§ 422.164 Adding, updating, and removing measures.

(a) *General.* CMS adds, updates, and removes measures used to calculate the Star Ratings as provided in this section. CMS lists the measures used for a particular Star Rating each year in the Technical Notes or similar guidance document with publication of the Star Ratings.

(b) *Review of data quality.* CMS reviews the quality of the data on which performance, scoring and rating of a measure is based before using the data to score and rate performance or in calculating a Star Rating. This includes review of variation in scores among MA organizations and Part D plan sponsors, and the accuracy, reliability, and validity of measures and performance data before making a final determination about inclusion of measures in each year's Star Ratings.

(c) *Adding measures.* (1) CMS will continue to review measures that are nationally endorsed and in alignment with the private sector, such as measures developed by National Committee for Quality Assurance (NCQA) and the Pharmacy Quality Alliance (PQA), or endorsed by the National Quality Forum for adoption and use in the Part C and Part D Quality Ratings System. CMS may develop its own measures as well when appropriate to measure and reflect performance specific to the Medicare program.

(2) In advance of the measurement period, CMS will announce potential new measures and solicit feedback through the process described for changes in and adoption of payment and risk adjustment policies in section 1853(b) of the Act and then subsequently will propose and finalize new measures through rulemaking.

(3) New measures added to the Part C Star Ratings program will be on the display page on *www.cms.gov* for a minimum of 2 years prior to becoming a Star Ratings measure.

(4) A measure will remain on the display page for longer than 2 years if CMS finds reliability or validity issues with the measure specification.

(d) *Updating measures*—(1) *Non-substantive updates.* For measures that are already used for Star Ratings, CMS will update measures so long as the changes in a measure are not substantive. CMS will announce non-substantive updates to measures that occur (or are announced by the measure steward) during or in advance of the measurement period through the process described for changes in and adoption of payment and risk adjustment policies in section 1853(b) of the Act. Non-substantive measure specification updates include those that—

(i) Narrow the denominator or population covered by the measure;

(ii) Do not meaningfully impact the numerator or denominator of the measure;

(iii) Update the clinical codes with no change in the target population or the intent of the measure;

(iv) Provide additional clarifications:

(A) Adding additional tests that would meet the numerator requirements;

(B) Clarifying documentation requirements;

(C) Adding additional instructions to identify services or procedures; or

(v) Add alternative data sources.

(2) *Substantive updates.* For measures that are already used for Star Ratings, in the case of measure specification updates that are substantive updates not subject to paragraph (d)(1) of this section, CMS will propose and finalize these measures through rulemaking similar to the process for adding new measures. CMS will initially solicit feedback on whether to make substantive measure updates through the process described for changes in and adoption of payment and risk adjustment policies in section 1853(b) of the Act. Once the update has been made to the measure specification by the measure steward, CMS may continue collection of performance data for the legacy measure and include it in Star Ratings until the updated measure has been on display for 2 years. CMS will place the updated measure on the display page for at least 2 years prior to using the updated measure to calculate and assign Star Ratings as specified in paragraph (c) of this section.

(e) *Removing measures.* (1) CMS will remove a measure from the Star Ratings program as follows:

(i) When the clinical guidelines associated with the specifications of the

measure change such that the specifications are no longer believed to align with positive health outcomes; or

(ii) A measure shows low statistical reliability.

(iii) The measure steward other than CMS retires a measure.

(2) CMS will announce in advance of the measurement period the removal of a measure based upon its application of this paragraph (e) through the process described for changes in and adoption of payment and risk adjustment policies in section 1853(b) of the Act in advance of the measurement period.

(f) *Improvement measure.* CMS will calculate improvement measure scores based on a comparison of the measure scores for the current year to the immediately preceding year as provided in this paragraph (f); the improvement measure score would be calculated for Parts C and D separately by taking a weighted sum of net improvement divided by the weighted sum of the number of eligible measures.

(1) *Identifying eligible measures.* Annually, the subset of measures to be included in the Part C and Part D improvement measures will be announced through the process described for changes in and adoption of payment and risk adjustment policies in section 1853(b) of the Act. CMS identifies measures to be used in the improvement measures if the measures meet all of the following:

(i) CMS will include only measures available for the current and previous year in the improvement measures and that have numeric value scores in both the current and prior year.

(ii) CMS will exclude any measure for which there was a substantive specification change from the previous year.

(iii) CMS will exclude any measures that are already focused on improvement in MA organization performance from year to year.

