Implications of Variation in the Relationships between Beneficiary Characteristics and Medicare Advantage CAHPS Measures

Laura A. Hatfield and Alan M. Zaslavsky

Objective. To study how differences in quality score adjustments across Medicare Advantage contracts change comparisons for individuals and contracts.

Data Sources. Responses to the Medicare Advantage implementation of the Consumer Assessments of Healthcare Providers and Systems (CAHPS) survey from 2010 to 2014.

Study Design. We fit national-and state-level hierarchical models to predict CAHPS scores for individuals and contracts, adjusted for self-reported education, general health, and mental health. We allow the effects of these variables on quality measures to vary across contracts with a hierarchical model.

Data Collection/Extraction Methods. We perform secondary data analysis.

Principal Findings. For average consumers, standard adjustment is sufficient to represent variation in contract quality standardized to a common population. For people with characteristics far from average, personalized reporting using their characteristics and contract-specific coefficients can substantially change the expected quality measures across contracts. This effect is stronger when comparing among contracts within a state than across all contracts.

Conclusions. Customized reporting may help consumers select the best Medicare Advantage plan, but policies should protect against unintended consequences.

Key Words. Case mix adjustment, quality measurement, Medicare Advantage, personalized reporting

Since 1997, Medicare Advantage (MA) enrollees have reported on their experiences with health care and plan services via the Consumer Assessments of Healthcare Providers and Systems (CAHPS[®]) survey (Goldstein et al. 2001). These scores influence policy and consumer choices in several ways. The Medicare Plan Finder website (<https://www.medicare.gov/find-a-plan/>

questions/home.aspx) reports CAHPS scores to inform beneficiaries choosing among MA plans (Reid et al. 2013). CAHPS scores contribute to the “star rating” system that determines quality incentive payments from Centers for Medicare & Medicaid Services to MA plans. These ratings also guide quality improvement efforts (Smith et al. 2001; Glazer et al. 2008; Price et al. 2014) and enable comparisons of MA to Traditional Medicare (Elliott et al. 2011; Ayanian et al. 2013).

Patient and consumer survey responses about experiences, quality, and satisfaction correlate with personal characteristics like health, education, and age, even among people with the same plan or provider (Zaslavsky et al. 2001; Simon et al. 2009; Rahmqvist and Bara 2010). These associations combine reporting tendencies with real differences in experiences. For example, sicker patients tend to report more problems with coordination, perhaps because they have more complex needs and more frequent contact with providers. More educated patients also report more problems, perhaps because they have higher expectations rather than because they receive lower-quality care (Sofaer and Firminger 2005).

A contract enrolling more people with characteristics (such as poor health) that predict lower scores will fare worse on survey-based quality measures than a contract whose members tend toward more favorable responses, even if it provides services of the same quality (Martino et al. 2011; Paddison et al. 2012). Case mix adjustment removes differences in contract quality measures predictable from personal characteristics, then estimates differences among plans as if they all enrolled similar populations. Standard CAHPS methodology does this by regression adjustment in a linear model (Zaslavsky et al. 2001; Kim, Zaslavsky, and Cleary 2005). This leads to valid comparisons if the effects of personal characteristics on quality measures are the same across contracts. However, there is considerable evidence that this is not the case for patient care experience measures in several settings (Zaslavsky, Zaboriski, and Cleary 2000b; Damman et al. 2009; Lyratzopoulos et al. 2012). For example, while sick patients tend to report more problems than healthy patients, the magnitude of the gap between sick and healthy varies across health plans (Zaslavsky and Cleary 2002), violating the assumptions of the usual CAHPS case mix models (Elliott et al. 2001).

Address correspondence to Laura A. Hatfield, Ph.D., Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115; e-mail: hatfield@hcp.med.harvard.edu. Alan M. Zaslavsky, Ph.D., is with the Department of Health Care Policy, Harvard Medical School, Boston, MA.

Even if this heterogeneity changes adjusted plan quality measures relatively little, it can affect predicted quality for individuals substantially, because characteristics of individuals vary more than plan averages. A single comparison averaged across all patients may be a poor prediction of some individuals' experiences; for example, a plan might perform relatively well on quality measures for care of healthy people but not for sick people. In some previous work, separating reports by patient types provided more additional information than separately reporting responses by measure topic (O'Malley et al. 2005).

To study how heterogeneity of these quality measure adjustment coefficients affects comparisons among contracts, we analyzed data from the MA implementation of the CAHPS survey from 2010 to 2014. Unlike previous studies that analyzed one predictor for one quality measure at a time, we jointly model variation across contracts in adjustment coefficients for multiple quality measures and identify the total effect on adjusted quality measures both for contracts and for predictions of quality experienced by different kinds of people.

