

# Adjusting for Subgroup Differences in Extreme Response Tendency in Ratings of Health Care: Impact on Disparity Estimates

*Marc N. Elliott, Amelia M. Haviland, David E. Kanouse, Katrin Hambarsoomian, and Ron D. Hays*

---

**Objective.** Adjust for subgroup differences in extreme response tendency (ERT) in ratings of health care, which otherwise obscure disparities in patient experience.

**Data Source.** 117,102 respondents to the 2004 Consumer Assessment of Healthcare Providers and Systems (CAHPS) Medicare Fee-for-Service survey.

**Study Design.** Multinomial logistic regression is used to model respondents' use of extremes of the 0–10 CAHPS rating scales as a function of education. A new two-stage model adjusts for both standard case-mix effects and ERT. Ratings of subgroups are compared after these adjustments.

**Principal Findings.** Medicare beneficiaries with greater educational attainment are less likely to use both extremes of the 0–10 rating scale than those with less attainment. Adjustments from the two-stage model may differ substantially from standard adjustments and resolve or attenuate several counterintuitive findings in subgroup comparisons.

**Conclusions.** Addressing ERT may be important when estimating disparities or comparing providers if patient populations differ markedly in educational attainment. Failures to do so may result in misdirected resources for reducing disparities and inaccurate assessment of some providers. Depending upon the application, ERT may be addressed by the two-stage approach developed here or through specified categorical or stratified reporting.

**Key Words.** Health disparities, education, vulnerable populations, response bias

---

Reducing disparities in health and health care by race/ethnicity and socioeconomic status (SES) is a significant health policy goal; accurate measurement of these disparities is a critical first step (Institute of Medicine 2002; National Research Council 2004). Without accurate measurement, policy makers will not be able to target the patients or providers in greatest need of intervention nor to assess the effectiveness of such interventions.

Consumer evaluations of health care, such as those from the Consumer Assessments of Healthcare Providers and Systems (CAHPS<sup>®</sup>) project sponsored by the Agency for Healthcare Research and Quality (AHRQ) and the Centers for Medicare and Medicaid Services (CMS), are a vital source of information for understanding disparities in health and health care by race/ethnicity, SES, disability or other characteristics defining vulnerable subgroups (Weech-Maldonado et al. 2001, 2003, 2004, 2008b; Bernard, Brody, and West 2004b; Onstad 2005). Dimensions such as the courtesy and respect with which patients are treated and the clarity of communication are best assessed through consumer reports about their care. Disparities in the CAHPS composites (but not the global ratings) are targeted by Healthy People 2010 (Office of Disease Prevention and Health Promotion 2008), both as a means of reducing disparities in health outcomes and because there is an inherent interest in guaranteeing patient access to health care that is respectful and allows them to participate in treatment and care decisions. The World Health Organization (WHO) notes that there are “nonfinancial dimensions of quality of care that are important because they reflect respect for human dignity” (de Silva and Valentine 2000), a sentiment echoed by U.S. policy makers (Frist 2005).

Unfortunately, the contribution of patient experience surveys to understanding health disparities has been limited by measurement difficulties that arise from systematic differences in how patients respond to surveys. In particular, the 0–10 CAHPS global ratings, which have potential as summative measures for evaluating national and health-plan initiatives (e.g., National Health Plan Collaborative 2006) to reduce racial and economic disparities in health and health care, have had only limited use, perhaps because of awareness of measurement limitations. In this work, we seek to reduce the extent to which differential use of response scales obscures disparities in patient experience by race/ethnicity and SES, thereby improving our ability to appropriately address those disparities. Because the elderly use health care disproportionately and are potentially vulnerable, we pursue this approach using Medicare CAHPS data.

---

Address correspondence to Marc N. Elliott, Ph.D., RAND, 1776 Main St., PO Box 2138, Santa Monica, CA 90407-2138. e-mail: [elliott@rand.org](mailto:elliott@rand.org). Amelia M. Haviland, Ph.D., is with RAND, Pittsburgh, PA. David E. Kanouse, Ph.D., Katrin Hambarsoomian, M.S., and Ron D. Hays, Ph.D., are with RAND, Santa Monica, CA.

## COUNTERINTUITIVE RESULTS IN MEASURING DISPARITIES IN PATIENT EXPERIENCE

Several previous analyses of patient experience surveys have found counterintuitive patterns among subgroups, even after standard case-mix adjustment (CMA). The patterns include less positive evaluations for those with supplementary insurance (Bernard, Brody, and West 2004a) and those with higher income (Hetherington, Hopkins, and Roemer 1975), as well as more positive evaluations for African Americans (Bashshur, Metzner, and Worden 1967; Morales et al. 2001; Weech-Maldonado et al. 2001, 2003, 2004; Dayton et al. 2006) than for non-Hispanic whites.

Blacks and those with lower income do not receive care that better adheres to clinical guidelines (McGlynn et al. 2003); hence, one might not expect a priori that they would report better care experiences. The most commonly advanced explanation for these patterns has been that the experiences of disadvantaged groups result in lower expectations of care. In Medicare data, the primary focus of this work, beneficiaries have transitioned to more uniform coverage under Medicare from various levels of prior insurance coverage, where expectations about health care may have been formed (House et al. 1994). Patient expectations in turn affect satisfaction with inpatient and outpatient encounters (Jackson, Chamberlin, and Kroenke 2001; Noble et al. 2006).

