Managing Chronic Illness in Managed Care Settings

Measuring the Quality of Diabetes Care Using Administrative Data: Is There Bias?

Nancy L. Keating, Mary Beth Landrum, Bruce E. Landon, John Z. Ayanian, Catherine Borbas, and Edward Guadagnoli

Objectives. Health care organizations often measure processes of care using only administrative data. We assessed whether measuring processes of diabetes care using administrative data without medical record data is likely to underdetect compliance with accepted standards for certain groups of patients.

Data Sources/Study Setting. Assessment of quality indicators during 1998 using administrative and medical records data for a cohort of 1,335 diabetic patients enrolled in three Minnesota health plans.

Study Design. Cross-sectional retrospective study assessing hemoglobin A1c testing, LDL cholesterol testing, and retinopathy screening from the two data sources. Analyses examined whether patient or clinic characteristics were associated with underdetection of quality indicators when administrative data were not supplemented with medical record data.

Data Collection/Extraction Methods. The health plans provided administrative data, and trained abstractors collected medical records data.

Principal Findings. Quality indicators that would be identified if administrative data were supplemented with medical records data are often not identified using administrative data alone. In adjusted analyses, older patients were more likely to have hemoglobin A1c testing underdetected in administrative data (compared to patients <45 years, OR 2.95, 95 percent CI 1.09 to 7.96 for patients 65 to 74 years, and OR 4.20, 95 percent CI 1.81 to 9.77 for patients 75 years and older). Black patients were more likely than white patients to have retinopathy screening underdetected using administrative data (2.57, 95 percent CI 1.16 to 5.70). Patients in different health plans also differed in the likelihood of having quality indicators underdetected.

Conclusions. Diabetes quality indicators may be underdetected more frequently for elderly and black patients and the physicians, clinics, and plans who care for such patients when quality measurement is based on administrative data alone. This suggests that providers who care for such patients may be disproportionately affected by public release of such data or by its use in determining the magnitude of financial incentives.

Key words. Quality measurement, quality indicators, diabetes care

In recent years, efforts to measure and improve the quality of care have expanded to include care for chronic medical conditions, such as diabetes mellitus, that are primarily treated in outpatient settings. In addition to providing information for purchasers of health care (Epstein 1995; Iglehart 1996; Marshall et al. 2000; National Committee for Quality Assurance 2000) and internal efforts to improve quality (Petitti et al. 2000; Kiefe et al. 2001), quality data are increasingly being included in publicly released performance reports (Epstein 2000; Jencks et al. 2000; Bost 2001) and as criteria for determining compensation for providers by medical groups and health plans (Schlackman 1993; Pedersen et al. 2000; Kowalczyk 2001a; Sussman et al. 2001). Because the results of quality measurements may have widespread consequences, it is important that the data collected be accurate and unbiased.

Quality measures for diabetes care are collected primarily from two sources: medical records data and administrative data (which are used primarily for billing purposes). Although some efforts incorporate data from both of these sources (National Committee for Quality Assurance 2000; Diabetes Quality Improvement Project 2001), others may use administrative data exclusively to measure quality of care (Jencks et al. 2000; Kowalczyk 2001b) because these data are typically more readily available and less expensive to collect than data from medical records (Iezzoni 1997).

It is important to know whether quality assessments that rely on administrative data alone fail to detect compliance with accepted standards and whether this problem varies systematically for certain groups of patients, particularly based on age, sex, race, or socioeconomic status. This could be the case, for example, if certain groups of patients are more frequently cared for at clinics that lack sophisticated systems that might facilitate more complete billing. The accuracy of quality assessments is crucial to determine whether

This work was supported by grant no. HS09936 from the Agency for Healthcare Research and Quality and grant no. HS98-005 from the American Association of Health Plans. Additional support was provided by the Marshall J. Seidman Center for Studies in Health Economics and Health Policy.

Address correspondence to Nancy L. Keating, M.D., M.P.H., Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115. Dr. Keating is also with the Division of General Internal Medicine (Section on Health Services and Policy Research), Department of Medicine, Brigham and Women's Hospital, as is John Z. Ayanian, M.D., M.P.P. With the Department of Health Care Policy, Harvard Medical School, are Drs. Keating and Ayanian, as well as Mary Beth Landrum, Ph.D., Bruce E. Landon, M.D., M.Sc., and Edward Guadagnoli, Ph.D. Dr. Landon is also with the Division of General Medicine and Primary Care, Beth Israel Deaconess Medical Center. Catherine Borbas, Ph.D., M.P.H., is with the Healthcare Evaluation and Research Foundation, Inc., St. Paul, MN.

disparities in quality of care persist (Fiscella et al. 2000). Moreover, if quality of care is systematically underdetected for some groups of patients when administrative data alone are used, then quality profiles may be unfairly biased against providers who more frequently care for such patients.

