



Overview of the 2019 HHS-RADV White Paper

Center for Consumer Information & Insurance Oversight (CCIIO)

Centers for Medicare & Medicaid Services (CMS)

Department of Health and Human Services (HHS)

December 2019



Agenda

- Purpose & Background
- Enrollee Sampling
- Outlier Detection
- Error Rate Calculation
- Application of HHS-RADV Results
- Next Steps

Purpose & Background

Purpose & Background

Patient Protection and Affordable Care Act (PPACA):

- Section 1343: Established a risk adjustment (RA) program to provide payments to health insurance issuers that attract high-risk enrollees to reduce the incentives for issuers to avoid those enrollees, and to lessen the potential influence of risk selection on the premiums that issuers charge
- Section 1321(c)(1): the Department of Health and Human Services (HHS) is responsible for operating the program on behalf of any states that do not elect to do so

Purpose & Background

To ensure the integrity of the RA program:

- CMS performs risk adjustment data validation (HHS-RADV) to validate the accuracy of data submitted by issuers for the purposes of RA transfer calculations

HHS-RADV Regulations:

- 45 C.F.R § 153.350: RADV Standards for a RA program
- 45 C.F.R. § 153.630: Requirements for HHS-RADV for HHS-operated RA

Purpose & Background

HHS-RADV:

- Serves as an audit of the information used in establishing an enrollee's risk score for purposes of calculating the issuer's plan liability risk score (PLRS) under the risk adjustment (RA) program
- Uses a multi-step process called error estimation to calculate error rates that are used to adjust outlier issuers' risk scores and RA transfers for the applicable state market risk pool(s)

Purpose & Regulatory Background



6 Steps to HHS-RADV:

1. Select a sample of an issuer's enrollees
2. Conduct the initial validation audit (IVA)
3. Conduct the second validation audit (SVA)
4. Use the IVA and SVA findings to determine error estimation
5. Allow discrepancies and appeals
6. Apply HHS-RADV results to RA transfers

Purpose & Background

Error Estimation Process

- Categorize Hierarchical Condition Categories (HCCs) for all participating issuers by sampled EDGE server data & IVA/SVA sample results (Low/Medium/High)

- Calculate national metrics (weighted mean, confidence intervals) to determine which issuers are outliers in each HCC grouping

- Determine group adjustment factor for outlier issuers
- Adjust EDGE risk score factors for outlier issuers
- Determine Error Rate for outlier issuers

Categorize
HCCs into
Groups



Determine
Outliers



Calculate
Error Rates

Purpose & Background

- 2015 & 2016 Benefit Years HHS-RADV were pilot years
- 2017 Benefit Year and beyond HHS-RADV will be used to adjust RA Transfers

Purpose & Background

- Discussion Paper Purpose: is to outline and seek feedback on certain HHS-RADV issues:
 - Enrollee Sampling
 - Outlier Detection
 - Error Rate Calculation
 - Application of HHS-RADV Error Rates
- Comments on the options outlined in this paper will help inform potential future rulemaking

Purpose & Background

- Paper Options:
 - Were developed based on:
 - HHS's ongoing internal analysis of potential refinements to the HHS-RADV program for future benefit years
 - Comments received on HHS-RADV through notice-and-comment rulemaking and through listening sessions with stakeholders
 - Were mostly tested using 2017 Benefit Year HHS-RADV data
 - Will continue to be tested to inform any potential future rulemaking

Enrollee Sampling

Enrollee Sampling

45 C.F.R § 153.350(a): Requires states, or HHS on behalf of states, to validate a statistically valid sample of risk adjustment data submitted by issuers each year

Enrollee Sampling

Goals for HHS-RADV sample size refinement:

- Ensure samples accurately represent issuer enrollee populations
- Increase the number of samples that meet the 10 percent precision target for a two-sided 95 percent confidence interval
- Minimize the administrative and financial burden on issuers

Enrollee Sampling

Metrics to evaluate sample size:

Precision	Measurement of how close in value sampled observations are likely to be to one another. Refers to the dispersion of a set of observations.
Accuracy	Property of being close to a target or true value. Measures how well the sample measurements match the true population value.