METHODS

Survey Sample

Medicare provides health insurance to almost all elderly (≥ 65 years old) legal residents of the United States. During our study period (2010–2014), about 30 percent of beneficiaries enrolled in MA, a program in which private health insurers provide benefits through contracts with Medicare. Each contract covers a specific geographic area, typically a whole state, a collection of counties in a state, or a combination of such areas across a few adjacent states. The contract is our unit of analysis, comprising all survey responses from enrollees of an MA contract in a given year.¹

Annual samples of 800 respondents were drawn from most contracts, fewer if eligible enrollment was low and more if the contract requested a larger sample and paid the additional survey costs. Response rates averaged 49 percent, for a total of 1,025,666 valid responses. Further details of survey methods, nonresponse patterns, measure correlations, and organizational and geographical variation in scores appear elsewhere (Goldstein et al. 2001; Zaslavsky, Zaborski, and Cleary 2004; Elliott et al. 2011; Farley et al. 2011; Dembosky et al. 2013; Hays et al. 2014). We excluded 129,450 responses from beneficiaries younger than 65 years; disability qualifies this younger

group for Medicare, and the relationships of needs to health status are likely different for them. We also excluded 66,126 responses with missing covariates values and another 11,194 responses from very small samples (<150 respondents in a contract-year), leaving 818,896 individual responses in 2,216 contract-years across 592 contracts. For our model with contracts clustered within states, we identified contract-state units with ≥ 150 responses. We excluded 26 contracts that spanned multiple states in a year and 10 that were the only contract in a given state and year. This left 2,176 contract-states in 218 state-years (median: 8 contracts per state-year; interquartile range: 5–12).

Variables

During our study period, the MA-CAHPS survey included numerous items about experiences with care and about getting prescription drugs (for those with that benefit), described in full on the survey website (Centers for Medicare & Medicaid Services, 2015). We analyze 11 quality measures, a combination of single survey items and composites that sum responses from several survey items. The left side of Table 1 summarizes the quality measures, their constituent survey items, and numerical ranges. These measures are used for consumer reporting and star summaries, with a few differences described in Appendix SA2.

Previous analyses (Zaslavsky et al. 2001; Martino et al. 2011) identified three personal characteristics strongly associated with survey responses and thus influential for adjustment: self-reported general and mental health status (poor, fair, good, very good, or excellent coded as 1–5) and self-reported education (no high school to graduate education coded as 1–6). Therefore, we consider models that adjust for these three covariates.

Hierarchical Modeling

Our approach approximates a set of mixed models, one for each quality measure in each contract with random coefficients for contracts and a multivariate second-level distribution governing the relationships among these coefficients.

Letting q index quality measure, h index contract, and k index respondents within contract, the mixed model for each measure is

$$E(y_{hkq}) = \delta_q \mathbf{z}_{hk} + \boldsymbol{\theta}_{hq} \mathbf{x}_{hk}$$

Table 1: Summary of the Quality Measures Analyzed, Their Constituent CAHPS Survey Items, Ranges, and Global Adjustment Coefficients

<i>Quality Measure</i>	<i>Range</i>	<i>Global Adjustment Coefficients</i>		
		<i>General Health</i>	<i>Mental Health</i>	<i>Education</i>
MD communication	4–16	0.79	1.20	−0.09
Explains	1–4			
Listens	1–4			
Respects	1–4			
Spends time	1–4			
Rate care	0–10	0.90	0.89	−0.18
Coordination of care	6–23	0.32	0.56	−0.15
Help managing care	1–3			
MD recommends	1–4			
Talk about meds	1–4			
Quick test results	1–4			
Test follow-up	1–4			
Specialist coordination	1–4			
Getting care quickly	2–8	0.29	0.36	0.05
Illness care	1–4			
Routine care	1–4			
Getting needed care	2–8	0.52	0.71	−0.20
Get appointment	1–4			
Get care	1–4			
Plan customer service	3–12	0.34	0.45	−0.20
Get info	1–4			
Courteous	1–4			
Easy forms	1–4			
Rate plan	0–10	0.47	0.47	−0.35
Rate drug plan	0–10	0.45	0.40	−0.43
Getting drug info	4–16	0.33	0.38	−0.05
Drug info	1–4			
Courteous	1–4			
Drug cost info	1–4			
Coverage into	1–4			
Getting needed drugs	2–8	0.35	0.62	−0.13
Easy to fill	1–4			
Easy by mail	1–4			
Flu shot	0–1	−0.39	−0.01	0.14

Note. Coefficients are scaled to represent the effect of a unit change in the predictor relative to (divided by) the standard deviation of the contract intercepts (adjusted contract means) of each quality measure.

where the parameters δ and θ are vectors of fixed and random coefficients of the respective covariate vectors $\mathbf{z}$ and $\mathbf{x}$, and the variables in $\mathbf{z}$ span the variables in $\mathbf{x}$.² The concatenated vectors of random coefficients for each contract,

$\theta_h = (\theta_{hq} \mid q = 1, \dots, Q)$, are related through a level-2 distribution,

$$\theta_h \sim N(\alpha, \Sigma) \quad (1)$$

where α is a vector of level-2 means and Σ is the variance-covariance matrix of contract random coefficients. To estimate this model, we use a three-step process.

