Patterns of Unit and Item Nonresponse in the CAHPS[®] Hospital Survey

Marc N. Elliott, Carol Edwards, January Angeles, Katrin Hambarsoomians, and Ron D. Hays

Objective. To examine the predictors of unit and item nonresponse, the magnitude of nonresponse bias, and the need for nonresponse weights in the Consumer Assessment of Health Care Providers and Systems (CAHPS[®]) Hospital Survey.

Methods. A common set of 11 administrative variables (41 degrees of freedom) was used to predict unit nonresponse and the rate of item nonresponse in multivariate models. Descriptive statistics were used to examine the impact of nonresponse on CAHPS Hospital Survey ratings and reports.

Results. Unit nonresponse was highest for younger patients and patients other than non-Hispanic whites ($p < .001$); item nonresponse increased steadily with age ($p < .001$). Fourteen of 20 reports of ratings of care had significant ($p < .05$) but small negative correlations with nonresponse weights (median -0.06 ; maximum -0.09). Nonresponse weights do not improve overall precision below sample sizes of 300–1,000, and are unlikely to improve the precision of hospital comparisons. In some contexts, case-mix adjustment eliminates most observed nonresponse bias.

Conclusions. Nonresponse weights should not be used for between-hospital comparisons of the CAHPS Hospital Survey, but may make small contributions to overall estimates or demographic comparisons, especially in the absence of case-mix adjustment.

Key Words. Bias, weighting, missing data, mean-squared error, design effects, satisfaction

The CAHPS[®] hospital project is an effort jointly sponsored by the Centers for Medicare and Medicaid Services and the Agency for Healthcare Research and Quality to collect data that objectively measure patients' perceptions of hospital care and services. The CAHPS consortium has developed a standardized survey instrument for collecting information from recently discharged adult patients about care they received in medical, surgical, or obstetric sections of acute care hospitals. In this paper, we describe unit and item nonresponse patterns in the pilot version of the CAHPS Hospital Survey, which

was conducted in Arizona, Maryland, and New York. A subset of items on the pilot survey will be used in national implementation (see Goldstein et al. 2005).

Surveys generally do not yield complete responses from every individual sampled from the population. In certain situations, nonresponse can bias the survey findings if appropriate adjustments are not made. There are two basic types of survey nonresponse. Unit nonresponse is the failure of a member of the sample to respond to the survey as a whole. Item nonresponse is the failure of a unit respondent to answer one or more survey items that the respondent is eligible to answer. In this analysis, we examine and model patterns of both unit and item nonresponse to the pilot CAHPS Hospital Survey to assess the potential impacts of nonresponse bias and the corresponding adjustments. (Throughout this paper, nonresponse and nonresponse bias refer to unit nonresponse unless otherwise specified.)

There are other sources of bias that have the potential to affect analyses of the CAHPS hospital data. In particular, the mix of services and patient characteristics often varies significantly among hospitals. Thus, comparisons of CAHPS Hospital Survey data across hospitals or across different groups of patients may be biased unless case-mix adjustments are made to take such differences into account. It is important that nonresponse analyses of the CAHPS Hospital Survey data be made in conjunction with case-mix analyses, as one type adjustment may affect biases targeted by the other.

Thus, once eligible discharges have been selected from each hospital, there are at least two links from this targeted group of patients to the data ultimately obtained, each of which can threaten equitable comparisons among hospitals. Case-mix adjustments attempt to account for response bias among respondents—i.e., the systematic lack of correspondence between actual health care experiences and reports thereof among respondents (see O'Malley et al. 2005). Nonresponse analyses, the subject of the present manuscript, attempt to account for nonresponse bias resulting from differences between respondents and the full set of those targeted by the survey. Together, case-mix adjustment and nonresponse analyses enable survey estimates to more accurately reflect consumer experience of hospital performance and to provide more informative comparisons with consumers, hospitals, and other stakeholders engaging in quality improvement efforts.

Address correspondence to Marc N. Elliott, Ph.D., PO Box 2138, 1776 Main Street, Santa Monica, CA 90401. Dr. Elliott, Carol Edwards, B.A., Katrin Hambarsoomians, M.S., and Ron D. Hays, Ph.D., are also with RAND at Santa Monica, CA. January Angeles, M.P.P., is with the American Institutes for Research (AIR), Washington, DC.

BACKGROUND

The Problem of Nonresponse Bias

Nonresponse bias may be defined as systematic error in estimates (e.g., the proportion of patients at a given hospital whose bed pans were always changed promptly during stays in the last 6 months) that are attributable to systematic differences between the responses of those who do respond and the responses that would have been obtained from nonrespondents had they responded.

The overall precision of survey-based estimates is typically measured by mean-squared error (MSE). This term, which measures the expected squared deviation of a survey estimate from the true value, is the sum of the sampling variance of the estimate and the square of the bias of the estimate. Bias can be a particularly important source of survey error because, unlike variance, it does not decrease with increasing sample size. For example, a 1932 survey of several million eligible voters by the Literary Digest estimated the proportion that would vote for Franklin D. Roosevelt with an error of more than 30 percentage points, because of the presence of substantial bias. Bias, including nonresponse bias, is of the greatest concern when sample sizes are large, because bias can easily become the dominant factor in limiting overall precision.

Nonresponse bias is not an automatic consequence of selective nonresponse, but is related to the “product” of the degree of selective nonresponse and the correlation of this pattern with the quantity being estimated. Substantial nonresponse bias requires that both of these conditions be met. This means that nonresponse bias is inherently specific to each outcome under consideration.

We provide a simplified example to motivate and illustrate our approach. Imagine that our population consists of two strata in proportions q and p . Such strata are not typically identified a priori. The true population mean for the parameter we are interested in estimating is higher by the amount a in the second stratum than in the first. Imagine further that we have complete response for the first stratum, but incomplete response for the second stratum. We will label the proportion of the second stratum not represented as u so that up is the proportion of the entire population not represented. The resulting estimate for the population mean will have a nonresponse bias with magnitude

$$b = (aupq)/(1 - up).$$

This bias is approximately a product of the extent of underrepresentation (u) and the difference (a) in the underrepresented and fully represented strata.

The bias increases linearly with a and more than linearly with u . We will apply this formula to illustrate two scenarios with a dichotomous outcome.

Scenario A: Twenty percent underrepresentation of a 10 percent stratum that differs in responses by 15 percentage points, resulting in a 0.3 percent nonresponse bias.

