Blue Cross Blue Shield of Massachusetts Pay-for-Equity Technical Methods, 2023

This document presents Blue Cross Blue Shield of Massachusetts' (BCBSMA's) technical approach to designing a pay-for-equity financial incentive for provider organizations participating in BCBSMA's Alternative Quality Contract (AQC). Like many Accountable Care Organization (ACO) contracts, the AQC features financial incentives linked to provider organization performance on measures of quality and total costs of care. In 2023, BCBSMA began to introduce financial incentives within the AQC that are explicitly linked to measures of racial and ethnic equity of care. These pay-for-equity incentives are treated similarly to measures of overall quality of care within the structure of the AQC; they constitute a percentage of a new quality-equity score, which appears in the AQC in all the same places that the quality score formerly appeared (i.e., as a stand-alone performance incentive and as a factor that affects the risk share for the cost component of the AQC).

This document is intended to help readers understand BCBSMA's technical approach to pay-for-equity and the statistical and methodological issues involved. It contains 3 chapters:

1. Calculating the Equity Incentive Measure (EIM) Score
2. Determining a Measure's Eligibility for the EIM Score
3. Determining the Curve Shape for the EIM Score

As context for these methodological materials, there are a few key concepts to keep in mind. First, BCBSMA's pay-for-equity program might change over time. This document describes only the first cohort of contracts in the program (i.e., those that went into effect on January 1, 2023). Second, the equity incentive measures (EIMs) that are the basis of the program are a race- and ethnicity-stratified subset of the quality measures already present in the AQC. In other words, every EIM has a corresponding overall quality measure incentive. For example, in a given AQC there is an EIM that incentivizes reduction of racial and ethnic inequities in blood pressure control, then the AQC also must include an incentive to improve overall performance (based on aggregated data for members of all races and ethnicities) on the same measure of blood pressure control. Third, the program incentivizes each provider organization to reduce racial and ethnic inequities among BCBSMA members that exist within their organization. BCBSMA's strategies to close between-provider organization racial and ethnic inequities are distinct from and complementary to the current pay-for-equity program; these between-provider strategies are outside the scope of this document. Fourth, while BCBSMA's pay-for-equity program focuses on racial and ethnic inequities, the methods described herein can be applied to future changes in how races and ethnicities are categorized, to other bases of inequities (e.g., language, sexual orientation, gender identity, national origin, disability status), and to intersections between current and future bases of inequities. All that is necessary to apply these methods is to have a basis for stratifying quality measure performance data and a goal of reducing any inequities that are measured via this stratification.

Chapter 1. Calculating the Equity Incentive Measure Score

Section I. Data Elements

Let i be the index for measurement year, where the measurement year can either be baseline ("base") or applicable quality measurement period ("MP"). The latter represents any of the years in which the contract is in force. For four racial/ethnic groups W, X, Y, Z, the Equity Incentive Measure (EIM) score calculation involves the following data elements collected at baseline year and at applicable quality measurement period (Measurement Period; there is one Measurement Period for each year of the contract) for the EIM in question:

- $\mathrm{n}_{\mathrm{W,i}}$ is the i denominator value for members classified to racial/ethnic group W
- $\mathrm{p}_{\mathrm{W,i}}$ is the i stratified performance for members classified to racial/ethnic group W
- $\mathrm{n}_{\mathrm{X,i}}$ is the i denominator value for members classified to racial/ethnic group X
- $\mathrm{p}_{\mathrm{X,i}}$ is the i stratified performance for members classified to racial/ethnic group X
- $\mathrm{n}_{\mathrm{Y,i}}$ is the i denominator value for members classified to racial/ethnic group Y
- $\mathrm{p}_{\mathrm{Y,i}}$ is the i stratified performance for members classified to racial/ethnic group Y
- $\mathrm{n}_{\mathrm{Z,i}}$ is the i denominator value for members classified to racial/ethnic group Z
- $\mathrm{p}_{\mathrm{Z,i}}$ is the i stratified performance for members classified to racial/ethnic group Z

An alternative visualization of the data needed to compute the EIM score is:

Racial/Ethnic Group	Denominator		Stratified Performance	
	Baseline	Measurement Period	Baseline	Measurement Period
W	$\mathrm{n}_{\mathrm{W,base}}$	$\mathrm{n}_{\mathrm{W,MP}}$	$\mathrm{p}_{\mathrm{W,base}}$	$\mathrm{p}_{\mathrm{W,MP}}$
X	$\mathrm{n}_{\mathrm{X,base}}$	$\mathrm{n}_{\mathrm{X,MP}}$	$\mathrm{p}_{\mathrm{X,base}}$	$\mathrm{p}_{\mathrm{X,MP}}$
Y	$\mathrm{n}_{\mathrm{Y,base}}$	$\mathrm{n}_{\mathrm{Y,MP}}$	$\mathrm{p}_{\mathrm{Y,base}}$	$\mathrm{p}_{\mathrm{Y,MP}}$
Z	$\mathrm{n}_{\mathrm{Z,base}}$	$\mathrm{n}_{\mathrm{Z,MP}}$	$\mathrm{p}_{\mathrm{Z,base}}$	$\mathrm{p}_{\mathrm{Z,MP}}$

Section II. Definitions and Calculations

1. Calculations completed at baseline using baseline data, before the Measurement Period begins:

For a specific EIM, start by identifying the reference group and by determining which of the racial/ethnic groups will be included in the EIM score calculation. The reference group for this EIM is the racial/ethnic stratum with the largest baseline denominator for this EIM. Selecting the largest denominator as the reference ensures that at least one of the rates used to calculate the inequity will be based on a relatively large value. To determine which racial/ethnic strata are included in the calculation, start by identifying those with baseline denominator sizes greater than or equal to 90 (i.e., identify which of $\mathrm{n}_{\mathrm{W,base}}, \mathrm{n}_{\mathrm{X,base}}, \mathrm{n}_{\mathrm{Y,base}}, \mathrm{n}_{\mathrm{Z,base}}$ are ≥ 90)[1]. Remove any strata with baseline denominators less than 90. If only one racial/ethnic stratum has a baseline denominator greater than or equal to 90, then the entire measure is not eligible for an EIM score. More details regarding the conditions used to determine whether a measure is eligible for an EIM score are in Chapter 2.

For illustrative purposes, assume Group Z is the reference group and assume that all four racial/ethnic groups will be included in the EIM score calculation for this measure (because their baseline denominators all are ≥ 90). However, each of the next definitions can be adapted to cases where two or three racial/ethnic groups are included in the EIM score calculation.

[1] Values other than 90 also were tested.

Minimum Denominator Requirement (MDR): MDRs are calculated at 70% - 100% of denominators in the baseline year (MY2019) based on simulations examining the degree of variability in the measure[2]. The selection of 70, 80, 90, or 100% is based on the simulation described in Chapter 2. Specifically, defining $\mathrm{r_{MDR}}$ as a constant where $\mathrm{r_{MDR}} \in \{0.7, 0.8, 0.9, 1.0\}$, the MDR is equal to $\mathrm{r_{MDR}(n_{t,base})}$ for each $\mathrm{t} \in \{\mathrm{W, X, Y, Z}\}$.

