the associations between engagement and improvement in alcohol and legal ASI composites were found to be significant in the outpatient sample but not in the inpatient–residential sample. Engagement is operationalized differently for these samples (two outpatient SUD visits within 30 days of a first outpatient versus from discharge from inpatient–residential treatment). However, it is unknown to what extent the different engagement definitions are responsible for the different results.

The following scenario illustrates one possible explanation for the weaker results in inpatient–residential settings. Consider two patients who both received 30 days of very successful residential treatment. One lives close to the VA medical center and the other lives 150 miles away. Because of the proximity of continuing care, the first attends two outpatient follow-up visits in the postdischarge month and thereby meets the engagement criteria. The second patient decides he is doing well enough to skip the aftercare appointments 150 miles away or seeks care outside the VA system, thereby failing to engage. The difference in days during which care was received is small (i.e., 32 vs. 30) and the magnitude of difference in hours of care is even less. Generalizing from this scenario, one would not be surprised to find no or very small outcome differences between engaged and unengaged patients in a residential sample. Changing the operationalization of initiation and engagement for inpatients and residential patients (i.e., requiring a minimum length of stay for initiation and allowing engagement before discharge, if the stay is sufficiently long) might enhance the association between these measures and clinical outcomes and make the association comparable to those found in outpatient samples.

The overall effect of engagement, though significant, was smaller for the drug composite compared to the alcohol and legal composites; when the overall sample was divided into inpatient–residential and outpatient, the magnitude of the engagement effect was not significant in either. Because of the small overall effect size and the reduced power of the stratified analyses, the proper interpretation of the results is that engagement was very modestly associated with improvement in the drug composites when the data were pooled across all settings.

## Limitations

While it is tempting to think of the engagement–outcome link in terms of causation, unobserved variables are likely to drive both engagement and outcomes but to an unknown extent. Even if engagement does cause better outcomes, estimation of treatment effects from observational data is difficult because exposure to treatment is nonrandom and often confounded with patient characteristics. Statistical controls for many patient characteristics were used, but data were not available on several potentially confounding variables, such as motivation and 12-step group involvement. Clearly, a statistical or design (randomization) control for motivation would greatly increase our confidence that the association between engagement and symptom improvement is causal rather than an artifact of self-selection. Similarly, patients who are more likely to engage in formal treatment also might be more likely to become involved in 12-step or other self-help groups. Studies that have measured both kinds of care often have found no association between continuing formal care once the effects of 12-step involvement are included.<sup>32</sup>

Also, generalization of these findings to facilities outside the VA system should be made with caution. There may be relevant case-mix factors (e.g., average age, proportion of women patients, intake symptom severity) that distinguish facilities in the VA system from those in other health care systems. VA also is more integrated, so transitions from inpatient to outpatient settings are made more reliably than in nonintegrated care. These differences may affect the relationships between meeting the criteria and clinical outcomes. The results obtained here hopefully will motivate similar analyses outside the VA system to determine if or to what extent these findings generalize.

The results should be interpreted in recognition of the time between measurement points. Average length of follow-up was about 7 months, whereas the engagement indicator is focused on the first month or so of treatment. A number of factors, including many not associated with treatment, could have intervened between these two time points to affect patient outcomes. Assessing patient outcomes closer to the end of treatment would arguably produce a better indicator of treatment effectiveness.

### **Future directions**

Our current research is focused on examining alternative specifications of the HEDIS measures to determine if associations with patient- and facility-level outcomes can be boosted. For example, engagement could be refined to include more than two visits in the month after initiation. Engagement could be changed from a dichotomous variable to a count of visits in the month after initiation. Currently, the HEDIS specifications rely on combinations of CPT codes and primary or secondary diagnosis codes to identify SUD care within claims and administrative data. It is possible that this system captures care that is not related to SUD, a technical question that is currently being addressed. Perhaps limiting the definition of SUD care to include only those visits for which an SUD is the primary diagnosis would increase the associations between the performance measures and outcomes.

## **Implications for Behavioral Health**

Performance indicators that assess important processes of care, that are linked to downstream patient outcomes, and that can be feasibly implemented across health care systems are critical to policy makers, administrators, and clinical managers for monitoring and improving care. The HEDIS engagement indicator for measuring the quality of SUD treatment has been widely adopted, largely in absence of evidence linking it to patient outcomes. The study presented here is the first to examine the strength of association between meeting the engagement criterion and patient-level changes in alcohol, drug, and legal symptoms in a nationwide health care system that includes both outpatients and patients treated in inpatient–residential programs.

