By J. Michael McWilliams, Gabe Weinreb, Mary Beth Landrum, and Michael E. Chernew

DOI: 10.1377/hlthaff.2023.01351
HEALTH AFFAIRS 44,
NO. 1 (2025): 48–57

Use Of Patient Health Survey Data For Risk Adjustment To Limit Distortionary Coding Incentives In Medicare

J. Michael McWilliams (mcwilliams@hcp.med.harvard.edu), Harvard University and Brigham and Women's Hospital, Boston, Massachusetts.

Gabe Weinreb, Harvard University.

Mary Beth Landrum, Harvard University.

Michael E. Chernew, Harvard University.

ABSTRACT A core problem with the current risk-adjustment system in Medicare Advantage and accountable care organization (ACO) programs—the Hierarchical Condition Categories (HCC) model—is that the inputs (coded diagnoses) can be influenced for gain by risk-bearing plans or providers. Using existing survey data on health status (which provide less manipulable inputs), we found that the use of a hybrid risk score drawing from survey data and a scaled-back set of HCCs would, in addition to mitigating coding incentives, modestly lessen risk-selection incentives, strengthen payment incentives to deliver efficient care, allocate payment across ACOs more efficiently according to markers of population health that are not as affected by practice patterns or coding efforts, and redistribute payment in a manner that supports equity goals. Although sampling error and survey nonresponse present challenges, analyses suggest that these should not be prohibitive. Overall, our proof-of-concept analysis suggests that using survey data to improve risk-adjustment performance is a promising strategy that merits further development.

In Medicare Advantage (MA), Medicare pays managed care plans a capitated payment to assume financial responsibility for an enrollee's covered medical expenses. To set payment prospectively at the expected level of spending for an enrollee, the Centers for Medicare and Medicaid Services (CMS) uses the Hierarchical Condition Categories (HCC) risk-adjustment model. Annual fee-for-service spending per beneficiary in traditional Medicare is predicted as a function of diagnosis codes recorded in medical claims (and other demographic and enrollment variables); the resulting coefficients, or weights, are then applied to an enrollee's historical diagnoses recorded in MA encounter data to calculate an HCC risk score, which in turn determines plan payment for the enrollee.

This system creates strong incentives for MA plans to encourage coding of more diagnoses, and it is well documented that HCC scores are meaningfully influenced by variation in coding practices.[1] Only some of this variation is due to fraudulent coding; much of it reflects variation in the thoroughness of coding or interpretation of clinically ambiguous diagnoses. Nevertheless, the incentive to code intensively has three major consequences in MA. First, more aggressive coding increases MA payments above traditional Medicare levels of spending. CMS reduces the HCC-based plan payment rates in MA by 5.9 percent to account for coding differences between the programs, but the difference is much larger.[2,3] The large subsidies to plans that result from this difference are shared with enrollees in the form of lower premiums and more generous benefits, but only partially. Plans have limited incentive to pass the subsidies along to enrollees in full because they face limited competition,[4] including from traditional Medicare. As coverage in

MA becomes more generous, the relative appeal of traditional Medicare diminishes, rendering it less able to apply competitive pressure on MA plans to share further increases in subsidies from coding with enrollees.[5]

Second, efforts to code consume resources without producing clear benefits for enrollees. Thus, even if CMS fully adjusted payment for the coding difference between MA and traditional Medicare, the incentive to code would still divert resources to socially wasteful activities. Third, allocating payment as a function of coding practices instead of population health care needs is inefficient and may exacerbate health disparities. Underserved populations may have fewer coded diagnoses despite similar or worse health status, either because access barriers constrain their use of care (limiting opportunities for diagnoses to be recorded) or because their providers have fewer resources to devote to coding efforts.[6–8]

For these reasons, reforming the HCC model to limit coding incentives has become a major focus of Medicare payment policy. Approaches to do so, however, must navigate a potential trade-off between mitigating incentives to code and exacerbating incentives for plans to engage in favorable risk selection. For example, paring back the number of diagnoses included in the model or lengthening look-back periods for gathering diagnoses may limit the gains from coding more diagnoses (or coding the same diagnoses more persistently). However, these measures may also erode the predictive accuracy of the HCC model, affording plans more opportunity to profit from attracting enrollees with overpredicted spending or avoiding those with underpredicted spending.

This trade-off is on display in attempts by Medicare accountable care organization (ACO) models, including the Medicare Shared Savings Program and ACO REACH (Realizing Equity, Access, and Community Health) model, to address the coding issue by capping risk-score growth at the ACO level. This approach limits rewards for increased coding for a given population, but it penalizes ACOs whose true population health risk rises faster than the cap, discouraging ACOs from taking on patients with high HCC scores. In addition, the approach still allows ACOs to gain from coding more intensively if they simultaneously shift their populations toward patients with lower HCC scores (to create more room for coding under the cap). Although these incentives might not elicit widespread risk selection by ACOs at the patient level,[9] a cap on risk-score growth, at the very least, entrenches payment differences due to baseline differences in coding practices, putting less aggressive coders at a permanent disadvantage and perpetuating an inefficient allocation of resources.

