

How Robust Are Health Plan Quality Indicators to Data Loss? A Monte Carlo Simulation Study of Pediatric Asthma Treatment

*Bruce Stuart, Puneet K. Singhal, Laurence S. Magder,
and Ilene H. Zuckerman*

Objectives. (1) To test the robustness of a health plan quality indicator (QI) for persistent asthma to various forms of data loss and (2) to assess the implications of the findings for other health plan quality measures.

Data Sources/Study Settings. Maryland Medicaid fee-for-service (FFS) claims. Children with asthma ($n = 5,804$) were selected from Medicaid enrollment records and medical and pharmacy FFS claims filed between June 1996 and December 1997.

Study Design. A variant of a HEDIS measure for treatment of persistent asthma (the percent of asthma patients filling two or more rescue medications who also filled a controller medication) was selected to test the robustness of proportion-based QIs to loss of data. Data loss was simulated through a series of Monte Carlo experiments.

Data Collection/Extraction Methods. Merged FFS medical and prescription claims.

Principal Findings. The asthma QI measure was highly robust to systematic and random data loss. The measure declined by less than 2 percent in the presence of up to a 35 percent data loss. Redundancy in the numerator of the QI significantly increased the robustness of the measure to data loss.

Conclusions. A HEDIS-related QI measure for persistent asthma is robust to data loss. The findings suggest that other proportion-based quality indicators, particularly those in which plan members have multiple opportunities to meet the numerator criterion, are likely to reflect true levels of health plan quality in the face of incomplete data capture.

Key Words: Quality indicator, HEDIS, data loss, asthma, managed care, Medicaid

Quality Indicators (QIs) have become benchmarks for evaluating health plan performance. Typically, QIs are based on evidence-driven measures of quality of care for plan populations as a whole (e.g., percentage of members receiving an annual flu shot) or subpopulations with particular diseases (e.g., percentage of diabetics receiving annual eye exams). The Health Plan Employer Data and Information Set (HEDIS) measures developed by the

National Committee for Quality Assurance (NCQA) stand at the forefront of health plan quality measures. The HEDIS measures are designed to permit a systematic and standardized approach to plan performance measurement. Depending upon the type of measure, health plans can choose to use administrative data (primarily, encounter data), medical record reviews, or client surveys to calculate these measures. Most plans find encounter data to be most efficient to use. However, several studies have reported that health plans face persistent problems in obtaining complete encounter data (Gold et al. 1995; Aizer, Felt, and Nelson 1996; National State Auditors Association 2002). In a survey of 108 managed care plans from 20 metropolitan areas nationwide, Gold et al. (1995) found that less than a quarter of plans received 90 percent or more of encounter data from their contracted physicians. A recent audit of four Medicaid managed care programs by the National State Auditors Association (2000) found that about one-third of the encounter data were lost. These circumstances raise an obvious question of how valid HEDIS measures are if the underlying data sources from which they are computed are incomplete.

One important factor affecting the robustness of a QI measure to data loss is the manner in which the measure is constructed. In a study by Dresser et al. (1997), HEDIS rates for cervical cancer screening from administrative data were very close to those obtained from chart reviews, whereas the rates for pediatric immunization and prenatal care differed greatly. In the case of cervical cancer screening, the authors speculated that because both the clinician who performs the Pap test and the clinician who reads it may have recorded the event, the likelihood that true events are accurately measured is high even in the presence of data loss; that is, redundancy improves validity. On the other hand, the robustness of a QI measure to data loss is reduced if it is based on multiple events. In the Dresser study, for example, the QI measure for cancer screening required just one Pap test over three years but the QI

Supported by a grant from the Agency for Healthcare Research and Quality, "Impact of Managed Care Organization Policies on the Quality of Pediatric Asthma Care," no. U01 HS09950.