Minimum Performance Required (MPR): This is the minimum accepted value for the Measurement Period Stratified Performance for each race/ethnicity category[3]. Use $\mathrm{n_{t,base}}$, $\mathrm{p_{t,base}}$ for each $\mathrm{t} \in \{\mathrm{W, X, Y, Z}\}$ to calculate Bonferroni-corrected two-sided $(1 - \frac{\alpha}{\mathrm{num_{RE}}})$% Wald confidence intervals. $\alpha = 0.05$ and $\mathrm{num_{RE}}$ is equal to the number of racial/ethnic strata that remains after applying the 90-value minimum baseline denominator size. Assuming greater values of p indicate better levels of performance, the lower bound of this confidence interval for each $\mathrm{t} \in \{\mathrm{W, X, Y, Z}\}$ is equivalent to the MPR for the corresponding race/ethnicity. If lesser values of p indicate better levels of performance (as in "lower-is-better" measures), then the upper bound of this confidence interval would be used, and the MPR would function as a maximum.

Baseline Category Inequity: The absolute value of Baseline Stratified Performance differences between each racial/ethnic group and the reference group. The Baseline Category Inequities are: $\mathrm{abs(p_{W,base} - p_{Z,base})}$, $\mathrm{abs(p_{X,base} - p_{Z,base})}$, $\mathrm{abs(p_{Y,base} - p_{Z,base})}$.

Baseline Denominator Weights: The values obtained by dividing the Baseline Denominators for all groups except the reference group by the sum of the Baseline Denominators for all groups except the reference group. Baseline Denominator Weights can take values between 0 and 1.

$$\begin{aligned}&\text{Baseline Denominator Weights}\\ &= \{\frac{\mathrm{n_{W,base}}}{\mathrm{n_{W,base} + n_{X,base} + n_{Y,base}}}, \frac{\mathrm{n_{X,base}}}{\mathrm{n_{W,base} + n_{X,base} + n_{Y,base}}}, \frac{\mathrm{n_{Y,base}}}{\mathrm{n_{W,base} + n_{X,base} + n_{Y,base}}} = \{\mathrm{d_W, d_X, d_Y}\}\end{aligned}$$

Baseline Category Inequity Weights: The values obtained by dividing each of the Baseline Category Inequities by the sum of the Baseline Category Inequities. Baseline Category Inequity Weights can take values between 0 and 1.

$$\begin{aligned}&\text{Baseline Category Inequity Weights}\\ &= \left\{\frac{\mathrm{abs(p_{W,base} - p_{Z,base})}}{\mathrm{b}}, \frac{\mathrm{abs(p_{X,base} - p_{Z,base})}}{\mathrm{b}}, \frac{\mathrm{abs(p_{Y,base} - p_{Z,base})}}{\mathrm{b}}\right\} \text{where}\\ &\mathrm{b} = \mathrm{abs(p_{W,base} - p_{Z,base})} + \mathrm{abs(p_{X,base} - p_{Z,base})} + \mathrm{abs(p_{Y,base} - p_{Z,base})}\\ &= \{\mathrm{b_{W-Z}, b_{X-Z}, b_{Y-Z}}\}\end{aligned}$$

Equity Weights: The weights applied to each racial/ethnic category when calculating Baseline Weighted Average Inequity and Weighted Average Inequity. For a given measure, the Equity Weights for each racial/ethnic stratum are the equally weighted average of the Baseline Denominator Weights and the Baseline Category Inequity Weights for that stratum. This was done to place greater weight on racial/ethnic inequities that are larger in magnitude and/or impact a relatively larger number of members (larger baseline denominators).

$$\text{Equity Weights} = \left\{\frac{\mathrm{d_W + b_{W-Z}}}{2}, \frac{\mathrm{d_X + b_{X-Z}}}{2}, \frac{\mathrm{d_Y + b_{Y-Z}}}{2}\right\} = \{\mathrm{w_{W-Z}, w_{X-Z}, w_{Y-Z}}\}$$

[2] We used the simulation in Chapter 2 to inform this decision since some measures with large enough baseline denominators and/or large enough baseline inequities could afford more significant drops in the denominator sizes without major changes to the variability of the measure while other measures could not.

[3] The MPR is incorporated to ensure that rewards are not given for equity improvements that are a result of decreasing performance for some racial/ethnic groups (ex. If White performance is higher than Hispanic performance, Hispanic-White inequity could theoretically be improved by decreasing performance for White members).

Baseline Weighted Average Inequity: The product obtained by multiplying the Equity Weights by the Baseline Category Inequities.

$$\begin{aligned}&\mathrm{Baseline\ Weighted\ Average\ Inequity}\\&\quad= \mathrm{w}_{\mathrm{W-Z}}\mathrm{abs}(\mathrm{p}_{\mathrm{W,base}}-\mathrm{p}_{\mathrm{Z,base}})+\mathrm{w}_{\mathrm{X-Z}}\mathrm{abs}(\mathrm{p}_{\mathrm{X,base}}-\mathrm{p}_{\mathrm{Z,base}})+\mathrm{w}_{\mathrm{Y-Z}}\mathrm{abs}(\mathrm{p}_{\mathrm{Y,base}}-\mathrm{p}_{\mathrm{Z,base}})\end{aligned}$$

2. Calculations completed after the Measurement Period ends using Calculations from Section II.1 and Measurement Period data:

Category Inequity: The absolute value of Measurement Period Stratified Performance differences between each racial/ethnic group and the reference group. The Category Inequities are: $\mathrm{abs}(\mathrm{p}_{\mathrm{W,MP}}-\mathrm{p}_{\mathrm{Z,MP}})$, $\mathrm{abs}(\mathrm{p}_{\mathrm{X,MP}}-\mathrm{p}_{\mathrm{Z,MP}})$, $\mathrm{abs}(\mathrm{p}_{\mathrm{Y,MP}}-\mathrm{p}_{\mathrm{Z,MP}})$.

Weighted Average Inequity: The product obtained by multiplying the Equity Weights by the Category Inequity for each racial/ethnic group.

$$\mathrm{Weighted\ Average\ Inequity} = \mathrm{w}_{\mathrm{W-Z}}\mathrm{abs}(\mathrm{p}_{\mathrm{W,MP}}-\mathrm{p}_{\mathrm{Z,MP}})+\mathrm{w}_{\mathrm{X-Z}}\mathrm{abs}(\mathrm{p}_{\mathrm{X,MP}}-\mathrm{p}_{\mathrm{Z,MP}})+\mathrm{w}_{\mathrm{Y-Z}}\mathrm{abs}(\mathrm{p}_{\mathrm{Y,MP}}-\mathrm{p}_{\mathrm{Z,MP}})$$

Section III. Calculating the EIM Score for a measure identified as eligible for an EIM Score

This section describes how we calculate the EIM score for any measure eligible to receive one[4].

