

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2012.01452.x
RESEARCH ARTICLE

Reporting CAHPS and HEDIS Data by Race/Ethnicity for Medicare Beneficiaries

Steven C. Martino, Robin M. Weinick, David E. Kanouse, Julie A. Brown, Amelia M. Haviland, Elizabeth Goldstein, John L. Adams, Katrin Hambarsoomian, David J. Klein, and Marc N. Elliott

Objective. To produce reliable and informative health plan performance data by race/ethnicity for the Medicare beneficiary population and to consider appropriate presentation strategies.

Data Sources. Patient experience data from the 2008–2009 Medicare Advantage (MA) and fee-for-service (FFS) CAHPS surveys and 2008–2009 HEDIS data (MA beneficiaries only).

Study Design. Mixed effects linear (and binomial) regression models estimated the reliability and statistical informativeness of CAHPS (HEDIS) measures.

Principal Findings. Seven CAHPS and seven HEDIS measures were reliable and informative for four racial/ethnic subgroups—Whites, Blacks, Hispanics, and Asian/Pacific Islanders—at sample sizes of 100 beneficiaries (200 for prescription drug plans). Although many plans lacked adequate sample size for reporting group-specific data, reportable plans contained a large majority of beneficiaries from each of the four racial/ethnic groups.

Conclusions. Statistically reliable and valid information on health plan performance can be reported by race/ethnicity. Many beneficiaries may have difficulty understanding such reports, however, even with careful guidance. Thus, it is recommended that health plan performance data by subgroups be reported as supplemental data and only for plans meeting sample size requirements.

Key Words. CAHPS, HEDIS, patient experience, public reporting, racial/ethnic differences

Public reporting on the quality of health plans and providers has become increasingly common as its advocates stress its potential to reduce costs and improve quality by informing consumer choice and motivating quality improvement initiatives (Berwick, James, and Coye 2003; Hibbard, Stockard,

and Tusler 2003). Despite momentum in this area, evidence that public reports aid consumer decision making is mixed (Fung et al. 2008; Harris and Buntin 2008; Faber et al. 2009; Kolstad and Chernew 2009).

For public reporting to support consumer decision making, consumers must be able to interpret and act on reported measures (Christianson et al. 2010; Hibbard, Greene, and Daniel 2010). To date, all public reporting of quality data has been based on care experienced by the *average* health plan enrollee or hospital patient. This practice makes sense if the experiences of the average enrollee or patient are applicable to all individuals using them for decision making. In practice, however, the best plans and providers for some groups may not be the best for all (Keenan et al. 2009; Elliott et al. 2011).

Evidence that patient experience within a given plan or hospital may vary with patient characteristics has been particularly evident for racial/ethnic subgroups. For example, in an investigation of racial/ethnic differences in consumer assessments of commercial and Medicare managed care plans, Lurie et al. (2003) found substantial plan-to-plan variation in the gap between whites and other racial/ethnic groups in experience with, access to, and use of care. That is, whereas some plans showed no differences between whites and racial/ethnic minority groups on these variables, other plans showed large differences. Similarly, Trivedi et al. (2006) found substantial variation within as well as between Medicare plans in racial/ethnic disparities on each of four Healthcare Effectiveness Data and Information Set (HEDIS) outcome measures. More recently, Elliott et al. (2010) found that the relative standings of hospitals' patient experience scores were highly heterogeneous for patients of different race/ethnicity. Together, these findings suggest that a one-size-fits-all approach to public reporting based on the experiences of the average health plan enrollee or hospital patient might not be optimal for all purposes.

Performance data stratified by race/ethnicity may have greater utility for consumers when they are selecting health plans and providers. Stratified performance reports could also raise awareness of racial/ethnic disparities in care and spur efforts to improve care for subgroups of patients who experience poorer access, treatment, and quality of care (Fiscella et al. 2000). Each of

Address correspondence to Steven C. Martino, Ph.D., RAND, 4570 Fifth Avenue, Suite 600, Pittsburgh, Pennsylvania 15213-2665, e-mail: martino@rand.org. Robin M. Weimick, Ph.D., is with RAND, Arlington, VA. David E. Kanouse, Ph.D., Marc N. Elliott, Ph.D., and Julie A. Brown, B.A., are with RAND, Santa Monica, CA. Amelia M. Haviland, Ph.D., is with Carnegie Mellon University, Pittsburgh, PA. Elizabeth Goldstein, Ph.D., is with Centers for Medicare & Medicaid Services, Baltimore, MD. John L. Adams, Ph.D., and Katrin Hambarsoomian, M.S., are with RAND, Santa Monica, CA. David J. Klein, M.S., is with Children's Hospital, Boston, MA.

these goals is evident in the Medicare Improvements for Patients and Providers Act of 2008 (MIPPA), which proposed the collection of data for measuring, evaluating, and reporting on health disparities within the Medicare population. To the best of our knowledge, health care quality measures have not previously been publicly reported by race/ethnicity.