First, we estimate the global coefficients (i.e., fixed effects) using the model $E(y_{hkg}) = \delta_g z_{hk}$. That is, for each quality measure,³ we fit a single model with global adjustment for the covariates. This produces an adjusted mean for each contract as if it had enrolled a population with the same mean characteristics (general and mental health status and education) as the national sample.

Next, we estimate the coefficients for each quality measure and contract under the model $E(y_{hkg}) - \delta_g z_{hk} = \theta_{hg} x_{hg}$ fitted to the residuals from the previous step. The fitted $\widehat{\theta}_h$ are estimates of the contract-level random effects, that is, the deviations of each contract's coefficients from the average adjustment model. We also compute robust estimates $\widehat{V}_h$ of the sampling variance-covariance matrix of θ_h , extending standard survey sampling methods for regression coefficients (Binder and Pataak 1994) to our stacked regression models.

Finally, we estimate the parameters of hierarchical model (1) fitted to the estimated coefficient vectors from each contract $\widehat{\theta}_h$ and their estimated covariance matrices $\widehat{V}_h$ as in a multivariate version of Fay and Herriot (1979),

$$\text{Sampling model : } \widehat{\theta}_h \sim N(\theta_h, \widehat{V}_h) \quad (2)$$

$$\text{Population model : } \theta_h \sim N(\alpha, \Sigma)$$

The contract coefficients θ_h and their covariance matrix Σ are the primary parameters of interest.

In a secondary analysis, we expand model (2) to a three-level model that nests contracts within state-years to assess variation among contracts that are potentially competing. In this analysis, the population model gains an additional level,

$$\begin{aligned} \theta_h &\sim N(\gamma_{g(h)}, \Sigma_\theta) \\ \gamma_g &\sim N(\alpha, \Sigma_\gamma) \end{aligned} \quad (3)$$

where $g(h)$ is the state-year in which contract h operates and γ_g is a state-year mean vector of the same length as θ_h . This partitions the variation into a component for state and a component for contract within state. The relative magnitudes of Σ_θ and Σ_γ indicate how much of the variation among contracts is a

consequence of variation within states. We fit both models by maximum likelihood using an EM algorithm. For further details of the computational methods, see Appendix SA2.

To enable meaningful comparison across measures, we report all our results on the scale of across-contract standard deviations (SDs) of the adjusted contract means (i.e., the SD of the contract-level random intercepts). Adjusted means are the basis for reported comparisons among contracts, so the potential effect of a variable on such comparisons depends on this relative magnitude. After rescaling, a 1-unit difference in a coefficient implies a 1 SD shift in the quality experience of a person with health or education 1 unit from average. A difference of ± 1 SD can move a contract from median performance to the 16th or 84th percentile.

Using coefficients and covariance matrices estimated under these models, we describe variation in quality and adjustment coefficients among contracts and how this affects comparisons among contracts for individual beneficiaries at the national and state level. We present our results using a range of summary statistics and distributional displays, the details of which are explained in the corresponding results sections below.

RESULTS

Global Adjustments and Contract Composition

Table 1 gives the estimated global coefficients, which are broadly consistent with previous analyses: healthier and less educated people report higher quality on surveys (Zaslavsky et al. 2001). For most measures, the health coefficients are substantially larger than those for education. Education coefficients for the plan rating item, however, are as large as health coefficients, though in the opposite direction. The rate care and the MD communication measures are most strongly associated with general and mental health. Unlike other measures, flu shot is more likely to be reported by less healthy patients, reflecting its clinical importance to such patients, but is unrelated to mental health.

Effects of Across-Contract Coefficient Variation

Comparing Across-Contract Variation in Adjustment Coefficients to Global Coefficients. If global coefficients dominate across-contract variation, the adjustments will not be affected much by contract-level variation in coefficients. Figure 1 compares the global coefficients to across-contract variation.

Figure 1: Global Coefficients (Dots) and Across-Contract Coefficient Standard Deviations (SDs) (Segments), Scaled to Show Impact of One-Unit Covariate Change on Each Quality Measure

Notes: Segments extend ± 2 SDs of across-contract coefficient variation, with tick marks at ± 1 SD. Tick marks plotted in black signify that the ± 1 -SD line crosses zero relative to the global coefficient.

[Correction added on 7 June 2017, after first online publication: the “Mental health” column of Figure 1 was incomplete and has now been amended.]

Each dot represents the global coefficient, that is, the effect on each quality measure of a 1-unit change in the covariate. The line segments extend ± 2 SDs of across-contract coefficient variation in either direction, with a vertical tick mark at ± 1 SD of across-contract coefficient variation. These intervals represent the distribution of contract coefficients around the global coefficient.

