Assessing Quality of Diabetes Care by Measuring Longitudinal Changes in Hemoglobin A1c in the Veterans Health Administration

Wes Thompson, Hongwei Wang, Minge Xie, John Kolassa, Mangala Rajan, Chin-Lin Tseng, Stephen Crystal, Quanwu Zhang, Yehuda Vardi, Leonard Pogach, and Monika M. Safford

Context. A1c levels are widely used to assess quality of diabetes care provided by health care systems. Currently, cross-sectional measures are commonly used for such assessments.

Objective. To study within-patient longitudinal changes in A1c levels at Veterans Health Administration (VHA) facilities as an alternative to cross-sectional measures of quality of diabetes care.

Design. Longitudinal study using institutional data on individual patient A1c level over time (October 1, 1998–September 30, 2000) with time variant and invariant covariates.

Setting. One hundred and twenty-five VHA facilities nationwide, October 1, 1998–September 30, 2000.

Patients. Diabetic veteran users with A1c measurement performed using National Glycosylated Hemoglobin Standardization Project certified A1c lab assay methods.

Exposures. Characteristics unlikely to reflect quality of care, but known to influence A1c levels, demographics, and baseline illness severity.

Main Outcome Measure. Monthly change in A1c for average patient cared for at each facility.

Results. The preponderance of facilities showed monthly declines in within-patient A1c over the study period (mean change of -0.0148 A1c units per month, range -0.074 to 0.042). Individual facilities varied in their monthly change, with 105 facilities showing monthly declines (70 significant at .05 level) and 20 showing monthly increases (5 significant at .05 level). Case-mix adjustment resulted in modest changes (mean change of -0.0131 case-mix adjusted A1c units per month, range -0.079 to 0.043). Facilities were ranked from worst to best, with attached 90 percent confidence intervals. Among the bottom 10 ranked facilities, four remained within the bottom decile with 90 percent confidence.

Conclusions. There is substantial variation in facility-level longitudinal changes in A1c levels. We propose that evaluation of change in A1c levels over time can be used as

a new measure to reflect quality of care provided to populations of individuals with chronic disease.

Key Words: Diabetes, longitudinal analysis, quality of care, glycemic control, performance measures

Improving the quality of health care is a fundamental concern in the Institute of Medicine's landmark 2001 report *Crossing the Quality Chasm* (Committee on Quality Health Care in America, Institute of Medicine 2001). Central to this are measures used to evaluate quality of care. Current efforts at measuring quality-of-care focus primarily on processes of care because comparisons of receipt of processes are relatively straightforward. With diabetes patients, for example, glycosylated hemoglobin (A1c) testing is a central component of the widely implemented Diabetes Quality Improvement Project performance measures. Intermediate health outcomes are also desirable for use in such assessments because they represent a more distal effect of care and are more directly linked to risks of outcomes than processes of care. One such intermediate health outcome for diabetes patients is A1c level, which has been associated with microvascular complications (Diabetes Control and Complications Trial Research Group 1993; U.K. Prospective Diabetes Study [UKPDS] Group 1998). Especially as they can be lowered with medications, A1c levels are considered evidence-based performance measures for quality-of-care assessments. Consequently, A1c testing is recommended at least annually for all diabetes patients and therefore provides a readily available pool of information.

However, A1c values are also related to patient-level factors extraneous to health care systems (Chin et al. 2000), complicating attribution of less-than-optimal A1c values to lapses in quality of care. Indeed, even in the high-quality

The first three authors contributed equally to this paper.

Address correspondence to Monika M. Safford, M.D., Deep South Center on Effectiveness at the Birmingham VA Medical Center and, The University of Alabama at Birmingham, MT 643, 1717 11th Avenue South, Birmingham, AL 35294-4410. Wes Thompson, Ph.D., is with the Department of Statistics, University of Pittsburgh, Pittsburgh, PA. Hongwei Wang, Ph.D., Minge Xie, Ph.D., John Kolassa, Ph.D., Yehuda Vardi, Ph.D. (deceased), are with the Department of Statistics Hill Center, Busch Campus Rutgers, The State University of New Jersey, Piscataway, NJ. Mangala Rajan, M.B.A., Chin-Lin Tseng, Dr.P.H., Quanwu Zhang, Ph.D., Leonard Pogach, M.D., M.B.A., are with the Center for Healthcare Knowledge Management, VANJ Healthcare System, East Orange, NJ. Stephen Crystal, Ph.D., is with the Institute for Health, Healthcare Policy, & Aging Research, Rutgers University, New Brunswick, NJ.

environment of landmark clinical trials (Wright et al. 2002), A1c levels rose over time within most individuals. This reflects in part the limitation of currently available treatments and in part other factors such as patient age or severity of diabetes. To account for this, A1c levels can be case-mix adjusted to result in more meaningful quality-of-care assessments (Iezzoni 1997; Zhang et al. 2000). However, case-mix adjustment methodology is not yet well established for intermediate health outcomes such as A1c levels, and most attention has been focused on cross-sectional analyses.