Realizing the goals of stratified public reporting by race/ethnicity requires addressing some challenging issues (Elliott et al. 2010). One issue is how to achieve adequate precision of measurement without increasing sample size and accompanying costs. Although the sample sizes recommended in current performance reporting systems (e.g., the Consumer Assessment of Healthcare Providers and Systems [CAHPS] surveys) allow for accurate measurement of the average enrollee or patient's experiences by plan or hospital, they typically would not allow for accurate measurement of the experiences of specific racial/ethnic minority groups within plans/hospitals. This article describes how we dealt with this issue in an effort to produce and report racial/ethnic group-specific quality of care data for the MA and FFS Medicare beneficiary populations. In making recommendations about how to report these data, we also considered—although we did not investigate—challenges that consumers may face in interpreting and using racial/ethnic data on quality.

METHODS

Overarching Aims

Our approach was guided by several principles—most important, identifying group-specific quality of care measures that are both reliable and informative. *Reliability* is the extent to which scores on a particular measure for a given racial/ethnic group distinguish true differences in care among health care entities (e.g., plans) versus differences due to sampling error. Experts in measurement recommend reliabilities of 0.85 or higher for high-stakes applications such as pay-for-performance (Roland et al. 2009; Adams et al. 2010; Elliott et al. 2010). We sought reliabilities of 0.70 or higher, a criterion considered acceptable for lower stakes applications, such as providing the type of supplementary data that was our focus.

The second principle was that of informativeness: the extent to which racial/ethnic-specific data for health care entities provide more information about the relative performance of those entities than what can be inferred from relative performance information for the average beneficiary. Although evidence suggests that average health care quality and patient experiences

often vary with race/ethnicity (Morales et al. 2001; Weech-Maldonado et al. 2003, 2004; Goldstein et al. 2010), differences that are consistent across health care entities do not provide a basis for reporting plan scores by race/ethnicity. Reporting by race/ethnicity is only warranted when racial/ethnic differences vary substantially among plans, so that two plans with the same average scores might have very different scores for members of a given racial/ethnic group. In this study, informativeness was calculated as the proportion of information about plan performance for a minority group that is obscured if all that is available is the measure of plan performance for whites. For most plans, the average performance for whites approximates the average performance for all racial/ethnic groups combined, given whites' greater prevalence among beneficiaries. For each performance measure, we sought a 20 percent or higher information gain for at least one racial/ethnic group, a value indicating at least moderate informativeness. The criteria of reliability and informativeness should be jointly considered; subgroup reporting is helpful only when there is a gain in group-specific information that outweighs the loss in sample size, precision, and reliability that comes with drilling down to the subgroup level.

We also aimed for consistency in our approach to providing racial/ethnic-specific performance data. In particular, our goal was to report a consistent set of measures across racial/ethnic groups for all plans and to develop reporting requirements that were consistent across measures. For example, for the sake of consistency across racial/ethnic groups, a measure that had good reliability for some groups but poor reliability for others was excluded from reporting for all groups. Similarly, rather than devising sample size requirements that varied across measures or racial/ethnic groups, we sought to establish requirements that could be applied uniformly.

Data Sources

Data for this study came from two sources: the Medicare CAHPS (MCAHPS) survey and HEDIS data. MCAHPS and HEDIS data were pooled across the years 2008 and 2009, to increase available sample size for analysis and increase the reliability of estimates.¹

MCAHPS Survey

CMS conducts the MCAHPS survey to collect, analyze, and report data on beneficiaries' experiences with care and services. All types of Medicare coverage are included in this survey: Medicare Advantage (MA), either with-

out Part D prescription drug coverage (MA-Only) or with such options (MA-PD), original fee-for-service Medicare without Part D coverage (FFS-Only), and original fee-for-service Medicare with a stand-alone prescription drug plan (FFS-PD). The MCAHPS survey is a mail survey with telephone follow-up based on a stratified random sample of Medicare beneficiaries, with states as strata for FFS-Only beneficiaries and contracts as strata for all others. A Medicare contract, which might be commonly called a plan, is a set of offerings from a single health plan sponsor in a specific geographic area. From this point forward, we refer to these entities, properly, as contracts rather than plans. The 2008 (2009) MCAHPS survey attempted to contact 671,280 (672,919) Medicare beneficiaries and received responses from 407,543 (415,902), for a 60.7 percent (61.8 percent) response rate. The survey represented all FFS beneficiaries and MA beneficiaries from the 575 (541) MA contracts with more than 600 enrollees. Across the 2 years, a total of 492,495 responses were received from MA beneficiaries and 336,438 responses from FFS beneficiaries (238,195 of whom had PDP coverage). Although the larger set of analyses includes data from all types of Medicare beneficiaries, analyses presented in this article are restricted to MA-only, MA-PD, and FFS-PD.