When the ± 1 -SD interval crosses 0, the probability that a contract-level coefficient will reverse the sign of the global coefficient exceeds ≥ 16 percent; we highlight these cases with black tick marks. For example, the global effect of education on the plan customer service measure is -0.20 (recall Table 1). The SD of the contract-level coefficients is 0.25 so the ± 1 SD interval around the global coefficient is $[-0.45, 0.05]$. This implies that for a nontrivial proportion of contracts, the direction of the relationship between education and the plan customer service measure is reversed; that is, people with more education rate plan customer service *higher* than those with less education.

The ± 1 -SD line crosses the axis for eight education coefficients, two general health coefficients, and two mental health coefficients. Two customer

service measures—plan customer service and getting drug info—have modest global coefficients of general and mental health, but large variation in contract coefficients. This is consistent with the high degree of control that contracts have over provision of these services, which they provide directly. By contrast, the large global coefficients for health on MD communication, rate care, flu shot, and so on dominate the variation across contract coefficients, indicating that these effects are consistent across contracts. This is particularly sensible in the case of flu shot, given its clinical rationale. (For administrative reasons, the flu shot item is not currently case mix adjusted.)

CAHPS Quality Measure Adjustment. Next we consider the joint and separate contributions of global- and contract-level coefficients to adjusted quality measures for contracts. Figure C.1 in Appendix SA2 displays the distribution of quality measure adjustments using global adjustment only, contract-level adjustment only, and both. The variation comes both from the differences in mean plan characteristics and (except for the global adjustment only) from the across-contract coefficient variation. Using global coefficients only, the standard deviation of adjustments across contracts ranged from 6 percent (of across-contract SD of means) for the flu shot measure to 42 percent for the MD communication and rate care measures. The contract-specific coefficients produce a narrower distribution of incremental adjustments, ranging from 8 percent (flu shot) to 27 percent (getting drug info) of across-contract SD of means. Only the getting drug info and plan customer service measures had more variation in adjustments due to contract-level coefficients than global. Combining contract-level and global coefficients, we get adjustment variation of similar magnitude to that using global coefficients only, from 11 percent to 42 percent of across-contract mean SD with the same pattern across measures. Thus, adding contract-specific coefficients has a modest impact on the overall adjustment to contract scores for most measures.

Individual Covariate Variation and Effects of Coefficient Variation. Individual characteristics vary much more than contract averages for their enrolled populations. Individual values of the education, general, and mental health adjustment variables have standard deviations 1.4, 1.0, and 1.0 respectively. These are several times larger than the corresponding SDs of contract means, which are 0.53, 0.26, and 0.25. However, individual-level correlations among covariates (+0.25 between education and health, +0.54 between mental and general

health) are weaker than correlations among contract means (between +0.76 and +0.89).

Consider then how individuals' characteristics affect comparisons between two contracts for an individual seeking to choose the one where her predicted experiences would be better. Under the global coefficient model, her predicted difference equals the difference between adjusted contract means and does not depend on her characteristics. Under our mixed model, the difference additionally depends on the contract-specific coefficients and individual characteristics.

Contract coefficients matter more for individuals with characteristics further from average. There are $5 \times 5 \times 6 = 150$ possible individual covariate combinations and thus the same number of unique predicted quality adjustments. For every covariate combination $\mathbf{x}$, we compute the across-contract variance of the adjustments to quality measures that are omitted from standard adjustment, i.e., $\text{Var}(\theta_{hq}\mathbf{x}) = \mathbf{x}'\Sigma_q\mathbf{x}$, where Σ_q is the 3×3 sub-block of Σ that corresponds to the across-contract variance-covariance matrix for the education, general health, and mental health coefficients for outcome q . We call this the variance of prediction error and plot quantiles of its distribution in Figure 2, weighted by each covariate combination's proportion in the whole population. For most measures, an average person's SD of prediction error will be 48–115 percent of the across-contract SD of the mean, while the most extreme covariate combinations have SD of prediction error of 87–243 percent of across-contract SD of the mean. Two measures—plan customer service and getting drug info—show much larger effects, especially for individuals with extreme covariate combinations. A few covariate combinations have consistently higher prediction error variance, namely, those with the poorest mental health and good physical health (data not shown).

To understand the impact on contract choice of allowing coefficients to vary across contracts, we compute the probability that a high-quality contract under standard adjustment also has high predicted quality for an individual with particular covariate values. The conditional probability that a contract's predicted quality for an individual, adjusted using contract-level coefficients, is in the top 20 percent, given that it is in the top 20 percent of contracts under the standard global adjustment, depends on the individual's characteristics. Figure 3 shows the distribution of these probabilities for the 150 covariate combinations, weighted by each combination's proportion in the population. Large values indicate that personal characteristics and contract coefficients have little effect on the contract rankings, while small values indicate that some top-performing contracts are not serving certain consumers well. The

Figure 2: Distribution of Across-Contract SD of Prediction Error for Individuals with Different Covariate Combinations

Notes: Each bar shows quantiles (across all covariate combinations, weighted by their proportions in the population) of the across-contract SD of prediction error, divided by the across-contract SD of the outcome means.

getting drug information measure provides the most striking illustration of how contract-level coefficients impact rankings: for half the population, the probability is less than 60 percent that a contract ranked in the top 20 percent globally will also be ranked in the top 20 percent when personal characteristics and contract-specific coefficients are considered.