In addition to cross-sectional analyses, another potentially useful strategy would be to evaluate each patient's A1c level over time within a health care system. Multiple measures on individuals over time, termed "longitudinal data," (Diggle, Liang, and Zeger 1996) are crucial for assessing change over time within patients. Longitudinal data enable differentiation of cohort effects (patient population changes over time) from within-subject time effects. For example, more aggressive screening may identify individuals earlier in their disease course. Therefore, serial cross-sectional assessments could result in declining A1c levels that are because of selection biases and not to better diabetes care. By following individuals' glycemic control over time, these sorts of cohort effects can effectively be separated from the effect of the health system. Longitudinal methods may give a better sense of how a patient's health care system is treating the disease over time than do cross-sectional data.

Longitudinal data can be analyzed using the random-effects growth curve model (Bryk and Raudenbush 1992). This method can be used to model an average A1c growth curve for all patients utilizing a given health care system for diabetes care, describing how its patients' A1c values changed on average over time. The deviation of patient-level change in A1c from the average A1c growth curve for a given health care system is accounted for by random effects in the model. These random effects incorporate the correlation of multiple A1c measurements made on each patient (Diggle, Liang, and Zeger 1996). This use of the growth curve methodology is attractive because it should result in meaningful assessments of quality of care as reflected in how health care systems influence patient A1c levels over time, or in other words, the patient response to the clinical setting.

We studied the longitudinal growth curve model as a new approach to evaluating quality of care among diabetic veterans in 125 of the Veterans Health Administration's (VHA) facilities nationwide. We hypothesized that some facilities would be more or less successful at controlling A1c levels over time, as reflected in facility-level average monthly A1c growth curves over

2 years. We also examined the importance of case-mix adjustment in the longitudinal context.

METHODS

Data Sources and Study Population

We used data from the Healthcare Data Analysis Information Group in Milwaukee, Wisconsin for pharmacy and laboratory results and the central data repository for the VHA in Austin, TX for demographics and inpatient and outpatient utilization with associated International Classifications of Diseases, 9th edition (ICD-9) codes. Veterans who used the VHA in Fiscal Year 1999 (FY99) or Fiscal Year 2000 (FY00) and in at least 1 year received a diabetes medication, had an ICD-9 code 250.xx (diabetes) associated with more than one outpatient encounter, or had any inpatient encounter were considered to have diabetes (Hebert et al. 1999; Health Plan Employer Data and Information Set [HEDIS[®]] 2004). This identified 1,476,921 A1c tests on 486,558 patients.

Hemoglobin A1c is the percentage of hemoglobin with an attached glucose moiety. As red blood cells, the repositories of hemoglobin, have an average lifespan of 120 days, A1c values reflect average ambient glucose values over the past 120 days, although the last 30 days' average glucose disproportionately determines the measured A1c (Tahara and Shima 1995). Thus, we only included A1c values for each individual that were separated by at least 30 days, excluding 38,186 tests. We also eliminated 1,481 A1c values that fell outside the plausible physiologic range (3–18 percent).

Considerable lab assay methodology variations for A1c testing have been described within the United States and VHA (Little et al. 2001; Safford et al. 2003). Because the National Glycohemoglobin Standardization Program (NGSP) is currently implementing nationwide standardization, such variation is likely to diminish in the near future. In FY99 and FY00, however, many labs in the United States, including several VHA facilities, were using methods not certified by the NGSP. Other VHA facilities were using multiple methods, some certified and others not. To minimize this variation, we excluded all 618,877 uncertified A1c tests from the analysis.

Many VHA patients receive care at multiple facilities within a given year; for example, 11 percent of the patients in our sample had A1c tests at more than one facility. For evaluation of facility-level differences in quality of care, each patient was assigned a “home facility” for the study period (FY99

and FY00), defined as the facility at which the patient made the most outpatient diabetes-related visits.

Some VHA facilities were reorganized in FY00, some did not provide lab assay method information, and some facilities reported lab data incompletely (Kerr et al. 2002). This resulted in the exclusion from the sample data of 32,342 tests at 28 facilities. Additionally, 19 facilities had fewer than 100 A1c tests over the study period, far fewer than the typical VHA facility. Because of this scarcity of data, it was judged that estimates of growth curves for these facilities would be too unstable to compare reliably to larger VHA facilities. Hence, they and their 1,274 tests were dropped from the analysis.

The final study sample consisted of 284,895 patients from 125 facilities with 816,721 A1c tests. Approval from our institution's Institutional Review Board was obtained.