In this study, we first compared assessments of the quality of care for diabetes that result from using either administrative data only, medical record data only, or the combination of both sources. We next assessed whether quality measurement using administrative data without medical records data is likely to systematically underdetect quality indicators for certain groups of patients based on age, sex, race, or income. Finally, because we were also concerned that there may be systematic differences in administrative processing of claims among clinics (for example, some clinics may submit less complete administrative data to the health plans), we explored whether such differences could be explained by the clinic where a patient receives care.

METHODS

Study Population

We collaborated with three health plans in Minnesota to examine quality of care for patients with diabetes. Using administrative data, we identified all patients aged 18 years or older with type 1 or 2 diabetes mellitus as defined by having two or more encounters listing an ICD-9-CM code for diabetes mellitus (250.xx), diabetic polyneuropathy (357.xx), diabetic retinopathy (362.0–362.0x), or diabetic cataract (366.41) during the 18-month period from July 1, 1997, through December 31, 1998. We based our identification on primary and secondary diagnoses for outpatient encounters and on primary diagnosis for inpatient encounters. We excluded any patients with a code for end-stage renal disease, patients with a break in enrollment of more than 30 days, and women who were pregnant during the 18-month period because they may have had gestational diabetes and because their standard of care may differ from other patients. The sample included 1,335 patients from the three health plans ($n = 792$ from Plan A, $n = 250$ from Plan B, and $n = 293$ from Plan C). The study protocol was approved by the Harvard Medical School Committee on Human Studies and by participating health plans.

Data Collection

In addition to administrative data supplied by the health plans, trained abstractors collected data from the medical records of the physician providing

most or all of the diabetes care for each patient in the sample. We used administrative encounter data to select the physician with whom the patient had the most outpatient visits with a diagnosis code for diabetes during the 18-month period. For cases in which a patient had the same number of such visits to two or more physicians, we selected the physician with the greatest number of total (diabetes and nondiabetes) visits. If more than one physician provided the same number of encounters, we selected the primary care physician, or if no primary care physician provided care, we selected the physician most likely to provide diabetes care based on their speciality (for example, if a patient saw an oncologist and an endocrinologist, we selected the endocrinologist). Medical records were not abstracted for 183 patients because they could not be located or patients did not consent; those whose records were abstracted were more often from Plan B ($p < .001$) and more often older ($p < .001$). We limited analyses to the 1,152 patients for whom we had both administrative data and medical records data. As part of a larger study, we also surveyed patients in the sample to collect information about their demographic characteristics and experiences with care. The survey response rate was 65.5 percent.

Quality Indicators

We assessed quality of care during the one-year period from January 1, 1998, through December 31, 1998, for three widely accepted diabetes indicators of quality (hemoglobin A1c testing, low-density lipoprotein [LDL] cholesterol testing, and assessment for retinopathy) (American Diabetes Association 2002) that could be reasonably collected from either administrative data or medical records. From administrative data, we identified hemoglobin A1c testing (Current Procedure Terminology [CPT] code 83036) and LDL cholesterol testing (CPT code 80061 or 83721) if there was a claim for the test during 1998 and we documented assessment for retinopathy if the patient had a claim for an ophthalmology or optometry visit including eye examination (CPT codes 92002, 92004, 92012, 92014, 92018, 92019, 92225, 92226, 92235, 92250) during 1998 (FACCT—Foundation for Accountability 1997).

From the medical records, we documented hemoglobin A1c testing and LDL cholesterol testing if there was evidence that the test was performed during 1998 and assessment for retinopathy if it was noted that the patient had a dilated retinal exam performed by an ophthalmologist or optometrist or had retinal photography at any time during 1998. Six experienced reviewers conducted the medical record abstractions; re-review of a sample of charts by the trainer demonstrated ≥ 90 percent agreement. The kappa statistics for the

indicators were .84 for Hemoglobin A1c testing, .83 for LDL cholesterol testing, and .64 for retinopathy assessment.

Patient Characteristics

We documented each patient's age and sex from the medical records and from administrative data. Information about race was collected from the patient survey and supplemented with medical records data for survey nonrespondents. We linked zip codes (obtained from administrative data) with 1990 U.S. Census data to obtain the median household income for the area of residence as a proxy for socioeconomic status.