State-Level Model Results

Each beneficiary chooses among contracts serving her own area, not the entire national selection of contracts. To see how this restriction changes the effects of beneficiary characteristics on adjusted comparisons among contracts, we fit a three-level model (3) that partitions variation among contracts into within-

Figure 3: Distribution of the Conditional Probability That a Top 20 Percent Contract for the Average Person Is Also a Top 20 Percent Contract for a Given Individual

Notes: Boxes cover the 25th to 75th quantiles with a horizontal bar at the median; whiskers extend to the extremes.

state and between-state components. We examined the ratio of within-state to total variation for the adjusted means and the three adjustment coefficients, extending previous research that examined this ratio only for contract means (Zaslavsky, Zaborski, and Cleary 2004). More than half of the variation is at the within-state level for every measure-predictor pair; the ratio of within-state to total variance ranges from 0.59 (education coefficient for flu shots) to 0.84 (health coefficient for plan customer service). For the adjusted means, the within-state fraction of variation ranges from 0.40 (coordination of care) to 0.75 (plan customer service). In 29 out of 33 measure-predictor pairs, the within-state fraction is larger for the coefficients than for the mean.

Extending work on variation in contract means at the state and national level (Zaslavsky, Zaborski, and Cleary 2004), we examine the implications of

coefficient variation on comparisons among contracts at these two levels. Appendix SA2 Figure C.2 compares the ratios of variation (across-contract SD of coefficients divided by across-contract SD of means) at the state and national levels for each quality measure and coefficient type. For most measures, the ratio is larger in the within-state level of the model, suggesting that the impact of coefficient variation is larger within states than in the national model, but the increase is uniform across measures and predictors (i.e., the ranks are mostly preserved). A notable exception is the coordination of care measure, for which the coefficient-to-mean variation ratio is among the smallest in the two-level model, but second or third largest in the within-state component of the three-level model. Thus, comparing expected coordination of care scores across contracts is markedly more affected by coefficient heterogeneity when considering contracts in the same state versus all contracts.

Reliability and Temporal Stability of Coefficient Estimates

Our methods focus on adjustment coefficient variation across contracts, which depends on our ability to estimate contract-specific coefficients with adequate precision. To quantify this precision, we compute the intercontract reliability: the ratio of modeled population variance to the sum of population and sampling variances (total variance). Reliability assumes a distinct value for each combination of contract, predictor (or intercept), and measure. Appendix SA2 contains further computational details and definitions of reliability and Appendix SA2 Figure C.3 shows the distribution of intercontract reliability for each coefficient and measure. Adjusted contract means (random intercepts) are the most reliably estimated, with population variation comprising 77–92 percent of total variation for the nine publicly reported measures (median across contracts). This reflects success in achieving the major design objective of the survey, namely to estimate adjusted contract means with enough precision to distinguish relative quality across contracts. Reliability is smaller for the adjustment coefficients, indicating less reliable contract-level estimates. Education coefficients are generally more reliable than health coefficients, with a few contracts achieving 90 percent reliability but median reliabilities ≥ 30 percent only for flu shot, rate plan, and rate drug plan. Median reliabilities for the health coefficients are generally below 25 percent.

Over our 5-year observation period, the contract intercepts are highly correlated (Appendix SA2 Figure C.4). For most adjusted means (i.e., random intercepts), the autocorrelation is greater than 0.9 at lags of 1 and 2 years. Many of the contract-level adjustment coefficients also have fairly high

autocorrelations, with median lag-1 and lag-2 autocorrelations (over measures) between 0.5 and 0.75 for education and general health status coefficients. Indeed, the estimated correlation of the coefficient in the fifth year of our data with the mean of the coefficients in the preceding 4 years is 0.6 for education and 0.7 for general health status, suggesting a fair degree of temporal stability in the contract-specific coefficients. This suggests that using results from previous years as predictors may improve coefficient estimates. Using the mean of estimates from the previous 4 years in this way more than doubled median reliability for six of the general health status coefficients and made substantial improvements in many others.

DISCUSSION

Conventional case mix adjustment for health care experience surveys assumes that patient characteristics affect responses the same way in every contract. Using contract-specific coefficients in a multiple-outcome model, we demonstrated how coefficient variation affects quality comparisons among MA health contracts. Contract-level coefficients change the overall adjustments modestly, contributing much less variation than fixed coefficients, which themselves represent a small fraction of across-contract variation. Thus, fixed coefficients are sufficient for generating adjusted mean quality measures for contracts.