## RESPONSE TENDENCIES AS A POSSIBLE EXPLANATION

Paulhus (1991) defines response bias as “a systematic tendency to respond to a range of questionnaire items on some other basis than the specific item content” (p. 17). If subgroups use response scales differently, response tendencies can obscure true disparities in patient experience by race/ethnicity and SES. We consider whether differing response tendencies, beyond those typically modeled, explain the counterintuitive patterns observed.

The response tendency that has received the most attention in consumer health care evaluations is positive response tendency (PRT), a tendency for some respondents to evaluate care more positively than others, given the same underlying experiences. This form of response bias can be addressed through standard regression-based CMA (Cleary and McNeil 1988; Hall and Dornan 1990; Kane, Maciejewski, and Finch 1997; Elliott et al. 2001), which assumes that certain patient characteristics are associated with a linear shift in response

tendency and seeks to offset this shift by subtracting the estimated bias from mean scores.

For example, higher educational attainment has consistently been associated with less positive evaluations of health care (Fox and Storms 1981; Elliott et al. 2001; Zaslavsky et al. 2001; O'Malley et al. 2005). A priori, one might suspect that those with better education receive better health care; hence, one might interpret a negative association between education and health care ratings as reflecting response bias. Similarly, older respondents typically evaluate care more positively than younger respondents (Elliott et al. 2001); again, this pattern probably reflects PRT more than consistently better care for older respondents.

PRT may not capture all important differences in response tendency across demographic subgroups. For example, digit preference, a tendency to systematically “round” answers to certain preferred digits (e.g., numbers ending in 0 or 5) in the absence of instructions to do so, can bias parameter estimates from surveys (Ridout and Morgan 1991; Klesges, DeBon, and Ray 1995; Crawford, Johannes, and Stellato 2002). Here we will focus on another form of response bias, *extreme response tendency* (ERT), which may be especially important when comparing consumer evaluations of health care.

## ERT

ERT (Hamilton 1968) is a systematic tendency for some respondents to prefer the endpoints of a response scale more than other respondents, given the same underlying experiences. A subgroup with higher ERT will have higher probabilities of endorsing both highly positive and highly negative values as opposed to intermediate values, relative to a reference group with lower ERT. ERT has been demonstrated to be stable for a given respondent across a broad variety of attitude items and over time in a large panel survey from the marketing literature (Greenleaf 1992b). Greenleaf (1992a) decomposes survey-reported attitudes into ERT and non-ERT components and finds that ERT components fail to predict corresponding reported behaviors, whereas the non-ERT components do predict these same behaviors, suggesting that the ERT component is in fact uninformative response bias.

Some subgroups are more likely to exhibit ERT than others; ERT has been correlated in a variety of settings with health insurance status, race/ethnicity, nationality, household income, and education. Medicaid beneficiaries demonstrate greater ERT than do commercially insured state employees in a statewide sample (Damiano et al. 2004) and commercial enrollees in a

national sample (Weech-Maldonado et al. 2008a). Damiano et al. (2004) found that when compared with commercially insured state employees, Iowa Medicaid enrollees had greater odds of providing 0–10 CAHPS health care ratings of “10” and “0–4” relative to a reference group of 5–8. Similarly, Weech-Maldonado et al. (2008a) found that Hispanics and white Medicaid-managed care enrollees had greater odds than white commercial enrollees of “0–4” and “10” responses on the same set of items. Greater ERT has been found for U.S. blacks relative to whites (Hui and Triandis 1985; Dayton et al. 2006). Greenleaf (1992b) found evidence of increasing ERT with age, but Hesterly (1963) suggests that such a pattern primarily involves greater ERT among adults older than 60 versus younger adults.

SES may be the construct most closely tied to ERT. Greenleaf (1992b) found that ERT decreases with education and household income (even in the same multivariate model), and in other work, negative correlations of  $-0.3$  to  $-0.5$  were found between measures of ERT and intelligence test scores (Hamilton 1968; Wilkinson 1970).

## CONSEQUENCES OF ERT

Schuman and Presser (1996) describe ERT as one of the most problematic sources of response bias in survey research, especially with skewed data (Baumgartner and Steenkamp 2001). Under these circumstances, ERT may masquerade as PRT: a symmetric tendency to endorse extremes (e.g., equal odds ratios of positive and negative extremes relative to a middle reference category) can cause a net positive shift in the mean. Misinterpretation of ERT as PRT could lead to adjustments that do not properly account for actual response bias.

The CAHPS measures we consider in this manuscript (0–10 global ratings) are consistently negatively skewed. For example, Landon et al. (2004) found that almost two-thirds of 0–10 ratings among Medicare beneficiaries fell in the highest 2 of 11 categories (9 or 10) for these CAHPS measures. Under these circumstances, standard CMA, which adjusts only for PRT, would not adjust enough at the positive end of the scale and would adjust in the wrong direction at the negative end if ERT differences were present.

## PREVIOUS APPROACHES TO MEASURING AND ADJUSTING FOR ERT

Previous approaches to measuring and adjusting for ERT have generally focused on the *proportion* of extreme responses (Schuman 1973; Hui

and Triandis 1985). Such methods are only appropriate when the underlying data have a symmetric rather than skewed distribution. Greenleaf (1992b) developed a measure of ERT for six-item categorical response scales based on the standard deviation of responses across items for a given respondent and a related adjustment (Greenleaf 1992a) that essentially attenuates or disattenuates according to whether the standard deviation is high or low around the respondent's own cross-item mean, noting that these approaches do not perform well with skewed data. With strongly negatively skewed CAHPS data, the standard deviation across items is strongly negatively associated with the mean and thus confounds ERT with PRT.