Performance measures are designed and implemented with the intention of improving clinical outcomes, so research documenting this link, or lack thereof, is critical.<sup>4</sup> Ideally, performance measures would assess processes of care that are associated with outcomes at both the patient and facility levels, but, as noted earlier, this often is not the case.<sup>5,8,29</sup> In this study of VA patients entering SUD treatment, those meeting the engagement criteria experienced significantly more improvement in ASI alcohol, drug, and legal composite scores compared to patients who did not engage. The effect of engagement on improvement in the ASI alcohol and legal composites was of greater magnitude in outpatient settings compared to inpatient settings. This evidence linking the HEDIS engagement quality indicator to several clinically meaningful patient outcomes adds to the literature documenting the clinical benefits of treatment entry and engagement. The results support and extend a previous study in a statewide sample of outpatients linking patient-level engagement to better legal outcomes in some domains.<sup>14</sup> The benefits accrued by those who engaged in the current study were statistically significant yet clinically modest. These results only indirectly support the use of the HEDIS engagement measure for its intended purpose—discriminating quality at the facility level—but confirm that the processes of care that compose the measure are modestly associated with important patient outcomes.

How are these data useful to administrators and clinical managers? This and other studies of the HEDIS measures<sup>9,14</sup> provide an evidentiary basis for all health care system stakeholders to judge the appropriate uses of these performance measures. Patients, clinicians, and administrators benefit from knowing that patients who meet the engagement criterion have modestly more improvement

than patients who do not. They would benefit from remembering that this relationship is not necessarily causal. Also, knowing that, at least within the VA, facility-level rates of engagement are weakly or not related to facility-level outcomes<sup>9</sup> should temper administrators' enthusiasm for using these measures in pay-for performance systems or consumers' inclination to use them for choosing treatment facilities.<sup>4</sup>

## Acknowledgements

This study was funded by the Center for Substance Abuse Treatment (Contract 270-02-7120), and the VA Office of Research and Development Health Services Research and Development Service (grants no. SUS 99-015, MRP-05-168-1, IIR-07-092-1). Keith Humphreys is a member of the Washington Circle and Alex Harris is a member of the Washington Circle Public Sector Workgroup. The opinions expressed are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs or the Center for Substance Abuse Treatment.

