Supplementary Material

Quality Disclosure and Regulation: Scoring Design in Medicare Advantage

Benjamin Vatter — April 30, 2024

I Data construction

This work combines two data sources: public data from CMS and individual-level data from the Medicare Current Beneficiary Survey (MCBS). Here I briefly describe the cleaning steps used when constructing the data used for estimation. The data and procedure used for constructing the quality ratings are discussed in the next section.

I.A Public Data

I construct a plan-county-year-level panel by combining several publicly available data sets. First, I use public *Enrollment* files to recover the average number of enrollees per plan-county-year. Plans with fewer than 11 enrollees have a missing enrollment number in this data. When needed for specific computations, I replace this missing value with 6. These plans are small and do not affect major computations or estimations. I use the *Contract* definition files to recover parent organization names. This requires extensive data cleaning as names are often misspelled or change across years. I cross-reference these with the public *Landscape* files containing organization names. I also use these files to construct the panel's core, as the landscape files list plans by county with their premiums, deductibles, and contract numbers. I drop a small fraction of plans that disappear between September of the previous year and March of the current. These are likely to be plans removed by CMS due to regulation issues or plans that failed to enroll any consumer. I also remove part-D only plans. I combine the landscape file with the *Plans* files to incorporate additional details, including the star ratings. I use the *Penetration* files to add information regarding the population of each county, including the number of Medicare eligibles and the total number enrolling in MA. I also add information from the *Dual Eligibles* files to count the number of dual Medicare/Medicaid eligible populations per county-year. I use the *Payment* files to add information about total payments to each plan from CMS in terms of subsidies and rebates. I derive the local benchmarks for each plan using the *Ratebook* files. I use the detailed plan-level *Benefits* files to add information on each plan's cost-sharing attributes. Using the detailed *Bid* data, I add each plan's bid, rebate, premiums, benchmark, the share of rebates allocated to each premium reduction, and benefits.

Using this detailed panel, I exclude Special Needs Plans, dual Medicare/Medicaid plans, and those offered exclusively to some employers and not the common market. I also exclude

a small fraction of plans that were not approved to operate in a county despite submitting a bid. I adjust all dollar values according to the 2015 Medical CPI.

I.B MCBS

The MCBS is a collection of files in different formats for different years. This data poses several challenges. First, column names, placement of variables among files, and categorical variable definitions are inconsistent across years. Second, some variables exist in the SAS version of files and not in the flat text files and vice-versa. This involves a lengthy manual work linking columns across years, which I did by starting from a crosswalk provided by the MCBS for connecting 2015 and 2013 data. I automated a process that reads the SAS description files, recovered each column's format, and applied it directly to categorical variables in the flat text files. This also flagged missing columns, which I recovered separately.

I merged the MCBS individual-level data with the public plan-level data using the contract and plan names when available. For 2009-2011 and 2015, the MCBS include the part C plan name, which provides an accurate match. In 2012-2013 I used the part D contract number and an "enrolled in MA" indicator. As most MA plans bundle part D coverage, these contract names are almost always MAPDP contracts. 731 out of 12392 beneficiary-years have more than one contract number in the data. I keep the first contract (chronologically) observed in the aggregate data for those. Overall, 98.9% of single contract consumers have a match in the aggregate data, and only 9 beneficiary-years with more than one contract have no match in the aggregate data. This results in 29312 beneficiary-years to match at the plan level. 15088 of those have more than one plan associated with their contract. I first match by premium, matching 7858. 1477 are then matched based on dental benefits and 9 on eye benefits (both of these columns exist in the MCBS and the aggregate data). The remaining 3207 beneficiary-year are assigned to the most popular plan within their contract-year-county.

II Scoring design regulation

This section provides additional details on the construction of the MA Star Ratings and quality regulation in the market.

II.A Included and excluded measures

CMS first classifies measures as either included in the rating or excluded, and then into one of five categories: outcomes (e.g., improving or maintaining physical health), intermediate