Weech-Maldonado et al. (2008a) note that odds ratios from multinomial logistic regression (a "difference-of-differences" comparison of the *odds* of extreme relative to nonextreme responses between two groups) effectively measures ERT even in the presence of skewed data, distinguishing ERT from PRT where a simple comparison of the proportion of extreme responses would not (also see Damiano et al. 2004).

## EDUCATIONAL ATTAINMENT AS A PROXY FOR ERT

Directly measuring ERT for individual subjects (e.g., Greenleaf 1992a) often requires both a very large number of items (i.e., 100 or more) and data that are not skewed. An approach based on a valid and widely available proxy for ERT might therefore be useful.

As noted earlier, SES in general, and perhaps educational attainment in particular, explains much of the observed variation in ERT. Educational attainment is a reliable and valid indicator of SES for virtually all adults and is a more meaningful measure of SES than income for the Medicare population, where current employment varies considerably. In this age group, educational attainment has the further advantage over income of being largely unaffected by current health status (Smith 2004). Although, as noted above, there is some evidence that age, ethnicity, and nationality are associated with ERT, these patterns generally correspond to the direction of effects that would be expected from educational differences. For example, in the present data, after considering education, we found no consistent association of age with ERT (specific results not reported). We therefore focus on educational attainment in the approach we develop here.

## A NEW APPROACH TO MODELING AND ADJUSTING FOR ERT

As we demonstrate below, greater educational attainment appears to be associated with both lower PRT and lower ERT. These two types of response bias suggest a two-stage adjustment. As a heuristic, we provide a two-stage conceptual model of the process that might result in these observed patterns and contrast it with the implicit one-stage model behind the use of standard CMA to adjust for PRT. The one-stage model might posit that individuals make a single-step judgment of the quality of their health care experience, and then PRT has a role in how they translate that judgment into a 0–10 rating.

We propose a two-stage heuristic model in which individuals first make a rough (implicit) categorization of their health care experience as *poor*, *fair*, or *good*. We propose that these might correspond to “super-categories” of 0–6, 7–8, and 9–10 ratings, respectively, with the “good” category comprising approximately the top seven deciles of the distribution of CAHPS scores, the “fair” category containing the next two deciles, and the “poor” category representing the lowest decile. This first stage would reflect actual health care experiences and PRT, but not ERT. In a second stage, individuals would select the specific rating within the previously chosen super-category to express their experiences (10 versus 9, 0 versus 1–6, etc.). This stage would reflect actual health care experiences and ERT, but not PRT. We devise a pattern of nonmonotonic adjustment for subgroup comparisons implied by this conceptual model and examine its impact on comparisons of CAHPS ratings among subgroups.

## METHODS

### *Data*

The 2004 CAHPS Medicare Fee-for-Service (FFS) survey instrument was fielded to a national probability sample of 168,000 original (FFS) Medicare beneficiaries. Beneficiaries qualify for Medicare by being older than 65 or legally disabled. FFS beneficiaries, those not enrolled in Medicare Advantage, the managed care version of Medicare, represent 85 percent of all Medicare beneficiaries (36 of 41 million total beneficiaries). The survey was distributed by mail, with phone follow-up, and resulted in 116,307 completed surveys, a 70 percent response rate among eligibles.

### *Measures*

We analyze three global ratings from the CAHPS instrument pertaining to personal doctor, overall health care, and health plan using an 11-point response scale verbally anchored at 0 (“worst possible”) and 10 (“best possible”).

We employ the standard Medicare CAHPS case-mix adjusters (Elliott, Hambarsoomians, and Edwards 2005; Elliott et al. 2008): age (<45, 45–64, 65–69, 70–74, 75–79, 80–84, and 85+ years), self-rated overall health (poor, fair, good, very good, and excellent), self-rated mental health (same five categories), education (no high school; some high school, but did not complete high school; high-school graduate or general educational development diploma [GED]; 1–3 years of college; and four-year college graduate), assistance with survey, and a Medicaid eligibility indicator. Dummies corresponding to 276 geographic units of contiguous counties within states are also included.

### *Statistical Analysis*

There were three phases of statistical analysis. First, we *establish* ERT in these data as a function of education. Second, we describe and illustrate a method to *adjust* for both ERT and PRT. Third, we *compare* our adjusted results with previous counterintuitive patterns found with standard adjustments.

#### *Establishing ERT as a Function of Educational Attainment*

We first demonstrate variation in ERT by educational attainment using a multinomial logistic regression for each of the three global ratings. The 0–10 response scale is grouped into four categories: 0–5, 6–8 (the omitted category), 9, or 10. This categorization deliberately separates the least extreme response from each of the two proposed super-categories (6 from 0–6 and 9 from 9–10) in order to contrast usage within super-categories by education.

#### *Adjusting for ERT then PRT*

Traditional CMA employs linear regression with case-mix adjusters that include educational attainment. Thus, adjustments are a function of respondents’ demographic characteristics, but not of the responses themselves. The two-stage process described here differs in that adjustments for education are a function of both education and the response (rating) itself. In the first stage, we adjust for education. In the second stage, we adjust for all other factors using the standard CMA mechanism. We split the stage 1 education adjustment into part 1a, which handles the ERT component of education, and part 1b, which handles the PRT component of education. Those with high school degrees



but no college attendance are used as the reference group for educational adjustments.