Address correspondence to Bruce Stuart, Ph.D., Professor and Director, The Peter Lamy Center on Drug Therapy and Aging, University of Maryland School of Pharmacy, 515 W. Lombard Street, Room 158, Baltimore, MD 21201. Puneet K. Singhal, B.Pharm., M.S., is a Graduate Assistant at the University of Maryland School of Pharmacy, Baltimore. Laurence S. Magder, Ph.D., is an Associate Professor, University of Maryland School of Medicine, Baltimore. Ilene H. Zuckerman, Pharm.D., is an Associate Professor, University of Maryland School of Pharmacy, Baltimore.

measure for pediatric immunizations required nine visits over a two-year period.

Many HEDIS measures are computed as proportions of patients receiving appropriate care. Examples include beta blocker treatment after heart attack, cholesterol management after acute cardiovascular event, comprehensive diabetes care (annual eye exam, hemoglobin A1c testing, lipid screening, nephropathy monitoring), follow-up after hospitalization for mental illness, use of appropriate medications in those with persistent asthma, and prenatal and postpartum care (<http://www.ncqa.org>). Each of these measures requires identification of a particular group of individuals, constituting the denominator of the proportion, based upon service encounter information (e.g., a prescription for insulin for identifying diabetics). The rate of a specific *QI* is then determined based upon another service encounter or encounters (e.g., eye exam), which constitutes the numerator of the proportion. Proportion-based *QI* measures are inherently more robust to data loss than count-based measures because any loss is likely to reduce the value of both the numerator and the denominator. In the special case where data loss is proportional in the numerator and denominator, the measure itself will be unaffected by the loss. There are, however, both empirical and mathematical reasons to suspect that proportion-based measures will degrade in the presence of data loss, particularly if the loss is severe. For example, if the numerator is a relatively rare event compared to the denominator, a given absolute loss of data will have a proportionally greater impact on the numerator, driving the fraction downward. On the other hand, as noted in the Dresser study, redundancy in the numerator makes the measure less sensitive to data loss.

This article examines the robustness of a particular proportion-based quality indicator to various threats of data loss. We also assess the implications of the findings for other proportion-based HEDIS measures. The motivation for the analysis was a study of changes in quality of asthma treatment for a population of Medicaid children transitioning from fee-for-service (FFS) to managed care. The quality indicator was based on FFS claims in the before period and encounter data in the follow-up period. We had high confidence in the completeness of claims data, but much less confidence with the encounter data. Our challenge was to determine the validity of our quality indicator in the face of presumed, but unknown levels of data loss from the managed care plans. The remainder of the paper is organized as follows. The next section describes traditional approaches to dealing with missing data problems and explains why these did not address our particular concerns. Sections

describing the study setting, methods, and results follow. We close with a discussion of the applicability of our methods for assessing the robustness of other HEDIS quality indicators to data loss.

TRADITIONAL APPROACHES TO DEALING WITH MISSING DATA

Various methods have been developed to deal with missing data problems. These methods may be broadly classified into four categories: (1) complete case analysis, (2) imputation, (3) maximum likelihood, and (4) weighting methods (Kalton 1983; Little and Rubin 1987; Allison 2001). Complete case analysis, also known as listwise deletion, uses only those cases that do not have missing observations on any of the study variables. This method is appropriate when the proportion of missing data is not large and data are missing completely at random (MCAR). The MCAR assumption holds when the missingness is unrelated to the missing values and also unrelated to any other variable in the dataset (Allison 2001). Imputation methods are used to fill in the missing values by using information from complete cases. Missing values can be imputed by either using unconditional means method (taking a simple mean of the variable with missing data) or by using conditional means method (regressing variable with missing values on other variables in the dataset to impute predicted value of the missing data). Estimation can further be improved by using multiple imputation techniques (e.g., hot deck) which, instead of imputing a single set of draws for the missing values, creates multiple datasets each containing a different set of draws of the missing values from their predictive distribution (Little and Rubin 2000). Another approach to multiple imputation uses maximum likelihood methods that model the process by which missing data are generated (Little and Rubin 2000). Both imputation and maximum likelihood approaches assume that the missing data are missing at random; that is, that missingness does not depend on the missing values (Little and Rubin 2000). A final approach to dealing with missing data is designed to produce accurate population-level statistics from surveys in the face of systematic differences in sample capture rates by assigning higher weights to groups with lower response rates.