Because the core problem is that the inputs of the HCC model (predictors of spending) can be influenced by plans or providers, the optimal approach to limiting rewards for coding would be to use predictors that cannot be as easily altered. This is challenging because of the paucity of detailed information on enrollee health that is consistently collected and not subject to the influence of self-interested actors. Some such data are available, however, and could be augmented. In this study, we explored the potential for using existing survey data on health status to mitigate coding incentives while minimizing the exacerbation of risk-selection incentives.

Study Data And Methods

CONCEPTUAL OVERVIEW Although survey data may be available for only a sample of a population, they may be used in concert with other information to formulate a population-level risk score for an organization (plan or ACO) that cannot be as readily manipulated by the organization. If such a survey-based risk score encompassed as much information as the HCC score about a person's expected health care costs resulting from health status and care needs, replacing the organization's mean HCC score with its mean survey-based score as the basis for risk-adjusting population-based payment could minimize coding incentives without worsening risk-selection incentives, on average. If a survey-based score did not capture as much of this information, its use might still have value, but there would be a trade-off between coding and selection incentives; the predictive accuracy (or fit) of risk-adjusted payments would degrade as increasing weight is placed on the survey-based score.

Because diagnoses used to compute HCC scores are gathered from medical claims, more service use tends to increase HCC scores (by supplying coded diagnoses), and efforts to code diagnoses may involve more service use (for example, encouraging the use of more evaluation and management services to record more diagnoses). This bidirectional relationship between spending and HCC scores has three implications. First, substituting survey assessments of health for HCCs may also improve the power of incentives encouraging care efficiency in a population-based payment system.[10] Second, one may expect some loss of fit even if substituting survey data for claims-based diagnoses does not have an undesirable impact on risk-selection incentives; the HCC score may better fit spending data, but this does not necessarily mean that it is a better risk adjuster. Third, judging the relative performance of a risk-adjustment system that relies more on survey data as inputs and less on HCCs requires measures of disease burden that are influenced less than total spending by practice patterns or coding efforts. These could include health outcomes (for example, mortality), service use that is less discretionary (for example, hospitalizations), and service use that should affect HCC scores less and be less affected by coding efforts (for example, medication use). In the absence of an observable gold-standard basis for distributing payment, a range of measures can be used to compare the allocative efficiency of different risk scores.

OVERVIEW OF ANALYSIS We explored the performance of three alternative risk scores for Medicare beneficiaries. First, as in prior work,[11] we formulated a risk score that drew from Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey assessments of health status, functional status, and chronic conditions, in addition to demographic and enrollment variables from the Medicare Beneficiary Summary File that are included in the HCC score. This CAHPS-based score reflects a substitution of survey information for all HCCs. Second, we created a hybrid CAHPS-HCC score by combining the inputs for the CAHPS-based score with a scaled-back set of HCCs. Specifically, we excluded ten condition groups that have been found to account for much of the difference in (version 28) HCC scores between MA and traditional Medicare;[12] all other conditions were included as inputs (see online appendix exhibit 1).[13] We did this as a proof of concept—a useful exercise to understand the consequences of substituting CAHPS data for a targeted set of HCCs. We do not contend that the resulting hybrid score is optimal, and we did not explore other combinations of inputs. Third, we constructed a risk score from the demographic and enrollment variables included in the HCC score and the subset of HCCs included in the hybrid score, omitting the CAHPS inputs. We refer to this as the HCC subset score.

Although our study was motivated by applications to MA, we focused on respondents to the Medicare Fee-for-Service (FFS) CAHPS for whom we could construct these risk scores, and we compared their properties with those of the HCC score. The FFS CAHPS also includes more health items than the MA CAHPS.[14] We further focused on respondents served by ACOs in the Medicare Shared Savings Program, which present policy-relevant units of analysis. Given that the Medicare Shared Savings Program and MA face similar challenges in risk adjustment, findings from our ACO-focused analysis (necessitated by data constraints) should inform consideration of using survey data for risk adjustment in MA.

Our main analyses considered the implications of using an ACO's mean CAHPS-based or hybrid risk score in lieu of its mean HCC score as the basis for risk-adjusting population-based payment. First, to understand the potential effects of using the alternative risk scores on selection incentives at the beneficiary level, we assessed the fit of each risk score to actual spending at the beneficiary level.

Second, to understand the effects of using the alternative risk scores on how payment is allocated across ACOs according to population health status and care needs, we compared the means of the different risk scores within quintiles of ACOs sorted by their mean value of an outcome, using a range of outcomes that varied in the extent to which they were influenced by practice patterns or coding practices.

Results of ACO-level analyses assessing allocative efficiency may differ from those of beneficiary-level analyses assessing selection incentives because the drivers of spending variation may differ at the organization and beneficiary levels. For example, a survey-based risk score may worsen undercompensation for high-cost beneficiaries when compared with the HCC score, creating incentives for ACOs to avoid beneficiaries with HCC scores that are higher than their survey-based scores, but this could nevertheless result in a more efficient allocation of resources across ACOs. This would be the case if variation across ACOs in spending and HCC scores were driven primarily by variation in overuse, not population health, or if ACO spending variation were driven by dimensions of population health better captured by the survey-based score. The relative performance of the survey-based score would then depend on the extent to which its use would elicit risk selection. Greater undercompensation or overcompensation at the beneficiary level might or might not translate directly into greater risk selection, depending on the costs of selection strategies and the extent to which organizations are constrained in the patients they can attract or avoid.