(1a) *Adjusting for Education-Based ERT*

We adjust for ERT by educational attainment within each of the three proposed rating super-categories: 0–6, 7 or 8, and 9 or 10. This adjustment, which is subtracted from the respondent's rating, varies only by educational attainment and the super-category of the chosen rating. It is calculated as the difference between (a) the mean rating among those who shared the respondent's educational attainment *and* chose the same super-category and (b) the mean rating among those with a high school degree who chose the same super-category. The within super-category adjustment for individual  $k$  in super-category  $i$  with educational attainment  $j$  is thus

$$W_{ijk} = \bar{s}_{ij} - \bar{s}_{iHS}$$

(1b) *Adjusting for Education-Based PRT*

The second part of the educational adjustment removes the PRT component that would exist even if response did not differ by education within super-category. This adjustment does not depend upon the rating itself, but varies only by education.

The educational PRT component is calculated as the difference between the mean rating by education that would result if education affected super-category choice but there were no response differences *within* super-categories. In particular, this “between super-category” effect of education is calculated as the difference in a weighted mean based on the super-category choices of those sharing the respondent's education and a weighted mean based on the super-category choices of those with a high school education. The between super-category adjustment for individual  $k$  in super-category  $i$  with educational attainment  $j$  is thus

$$B_{ijk} = \sum_i \bar{s}_{iHS} (\Pr(i|HS) - \Pr(i|j))$$

The total adjustment for education is the sum of these two adjustments.

(2) *Adjusting for PRT from Age, Health, Medicaid Status, and Proxy Use*

In the second stage, education-adjusted ratings are outcomes in a standard CMA model of PRT that omits education, which has already been incorporated. More specifically, ratings minus the education adjustments are the outcomes in a series of linear regressions with all standard case-mix adjusters other than education serving as predictors, along with dummies for geographic region. The second-stage adjustment is thus the difference

between the predicted outcome for the individual from this final model and the corresponding prediction at the mean of sample case mix characteristics other than education. The full adjustment is the sum of these two adjustments. We compare the two-stage adjustments with traditional adjustments that employ the same CMA variables.

*Comparing Subgroup Patterns with Those from Traditional Adjustments*

Finally, we compare patterns of subgroup results after the proposed two-stage adjustments with the patterns after traditional (one-stage) adjustments. Specifically, we compare adjusted outcomes for subgroups defined by insurance status and race/ethnicity based on three pairs of linear regressions (one for each of three outcomes and two methods) that simultaneously add dummy variables for insurance status (additional private insurance, Medicaid eligible, neither/reference), and race/ethnicity (black, Hispanic, non-Hispanic white/reference, other) to the predictors in the second-stage linear regression model for the new method as well as to the traditional CMA regression model. The coefficients on these demographic dummies represent estimates of multivariate adjusted subgroup differences from the reference category. We also include gender in the models because the insurance subgroups differ in gender composition.

## RESULTS

The distributions of the three 0–10 global ratings appear in Table 1a. Of the 116,307 respondents, all were eligible to rate Medicare, but only the 90 percent with a personal doctor or nurse and the 77 percent with some health care utilization in the past 6 months were eligible to rate personal doctors/nurses (hereafter “doctor” for brevity) and health care received, respectively. Of those who were eligible to respond, missingness for each of the three ratings was 4.3–5.5 percent. The distributions show the strong negative skewness that is typical of consumer health care evaluations, with 36–50 percent of all respondents selecting the best possible response. All three global rating questions have monotonic increases in the number of responses in each category from 0 to 10, with two exceptions. First, all three measures include more responses at five than for six, likely indicating some digit preference; second, for two of the ratings, eight is selected more often than nine.

The distributions of the standard case-mix adjustors appear in Table 1b. Approximately a quarter of the sample lacks a high school diploma, whereas

Table 1a: Distribution of the Three 0–10 Global Ratings

Ratings	N	Distribution of Responses (%)										
		0	1	2	3	4	5	6	7	8	9	10
Doctor/nurse	100,356	0.1	0.2	0.2	0.4	0.6	4.0	3.0	7.0	18.6	16.2	49.8
Health care received	90,254	0.2	0.2	0.2	0.5	0.7	2.9	2.5	5.9	17.3	21.4	48.2
Medicare	109,858	0.8	0.5	0.7	1.3	1.8	9.6	5.4	9.6	18.3	15.9	36.1

just over one-sixth completed a four-year college degree. About 10 percent are under the age of 65, all of whom are eligible for Medicare through disability. Nearly half of the sample is between 65 and 74 years of age, with about 10 percent older than 84 years. *Good* overall health and *very good* mental health are the modal and median response categories.