Variables

The outcome variable of interest was patient A1c level. Available covariates included the number and dates of all A1c lab tests for each individual, lab methodology, and case-mix adjustment variables (which included demographic information, marital status, diabetes severity, and the Charlson index of comorbid illnesses). The Charlson index was selected to capture potential illness-related influences on treatment decisions to achieve glycemic control.

We operationalized two domains of time: "relative time" and "absolute time." Relative time was used to model individual patients' pattern of change in A1c values over the course of the study period. The relative time associated with a patient's given A1c value was the number of months elapsed from the patient's first recorded A1c test during the study period to the given A1c test date. Absolute time was defined as the number of days elapsed from the start of FY99 to the given A1c test date, which was used to model the seasonal population-level fluctuations in A1c that we observed in preliminary analyses (see Figure 1).

We found small but significant differences among the NGSP certified lab methods. Thus, we included A1c lab methodology in the model, with categories denoted by method B (Biorad, Hercules, CA), method T (Tosoh, Tokyo), and other.

Available demographic factors included age, sex, race/ethnicity, and marital status. Age has been associated with A1c levels, with older individuals

Figure 1: Mean Monthly A1c for Veterans Health Administration Veteran Users with Diabetes in Fiscal Year 1999 and Fiscal Year 2000, by Month

having lower values (Harris et al. 1999; Zhang et al. 2000). Therefore, we included age at onset of the study period as a baseline adjuster. Patient sex was also included in the model. Race was categorized as white, African American, Hispanic, Native American, Asian, other, and unknown. Marital status (married or not married) was included because it has been associated with A1c levels in older men (Zhang et al. 2000; Safford et al. 2003) and may reflect social support.

Insulin use has been closely associated with duration and severity of diabetes (UKPDS Group 1998). The patient's first diabetes treatment recorded in the study period (insulin, no insulin, and missing) was used as a measure of baseline diabetes severity.

The Charlson index reflects 1-year mortality risk based on the presence of selected illnesses (Charlson et al. 1987). This index has been adapted for use in administrative data (Deyo, Cherkin, and Ciol 1992), which we used to construct Charlson index scores to capture general illness burden. Charlson score was used as a time invariant measure of comorbidity, similar to other reports (Berlowitz et al. 1998; Chin et al. 2000; Zhang et al. 2000; Safford et al. 2003).

ANALYSIS

The models ascribed to each facility an expected trajectory (“growth curve”) of A1c values for patients utilizing that facility as their home facility. We used the estimated facility-level growth curve slopes to compare outcomes of diabetes care among the facilities, both with and without case-mix adjustment. The hierarchical linear models used are described below; more detailed descriptions are available at www.stat.rutgers.edu/~slothrop/VA/Appendix.htm

Two parts, along with a random error term, give the general framework of the model:

$$\text{A1c level} = [\text{time variant terms}] + [\text{time invariant terms}] + \text{error}$$

As with the usual linear regression models, the error terms were treated as independent, mean-zero normal variables.

The time invariant terms in the model included the certified A1c lab method and the case-mix adjustment variables. The time variant terms included seasonal and growth curve terms. The seasonal term consisted of a linear combination of sine and cosine with a period of 1 year; inclusion of both sine and cosine allowed the starting point of the cycle to be empirically determined. The growth curve term was linear in relative time. Each patient in the study was ascribed an individual growth curve for A1c values over relative time, with coefficients containing both fixed and random components. The fixed component of the growth curve coefficients, which differed by facility, described the pattern of change over time for an average patient in that patient’s home facility. The random component of the growth curve coefficients modeled deviations of the patient-specific growth curve from the facility-level average growth curve. The random effects were treated as independent mean-zero normal variables.

The random-effects model required estimation of random effects for each patient. With 284,895 patients, this entails intensive computational effort. Because of this practical limitation, generalized estimating equations (GEEs) were used to fit the model rather than implementing the random-effects model directly (Liang and Zeger 1986). The GEE approach relieved computing load by avoiding estimation of random effects. For continuous responses such as A1c levels, this approach produces consistent parameter estimates for the fixed effects that are identical to those produced by fitting the random-effects model (Diggle, Liang, and Zeger 1996).

A common term for model evaluation is R^2 , which is interpreted as the proportion of variation explained by the regression model. In the longitudinal context, no generally accepted counterpart of R^2 exists and so we defined “working” R^2 . For each patient, we computed the mean observed A1c value and the mean predicted A1c value. The working R^2 was computed as in the usual regression setting with independent observations, using the patient mean observed and predicted values instead (see Appendix for details).