None of these traditional methods is appropriate in circumstances where the degree, or even the presence of missingness, is unknown. But this is precisely the problem that we faced in assessing the usefulness of health plan encounter data. Unlike situations where missing data are obvious (e.g., missing

lab values for tests known to have been conducted) zero encounters in a health plan dataset could represent either a true finding of nonuse or an artifact of missing data. Thus, unless there is an independent way to verify whether all encounters on behalf of the plan enrollees are recorded, one cannot distinguish which observations are truly complete (or incomplete). This rules out traditional imputation or weighting methods because they all require at least one set of complete data. Moreover, these methods make strong, generally untestable assumptions about the nature of missing data and its mechanism which, if wrong, will produce biased results (Little and Rubin 2000; Allison 2001).

Another approach to the missing data problem was recently developed by Le Corfec, Chevret, and Costagliola (1999). They used a Monte Carlo simulation technique to assess whether lost follow-up data in an HIV clinical trial would bias various summary statistics relating to the outcomes of the trial. Using a longitudinal clinical trial dataset with no missing values, they simulated known patterns of data loss to establish which statistics were robust to lost follow-up for 17 percent and 25 percent of the patient population. By comparing the outcomes of the simulated tests with data loss to the actual test results in the complete dataset, they were able to identify three summary statistics that were insensitive to data loss. This approach is appealing because it provides a strategy that analysts can use to select measures vetted on the basis of their robustness to data loss before they undertake the study.

STUDY SETTING

The genesis of our work is a grant funded by the Agency for Healthcare Research and Quality (AHRQ) dubbed Project INHALE (INitiatives to Help Asthmatics Live Easier). The primary project aim was to track changes in treatment quality for children with asthma enrolled in Maryland Medicaid during a transition from fee-for-service (FFS) to managed care under a program known as HealthChoice. Initiated on July 1, 1997, the HealthChoice program required most Medicaid recipients in the state to enroll in one of nine managed care plans (later dropping to five plans). Study subjects ($n = 5,804$) were selected based on a review of Medicaid enrollment records and medical and pharmacy FFS claims filed between June 1, 1996, and December 31, 1997. Selection criteria included age (between 5 and 18 years old at time of managed care organization [MCO] enrollment), a minimum of three months of continuous Medicaid enrollment during the baseline FFS period, and

evidence of asthma at baseline. Evidence of asthma was based on meeting one or more of the following three criteria: (1) ≥ 1 medical or hospitalization claim with an asthma diagnosis (ICD9 493.x) and ≥ 1 claim for an asthma related drug; (2) ≥ 2 medical or hospitalization claims with an asthma diagnosis; or (3) ≥ 3 pharmacy claims for an asthma related drug. Children with a diagnosis of cystic fibrosis (ICD9 277) were excluded from the study.

The project used fee-for-service claims, plan enrollment data, and MCO encounter forms to follow members of the study cohort through the transition into HealthChoice. Data files were created with the person-month as the unit of analysis. Variables tracked each month included Medicaid eligibility, setting (FFS, MCO with plan indicators), index dates for MCO enrollment and disenrollment, and asthma-specific services received (prescriptions, medical encounters, hospital and emergency department visits). However, as the study progressed, it became evident that data capture from the HealthChoice MCOs was inconsistent and erratic. This raised concern that measurement of plan performance indicators developed to test changes in quality of care over the transition would be biased by missing data in the follow up period.