Table 1b: Descriptive Statistics of Respondent Characteristics ( $n = 116,307$ )

Predictors/CMA	%
Educational attainment	
Less than high school	24.6
High school	37.1
Some college	21.1
College degree or higher	17.3
Age	
≤ 44	2.0
45–64	9.4
65–69	23.5
70–74	22.4
75–79	19.3
80–84	13.8
85+	9.7
Self-rated overall health	
Excellent	6.0
Very good	20.9
Good	34.2
Fair	28.5
Poor	10.4
Self-rated mental health	
Excellent	22.3
Very good	31.0
Good	29.7
Fair	13.4
Poor	3.7
No proxy assistance	83.6
Dually eligible (for Medicaid)	11.7

Table 2 illustrates differences in ERT by educational attainment for each of the three global ratings, displaying the adjusted log odds of selecting a low (0–5) or either of the two highest categories (9 or 10) versus the omitted middle category (6–8) by educational attainment. Across all three outcomes, those with higher educational attainment have significantly lower odds of selecting response options at the highest (10) or lowest (0–5) ends of the scale. This pattern is monotonic for all three outcomes and is always significant for the low (0–5) and top (10) response categories. The small and inconsistent odds ratios for the 9 category suggest that this category, unlike the 10 category, does not represent “extreme response” relative to the reference category of 6–8.

In Table 3 we compare the proposed ERT/PRT education adjustment with the standard PRT-only education adjustment from CMA, both relative to a reference group of high school degree only. With standard CMA, everyone with the same level of educational attainment is treated equally, with small adjustments for those with no high school degree and positive and increasing adjustments as educational attainment increases beyond a high school diploma.

For those with at least a four-year college degree who responded “9” or “10,” the proposed ERT/PRT approach provides positive adjustments that are about twice as large as the standard approach. Those with similar education who responded 0–6 receive zero or negative adjustments with the ERT/PRT approach, as compared with the standard large positive adjustment with standard CMA. The patterns are similar but somewhat less dramatic for those with some college. For those without a high school degree, the pattern of adjustments is the opposite for the highest educational group. In particular, those without a high school diploma who provide 0–6 ratings have considerably more positive adjustments under ERT/PRT than under standard CMA. Adjustments that incorporate ERT differ markedly from traditional adjustments; this may affect comparisons of subgroups that differ in education.

Table 4 compares OLS estimates of differences in global ratings by subgroup obtained under standard CMA with OLS estimates of those same differences obtained after adjusting for ERT. Eight of the nine disparity estimates move away from the original counterintuitive finding, four of them by 0.1 units or more.

## DISCUSSION

Educational attainment, which is negatively associated with ERT, provides a unifying explanation for previously established patterns of ERT by race/ethnicity and insurance status in CAHPS surveys (Damiano et al.

Table 2: CMA-Adjusted Log Odds of Response Relative to Responses of 6–8 and High School Education by Educational Attainment

	Rate Personal Doctor or Nurse (n = 100,356)			Rate All Health Care from All Providers (n = 90,254)			Rate All Your Experience with Medicare (n = 109,858)		
	Rating of 0–5	Rating of 9	Rating of 10	Rating of 0–5	Rating of 9	Rating of 10	Rating of 0–5	Rating of 9	Rating of 10
	5%	16%	50%	5%	21%	48%	15%	16%	36%
% of All									
< High school degree	25	0.12**	0.29***	0.23***	-0.01	0.29***	0.19***	0.06*	0.41***
High school graduate	37	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Some college or a 2-year degree	21	-0.13**	0.01	-0.12*	0.02	-0.28***	-0.08**	-0.07**	-0.37***
BA or more	17	-0.39***	0.04	-0.43***	0.04	-0.72***	-0.22***	-0.09***	-0.85***

Ratings of 6–8 omitted.

\*\*\*p < .001.

\*\*p < .01.

\*p < .05.

Table 3: Comparison of Conventional and ERT-Based Adjustments by Education, Response Category, and Rating Type

Education	Response Category	Doctor/Nurse (n = 100,356)		Health Care (n = 90,254)		Medicare (n = 109,858)	
		Usual Adjust	New Adjust	Usual Adjust	New Adjust	Usual Adjust	New Adjust
< High school	0-6	-0.08	0.04	0.03	0.22	-0.22	-0.02
	7-8	-0.08	-0.09	0.03	0.03	-0.22	-0.23
	9-10	-0.08	-0.13	0.03	-0.02	-0.22	-0.30
Some college	0-6	0.06	0.01	0.08	0.06	0.20	0.20
	7-8	0.06	0.07	0.08	0.09	0.20	0.22
	9-10	0.06	0.10	0.08	0.13	0.20	0.25
College+	0-6	0.17	0.00	0.10	-0.05	0.26	0.11
	7-8	0.17	0.16	0.10	0.09	0.26	0.28
	9-10	0.17	0.29	0.10	0.25	0.26	0.41

2004; Weech-Maldonado et al. 2008a). We describe a two-stage model of response to a 0-10 global health care item and a corresponding two-stage adjustment that differs markedly from traditional adjustments and can accommodate the skewed data typical of consumer evaluations of health care.

As seen in Table 4, these adjustments result in more plausible comparisons of subgroups than do the standard adjustments. In two instances (underlined in Table 4) counterintuitive subgroup differences reverse signs with the new approach. With the standard approach, Medicaid eligibility was count-

Table 4: Comparison of OLS Coefficients from Complete ERT/PRT Adjustments to Standard CMA for Comparisons of Beneficiary Subgroups, also Adjusting for Gender

Outcome	Adjustment	Demographic Predictor		
		Medicaid Eligible	Black	Private Insurance
Doctor (RMSE 0.24) (n = 100,356)	Old	<u>+0.099</u>	<b>+0.251</b>	<u>-0.057</u>
	New	<u>-0.102</u>	<b>+0.136</b>	<u>+0.047</u>
Health care (RMSE 0.31) (n = 90,254)	Old	-0.352	<b>-0.063</b>	-0.001 m
	New	-0.347	<b>-0.142</b>	+0.070
Medicare (RMSE 0.45) (n = 109,858)	Old	<b>+1.181</b>	<b>+0.122</b>	-0.049
	New	<b>+1.032</b>	<b>+0.043</b>	-0.139

p < .05 for all effects of demographic predictors on outcomes, with the exception of cases marked m, for which .10 > p > .05.