Analysis

We first calculated the proportion of patients who had the test or assessment of interest using each data source or a combination of the two data sources (our gold standard). We conducted simple frequencies to assess how rankings of our three plans might vary depending on the data source used to measure each indicator. We next assessed whether use of administrative data without medical record data was more likely to systematically underdetect quality indicators for certain groups of patients that would have been detected if medical records data were also collected. We were interested in this, and not in possible underdetection of care using medical record data alone, because health care organizations are unlikely to rely on medical records data alone for quality measurement. Thus, for each quality indicator, the gold standard was evidence of the indicator in either the medical records or administrative data or both, and we created dependent variables to indicate that the test or assessment was performed based on the gold standard, but not based on administrative data alone.

We used the χ^2 test to identify patients' characteristics associated with underdetection of each quality indicator when administrative data were not supplemented with medical records data. We then used logistic regression to assess whether patients' age in years (<45, 45–54, 55–64, 65–74, 75 and older), sex, race (white, black, Hispanic, Asian, other or multiracial, unknown), income (median household income in the zip code of residence, by quartile), and health plan were associated with underdetection of quality indicators. We also adjusted for the specialty of the physician documenting in the chart we abstracted (primary care physician, endocrinologist, more than one physician, other). We used generalized estimating equations to control for clustering of patients within clinics.

Finally, we used conditional logistic regression models to assess for confounding of the associations between patient characteristics and under-detection of quality indicators using administrative data alone by clinic (Localio et al. 2001). These models account for the clustering of similar types of patients within clinics when estimating patient-level effects by stratifying on clinic, and thus, allowed us to determine whether the associations identified are within-clinic or among-clinic associations. For example, certain clinics may predominantly serve black patients or older patients and those clinics may also be more likely to submit incomplete claims, thus leading to underreporting of quality indicators using administrative data.

Sensitivity Analyses

First, we repeated analyses among patients from the entire sample of 1,335 patients even though some did not have medical records data because quality measures using administrative data only would include such patients. Second, because data on race were missing for 125 patients, we created an indicator variable to designate missing race information, and included this variable in the models described above. In a series of analyses, we (a) excluded patients with missing race from analyses, (b) classified patients with missing race as white patients because the patients with missing race resided in areas of predominantly white residents (mean proportion of white residents in zip codes of residence for patients with missing information on race >95 percent), and (c) weighted the analyses to reflect the probability that patients were missing information on race (Little and Rubin 1987; Wells et al. 2000). Finally, to determine whether findings were influenced by the likelihood that patients see multiple providers (especially because some of the providers may not document in the chart we abstracted), we repeated multivariable models including a variable for the number of unique providers seen during the year, as determined from claims data.

All analyses were conducted using *SAS* statistical software, version 8.0 (SAS Institute, Inc., Cary, North Carolina), with the exception of the sensitivity analysis that included sample weights, which was conducted using *SUDAAN* software, version 7.5.6 (Research Triangle Institute, Research Triangle Park, North Carolina).

RESULTS

The 1,152 patients with both administrative and medical records data had a mean age of 62 years (SD 15.3) and 54 percent were women. Patient

Table 1: Proportion of Patients with Each Quality Indicator by Data Source, $N = 1,152$

<i>Quality Indicator</i>	<i>Administrative Data*</i>	<i>Medical Record Data*</i>	<i>Administrative Data or Medical Record Data (Gold Standard)</i>
Hemoglobin A1c	72% (79%)	88% (97%)	91%
LDL cholesterol	36% (72%)	44% (88%)	50%
Retinopathy screening	68% (90%)	34% (45%)	76%

*Proportion in parentheses represents the proportion of the gold standard cases identified.

characteristics varied by plan; patients in Plan B were younger ($p < .001$) than patients in Plans A and C, and patients in Plan C were more often women ($p < .001$) and more often of black or Asian race ($p < .001$) compared to patients in Plans A and B.

The proportion of patients with care in compliance with accepted standards based on administrative data, medical records data, or both (the gold standard) is included in Table 1. Administrative data alone and medical records data alone each identified patients with quality indicators present in one source but not the other source. Medical records were a better source than administrative data for hemoglobin A1c and LDL cholesterol testing, but administrative data captured more ophthalmologic examinations. Table 2 demonstrates how quality rankings might vary for the three plans depending on the quality indicator measured and the source of measurement. Although rankings did not change for LDL measurement, the ranking for Plan B for hemoglobin A1c testing changed from 3 when both data sources were used to 1 when only administrative data were used.