First, complete all baseline calculations for this EIM among the eligible racial/ethnic strata detailed in Section II.1.

For illustrative purposes, consider a particular measure that has been identified as eligible to receive an EIM score. Assume Group Z is the reference group for this measure and assume that all four racial/ethnic groups are being included in the EIM score calculation (because their baseline denominators are all ≥ 90*).*

Then, once Measurement Period data are available,

1. Determine if the Measurement Period denominators ($\mathrm{n}_{\mathrm{W,MP}}, \mathrm{n}_{\mathrm{X,MP}}, \mathrm{n}_{\mathrm{Y,MP}}, \mathrm{n}_{\mathrm{Z,MP}}$) are each greater than or equal to the corresponding MDR. If this is not the case, then this measure is no longer eligible to receive an EIM score, and no further calculation is required[5]. This EIM will be excluded from the Aggregated Weighted EIM Score.

For illustrative purposes, assume that all Measurement Period denominator sizes are greater than or equal to their corresponding MDRs to proceed with the calculation of this specific EIM score.

2. Determine whether the Measurement Period Stratified Performance, ($\mathrm{p}_{\mathrm{W,MP}}, \mathrm{p}_{\mathrm{X,MP}}, \mathrm{p}_{\mathrm{Y,MP}}, \mathrm{p}_{\mathrm{Z,MP}}$), falls below the corresponding MPR for each of the racial/ethnic strata that remain. Assuming larger values of Stratified Performance indicate higher levels of performance, any Measurement Period Stratified Performance value falling below the corresponding MPR represents a statistically significant performance decline relative to baseline year. For a given EIM, if Measurement Period Stratified Performance $\mathrm{p}_{\mathrm{t,MP}}$ for any stratum is less

[4] As a reminder, an EIM score is only calculated on the subset of Ambulatory Care Quality Incentive Measures (AIMs) that are eligible for an EIM score (using the two conditions introduced in Chapter 2).

[5] The measure is no longer eligible for two potential reasons: 1) Large drops in the measurement period denominators may indicate that the variability in measuring improvements in inequities is too large and 2) Significant drops in measurement period denominators for particular racial/ethnic categories may hint that the group is intentionally dropping certain members to perform better on the measure.

than (worse than) its corresponding MPR, the entire measure receives an EIM score of 0 and no further calculation is required.

For illustrative purposes, assume that all Measurement Period Stratified Performance values exceed their MPRs to proceed with the calculation of the EIM Score.

3. Calculate the Category Inequity for each racial/ethnic group using reference group Z.

	Denominator		Stratified Performance		MDR	MPR	Category Inequity		Equity Weights
	Baseline	**Measurement Period**	**Baseline**	**Measurement Period**			**Baseline**	**Measurement Period**	
W	$n_{W,base}$	$n_{W,MP}$	$p_{W,base}$	$p_{W,MP}$	$r_{MDR}n_{W,base}$	Precalculated using $n_{t,base}$, $p_{t,base}$ for t $\in \{W, X, Y, Z\}$	$abs(p_{W,base} - p_{Z,base})$	$abs(p_{W,MP} - p_{Z,MP})$	w_{W-Z}
X	$n_{X,base}$	$n_{X,MP}$	$p_{X,base}$	$p_{X,MP}$	$r_{MDR}n_{X,base}$		$abs(p_{X,base} - p_{Z,base})$	$abs(p_{X,MP} - p_{Z,MP})$	w_{X-Z}
Y	$n_{Y,base}$	$n_{Y,MP}$	$p_{Y,base}$	$p_{Y,MP}$	$r_{MDR}n_{Y,base}$		$abs(p_{Y,base} - p_{Z,base})$	$abs(p_{Y,MP} - p_{Z\ MP})$	w_{Y-Z}
Z	$n_{Z,base}$	$n_{Z,MP}$	$p_{Z,base}$	$p_{Z,MP}$	$r_{MDR}n_{Z,base}$		NA	NA	NA
Baseline Weighted Average Inequity					$w_{W-Z}abs(p_{W,base} - p_{Z,base}) + w_{X-Z}abs(p_{X,base} - p_{Z,base}) + w_{Y-Z}abs(p_{Y,base} - p_{Z,base})$				
Weighted Average Inequity					$w_{W-Z}abs(p_{W,MP} - p_{Z,MP}) + w_{X-Z}abs(p_{X,MP} - p_{Z,MP}) + w_{Y-Z}abs(p_{Y,MP} - p_{Z,MP})$				

4. Multiply the Category Inequity by the Equity Weights to yield the Weighted Average Inequity.

5. Use the Baseline Weighted Average Inequity and Weighted Average Inequity to calculate the percent reduction in weighted baseline inequity (PRW).

$$\text{PRW} = \frac{\text{Baseline Weighted Average Inequity } - \text{Weighted Average Inequity}}{\text{Baseline Weighted Average Inequity}}$$

6. Compute the EIM score: $\text{EIM Score} = 6.667$ x PRW when PRW $<$ 0.75 and EIM Score = 5 when PRW $\geq$ 0.75. If the EIM score is less than 0, set it equal to 0. If the EIM score is greater than 5, set it equal to 5. This is the EIM score formula when using imputed data; when self-reported race/ethnicity are available for substantial and similar proportions of members in the Baseline and Measurement Periods, a different EIM score formula may be applied. For more details on how the EIM score formula was determined, refer to Chapter 3.

Example:

Suppose that the data available at baseline are:

	Denominator		Stratified Performance	
	Baseline	**Measurement Period**	**Baseline**	**Measurement Period**
Asian	300		0.787	
Black	600		0.697	
Hispanic	500		0.623	
White	1800		0.728	

Using the baseline data, we can determine that 1) the reference group is White for this example because it has the largest baseline denominator and that 2) all four racial/ethnic groups are maintained in the calculation because the baseline denominator values are each greater than 90. Additionally, we assume this measure has been identified as eligible for an EIM score following the simulation procedure detailed in Chapter 2 and that $r_{MDR} = 0.8$.

Using this information, we can calculate the MDRs, the MPRs, the Baseline Category Inequities, the Equity Weights, and the Baseline Weighted Average Inequity.

	Denominator		Stratified Performance		MDR	MPR	Category Inequity		Equity Weights
	Baseline	**Measurement Period**	**Baseline**	**Measurement Period**			**Baseline**	**Measurement Period**	
Asian	300		0.787		240	0.728	0.059		0.258
Black	600		0.697		480	0.650	0.031		0.294
Hispanic	500		0.623		400	0.569	0.105		0.448
White	1800		0.728		1440	0.702	NA		NA
Baseline Weighted Average Inequity							0.071		
Weighted Average Inequity									

Once the Measurement Period data (Denominators, Stratified Performance) are available, Steps 1-6 in Section III can be completed.