Scenario B: Forty percent underrepresentation of a 20 percent stratum that differs in responses by 25 percentage points, resulting in a 1.7 percent nonresponse bias.

On the other hand, if the underrepresented stratum was unrelated to all outcomes ($a = 0$, e.g., patients with social security numbers ending in “3”), the resulting bias would be zero for all outcomes.

Biases of 0.3–1.7 percent may seem small, but they are large when considered in terms of the sample size required to offset these losses of precision. Effective sample size (ESS) is the size of the unbiased simple random sample that is equivalent in precision to a given sample. If we use ε to represent standardized bias ($\varepsilon = b/\sigma$), it can be shown that the ESS for an estimate is $n/(1 + n\varepsilon^2)$. Let us assume that the dichotomous outcome in scenarios A and B has a mean of 50 percent. In that case, the relatively mild bias of scenario A translates into $\varepsilon = 0.006$ ($0.3/50$) and the more substantial bias of scenario B translates into $\varepsilon = 0.034$ ($1.7/50$). Scenario A bias turns a sample size of 20,000 for a population into an ESS of 11,628 and turns a single hospital sample size of 300 into an ESS of 297. The more severe bias in scenario B results in an ESS of 829 and 223, respectively.

Biases are particular to the comparison being made, because the bases of selective nonresponse may have different associations with different estimates. In the CAHPS Hospital Survey, comparisons of hospitals are of particular interest. Bias that affects hospitals differentially may therefore be of more concern than bias that affects all hospitals equally.

Solutions to Nonresponse Bias

Perhaps the most effective approach for limiting nonresponse bias is to limit selective nonresponse at the design stage, as analytic approaches will only correct bias to the extent that the nonresponse is well modeled. Both specifying the correct functional form and obtaining the relevant predictors for nonrespondents can be challenging. Because very high response rates place some limit on the maximum amount of selective nonresponse that can occur, there is a general belief that higher response rates are likely to result in lower levels of selective nonresponse and hence nonresponse bias (Groves and

Couper 1998). Low response rates, however, do not necessarily imply selective nonresponse and fairly high response rates can still contain substantial selective nonresponse (Bootsma-van der Wiel et al. 2002). Even beyond response rate effects, there is evidence that multimode approaches, such as mail survey with telephone follow-up of initial nonrespondents, reduce nonresponse bias, because certain members of the population are more likely to respond to each mode of data collection (Fowler et al. 2002; Zaslavsky, Zaborski, and Cleary 2002).

The most common analytic approach to reducing the effects of nonresponse bias is nonresponse weighting. This approach uses information available for both respondents and nonrespondents to model probabilities of response conditional on these available variables. Under the assumption that respondents and nonrespondents do not differ in the quantity being measured after conditioning on these variables, nonresponse weights that are inversely proportional to these predicted probabilities are applied to respondents in order to represent the nonrespondents. If these assumptions are met, nonresponse bias is eliminated. In practice nonresponse bias is reduced to the extent that the assumptions are approximately true. For example, Fowler et al. (2002) found that nonresponse weighting of mail respondents did not reproduce the results obtained with phone follow-up of nonrespondents.

While nonresponse weighting is likely to reduce nonresponse bias, it does not necessarily improve the total precision of estimation MSE, because weighted estimation increases the variance of estimates. Weighting multiplies variances (and divides ESS) by a factor called the design effect (DEFF), which in the case of weighting is approximately equal to $1 + CV^2$, where CV is the coefficient of variation of the weights in the sample. In our simplified model of nonresponse used above, it can be shown that DEFF is approximately $1 + pqu^2$ with known strata. As the formula implies, variance inflation from nonresponse weighting increases quadratically with selective response, but there is a corresponding reduction in squared bias (and an improvement in MSE) only to the extent that this selective nonresponse is correlated with the outcome of interest.¹ Creating nonresponse weights for a pattern of underrepresentation uncorrelated with outcomes of interest would decrease precision without affecting nonresponse bias. There are a variety of other techniques used to model nonresponse, including approaches that use late respondents to model nonrespondents, on the assumption that late respondents are the respondents who most resemble nonrespondents (e.g., Paganini-Hill et al. 1993).

Predictors of Nonresponse to Health Surveys

A number of prior studies examined the characteristics of individuals least likely to respond to health surveys. (The studies cited below refer to surveys conducted in the U.S. unless otherwise noted.) Many studies have found nonrespondents more likely to be male (Ware and Berwick 1990; Hays et al. 1991; Mishra et al. 1993; Barkley and Furse 1996; Burroughs et al. 1999). Some studies have found higher nonresponse for younger patients (Hays et al. 1991; Barkley and Furse 1996), whereas others have found the opposite (Ware and Berwick 1990; Burroughs et al. 1999; Zaslavsky, Zaboriski, and Cleary 2002). These apparent inconsistencies in age patterns may be the result of a curvilinear relationship in which nonresponse is highest for the young and the very old. Using a mixed-mode survey, Zaslavsky, Zaboriski, and Cleary (2002) found higher nonresponse for all race/ethnic groups relative to non-Hispanic whites, with high nonresponse independently predicted by zip codes that were predominantly Asian, Hispanic, and urban among Medicare beneficiaries.

Several studies have found that nonrespondents to health surveys are less healthy than respondents. For example, Cohen and Duffy (2002) and Paganini-Hill et al. (1993) found that among those alive at the time of an initial health survey, nonresponse predicted substantially shorter survival in Scotland and California. Similar results have been found using other measures of health (Mishra et al. 1993; Hoeymans et al. 1998; Hoff et al. 1999).

The pattern with respect to utilization is mixed, with Grotzinger, Stuart, and Ahern (1994) finding more inpatient utilization among nonrespondents, Lamers (1997) finding no inpatient utilization differences, but less outpatient utilization among nonrespondents, and Gasquet, Falissard, and Ravaud (2001) finding fewer inpatient stays among nonrespondents.

The limited available information suggests that nonrespondents may have less favorable perceptions of care than respondents. Rubin (1990) reviewed two studies in which nonrespondents appeared to be less satisfied with care than respondents, although one study was based on proxy respondents and both were limited to psychiatric populations. Another survey of patients (Barkley and Furse 1996) found earlier respondents to be more satisfied with their care than late respondents, which has implications for nonresponse bias only to the extent that late respondents resemble nonrespondents in this regard. Zaslavsky, Zaboriski, and Cleary (2002) found that nonresponse rates were higher in Medicare managed care plans with lower ratings, but this does not necessarily imply a within-plan effect at the patient level.