Table 3 displays the unadjusted associations between patients' characteristics and the underdetection of indicators if assessments were made using administrative data alone. Patients who were older, black, or Asian, and members of Plan C were more likely than other patients to have their hemoglobin A1c testing and LDL testing underdetected if administrative data were not supplemented with medical records data. For retinopathy screening, younger patients, black, Asian, or patients of other races, as well as patients with low incomes and patients from Plans B and C were more likely to have the indicator underdetected.

In multivariable analyses, patients 65 and older compared to patients under age 45 were more likely to have hemoglobin A1c testing underdetected

Table 2: Changes in Rankings of Measured Quality Indicators Based on Data Source Examined

	<i>Administrative Data Only</i>	<i>Medical Record Data Only</i>	<i>Administrative Data or Medical Record Data (Gold Standard)</i>
Hemoglobin A1c			
Plan A	2	1	1
Plan B	1	3	3
Plan C	3	2	2
LDL Cholesterol			
Plan A	2	2	2
Plan B	1	1	1
Plan C	3	3	3
Retinopathy Screening			
Plan A	1	2	1
Plan B	3	1	3
Plan C	2	3	2

in administrative data, as were patients in Health Plan C compared to Plan A (Table 4). Patients whose medical records had more than one physician documenting were less likely to have A1c testing underdetected in administrative data. Patients who were members of Plan C were more likely than patients of Plan A to have their LDL cholesterol underdetected using administrative data (Table 4). Black patients were more likely than white patients to have their retinopathy screening underdetected using administrative data, as were patients from Plan B compared to Plan A. Patients in the three highest quartiles of income were less likely than patients in the lowest quartile to have their retinopathy screening underdetected using administrative data, although only patients in the third highest income quartile differed significantly from patients in the lowest quartile (Table 4).

Finally, we used conditional logistic regression to assess for confounding of associations between patient characteristics and underdetection of quality indicators by clinic. In this set of models, we found that the strength of the associations between black race and underdetection of retinopathy screening was reduced slightly (by 9 percent) and was no longer statistically significant ($p = .13$). The associations of age with hemoglobin A1c testing were also reduced slightly (by 15 percent to 16 percent) but remained statistically significant. The association of two or more physicians documenting in the record and A1c decreased by 25 percent and was no longer statistically significant ($p = .06$). The association between Plan C and hemoglobin A1c

Table 3: Characteristics Associated with Underdetection of Quality Indicators Using Administrative Data Only, Unadjusted

Characteristic	Hemoglobin A1c				LDL Cholesterol				Retinopathy Screening			
	Number	Number of Patients with Indicator Using Administrative and Medical Records	Proportion Underdetected Using Only Administrative Data (20.6%)	p-Value	Number of Patients with Indicator Using Administrative and Medical Record	Proportion Underdetected Using Only Administrative Data (27.5%)	p-Value	Number of Patients with Indicator Using Administrative and Medical Record	Proportion Underdetected Using Only Administrative Data (10.1%)	p-Value		
	N = 1,152	N = 1,044			N = 572			N = 870				
Age in Years												
<45	192	176	12.5	<.001	944	27.7	<.001	131	12.2	.09		
45–54	197	182	13.2		98	20.4		132	13.6			
55–64	242	225	10.2		159	18.9		178	13.5			
65–74	247	222	29.7		126	38.1		208	5.3			
≥ 75	274	239	33.5		95	34.7		221	8.6			
Sex												
Male	527	475	16.4	.002	270	25.2	.25	396	10.4	.83		
Female	625	569	24.1		302	29.5		474	9.9			
Race												
White	835	753	17.8	<.001	437	25.9	.04	644	8.9	<.001		
Black	48	40	40.0		18	50.0		35	25.7			
Hispanic	24	23	17.4		12	33.3		17	5.9			
Asian	71	67	44.8		30	46.7		51	17.6			
Other/multiracial	49	43	23.3		22	22.7		35	22.9			
Unknown	125	118	17.8		53	22.6		88	4.6			