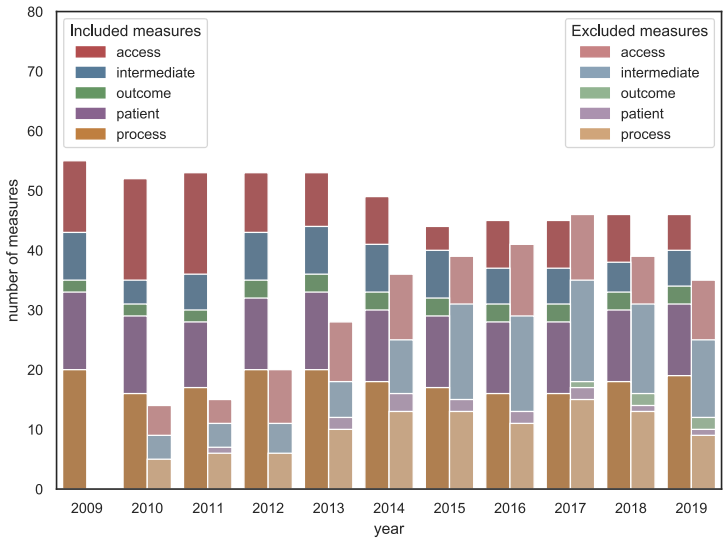


Figure 1: Number of included and excluded quality measures

Note: The weight assigned to each included measure before 2012 was equal to 1. Starting in 2012, it was set to 3 for measures categorized as Outcomes and Intermediate Outcomes, 1.5 for patient and access, and 1 for Process. Excluded measures weight zero.

outcomes (e.g., controlling blood pressure), patient experience and complaints (e.g., consumer rating of the plan's customer service), access (e.g., the processing time for appeals), and process (e.g., colorectal cancer screenings).[1] The correlation across and among the included and excluded groups, by category, is presented in Table 2. Included quality measures belonging to the same category are considered to represent a similar type of quality and are assigned an identical weight in the overall rating. As shown in Figure 1, the number of measures in each category has varied over the years. As there's a single weight for a category, this has created significant variation in the information represented by the Star Ratings.

II.B The Quality Bonus Payment (QBP) program

In addition to the star ratings, CMS also provides direct pecuniary incentives for insurers to improve their scores. The Affordable Care Act of 2010 mandated this Quality Bonus Payments (QBP) system, which took effect in the enrollment-year 2012. QBP introduced two changes to the bidding system of insurers. First, it modified the percentage paid in rebates for contracts bidding below the benchmark from a uniform 75% to a formula increasing in the previous year's star rating. Second, it introduced a *reward factor* that increases the

[1]Most quality measures can be thought of as "the fraction of enrollees obtaining service X." For example, CMS computes the outcome measure *improving or maintaining physical health* as the fraction of the population that saw their physical health improve or sustained during the last year.

Table 1: Quality Payment Rules

	Year				
Star Rating	2009-2011	2012	2013	2014	2015-2019
			Benchmark Bonus		
≤ 2.5	0.0%	0.0%	0.0%	0.0%	0.0%
3.0	0.0%	3.0%	3.0%	3.0%	0.0%
3.5	0.0%	3.5%	3.5%	3.5%	0.0
4.0	0.0%	4.0%	4.0%	5.0%	5.0%
4.5	0.0%	4.0%	4.0%	5.0%	5.0%
5.0	0.0%	5.0%	5.0%	5.0%	5.0%
New Plans	0.0%	3.0%	3.0%	3.5%	3.5%
Low Enrollment Plans	0.0%	3.0%	3.0%	3.0%	3.5%
			Rebate		
≤ 2.5	75%	66.67%	58.33%	50%	50%
3.0	75%	66.67%	58.33%	50%	50%
3.5	75%	71.67%	68.33%	65%	65%
4.0	75%	71.67%	68.33%	65%	65%
4.5	75%	73.33%	71.67%	70%	70%
5.0	75%	73.33%	71.67%	70%	70%
New Plans	75%	71.67%	68.33%	65%	65%
Low Enrollment Plans	75%	73.33%	58.33%	50%	65%