Third, we conducted additional analyses to explore two key challenges in the use of survey data for risk adjustment: sampling error and nonresponse bias. Fourth, we described the distributional consequences of using the CAHPS-based or hybrid risk score in lieu of the HCC score—specifically, how the resulting payment redistribution would affect various sociodemographic groups.

STUDY POPULATION Our study population included Medicare beneficiaries who responded to the 2010–19 annual cross-sectional FFS CAHPS surveys, were enrolled in Medicare in the year preceding the survey year (allowing the derivation of HCC scores), and were not enrolled in MA in either year. Using linked claims data for the year preceding the survey and Medicare Shared Savings Program Provider-level files (identifying each ACO's constituent providers), we identified ACO-assigned respondents as those for whom an ACO accounted for more of the respondent's office visits with a primary care physician than any other ACO or non-ACO provider. We used the year preceding the survey for ACO assignment to align assignment with the provider coding practices contributing to the respondent's HCC score.

We limited our analysis to respondents assigned to ninety-five ACOs with at least 600 assigned respondents per ACO (pooled over the period 2010–19) to approximate sample sizes that would be generated from two or more years of the ACO CAHPS survey, which is administered separately from the FFS CAHPS survey to a sample of 860 beneficiaries per ACO, with an average response rate of approximately 40 percent.[11] We did not have access to ACO CAHPS data for this study.

Study Variables

▶ **Construction of risk scores:** To calculate each of the four risk scores for each survey respondent in each year, we first used data on all respondents to the FFS CAHPS survey (not limited to ACO-assigned respondents) in a baseline year ($N = 85{,}395$) to fit linear regression models of total (Parts A and B) Medicare spending. To construct the HCC score, we modeled spending in the baseline year as a function of the following variables obtained from linked Medicare Beneficiary Summary File and claims data in the prior year: age, sex, disability, dual eligibility for Medicaid, and all HCC condition indicators. To construct the HCC subset score, we omitted the HCCs described in appendix exhibit 1.[13]

To calculate the CAHPS-based risk score for each respondent, we modeled spending in the baseline year for respondents as a function of the same set of demographic and enrollment variables included in the HCC model and the following self-reported health variables from the baseline year survey (in lieu of HCCs): general health and mental health status, difficulty with six activities of daily living (ADLs), and five self-reported diagnoses. We also included interactions between predictors. To calculate the hybrid score, we added to this model the HCCs included in the HCC subset score. We then applied the coefficients from these models to derive each of the four risk scores for each ACO-assigned respondent in each survey year.

For each score, we normalized predictions to a common year (2019) to control for secular trends in predictors, and we derived a relative value with mean of 1.0 by dividing each respondent's prediction by the mean among all ACO-assigned respondents (see the methods appendix for more details on risk-score derivation).[13]

▶ **Outcome variables:** We examined five measures. First, we assessed mortality from the Medicare Beneficiary Summary File (through the year after the survey because CAHPS respondents are all alive when surveyed). Second, among Part D enrollees, we assessed the number of prescription drug fills during the survey year as a proxy for the number and severity of clinical conditions. Third, we assessed the number of hospitalizations during the survey year. Fourth and fifth, we assessed total annual evaluation and management services and total Parts A and B Medicare spending for the respondent during the survey year. Relative to the other measures, we expected these two metrics to be less specific to health status and health care needs and more influenced by practice patterns and coding efforts, but also more sensitive to condition prevalence and associated costs potentially missed by the CAHPS-based and hybrid scores. Outcomes were also normalized to 2019 to account for secular trends.

Analysis

▶ **Main analyses:** Within each decile of beneficiaries sorted by annual per beneficiary Medicare spending, we calculated the mean difference between predicted and actual spending, alternately using each of the four risk scores to determine predicted spending. Among ACOs in the highest (or lowest) quintile for a given outcome, we computed the mean beneficiary-weighted risk score for each of the four risk scores. We did this separately for each outcome. To assess the information contained in each risk score, we regressed each of the five outcomes separately on each risk score at the beneficiary and ACO levels, and we report the proportion of variance explained by each score. All analyses were weighted to account for observable patterns of survey nonresponse (see "Survey Non-response" in the appendix).[13]

▶ **Nonresponse bias and sampling error:** Use of the CAHPS-based or hybrid risk score might proceed as follows. First, CMS could calculate a score ratio (HCC/CAHPS-based or HCC/hybrid) for each ACO based on its population of survey respondents in the prior one to three years. This ratio would describe how the HCC score related to the alternative score for the same sample of beneficiaries. To estimate what an ACO's CAHPS-based or hybrid score would be for its full population (which is unobserved), CMS could then divide the ACO's HCC score for its full population (which is known) by this ratio. The ratio thus would serve as a correction factor: If an ACO's HCC score for its surveyed population exceeded what was expected based on the alternative score, its full population HCC score would be reduced accordingly.