Continued

Table 3: Continued

Characteristic	Hemoglobin A1c			LDL Cholesterol			Retinopathy Screening		
	Number	Indicator Using Administrative and Medical Data 215/1,044	Proportion Undetected Using Only Administrative Data 215/1,044 (20.6%)	Number of Patients with Indicator Using Administrative and Medical Data N = 572	Proportion Undetected Using Only Administrative Data N = 157/572 (27.5%)	p-Value	Number of Patients with Indicator Using Administrative and Medical Data N = 88/870	Proportion Undetected Using Only Administrative Data N = 88/870 (10.1%)	p-Value
Median Household Income (quartiles)									
\$25,441 or less	260	224	21.0	129	24.8	.39	189	14.8	.03
\$25,442 to \$32,330	247	222	16.2	110	27.3		189	7.9	
\$32,331 to \$39,407	234	223	17.6	124	25.0		203	6.9	
More Than \$39,407	264	241	15.2	135	23.0		189	8.4	
Specialty of Physician									
Whose Chart Reviewed									
Internist or FP	896	811	22.6	439	28.7	.02	664	11.1	.14
Endocrinologist	90	90	14.4	50	26.0		75	10.7	
Two or More Providers	129	111	11.7	66	22.7		103	3.9	
Documenting									
Other	37	32	18.7	17	17.6		28	7.1	
Health Plan									
Plan A	711	650	15.8	352	25.6	<.001	579	6.4	<.001
Plan B	162	142	6.3	92	17.4		99	21.2	
Plan C	279	252	41.1	128	39.8		192	15.6	

*Denominator includes all patients with indicator based on medical record and/or administrative data (N = 1,044).

†Denominator includes all patients with indicator based on medical record and/or administrative data (N = 572).

‡Denominator includes all patients with indicator based on medical record and/or administrative data (N = 870).

Table 4: Characteristics Associated with Underdetection of Quality Indicators Using Administrative Data Only, Adjusted*

Characteristic	Hemoglobin A1c		LDL Cholesterol		Retinopathy Screening	
	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Age in Years						
<45	–	Ref	–	Ref	–	Ref
45–54	1.12	0.61–2.08	0.61	0.31–1.19	1.54	0.80–3.00
55–64	0.78	0.43–1.41	0.55	0.29–1.04	1.43	0.80–2.53
65–74	2.94	1.10–7.88[†]	1.48	0.81–2.70	0.49	0.23–1.06
≥75	4.20	1.80–9.81[†]	1.39	0.73–2.67	1.25	0.52–2.98
Sex						
Male	–	Ref	–	Ref	–	Ref
Female	1.15	0.84–1.59	0.96	0.63–1.45	0.79	0.53–1.18
Race						
White	–	Ref	–	Ref	–	Ref
Black	1.67	0.80–3.50	1.85	0.73–4.72	2.51	1.14–5.56[†]
Hispanic	0.67	0.22–2.07	0.95	0.27–3.32	0.40	0.03–6.34
Asian	1.21	0.73–2.00	1.55	0.72–3.34	2.06	0.73–5.84
Other/multiracial	0.96	0.37–2.49	0.66	0.30–1.42	1.92	0.88–4.17
Unknown	1.17	0.78–1.74	0.88	0.43–1.78	0.45	0.17–1.22
Quartile of Median Household Income						
Less than \$25,504	–	Ref	–	Ref	–	Ref
\$25,504 to \$32,515	0.84	0.57–1.23	1.04	0.60–1.79	0.51	0.24–1.12
\$32,516 to \$39,995	1.06	0.69–1.62	0.73	0.46–1.18	0.46	0.24–0.87[†]
More Than \$39,995	1.07	0.65–1.78	0.81	0.47–1.41	0.72	0.39–1.34
Specialty of Physician Documenting in Chart						
Internist or family practitioner	–	Ref	–	Ref	–	Ref
Endocrinologist	0.99	0.41–2.39	1.17	0.61–2.25	0.96	0.42–2.20
Two or More Providers Documenting	0.52	0.29–0.95[†]	0.75	0.44–1.27	0.34	0.07–1.70
Other	0.92	0.34–2.51	0.55	0.17–1.75	0.60	0.17–2.14
Health Plan						
Plan A	–	Ref	–	Ref	–	Ref
Plan B	0.64	0.29–1.41	0.74	0.39–1.42	3.74	1.99–7.05[†]
Plan C	3.24	2.01–5.22[†]	1.90	1.16–3.13[†]	1.98	0.85–4.20

*Controlling for variables listed and account for clustering of patients within clinics.

[†] $p < .05$

testing did not change, but the association between Plan C and LDL-cholesterol testing decreased by 69 percent and was no longer statistically significant ($p = .59$). The strength of the association between Health Plan B

and retinopathy screening decreased by 22 percent but remained statistically significant. These findings suggest that some of the underreporting of hemoglobin A1c and retinopathy screening among black patients and older patients and the better reporting among patients with two or more physicians documenting in the chart could be accounted for by the clinics in which they were treated. However, significant associations with underreporting of quality measures persisted within clinics for patient age and health plan.