METHODS

Our primary quality indicator for plan performance in prescribing asthma medications was based on the proportion of use (measured as prescription fills) of two types of asthma medications; rescue medications used to treat acute exacerbations of the disease, and controller medications designed to reduce the incidence of the acute attacks. The measure reflects treatment guidelines developed by the National Heart, Lung, and Blood Institute, National Asthma Education and Prevention Program (1997) for individuals suffering from persistent asthma. Lack of controller medication use in persistent asthma has been linked to significant morbidity and mortality. The QI is measured over a six-month period, thereby accounting for multiple (e.g., at home and school) and lost inhalers. Evidence of prescription fills for two or more canisters of beta₂-agonists (rescue medications) in a six-month period has been used as an indicator of persistent and potentially uncontrolled asthma (Zuckerman et al. 2000). Our QI measure is thus defined as the proportion of study subjects filling two or more prescriptions of beta₂-agonists who also filled at least one prescription of controller medication during the same six-month period. Higher values on this measure indicate better quality of asthma care.

We expected that this QI measure would be robust to incomplete reporting since it is based upon a proportion of two types (controller and rescue) of asthma medication and there is no a priori reason to believe that prescription records for one type are more likely to be recorded compared to the other type. However, for our QI measure, the numerator and denominator were determined by any fill over six months, and it was not obvious how various types and degrees of loss of monthly data would affect the proportion.

The first step in testing the robustness of our QI measure was to better understand the types and magnitude of potential data loss during the MCO period. For this purpose, we plotted the percent of subjects filling any prescription medication (not just asthma drugs) during each month pre- and post-MCO enrollment and the number of prescriptions filled by those who filled at least one prescription. Visual inspection of these time plots led us to classify potential data loss into two major categories:

1. *Systematic data loss*, where observations were unreported for all or most subjects in a given period. This type of data loss was more prevalent during the months immediately following the transition into the HealthChoice program. The leitmotif of systematic data loss was the presence of unexplained valleys in a utilization time series.
2. *Random data loss*, where there were unexplained gaps in observed use of services for some subjects during certain periods of time. While these processes may not be truly random, there is no obvious structure upon which to build an explanatory model. Random loss will generate a utilization time series that is below the true rate (and parallel to it if the extent of the loss is constant). An implausibly low mean utilization time series with unexplained peaks and valleys would suggest the presence of both systematic and random data loss.

The next task was to devise tests to determine the robustness of our QI measure to both types of potential data loss. We used a strategy similar to that used by Le Corfec et al. (1999). We ran two series of Monte Carlo simulations using just FFS prescription claims data for all study subjects continuously enrolled for six months, from December 1996 through May 1997. We used FFS prescription data since we expected these data to be complete due to their tie with reimbursement. The initial set of runs was designed to test the effects of systematic data loss from missing entire reporting periods. We first established the true QI rate across the entire FFS baseline dataset and then sequentially eliminated one to five months of data based on the following steps:

1. A random number generator (RANUNI function in *SAS* with the seed value of 0) was used to select a particular month (December to May) for which prescription records would be deleted.
2. Data were systematically deleted for every individual for that month.
3. The asthma QI rate was determined from the remaining data.
4. Steps 1 to 3 were repeated one thousand times.
5. The mean and standard error of the sample distribution were calculated for the one thousand iterations.
6. Steps 1 to 5 were repeated, deleting two, three, four, and, finally five months of data.

The robustness of the asthma QI measure to systematic data loss could then be empirically measured in terms of changes in the sample statistics as additional months of FFS claims were deleted.

The second set of Monte Carlo simulations tested the impact of random data loss. The procedures are analogous to those just described except that individual person-months of data (rather than entire months) were deleted:

1. All person-months were sorted randomly using a random number generator (RANUNI function in *SAS* with the seed value of 0).
2. Prescription records for 5 percent of person-months were deleted randomly.
3. The asthma QI rate was calculated from the remaining data.
4. Steps 1 to 3 were repeated one thousand times.
5. The mean and standard error of the sampling distribution were calculated for the one thousand iterations.
6. Steps 1 to 5 were repeated, randomly deleting additional person-months in 5 percent increments until 95 percent of the data were deleted.