The validity of this approach rests on the assumption that the relative difference in scores observed for an ACO's survey respondents would be similar for its nonrespondents. We probed this assumption by comparing ACO risk scores and score ratios with versus without weighting for nonresponse. We found, for example, that nonresponse weighting corrected some underestimation of both CAHPS-based and HCC scores for ACOs with higher scores, consistent with lower response rates among beneficiaries in poorer health. However, weighting for nonresponse affected ACOs' HCC/CAHPS-based score ratios minimally (appendix exhibit 2).[13] This suggests that a biased sample of survey respondents may yield an unbiased estimate of an organization's score ratio because estimates of the numerator and denominator scores were proportionally biased.

Even if this approach supports an unbiased estimate of an alternative risk score for an ACO's full population, sampling error reduces the reliability of that estimate. Although random error in payment would average out over time, it may nevertheless be disruptive in the short term and might substantially disadvantage some unlucky ACOs over the longer term. Thus, we quantified the sampling error with which the score ratios would be estimated for ACOs based on different sample sizes (see "Sampling Error" in the appendix).[13]

DISTRIBUTIONAL IMPLICATIONS To assess the distributional implications of employing the CAHPS-based or hybrid score in lieu of the HCC score in risk-adjusting ACO payments (benchmarks), we sorted ACOs into quintiles on the basis of their score ratios, and we compared population characteristics between ACOs in the top versus bottom quintiles.

LIMITATIONS Our study had several limitations. First, the results of our analysis of ninety-five ACOs might not generalize to the full Medicare Shared Savings Program or other ACO models, and our results may also differ from those that would be observed for MA plans, where coding incentives are stronger and variation in HCC scores wider.[4] Nevertheless, the approach we implemented for ACOs could be applied in the MA setting to explore the use of survey data to limit coding incentives while preserving or improving the performance of risk adjustment on other dimensions. Existing sources of survey data in MA include MA CAHPS and the Medicare Health Outcomes Survey, which includes a more detailed assessment of health and functional status than CAHPS.[15]

Partially substituting survey data can mitigate distortionary coding incentives while improving other aspects of risk adjustment.

Second, nonresponse bias may distort risk adjustment that relies on survey data. We found supportive evidence for an approach that would leverage the quantifiable impact of nonresponse bias on HCC scores, which are observable for both the full population and survey samples, to correct for nonresponse bias when deriving alternative risk scores from a survey sample. However, it is possible that nonresponse bias may affect different risk scores differently in ways that are challenging to adjust for. Policy makers could help address this potential limitation by devoting resources to encouraging survey response. The average rate of response to the 2010–19 FFS CAHPS surveys was 48.8 percent among beneficiaries eligible for ACO assignment.

Third, our analysis of FFS CAHPS data had to pool across many years to obtain sample sizes commensurate with those generated by the annual ACO or MA CAHPS surveys. It would not be desirable to pool more than a few years in a policy application.

Finally, our study was premised on the assumption that survey data collected by an independent party would be less manipulable by providers or plans than diagnoses recorded in medical claims. This is a reasonable assumption, as an ACO or plan would not know ex ante which beneficiaries to target for coaching (the survey sample would be unknown), and responses to some items (for example, difficulties with ADLs) would seem hard to influence. Also, gains from strategically encouraging sicker beneficiaries to respond to surveys should be limited by the correction approach described above and by strengthening efforts to limit nonresponse.

Moreover, efforts to influence survey data could be prohibited and penalized more cleanly because the data, unlike HCCs, would not be part of claims processing. However, we could not directly assess the relative manipulability of survey data.

Study Results

BENEFICIARY-LEVEL ANALYSES: SELECTION INCENTIVES The study sample included 95,559 FFS CAHPS respondents. Relative to the HCC score, the CAHPS-based score and HCC subset score exhibited greater undercompensation in higher deciles of actual Medicare spending at the beneficiary level and greater overcompensation in lower deciles, whereas the hybrid score reduced both undercompensation in higher deciles and overcompensation in lower deciles (exhibit 1). This suggests that substituting CAHPS data on health for a set of variably coded HCCs would modestly diminish risk-selection incentives.

ACO-LEVEL ANALYSES: ALLOCATIVE EFFICIENCY The analysis included ninety-five ACOs with a mean of 1,005 respondents per ACO (range, 601–2,392 respondents; median, 868 respondents). The mean hybrid score was similar to the mean HCC score among ACOs in the highest quintile of total per beneficiary spending (exhibit 2) and slightly lower than the mean HCC score among ACOs in the lowest quintile of spending (exhibit 3). This suggests that use of the hybrid score instead of the HCC score would result in similar undercompensation for ACOs with high spending and slightly less overcompensation for ACOs with low spending, or slightly improved allocative efficiency overall if actual spending were the desired basis for allocating payment. Of course, it is not. A prospective payment system should neither reward higher spending resulting from greater low-value care provision nor penalize lower spending resulting from more efficient care. To better understand the allocative efficiency of a hybrid score, we therefore turn to alternative markers of population disease burden.