	Denominator		Stratified Performance		MDR	MPR	Category Inequity		Equity Weights
	Baseline	**Measurement Period**	**Baseline**	**Measurement Period**			**Baseline**	**Measurement Period**	
Asian	300	308	0.787	0.784	240	0.728	0.059	0.054	0.258
Black	600	620	0.697	0.710	480	0.650	0.031	0.020	0.294
Hispanic	500	520	0.623	0.680	400	0.569	0.105	0.050	0.448
White	1800	1810	0.728	0.730	1440	0.702	NA	NA	NA
Baseline Weighted Average Inequity							0.071		
Weighted Average Inequity							0.042		

1. The measure is eligible to receive an EIM score because the Measurement Period Denominators are each greater than or equal to their corresponding MDRs.
2. Calculation of the EIM score proceeds because none of the Measurement Period Stratified Performance values are less than the corresponding MPRs.
3. The Category Inequity values are $0.054, 0.020, 0.050$.
4. The Weighted Average Inequity is 0.042.
5. The PRW is $(0.071 - 0.042)/0.071 = 0.408$
6. The EIM score is $0.408 * 6.667 = \mathbf{2.7}$ since $0.408 < 0.75$.

Chapter 2. Determining a Measure’s Eligibility for the EIM Score

The objective of Chapter 2 is to describe how we determine which subset of the Ambulatory Care Quality Incentive Measures (AIMs) are also eligible for an EIM score for each provider group. This step occurs before contracting.

Not all AIMs are eligible for an EIM score due to sample size and baseline inequity considerations (which together determine the magnitude of random error in—analogous to the reliability of—each EIM measurement). While an AIM score for a provider group is based on between-provider comparisons among all its members eligible for a measure, the EIM score stratifies all its members eligible for a measure by race/ethnicity and examines improvements in relative stratified performance rates between baseline and the measurement period. As a result, the EIM score is a comparison of rates that are based on considerably smaller denominator counts than the AIM score. This leads to increased variability in the EIM score relative to the AIM score for a specific provider. This chapter describes standards for determining which AIM measures satisfy certain minimum denominator counts and maximum estimation errors accepted to be eligible for inclusion in the set of EIMs for each provider group. The approach described here is analogous to the approach we use to ensure reliability for each AIM.

Section I. Overview

The subset of AIMs eligible for an EIM score for each provider group’s contract is selected based on measures that have enough members in each racial/ethnic subgroup to draw meaningful conclusions about the provider group’s performance for each racial/ethnic subgroup (Condition 1) and similar expected levels of error in calculating the EIM score as what is currently accepted in calculating the AIM score (Condition 2). To understand the expected levels of error in calculating the EIM score, a simulation is conducted for each provider group that considers a range of potential improvement scenarios given the baseline data for each measure, estimates the corresponding EIM score, and then determines the distance between the estimated EIM score and the known, true value of the EIM score. If the estimated EIM scores are close to their true value, then the error in calculating the EIM score for this measure is low. If the level of error for this EIM score is low enough, as determined by comparing this to the level of error accepted in calculating the AIM scores at their minimum denominators, then this AIM is eligible for an EIM score for the provider group.

We performed the simulation to determine the set of eligible EIMs using 2020 baseline data for multiple large provider groups. Summarizing the results across the provider groups shows that AIMs that are eligible for EIM scores tend to be measures that have large baseline racial/ethnic inequities and large denominators even when stratified by race/ethnicity.

Section II. Detailed Description

The set of measures eligible for an EIM score for a provider group is a subset of the AIMs applied during the Measurement Period and that are calculable (or already calculated) at baseline. To be eligible for an EIM score, each AIM must satisfy the following two conditions:

Condition 1. At least two racial/ethnic strata each have baseline denominator ≥ 90.

Condition 2. The measure has an average root mean squared error (RMSE) that is comparable or less than the RMSE tolerated by the AIMs at their MDRs after completing the simulations detailed in Section II 1-3[6].

If either of these two conditions is not satisfied for a provider group, then this AIM is not eligible for an EIM score for the provider group. Details regarding Condition 2 are outlined in Section II 1-3.

[6] Because the EIM score is based on within-provider improvements to inequities (i.e., the weighted average of differences between multiple proportions at two points in time), the reliability calculation method used for AIM scores (i.e., a comparison of single proportions between providers at one point in time) could not be used. Because both EIMs and AIMs are nonetheless on the same 0-5 scale, we wanted the magnitude of random error tolerated for EIM scores to be comparable to the magnitude of random error tolerated at the minimum denominator threshold for AIM scores. Further details are available in Section II.2.

1. Condition 2 – Simulation Set-Up

Condition 2 being satisfied implies that the amount of measurement error associated with an EIM score for the provider group is comparable or less than the amount of error currently tolerated by the AIMs scores. Because verification that Condition 2 is satisfied occurs before the Measurement Period, a Monte Carlo simulation examining a range of potential improvement patterns and Measurement Period Denominators is conducted. This simulation compares the true, known value of the EIM score for a specific assumed improvement pattern (g^*) to a set of drawn EIM scores obtained by taking m draws of the Measurement Period data from the corresponding binomial probability distribution $\{g_1, g_2, \dots, g_m\}$, given this assumed improvement pattern and Measurement Period denominators. Formally, this comparison is made for a specific set of Measurement Period Denominators and an assumed improvement pattern using $\text{RMSE} = \sqrt{\text{MSE}}$ where $\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}(g_i - g^*)^2 = (\bar{g} - g^*)^2 + \frac{1}{m}\sum_{i=1}^{m}(g_i - \bar{g})^2 = \text{Bias}^2 + \text{Variance}$. As a result, RMSE is a measure of error that quantifies the estimator's bias and variability. The metric from this simulation used to determine if Condition 2 is satisfied is the average RMSE, obtained by averaging the RMSE across a range of improvement scenarios for a fixed set of Measurement Period denominators.

Because Measurement Period data are not available, conducting the simulation for each provider group and calculating the average RMSE for each measure within each provider group requires certain assumptions regarding the Measurement Period Denominators and Measurement Period Stratified Performance rates:

For this explanation, assume that $\text{num}_{\text{RE}} = 4$ and so all racial/ethnic groups are part of the calculation. If this is not the case, these steps are completed in the same manner using the reduced set of racial/ethnic groups.