## References

1. AcademyHealth. Glossary of terms commonly used in health care. <http://www.academyhealth.org/publications/glossary.pdf>. Accessed April 22, 2008.
2. Garnick D, Horgan C, Chalk M. Performance Measures for Alcohol and Other Drug Services. National Institute of Alcohol Abuse and Alcoholism. <http://pubs.niaaa.nih.gov/publications/arh291/19-26.htm>. Accessed September 12, 2006.
3. Donabedian A. *The Definition of Quality and Approaches to Assessment*. Ann Arbor: Health Administration Press; 1980.
4. Horn SD. Performance measures and clinical outcomes. *JAMA*. 2006;296(22):2731-2732. doi:10.1001/jama.296.22.2731.
5. Hofer TP, Hayward RA, Greenfield S, et al. The unreliability of individual physician "report cards" for assessing the costs and quality of care of a chronic disease. *JAMA*. 1999;281(22):2098-2105. doi:10.1001/jama.281.22.2098.
6. Department of Veterans Affairs Office of Quality and Performance VA/DOD. Clinical Practice Guideline for the Management of Substance Use Disorders: Guideline Summary. <http://www.oqp.med.va.gov/cpg/SUD/G/SUD%20Specialty%20Summary.pdf>. Accessed July 26, 2005.
7. National Committee for Quality Assurance. *HEDIS 2006 Volume 2: Technical Specifications*. Washington, DC: National Committee for Quality Assurance; 2006.
8. Harris A, Humphreys K, Bowe T, et al. Measuring the quality of substance use disorder treatment: Assessing the VA continuity of care quality measure. *Journal of Substance Abuse Treatment*. in press.
9. Harris AH, Humphreys K, Finney JW. Veterans Affairs facility performance on Washington Circle indicators and case mix-adjusted effectiveness. *Journal of Substance Abuse Treatment*. 2007;33(4):333-339. doi:10.1016/j.jsat.2006.12.015.
10. Greenberg A, Rosenheck R, Seibyl C. Continuity of care and clinical effectiveness: Outcomes following residential treatment for severe substance abuse. *Medical Care*. 2002;40(3):246-259. doi:10.1097/00005650-200203000-00008.
11. Greenberg GA, Rosenheck RA. Continuity of care and clinical outcomes in a national health system. *Psychiatric Services*. 2005;56(4):427-433. doi:10.1176/appi.ps.56.4.427.
12. Rost K, Dickinson LM, Fortney J, et al. Clinical improvement associated with conformance to HEDIS-based depression care. *Mental Health Services Research*. 2005;7(2):103-112. doi:10.1007/s11020-005-3781-1.
13. McCorry F, Garnick DW, Bartlett J, et al. Developing performance measures for alcohol and other drug services in managed care plans. *Joint Commission on Quality Improvement*. 2000;26:633-643.
14. Garnick DW, Horgan C, Lee M, et al. Are Washington Circle performance measures associated with decreased criminal activity following treatment. *Journal of Substance Abuse Treatment*. 2007;33(4):341-352. doi:10.1016/j.jsat.2007.03.002.
15. Tiet QQ, Byrnes HF, Barnett P, et al. A practical system for monitoring the outcomes of substance use disorder patients. *Journal of Substance Abuse Treatment*. 2006;30(4):337-347. doi:10.1016/j.jsat.2006.03.002.
16. McLellan A, Kushner H, Metzger D, et al. The fifth edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment*. 1992;9:199-213. doi:10.1016/0740-5472(92)90062-S.
17. Rosen CS, Henson BR, Finney JW, et al. Consistency of self-administered and interview-based Addiction Severity Index composite scores. *Addiction*. 2000;95(3):419-425. doi:10.1046/j.1360-0443.2000.95341912.x.
18. Hedeker D, Gibbons RD. Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*. 1997;2(1):64-78. doi:10.1037/1082-989X.2.1.64.
19. Little RJA. Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*. 1993;88:125-133. doi:10.2307/2290705.
20. Rubin DB. Inference and missing data. *Biometrika*. 1976;63:581-592. doi:10.1093/biomet/63.3.581.
21. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods*. 2002;7(2):147-177. doi:10.1037/1082-989X.7.2.147.
22. Schafer JL. *Analysis of Incomplete Multivariate Data*. New York: Chapman & Hall; 1997.
23. Daniels M, Hogan J. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. New York: Chapman & Hall; 2008.

24. R Foundation for Statistical Computing. *R: A Language and Environment for Statistical Computing [computer program]*. Vienna: R Foundation for Statistical Computing; 2005.
25. van Buuren S, Oudshoorn CGM. *Multivariate Imputation by Chained Equations. R package version 1.14* [computer program]. 2005.
26. Glynn RJ, Laird NM, Rubin DB. Selection modeling versus mixture modeling with nonignorable nonresponse. In: Wainer H, ed. *Drawing Inferences from Self-selected Samples*. Mahwah: Lawrence Erlbaum Associates Publishers; 1986:115–142.
27. McCarty D. Performance measurement for systems treating alcohol and drug use disorders. *Journal of Substance Abuse Treatment*. 2007;33(4):353–354. doi:10.1016/j.jsat.2007.04.002.
28. Robinson W. Ecological correlations and the behavior of individuals. *American Sociological Review*. 1950;15:351–357. doi:10.2307/2087176.
29. Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *International Journal of Epidemiology*. 1989;18(1):269–274. doi:10.1093/ije/18.1.269.
30. Werner RM, Bradlow ET. Relationship between Medicare’s hospital compare performance measures and mortality rates. *JAMA*. 2006;296(22):2694–2702. doi:10.1001/jama.296.22.2694.
31. McLellan A, Chalk M, Bartlett J. Outcomes, performance, and quality—what’s the difference. *Journal of Substance Abuse Treatment*. 2007;32:331–340. doi:10.1016/j.jsat.2006.09.004.
32. Witbrodt J, Bond J, Kaskutas LA, et al. Day hospital and residential addiction treatment: randomized and nonrandomized managed care clients. *Journal of Consulting and Clinical Psychology*. 2007;75(6):947–959. doi:10.1037/0022-006X.75.6.947.