Enrollee Sampling

Current Initial Validation Audit (IVA) Sample Sizes:

Issuer Population Size (N)	IVA Sample Size (n)
$N \geq 4,000$	$n = 200$
$50 \leq N < 4,000$	$n = 200 * \text{Finite Population Correction (FPC)}$ $\text{FPC} = (N - 200) / N$ If $(200 * \text{FPC}) < 50$, $n = 50$
$N < 50$	$n = N$

HHS chose a sample size of 200 enrollees for most issuers based on sample size precision analyses conducted using proxy risk score data from the Medicare Advantage RADV (MA-RADV) program.

Enrollee Sampling

3 criteria currently used to help identify small issuers:

1. **Total annual premiums:** Beginning with 2018 Benefit Year HHS-RADV, issuers at or below the \$15 million premium materiality threshold only have an IVA approximately every three years (barring any risk-based triggers that warrant more frequent audits).
2. **Enrollee population:** Issuers with enrollee populations below 4,000 have smaller sample sizes.
3. **Billable member months:** Issuers with 500 or fewer billable member months are exempt from HHS-RADV.

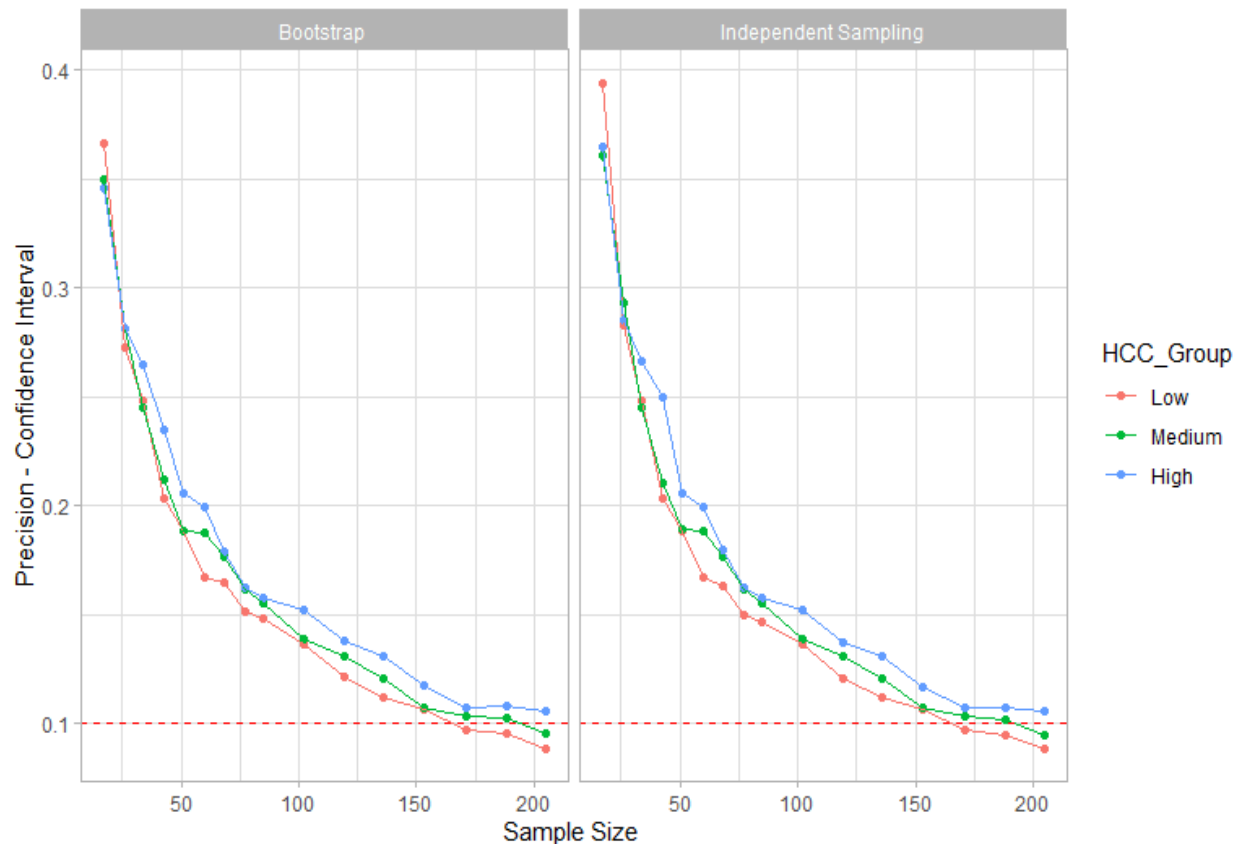
Enrollee Sampling

- Stratification of a population prior to sampling and selecting more cases from strata with greater variance can increase the likelihood that the sample achieves targeted levels of confidence and precision relative to a simple random sample for which no stratification is performed.
- HHS calculates the individual sample size per stratum using the Neyman optimal allocation method.

HCC Stratum	Age	Risk Level	Stratum
1 or More HCC(s)	Adult	Low	1
		Medium	2
		High	3
	Child	Low	4
		Medium	5
		High	6
	Infant	Low	7
		Medium	8
		High	9
No HCCs	All	N/A	10