In sensitivity analyses, expanding the sample to include all 1,335 patients, even those whose medical records were not abstracted (because claims-based quality measurement would also include such patients), results were similar (data not shown). In additional sensitivity analyses, results were also similar when we excluded patients with missing race from analyses, classified them as white, or weighted the data for propensity to have missing race (data not shown). Finally, when we included a variable for the number of unique providers seen during 1998, having seen more providers was associated with less underdetection of hemoglobin A1c testing, but was not associated with underdetection of LDL testing or retinopathy screening. In addition, inclusion of this variable in the model did not change any of the other associations reported.

DISCUSSION

As efforts to measure the quality of diabetes care proliferate and their uses expand, assuring collection of accurate and unbiased data becomes increasingly important. In this study, we collected diabetes quality data for a large cohort of patients using two sources (administrative and medical records data) for three widely used quality indicators. We found that different data sources provide varying evidence of quality depending on the indicator being measured and that older patients were more likely than younger patients to have hemoglobin A1c testing underdetected, black patients were more likely than white patients to have retinopathy screening underdetected, and that underdetection of quality indicators varied according to the indicator being measured for patients from different health plans.

Our findings about the variability of different sources used for quality measurement are consistent with a recent report comparing diabetes quality measures collected from administrative data, medical records data, and self-report (Fowles et al. 1999). Our study builds on this prior report by demonstrating that underdetection of compliance with quality standards was

associated with patient characteristics when indicators are measured with administrative data only. This finding is particularly important considering that such data may serve as the basis for publicly released performance reports or physician compensation (Schlackman 1993; Epstein 2000; Jencks et al. 2000; Pedersen et al. 2000; Bost 2001; Kowalczyk 2001a; Sussman et al. 2001). Health care organizations using administrative data for such efforts should recognize that they might unfairly penalize physicians or clinics who predominantly care for disadvantaged patients, such as elderly or black patients. It is possible that increases in quality measurement using only administrative data may prompt providers and plans to invest resources in improving the completeness of their data, which would be a good outcome of expanding claims-based measurement systems. However, some providers and plans may lack resources to do so, and these may be the same providers who more often care for disadvantaged patients.

There are potential explanations for our findings at both the plan and clinic level. At the health plan level, there may be differences in benefits covered (for example, if eye examinations are not routinely covered, fewer claims may be submitted). This might explain why patients in Plan B were most likely to have retinopathy screening underdetected using administrative data, but were least likely to have hemoglobin A1c measurements underdetected. Other potential explanations include the frequency of out-of-network care and coverage policies for such care, the organization's size, the role of specialists versus primary care physicians in disease management, and variations in claims processing. Additionally, the quality of administrative data may vary within a health care organization depending on the organization's contracts with providers and groups. For example, encounter data from medical groups paid by capitation may be less reliable than data from medical groups paid by fee-for-service arrangements where claims must be submitted for reimbursement.

The results of our conditional logistic regression models suggest that at least part of the association between race and underdetection of quality indicators in administrative data can be explained by the clinics where patients received care. Clinics that care for minority patients may be more likely to submit incomplete or inaccurate claims or health plan systems may allow more seamless transfer of data from some clinics than others. Our finding that the association of age and underdetection of hemoglobin A1c testing was partially explained by clinic type may reflect differences in claims processing for Medicare beneficiaries compared to younger patients within clinics. Alternatively, the finding of better detection of indicators in claims of patients

with two or more physicians documenting in the chart, which was partly explained by accounting for clinic, suggests that clinics where patients see more than one provider (often multispecialty groups) may submit more complete claims. Our finding that patients who saw more unique doctors were less likely to have their hemoglobin A1c underdetected may reflect the possibility that care abstracted from one medical record was insufficient to identify hemoglobin A1c values that may have been documented in other records located at other clinics. Unfortunately, we did not have any details on the characteristics of the clinics studied.

Our findings may have implications for the field of health disparities research, where much of the literature documenting disparities in care as a function of race, ethnicity, and socioeconomic status is based on administrative data (Ayanian et al. 1993; Guadagnoli et al. 1995; Klabunde et al. 1998; Bach et al. 1999; Einbinder and Schulman 2000; Kressin and Petersen 2001; Bradley, Given, and Roberts 2002; Potosky et al. 2002; Shavers and Brown 2002). Although most of this literature focuses on disparities in use of procedures, which are costly and thus more likely to be reported in administrative data than the indicators we studied, our findings nevertheless highlight the need to validate administrative data sources.

As noted above, performance on some quality indicators may vary by data source depending on the indicator, and both administrative data and medical records can provide inaccurate measures of quality (Fowles et al. 1999; Luck et al. 2000). For example, in our study, although medical records may be the best source for some indicators, eye examinations were identified more often with administrative data, likely because they are performed by other providers and may not be documented in the record of the patients' primary diabetes care provider. It is possible that patient and clinic characteristics could be associated with underdetection of compliance with quality standards based on medical records data that are not supplemented with administrative data; however, we did not explore this analysis because health care organizations are unlikely to rely on only medical record data for quality measurement.