Appendix Table A1 contains detail on the 13 CAHPS measures considered for analysis. These include three doctor performance measures, four measures related to health plan and care received, two beneficiary-reported immunization items,² three prescription drug measures, and a customer service composite. The five global ratings were excluded based on concerns about the validity of these items for making comparisons by race/ethnicity. In particular, there is evidence that response tendencies to CAHPS global ratings vary with race/ethnicity (Weech-Maldonado et al. 2008; Elliott et al. 2009a,b), such that comparisons of mean global ratings across races/ethnicities could be misleading. Such variation in response tendency is not evident for other CAHPS measures (Weinick et al. 2011). The remaining eight measures were subjected to the analyses described below.

HEDIS Measures

HEDIS consists of health care process measures and intermediate outcome measures that are based on administrative data, supplemented in some cases by information obtained from individuals' medical records (National Committee for Quality Assurance 2011). Our analysis used data from 5.7 million records of care provided to Medicare beneficiaries in 382 MA plans in 2008–2009. HEDIS data were not available for FFS beneficiaries; however,

we did limit our consideration to 16 HEDIS measures for which FFS equivalents exist. Detail on these measures, each of which focuses on use of effective care, is presented in Appendix Table A2.

As the appendix shows, there are four measures for monitoring patients who take specific medications. These measures were combined to create a fifth (summary) measure with larger sample sizes and greater reliability than would be obtained from its four constituent measures, which we excluded from further consideration. The remaining 12 measures were subjected to the analyses described below.

Measurement/Estimation of Race/Ethnicity in CAHPS and HEDIS

The MCAHPS survey asks beneficiaries whether they are of Hispanic or Latino origin or descent and subsequently asks their race, with response options of “White,” “Black or African American,” “Asian,” “Native Hawaiian or other Pacific Islander,” and “American Indian or Alaska Native” and the opportunity to mark one or more responses. Ten percent of beneficiaries were missing data on race/ethnicity and were excluded from our analysis. If a beneficiary identified as Hispanic, we classified the beneficiary as Hispanic regardless of races endorsed. If a beneficiary did not identify as Hispanic and endorsed exactly one race or endorsed Asian and Pacific Islander and no other race, we classified the beneficiary as American Indian/Alaskan Native (AI/AN), black, API, or white. If a beneficiary endorsed two or more races (except Asian and Pacific Islander), we classified the beneficiary as multiracial. Thus, we had six mutually exclusive racial/ethnic categories.

Whereas MCAHPS data include self-reported race/ethnicity, HEDIS data do not. CMS has administrative data that can be linked to beneficiary HEDIS records. Although these administrative data on race/ethnicity are suitable for identifying black beneficiaries, they do not perform well in identifying Hispanics and Asian/Pacific Islanders (Elliott et al. 2009a, b). This limitation makes the data unsuitable for comparing HEDIS scores by race/ethnicity. To address this limitation, we adapted a new method for inferring race/ethnicity from surname and residential address data while taking advantage of the racial/ethnic information present in the CMS administrative data. In a validation of this method that used data from approximately 2 million commercially insured beneficiaries (Elliott et al. 2009a, b), indirectly estimated race/ethnicity had 93 percent concordance overall with self-reported race-ethnicity. Although the method performs well for identifying white (93 percent concordance in Elliott et al. 2009a, b), black (93 percent

concordance), Hispanic (95 percent concordance), and Asian/Pacific Islander (API; 94 percent concordance) groups, it is not yet practical for AI/AN or multiracial groups. Given that, and considering that AI/AN and multiracial group members constitute very low percentages of the MCAHPS sample (<1 and 2 percent, respectively), we excluded these two groups from our analysis. Thus, our analysis focused on four groups: Hispanics, blacks, Asian/Pacific Islanders, and whites.

Analytic Approach

For each contract represented in the 2008–2009 MCAHPS survey, we estimated the contract-level reliability and informativeness of eight CAHPS measures (see Appendix Table A1) for Medicare Advantage contracts and of two prescription drug composites for free-standing PDP contracts via mixed linear regression models. The models included fixed effects for a set of standard case mix adjustment variables (Zaslavsky et al. 2001; O'Malley et al. 2005),³ three racial/ethnic group indicators (omitting white), and an indicator of survey year. These models also included random effects for contract (MA and PDP) and race/ethnicity by contract interactions. For each contract represented in the 2008–2009 HEDIS data sets, we estimated the contract-level reliability and informativeness of HEDIS measures via mixed effects binomial regression models, using estimated probabilities of being in each of four racial/ethnic groups—Hispanic, black, white, and API—in place of racial/ethnic group indicators. These models included fixed effects for racial/ethnic group probabilities and data collection year and random effects for contract (MA) and race/ethnicity by contract interactions. Following current practice, we did not use case mix adjustment in analyzing HEDIS measures.