Assumption 1. Measurement Period denominators vary by a constant, r, relative to Baseline denominators:

- $n_{W,MP} = r(n_{W,base})$
- $n_{X,MP} = r(n_{X,base})$
- $n_{Y,MP} = r(n_{Y,base})$
- $n_{Z,MP} = r(n_{Z,base})$

If $r > 1$, then the denominators for all racial/ethnic groups in the Measurement Period are greater than at baseline; if $r < 1$, then the denominator value for all racial/ethnic groups in the Measurement Period is less than at baseline. In general, we consider r ranging between 0.70 to 1.20 because analyses comparing changes in denominator sizes between 2018 and 2019 showed this was a reasonable range for r. Specifically, these analyses involved computing $n_{t,2019}/n_{t,2018}$ for each $t \in \{\text{Asian}, \text{Black}, \text{Hispanic}, \text{White}\}$, measure, and provider group. Then, for each measure and each $t \in \{\text{Asian}, \text{Black}, \text{Hispanic}, \text{White}\}$, the minimum and maximum of $n_{t,2019}/n_{t,2018}$ across all provider groups was calculated. The average minimum [maximum] of $n_{t,2019}/n_{t,2018}$ across all measures was equal to 0.89 [1.15] for Asian members, 0.80 [0.89] for Black members, 0.90 [1.12] for Hispanic members, 0.82 [1.11] for White members. As a result, it seemed reasonable to assume the denominator change ratio r between the baseline and Measurement Period would usually fall within the 0.70 to 1.20 range[7].

Assumption 2. The simulation assumes that provider groups achieve equity improvements by improving Stratified Performance among the racial/ethnic groups for which baseline performance is worse by more than Stratified Performance among racial/ethnic groups for which baseline performance is better.

Specifically, provider groups first improve performance for the racial/ethnic group receiving the lowest baseline performance until it matches the performance for the racial/ethnic group receiving the third-highest performance.

[7] Although a constant denominator change ratio r across all racial/ethnic groups is a significant assumption, if a different constant had been selected for each racial/ethnic group, the range of simulation configurations to consider would have been too extensive. As a result, we kept r constant for all racial/ethnic groups to examine the most extreme case where all groups are experiencing declines in their denominators.

Provider groups then improve performance for these two racial/ethnic groups until the performance for each matches the performance for the racial/ethnic group receiving the second-highest performance. Finally, provider groups improve performance for these three racial/ethnic groups until the performance for each matches the performance of the racial/ethnic group receiving the highest performance (at which point zero inequities remain on this AIM)[8].

Here are the detailed improvement scenario steps for a measure in which higher values of p indicate better performance.

1) Using the set of Baseline Stratified Performance rates, $\{\mathrm{p_{W,base}}, \mathrm{p_{X,base}}, \mathrm{p_{Y,base}}, \mathrm{p_{Z,base}}\}$, sort these such that $\mathrm{p_{1,base}}$ is the racial/ethnic group receiving the highest performance, $\mathrm{p_{2,base}}$ is the racial/ethnic group receiving the second-highest performance, $\mathrm{p_{3,base}}$ is the racial/ethnic group receiving the third-highest performance, and $\mathrm{p_{4,base}}$ is the racial/ethnic group receiving the lowest performance.
2) Calculate the difference between the groups receiving the highest and lowest performances, $\mathrm{p_{1,base}} - \mathrm{p_{4,base}}$, rounded to the nearest thousandth. Divide this difference in proportions by 0.001 and add 1. The number obtained is equal to the number of rows in a matrix representing the potential combinations of Measurement Period Stratified Performance values given the assumed improvement pattern. The number of columns is equal to $\mathrm{num_{RE}}$.
3) For the column corresponding to the racial/ethnic group receiving the highest performance, assume that all entries of the matrix are equal to $\mathrm{p_{1,base}}$. This means that for this group, the Measurement Period Stratified Performance will be assumed to be equal to the Baseline Stratified Performance in all simulation settings.
4) For the racial/ethnic group receiving the lowest performance, its corresponding column in the matrix representing potential Measurement Period Stratified Performance should be equal to the sequence from $\mathrm{p_{4,base}}$ to $\mathrm{p_{1,base}}$, increasing by 0.001 in each row.
5) For the column corresponding to the group receiving the third highest performance, all rows for which the group receiving the lowest Stratified Performance has performance less than or equal to $\mathrm{p_{3,base}}$ should be equal to $\mathrm{p_{3,base}}$. However, once the group receiving the lowest Stratified Performance catches up to the group receiving the third highest performance, both groups should move together to improve their performance until they reach $\mathrm{p_{1,base}}$.
6) A similar pattern is assumed for the column corresponding to the racial/ethnic group receiving the second highest performance: once the groups receiving the lowest and third highest performance reach $\mathrm{p_{2,base}}$ performance, the Measurement Period Stratified Performance for the groups receiving the 2nd, 3rd and 4th highest performances are assumed to improve up until they reach that of the group receiving the highest performance.

[8] This orderly improvement path may not be the exact one a provider group may take; the objective of the simulation was to examine variability across different levels of improvement. It is not possible to know exactly how any provider group might structure its improvement efforts in a new program. But for the purpose of the simulation, we needed to make an assumption about the pattern of improvement. Because communication with provider groups will incentivize them to focus on improving large baseline inequities or inequities with large Equity Weights, we felt it was reasonable to understand variability for this improvement pattern. We did consider a different improvement path in which each inequity is reduced at a fixed percentage (ex. 10%, 20%, …). Compared to the selected improvement path, the RMSEs resulting from this alternative improvement path were very similar.

The matrix below illustrates the full set of potential Measurement Period Stratified Performance (MPSP in the tables) values, given this assumed pattern of improvement. At the first row of this matrix, the EIM score is 0 because all Measurement Period Stratified Performance values are equal to their baseline values (and so inequities stay equal as well). In the last row of this matrix, the EIM score is equal to 5 because there are no inequities remaining.

Number of rows is equal to $\left(\left(p_{1,\text{base}} - p_{4,\text{base}}\right)/0.001\right) + 1$

Potential MPSP for Group Receiving Highest Performance	Potential MPSP for Group Receiving 2nd Highest Performance	Potential MPSP for Group Receiving 3rd Highest Performance	Potential MPSP for Group Receiving Lowest Performance
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{3,\text{base}}$	$p_{4,\text{base}}$
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{3,\text{base}}$	$p_{4,\text{base}} + 0.001$
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{3,\text{base}}$	$p_{4,\text{base}} + 0.002$
…	…	…	…
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{3,\text{base}}$	$p_{3,\text{base}}$
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{3,\text{base}} + 0.001$	$p_{3,\text{base}} + 0.001$
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{3,\text{base}} + 0.002$	$p_{3,\text{base}} + 0.002$
…	…	…	…
$p_{1,\text{base}}$	$p_{2,\text{base}}$	$p_{2,\text{base}}$	$p_{2,\text{base}}$
$p_{1,\text{base}}$	$p_{2,\text{base}} + 0.001$	$p_{2,\text{base}} + 0.001$	$p_{2,\text{base}} + 0.001$
$p_{1,\text{base}}$	$p_{2,\text{base}} + 0.002$	$p_{2,\text{base}} + 0.002$	$p_{2,\text{base}} + 0.002$
…	…	…	…
$p_{1,\text{base}}$	$p_{1,\text{base}}$	$p_{1,\text{base}}$	$p_{1,\text{base}}$

Number of columns is equal to num_{RE}

The three examples below show how the simulation is set up for specific values of r and the assumed improvement pattern. Each example follows the steps outlined in Chapter 1 Section III.