Among ACOs in the top quintile of mortality rates, prescription drug fills, or hospitalization rates—outcomes that should reflect practice patterns and coding efforts less than total spending—the mean hybrid score was consistently higher than the mean HCC score (as was the mean CAHPS-based score for two of these outcomes). In contrast, the hybrid score was lower than the HCC score among ACOs in the top quintile of evaluation and management service use, which may be most influenced by discretionary service use and coding efforts (exhibit 2). For each outcome, the expected reverse pattern was observed for ACOs in the lowest quintile (exhibit 3).

Likewise, ACO-level regressions of outcomes on each risk score revealed that ACOs' mean hybrid scores were more strongly predictive of population mortality, prescription drug use, and hospitalization rates than mean HCC scores, but less strongly predictive of evaluation and management service use (appendix exhibit 3).[13] Taken together, these findings suggest that use of the hybrid score in lieu of the HCC score would vary payment across ACOs to a greater extent based on population health care needs and to a lesser extent based on practice patterns, con-

EXHIBIT 1

Comparison of risk-score fit to total Medicare spending at the beneficiary level: mean predicted versus actual spending for each risk score, by decile of actual spending

	Decile of actual spending									
	1	2	3	4	5	6	7	8	9	10
Actual spending ($)	328	834	1,353	1,987	2,834	4,039	5,919	9,801	19,770	60,676
Predicted spending ($)										
HCC score	6,440	7,165	7,910	8,496	9,268	9,860	10,985	12,482	14,155	20,780
CAHPS-based score	8,339	8,635	9,004	9,324	9,850	10,394	11,092	12,215	13,205	15,484
HCC subset score	6,968	7,564	8,209	8,733	9,321	9,868	10,842	12,162	13,677	20,199
Hybrid score	6,365	6,993	7,713	8,250	9,005	9,713	10,931	12,579	14,362	21,633
Predicted less actual spending ($)										
HCC score	6,112	6,331	6,558	6,509	6,434	5,822	5,066	2,682	−5,614	−39,896
CAHPS-based score	8,010	7,801	7,651	7,337	7,017	6,356	5,173	2,415	−6,564	−45,192
HCC subset score	6,640	6,730	6,856	6,745	6,487	5,829	4,923	2,361	−6,092	−40,477
Hybrid score	6,037	6,159	6,360	6,262	6,171	5,674	5,012	2,778	−5,408	−39,044

SOURCE Authors' analysis of 2010–19 Fee-for-Service Consumer Assessment of Healthcare Providers and Systems (FFS CAHPS) survey data and Medicare enrollment and claims data for survey respondents. **NOTES** Estimates are adjusted for year and weighted for nonresponse as described in the study methods. The various scores are defined in the text. "Total spending" is per beneficiary spending on Medicare Parts A and B. HCC is Hierarchical Condition Categories.

EXHIBIT 2

Comparison of risk-score allocative impact at the accountable care organization (ACO) level: mean of each risk score for ACOs in the highest quintile of selected markers of population disease burden

SOURCE Authors' analysis of 2010–19 Fee-for-Service Consumer Assessment of Healthcare Providers and Systems (FFS CAHPS) survey data and Medicare enrollment and claims data for survey respondents. **NOTES** Mean ACO risk scores are adjusted for year and nonresponse and averaged across ACOs within a quintile using the ACO-assigned population size as weights. "Total spending" is per beneficiary spending on Medicare Parts A and B. We omitted survey year 2019 from analyses of mortality because of COVID-19-related deaths in 2020.

sistent with modest potential improvements in allocative efficiency.