Within each racial/ethnic group, reliability was estimated as the ratio of the variance in ratings between contracts over the sum of the between-contract variance and sampling error (Hays et al. 1999; Solomon et al. 2002). Informativeness was estimated in two steps. First, we computed disattenuated within-contract correlations of each CAHPS and HEDIS score for each racial/ethnic minority group with the corresponding score for whites. These correlations were calculated as the square root of the ratio of the contract variance component to the sum of the contract and contract by racial/ethnic variance components. We then estimated the informativeness of each measure by squaring the within-contract correlation

between the score for the minority group and the score for whites and subtracting that value from one.

RESULTS

Reliability

CAHPS Measures. For MA contracts, 100 completed cases sometimes provided reliability exceeding 0.80, usually provided reliability exceeding 0.70, and always provided reliability exceeding 0.60 for all four racial/ethnic subgroups and measures except the doctor communication measure. Doctor communication had reliability <0.60 for three of four subgroups and required 300 completed cases to achieve this criterion for all subgroups. Poor reliability at sample sizes adequate for other measures led us to exclude the doctor communication measure from reporting. Neither Part D measure achieved 0.60 reliability for all four subgroups for free-standing PDPs at 100 completed cases, but both had reliability at 200 completed cases that was similar to what was achieved in MA with 100 completed cases. Thus, we adopted 100 completed cases as the minimum sample size for a given racial/ethnic group for a given measure for MA contracts and 200 completes for free-standing PDP contracts.

Although higher reliability could have been achieved by setting more stringent sample size criteria (i.e., requiring a greater number of completed cases), the criteria adopted represent a trade-off between reliability and the number of contracts that would be reportable for racial/ethnic minority groups. Table 1 shows the number of contracts for which 100, 200, or 300 completed CAHPS surveys were available for each of the four racial/ethnic subgroups. Although the number of responses for each CAHPS measure is the more relevant criterion, the number of completed surveys is a convenient summary for CAHPS measures, as most beneficiaries who complete the survey also respond to each measure. About 91 percent of the 459 MA contracts had 100 or more completes from whites, about 27 percent had 100 or more completes from each of blacks and Hispanics, and about 5 percent had 100 or more completes from API.

Although these proportions may seem low, they provide coverage for a majority of beneficiaries of the respective racial/ethnic subgroups (65 percent of API, 85 percent of blacks, 91 percent of Hispanics, and >99 percent of whites) because the contracts reaching this threshold tended to be large contracts with larger total sample sizes and higher proportions of the minority

Table 1: Percent of Contracts with a Minimum of 100, 200, or 300 Completed 2008–2009 CAHPS Surveys for Each Racial/Ethnic Group and the Percent of Beneficiaries in Contracts with a Minimum of 100, 200, and 300 Complete Surveys per Contract

<i>Racial/ Ethnic Group</i>	<i>Coverage Type</i>	<i>Total No. of Contracts</i>	<i>Percent of Contracts with a Minimum of k Complete Surveys per Contract</i>			<i>Percent of Beneficiaries in Contracts with a Minimum of k Complete Surveys per Contract</i>		
			<i>k = 100</i>	<i>k = 200</i>	<i>k = 300</i>	<i>k = 100</i>	<i>k = 200</i>	<i>k = 300</i>
Hispanic	MA	459	27	16	10	91	81	71
	FS-PDP	82	43	29	22	96	92	84
Black	MA	459	28	13	7	85	69	53
	FS-PDP	82	38	27	20	97	93	88
API	MA	459	5	3	2	65	59	54
	FS-PDP	82	17	11	6	82	67	36
White	MA	459	91	84	76	100	99	98
	FS-PDP	82	95	93	88	100	100	100

group in question (see Table 1). As free-standing PDP contracts had larger average sample sizes than MA contracts, a threshold of 200 completed cases corresponds to similar proportions of contracts for blacks, Hispanics, and whites and a greater proportion (11 percent) for API. Similar proportions of Hispanic, white, and API beneficiaries are covered, as are a greater proportion (93 percent) of black beneficiaries. Table 2 shows the number (and percentage) of MA contracts reaching a threshold of 100 completed cases for each of the seven retained CAHPS measures. These values are similar to the number (and percentage) of contracts with 100 completed surveys except for the two

Table 2: Number (Percentage) of Medicare Advantage Contracts (out of 459) with at Least 100 Responses on a Given Racial/Ethnic Group and Measure, Pooled 2008–2009 CAHPS Data

	<i>Getting Needed Care</i>	<i>Getting Care Quickly</i>	<i>Customer Service</i>	<i>Getting Needed Prescription Drugs</i>	<i>Getting Needed Prescription Drug Information</i>	<i>Had Flu Shot in Last Year</i>	<i>Had Pneumonia Shot</i>
Hispanic	89 (19)	105 (23)	62 (14)	109 (24)	41 (9)	106 (23)	95 (21)
Black	67 (15)	110 (24)	43 (9)	103 (22)	19 (4)	97 (21)	90 (20)
API	15 (3)	19 (4)	10 (2)	19 (4)	6 (1)	19 (4)	16 (3)
White	400 (87)	409 (89)	359 (78)	388 (85)	280 (61)	399 (87)	397 (86)

measures applicable to smaller proportions of beneficiaries: customer service and getting information about prescription drugs.