erintuitively associated with higher doctor ratings and additional private insurance with lower doctor ratings. The newly proposed approach reverses both of these patterns. In another instance, (italicized in Table 4), a nonsignificant result becomes significant in the intuitive direction (more positive health care experiences with additional private insurance) using the new approach. In three other comparisons of blacks and whites and one comparison of Medicaid-eligible and noneligible beneficiaries (boldface in Table 4), the new approach substantially changes the estimated subgroup difference in the intuitive direction with no change in sign or statistical significance. For instance, while blacks still give their doctors higher ratings than whites, the estimated differences are half as large. In one instance (Medicaid-eligible ratings of health care), the difference was negligible. In only one case of nine did the proposed new approach result in a possibly counterintuitive pattern (lower ratings of Medicare for those with additional private insurance increasing in its magnitude).

The magnitude of education-based ERT adjustments is related to differences in education across the groups being compared. As a heuristic, one can compute mean education (<HS = 0, HS degree = 1, some college = 2, and BA or more = 3). In the present data, blacks and whites differ by  $\frac{1}{2}$  of a unit and beneficiaries eligible for Medicaid differ by  $\frac{3}{4}$  of a unit from those not eligible for Medicaid on this metric. Differences that are somewhat smaller than these, such as differences of  $\frac{1}{3}$  of a unit in this mean education measure, are large enough to affect conclusions regarding disparities.

More broadly, adjustment of these disparity estimates for ERT suggests that while factors such as patient expectations may have some role in explaining apparent paradoxes in racial/ethnic and SES disparities in patient experience, many of these paradoxes can be explained in significant part by differences in ERT. Future research involving the use of vignettes describing specific care scenarios could provide additional insight into the role of patient expectations.

## CONCLUSION

Despite evidence that the CAHPS surveys yield fairly similar levels of reliability and validity within racial and ethnic subgroups (e.g., Fongwa et al. 2006), we have shown that ERT, which has received only limited attention in the patient experience literature, can obscure measurement of disparities in patient experience with health care by race/ethnicity and SES. While we focus here on Medicare CAHPS data, these same issues apply when 0–10 global ratings are used on other CAHPS surveys, such as commercial health plan surveys and the CAHPS Hospital Survey (HCAHPS).

ERT is most important when comparing groups that differ substantially in mean educational attainment. This includes both explicit estimation of disparities by race/ethnicity or socioeconomic status and comparison of plans, hospitals, or providers that serve populations with especially low educational attainment (such as Medicare Special Needs Plans [SNPs] for Medicaid-eligible Medicare beneficiaries) or especially high educational attainment (specialists performing elective surgery).

Errors in estimating disparities may result in misidentifying the most pressing needs for interventions and may hamper evaluation of such initiatives. Errors in the evaluation of providers or other health care entities degrade the quality of information available to patients and may have financial consequences to providers as pay-for-performance initiatives increasingly incorporate patient experiences measures.

The consequences of not adjusting for ERT will vary. In comparisons of entities serving populations with similar education attainment, the consequences may be modest, but as noted above, the consequences may be substantial when comparing a plan serving a vulnerable population with low educational attainment to other plans serving those with greater educational attainment. Because ERT adjustments vary by the scores received, ERT adjustments for plans serving low-SES individuals would be more *positive* for lower scores and more *negative* for higher scores compared with standard adjustment. Thus the scores of a plan serving lower-SES individuals that appeared to score poorly would *benefit* from ERT adjustment (as the true performance would be better than was apparent without considering ERT), but a particularly high-scoring plan serving lower-SES individuals would be (appropriately) adjusted *downward*, because some of those high scores reflect their beneficiaries' greater ERT. Thus in a pay-for-performance setting, ERT adjustment might prevent understating the performance of the subset of those plans (or other entities) serving low SES individuals whose apparent performance was lowest—the very entities that might be most financially vulnerable to pay-for-performance.

In those situations in which ERT is likely to matter, a variety of options are available for limiting its influence. Policy makers and other stakeholders may prefer different approaches in different contexts. In this paper we present a two-stage adjustment for ERT and PRT that builds upon existing CMA for PRT. This approach may be straightforward to implement in some settings, such as those that currently employ CMA, but may be too complex for applications where parsimony is especially important.

Alternative approaches for limiting the influence of ERT might include grouping outcomes in specific ways or stratifying reporting by educational



attainment. Categorical reporting (with or without standard CMA) that does not split the hypothesized super-categories (e.g., reporting the proportion of 9s or 10s, not the proportion of 10s), as recommended by Damiano et al. (2004) and Weech-Maldonado et al. (2004), is one approach. If sample sizes permitted sufficient precision, one could stratify by educational attainment, rather than using regression adjustment, in settings such as public reporting of commercial health plans, where regression-adjustment is not currently employed (such as NCQA's public reporting of CAHPS commercial health plan survey data).

As consumer evaluations rapidly expand in scope, breadth of implementation, and uses (Darby, Crofton, and Clancy 2006), we have increasing opportunity to measure and address differences in patient experience. Doing so credibly and effectively will require careful attention to the best ways to capture these experiences from patient reports.