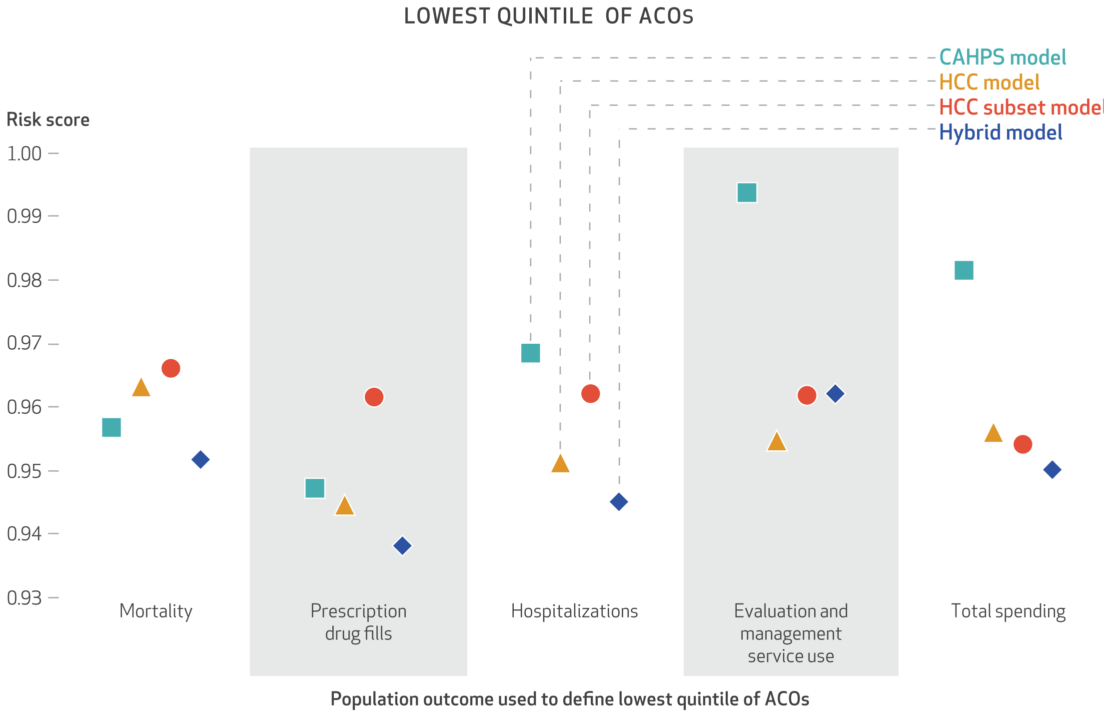
DISTRIBUTIONAL IMPLICATIONS Mean risk scores varied considerably across ACOs (tenth to ninetieth percentile spread: 0.92–1.07 for HCC and hybrid scores and 0.94–1.06 for CAHPS-based scores), as did the score ratios described in exhibit 4. ACOs with high HCC/CAHPS-based or HCC/hybrid score ratios tended to have higher HCC scores, on average (exhibit 4), but spanned a wide range of HCC scores that overlapped with ACOs with low score ratios (appendix exhibits 4 and 5).[13] This would be expected, for example, from more intensive coding practices by ACOs with a wide range of true population illness burden.

Population characteristics were well balanced across ACOs in higher versus lower quintiles of score ratios, with some notable exceptions (exhibit 4). ACOs with higher HCC/CAHPS score ratios served populations in more urban counties with lower rates of disability as the original reason for Medicare enrollment; therefore, use of the CAHPS-based score would redistribute some payment toward ACOs in less urban counties that have higher shares of beneficiaries with disabilities, on average. Likewise, use of the hybrid score would redistribute payment toward ACOs with higher shares of Black beneficiaries (and non-White beneficiaries more broadly) and would not meaningfully redistribute payment across any of the other sociodemographic groups we examined.

SAMPLING ERROR Our assessment of sampling error suggests that three years of currently available CAHPS data would support estimation of an ACO's HCC/CAHPS-based score ratio that was within 3.8 percent of its true score ratio, and estimation of an ACO's HCC/hybrid score ratio that was within 1.6 percent of its true score ratio, 90 percent of the time (appendix exhibits 6–9).[13] Larger sample sizes would result in less sampling error.

EXHIBIT 3

Comparison of risk-score allocative impact at the accountable care organization (ACO) level: mean of each risk score for ACOs in the lowest quintile of selected markers of population disease burden



SOURCE Authors' analysis of 2010–19 Fee-for-Service Consumer Assessment of Healthcare Providers and Systems (FFS CAHPS) survey data and Medicare enrollment and claims data for survey respondents. **NOTES** Mean ACO risk scores are adjusted for year and nonresponse and averaged across ACOs within a quintile, using the ACO-assigned population size as weights. "Total spending" is per beneficiary spending on Medicare Parts A and B. We omitted survey year 2019 from analyses of mortality because of COVID-19-related deaths in 2020.

Discussion

Collectively, our findings suggest that partially substituting survey data on health status for claims-based diagnoses as the basis for risk adjustment of population-based payments can mitigate distortionary incentives to code diagnoses while also helping mitigate risk-selection incentives, strengthen the power of prospective payment incentives, improve allocative efficiency, and support equity goals.

Specifically, we found that a hybrid risk score drawing from survey data and a scaled-back set of HCCs would slightly lessen under- or overcompensation for Medicare beneficiaries with high or low spending, respectively, and would align population-based payments to ACOs more closely with markers of population disease burden that are influenced less than total medical spending by organizations' practice patterns or coding efforts. At an ACO level, the hybrid score was a stronger predictor of population mortality rates, hospitalization rates, and prescription drug use than the HCC score, but a weaker predictor of evaluation and management service use. These findings suggest that organizational variation in the provision of discretionary services, including services induced by efforts to gather diagnoses, contributes to organizational variation in spending and HCC scores. Accordingly, a risk-adjustment system relying less exclusively on claims-based diagnoses as inputs and instead incorporating inputs from survey data may better preserve a population-based payment system's rewards for efficient care by lessening the problematic endogeneity of HCC scores. Our findings therefore highlight potential gains from using survey data to improve the allocative efficiency of population-based payments and discourage organizations from providing unnecessary care or devoting resources to coding.