Example:

Let $r = 1.0$ and suppose that the data available at baseline for the provider group are:

	Denominator		Stratified Performance		MDR	MPR	Category Inequity		Equity Weights
	Baseline	Measurement Period	Baseline	Measurement Period			Baseline	Measurement Period	
Asian	180		0.650		144	0.565	0.100		0.657
Black	150		0.720		120	0.632	0.030		0.343
Hispanic	85		0.680		NA	NA	NA		NA
White	2000		0.750		1600	0.727	NA		NA
Baseline Weighted Average Inequity							0.076		
Weighted Average Inequity									

Note that White is the reference group in this example because its baseline denominator is the largest. Also, because $n_{\text{Hispanic,base}} < 90$, Hispanic members are excluded from calculations.

1. Because $r = 1.0$, $n_{\text{White,MP}} = n_{\text{White,base}}$, $n_{\text{Black,MP}} = n_{\text{Black,base}}$, $n_{\text{Asian,MP}} = n_{\text{Asian,base}}$, so the Measurement Period Denominators are each greater than the MDR.
2. Based on the assumed improvement pattern, the Measurement Period Stratified Performance rates are each greater than the MPRs across all combinations of the matrix below.

Set-up step: The White racial/ethnic group receives the highest performance, the Asian racial/ethnic group receives the lowest performance.

Number of rows is equal to 101

Potential MPSP for Group Receiving Highest Performance White	**Potential MPSP for Group Receiving 2nd Highest Performance** Black	**Potential MPSP for Group Receiving Lowest Performance** Asian
0.750	0.720	0.650
0.750	0.720	$0.650 + 0.001$
0.750	0.720	$0.650 + 0.002$
…	…	…
0.750	0.720	0.720
0.750	$0.720 + 0.001$	$0.720 + 0.001$
0.750	$0.720 + 0.002$	$0.720 + 0.002$
…	…	…
0.750	0.750	0.750

Number of columns is equal to $\mathrm{num}_{\mathrm{RE}} = 3$

We can then proceed with Steps 3-6 of Chapter 1 Section III to calculate the EIM score for each row of the potential Measurement Period Stratified Performance rates and other r values.

Example:

Let $\mathrm{r} = 1.2$ and suppose that the data available at baseline for the provider group are:

	Denominator		**Stratified Performance**		**MDR**	**MPR**	**Category Inequity**		**Equity Weights**
	Baseline	**Measurement Period**	**Baseline**	**Measurement Period**			**Baseline**	**Measurement Period**	
Asian	180		0.850		144	0.784	0.100		0.418
Black	150		0.790		120	0.707	0.040		0.235
Hispanic	250		0.800		200	0.737	0.050		0.347
White	2000		0.750		1600	0.726	NA		NA
Baseline Weighted Average Inequity							0.069		
Weighted Average Inequity									

Note that White is the reference group in this example because its baseline denominator is the largest. All four racial/ethnic groups are included in calculations because the baseline denominators for each are greater than 90.

1. Because $\mathrm{r} = 1.2$, $\mathrm{n}_{\mathrm{Asian,MP}} = 1.2\mathrm{n}_{\mathrm{Asian,base}}$, $\mathrm{n}_{\mathrm{Black,MP}} = 1.2\mathrm{n}_{\mathrm{Black,base}}$, $\mathrm{n}_{\mathrm{Hispanic,MP}} = 1.2\mathrm{n}_{\mathrm{Hispanic,base}}$, $\mathrm{n}_{\mathrm{White,MP}} = 1.2\mathrm{n}_{\mathrm{White,base}}$, so the Measurement Period Denominators are each greater than the MDR.
2. Based on the assumed improvement pattern, the Measurement Period Stratified Performance rates are each larger than the MPRs across all combinations of the matrix below.

Set-up step: The Asian racial/ethnic group receives the highest performance, the White racial/ethnic group receives the lowest performance.

Potential MPSP for Group Receiving Highest Performance Asian	Potential MPSP for Group Receiving 2nd Highest Performance Hispanic	Potential MPSP for Group Receiving 3rd Highest Performance Black	Potential MPSP for Group Receiving Lowest Performance White
0.850	0.800	0.790	0.750
0.850	0.800	0.790	$0.750 + 0.001$
0.850	0.800	0.790	$0.750 + 0.002$
…	…	…	…
0.850	0.800	0.790	0.790
0.850	0.800	$0.790 + 0.001$	$0.790 + 0.001$
0.850	0.800	$0.790 + 0.002$	$0.790 + 0.002$
…	…	…	…
0.850	0.800	0.800	0.800
0.850	$0.800 + 0.001$	$0.800 + 0.001$	$0.800 + 0.001$
0.850	$0.800 + 0.002$	$0.800 + 0.002$	$0.800 + 0.002$
…	…	…	…
0.850	0.850	0.850	0.850

Number of rows is equal to 101

Number of columns is equal to $\mathrm{num}_{\mathrm{RE}} = 4$

We can then proceed with Steps 3-6 of Chapter 1 Section III to calculate the EIM score for each row of the potential Measurement Period Stratified Performance rates and other values of r.

Example:

Let $\mathrm{r} = 0.8$ and suppose that the data available at baseline for the provider group are:

	Denominator		Stratified Performance		MDR	MPR	Category Inequity		Equity Weights
	Baseline	Measurement Period	Baseline	Measurement Period			Baseline	Measurement Period	
Asian	180		0.850		144	0.784	0.100		0.418
Black	150		0.790		120	0.707	0.040		0.235
Hispanic	250		0.700		200	0.628	0.050		0.347
White	2000		0.750		1600	0.726	NA		NA
Baseline Weighted Average Inequity							0.069		
Weighted Average Inequity									

Note that White is the reference group in this example because its baseline denominator is the largest. All four racial/ethnic groups are included in calculations because the baseline denominators for each are greater than 90.

1. Because $\mathrm{r} = 0.8$, $\mathrm{n}_{\mathrm{Asian,MP}} = 0.8\mathrm{n}_{\mathrm{Asian,base}}$, $\mathrm{n}_{\mathrm{Black,MP}} = 0.8\mathrm{n}_{\mathrm{Black,base}}$, $\mathrm{n}_{\mathrm{Hispanic,MP}} = 0.8\mathrm{n}_{\mathrm{Hispanic,base}}$ $\mathrm{n}_{\mathrm{White,MP}} = 0.8\mathrm{n}_{\mathrm{White,base}}$, so the Measurement Period Denominators are each equal to the MDR. We can proceed with the calculation because none of the Measurement Period Denominators are less than their corresponding MDRs.
2. Based on the assumed improvement pattern, the Measurement Period Stratified Performance rates are each larger than the MPRs across all combinations of the matrix below.