Coefficient variation matters more for predictions for individuals, whose characteristics are more variable than contract means are. Standard CAHPS quality reports present means and comparisons averaged over a typical distribution of characteristics, but comparisons can differ for beneficiaries at the extremes of the distribution. The magnitude of this interaction between contract coefficients and beneficiary characteristics varies across measures. A person with characteristics modestly different from average who is comparing expected quality on getting drug information and plan customer service measures should expect predictions of their personal experiences to differ from those for the plan average by 50 percent or more of the typical difference among contract means. For other measures, particularly flu shot, the differences in relative quality are much smaller.

Studying differences across measures can yield insight into possible mechanisms. Coefficients that vary least across contracts are likely driven by mechanisms that apply to all beneficiaries and are not influenced much by the

plans. In particular, the strong and stable association between poor general health and the flu shot measure likely reflects clinician adherence to recommendations to immunize vulnerable patients.

Similarly, worse mental health and general health are generally associated with less favorable reports, reflecting the greater potential for unmet needs among beneficiaries in poor health as well as the negative effects of poor mental health on patient efficacy. Although health status and patient-perceived experiences with care might mutually influence each other, the endogenous effects of health care on overall health are likely to be small over the 6-month period of the survey. In contrast, the effects of health on experiences are immediate, especially for those in poor health who frequently seek services.

Conversely, the variability of coefficients for plan customer service and getting drug information suggests that these services provided directly by the plan vary more across contracts, affecting beneficiaries differentially by health status and education. For example, contracts may emphasize providing information on paper, on the Internet, or by telephone; offer written materials of varying literacy levels; train service operators differently; and offer more or less accessible information for patients with chronic conditions. Enrollees with more education rate their care and health plans less favorably, likely because they have higher expectations. However, the education coefficients for getting needed care, coordination of care, and plan customer service measures are less consistent and more likely to reverse sign in some contracts. This may be because barriers to care for less educated beneficiaries vary so much depending on availability of information and ease of access.

Our adjustment models could be extended to include observable plan features that might explain some of the variation in coefficients, that is, modeling varying coefficients as functions of plan characteristics. But readily available plan characteristics like size, network model, and tax status do not have evident mechanistic (as opposed to merely empirical) relationships with quality, while only very limited structural measures of organizational functioning are available. Furthermore, we believe that on (correct) policy grounds, CMS is unlikely to adjust contract scores for structural characteristics. Nonetheless, future studies should consider how plan characteristics relate to quality adjustment coefficients and overall quality.

Our findings have potential practical implications. Models like ours might be used to personalize information for decision making, as has been suggested for hospital choice (Elliott et al. 2010). A beneficiary comparing contracts on the Medicare Compare website might anonymously submit

information on her health and education and receive a customized report of predicted scores in locally available contracts. This could be particularly useful in contract choice because coefficient variation affects within-state comparisons among potentially competing contracts more than at the national level. However, personalized reporting complicates the already cognitively demanding task of plan choice (McWilliams et al. 2011); better quality information will only improve consumer experiences if individuals can understand and use the information (Martino et al. 2013).

For health plans, tailored quality reports may aid quality improvement efforts. For example, contract administrators could learn that their less educated enrollees have more difficulty understanding plan written materials or that those in poor health cannot get timely appointments. On the other hand, quality reports that depend on beneficiary characteristics might facilitate selection. General health status reported in CAHPS predicts medical spending (Zaslavsky and Buntin 2002), even after applying current risk adjustment models for plan reimbursement (Rose, Zaslavsky, and McWilliams 2016). Thus, contracts might design their services to selectively reduce adjusted measures for enrollees whose characteristics predict high residual costs (Glazer et al. 2008), exploiting the customized reports as a mechanism to transmit a signal that discourages their future enrollment. This selection could be discouraged by including the patient characteristics used for personalized reporting in the risk adjustment model.

Finally, for Medicare administrators, our contract- and subgroup-specific estimates allow monitoring of quality experiences of vulnerable groups such as those with less education or poor mental health. This might identify contracts with particularly poor predicted scores for these groups (for oversight and action) or particularly good scores (to identify exemplary practices and allow steering). Or it might reveal markets that lack high-quality contracts for specific beneficiary types, prompting adjustment of incentives. Adverse selection engendered by differential quality, possibly facilitated by enhanced reporting that publicizes such quality differences, could also prompt action to preserve competition. Otherwise contracts that provide especially good care for sicker and more expensive patients could suffer a “death spiral” of financial stress and adverse selection.

This research has several limitations. First, we restricted our analysis to three adjustment covariates, although these are the most important predictors among those available on the survey. Second, survey nonresponse was substantial and was associated with relevant personal characteristics (Klein et al. 2011), although its differential effect on coefficients across contracts is

unknown. Finally, the practical application of the analysis is limited by the poor reliability of estimated contract-specific coefficients for some of the measures, although this might be improved through survey redesign and estimation methods using data from multiple years.