Item Nonresponse

Item nonresponse involves loss of information and, if not carefully handled, can result in substantial biases as well. The shortcomings of naïve approaches, such as complete case analysis, available case analysis, and unconditional mean imputation, are well known. Superior approaches include stochastic regression imputation, maximum likelihood methods, and multiple imputation (see Little and Rubin 1987; Little and Schenker 1995, for discussion of these techniques).

Several studies have found higher rates of item nonresponse to health surveys among those with poor health, cognitive impairment, or physical impairment (e.g., Colsher and Wallace 1989; Guadagnoli and Cleary 1992). Whereas Colsher and Wallace (1989) found respondent age independently associated with item nonresponse rate among the community-dwelling elderly and Sherbourne and Meredith (1992) found a similar increase in item nonresponse among patients with chronic conditions, Guadagnoli and Cleary (1992) found no independent effect of age among recently discharged surgical patients.

METHOD

Sample and Survey Administration

The pilot CAHPS Hospital Survey collected data on 49,812 adult patients who were discharged to home between December 2002 and January 2003 from the medical, surgical, and obstetrics services of 132 hospitals in three states (Arizona, Maryland, and New York). The sample excluded patients with psychiatric diagnoses or who were missing administrative data necessary for sampling and tracking. The participating hospitals included 24 core hospitals (20,376 patients, for an average of 849 per hospital) and 108 noncore hospitals (29,436 patients, for an average of 273 per hospital). Core hospitals were deliberately selected among the full set for a variety of reasons, including the capacity to recruit a large sample and variation in hospital characteristics across the set. Core hospitals sampled approximately 300 patients from each of the three services (900 total); other hospitals sampled approximately 300 patients from the three services combined. Sampled totals fell below these goals only when insufficient patients were available. Likewise, patients who had previously received a different hospital satisfaction survey were sampled only when other eligible patients were unavailable. All patients were initially contacted by mail, with phone follow-up for nonrespondents in core hospitals (9,504 respondents, an average of 396 per hospital, for a 47 percent response

rate²), and mail follow-up for other hospitals (10,216 respondents, an average of 95 per hospital, for a 35 percent response rate³). Note that the design confounds mode of response and hospital characteristics. The survey was available in Spanish as well as in English; 3 percent of respondents chose to complete in Spanish. For additional details regarding the pilot CAHPS Hospital Survey data collection process, please see Goldstein et al. (2005).

Administrative Variables

A number of administrative variables describing patient and institution characteristics were available for all sampled patients. Patient variables used in the present study were age (18–24, 25–34, 35–44, 45–54, 55–64, 65–74, 75–79, 80 years or older), female gender, race/ethnicity (non-Hispanic white, black, Hispanic, Asian American, Native American, unknown, or missing), an indicator for whether Spanish was spoken at home; length of stay (1, 2–3, 4–7, 8–14, 15 or more nights), admission source (emergency room versus all others, which include standard referrals and transfers), Major Diagnostic Group (MDC), with the seven categories that were observed in less than 1 percent of cases in the sampling frame collapsed into a single “other” category, resulting in a total of 17 categories,⁴ and discharge status (discharged sick, left against medical advice, standard discharge to home). Institutional characteristics used in this study included state (Arizona, Maryland, and New York), service (surgery, obstetrics, and medical), and an indicator for being served by a core hospital.

Survey-Based Variables

The pilot version of the CAHPS Hospital Survey fielded had 66 items, 32 of which were retained for a shortened survey, here referred to as CHS-32. Survey-based variables used in the present study fall into three categories: actual responses to survey items, derived variables describing the proportion of items with missing responses, and a variable measuring the number of days between when the first survey was mailed out and when the survey was completed (by either mode). Actual survey responses used in the present study for evaluating the impact of weights are the responses to the 16 reports of care items (14 with Never/Sometimes/Usually/Always responses and two with dichotomous Yes/No responses) and four global rating items (three on a 0–10 scale, one on a four-point scale) that were retained in CHS-32; item nonresponse measures include some items not retained on the CHS-32. These items will be described by their placement in the original survey. Item nonresponse was measured by two derived variables that include some items that were

subsequently dropped from the survey. Items regarding pain were not part of item nonresponse calculations because skip patterns differed by mode.

The first measure of item nonresponse, *proportion missing*, is the proportion of missing items for a given respondent among the 42 items on the original survey asked of all respondents.⁵ These include ratings and reports of care, screener items, and demographic information. Twenty of these 42 items were retained in the CHS-32. The second measure of item nonresponse, *proportion of report items missing*, is the proportion of missing items for a given respondent among the *applicable* subset of 30 report items on the original survey.⁶ The second measure differs from the first in that it does not consider screeners or demographic items. The denominator for the second measure is determined by responses to the screener items. Twelve survey items were common to the CHS-32 and the pilot survey.

Analysis of Unit Nonresponse

We used a series of multivariate logistic regression models to analyze the factors associated with unit nonresponse. The first model included the full set of 11 administrative variables described above, with 41 predictor degrees of freedom. In the subsequent models, we altered the model to consider possible interactions and to determine the most parsimonious specifications. In the second model, we parsed the original parameterization by collapsing categories with similar coefficients. For the third model, we examined all possible two-way interactions among terms in the second model. We performed block tests of significance on the pairwise combinations of the 11 categorical variables, and retained those blocks that were significant at $p < .01$. (Age and length of stay were parameterized with a single degree of freedom each within interaction terms.) Inverse probability weights were generated from the predictions of each of the three models. DEFFs for these nonresponse weights were calculated in order to assess the variance inflation from the use of non-response weights.

To estimate the effects of these three sets of weights on nonresponse bias, simple Pearson correlations were computed between these nonresponse weights and the set of 20 reports and global rating items included in the CHS-32. A true correlation of zero would mean that the probability of response to the survey, as modeled in our logistic regressions, was unrelated to an outcome. This would imply that the weights do not eliminate any response bias. To the extent that we have correctly and completely modeled nonresponse, this implies that no nonresponse bias is present in uncorrected estimates of that

outcome. The absolute magnitude of a true nonzero correlation would increase with the amount of nonresponse bias corrected by the weights. A positive correlation between weights and an outcome would mean that scores are higher for those identified as less likely to respond, and by extension among nonrespondents, if the assumptions of nonresponse weighting hold. This would imply that unadjusted scores underestimate true values, while a negative correlation would mean the opposite. Statistically significant correlations provide evidence against the null hypothesis that weights accomplish no bias reduction. Additionally, the overall means for the 16 retained reports of care and the four global ratings were computed with and without the weights.