## ACKNOWLEDGMENTS

*Joint Acknowledgment/Disclosure Statement:* This study was funded by CMS contract HHSM-500-2005-000281 to RAND. Marc Elliott is supported in part by the Centers for Disease Control and Prevention (CDC U48/DP000056). No authors had any other conflicts of interest to report. The authors would like to thank CMS Project Officer Elizabeth Goldstein for her support. The authors would like to thank Jacquelyn Chou, Kate Sommers-Dawes, and Scott Stephenson for assistance with the preparation of the manuscript.

*Disclaimers:* none.

## REFERENCES

- Bashshur, R. L., C. A. Metzner, and C. Worden. 1967. "Consumer Satisfaction with Group Practice, the CHA Case." *American Journal of Public Health and the Nation's Health* 57 (11): 1991-9.
- Baumgartner, H., and J.-B.E.M. Steenkamp. 2001. "Response Styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research* 38 (2): 143-56.
- Bernard, S. L., E. R. Brody, and N. D. West. 2004a. *Medicare CAHPS Fee-for-Service Survey, Subgroup Analysis Results (2000-2003)*. Research Triangle Park, NC: RTI International.
- . 2004b. *Medicare CAHPS Fee-for-Service Survey, Subgroup Analysis Results (2000-2003)*. Research Triangle Park, NC: RTI International.
- Cleary, P. D., and B.J. McNeil. 1988. "Patient Satisfaction as an Indicator of Quality of Care." *Inquiry* 25 (1): 25-36.

- Crawford, S. L., C. B. Johannes, and R. K. Stellato. 2002. "Assessment of Digit Preference in Self-Reported Year at Menopause: Choice of an Appropriate Reference Distribution." *American Journal of Epidemiology* 156 (7): 676–83.
- Damiano, P. C., M. N. Elliott, M. Tyler, and R. D. Hays. 2004. "Differential Use of the CAHPS 0-10 Global Rating Scale by Medicaid and Commercial Populations." *Health Services and Outcomes Research Methodology* 5 (3–4): 193–205.
- Darby, C., C. Crofton, and C. M. Clancy. 2006. "Consumer Assessment of Health Providers and Systems (CAHPS): Evolving to Meet Stakeholder Needs." *American Journal of Medical Quality* 21 (2): 144–7.
- Dayton, E., C. Zhan, J. Sangl, C. Darby, and E. Moy. 2006. "Racial and Ethnic Differences in Patient Assessments of Interactions with Providers: Disparities or Measurement Biases?" *American Journal of Medical Quality* 21 (2): 109–14.
- de Silva, A., and N. Valentine. 2000. "Measuring Responsiveness: Results of a Key Informants Survey in 35 Countries." GPE Discussion Paper Series: No. 21. Geneva, Switzerland: World Health Organization.
- Elliott, M. N., M. K. Beckett, K. Chong, K. Hambarsoomians, and R. D. Hays. 2008. "How Do Proxy Responses and Proxy-Assisted Responses Differ from What Medicare Beneficiaries Might Have Reported about their Health Care?" *Health Services Research* 4 (3): 833–48.
- Elliott, M. N., K. Hambarsoomians, and C. A. Edwards. 2005. "Analysis of Case-Mix Strategies and Recommendations for Medicare Fee-for-Service CAHPS." Case-Mix Adjustment Report. Santa Monica, CA: The RAND Corporation.
- Elliott, M. N., R. Swartz, J. Adams, K. L. Spritzer, and R. D. Hays. 2001. "Consumer Evaluations of Health Plans and Health Care Providers: Case-Mix Adjustment of the National CAHPS Benchmarking Data 1.0: A Violation of Model Assumptions?" *Health Services Research* 36 (3): 555–73.
- Fongwa, M. N., W. Cunningham, R. Weech-Maldonado, P. R. Gutierrez, and R. D. Hays. 2006. "Comparison of Data Quality for Reports and Ratings of Ambulatory Care by African Americans and White Medicare Managed Care Enrollees." *Journal of Aging and Health* 18 (5): 707–21.
- Fox, J. G., and D. M. Storms. 1981. "A Different Approach to Sociodemographic Predictors of Satisfaction with Healthcare." *Social Science and Medicine* 15A: 557–64.
- Frist, W. H. 2005. "Overcoming Disparities in U.S. Health Care." *Health Affairs* 24 (2): 445–51.
- Greenleaf, E. A. 1992a. "Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles." *Journal of Marketing Research* 29: 176–88.
- . 1992b. "Measuring Extreme Response Style." *Public Opinion Quarterly* 56 (3): 328–51.
- Hall, J. A., and M. C. Dornan. 1990. "Patient Sociodemographic Characteristics as Predictors of Satisfaction with Medical Care: A Meta-Analysis." *Social Science and Medicine* 30 (7): 811–8.
- Hamilton, D. C. 1968. "Personality Attributes Associated with Extreme Response Style." *Psychological Bulletin* 69 (3): 192–203.