In summary, variation in the adjustment coefficients has modest effects on adjustment of contract scores but substantial effects on comparisons for individuals or subgroups defined by levels of health and education. Considering these effects can lead to more relevant quality reporting and monitoring.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: LAH and AMZ receiving funding for this work from AHRQ through a grant to Yale University (U18HS016978-05) and CMS through a contract with RAND (9920120015). The funding agreements ensured the independence of the work.

Disclosures: No other disclosures.

Disclaimers: None.

NOTES

1. We refer to our units of analysis as “contracts,” but when referring to the organization or quoting quality measures that ask respondents to report on their health or prescription drug “plans,” we retain that term.
2. In our application, we include columns for general health, mental health, and education in both z and x , but only x contains an intercept column. Thus, the contract-level intercepts are on the original scale, rather than re-centered around zero.
3. In this exposition, we do not distinguish between single-item and composite measures. For further detail, see Appendix SA2.

REFERENCES

- Ayanian, J. Z., B. E. Landon, A. M. Zaslavsky, R. C. Saunders, L. G. Pawlson, and J. P. Newhouse. 2013. “Medicare Beneficiaries More Likely to Receive Appropriate Ambulatory Services in HMOs Than in Traditional Medicare.” *Health Affairs (Millwood)* 32 (7): 1228–35.
- Binder, D. A., and Z. Pataki. 1994. “Use of Estimating Functions for Estimation from Complex Surveys.” *Journal of the American Statistical Association* 89 (427): 1035–43.

- Centers for Medicare & Medicaid Services. 2015. Baltimore, MD [accessed on January 26, 2015]. Available at www.ma-pdpcahps.org
- Damman, O. C., J. H. Stubbe, M. Hendriks, O. A. Arah, P. Spreeuwenberg, D. M. Delnoij, and P. P. Groenewegen. 2009. "Using Multilevel Modeling to Assess Case-Mix Adjusters in Consumer Experience Surveys in Health Care." *Medical Care* 47 (4): 496–503.
- Dembosky, J. W., A. M. Haviland, M. N. Elliott, P. Kallaur, C. A. Edwards, E. Sekscen-ski, A. M. Zaslavsky, and J. A. Brown. 2013. "Does Naming the Focal Plan in a CAHPS Survey of Health Care Quality Affect Response Rates and Beneficiary Evaluations?" *Public Opinion Quarterly* 77 (2): 455–73.
- Elliott, M. N., R. Swartz, J. Adams, K. L. Spritzer, and R. D. Hays. 2001. "Case-Mix Adjustment of the National CAHPS Benchmarking Data 1.0: A Violation of Model Assumptions?" *Health Services Research* 36 (3): 555–73.
- Elliott, M. N., W. G. Lehrman, E. Goldstein, K. Hambarsoomian, M. K. Beckett, and L. A. Giordano. 2010. "Do Hospitals Rank Differently on HCAHPS for Different Patient Subgroups?" *Medical Care Research and Review* 67 (1): 56–73.
- Elliott, M. N., A. M. Haviland, N. Orr, K. Hambarsoomian, and P. D. Cleary. 2011. "How Do the Experiences of Medicare Beneficiary Subgroups Differ between Managed Care and Original Medicare?" *Health Services Research* 46 (4): 1039–58.
- Farley, D. O., M. N. Elliott, A. M. Haviland, M. E. Slaughter, and A. Heller. 2011. "Understanding Variations in Medicare Consumer Assessment of Health Care Providers and Systems Scores: California as an Example." *Health Services Research* 46 (5): 1646–62.
- Fay, R. E., and R. A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74 (366): 269–77.
- Glazer, J., T. G. McGuire, Z. Cao, and A. Zaslavsky. 2008. "Using Global Ratings of Health Plans to Improve the Quality of Health Care." *Journal of Health Economics* 27 (5): 1182–95.
- Goldstein, E., P. D. Cleary, K. M. Langwell, A. M. Zaslavsky, and A. Heller. 2001. "Medicare Managed Care CAHPS: A Tool for Performance Improvement." *Health Care Financing Review* 22 (3): 101–7.
- Hays, R. D., S. Martino, J. A. Brown, M. Cui, P. Cleary, S. Gaillet, and M. Elliott. 2014. "Evaluation of a Care Coordination Measure for the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Medicare Survey." *Medical Care Research and Review* 71 (2): 192–202.
- Kim, M., A. M. Zaslavsky, and P. D. Cleary. 2005. "Adjusting Pediatric Consumer Assessment of Health Plans Study (CAHPS) Scores to Ensure Fair Comparison of Health Plan Performances." *Medical Care* 43 (1): 44–52.
- Klein, D. J., M. N. Elliott, A. M. Haviland, D. Saliba, Q. Burkhart, C. Edwards, and A. M. Zaslavsky. 2011. "Understanding Nonresponse to the 2007 Medicare CAHPS Survey." *Gerontologist* 51 (6): 843–55.
- Lyratzopoulos, G., M. N. Elliott, H. Barbiere, L. Staetsky, C. Paddison, A. Henderson, J. Campbell, and M. Roland. 2012. "Understanding Ethnic and Other Socio-Demographic Differences in Patient Experience of Primary Care: Evidence