To assess the extent to which the reduction of response bias through case-mix adjustment affected nonresponse bias, we performed parallel versions of these analyses on case-mix-adjusted residuals. Case-mix models were linear regressions parameterized by service dummies, linear age, a non-Hispanic black indicator, linear education, linear self-reported health status, an indicator of Spanish language spoken at home, and an indicator of circulatory MDC, along with interactions of service with all other case-mix variables.

Also of interest is the effect of nonresponse bias on comparisons between hospitals. In order to assess the magnitude of this bias, weighted and unweighted hospital-level estimates of one report item and one global rating were compared for the 24 core hospitals. These examples were selected because they had the two strongest correlations with nonresponse weights. These analyses were performed a total of six times, corresponding to combinations of the three sets of weights and unadjusted versus case-mix-adjusted means.

We used a multivariate regression model to examine the predictors of response time, the interval between when a case was fielded and when a complete response was received. The independent variables included the same set of administrative variables used to predict unit nonresponse. Pearson correlations were computed between the set of 20 retained report and global rating items and the measure of response time.

Analysis of Item Nonresponse

Descriptive statistics were computed for both measures of item nonresponse. Because both of these variables had distributions that were highly positively skewed, two multivariate ordered logistic regressions were used to predict each measure of item nonresponse from the same set of administrative predictors used in the unit nonresponse analyses.

RESULTS

Descriptive Statistics

The overall response rate was 39.6 percent (19,720 of 49,812). The second column of Table 1 describes the distribution of the 11 categorical predictors (41 degrees of freedom) used in the first multivariate model of unit nonresponse and the multivariate models of item nonresponse and time to response. The third column describes response rates within these 41 categories. These bivariate response rates ranged from less than 20 percent for patients who left the hospital against medical advice to over 50 percent for patients with male reproductive disorders.

Predictors of Unit Nonresponse

The fourth column of Table 1 presents the results of the first multivariate logistic regression predicting unit response. This model had a concordance of 65 percent, very similar to what was found by Zaslavsky, Zaborski, and Cleary (2002) when predicting nonresponse to the Medicare Managed Care CAHPS Survey among those beneficiaries with good contact information. The strongest predictors of unit nonresponse were age, race/ethnicity, and having left against medical advice. Nonresponse is highest for those 18–24, drops steadily to its lowest level at 65–74, and then rises by 80+ to the same level seen for those 45–54. Nonresponse is higher for males. Nonresponse is lower for non-Hispanic whites than for all other race/ethnic groups. Among Hispanics, nonresponse rates were lower for those who speak Spanish at home. The relationship between length of stay and nonresponse was somewhat complex, but there was a general tendency toward higher nonresponse among patients with longer stays. Nonresponse was higher for those admitted from the emergency room. Nonresponse varied by MDC, with nonresponse highest for injury/poisoning and lowest for health services/health status. Among more common MDCs, nonresponse was high for the respiratory MDC and low for the female reproductive MDC. Nonresponse was slightly higher for those discharged sick and considerably higher for those who left against medical advice. Nonresponse was lowest in Maryland, medical services, and core hospitals. Because some of the strongest predictors were very rare (left against medical advice, injury/poisoning MDC, health services/health status MDC), their importance in predicting overall nonresponse may be limited.

The second parsed model collapsed ages 55–79 in a single category, collapsed length of stay from 2 to 7 days, and combined surgical and obstetric services. It also combined all race/ethnic dummies, except for non-Hispanic

Table 1: Multivariate Predictors of Unit Nonresponse, Item Nonresponse, and Time to Response

	Analyses of Full Sample			Analyses of Respondents		
	Proportion of Sample (49,812)	Unadjusted Response Rate (%)	Coefficient for Unit Response ^{†,‡}	Coefficient for Proportion Missing of 42 Items Asked	Coefficient for Proportion Missing of Applicable Report Items [†]	Coefficient for Days Unit Response [§]
Age (years)						
18–24	0.11	25.7	– 0.43*** (0)	0.11 (0)	– 0.08 (0)	– 0.60 (0)
25–34	0.19	35.4				– 1.40*
35–44	0.14	35.1	0.13***	0.10	0.22***	– 2.60***
45–54	0.12	40.4	0.50***	0.13	0.27***	– 4.58***
55–64	0.12	46.6	0.77***	0.23***	0.34***	– 6.07***
65–74	0.14	49.7	0.89***	0.38***	0.57***	– 5.91***
75–79	0.07	48.1	0.83***	0.63***	0.78***	– 6.01***
80 +	0.11	39.3	0.50***	0.89***	1.00***	– 0.30
Female	0.66	39.9	0.20***	0.13***	0.24***	
Race/ethnicity						
Black	0.14	25.9	– 0.75***	0.52***	0.20***	3.92***
Hispanic	0.08	31.0	– 0.58***	0.66***	– 0.18	5.18***
Asian American	0.03	25.2	– 0.77***	0.56***	0.26*	2.95***
Native American	0.01	24.6	– 0.73***	0.45*	– 0.17	6.58***
Unknown or missing (Non-Hispanic white)	0.17	39.2	– 0.22***	0.14***	0.14***	3.66***
Spanish language	0.57	45.1	(0)	(0)	(0)	(0)
Length of stay (days)	0.02	38.8	0.39***	0.14	0.10	– 7.08***
1	0.22	38.4	(0)	(0)	(0)	(0)
2–3	0.45	39.8	0.14***	– 0.14***	– 0.26***	– 0.30
4–7	0.23	41.2	0.11***	– 0.19***	– 0.37***	– 0.19
8–14	0.07	37.9	– 0.09*	– 0.18***	– 0.48***	– 0.37
15 +	0.02	35.3	– 0.20***	– 0.02	– 0.26*	0.40
Admitted from emergency room	0.35	34.8	– 0.19***	0.10**	0.21***	0.56
Major diagnostic category (MDC)						
Nervous system	0.05	39.3	– 0.04	0.10	0.23***	0.40
Ear, nose, throat, mouth	0.01	37.8	– 0.08	– 0.03	0.15	2.05
Respiratory system	0.07	36.5	– 0.13***	– 0.08	0.00	– 1.13
Digestive system	0.10	40.6	0.03	– 0.03	0.01	0.74
Liver/pancreas	0.03	39.7	– 0.03	0.08	0.10	0.91
Musculoskeletal	0.08	46.9	0.08*	– 0.10	0.04	– 0.01
Skin/breast	0.03	36.9	– 0.15*	– 0.05	0.15	0.33
Metabolic disorders	0.03	36.1	– 0.13*	– 0.11	0.03	– 0.62