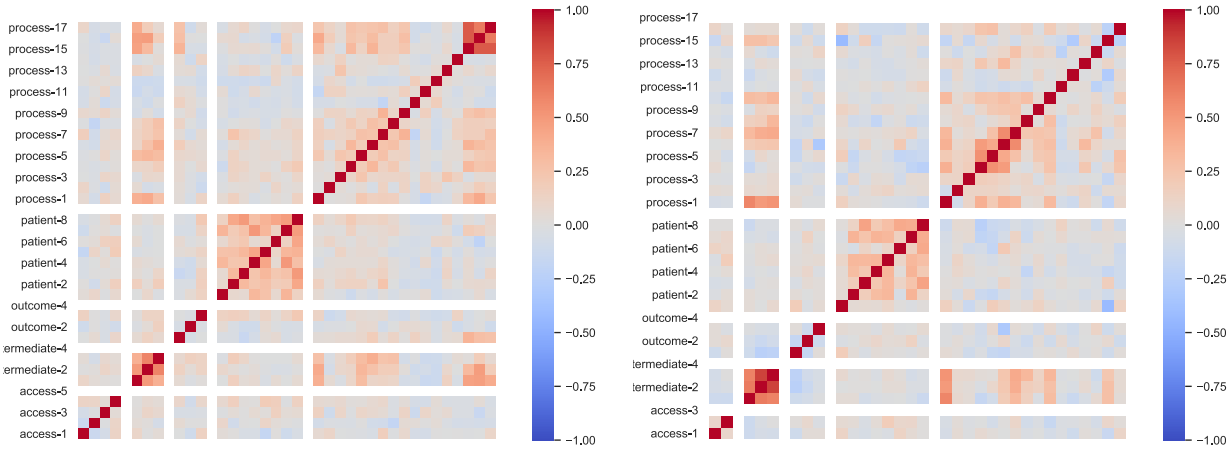
Note: New plans are those offered by a parent organization that has not had any MA contract(s) with CMS in the previous three years. A low enrollment contract is a contract that could not undertake Healthcare Effectiveness Data and Information Set (HEDIS) and Health Outcome Survey (HOS) data collections because of a lack of a sufficient number of enrollees to reliably measure the performance of the health plan (Advance notice of methodological changes, 2012)

Table 2: Correlation between included and excluded measures

	access	intermediate	outcome	patient	process
access	0.102	0.223	0.275	0.264	0.203
intermediate	-0.061	-0.364	0.404	0.193	0.658
outcome	0.375	-0.253	0.095	0.171	0.242
patient	-0.099	-0.267	0.168	0.354	0.238
process	0.142	0.486	-0.031	-0.288	0.033

Notes: On the diagonal: the correlation between included and excluded measures in each category. Above the diagonal: the correlation across those included. Below the diagonal: the correlation across those excluded..

benchmark for plans with a higher previous year rating. Overall, after 2012 the rebate paid



(a) Part C measures, 2014 (b) Part C measures, 2015

Figure 2: Correlation in measure-level quality within and across categories

Notes: These figures present the correlation within and across categories in the measure-level quality. Quality is adjusted and normalized as to span the unit interval and have consistent interpretation across years. As the figures show, there is a substantial correlation across measures within categories..

to a contract j bidding p_{jt} against a benchmark B_{jt} in year t is given by

$$\text{Rebate}_{jt}(p_{jt}) = \max\{\rho(r_{jt-1})(B_{jt}(1+\tau(r_{jt-1})) - p_{jt}), 0\} \tag{1}$$

where $\rho(r_{jt-1})$ and $\tau(r_{jt-1})$ are the *rebate share* and *reward factor* of a plan j that obtained a rating of r_{jt-1} in year $t-1$, respectively. Table 1 shows the formula for $\tau(\cdot)$ and $\rho(\cdot)$.

There is some discrepancy in the literature about the rating year that enters the QBP calculation. In particular, some have modeled it as depending on the current year (i.e., R_{jt} being a function of r_{jt}). However, two pieces of evidence clarify that this is not the case. First, the rating is required at bidding time as the rebate plays an important role in the bidding regulation. This occurs between June and September of the preceding year $(t-1)$. Second, the following year's ratings (r_{jt}) are released in mid-October, together with the open enrollment period. Therefore, it can not be that QBP rules use the same ratings as those used by consumers in the same year.

II.C Star rating formula and reconstruction from public data

CMS provides six data sets required to compute the star ratings (the names correspond roughly to the names of files within the CMS zipped rating files)

1. **Stars**: The per measure star levels computed by CMS

2. **Data**: The quality measurements for included measures
3. **Display**: The quality measurement for excluded measures
4. **Cutpoints**: The cutpoints used to transform measure-level data to measure-level stars
5. **Summary**: The summary stars of each contract
6. **Domain**: Domain-level stars of each contract

Importantly, two steps require work. First, the weights associated with each measure are not provided in the CSV files but must be recovered from a series of PDF files included in the zipped folders. Second, the direction of improvement is not always clear for each measure from the data. This is particularly true for excluded (display) measures. However, these can be recovered manually by examining CMS communication documents each year, announcing changes to the rating program. Some of the most difficult to pin down are (with their direction of improvement)