HEDIS Measures. Sample sizes of 100 per racial/ethnic group⁴ and contract produced reliabilities of 0.70 or higher for all four racial/ethnic subgroups for all HEDIS measures under consideration except for LDL-C screening as part of cardiovascular care and persistence of beta blocker treatment after a heart attack. These two measures were excluded because they had inadequate reliability for racial/ethnic subgroups at a sample size that was adequate for all other HEDIS measures.

Unlike for CAHPS measures, eligibility for many HEDIS measures is limited to only a small proportion of beneficiaries. Accordingly, three measures were excluded because they did not meet contract-specific sample size requirements, despite having adequate reliability at the required sample size: prescription therapy for rheumatoid arthritis and the two measures of antidepressant medication management. These three measures had total sample sizes (across contracts) that were less than half the lowest sample size for any of the seven measures that were retained: the four diabetes measures, two cancer screening measures, and the summary measure of monitoring patients taking long-term medication. Table 3 shows the number (and percentage) of MA contracts reaching a threshold of 100 completed cases for each of the seven retained HEDIS measures. For API and white beneficiaries, the percentages of contracts with 100 completed cases per measure are comparable to those observed for the CAHPS measures; for blacks and Hispanics, the percentages are considerably higher.

Table 3: Number (Percentage) of Medicare Advantage Contracts (out of 382) with at Least 100 Responses on a Given Racial/Ethnic Group and Measure, Pooled 2008–2009 Medicare Advantage HEDIS Data

	<i>Diabetes Care</i>						
	<i>LDL-C Screening</i>	<i>Blood Sugar Testing</i>	<i>Retinal Eye Exam</i>	<i>Medical Attention for Nephropathy</i>	<i>Breast Cancer Screening</i>	<i>Colon Cancer Screening</i>	<i>Long-Term Medication Management</i>
Hispanic	95 (25)	95 (25)	93 (24)	84 (22)	98 (26)	80 (21)	179 (47)
Black	136 (36)	136 (36)	137 (36)	118 (31)	148 (39)	90 (24)	260 (68)
API	24 (6)	24 (6)	25 (7)	15 (4)	44 (12)	33 (9)	112 (29)
White	333 (87)	333 (87)	332 (87)	290 (76)	295 (77)	272 (71)	357 (93)

Table 4: Percentage of Information about Minority Group Scores on CAHPS and HEDIS Measures Not Contained in Score for White

	<i>Hispanic</i>	<i>Black</i>	<i>API</i>
MA CAHPS measures			
Getting needed care	44	42	71
Getting care quickly	8	23	60
Customer service	23	21	55
Had flu shot in last year	39	17	14
Had pneumonia shot	36	23	38
Getting needed prescription drugs	48	36	63
Getting needed prescription drug information	44	39	39
FS-PDP CAHPS measures			
Getting needed prescription drugs	45	28	72
Getting needed prescription drug information	24	2	72
MA HEDIS measures			
Diabetes care: LDL-C screening	21	14	21
Diabetes care: Blood sugar testing	24	14	17
Diabetes care: Retinal eye exam	14	8	12
Diabetes care: Medical attention for nephropathy	26	15	31
Breast cancer screening	45	19	33
Colorectal cancer screening	26	10	23
Long-term medication management	19	15	23

Informativeness

CAHPS Measures. For each CAHPS measure deemed reliable at the adopted sample size, the top half of Table 4 shows the percentage of information about minority group scores that is lost if only the score for whites is available. These percentages were 20 or higher for 7 of 8 MA measures for API, Hispanics, and blacks, indicating at least moderate informativeness. Similar patterns were evident for free-standing PDPs. All measures were at least moderately informative for one of the three racial/ethnic minority subgroups and all but three measures (getting care quickly, flu shot, and getting needed information about prescription drugs) were informative for all three subgroups. Consequently, none of the seven CAHPS measures was removed for lack of informativeness.

HEDIS Measures. For each HEDIS measure that met its sample size criterion, the bottom half of Table 4 shows the percentage of information about minority group scores that is lost if only the score for whites is available. These percentages were 20 percent or higher for 5 of 7 measures for API and Hispanics; however, none of the measures reached the moderate

informativeness threshold for blacks. All measures were at least moderately informative for one of the three racial/ethnic minority groups except for the measure of retinal eye exam for diabetes patients. Nonetheless, we preserved this measure along with all the others to avoid breaking up the set of four HEDIS diabetes measures.