We explored the properties of a specific hybrid risk score as a proof of concept. The gains from this approach depend on the selection of inputs, which could be improved substantially in two ways. First, a deeper understanding of which diagnoses are more prone to coding variation that departs from variation in true disease prevalence could guide a refined selection of conditions to include in a hybrid score. Second,

EXHIBIT 4

Distributional implications of using a Consumer Assessment of Healthcare Providers and Systems (CAHPS)-based or hybrid risk score: population characteristics of accountable care organizations (ACOs) with high versus low Hierarchical Condition Categories (HCC) scores relative to their CAHPS-based or hybrid scores

	Quintile of HCC/CAHPS-based score ratio			Quintile of HCC/hybrid score ratio		
ACO population characteristics	First	Fifth	Fifth minus first	First	Fifth	Fifth minus first
Relevant score ratio	0.94	1.07	0.14****	0.97	1.04	0.07****
HCC score	0.96	1.05	0.09****	1.00	1.01	0.01
CAHPS-based score	1.02	0.98	−0.04***	1.04	0.97	−0.07****
Hybrid score	0.98	1.04	0.05***	1.03	0.97	−0.06***
Age, years	74.4	74.9	0.5*	74.6	74.7	0.1
Female, %	58.3	57.2	−1.2*	58.9	57.3	−1.5*
Race and ethnicity,[a] %						
American Indian or Alaska Native	0.3	0.1	−0.2**	0.3	0.2	−0.1
Asian or Pacific Islander	1.1	1.9	0.7	1.8	1.5	−0.3
Black	9.1	5.7	−3.4	11.8	2.9	−9.0***
Hispanic	3.1	4.6	1.5	4.6	2.9	−1.6
White	84.9	86.0	1.2	79.9	90.7	10.8***
Dually eligible for Medicaid, %	9.6	8.8	−0.8	10.9	10.3	−0.5
Disability as original reason for Medicare enrollment, %	18.1	15.0	−3.0**	17.5	16.8	−0.6
High school graduate,[b] %	89.2	91.2	1.9**	89.5	90.8	1.4
Urban county,[c] %	78.0	88.7	10.7**	85.3	83.9	−1.4

SOURCE Authors' analysis of 2010–19 Fee-for-Service (FFS) CAHPS survey data and Medicare enrollment and claims data for survey respondents. NOTES "Relevant score ratio" refers to the score ratio by which ACOs are sorted into quintiles. Age, sex, race and ethnicity, dual eligibility for Medicaid (including eligibility for full and partial Medicaid benefits), and disability were ascertained from Medicare Beneficiary Summary Files. The 2010–11 FFS CAHPS surveys excluded Part D enrollees (and thus all dually eligible beneficiaries in those years), lowering the proportion of dually eligible beneficiaries in the study sample. Estimates are adjusted for year and weighted for nonresponse as described in the study methods. [a]Based on Research Triangle Institute Race Codes. Racial groups are non-Hispanic. [b]Educational attainment was ascertained from the FFS CAHPS survey data. [c]Counties designated as metropolitan by Rural-Urban Continuum Codes. *$p < 0.10$ **$p < 0.05$ ***$p < 0.01$ ****$p < 0.001$

CAHPS surveys collect fairly limited information about health. More detailed survey data could be used, whether drawn from existing surveys (for example, the more in-depth Health Outcomes Survey for MA) or newly collected by augmenting current surveys. With additional resources, enhanced data collection might also include standardized collection of diagnoses (and perhaps physical exam or laboratory data). The costs of such efforts need be incurred only for a sample of beneficiaries, whereas current payment policy incentivizes MA plans and ACOs to devote resources to gathering diagnoses for all enrollees (for example, via home-based health risk assessments). Complementary reinsurance approaches could further mitigate coding and selection incentives.[16]

In terms of the distributional consequences of leveraging survey data for risk adjustment, use of the hybrid score we investigated would redistribute payment toward ACOs serving higher shares of historically marginalized racial and ethnic groups. This finding is consistent with access barriers limiting health care use by these groups, or less-intensive diagnosis coding by their providers conditional on use, resulting in HCC scores that are low for their health status.[6,8,17] Thus, the use of survey-based risk assessments might limit undesirable effects of coding variation on health care disparities and help align payment more closely with desirable levels of spending for underserved populations.

Although the distributional effects are important to understand, they should not be a decisive factor in determining whether to pursue the use of survey data to mitigate coding incentives, as targeted payment redistributions can be implemented separately to support equity goals.[6,18] Likewise, although some plans or providers may appropriate subsidies via coding to finance enhanced care for underserved groups, this is not a sound justification for perpetuating rewards for coding.

Finally, the use of survey data to improve risk adjustment would have to address challenges related to sampling error and nonresponse, but our findings suggest that these challenges may be readily addressable. Current collection of CAHPS data for samples of ACO populations is sufficient to estimate a hybrid score with reasonable sampling error. Future research could assess the reliability of hybrid scores for MA plans supported by the Health Outcomes Survey. Moreover, data collection could be augmented and steps taken to reverse declining response rates.[19] As the data infrastructure is being more fully developed, CMS could blend an ACO's or plan's HCC and hybrid risk scores in an initial phase of implementation, with lesser weight given to the survey-informed hybrid score for more sparsely sampled plans or ACOs.