The facility-specific intercepts of the growth curve model represented initial mean A1c at that facility and the slopes represented mean monthly A1c change. We used the facility slopes to compare diabetes outcomes. A facility with an expected patient A1c trajectory that brought A1c levels into control rapidly may have provided better care than facilities with trajectories that brought levels into control more slowly or not at all. A facility with a more negative slope was therefore considered to be better performing than one whose slope was either less negative or positive. For multiple comparisons, we constructed simultaneous confidence intervals (SCIs) for the facility slopes using the Studentized maximum modulus method (Hsu 1996). We assessed changes in facility slopes between the models with and without case-mix variables.

The facilities were also ranked based on the slope estimates, from most *rapid glycemic improvement*, or steepest negative slope, to least *rapid/glycemic deterioration*. We constructed confidence intervals for the ranks based on a slight modification of the simulation approach of Goldstein and Spiegelhalter (1996) and Hsu (1996) (see Appendix for details).

We also investigated an alternative explanation for the overall declining trend of the A1c values: those with worsening glycemic control preferentially leaving the VA system. For this analysis, we selected patients who were alive at the end of the 2-year period by examining the Beneficiary Identification and Records Locator Subsystem for vital status. We then defined two subgroups of patients: those who likely remained in the system, that is, they utilized services in both years of the observation period ($n = 142,426$); and those who left it (“out-migrators”), that is, they had utilization only in the first year ($n = 51,054$). We then fitted the hierarchical growth curve models as described for the main analysis to each subgroup separately for the first year only. Any difference between the average A1c decline rates in each of the two subgroups would indicate if patients who “out-migrated” were more often experiencing worsening control.

All statistical analyses were performed using SAS version 8.2. The level of significance used for inclusion of covariates was $p < .05$.

RESULTS

Sample Characteristics

Table 1 compares the sample used in this study to the overall population of VHA users with diabetes in FY99 and FY00. The mean age for subjects in the study sample was 64 years and 98 percent were male, 70 percent were white, 61 percent were married, and 31 percent used insulin at entry into the sample. These differed only slightly from the overall VHA diabetes patient population. For both groups, mean A1c levels decreased monotonically with age. Patients who were unmarried, using insulin, or were African American or Hispanic had higher A1c values compared with their counterparts. A1c levels were slightly but significantly higher in the winter than in the summer and monthly A1c values trended downward over the study period (see Figure 1).

Model Estimates

In the growth curve models, the facility slopes represented the average monthly change in patients' A1c in a given home facility. In the model without case-mix adjustment, the facility slopes averaged -0.0148 A1c units per month (SD 0.0193, range -0.074 to 0.042). Of the 125 facilities, 105 had negative slopes, but 20 had monthly rises in A1c. The 95 percent SCIs for the slopes (see Figure 2) exhibited significant differences among the estimates. Additionally, 70 of the negative slopes and five of the positive slopes had SCIs which did not cross zero.

Most facility slope estimates were also negative in the case-mix adjusted model. In this model, 102 of 125 facilities had negative adjusted slopes, with mean -0.0131 (SD 0.0188, range -0.079 to 0.043). The 95 percent SCIs for the case-mix adjusted slopes also exhibited significant differences among the estimates. As well, 68 of the negative slopes and five of the positive slopes had SCIs which did not cross zero. The graph of these slopes and SCIs was remarkably similar to Figure 2 and hence is not shown.

We found some differences between the growth curve parameter estimates with and without case-mix adjustment. As expected, the facility intercepts from both models, or initial mean A1c values for patients at that facility, averaged 7.17 (SE 0.252) and 7.84 (SE 0.265) for models with and without case-mix adjustment, respectively, an unsurprising difference reflecting the entry of more explanatory variables into the model. The facility slopes averaged -0.0148 (SE 0.0193) and -0.0131 (SE 0.0188), respectively, and were thus much more similar. After case-mix adjustment, three facilities changed from negative to positive slope and no facility changed from a positive to negative slope. The model without case-mix adjustment had

Table 1: Comparison of Study Sample and the Overall Population VHA Users with Diabetes in FY99 and FY00