DISCUSSION

Our study demonstrates that reliability and informativeness are useful criteria for identifying measures that may be used in reporting health care performance data by racial/ethnic subgroups. As the sample size for any specific racial/ethnic group within a plan is only a fraction of the total sample size for all members of a plan, plan scores for specific racial/ethnic groups are often less reliable than are overall plan scores. To some extent, this reduction in sample size can be addressed by pooling plan data over multiple years rather than the typical 1 year.⁵ Still, for racial/ethnic-specific scores to be more useful in selecting plans than overall scores (in fulfillment of the informativeness criterion), they must provide a more accurate guide to which plans are better for which groups than overall plan scores do. In particular, meaningful subgroup reporting requires that there is a gain in group-specific information that outweighs the loss in sample size, precision, and reliability, so that the net result is more accurate information with respect to relative plan experiences for the racial/ethnic group in question.

The minimum sample sizes required to ensure reliability determined which specific measures were reported for each combination of racial/ethnic group, plan, and measure. Applying these standards resulted in many plans having no reportable racial/ethnic data on some or all measures. Nevertheless, a large majority of black, Hispanic, and white beneficiaries are covered by the subset of plans for which group-specific data can be reported, and a smaller majority of API beneficiaries are covered. The somewhat poorer coverage of API beneficiaries is a result of a large proportion of the API population being spread thinly across many Medicare plans. Future work can investigate the possibility of oversampling using administrative measures of race/ethnicity to increase coverage of this group.

As reliable estimates could not be reported for all racial/ethnic groups for many plans, reports that present racial/ethnic data in a format parallel to the one used on the interactive portion of the Medicare site could lead users to focus on the absence of racial/ethnic group data for particular plans or mea-

tures and to draw unwarranted negative conclusions about the meaning of unreported data, as users have been found to do on other websites presenting comparative quality data (Gerteis et al. 2007). If so, reporting racial/ethnic group-specific data on the interactive portion of the Medicare website—where users would inevitably find substantial missing data for plans—would not be helpful.

Instead, we recommend presenting data for all plans nationally in a set of noninteractive tables that users could access on the Medicare website. These tables would provide data only for those plans and measures for which reporting criteria are met for each racial/ethnic group, thus providing access to all the useful information without potentially confusing or frustrating users with many blank cells. Users of the Medicare website who wished to retrieve these data could go to the section of the website with supplemental data and click on a button that brings up an introductory web page with a header, such as “Quality Scores for Specific Racial and Ethnic Groups.” This page would explain that “quality scores are available for different racial and ethnic groups for some measures that are reported for everyone in the Medicare Plan Finder.” It would go on to explain that data are provided for racial/ethnic groups “because there is evidence that quality and patient experiences with care may be different for different groups. These differences may make some plans better than others for some groups of people.” As this process that we envision may require a more thorough exploration of the Medicare website than some people may be inclined to undertake (e.g., elderly beneficiaries; Tanius et al. 2009; Yoon, Cole, and Lee 2009), it may be preferable to target these subgroup reports, at least in part, toward information intermediaries who can bring these reports to the attention of beneficiaries who might otherwise not see them (Shaller, Kanouse, and Schlesinger 2011).

In public reporting of comparative quality data, decisions about which data to present and how to present them should be based not only on what information is available and how reliable and informative it is but also on whether consumers are able to understand and use them to make better decisions. Our efforts demonstrate the feasibility of measuring plan performance by race/ethnicity, but we note that there are potential challenges in communicating these results to a broad range of consumers. Given the potential utility of these reports for beneficiaries and other stakeholders, additional research is needed to identify and refine reporting templates that meet the needs of a variety of audiences. For example, although we recommend reporting between-plan differences to aid consumers in choosing a plan that is best for

people like them (i.e., based on racial/ethnic background), reports of within-plan differences are likely to be more useful for quality improvement, payer oversight, consumer advocacy, and policy making. Additional research should also consider the incentives that reporting racial/ethnic data gives to health plans. One possibility is that plans may seek to increase minority group enrollment to ensure that they are present on the list of health plans shown when a beneficiary looks at racial/ethnic group-specific data. Conversely, some health plans may deemphasize enrollment by racial/ethnic minority group members to avoid public reporting of potentially unflattering quality data.