Similar to the test of systematic data loss, the deterioration in the sample statistics represented a measure of the robustness of our QI to random data loss.

Once we had established the robustness of our QI to each form of data loss, we examined the impact of redundancy and prevalence of events forming the numerator. As noted previously, redundancy (recurrent controller medication fills by those who filled these medications) should increase the robustness of the measure while reduced prevalence of controller medication fills should have the opposite effect. To examine the impact of redundancy, we removed subsequent controller medication fills for individuals after their

initial fill for these medications, and then repeated the Monte Carlo simulations for both systematic and random data loss as described above. To assess the impact of reduced prevalence of individuals meeting the numerator criterion, we randomly removed controller medications for half of the person-months in the dataset, and reran the simulations as described above.

We tested the statistical significance of decrements in the QI indicator to various levels of data loss for each set of Monte Carlo simulations at $p < .05$. We also tested for significant differences in QI levels at the same level of data loss between simulations using the different assumptions. We used SAS version 8.2 for all analyses (SAS Institute 1989).

RESULTS

Of the entire INHALE cohort of 5,804 children with asthma, 4,378 (75.4 percent) were continuously enrolled during FFS from December 1996 through May 1997 and served as the sample frame for our robustness testing. Of these, 1,242 children filled two or more prescriptions of beta₂-agonists and among these, 736 children also filled one or more prescriptions for a controller medication. Thus, the baseline QI measure for this sample was .59. The mean rescue medication fill for those who filled at least two of these medications was 3.5 ± 2.2 prescriptions during the six-month study period. The mean controller medication fill in those who filled at least one of these medications who also filled two or more rescue medications was 3.2 ± 2.7 prescriptions.

Tables 1 and 2 present findings from the various Monte Carlo tests. The first column in each table presents results using the actual proportions of rescue and controller medications derived from the baseline sample. The second column in each table shows the impact of removing redundancy from the numerator of each redundant case, while the third column demonstrates the impact of cutting the sample proportion of controller medication fills by half.

Focusing first on the full baseline sample results, it is clear that the asthma QI measure is highly robust to both systematic and random data loss. The decrement in the estimated proportion of children filling controller medications was just 1.7 percent with loss of all data from up to two of the six monthly observation periods (Table 1) and up to 35 percent loss of individual months of observation (Table 2). This small decline is not statistically different from zero at $p < .05$. Beyond 35 percent data loss, the simulations generate QI

Table 1: Monte Carlo Simulation Results to Test the Robustness of an Asthma Quality Indicator^a to Systematic Data Loss

Loss of All Data for (x) Months		Baseline Sample		Baseline Sample with Redundant Controller Medication Fills Removed		Baseline Sample with Half of Controller Medication Fills Removed	
		QI Rate	Percent Deterioration in QI rate	QI Rate	Percent Deterioration in QI rate	QI Rate	Percent Deterioration in QI rate
Number	%	Mean (SE)	%	Mean (SE)	%	Mean (SE)	%
0	0.0	.59 (.00)	0.0	.59 (.00)	0.0	.29 (.00)	0.0
1	16.66	.58 (.00)	1.7	.53 (.07)	10.2	.29 (.00)	0.0
2	33.33	.58 (.01)	1.7	.46 (.10)	22.0	.28 (.00)	3.4
3	50.00	.56 (.01)	5.1	.39 (.11)	33.9	.27 (.01)	6.9
4	66.66	.53 (.02)	10.2	.30 (.12)	49.2	.25 (.01)	13.8
5	83.33	.46 (.04)	22.0	.21 (.11)	64.4	.22 (.02)	24.2

^aThe asthma quality indicator is defined as the proportion of children with asthma filling two or more rescue medications who also filled at least one controller medication during a six-month observation period based on fee-for-service claims.