Continued

Table 1: Continued

Analyses of Full Sample			Analyses of Respondents			
Variable	Proportion of Sample (49,812)	Unadjusted Response Rate (%)	Coefficient for Unit Response ^{†,‡}	Coefficient for Proportion Missing of 42 Items Asked	Coefficient for Proportion Missing of Applicable Report Items [†]	Coefficient for Days Until Response [§]
Kidney/urinary tract	0.03	37.4	−0.13*	−0.01	0.02	1.18
Male reproductive	0.01	52.3	0.11	−0.26*	−0.13	−1.12
Female reproductive	0.05	48.5	0.15***	−0.20***	−0.05	−0.06
Infectious diseases	0.01	39.6	0.05	−0.18	−0.02	0.55
Injury/poisoning	0.01	30.4	−0.36***	−0.02	0.04	1.88
Health status/services	0.02	47.2	0.27***	0.02	0.26*	−1.37
Other	0.03	35.6	−0.11	−0.01	0.05	0.46
(Circulatory)	0.17	43.0	(0)		(0)	(0)
Discharge status						
Discharged sick	0.08	40.8	−0.14***	0.03	0.01	0.40
Left against medical advice	0.01	16.2	−0.83***	0.69**	0.08	−1.45
(Standard discharge to home)	0.91	39.6	(0)	(0)	(0)	(0)
State						
Arizona	0.23	40.6	−0.04	−0.12**	−0.09	−3.08***
Maryland	0.33	39.0	0.08**	−0.12***	−0.11**	−1.89***
(New York)	0.44	39.5	(0)	(0)	(0)	(0)
Service						
Surgery	0.34	46.5	0.32***	−0.10**	−0.02	−1.00**
Obstetrics	0.26	35.2	0.36***	−0.30***	0.10	−1.42*
(Medical)	0.40	36.5	(0)	(0)	(0)	(0)
Core hospital	0.42	45.6	0.43***	0.65***	0.26***	−0.87**
Concordance	NA	NA	0.652	0.541	0.570	NA
Adjusted-R ²	NA	NA	NA	NA	NA	0.025

[†]Log-odds.

[‡]First multivariate model of unit nonresponse.

[§]OLS coefficient.

*The MDC related to obstetrics is omitted here because it is identical to the obstetrics service indicator already in the model.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

white and unknown/missing, and combined all MDCs other than respiratory, female reproductive, and health status/services. This eliminated 19 degrees of freedom from the model and left the concordance at 65 percent. (Results for the second and third model are not shown.) In the third model, nine of 55 block tests of interactions were significant at less than .01: interactions of race, emergency room admission, and state with service; interactions of gender, length of stay, and core hospital with age; interactions of Spanish language and state with core hospital; and interaction of state with race/ethnicity. This model increased the concordance to 66 percent.

Nonresponse Weights

Table 2 describes the effects of nonresponse weights generated by each of the three models. As can be seen, results were remarkably consistent. The DEFFs from the nonresponse weights were modest, never exceeding 1.17. No correlations of the 20 evaluative survey items with nonresponse weights exceeded 0.10 in absolute value in the absence of case-mix adjustment. A majority of correlations (including all four global ratings) were significantly negative at $p < .05$ and several were significantly positive (including for all models Q49 regarding written information on what symptoms to look out for when at home, Q48 regarding talking about help after discharge, and Q18 regarding whether the room was quiet). The three strongest correlations were negative and came from Q15 (the MD rating), followed by Q11 and Q53 (MD showing respect and recommending the hospital). These results indicate that respondents who were most like nonrespondents tended to give slightly lower ratings and reports, particularly with respect to doctors.

Under model assumptions that nonrespondents have the same experiences on outcomes as respondents with the same values of variables in the nonresponse model, these results suggest a tendency toward higher nonresponse among those with less positive experiences with the hospital. These correlations are proportionate to the amount by which weights alter the estimated means, the largest effect being a reduction of the overall mean by 0.04 standard deviations and the median absolute effect being a change of 0.02 standard deviations for each of the three sets of weights without case-mix adjustment.

Combining this information with the DEFF, it can be shown that under standard assumptions, the minimum sample size for nonresponse weights to improve the MSE of single-population estimates for the median of the 20 rating and report items is between 330 and 400. At this break-even point, the

Table 2: Impact of Nonresponse Weights from First, Second, and Third Models

	<i>Design Effect</i>	<i>First Model (Untrimmed)</i>		<i>Second Model (Trimmed)</i>		<i>Fifth Model (Interactions, Trimmed)</i>	
		1.15		1.14		1.17	
No case-mix adjustment	Significant positive correlations with outcomes*	3		3		3	
	Significant negative correlations with outcomes*	14		14		15	
	Largest absolute correlation	0.09		0.09		0.10	
	Median absolute correlation	0.06		0.06		0.06	
	Largest bias reduction	0.04 standard deviations		0.04 standard deviations		0.04 standard deviations	
	Median bias reduction	0.02 standard deviations		0.02 standard deviations		0.02 standard deviations	
With case-mix adjustment	Minimum sample size for weights to improve MSE of median outcome	375		330		400	
	Percentage of nonresponse bias reduction accomplished by case-mix adjustment	65		74		63	
	Largest bias reduction	0.02 standard deviations		0.02 standard deviations		0.02 standard deviations	
	Median bias reduction	0.01 standard deviations		0.01 standard deviations		0.01 standard deviations	
	Minimum sample size for weights to improve MSE of median outcome	1,040		1,060		1,080	

* $p < .05$, of 20 outcomes.

MSE for half of the outcomes worsens slightly and the MSE for the other half improves slightly when nonresponse weights are applied, with no changes exceeding 0.04 percent of the individual-level MSE. The distributions of these small losses and gains in MSE are fairly symmetric across the set of 20 items at these break-even sample sizes, so that the magnitudes of improvements and decrements are similar. At all smaller sample sizes the impact of nonresponse weights more consistently increases MSE, whereas at larger sample sizes, they more consistently decrease it.