Two main limitations of our research and recommendations should be noted. First, our study did not consider the interaction between preferred language and race/ethnicity, which can be particularly important for Hispanics. For example, Weech-Maldonado et al. (2003, 2004) found that among Hispanics, language barriers had a larger negative impact on assessment of Medicare managed care than did ethnicity. Future work and larger sample sizes might allow for additional investigation of this issue. Second, we have proposed, excluding doctor communication, an important performance measure (Beck, Daughtridge, and Sloane 2002) in which consumers are especially interested (Sofaer et al. 2005), to balance the cost implications of the larger sample size that this measure would require. These limitations notwithstanding, our study responds to accumulating evidence that patient experience within a given health plan may vary with race/ethnicity (e.g., Lurie et al. 2003; Trivedi et al. 2006) and demonstrates the viability of an approach to quality reporting that has significant potential to enhance consumer decision making and spur quality improvement to reduce racial/ethnic disparities in care.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was funded in full by CMS contract HHSM-500-2005-000281 to RAND. Although prior approval and notification by CMS is not required, CMS was provided with an advanced copy of the manuscript as a courtesy. The authors thank Carol A. Edwards for her assistance with programming and data management.

Disclosures: None.

Disclaimers: None.

NOTES

1. To investigate whether results were sufficiently consistent across years to permit combination, we used mixed linear models to measure disattenuated correlations of contract performance across years; the very high correlations obtained (0.94 to >0.99) supported pooling data across years.
2. The immunization measures are considered to be HEDIS measures. Unlike other HEDIS measures, however, they are collected via beneficiary survey (the CAHPS survey, in particular).
3. Case mix adjustment variables included the beneficiary's age, education, general health status, general mental health status, and the use of a proxy to complete the CAHPS survey.
4. Sample sizes for racial/ethnic subgroups were measured as the sum of the predicted probabilities of membership in a given group among those eligible for a given HEDIS measure.
5. A potential drawback of pooling data over multiple years is that it may create the perception that the information is outdated and not reflective of a plan's current performance. It may be necessary, therefore, to correct such a misperception by pointing out the very high correlation of underlying scores (i.e., with sampling error removed) between consecutive years (see Note 1).

REFERENCES

- Adams, J. L., A. Mehrotra, J. W. Thomas, and E. A. McGlynn. 2010. "Physician Cost Profiling—Reliability and Risk of Misclassification." *New England Journal of Medicine* 362 (11): 1014–21.
- Beck, R. S., R. Daughtridge, and P. D. Sloane. 2002. "Physician-Patient Communication in the Primary Care Office: A Systematic Review." *Journal of the American Board of Family Medicine* 15 (1): 25–38.
- Berwick, D. M., B. James, and M. J. Coye. 2003. "Connections between Quality Measurement and Improvement." *Medical Care* 41 (suppl. 1): I30–8.
- Christianson, J. B., K. M. Volmar, J. Alexander, and D. P. Scanlon. 2010. "A Report Card on Provider Report Cards: Current Status of the Health Care Transparency Movement." *Journal of General Internal Medicine* 25 (11): 1235–41.
- Elliott, M. N., A. M. Haviland, D. E. Kanouse, K. Hambarsoomian, and R. D. Hays. 2009a. "Adjusting for Subgroup Differences in Extreme Response Tendency in Ratings of Health Care: Impact on Disparity Estimates." *Health Services Research* 44 (2 Part 1): 542–61.
- Elliott, M. N., P. A. Morrison, A. Fremont, D. M. McCaffrey, P. Pantoja, and N. Lurie. 2009b. "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities." *Health Services and Outcomes Research Methodology* 9 (2): 69–83.