- Hesterly, S. O. 1963. "Deviant Response Patterns as a Function of Chronological Age." *Journal of Consulting and Clinical Psychology* 27: 210–4.
- Hetherington, R. W., C. E. Hopkins, and M. I. Roemer. 1975. *Health Insurance Plans: Promise and Performance*. New York: Wiley.
- House, J. S., J. M. Lepkowski, A. M. Kinney, R. P. Mero, R. C. Kessler, and A. R. Herzog. 1994. "The Social Stratification of Aging and Health." *Journal of Health and Social Behavior* 35 (3): 213–34.
- Hui, C. H., and H. C. Triandis. 1985. "The Instability of Response Sets." *Public Opinion Quarterly* 49: 253–60.
- Institute of Medicine. 2002. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies.
- Jackson, J. L., J. Chamberlin, and K. Kroenke. 2001. "Predictors of Patient Satisfaction." *Social Science and Medicine* 52 (4): 609–20.
- Kane, R. L., M. Maciejewski, and M. Finch. 1997. "The Relationship of Patient Satisfaction with Care and Clinical Outcomes." *Medical Care* 35 (7): 714–30.
- Klesges, R. C., M. DeBon, and J. W. Ray. 1995. "Are Self-Reports of Smoking Rate Biased?" *Journal of Clinical Epidemiology* 48 (10): 1225–33.
- Landon, B. E., A. M. Zaslavsky, S. L. Bernard, M. J. Cioffi, and P. D. Cleary. 2004. "Comparison of Performance of Traditional Medicare vs Medicare Managed Care." *Journal of the American Medical Association* 291 (14): 1744–52.
- McGlynn, E. A., S. M. Asch, J. Adams, J. Keeseey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *New England Journal of Medicine* 348 (26): 2635–45.
- Morales, L. S., M. N. Elliott, R. Weech-Maldonado, K. L. Spritzer, and R. D. Hays. 2001. "Differences in CAHPS Adult Survey Ratings and Reports by Race and Ethnicity: An Analysis of the National CAHPS Benchmarking Data 1." *Health Services Research* 36 (3): 595–617.
- National Health Plan Collaborative. 2006. "Reducing Racial and Ethnic Disparities Improving Quality of Health Care." Phase 1 summary report. Hamilton, NJ: National Health Plan Collaborative.
- National Research Council. 2004. *Eliminating Health Disparities: Measurement and Data Needs*. Washington, DC: National Academies Press.
- Noble, P. C., M. A. Conditt, K. F. Cook, and K. B. Mathis. 2006. "The John Insall Award: Patient Expectations Affect Satisfaction with Total Knee Arthroplasty." *Clinical Orthopaedics and Related Research* 7: 35–43.
- Office of Disease Prevention and Health Promotion. 2008. "Healthy People 2010" [accessed on July 18, 2008]. Available at <http://www.healthypeople.gov/data/midcourse/html/focusareas/FA11TOC.htm>
- O'Malley, A. J., A. M. Zaslavsky, M. N. Elliott, L. Zaborski, and P. D. Cleary. 2005. "Case-Mix Adjustment of the CAHPS Hospital Survey." *Health Services Research* 40 (6 part 2): 2162–81.
- Onstad, K. 2005. "Research Brief: Racial and Ethnic Disparities in the Experiences of Health Care Consumers." National CAHPS Benchmarking Database, AHRQ Contract No. 290-01-0003.

- Paulhus, D. L. 1991. "Measurement and Control of Response Bias." In *Measures of Personality and Social Psychological Attitudes*, edited by J. P. Robinson, P. R. Shaver, and L. S. Wrightsman, pp. 17–59. New York: Academic Press.
- Ridout, M. S., and B. J. T. Morgan. 1991. "Modelling Digit Preference in Fecundability Studies." *Biometrics* 47 (4): 1423–33.
- Schuman, H. 1973. "A Comparison of Two Scales of Extreme Response Bias." *Public Opinion Quarterly* 37: 407–12.
- Schuman, H., and S. Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Thousand Oaks, CA: Sage.
- Smith, J. 2004. *Unraveling the SES-Health Connection*. New York: The Population Council.
- Weech-Maldonado, R., M. N. Elliott, L. S. Morales, K. Spritzer, G. N. Marshall, and R. D. Hays. 2004. "Health Plan Effects on Patient Assessments of Medicaid Managed Care among Racial/Ethnic Minorities." *Journal of General Internal Medicine* 19 (2): 136–45.
- Weech-Maldonado, R., M. N. Elliott, T. Oluwole, K. C. Schiller, and R. D. Hays. 2008a. "Survey Response Style and Differential Use of CAHPS Rating Styles by Hispanics." *Medical Care* 46 (9): 963–8.
- Weech-Maldonado, R., M. Fongwa, P. Gutierrez, and R. D. Hays. 2008b. "Language and Regional Differences in Evaluations of Medicare Managed Care by Hispanics." *Health Services Research* 43 (2): 552–68.
- Weech-Maldonado, R., L. S. Morales, M. Elliott, K. Spritzer, G. Marshall, and R. D. Hays. 2003. "Race/Ethnicity, Language, and Patients' Assessments of Care in Medicaid Managed Care." *Health Services Research* 38 (3): 789–808.
- Weech-Maldonado, R., L. S. Morales, K. Spritzer, M. N. Elliott, and R. D. Hays. 2001. "Racial and Ethnic Differences in Parents' Assessments of Pediatric Care in Medicaid Managed Care." *Health Services Research* 36 (3): 575–94.
- Wilkinson, A. E. 1970. "Relationship between Measures of Intellectual Functioning and Extreme Response Style." *Journal of Social Psychology* 81 (2): 271–2.
- Zaslavsky, A. M., L. B. 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: Other Contributions.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.