On average, case-mix adjustment eliminated about two-thirds of the bias eliminated by nonresponse weights. As a consequence, nonresponse weights would produce even smaller adjustments when applied after case-mix adjustment. After case-mix adjustment, break-even sample sizes rise to over 1,000; at sample sizes of 300 or less, weights for all models would inflate MSE for all 20 outcomes (results not shown).

To illustrate the effect of the third model nonresponse weights on comparisons between hospitals, we compare weighted and unweighted hospital-level means for the 24 core hospitals⁷ for two of the outcomes with the strongest correlations with the nonresponse weights: Q11 (MD shows respect) and Q15 (Rating of MD). The median hospital adjustments from weighting were 0.03 and 0.02 individual-level standard deviations, respectively. It is worth noting that these are somewhat smaller reductions within hospitals than were observed overall. This suggests that at standard sample sizes per hospital, weights would improve these comparative estimates for only one of the two most affected outcomes, and would harm a majority of these estimates, even in the absence of case-mix adjustment. Interestingly, case-mix adjustment only eliminates 11 and 14 percent of the bias corrected by these weights, respectively. Thus, the effects of weights on hospital comparisons are largely independent of case-mix adjustment.

The maximum inflation to MSE-based standard errors at sample sizes below the break-even thresholds discussed above is determined by the DEFF, and never exceeds 9 percent. The maximum reduction in MSE at large sample sizes corresponds to the maximum bias reductions reported in Table 2. While absolutely small, these can be large relative to the very small MSE-based standard errors of sample sizes much larger than 1,000.

Time to Response

Mean response time to the survey was 36 days, with a standard deviation of 19 days. The seventh column of Table 1 displays the results of a multivariate

linear regression predicting response time in days. This model had an adjusted R^2 of less than 0.03. The most important factors for response time were race/ethnicity/language and age. Length of response time decreased with age through 65–74, with the elderly responding about 6 days sooner than the youngest adults. Blacks, Native Americans, and Asian Americans responded 3–7 days later than non-Hispanic whites. Those who spoke Spanish at home responded 7 days sooner than those who spoke English at home. Combining these with race/ethnicity effects, Spanish-speaking Hispanics responded as quickly or more quickly than non-Hispanic whites, whereas English-speaking Hispanics responded about 5 days later than non-Hispanic whites. Response times were slower in New York, noncore hospitals, and for the medical service.

Correlations of the 20 evaluative items with response times were often mildly negative. Correlations for 11 of the items, including all four global ratings, were significantly negative ($p < .05$). The significant correlations ranged from -0.02 to -0.04 , with a median of -0.02 . The strongest correlation was with Q32 (regarding pain control). This pattern indicates that late respondents had slightly lower ratings and reports for some items. The correlations with response time had little correspondence to correlations with nonresponse weights. Because time to response was positively correlated with telephone mode, which is associated with more positive ratings and reports (see De Vries et al. 2005), the correlation of delay in response with ratings and reports is probably underestimated.

Item Nonresponse

The overall proportion of inappropriate missing responses for the 30 report items was 2 percent. The overall proportion of missing response for the 42 items asked of everyone was 4 percent. Table 3 summarizes the individual-level rates of item nonresponse. About one-fourth of respondents had at least one missing value for a report item for which they were eligible and about half of respondents failed to answer at least one of the 42 questions asked of all respondents. For the 30 report items, the proportion of applicable reports missing was 10 percent or more for 7 percent of respondents and 25 percent or more for 1 percent of respondents. For the 42 items asked of everyone, the proportion of reports missing was 10 percent or more for 11 percent of respondents and 25 percent or more for 3 percent of respondents.

The fifth and sixth columns of Table 1 display the results of multivariate ordered logistic regressions predicting the proportion of items missing for the 42 items that were asked of all respondents (fifth column) and for the up to

Table 3: Individual Rates of Item Nonresponse (%)

<i>Rate of Item Missingness (%)</i>	<i>Report Items (of Those for Which Respondent Was Eligible)</i>	<i>Items Asked of All Respondents</i>
0	78 %	50 %
1–4	7 %	27 %
5–9	8 %	12 %
10–24	6 %	8 %
25–49	1 %	2 %
≥ 50	0.4 %	0.4 %

30 report or global rating questions for which a given respondent was eligible (sixth column). These models had concordances of 54 and 57 percent, respectively. Age was the most important predictor of item nonresponse. For items asked of everyone, race/ethnicity, having left against medical advice, and core hospital status were also substantially predictive. Item nonresponse increased strongly and steadily with age. Non-Hispanic whites had the lowest proportion missing for items asked of everyone. The relationship of length of stay to the proportion of missing responses is curvilinear, with the highest rates of missingness for those with stays of 4–14 days. The proportion of missing responses is higher for those admitted from the emergency room, but varied little by MDC. Those who left against medical advice had very high rates of missingness of items asked of everyone. The proportion of missing responses was highest in core hospitals, especially for items asked of everyone.

DISCUSSION

We described patterns of unit nonresponse, time to response, and item nonresponse in the CAHPS Hospital Survey. Below we summarize these findings, compare them with the literature, and evaluate solutions, including nonresponse weighting, the use of response time to impute nonrespondents, and item imputation.

The young elderly (65–74) had the highest response rates. Those 65 and over had the quickest responses, but the proportion of missing responses increased steadily with age. This may reflect cooperativeness on the part of the elderly tempered with cognitive or health effects on item nonresponse, and, for the old elderly, unit nonresponse. Men had higher unit but lower item

nonresponse. Non-Hispanic whites had the lowest rates of unit and item nonresponse, and among the quickest responses. Those admitted from the emergency room had somewhat higher rates of unit and item nonresponse. Unit nonresponse varied somewhat by MDC. Those discharged sick had slightly higher rates of unit nonresponse. Those who left against medical advice had very high rates of unit and item nonresponse, suggesting that the act of leaving against medical advice is a marker for being uncooperative or less invested in the hospital experience.⁸ The medical service had the highest rates of unit and item nonresponse.

Our nonresponse findings on gender, race/ethnicity, and age are generally consistent with previous findings. Our findings are also consistent with the emerging evidence that nonrespondents rate their care less favorably than respondents. Our findings of increased item nonresponse with age are consistent with Colsher and Wallace (1989), but differ from the absence of a relationship noted in Guadagnoli and Cleary (1992).