Source: Maryland Medicaid fee-for-service claims data June 1996 and December 1997.

measures that are significantly different from the no-loss level. These tests suggest that systematic data loss leads to a somewhat greater degeneration in the validity of the QI measure at higher degrees of loss, but the rate of decrement is still low in both cases.

The middle column in Tables 1 and 2 shows what happens when redundancy in controller medication fills is removed. It is evident from this scenario that the QI measure is quite susceptible to both types of data loss in the absence of redundancy in the numerator. As seen in Table 1, the QI measure dropped by more than 10 percent with loss of one of the six monthly observation periods and 22 percent with the loss of two months (or a third of the sample). The susceptibility of the QI measure to random loss of person-month observations in the absence of redundancy was equally large (Table 2): a 10 percent loss of monthly observations resulted in a 5 percent decline in the QI measure, while a 20 percent loss drove the QI measure down by nearly 12

Table 2: Monte Carlo Simulation Results to Test the Robustness of an Asthma Quality Indicator^a to Random Data Loss

Loss of Data for (x) Person-Months	Baseline Sample		Baseline Sample with Redundant Controller Medication Fills Removed		Baseline Sample with Half of Controller Medication Fills Removed	
	QI Rate	Percent Deterioration in QI Rate	QI Rate	Percent Deterioration in QI Rate	QI Rate	Percent Deterioration in QI Rate
	Mean (SE)	%	Mean (SE)	%	Mean (SE)	%
0	.59 (.00)	0.0	.59 (.00)	0.0	.29 (.00)	0.0
5	.59 (.00)	0.0	.57 (.01)	3.4	.29 (.00)	0.0
10	.59 (.01)	0.0	.56 (.01)	5.1	.29 (.01)	0.0
15	.59 (.01)	0.0	.54 (.01)	8.4	.29 (.01)	0.0
20	.58 (.01)	1.7	.52 (.01)	11.8	.29 (.01)	0.0
25	.58 (.01)	1.7	.51 (.01)	13.5	.29 (.01)	0.0
30	.58 (.01)	1.7	.49 (.01)	16.9	.28 (.01)	3.4
35	.58 (.01)	1.7	.47 (.01)	20.3	.28 (.01)	3.4
40	.57 (.01)	3.3	.45 (.01)	23.7	.28 (.01)	3.4
45	.57 (.01)	3.3	.44 (.02)	25.4	.28 (.01)	3.4
50	.56 (.01)	5.1	.42 (.02)	28.8	.27 (.01)	6.9
55	.56 (.02)	5.1	.40 (.02)	32.2	.27 (.01)	6.9
60	.55 (.02)	6.8	.38 (.02)	35.6	.27 (.01)	6.9
65	.55 (.02)	6.8	.37 (.02)	37.3	.26 (.02)	10.3
70	.54 (.02)	8.5	.35 (.02)	40.7	.26 (.02)	10.3
75	.53 (.02)	10.2	.33 (.03)	44.1	.26 (.02)	10.3
80	.52 (.03)	11.9	.31 (.03)	47.5	.25 (.02)	13.8
85	.51 (.03)	13.6	.28 (.03)	52.5	.24 (.03)	17.2
90	.50 (.04)	15.3	.26 (.04)	55.9	.24 (.04)	17.2
95	.48 (.07)	18.6	.24 (.06)	59.3	.23 (.05)	20.7

^aThe asthma Quality Indicator is defined as the proportion of children with asthma filling two or more rescue medications who also filled at least one controller medication during a six-month observation period based on fee-for-service claims.

Source: Maryland Medicaid fee-for-service claims data June 1996 and December 1997.

percent. Generally, these declines were statistically different from zero at $p < .05$. In fact, a random loss of just 5 percent of the data produced a statistically significant decline from 59 to 57 percent in the QI measure. The drop was also significantly different from the result produced in the first set of simulations.