Set-up step: The Asian racial/ethnic group receives the highest performance, the Hispanic racial/ethnic group receives the lowest performance.

Number of rows is equal to 151

Potential MPSP for Group Receiving Highest Performance Asian	Potential MPSP for Group Receiving 2nd Highest Performance Black	Potential MPSP for Group Receiving 3rd Highest Performance White	Potential MPSP for Group Receiving Lowest Performance Hispanic
0.850	0.790	0.750	0.700
0.850	0.790	0.750	$0.700 + 0.001$
0.850	0.790	0.750	$0.700 + 0.002$
…	…	…	…
0.850	0.790	0.750	0.750
0.850	0.790	$0.750 + 0.001$	$0.750 + 0.001$
0.850	0.790	$0.750 + 0.002$	$0.750 + 0.002$
…	…	…	…
0.850	0.790	0.790	0.790
0.850	$0.790 + 0.001$	$0.790 + 0.001$	$0.790 + 0.001$
0.850	$0.790 + 0.002$	$0.790 + 0.002$	$0.790 + 0.002$
…	…	…	…
0.850	0.850	0.850	0.850

Number of columns is equal to $\mathrm{num}_{\mathrm{RE}} = 4$

We can then proceed with Steps 3-6 of Chapter 1 Section III to calculate the EIM score for each row of the potential Measurement Period Stratified Performance rates and other values of r between 0.70 to 1.20.

2. Condition 2 – Implementing the Simulation

For each AIM that satisfies Condition 1, the Monte Carlo simulation to determine if Condition 2 is satisfied for the provider group proceeds as follows:

i. Complete the relevant baseline calculations detailed in Chapter 1, Section II.1.

ii. For each of the rows in the matrix of potential Measurement Period Stratified Performance rates and for each value of $\mathrm{r} \in \{0.7, 0.8, 0.9, 1.0, 1.1, 1.2\}$, do the following:

   a. Calculate the Measurement Period Denominators for this value of r.

   b. Use the Baseline and Measurement Period Denominators and Stratified Performance to calculate the true EIM score, g^*, following the procedure detailed in Chapter 1, Section III.

   c. Use the Measurement Period denominators and Stratified Performance to draw Measurement Period data for each racial/ethnic stratum based on its probability distribution. We set the number of replications to $\mathrm{m} = 20{,}000$.

      For example, draw $20{,}000$ draws of $\mathrm{Binomial}(\mathrm{n}_{\mathrm{W,MP}}, \mathrm{p}_{\mathrm{W,MP}})$, $20{,}000$ draws of $\mathrm{Binomial}(\mathrm{n}_{\mathrm{X,MP}}, \mathrm{p}_{\mathrm{X,MP}})$, $20{,}000$ draws of $\mathrm{Binomial}(\mathrm{n}_{\mathrm{Y,MP}}, \mathrm{p}_{\mathrm{Y,MP}})$, and 20,0000 draws of $\mathrm{Binomial}(\mathrm{n}_{\mathrm{Z,MP}}, \mathrm{p}_{\mathrm{Z,MP}})$.

   d. For each set of draws, in combination with the observed baseline denominators and Stratified Performance Rates, calculate the drawn EIM score g_i following the procedure in Chapter 1, Section III.

   e. Compare the $\mathrm{m} = 20{,}000$ draws of the EIM score, $\{\mathrm{g}_1, \mathrm{g}_2, \dots, \mathrm{g}_{20{,}000}\}$, to the true EIM score from step ii b using $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$ where $\mathrm{MSE} = \frac{1}{\mathrm{m}}\sum_{\mathrm{i}=1}^{\mathrm{m}}(\mathrm{g}_\mathrm{i} - \mathrm{g}^*)^2$.

iii. Repeat step ii for each of the remaining potential combinations of Measurement Period Stratified Performance and all potential r values.

iv. Using the RMSE simulation metric recorded in step ii e for each of the combinations of r and the potential Measurement Period Stratified Performance rates (while excluding any combinations that lead to above a 95% reduction in PRW, since the objective was to learn about realistic improvements in equity that would take place in the initial years of the program), summarize the results by calculating the mean RMSE for each fixed value of

r. For a specific value of r, this is equal to the mean of the RMSEs from each of the combinations of Measurement Period Stratified Performance.

3. Using Simulation Results

If the average RMSE at any value of $\mathrm{r} \leq 1$ is comparable in magnitude to the RMSE tolerated by the AIMs at their MDRs, then this measure is eligible for equity calculations for the provider group because this means that the degree of error associated to this EIM is similar in magnitude to the degree of error tolerated in AIMs scores. Among measures that are eligible, the MDRs for that measure, $\mathrm{r}_{\mathrm{MDR}}(\mathrm{n}_{\mathrm{t,base}})$, are calculated by multiplying baseline denominators by the smallest value of $\mathrm{r} = 0.7, 0.8, 0.9, 1.0$ associated to the RMSE tolerated by the AIMs at their MDRs.

The RMSE tolerated by the AIMs at their MDRs was calculated based on a separate Monte Carlo simulation examining the degree of uncertainty currently accepted in the AIM scores. Each measure that receives an AIM score has a Minimum Denominator Required, Minimum Threshold and Upper Threshold. For each Provider and each of the measures that receives an AIM score, the Actual Provider Performance Level ($\mathrm{p}_{\mathrm{AIM}}$) and the AIM Minimum Denominator Required were used to draw 1000 draws from the binomial distribution and obtain a set of 1000 drawn performance rates, $\mathrm{p}_{\mathrm{AIM,i}}$ for $\mathrm{i} = 1, \ldots, 1000$. These drawn performance rates were then converted to a score between 0 and 5 (similar to the EIM score formula) by letting $\mathrm{PRW}_{\mathrm{AIM,i}} = \frac{\mathrm{p}_{\mathrm{AIM,i}} - \text{Minimum Threshold}}{\text{Upper Threshold} - \text{Minimum Threshold}}$ and using the formula: Score=6.667($\mathrm{PRW}_{\mathrm{AIM,i}}$) when $\mathrm{PRW}_{\mathrm{AIM,i}} < 0.75$ and Score = 5 when $\mathrm{PRW}_{\mathrm{AIM,i}} \geq 0.75$. If $\mathrm{p}_{\mathrm{AIM,i}}$ is less than the Minimum Threshold, then the Score is replaced with 0; if $\mathrm{p}_{\mathrm{AIM,i}}$ is greater than the Upper Threshold then the Score is replaced with 5. The 1000 drawn Scores were compared to the true Score (derived using the observed $\mathrm{p}_{\mathrm{AIM}}$ and the Score formula) using the RMSE. The mean RMSE was then calculated across all groups and all measures to inform the mean RMSE threshold used to determine a measure’s eligibility for an EIM score.