(iv) The Part C improvement measure will include only Part C measure scores; the Part D improvement measure will include only Part D measure scores.

(v) CMS excludes any measure that receives a measure-level Star Rating reduction for data integrity concerns for either the current or prior year from the improvement measure(s).

(2) *Determining eligible contracts.* CMS will calculate an improvement score only for contracts that have numeric measure scores for both years in at least half of the measures identified for use applying the standards in paragraphs (f)(1)(i) through (iv) of this section.

(3) *Special rules for calculation of the improvement score.* For any measure used for the improvement measure for which a contract received 5 stars in each of the years examined, but for which the measure score demonstrates a statistically significant decline based on the results of the significance testing (at a level of significance of 0.05) on the change score, the measure will be categorized as having no significant change and included in the count of measures used to determine eligibility for the measure (that is, for the denominator of the improvement measure score).

(4) *Calculation of the improvement score.* The improvement measure will be calculated as follows:

(i) The improvement change score (the difference in the measure scores in the 2-year period) will be determined for each measure that has been designated an improvement measure and for which a contract has a numeric score for each of the 2 years examined.

(ii) Each contract's improvement change score per measure will be categorized as a significant change or not a significant change by employing a two-tailed t-test with a level of significance of 0.05.

(iii) The net improvement per measure category (outcome, access, patient experience, process) would be calculated by finding the difference between the weighted number of significantly improved measures and significantly declined measures, using the measure weights associated with each measure category.

(iv) The improvement measure score will then be determined by calculating the weighted sum of the net improvement per measure category divided by the weighted sum of the number of eligible measures.

(v) The improvement measure scores will be converted to measure-level Star Ratings by determining the cut points

using hierarchical clustering algorithms in accordance with § 422.166(a)(2)(i) through (iii).

(vi) The Part D improvement measure cut points for MA–PDs and PDPs will be determined using separate clustering algorithms in accordance with §§ 422.166(a)(2)(iii) and 423.186(a)(2)(iii) of this chapter.

(g) *Data integrity.* (1) CMS will reduce a contract's measure rating when CMS determines that a contract's measure data are inaccurate, incomplete, or biased; such determinations may be based on a number of reasons, including mishandling of data, inappropriate processing, or implementation of incorrect practices that have an impact on the accuracy, impartiality, or completeness of the data used for one or more specific measure(s).

(i) CMS will reduce HEDIS measures to 1 star when audited data are submitted to NCQA with a designation of ''biased rate'' or BR based on an auditor's review of the data or a designation of ''nonreport'' or NR.

(ii) CMS will reduce measures based on data that an MA organization must submit to CMS under § 422.516 to 1 star when a contract did not score at least 95 percent on data validation for the applicable reporting section or was not compliant with CMS data validation standards/substandards for data directly used to calculate the associated measure.

(iii) For the appeals measures, CMS will use statistical criteria to estimate the percentage of missing data for each contract (using data from multiple sources such as a timeliness monitoring study or audit information) to scale the star reductions to determine whether the data at the independent review entity (IRE) are complete. CMS will use scaled reductions for the Star Ratings for the applicable appeals measures to account for the degree to which the IRE data are missing.

(A)(*1*) The data submitted for the Timeliness Monitoring Project (TMP) or audit that aligns with the Star Ratings year measurement period is used to determine the scaled reduction.

(*2*) For contract consolidations approved on or after January 1, 2022, if there is a contract consolidation as described at § 422.162(b)(3), the TMP or

audit data are combined for the consumed and surviving contracts before the methodology provided in paragraphs (g)(1)(iii)(B) through (O) of this section is applied.

(B) The determination of the Part C appeals measure IRE data reduction is done independently of the Part D appeals measure IRE data reduction.

(C) The reductions range from a one-star reduction to a four-star reduction; the most severe reduction for the degree of missing IRE data is a four-star reduction.

(D) The thresholds used for determining the reduction and the associated appeals measure reduction are as follows:

(*1*) 20 percent, 1 star reduction.
(*2*) 40 percent, 2 star reduction.
(*3*) 60 percent, 3 star reduction.
(*4*) 80 percent, 4 star reduction.

(E) If a contract receives a reduction due to missing Part C IRE data, the reduction is applied to both of the contract's Part C appeals measures.

(F) If a contract receives a reduction due to missing Part D IRE data, the reduction is applied to both of the contract's Part D appeals measures.