<i>Patient Characteristics</i>	<i>Study Sample (%) (SE)</i>	<i>Mean A1c (SD) of Study Sample*</i>	<i>Veterans with Diabetes in FY99 or FY00</i>
<i>N</i>	285,262	7.6 (1.8)	486,558
<i>Age (years)</i>			
Mean (SD)	64.1 (11.3)		64.5 (11.1)
<55	22.5	8.1 (2.2)	20.5
55–65	20.9	7.7 (1.7)	19.9
65–75	34.8	7.4 (1.5)	35.1
75+	18.5	7.2 (1.4)	18.7
Missing	3.3	8.1 (2.1)	5.8
<i>Sex</i>			
Female	2.5	7.7 (1.9)	2.5
Male	97.5	7.6 (1.8)	97.5
<i>Race/ethnicity</i>			
White	70.2	7.5 (1.6)	69.5
African American	15.3	7.9 (2.1)	14.9
Hispanic	2.8	7.9 (1.9)	2.8
Asian	0.5	7.9 (1.8)	0.4
Native American	0.4	7.9 (2.1)	0.3
Other	3.9	8.0 (2.1)	6.4
Unknown	6.8	8.1 (2.0)	5.6
<i>Marital status</i>			
Married	61.3	7.6 (1.7)	61.9
Unmarried	38.7	7.7 (1.9)	38.1
<i>Charlson scores</i>			
0 or 1	47.3	7.7 (1.8)	47.1
2	17.3	7.5 (1.7)	17.1
3 or 4	25.1	7.7 (1.7)	25.2
5 or more	10.4	7.5 (1.7)	10.6
<i>No. of A1c tests</i>			
Mean (SD)	3.17 (2.04)		3.04 (1.99)
1 or 2	49.0	7.6 (1.8)	70.1
3	18.1	7.5 (1.7)	10.6
4	13.2	7.6 (1.7)	7.7
5	8.7	7.7 (1.6)	5.1
6 or more	11.1	7.8 (1.6)	6.5
<i>Initial insulin use</i>	31.0	8.2 (1.8)	33.6

*A1c values by patient characteristic did not differ substantially for the overall population compared with the study sample, thus only the study sample A1c levels are presented.

VHA, Veterans Health Administration; FY99, Fiscal Year 1999; FY00, Fiscal Year 2000.

working $R^2 = 1.77$ percent and the case-mix adjusted model had working $R^2 = 14.45$ percent, reflecting considerable explanatory power for the case-mix adjustment variables.

Figure 2: Plot of Monthly A1c Change Rates, with 95 Percent Simultaneous Confidence Intervals, for Average Patient by Facility

Table 2 displays the case-mix adjusted model fixed parameter estimates. The case-mix parameter estimates largely mirrored the descriptive statistics. For example, the youngest patients (<55) had higher A1c than older patients. Being unmarried or using insulin was also associated with high A1c. Hispanics and African Americans had higher A1c compared with whites. Lab assay method B yielded slightly but significantly higher A1c than other certified lab assay methods. Seasonal effects were modest but significant in both models.

Higher-order polynomial growth curve models were also explored. These demonstrated little improvement over simple linear growth curve models. For the subanalysis on “out-migrators,” the average A1c decline rate was -0.033 (SE 0.0056) for patients in the “out-migrator” group and -0.027 (SE 0.0046) for the patients who remained in VA care. Normal-based statistical tests suggested that these two decline rates were significant (p -values $<.001$ for both), but there were no significant difference between them (p -value = .407). This suggested that the overall declining trend of the A1c values could not be attributed to patients with worsening control “out migrating.”

Table 2: Parameter Estimates for Case-Mix Variables, Seasonal Terms, and A1c Lab Method*

<i>Variable</i>	<i>Parameter Estimate (SD)*</i>	<i>95% Confidence Interval</i>
Age (years)		
<55	0.761 (0.0099)	(0.741, 0.780)
55–65	0.413 (0.0088)	(0.395, 0.430)
65–75	0.158 (0.0074)	(0.144, 0.173)
75+	—	—
Sex		
Female	−0.097 (0.0215)	(−0.139, −0.055)
Male	—	—
Race/ethnicity		
White	−0.262 (0.0553)	(−0.370, −0.154)
African American	0.015 (0.0560)	(−0.095, 0.124)
Hispanic	−0.021 (0.0568)	(−0.132, 0.090)
Asian	−0.159 (0.0569)	(−0.289, −0.030)
Native American	−0.063 (0.0694)	(−0.199, 0.073)
Other	−0.001 (0.0583)	(−0.124, 0.105)
Unknown	—	—
Marital status		
Married	−0.068 (0.0065)	(−0.081, −0.056)
Unmarried	—	—
Charlson scores		
0 or 1	0.191 (0.0102)	(0.172, 0.211)
2	0.117 (0.0113)	(0.095, 0.139)
3 or 4	0.197 (0.0105)	(0.176, 0.218)
5 or more	—	—
Diabetes treatment		
Insulin use	0.682 (0.0070)	(0.668, 0.695)
No insulin use	—	—
Lab method		
B	0.068 (0.0127)	(0.043, 0.093)
Others	−0.013 (0.0193)	(−0.051, 0.025)
T	—	—
Seasonal effects		
Sine term	0.018 (0.0020)	(0.014, 0.022)
Cosine term	−0.084 (0.0018)	(−0.088, −0.081)

*The parameter estimate indicates the level of A1c higher or lower than the reference (last) category.