Limitations of these findings include the following: (1) the importance of health status as a predictor of nonresponse in the literature and the limited health information available suggests there may be unmeasured variation in health status that determines additional unmeasured selective nonresponse; (2) there may be direct selection on the outcomes, with those with less positive experience being less likely to respond, even after controlling for these predictors; (3) nonresponse levels are probably higher for this 66-item instrument than will be the case for the CHS-32, and selective nonresponse patterns may differ for a shorter instrument; and (4) the patterns of nonresponse observed are likely to be somewhat specific to the mode of administration (two mail waves or mail with phone follow-up).

It is often thought that the mere existence of selective nonresponse necessitates the use of nonresponse weights for principled estimation. This is not necessarily the case. Under some circumstances, survey nonresponse produces little or no bias, and thus nonresponse weighting is at best an unnecessary complexity. In some cases, nonresponse weights may actually reduce the precision of estimates through variance inflation.

Inverse probability nonresponse weights generated from the multivariate model would produce a relatively modest variance inflation of 14–17 percent. Alternate parameterizations have little effect on variance inflation or bias reduction. These weights are mildly negatively correlated with a majority of report and rating items, suggesting that nonresponse weights would correct for a small bias from underrepresentation of patients with less positive experiences. In the absence of case-mix adjustment, these weights would slightly

improve overall estimates for sample sizes beyond 400. Case-mix adjustment on respondents eliminates a majority of the bias accounted for by nonresponse weights for overall estimates, so that weights would improve MSE only at sample sizes in the thousands. Nonresponse weights appear to account for no more bias in hospital comparisons than they do overall. At sample sizes of 300 or less, nonresponse weights are likely to worsen MSE for hospital comparisons. As case-mix adjustment does little to eliminate the effects of measurable nonresponse bias on hospital comparisons, this conclusion is independent of the use of case-mix adjustment.

While nonresponse weighting and main effects terms in regression imply a different functional form, there is a high degree of correspondence between the case-mix adjusters recommended for the CAHPS Hospital Survey and the important predictors of unit nonresponse. It may be that adding specified interaction terms could reduce remaining nonresponse bias at a lesser cost than that imposed by the DEFFs of weighting. In particular, one might consider as candidate case-mix adjusters interactions between existing case-mix variables and a nonresponse weight (without actually weighting or including a main effect of nonresponse weights). Future work might address these possibilities. There may also be value in developing a nonresponse weighting analog to the explanatory power (EP) statistic developed by Zaslavsky et al. (2001) for variable selection in case-mix adjustment models. Such a bias reduction power (BRP) statistic would aid variable selection in nonresponse models when collecting additional variables for the purpose of nonresponse modeling might be costly.

Of the 11 administrative predictors of response time, five or six generally follow the pattern in which those less likely to respond were also slower to respond, depending on the stringency of one's criterion for similarity. Five administrative predictors had patterns that were clearly inconsistent with the notion that late respondents resemble those with lower response rates. While mode of follow-up was confounded with the characteristics that distinguished core hospitals from other hospitals, the inconsistency of these associations provide no clear evidence for using late respondents to model nonrespondents. The fact that ratings and reports are less strongly correlated with timing of response than with probability of unit nonresponse further limits the potential of this approach to reduce nonresponse bias. Future work might consider percentiles of response within mode and institution.

Item nonresponse was generally low in this pilot version (2–4 percent). Levels of item nonresponse are likely to fall further for the shortened CHS-32, as items with poor cognitive properties are eliminated, the clarity of skip

patterns is improved, and total respondent burden is decreased. Nevertheless, the proportion of respondents with some missing items is too high (50 percent) to consider complete case analysis/listwise deletion, and will surely remain so. Even a 10 percent rate of casewise deletion would be unacceptable from the perspective of cost and statistical power alone, without even considering the selection bias such deletion might incur.

On the other hand, the proportions of missing items are low enough that the benefit from multiple imputation are likely to be limited (see Little and Schenker 1995), and single imputations based on stochastic regression, maximum likelihood, or Markov chain Monte Carlo methods (multiple imputation performed once) are likely to suffice. The PROC MI command in *SAS*, when set to a single imputation, provides one way to implement such an approach.

CONCLUSION

Pilot CAHPS Hospital Survey data suggest that there is moderate selective nonresponse that probably translates into a small amount of nonresponse bias in estimates of reports and ratings and hospital-level comparisons thereof in the absence of case-mix adjustment. The data do not support the use of late respondents to represent nonrespondents. The use of nonresponse weights would slightly reduce bias and improve the precision of CAHPS estimates for the total population and for subgroups for which there are more than 300 completed surveys. This threshold rises to 1,000 or more completed surveys when case-mix adjustment is used. For comparisons of hospitals with sample sizes near the recommended 300 completed surveys, nonresponse weights would likely slightly worsen precision even in the absence of case-mix adjustment. We therefore recommend that nonresponse weighting not be employed in the comparison of individual hospitals for the CAHPS Hospital Survey, but that it be considered for overall estimates or comparisons of large demographic or regional subgroups. The proportions of missing responses observed necessitate principled single imputation such as that provided by PROC MI in *SAS*, but do not require multiple imputation.

ACKNOWLEDGMENTS

The CAHPS II project is funded by the Agency for Health Care Quality and Research (AHQR) and the Center for Medicare and Medicaid Service (CMS)

through cooperative agreements with the RAND Corporation (5 U18 HS00924), AIR, and Harvard Medical School. User support is provided through a contract with Westat.

NOTES

1. It can be shown that nonresponse weights improve MSE only when the reduction standardized bias (ε above) exceeds the square root of $(\text{DEFF} - 1)/n$.
2. Twenty-nine percent after the mail wave.
3. If we remove the 989 sampled patients (2 percent of total), who subsequently were found to have died (750), become incapacitated (221), or become incarcerated (18), this response rate rises to 36 percent and all others reported do not change. These cases have not been excluded from analyses of unit nonresponse, as they represent members of the population of interest who became unavailable to the survey.
4. See Table 1 for 16 of these categories; the MDC related to obstetrics is not listed because it is identical to the obstetrics service indicator.
5. Q1–Q4, Q8, Q10, Q11–Q19, Q21, Q23, Q25–Q26, Q28, Q29, Q34, Q36, Q42–Q45, Q52, Q53, and Q56–Q65.
6. Q4–Q7, Q9, Q11–Q14, Q16–Q18, Q20, Q22, Q24, Q25, Q27, Q28, Q35, Q37–Q41, Q43, Q44, Q47–Q49, and Q51.
7. The analysis was restricted to hospitals with sufficient sample for reliable estimation of hospital-level means.
8. Leaving against medical advice is sufficiently rare (about 1 percent of sampled patients) that its impact on nonresponse weights is limited (results not shown).