1. Call center pharmacy hold time: decreasing
2. Appeals auto-forward: decreasing
3. Doctors who communicate well: increasing
4. MPF composite: increasing
5. Call center calls disconnected when customer calls drug plan: decreasing
6. Beneficiary access and performance problems: decreasing in 2010-2011, increasing in 2012-2014
7. Call center calls disconnected when calling pharmacist: decreasing in 2009
8. Complaints about enrollment: decreasing in 2009
9. Call center beneficiary hold time: decreasing in 2011
10. MPF price accuracy: increasing in 2013-2014

Some additional considerations have to be taken when processing the files. First, special care has to be taken with the names of measures, as they haven't been consistent over the years. Second, some measures change from being presented as percentages to equivalent

fractions. Third, some quality measures are duplicated among the parts C and D files. Contracts that offer part C and part D coverage must be adjusted to not double-count the measures. Some part D measures are only relevant for pure part D contracts and have to be excluded from MAPDP plans.

The adjustment factor discussed in the main text is composed of two components. First, CMS computes an *improvement star* for contracts that have maintained or improved their quality over the preceding years. The data and formula used to compute this are not always available. Hence I take it as given. As the model developed in this paper is static, this is not particularly troublesome. Second, CMS computes a *reward factor* based on how the weighted average and variance of a contract's measure-level stars compare with the same values of other contracts. Specifically, letting $\bar{s}_j$ and $WV(s_j)$ denote the weighted average and variance of measure-level stars for contract j, the reward factor is defined as

$$\text{reward}(s_j) = \begin{cases} 0.4 & WV(s_j) < Q30(WV(\boldsymbol{s})) \text{ and } \bar{s}_j > Q85(\bar{\boldsymbol{s}}) \\ 0.3 & Q30(WV(\boldsymbol{s})) \leq WV(s_j) < Q70(WV(\boldsymbol{s})) \text{ and } \bar{s}_j > Q85(\bar{\boldsymbol{s}}) \\ 0.2 & WV(s_j) < Q30(WV(\boldsymbol{s})) \text{ and } Q65(\bar{\boldsymbol{s}}) < \bar{s}_j \leq Q85(\bar{\boldsymbol{s}}) \\ 0.1 & Q30(WV(\boldsymbol{s})) \leq WV(s_j) < Q70(WV(\boldsymbol{s})) \text{ and } Q65(\bar{\boldsymbol{s}}) < \bar{s}_j \leq Q85(\bar{\boldsymbol{s}}) \\ 0 & \text{otherwise} \end{cases}$$

Where $Q30(\bar{\boldsymbol{s}})$ is the 30-th percentile of the average rating among all contracts for the year, and $Q65, Q70, Q85$ define analogously.

The continuous star rating of a plan is defined based on five values:

1. $\bar{s}_j$: the weighted average star rating of a plan across its measures, including the improvement star.
2. $\bar{s}_j^o$: the weighted average star rating of a plan, across its measures, without the improvement star.
3. reward(s_j): the reward factor computed using the improvement star.
4. reward$^o(s_j)$: the reward factor computed without using the improvement star.
5. improv$_j$: an indicator of whether CMS awarded the contract j an improvement star.

The continuous star rating of contract j is thus

$$cr_j = \begin{cases} \bar{s}_j^o + \text{reward}_j^o(s_j) & \text{if } \bar{s}_j^o + \text{reward}_j^o(s_j) \leq 1.75 \text{ or } \text{improv}_j = 0 \\ \bar{s}_j + \text{reward}_j(s_j) & \text{if } \bar{s}_j^o + \text{reward}_j^o(s_j) \in (1.75, 3.75) \text{ and } \text{improv}_j = 1 \\ \max\{\bar{s}_j + \text{reward}_j(s_j), \bar{s}_j^o + \text{reward}_j^o(s_j)\} & \text{if } \bar{s}_j^o + \text{reward}_j^o(s_j) \geq 3.75 \text{ and } \text{improv}_j = 1 \end{cases}$$

The overall star rating of contract j is cr_j rounded up to the nearest .5 decimal.