Facility Rankings

One method for comparing the quality of health care is through ranking performance based on some specified criterion. In this study, we ranked facilities based on the facility slopes, using the case-mix adjusted model

estimates. We limited ranking to those 105 facilities that had at least 50 patients with two or more A1c values to maximize clinical relevance. Figure 3 shows facility ranks with 90 percent two-sided CIs for the 105 facilities, allowing assessment of comparative performance with the specified certainty. Relative performance of some facilities was evaluated fairly precisely, others less so. For instance, the facility ranked 105th (most highly positive slope) had a 90 percent SCI of [101, 105], so that we can be confident it is a poor performer relative to other facilities. However, the facility ranked 102nd had a 90 percent SCI of [34, 105], so we can only ascertain that it is not in the first quartile of performers. For the top 10 performers (ranks 1–10) we obtained 90 percent upper CIs for the ranks and for the bottom 10 performers (ranks 96–105) we obtained 90 percent lower CIs for the ranks. Three out of the 10 best performers had CIs contained within the top decile and another four had CIs contained within the top quartile. Also, four of the 10 worst performers had CIs contained in the bottom decile, and another four had CIs contained in the bottom half.

We also examined change in facility ranks between the models with and without case-mix adjustment. Despite fairly modest differences in growth

Figure 3: Diabetes Quality-of-Care Rank Assignments from Best (1) to Worst (105), with 90 Percent Simultaneous Confidence Intervals

Quality of care as reflected in monthly change in A1c, using the fully adjusted model.

curve parameter estimates, some *substantial* changes in ranks were evident when case-mix variables were included. After case-mix adjustment, seven of 105 facilities improved their quartile of performance and nine moved to a lower quartile. Eighty-nine facilities remained at the same quartile with and without case-mix adjustment. Seven of 10 facilities remained at or above the 90th percentile of performance and eight of 10 facilities remained at or below the 10th percentile of performance while four of six remained at or above the 95th percentile and five of six remained at or below the 5th percentile.

DISCUSSION

Current approaches to comparison of health care systems based on quality-of-care measures rely on cross-sectional methodology because of the need to rely upon chart abstractions in many instances. As cross-sectional evaluation of intermediate outcomes, such as A1c levels, cannot measure within-patient change over time, it is difficult to assess care in stable populations of individuals with chronic disease. Furthermore, quality comparisons for diabetes care currently do not include adjustments for case mix or duration of health system care. For example, health care systems that serve largely older individuals, like Medicare beneficiaries, may have better A1c than those serving inner city minorities because of population age and ethnicity. Additionally, if newer enrollees are healthier, cross-sectional measures can appear better for similar nonquality reasons. Although currently used cross-sectional measures of quality of diabetes care have made major contributions and demonstrable progress toward improvement, they cannot evaluate continued health care given over time to individuals with chronic diseases (Zhang et al. 2000; Weiner and Long 2004). The longitudinal approach presented here measures changes in A1c over time within individuals, effectively overcoming some of the cross-sectional problems and capturing a new domain in quality measurement.

We described considerable facility-level variation in controlling A1c over time for patients within semi-autonomous units within this large national health care system. On average, case-mix adjusted facility-level A1c values decreased by -0.314 (range -1.90 to 1.03) over 2 years, suggesting that overall VHA diabetes care improved. This level of change is clinically meaningful; a patient with A1c of 7.5 percent who began diabetes care at a facility at the start of the study period would on average have had A1c of 7.2 percent by the end of the study period. The majority of 125 facilities (102

with case-mix adjustment, 105 without case-mix adjustment) demonstrated negative monthly change rates in A1c. The facilities with rising within-patient A1c levels over time might represent targets for quality improvement efforts, especially those whose slope estimates had SCIs that did not cross zero. Such slope estimate comparisons can be highly useful for quality improvement purposes.

We also assigned ranks to facilities based on slope estimates. Identifying facilities in the lowest decile of performance is a common approach when determining which facilities could improve quality of care. Fairness dictates high confidence in such rank assignments, although statistical variability remains a barrier to achieving high confidence in many situations. Even after case-mix adjustment and limiting rank assignment to facilities with sufficient data, we found considerable uncertainty in ranking, similar to other reports (Marshall and Spiegelhalter 1998). Nevertheless, four facilities identified in the lowest decile of performance had rank CIs remaining within the lowest decile of performance, indicating that these facilities likely did perform "worse." Use of ranks for accountability has been debated but may be highly useful for quality improvement purposes (Schneider and Epstein 1996; Marshall and Spiegelhalter 1998).