REFERENCES

- Barkley, W. M., and D. H. Furse. 1996. "Changing Priorities for Improvement: The Impact of Low Response Rates in Patient Satisfaction." *Journal on Quality Improvement* 22: 427–33.
- Bootsma-van der Wiel, A., E. van Exel, A. J. M. de Craen, J. Gussekloo, A. M. Lagaay, D. L. Knook, and R. G. J. Westendorp. 2002. "A High Response Is Not Essential to Prevent Selection Bias; Results from the Leiden 85-Plus Study." *Journal of Clinical Epidemiology* 55: 1119–25.
- Burroughs, T. E., A. R. Davies, J. C. Cira, and W. C. Dunagan. 1999. "Understanding Patient Willingness to Recommend and Return: A Strategy for Prioritizing Improvement Opportunities." *Joint Commission Journal on Quality Improvement* 25: 271–87.
- Cohen, G., and J. C. Duffy. 2002. "Are Nonrespondents to Health Surveys Less Healthy Than Respondents?" *Journal of Official Statistics* 18 (1): 13–23.

- Colsher, P. L., and R. B. Wallace. 1989. "Data Quality and Age: Health and Psycho-Behavioral Correlates of Item Non-Response and Inconsistent Responses." *Journal of Gerontology: Psychological Sciences* 44: 45-52.
- de Vries, H., M. N. Elliott, K. A. Hepner, S. D. Keller, and R. D. Hays. 2005. "Equivalence of Mail and Telephone Responses to the Hospital CAHPS[®] (HCAHPS) Survey." *Health Services Research* DOI:10.1111/j.1475-6773.2005.00479.x. Available at www.blackwell-synergy.com
- Fowler, F. J., P. M. Gallagher, V. L. Stringfellow, A. M. Zaslavsky, J. W. Thompson, and P. D. Cleary. 2002. "Using Telephone Interviews to Reduce Non-Response Bias to Mail Surveys of Health Plan Members." *Medical Care* 40 (3): 190-200.
- Gasquet, I., B. Falissard, and P. Ravaud. 2001. "Impact of Reminders and Method of Questionnaire Distribution on Patient Response to Mail-Back Satisfaction Survey." *Journal of Clinical Epidemiology* 54: 1174-80.
- Goldstein, L., C. Crofton, S. Garfinkel, and C. Darby. 2005. "Why Another Patient Survey of Hospital Care?" *Health Services Research* DOI:10.1111/j.1475-6773.2005.00477.x. Available at www.blackwell-synergy.com
- Grotzinger, K. M., B. C. Stuart, and F. Ahern. 1994. "Assessment and Control of Non-Response Bias in a Survey of Medicine Use by the Elderly." *Medical Care* 32: 989-1003.
- Groves, R. M., and M. P. Couper. 1998. *Non-Response in Household Interview Surveys*. New York: John Wiley and Sons.
- Guadagnoli, E., and P. D. Cleary. 1992. "Age-Related Non-Response in Surveys of Recently Discharged Patients." *Journals of Gerontology: Psychological Sciences* 47 (3): 206-12.
- Hays, R. D., C. Larson, E. C. Nelson, and P. B. Batalden. 1991. "Hospital Quality Trends: A Short-Form Patient-Based Measure." *Medical Care* 29 (7): 661-8.
- Hoeymans, N., E. J. M. Feskens, G. A. M. van den Bos, and D. Kromhout. 1998. "Non-Response Bias in a Study of Cardiovascular Diseases, Functional Status and Self-Rated Health among Elderly Men." *Age and Ageing* 27: 35-40.
- Hoff, R. A., R. A. Rosenheck, M. Meterko, and N. J. Wilson. 1999. "Mental Illness as a Predictor of Satisfaction with Inpatient Care at Veterans Affairs Hospitals." *Psychiatric Services* 50: 680-5.
- Lamers, L. M. 1997. "Medical Consumption of Respondents and Nonrespondents to a Mailed Health Survey." *European Journal of Public Health* 7: 267-71.
- Little, R. J. A., and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: John Wiley and Sons.
- Little, R. J. A., and N. Schenker. 1995. "Missing Data." In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, edited by G. Arminger, C. C. Clogg, and M. E. Sobel, pp. 39-76. New York: Plenum Press.
- Mishra, S. I., D. Dooley, R. Catalano, and S. Serxner. 1993. "Telephone Health Surveys: Potential Bias from Noncompletion." *American Journal of Public Health* 83: 94-9.
- O'Malley, J., A. M. Zaslavsky, M. N. Elliott, L. Zaborski, and P. D. Cleary. 2005. "Case-Mix Adjustment of the Hospital Consumer Assessment of Health Plans Study

- (HCAHPS[®]) Survey.” *Health Services Research* DOI:10.1111/j.1475-6773.2005.00470.x. Available at www.blackwell-synergy.com
- Paganini-Hill, A., G. Hsu, A. Chao, and R. K. Ross. 1993. “Comparison of Early and Late Respondents to a Postal Health Survey Questionnaire.” *Epidemiology* 4: 375–9.
- Rubin, H. R. 1990. “Patient Evaluations of Hospital Care: A Review of the Literature.” *Medical Care* 28: S3–9.
- Sherbourne, C. D., and L. S. Meredith. 1992. “Quality of Self-Report Data: A Comparison of Older and Younger Chronically Ill Patients.” *Journal of Gerontology* 47 (4): S204–11.
- Ware, J. E., and D. M. Berwick. 1990. “Conclusions and Recommendations.” *Medical Care* 28: S39–44.
- Zaslavsky, A. M., L. B. Zaboriski, and P. D. Cleary. 2002. “Factors Affecting Response Rates to the Consumer Assessment of Health Plans Study Survey.” *Medical Care* 40 (6): 485–99.
- Zaslavsky, A. M., L. B. Zaboriski, L. Ding, J. A. Shaul, M. J. Cioffi, and P. D. Cleary. 2001. “Adjusting Performance Measures to Ensure Equitable Plan Comparisons.” *Health Care Financing Review* 22: 109–26.