Chapter 3. Determining the Curve Shape for the EIM Score

In Step 6 of the procedure for calculating the EIM Score (Chapter 1, Section III), the formula for transforming the percent reduction in weighted baseline inequity (PRW) to an EIM Score is provided.

We set the conversion from PRW to EIM score for now to be generous to provider groups because race/ethnicity data are fully imputed and therefore contain measurement errors. The EIM Score = 6.667(PRW) when PRW $<$ 75% and EIM Score = 5 when PRW $\geq$ 75% (Figure C1). In other words, rather than require the PRW to be 100% to earn an EIM score equal to 5, its maximum, the PRW only needs to be 75% or more. This approach makes it easier to earn the maximum payout to account for the use of fully imputed race/ethnicity data. The Plan will consider increasing this threshold as more self-reported data become available.

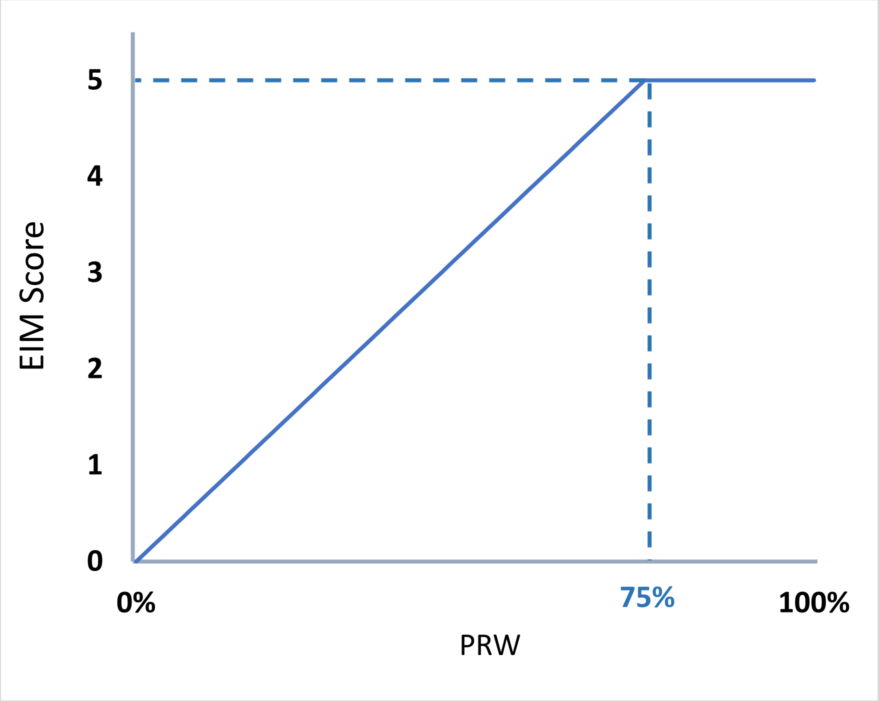

Figure C1. Relationship Between PRW, EIM Score

The decision to transform any PRW score greater than or equal to 75% (or equivalently 0.75) to an EIM Score of 5 was based on an analysis that examined book-of-business data for a subset of measures with some self-reported race/ethnicity data available. For this group of measures and using only members with self-reported race/ethnicity, we modeled a scenario in which the Plan was using fully imputed race/ethnicity data to determine baseline weighted inequity values and to track improvement, while the provider organization (under the assumption that the provider organization will have 100% self-reported race/ethnicity data during the Measurement Period) is using only true self-reported race/ethnicity data to guide and track internal improvement efforts. In this scenario, the Plan's calculations might not match the true PRW values.

For each measure, our analyses compared how simulated improvements to the true PRW calculated using self-reported race/ethnicity translated to changes to the PRW that is calculated using imputed race/ethnicity. We found that as the PRW calculated using only self-reported race/ethnicity data approached 100%, the PRW calculated using only imputed data approached 75% across most measures examined.

As an illustrative example, results for this analysis for a group of 22,184 members for the cervical cancer screening measure are shown in Figure C2. In other words, if the PO has internal access to self-reported data and is using this to guide their improvement, as they come close to achieving a PRW of 100%, the Plan, which is using imputed data, estimates the PRW to be closer to about 75%. Because the Plan is limited to using imputed race/ethnicity due to high levels of missingness for self-reported race/ethnicity data, the curve connecting the PRW to the EIM Score has been transformed to account for this underestimation. If 100% self-reported race/ethnicity data were available to the Plan and used to calculate the PRW, the curve displayed in Figure C1 would be closer to a straight line between 0 and 5. However, because imputed race/ethnicity is being used, the line currently displayed in Figure C1 reaches an EIM Score of 5 at PRW=75% and has a steeper slope between a PRW of 0 and 75% than it would if it were a straight line between 0 and 100%.

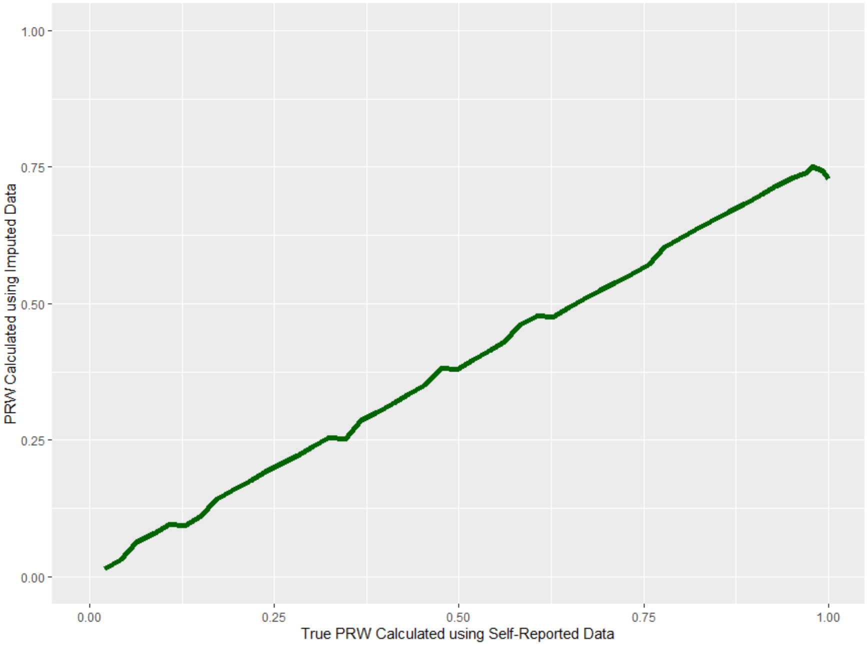

Figure C2. Analysis among Members eligible for Cervical Cancer Screening who Reported Race and Ethnicity: Comparing PRW Calculated using the Self-Reported Data When the Provider Uses Self-Reported Data to Guide Improvement Efforts with Corresponding PRW Calculated using Imputed Data