We found some effect of case-mix adjustment on diabetes care evaluation, in that the case-mix adjusted slope estimates were modestly different than those without case-mix adjustment. We observed changes in composition of deciles of rank, with five out of 20 facilities identified at or below the 10th percentile of performance without case-mix variables moving into a better decile in the case-mix adjusted model. Consequently, case-mix adjustment may be warranted for longitudinal evaluations of A1c rates of change in an accountability context.

One of the strengths of this study was the ability to describe a large population by using administrative data, which has been demonstrated to have high reliability compared with chart abstraction data (Kerr et al. 2002). Additionally, the VHA has effectively reduced access barriers to care and pharmacy services, diminishing variation because of accessibility that can overwhelm quality-of-care studies in other settings (Jha et al. 2003).

There are also some limitations to this study. First, the restriction of the study to the VHA, which serves primarily older men, possibly impacts the generalizability of our findings. Second, we controlled variation in A1c measurement by only using values from NGSP-certified methods, yet we found that one method had consistently slightly higher values than the other certified methods. Recent standards for NGSP certification have become more

stringent than those in place at the time of this study (National Glycohemoglobin Standardization Program [NGSP] 2004). The ongoing implementation of the NGSP nationally should improve future A1c comparability. Third, as an observational study based on existing administrative data, potentially important confounders were not available for analysis. For example, medication adherence may be a very important determinant, but is difficult to ascertain from administrative data, especially for insulin utilization. Also, an important problem in studies like this is loss to follow-up. We performed a subanalysis that demonstrated that patients leaving the system were similar in A1c control as those who remained in VA care. Thus, a “healthy survivor effect” was unlikely to account for the findings. Finally, for ranking purposes, we used only facilities with at least two measures on 50 or more individuals to maximize stability of estimates. Because enrollment in private sector health plans is on average about 2.5 years, the individuals with this much data may be limited for typical Health Employer Data Information Set samples of 411 chart reviews (HEDIS[®] 2004). However, increasing emphasis on frequent A1c measurement by the American Diabetes Association American Diabetes Association (2004) and the National Committee on Quality Assurance (HEDIS[®] 2004) is likely to increase these numbers in the future.

In summary, we demonstrated significant differences among VHA facilities in A1c change rates for average patients within a given facility. Additionally, our results indicate the likely importance of risk adjustment if the results are to be used for accountability purposes. This use of longitudinal data may offer distinct advantages over cross-sectional methods by distinguishing within-subject effects from population effects, thereby determining patient response to the clinical setting. We suggest evaluation and comparison of these within-patient change rates may prove to be a useful new approach to studying the quality of ongoing care to a population with chronic disease, and may identify those facilities that could be more closely evaluated for best (and more problematic) practices.

ACKNOWLEDGMENTS

This work was supported by a grant from the Veterans Administration Health Services Research and Development Program (IIR 00-072-1) (Dr. Pogach). The research is also partly supported by NSF SES-0241859 (Dr. Xie). Address questions about the statistical methodology to Dr. Xie at mxie@rutgers.edu.

REFERENCES

- American Diabetes Association. 2004. "Clinical Practice Recommendations 2004." *Diabetes Care* 27 (suppl): S92-3.
- Berlowitz, D. R., A. Ash, R. Friedman, E. Hickey, B. Kader, and M. A. Moskowitz. 1998. "Problems in Assessing Diabetes Control in an Ambulatory Setting." *American Journal of Medical Quality* 13 (2): 89-93.
- Bryk, A., and S. W. Raudenbush. 1992. *Hierarchical Linear Models for Social and Behavioral Research: Applications and Data Analysis Methods*. Newbury Park, CA: Sage.
- Charlson, M. E., P. Pompei, K. L. Ales, and C. R. MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5): 373-83.
- Chin, M. H., S. B. Auerbach, S. Cook, J. F. Harrison, J. Koppert, L. Jin, F. Thiel, T. G. Karrison, A. G. Harrand, C. T. Schaefer, H. T. Takashima, N. Egbert, S. C. Chiu, and W. L. McNabb. 2000. "Quality of Diabetes Care in Community Health Centers." *American Journal of Public Health* 90: 431-4.
- Committee on Quality Health Care in America, Institute of Medicine. 2000. *Crossing the Quality Chasm: A New Health System for the 21st Century*, pp 23-3. Washington, DC: National Academy Press.
- Deyo, R. A., D. C. Cherkin, and M. A. Ciol. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Databases." 1992. *Journal of Clinical Epidemiology* 45 (6): 613-9.
- Diabetes Control and Complications Trial Research Group. 1993. "The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-Term Complications in Insulin-Dependent Diabetes Mellitus." *New England Journal of Medicine* 329 (14): 977-86.
- Diggle, P. J., K. Y. Liang, and S. L. Zeger. 1996. *The Analysis of Longitudinal Data*. Oxford, U.K.: Oxford University Press.
- Goldstein, H., and D. J. Spiegelhalter. 1996. "League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance." *Journal of the Royal Statistical Society Series A* 159: 385-443.
- Harris, M. I., R. C. Eastman, C. C. Cowie, K. M. Flegal, and M. S. Eberhardt. 1999. "Racial and Ethnic Differences in Glycemic Control of Adults with Type 2 Diabetes." *Diabetes Care* 22 (3): 403-8.
- Health Plan Employer Data and Information Set (HEDIS®). 2004. [Accessed June 23, 2004]. Available at <http://www.ncqa.org/Programs/HEDIS/>
- Hebert, P. L., L. S. Geiss, E. F. Tierney, M. M. Engelgau, B. P. Yawn, and A. M. McBean. 1999. "Identifying Persons with Diabetes Using Medicare Claims Data." *American Journal of Medical Quality* 14: 270-7.
- Hsu, J. 1996. *Multiple Comparisons: Theory and Methods*. 1st edition. Boca Raton, LA: Chapman & Hall/CRC.
- Iezzoni, L. I. 1997. "Risk and Outcomes." In *Risk Adjustment for Measuring Healthcare Outcomes*, edited by L. I. Iezzoni, pp 1-42. Chicago: Health Administration Press.