Conclusion

Our study suggests substantial promise in using survey data to improve the performance of risk adjustment in Medicare's population-based payment systems. To inform next steps, future analysis could apply the methods we have laid out here to available data in MA and improve on the hybrid model we explored. ■

This research was presented at the Risk Adjustment Network Conference in the Hague, the Netherlands, November 3, 2023. This work was supported by grants from the Commonwealth Fund (Grant No. 20223746) and the National Institute on Aging, National Institutes of Health (Grant No. P01 AG032952). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the Commonwealth Fund, the National Institutes of Health, or any organization with which the authors are affiliated. To access the authors' disclosures, click on the Details tab of the article online.

NOTES

1 Geruso M, Layton T. Upcoding: evidence from Medicare on squishy risk adjustment. J Polit Econ. 2020; 12(3):984–1026.
2 Kronick R, Chua FM. Industry-wide and sponsor-specific estimates of Medicare Advantage coding intensity. Social Science Research Network [preprint on the Internet]. 2021 Nov 11 [cited 2024 Oct 30]. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3959446
3 Hammond S, Johnson A, Serna L. The Medicare Advantage program: status report [Internet]. Washington (DC): Medicare Payment Advisory Commission; 2024 Jan 12 [cited 2024 Oct 30]. Available from: https://www.medpac.gov/wp-content/uploads/2023/10/MedPAC-MA-status-report-Jan-2024.pdf
4 Medicare Payment Advisory Commission. Report to the Congress: Medicare payment policy [Internet]. Washington (DC): MedPAC; 2023 Mar. Chapter 11, The Medicare Advantage program: status report; [cited 2024 Oct 30]. Available from: https://www.medpac.gov/wp-content/uploads/2023/03/Ch11_Mar23_MedPAC_Report_To_Congress_SEC.pdf
5 McWilliams JM. The future of Medicare and the role of traditional Medicare as competitor. N Engl J Med. 2024;391(8):763–9.
6 McWilliams JM, Weinreb G, Ding L, Ndumele CD, Wallace J. Risk adjustment and promoting health equity in population-based payment: concepts and evidence. Health Aff (Millwood). 2023;42(1):105–14.
7 Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53.
8 Office of Attorney General Maura Healey. Examination of health care cost trends and cost drivers [Internet]. Boston (MA): The Office; 2022 [cited 2024 Oct 30]. Available from: https://www.mass.gov/files/documents/2022/11/02/2022-11-2%20COST-TRENDS-REPORT_PUB_DRAFT4_HQ.pdf
9 McWilliams JM, Hatfield LA, Landon BE, Chernew ME. Savings or selection? Initial spending reductions in the Medicare Shared Savings Program and considerations for reform. Milbank Q. 2020;98(3):847–907.
10 Geruso M, McGuire TG. Tradeoffs in the design of health plan payment systems: fit, power, and balance. J Health Econ. 2016;47:1–19.
11 Chernew ME, Carichner J, Impreso J, McWilliams JM, McGuire TG, Alam S, et al. Coding-driven changes in measured risk in accountable care organizations. Health Aff (Millwood). 2021;40(12):1909–17.
12 Kronick R, Chua F, Krauss R, Johnson L, Waldo D. A proposal to improve the equity and the efficiency of the Medicare Advantage risk payment system. Health Aff (Millwood). 2025;44(1):66–74.
13 To access the appendix, click on the Details tab of the article online.
14 Centers for Medicare and Medicaid Services. Fee-for-Service (FFS) CAHPS [Internet]. Baltimore (MD): CMS; [last updated 2024 Sep 10; cited 2024 Oct 30]. Available from: https://www.cms.gov/data-research/research/consumer-assessment-healthcare-providers-systems/fee-service-cahps
15 National Committee for Quality Assurance. HEDIS Medicare Health Outcomes Survey [Internet]. Washington (DC): NCQA; c 2024 [cited 2024 Oct 30]. Available from: https://www.ncqa.org/hedis/measures/hos/
16 McGuire TG, Schillo S, van Kleef RC. Reinsurance, repayments, and risk adjustment in individual health insurance: Germany, the Netherlands, and the US Marketplaces. Am J Health Econ. 2020;6(1):139–68.
17 Wallace J, Lollo A, Duchowny KA, Lavallee M, Ndumele CD. Disparities in health care spending and utilization among Black and White Medicaid enrollees. JAMA Health Forum. 2022;3(6):e221398.
18 Center for Medicare and Medicaid Innovation. ACO REACH [Internet]. Baltimore (MD): Centers for Medicare and Medicaid Services; [cited 2024 Oct 30]. Available from: https://www.cms.gov/priorities/innovation/innovation-models/aco-reach
19 Centers for Medicare and Medicaid Services. Medicare HOS survey status information [Internet]. Baltimore (MD): CMS; 2023 Aug [cited 2024 Oct 30]. Available from: https://www.hosonline.org/globalassets/hos-online/survey-results/mhos_survey_status_information_c23.pdf