- from the English General Practice Patient Survey.” *British Medical Journal Quality & Safety* 21 (1): 21–9.
- Martino, S. C., M. N. Elliott, D. E. Kanouse, D. O. Farley, Q. Burkhart, and R. D. Hays. 2011. “Depression and the Health Care Experiences of Medicare Beneficiaries.” *Health Services Research* 46 (6pt1): 1883–904.
- Martino, S. C., R. M. Weinick, D. E. Kanouse, J. A. Brown, A. M. Haviland, E. Goldstein, J. L. Adams, K. Hambarsoomian, D. J. Klein, and M. N. Elliott. 2013. “Reporting CAHPS and HEDIS Data by Race/Ethnicity for Medicare Beneficiaries.” *Health Services Research* 48 (2 Pt 1): 417–34.
- McWilliams, J. M., C. C. Afendulis, T. G. McGuire, and B. E. Landon. 2011. “Complex Medicare Advantage Choices May Overwhelm Seniors—Especially Those with Impaired Decision Making.” *Health Affairs* 30 (9): 1786–94.
- O’Malley, A. J., A. M. Zaslavsky, R. D. Hays, K. A. Hepner, S. Keller, and P. D. Cleary. 2005. “Exploratory Factor Analyses of the CAHPS Hospital Pilot Survey Responses across and within Medical, Surgical and Obstetric Services.” *Health Services Research* 40 (6(II)): 2078–95.
- Paddison, C., M. Elliott, R. Parker, L. Staetsky, G. Lyratzopoulos, J. L. Campbell, and M. Roland. 2012. “Should Measures of Patient Experience in Primary Care Be Adjusted for Case Mix? Evidence from the English General Practice Patient Survey.” *British Medical Journal Quality and Safety* 21 (8): 634–40.
- Price, A., M. N. Elliott, A. M. Zaslavsky, R. D. Hays, W. G. Lehrman, L. Rybowski, S. Edgman-Levitan, and P. D. Cleary. 2014. “Examining the Role of Patient Experience Surveys in Measuring Health Care Quality.” *Medical Care Research and Review* 71 (5): 522–54.
- Rahmqvist, M., and A. C. Bara. 2010. “Patient Characteristics and Quality Dimensions Related to Patient Satisfaction.” *International Journal for Quality in Health Care* 22 (2): 86–92.
- Reid, R. O., P. Deb, B. L. Howell, and W. H. Shrank. 2013. “Association between Medicare Advantage Plan Star Ratings and Enrollment.” *Journal of the American Medical Association* 309 (3): 267–74.
- Rose, S., A. M. Zaslavsky, and J. M. McWilliams. 2016. Variation in Accountable Care Organization Spending and Sensitivity to Risk Adjustment: Implications for Benchmarking. *Health Affairs* 35 (3): 440–8.
- Simon, G., C. Rutter, M. Crosier, J. Scott, B. H. Operskalski, and E. Ludman. 2009. “Are Comparisons of Consumer Satisfaction with Providers Biased by Nonresponse or Case-Mix Differences?” *Psychiatric Services* 60 (1): 67–73.
- Smith, F., M. Gerteis, N. Downey, J. Lewy, and S. Edgman-Levitan. 2001. “The Effects of Disseminating Performance Data to Health Plans: Results of Qualitative Research with the Medicare Managed Care Plans.” *Health Services Research* 36 (3): 643–63.
- Sofaer, S., and K. Firminger. 2005. “Patient Perceptions of the Quality of Health Services.” *Annual Review of Public Health* 26: 513–59.
- Zaslavsky, A. M., and M. J. Buntin. 2002. “Using survey Measures to Assess Risk Selection among Medicare Managed Care Plans.” *Inquiry* 39 (2): 138–51.

- Zaslavsky, A. M., and P. D. Cleary. 2002. "Dimensions of Plan Performance for Sick and Healthy Members on the Consumer Assessments of Health Plans Study 2.0 Survey." *Medical Care* 40 (10): 951–64.
- Zaslavsky, A. M., L. Zaborski, and P. D. Cleary. 2000b. "Does the Effect of Respondent Characteristics on Consumer Assessments Vary across Health Plans?" *Medical Care Research and Review* 57 (3): 379–94.
- Zaslavsky, A. M., L. B. Zaborski, and P. D. Cleary. 2004. "Plan, Geographical, and Temporal Variation of Consumer Assessments of Ambulatory Health Care." *Health Services Research* 39 (5): 1467–85.
- Zaslavsky, A., L. Zaborski, 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 (3): 109–26.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: Computational Details and Additional Figures.