- Jha, A. K., J. B. Perlin, K. W. Kizer, and R. A. Dudley. 2003. "Effect of the Transformation of the Veterans Affairs Health Care System on the Quality of Care." *New England Journal of Medicine* 348 (22): 2218–27.
- Kerr, E. A., D. M. Smith, M. M. Hogan, S. L. Krein, L. M. Pogach, T. P. Hofer, and R. A. Hayward. 2002. "Comparing Clinical Automated, Medical Record, and Hybrid Data Sources for Diabetes Quality Measures." *Joint Commission Journal on Quality Improvement* 28 (10): 555–65.
- Liang, K-Y., and S. L. Zeger. 1986. "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika* 73: 13–22.
- Little, R. R., C. L. Rohlfing, H. M. Wiedmeyer, G. L. Myers, D. B. Sacks, and D. E. Goldstein. 2001. "The National Glycohemoglobin Standardization Program: A Five-Year Progress Report." *Clinical Chemistry* 47: 1985–92.
- Marshall, E. G., and D. J. Spiegelhalter. 1998. "Reliability of League Tables of In Vitro Fertilisation Clinics: Retrospective Analysis of Live Birth Rates." *British Medical Journal* 316 (7146): 1701–4.
- National Glycohemoglobin Standardization Program. 2004. "List of Certified Laboratories" [accessed June 23, 2004]. Available at <http://www.missouri.edu/~diabetes/ngsp.html>
- Schneider, E. C., and A. M. Epstein. 1996. "Influence of Cardiac-Surgery Performance Reports on Referral Practices and Access to Care—A Survey of Cardiovascular Specialists." *New England Journal of Medicine* 335 (4): 251–6.
- Safford, M., L. Eaton, G. Hawley, M. Brimacombe, M. Rajan, H. Li and L. Pogach. 2003. "Disparities in Use of Lipid-Lowering Medications among People with Type 2 Diabetes Mellitus." *Archives of Internal Medicine* 163 (8): 922–8.
- Tahara, Y., and K. Shima. 1995. "Kinetics of HbA1c, Glycated Albumin, and Fructosamine and Analysis of Their Weight Functions against Preceding Plasma Glucose Level." *Diabetes Care* 18 (4): 440–7.
- U.K. Prospective Diabetes Study (UKPDS) Group. 1998. "Intensive Blood-Glucose Control with Sulphonylureas or Insulin Compared with Conventional Treatment and Risk of Complications in Patients with Type 2 Diabetes (UKPDS 33)." *Lancet* 352: 837–53.
- Weiner, M., and J. Long. 2004. "Cross-Sectional versus Longitudinal Performance Assessments in the Management of Diabetes." *Medical Care* 42 (2 suppl): II34–9.
- Wright, A., A. C. Burden, R. B. Paisey, C. A. Cull, and R. R. Holman. 2002. "U.K. Prospective Diabetes Study Group. Sulphonylurea Inadequacy: Efficacy of Addition of Insulin over 6 Years in Patients with Type 2 Diabetes in the U.K. Prospective Diabetes Study (UKPDS 57)." *Diabetes Care* 25 (2): 330–6.
- Zhang, Q., M. Safford, J. E. Ottenweller, G. Hawley, D. Repke, J. F. Burgess Jr., S. Dhar, H. Cheng, H. Naito, and L. M. Pogach. 2000. "Performance Status of Health Care Facilities Changes with Risk Adjustment of Hemoglobin A1c." *Diabetes Care* 23: 919–27.