Our study should be interpreted in light of several limitations. First, we studied patients who were members of three health plans in one state where patients receive relatively high-quality care (Jencks et al. 2000). The generalizability of our findings to other settings requires further study. Second, we obtained medical records from the office of the physician whom we determined to provide diabetes care for each patient, and for some patients we were unable to obtain records. More complete data collection might have

increased our rates of compliance with quality standards based on medical records, but it would have also likely identified more patients whose quality indicators were underdetected using only administrative data. Finally, we considered evidence for the indicator in either medical records data or claims data to be a gold standard (Fowles et al. 1999; National Committee for Quality Assurance 2000), and thus may have included a small number cases where the indicator was incorrectly identified as being performed.

In summary, our findings demonstrate that medical records data and administrative data provide differing information about quality of diabetes care. More important, our data suggest that certain groups of patients, such as elderly patients and black patients, may be more likely than other patients to have compliance with standards underdetected for some indicators if quality measurement with administrative data is not supplemented by medical records review. This finding suggests that providers, medical groups, or health plans who more often care for such patients may be disproportionately affected by public release of such data or by their use in determining financial incentives. Therefore, health care organizations that are collecting diabetes quality data for performance reports or financial incentives should consider using standardized indicators collected from both administrative data and medical records, such as those developed by the Diabetes Quality Improvement Project (Diabetes Quality Improvement Project 2001). Efforts to improve data collection, such as the development of comprehensive information systems incorporating key medical records data, will be important contributions to measuring and improving the quality of care delivered to patients.

ACKNOWLEDGMENTS

The authors are grateful to Yang Xu, M.S., and Robert Wolf, M.S., for their programming assistance.

REFERENCES

- American Diabetes Association. 2002. "American Diabetes Association: Clinical Practice Recommendations 2002." *Diabetes Care* 25 (supplement 1): S1–147.
- Ayanian, J. Z., I. S. Udvarhelyi, C. A. Gatsonis, C. L. Pashos, and A. M. Epstein. 1993. "Racial Differences in the Use of Revascularization Procedures after Coronary Angiography." *Journal of the American Medical Association* 269 (20): 2642–6.

- Bach, P. B., L. D. Cramer, J. L. Warren, and C. B. Begg. 1999. "Racial Differences in the Treatment of Early-Stage Lung Cancer." *New England Journal of Medicine* 341 (16): 1198-205.
- Bost, J. E. 2001. "Managed Care Organizations Publicly Reporting Three Years of HEDIS Measures." *Managed Care Interface* 14 (9): 50-4.
- Bradley, C. J., C. W. Given, and C. Roberts. 2002. "Race, Socioeconomic Status, and Breast Cancer Treatment and Survival." *Journal of the National Cancer Institute* 94 (7): 490-6.
- Diabetes Quality Improvement Project. 2001. Available at <http://www.dqip.org>. [Accessed September 24, 2001].
- Einbinder, L. C., and K. A. Schulman. 2000. "The Effect of Race on the Referral Process for Invasive Cardiac Procedures." *Medical Care Research and Review* 57 (supplement 1): 162-80.
- Epstein, A. 1995. "Performance Reports on Quality: Prototypes, Problems, and Prospects." *New England Journal of Medicine* 333 (1): 57-61.
- Epstein, A. M. 2000. "Public Release of Performance Data: A Progress Report from the Front." *Journal of the American Medical Association* 283 (14): 1884-6.
- FACCT—Foundation for Accountability. 1997. *FACCT Quality Measures: Diabetes Measurement Specifications (version 1.0)*, pp. 58-59. Portland, OR: FACCT.
- Fiscella, K., P. Franks, M. R. Gold, and C. M. Clancy. 2000. "Inequality in Quality: Addressing Socioeconomic, Racial, and Ethnic Disparities in Health Care." *Journal of the American Medical Association* 283 (19): 2579-84.
- Fowles, J. B., K. Rosheim, E. J. Fowler, C. Craft, and L. Arrichiello. 1999. "The Validity of Self-reported Diabetes Quality of Care Measures." *International Journal of Quality Health Care* 11 (5): 407-12.
- Guadagnoli, E., J. Z. Ayanian, G. Gibbons, B. J. McNeil, and F. W. LoGerfo. 1995. "The Influence of Race on the Use of Surgical Procedures for Treatment of Peripheral Vascular Disease of the Lower Extremities." *Archives of Surgery* 130 (4): 381-6.
- Iezzoni, L. I. 1997. "Assessing Quality Using Administrative Data." *Annals of Internal Medicine* 127 (8, part 2): 666-74.
- Iglehart, J. K. 1996. "The National Committee for Quality Assurance." *New England Journal of Medicine* 335 (13): 995-9.
- Jencks, S. F., T. Cuerdon, D. R. Burwen, B. Fleming, P. M. Houck, A. E. Kussmaul, D. S. Nilasena, D. L. Ordian, and D. R. Arday. 2000. "Quality of Medical Care Delivered to Medicare Beneficiaries: A Profile at State and National Levels." *Journal of the American Medical Association* 284 (13): 1670-6.
- Kiefe, C. I., J. J. Allison, O. D. Williams, et al. 2001. "Improving Quality Improvement Using Achievable Benchmarks for Physician Feedback: A Randomized Controlled Trial." *Journal of the American Medical Association* 285 (22): 2871-9.
- Klabunde, C. N., A. L. Potosky, L. C. Harlan, and B. S. Kramer. 1998. "Trends and Black/White Differences in Treatment for Nonmetastatic Prostate Cancer." *Medical Care* 36 (9): 1337-48.
- Kowalczyk, L. 2001a. "Harvard Pilgrim OK's Fee Hikes for Treating of 770,000 in Plan." *Boston Globe*, A1, June 20, 2001.