We can map the detailed description above to the simple formula in the main text. First, for each measure m, define $\psi_m(\cdot)$ as the step function that transforms quality measurements to measure-level stars, multiplied by the relative weight corresponding to the measure given its category. Second, define the adjustment factor as

$$\omega_j = \begin{cases} \text{reward}_j^o(s_j) & \text{if } \bar{s}_j^o + \text{reward}_j^o(s_j) \leq 1.75 \text{ or } \text{improv}_j = 0 \\ w^{\text{improve}} s_j^{\text{improve}} + \text{reward}_j(s_j) & \text{if } \bar{s}_j^o + \text{reward}_j^o(s_j) \in (1.75, 3.75) \text{ and } \text{improv}_j = 1 \\ w^{\text{improve}} s_j^{\text{improve}} + \text{reward}_j(s_j) & \text{if } \bar{s}_j^o + \text{reward}_j^o(s_j) \geq 3.75 \text{ and } \text{improv}_j = 1 \\ & \text{and } \bar{s}_j + \text{reward}_j(s_j) \geq \bar{s}_j^o + \text{reward}_j^o(s_j) \\ \text{reward}_j^o(s_j) & \text{otherwise} \end{cases}$$

where w^{improve} is the relative weight of the improvement star and s_j^{improve} is contract j's improvement star.

Finally, it is worth pointing out that there are three types of contracts for which CMS does not compute a star rating as above. First are new contracts of organizations that did not offer any contracts in the previous three years. For QBP benchmark bonuses, these were treated as three stars between 2012 and 2013 and as 3.5 stars afterward. For rebate purposes, these are treated as 3.5 stars throughout the years. Second, are new contracts of existing organizations. These are assigned the average star rating of the firm's other contracts, weighted by enrollment, in the earliest year possible, up to three years back. Finally, the third group is composed of contracts with low enrollment. For QBP benchmarks, they were treated as having three stars before 2015 and as 3.5 afterward. For rebate purposes, they were treated as 4.5 stars in 2012, 3 stars in 2013-2014, and 3.5 starting in 2015.

III Individual risk scores

Each year, CMS publishes SAS code that takes as inputs information about each beneficiary's date of birth, dual-eligibility status, gender, reason for Medicare entitlement, and a history of physician, inpatient, and outpatient diagnosis codes. The code maps this information using precomputed parameters into an individual's risk score. As CMS computes consumers' risk scores using more data than what is available in the MCBS, I perform an ex-post adjustment using the plan's landscape files. By linking each MCBS beneficiary with its chosen plan, and by comparing the predicted and observed plan average risk score (which is contained in the landscape files) I scale enrollees' score to match their plans' reported level.

IV Predicted spending model

Using the MCBS linked files, I compute each consumer's total spending on eight categories: inpatient care, outpatient care, physician visits, home health services, durable medical devices, skilled nursing facilities, and hospice. I split consumers into age groups as: $< 65, [65, 75), [75, 85), [95, \infty)$. Letting Y_{ikt} denote total spending by consumer i in category k in year t, I estimate a two-stage saturated spending model for each category:

$$\mathbb{1}\{Y_{ikt} > 0\} = \eta^F_{f(i),k} + \eta^T_{t,k} + \eta^S_{s(i),k} + \boldsymbol{\alpha}'_{i,k}\boldsymbol{x}_{it} + \nu_{ikt} \tag{2}$$

$$\ln(Y_{ikT}) = \delta^F_{f(i),k} + \delta^T_{t,k} + \delta^S_{s(i),k} + \boldsymbol{\beta}'_{i,k}\boldsymbol{x}_{it} + \epsilon_{ikt} \tag{3}$$

Where η^F, δ^F are fixed-effects for the insurer covering consumer i in period t, denoted $f(i)$, η^T, δ^T are year fixed effects, and η^S, δ^S are state fixed effects. $\boldsymbol{x}_{it}$ contains all chronic conditions reported on the MCBS, and all indicators of impairement or difficulty with activities of daily living (ADL). In addition it contains the following factors: smoker, obese, underweight, has urininary control, has had pneumonia shot within the last year, has had flu shot within the last year, frequent faller, any falls last year, depressed, age, sex, self-reported health status, asian indicator, black indicator, hispanic indicator, is working indicator, suffers from memory loss, on low income subsidy, medicaid eligibility, medicare status, has pc, had prostate check within the last 4 years, has had a mammogram within the last 4 years, had a pap smear within the last 4 years, has had blood prostate check within last 4 years, graduated high school, attained a college degree or higher, income, age group, weight, and interaction between age and race, age and falls, gender and gender-specific cancer diagnostic checks.

Regression (2) is a probit model, subject to the appropriate assumption on ν. Regression

(3) is a saturated high-dimensional linear regression. Predicted spending is computed by forming the predicted probability of having any spending based on the probit model, times the expected magnitude of spending based on the second stage. The estimation and prediction is done separately across categories of spending and then summed up to form total consumer predicted spending.