On the other hand, as the findings in the third column of each table attest, the QI measure was very robust to halving the proportion of study subjects meeting the numerator criterion. For systematic and random data losses of up to a third of the entire sample, there was a nonsignificant 3.4

percent decline in the simulated QI rate. Only at loss rates exceeding 50 percent of the entire sample was the rate of deterioration greater in these simulations compared to full-sample Monte Carlo simulations.

DISCUSSION

The potential of missing encounter data can present serious challenges in using HEDIS measures to determine true plan performance. The problem is exacerbated by the fact that traditional ways of handling missing data (e.g., listwise deletion, imputation, maximum likelihood, and weighting) are applicable only when the source and extent of the loss is known. That is generally not the case when lack of data may represent either a true nonuse of the service in question or a failure to report an event that did in fact occur. In this circumstance, sensitivity analysis is appropriate to determine the degree to which conclusions can be affected by various types and degrees of data loss. Computer simulation experiments are often a useful approach in this regard. In particular, when comparable complete data are available, the effect of various types of data loss can be assessed empirically by simulating the data loss from the complete data. This approach provides a strategy that analysts can use to proactively select measures on the basis of their robustness to data loss. In using this approach we built on the work of Le Corfec et al. (1999), who were interested in assessing the robustness of statistical tests to loss of follow-up on HIV patients enrolled in clinical trials. Our approach used fee-for-service claims data (assumed to be complete) to estimate the robustness of an asthma QI indicator to both known and suspected loss of encounter records from Medicaid managed care plans in the state of Maryland.

Our simulation procedures can be readily applied to other types of quality indicators. As with our asthma QI, a number of HEDIS plan performance measures are proportion-based. These measures offer some inherent protection from data loss to the extent that missing records will typically affect both numerator and denominator values. Our results suggest that the single most important element in a robust indicator is redundancy in the numerator. In our test, the average child with evidence of asthma controller drug use filled 3.5 prescriptions for these medications during a six-month period. Random data loss of up to 35 percent of all monthly observations resulted in a mean decline in the QI indicator of a mere 1.7 percent when all controller medications fills were recorded in the dataset. However, when we eliminated repeat controller fills, a 35 percent data loss

lowered the estimated QI measure by more than 20 percent. This finding provides empirical verification for an inference drawn by Dresser et al. (1997) that the HEDIS measure for cervical cancer screening is less susceptible to data loss than a QI for prenatal care and immunizations due to redundancy in the event recording. Extending this logic, we would expect that other HEDIS measures with a high likelihood of redundant measures of a numerator event would be more robust to data loss compared to those measures with a nonredundant numerator. For example, the HEDIS guideline for treating myocardial infarction calls for at least one prescription for a beta-blocker following a heart attack. We would expect health plan reporting on this measure to be highly robust to missing data because those who fill one prescription for beta-blockers are likely to do so on a regular basis. We would have a similar expectation for the HEDIS guidelines on cholesterol management. On the other hand, we would expect that the HEDIS measure for an annual eye exam for diabetics would be less robust to data loss because diabetics rarely get more than one annual eye exam, if that.

In our Medicaid sample we found that nearly 60 percent of the children with evidence of persistent asthma were getting the recommended controller medications. To test whether a high prevalence rate increased the robustness of the QI measure to data loss, we estimated one set of simulations with controller medication fills removed for half of the sample. The results were virtually unchanged. This would suggest that HEDIS measures with medium to high compliance rates should be relatively robust to data loss on that score. Although we did not perform additional tests with lower prevalence rates, we would expect to find that HEDIS measures with low numerator rates would be more susceptible to data loss. For instance, the HEDIS guideline for the follow-up care after hospitalization for mental illness specifies an ambulatory care visit within seven days of discharge. Such a narrow time window is likely to make the measure less stable in the presence of data loss.