- . 2001b. "With Costs Up, HMOs to Monitor Patients: Blue Cross-Blue Shield, Tufts OK Claim Analysis." *Boston Globe*, A1, April 8, 2001.
- Kressin, N. R., and L. A. Petersen. 2001. "Racial Differences in the Use of Invasive Cardiovascular Procedures: Review of the Literature and Prescription for Future Research." *Annals of Internal Medicine* 135 (5): 352–66.
- Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.
- Localio, A. R., J. A. Berlin, T. R. Ten Have, and S. E. Kimmel. 2001. "Adjustments for Center in Multicenter Studies: An Overview." *Annals of Internal Medicine* 135 (2): 112–23.
- Luck, J., J. W. Peabody, T. R. Dresselhaus, M. Lee, and P. Glassman. 2000. "How Well Does Chart Abstraction Measure Quality? A Prospective Comparison of Standardized Patients with the Medical Record." *American Journal of Medicine* 108 (8): 642–9.
- Marshall, M. N., P. G. Shekelle, S. Leatherman, and R. H. Brook. 2000. "The Public Release of Performance Data: What Do We Expect to Gain? A Review of the Evidence." *Journal of the American Medical Association* 283 (14): 1866–74.
- National Committee for Quality Assurance. 2000. *The State of Managed Care Quality 2000*. Washington, DC: National Committee for Quality Assurance.
- Pedersen, C. A., E. C. Rich, J. Kralewski, R. Feldman, B. Dowda, and T. S. Bernhardt. 2000. "Primary Care Physician Incentives in Medical Group Practices." *Archives of Family Medicine* 9 (5): 458–62.
- Petitti, D. B., R. Contreras, F. H. Ziel, et al. 2000. "Evaluation of the Effect of Performance Monitoring and Feedback on Care Process, Utilization, and Outcome." *Diabetes Care* 23 (2): 192–6.
- Potosky, A. L., L. C. Harlan, R. S. Kaplan, K. A. Johnson, and C. F. Lynch. 2002. "Age, Sex, and Racial Differences in the Use of Standard Adjuvant Therapy for Colorectal Cancer." *Journal of Clinical Oncology* 20 (5): 1192–202.
- Schlackman, N. 1993. "Evolution of a Quality-Based Compensation Model: The Third Generation." *American Journal of Medical Quality* 8 (2): 103–10.
- Shavers, V. L., and M. L. Brown. 2002. "Racial and Ethnic Disparities in the Receipt of Cancer Treatment." *Journal of the National Cancer Institute* 94 (5): 334–57.
- Sussman, A. J., D. G. Fairchild, J. Coblyn, and T. A. Brennan. 2001. "Primary Care Compensation at an Academic Medical Center: A Model for the Mixed-Payer Environment." *Academic Medicine* 76 (7): 693–9.
- Wells, K. B., C. Sherbourne, M. Schoenbaum, N. Duan, L. Meredith, J. Unutzer, J. Miranda, M. F. Carney, and L. V. Rubenstein. 2000. "Impact of Disseminating Quality Improvement Programs for Depression in Managed Primary Care: A Randomized Controlled Trial." *Journal of the American Medical Association* 283 (2): 212–20.