- 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. Hambarssomian, 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.
- Faber, M., M. Bosch, H. Wollersheim, S. Leatherman, and R. Grol. 2009. "Public Reporting in Health Care: How Do Consumers Use Quality-of-Care Information? A Systematic Review." *Medical Care* 47 (1): 1–8.
- Fiscella, K., P. Franks, M. R. Gold, and C. M. Clancy. 2000. "Inequality in Quality: Addressing Socioeconomic, Racial, and Ethnic Disparities in Health Care." *Journal of the American Medical Association* 282 (19): 2579–84.
- Fung, C. H., Y.-W. Lim, S. Mattke, C. Damberg, and P. G. Shekelle. 2008. "Systematic Review: The Evidence that Publishing Patient Care Performance Data Improves Quality of Care." *Annals of Internal Medicine* 148 (2): 111–23.
- Gerteis, M., J. S. Gerteis, D. Newman, and C. Koepke. 2007. "Testing Consumers' Comprehension of Quality Measures Using Alternative Reporting Formats." *Health Care Financing Review* 28 (3): 31–45.
- Goldstein, E., M. N. Elliott, W. G. Lehrman, K. Hambarsoomian, and L. A. Giordano. 2010. "Racial/Ethnic Differences in Patients' Perceptions of Inpatient Care Using the HCAHPS Survey." *Medical Care Research and Review* 67 (1): 74–92.
- Harris, K. M., and M. B. Buntin. 2008. *Choosing a Health Care Provider: The Role of Quality Information (Research Synthesis Report No. 14)*. Princeton, NJ: The Robert Wood Johnson Foundation.
- Hays, R. D., J. A. Shaul, V. S. L. Williams, J. S. Lubalin, L. D. Harris-Kojetin, S. F. Sweeny, and P. D. Cleary. 1999. "Psychometric Properties of the CAHPS 1.0 Survey Measures." *Medical Care* 37 (3): MS22–31.
- Hibbard, J. H., J. Greene, and D. Daniel. 2010. "What Is Quality Anyway? Performance Reports that Clearly Communicate to Consumers the Meaning of Quality of Care." *Medical Care Research and Review* 67 (3): 275–93.
- Hibbard, J. H., J. Stockard, and M. Tusler. 2003. "Does Publicizing Hospital Performance Stimulate Quality Improvement Efforts?" *Health Affairs* 22 (2): 84–94.
- Keenan, P. S., M. N. Elliott, P. D. Cleary, A. M. Zaslavsky, and B. E. Landon. 2009. "Quality Assessments by Sick and Healthy Beneficiaries in Traditional Medicare and Medicare Managed Care." *Medical Care* 47 (8): 882–8.
- Kolstad, J. T., and M. E. Chernew. 2009. "Consumer Decision Making in the Market for Health Insurance and Health Care Services." *Medical Care Research and Review* 66 (1): 28S–52S.
- Lurie, N., C. Zhan, J. Sangl, A. S. Bierman, and E. S. Sekscenski. 2003. "Variation in Racial and Ethnic Differences in Consumer Assessments of Health Care." *American Journal of Managed Care* 9 (7): 502–9.

- Morales, L. S., M. N. Elliott, R. Weech-Maldonado, K. L. Spritzer, and R. D. Hays. 2001. "Differences in CAHPS Adult Survey Reports and Ratings by Race and Ethnicity: An Analysis of the National CAHPS Benchmarking Data 1.0." *Health Services Research* 36 (3): 595–617.
- National Committee for Quality Assurance. 2011. "What Is HEDIS?" [accessed on November 18, 2011]. Available at <http://www.ncqa.org/tabid/187/Default.aspx>
- 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.
- Roland, M., M. N. Elliott, G. Lyratzopoulos, J. Barbiere, R. A. Parker, P. Smith, P. Bower, and J. Campbell. 2009. "Reliability of Patient Responses in Pay for Performance Schemes: Analysis of National General Practitioner Patient Survey Data in England." *British Medical Journal* 339 (7727): b3851.
- Shaller, D., D. Kanouse, and M. Schlesinger. 2011, March 23. "Meeting Consumers Halfway: Context-Driven Strategies for Engaging Consumers to Use Public Reports on Health Care Providers." Paper commissioned for the AHRQ summit on public reporting.
- Sofaer, S., C. Crofton, E. Goldstein, E. Hoy, and J. Crabb. 2005. "What Do Consumers Want to Know about the Quality of Care in Hospitals?" *Health Services Research* 40 (6): 2018–36.
- Solomon, L. S., A. M. Zaslavsky, B. E. Landon, and P. D. Cleary. 2002. "Variation in Patient-Reported Quality among Health Care Organizations." *Health Care Financing Review* 23 (4): 85–100.
- Tanius, B. E., S. Wood, Y. Hanoch, and T. Rice. 2009. "Aging and Choice: Applications to Medicare Part D." *Judgment and Decision Making* 4 (1): 92–101.
- Trivedi, A. N., A. M. Zaslavsky, E. C. Schneider, and J. Z. Ayanian. 2006. "Relationship between Quality of Care and Racial Disparities in Medicare Health Plans." *Journal of the American Medical Association* 296 (16): 1998–2004.
- Weech-Maldonado, R., L. S. Morales, M. N. Elliott, K. L. 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., M. N. Elliott, L. S. Morales, K. L. 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, A. Oluwole, K. C. Schiller, and R. D. Hays. 2008. "Survey Response Style and Differential Use of CAHPS Rating Scales by Hispanics." *Medical Care* 46 (9): 963–8.
- Weinick, R. M., M. N. Elliott, A. E. Volandes, L. Lopez, Q. Burkhart, and M. Schlesinger. 2011. "Using Standardized Encounters to Understand Reported Racial/Ethnic Disparities in Patient Experiences with Care." *Health Services Research* 46 (2): 491–509.
- Yoon, C., C. A. Cole, and M. P. Lee. 2009. "Consumer Decision Making and Aging: Current Knowledge and Future Directions." *Journal of Consumer Psychology* 19 (1): 2–16.

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.

Table A1: Detail on Thirteen CAHPS Measures Considered for Racial/Ethnic Subgroup Reporting.

Table A2: Detail on Sixteen HEDIS Measures Considered for Racial/Ethnic Subgroup Reporting.

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.