One hopes that the problem of incomplete and missing encounter data will diminish as health plan information technology improves. Until that day arrives, prudence calls for having a clear understanding of the potential impact of data loss on health plan quality indicators. The methods described in this article represent a general approach rather than a definitive set of rules to address the problem. We tested the sensitivity of our QI measure to a limited selection of potential forms of data loss and it is possible that other types of systematic patterns of missingness would produce different results. For example, individuals with a missing month of data might be more likely to be missing observations for adjacent months (we tested that possibility in a

separate set of Monte Carlo simulations and found no difference from purely random losses). Our test was limited to pharmaceutical claims, which are more likely to be accurately reported than other forms of encounter data given the refined point-of-service technology used by the major pharmaceutical benefit managers. We did not test the robustness of our QI measure under circumstances in which either the numerator or the denominator is a rare event. However, the ready availability of claims files and a growing collection of validated encounter datasets will enable future researchers to test these possibilities and various other extensions of the methods describe here.

CONCLUSIONS

Continuing concerns with managed care encounter data systems raise the possibility that reported HEDIS quality measures will be biased by missing data. We used Monte Carlo simulation techniques to test the robustness to data loss of a quality indicator for treating pediatric asthma in a state Medicaid program as it transitioned from fee-for-service to manage care. The measure, a variant on the HEDIS guideline for treatment of persistent asthma, was based upon the proportion of fills of two types of asthma medications. This measure proved highly robust to most forms of data loss tested. Robustness was dependent upon redundancy in events related to the numerator in the fraction, but was relatively insensitive to the prevalence of the numerator value or to the structure of the data loss. The techniques described have general applicability to other health plan quality measures and to other settings than Medicaid. That said, it is important to emphasize that no simulation exercise can substitute for accurate and appropriate data collection.

REFERENCES

- Aizer, A., S. Felt, and L. Nelson. 1996. *Experience in Collecting Selected HEDIS 2.0 Measures for the Medicaid Population*. Washington, DC: Henry J. Kaiser Family Foundation.
- Allison, P. D. 2001. *Missing Data*. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-136 Thousand Oaks, CA: Sage.
- Dresser, M. V. D., L. Feingold, S. L. Rosenkantz, and K. L. Coltin. 1997. "Clinical Quality Measurement: Comparing Chart Review and Automated Methodologies." *Medical Care* 35(6): 539-52.

- Gold, M. R., R. Hurley, T. Lake, T. Ensor, and R. Berenson. 1995. "A National Survey of the Arrangements Managed-Care Plans Make with Physicians." *New England Journal of Medicine* 333(25): 1678–83.
- Kalton, G. 1983. *Compensating for Missing Survey Data*. Ann Arbor: Survey Research Center, Institute for Social Research, The University of Michigan.
- Le Corfec, E., S. Chevret, and D. Costagliola. 1999. "Visit-Driven Endpoints in Randomized HIV/AIDS Clinical Trials: Impact of Missing Data on Treatment Difference Measured on Summary Statistics." *Statistics in Medicine* 18(14): 1803–17.
- Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.
- . 2000. "Causal Effects in Clinical and Epidemiological Studies via Potential Outcomes: Concepts and Analytical Approaches." *Annual Review of Public Health* 21: 121–45.
- National Heart, Lung, and Blood Institute, National Asthma Education and Prevention Program. 1997. *Expert Panel Report 2: Guidelines for the Diagnosis and Management of Asthma*, publication no. 97-4051 Bethesda, MD: U.S. Department of Health and Human Services.
- National State Auditors Association. 2002. "2000 Joint Performance Audit: Medicaid Managed Care Encounter Data [accessed March 2003]." Albany, NY: National State Auditors Association. Available at http://www.nasact.org/techupdates/downloads/JAudits/06_02-Medicaid_Managed-Care.pdf.
- SAS Institute. 1989. *SAS*, statistical software, release 8.2 Cary, NC: SAS Institute.
- Zuckerman, I. H., B. Stuart, L. S. Magder, M. E. Bollinger, and S. R. Weiss. 2000. "Adherence of Asthma Treatment Guidelines among Children in the Maryland Medicaid Program." *Current Therapeutic Research Clinical and Experimental* 61: 912–24.