(G) The scaled reduction is applied after the calculation for the appeals measure-level Star Ratings. If the application of the scaled reduction results in a measure-level star rating less than 1 star, the contract will be assigned 1 star for the appeals measure.

(H) The Part C Calculated Error is determined using the quotient of number of cases not forwarded to the IRE and the total number of cases that should have been forwarded to the IRE. (The number of cases that should have been forwarded to the IRE is the sum of the number of cases in the IRE during the data collection or data sample period and the number of cases not forwarded to the IRE during the same period.)

(I) The Part D Calculated Error is determined by the quotient of the number of untimely cases not auto-forwarded to the IRE and the total number of untimely cases.

(J) The projected number of cases not forwarded to the IRE in a 3-month period is calculated by multiplying the number of cases found not to be forwarded to the IRE based on the TMP or

audit data by a constant determined by the data collection or data sample time period. The value of the constant will be 1.0 for contracts that submitted 3 months of data; 1.5 for contracts that submitted 2 months of data; and 3.0 for contracts that submitted 1 month of data.

(K) Contracts are subject to a possible reduction due to lack of IRE data completeness if both of the following conditions are met:

(*1*) The calculated error rate is 20 percent or more.

(*2*) The projected number of cases not forwarded to the IRE is at least 10 in a 3-month period.

(L) A confidence interval estimate for the true error rate for the contract is calculated using a Score Interval (Wilson Score Interval) at a confidence level of 95 percent and an associated z of 1.959964 for a contract that is subject to a possible reduction.

(M) A contract's lower bound is compared to the thresholds of the scaled reductions to determine the IRE data completeness reduction.

(N) The reduction is identified by the highest threshold that a contract's lower bound exceeds.

(O) CMS reduces the measure rating to 1 star for the applicable appeals measure(s) if a contract fails to submit Timeliness Monitoring Project data for CMS's review to ensure the completeness of the contract's IRE data.

(2) CMS will reduce a measure rating to 1 star for additional concerns that data inaccuracy, incompleteness, or bias have an impact on measure scores and are not specified in paragraphs (g)(1)(i) through (iii) of this section, including a contract's failure to adhere to HEDIS, HOS, or CAHPS reporting requirements.

(h) *Review of sponsors' data.* (1) An MA organization may request that CMS or the IRE review its' contract's appeals data provided that the request is received by the annual deadline set by CMS.

(2) An MA organization may request that CMS review its' contract's Complaints Tracking Module (CTM) data provided that the request is received by the annual deadline set by CMS for the applicable Star Ratings year.

(i) [Reserved]

[83 FR 16725, Apr. 16, 2018, as amended at 84 FR 15829, Apr. 16, 2019; 85 FR 19290, Apr. 6, 2020; 86 FR 6097, Jan. 19, 2021; 87 FR 27895, May 9, 2022; 88 FR 22332, Apr. 12, 2023]

§ 422.166 Calculation of Star Ratings.

(a) *Measure Star Ratings*—(1) *Cut points.* CMS will determine cut points for the assignment of a Star Rating for each numeric measure score by applying either a clustering or a relative distribution and significance testing methodology. For the Part D measures, CMS will determine MA–PD and PDP cut points separately.

(2) *Clustering algorithm for all measures except CAHPS measures.*

(i) The method maximizes differences across the star categories and minimizes the differences within star categories using mean resampling with the hierarchal clustering of the current year's data. Effective for the Star Ratings issued in October 2023 and subsequent years, prior to applying mean resampling with hierarchal clustering, Tukey outer fence outliers are removed. Effective for the Star Ratings issued in October 2022 and subsequent years, CMS will add a guardrail so that the measure-threshold-specific cut points for non-CAHPS measures do not increase or decrease more than the value of the cap from 1 year to the next. The cap is equal to 5 percentage points for measures having a 0 to 100 scale (absolute percentage cap) or 5 percent of the restricted range for measures not having a 0 to 100 scale (restricted range cap). New measures that have been in the Part C and D Star Rating program for 3 years or less use the hierarchal clustering methodology with mean resampling with no guardrail for the first 3 years in the program.

(ii) In cases where multiple clusters have the same measure score value range, those clusters would be combined, leading to fewer than 5 clusters.

(iii) The clustering algorithm for the improvement measure scores is done in two steps to determine the cut points for the measure-level Star Ratings. Clustering is conducted separately for improvement measure scores greater than or equal to zero and those with