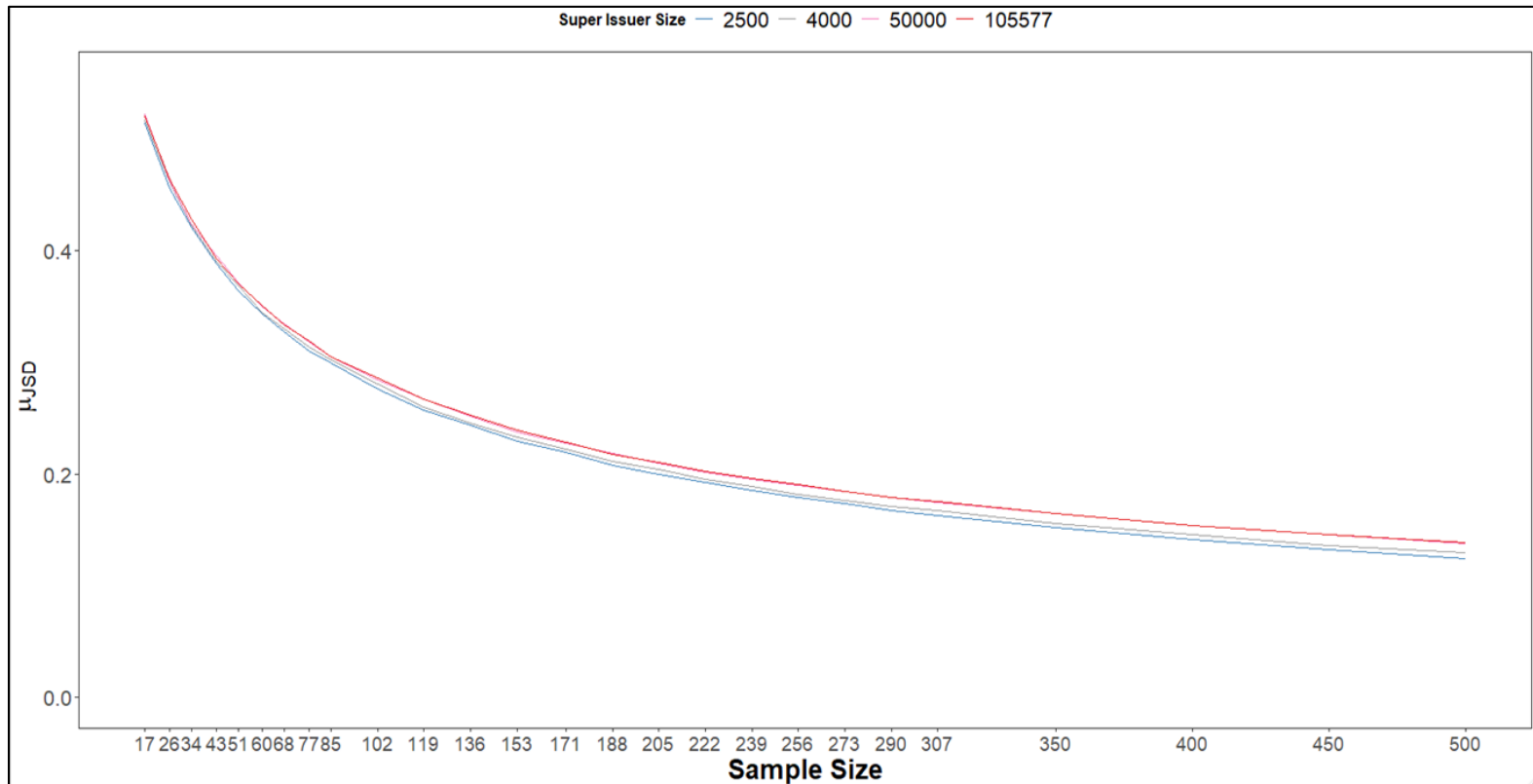
Enrollee Sampling

Precision improves (decreases in value) as sample size increases, and the current sample size of 200 enrollees can achieve the 10 percent precision target.



Enrollee Sampling

When comparing the probability of finding specific HCCs between samples and simulated populations at different sample sizes, there are small marginal gains in the alignment of the sample and simulated population HCC frequency distributions beyond a sample of 200 enrollees.



Enrollee Sampling

Options Explored:

1. Vary sample size based on issuers' distance from the HCC group failure rate outlier threshold and precision.
2. Re-evaluate the standard sample size using national average HHS-RADV error rates instead of proxy data from MA-RADV.
3. Consider other sampling options and measures to reduce burden on issuers with small populations

In response to large issuers' requests for larger sample sizes, HHS is also considering allowing issuers to elect larger sample sizes.

Enrollee Sampling

Option	Pros	Cons
1. Vary sample size based on HCC group failure rates and precision	<ul style="list-style-type: none"> - Larger samples could improve precision and/or accuracy - Opportunity to retrieve more accurate and complete medical records 	<ul style="list-style-type: none"> - Some issuers may not have enough enrollees with HCCs from which to sample to meaningfully improve precision or accuracy - Requires using data from 2 years prior
2. Use national average HHS-RADV instead of MA-RADV data	<ul style="list-style-type: none"> - More representative data from HHS-RADV issuers 	<ul style="list-style-type: none"> - May want to wait until we have more years of HHS-RADV error rate data - Requires using data from 2 years prior
3. Require a sample size of 200 or alternative for issuers with small populations	<ul style="list-style-type: none"> - Larger samples could improve precision and/or accuracy - Opportunity to retrieve more accurate and complete medical records - Potential new exemption for small issuers would reduce burden 	<ul style="list-style-type: none"> - Calculated cutoff value for sample size of 200 based on 1 year of HHS-RADV data and MA-RADV data only - Potential for gaming under exemption
Allow issuers to elect larger sample sizes	<ul style="list-style-type: none"> - Customized sample sizes 	<ul style="list-style-type: none"> - Increasing sample size may not meaningfully improve precision and accuracy

Outlier Detection

Outlier Detection

- Issue 1: Examines alternative methodologies to identify which issuers, if any, have failure rates that are very different from the national average
- Issue 2: Examines alternative methodologies to account for HCC hierarchies in identifying outliers

Outlier Detection – Issue 1

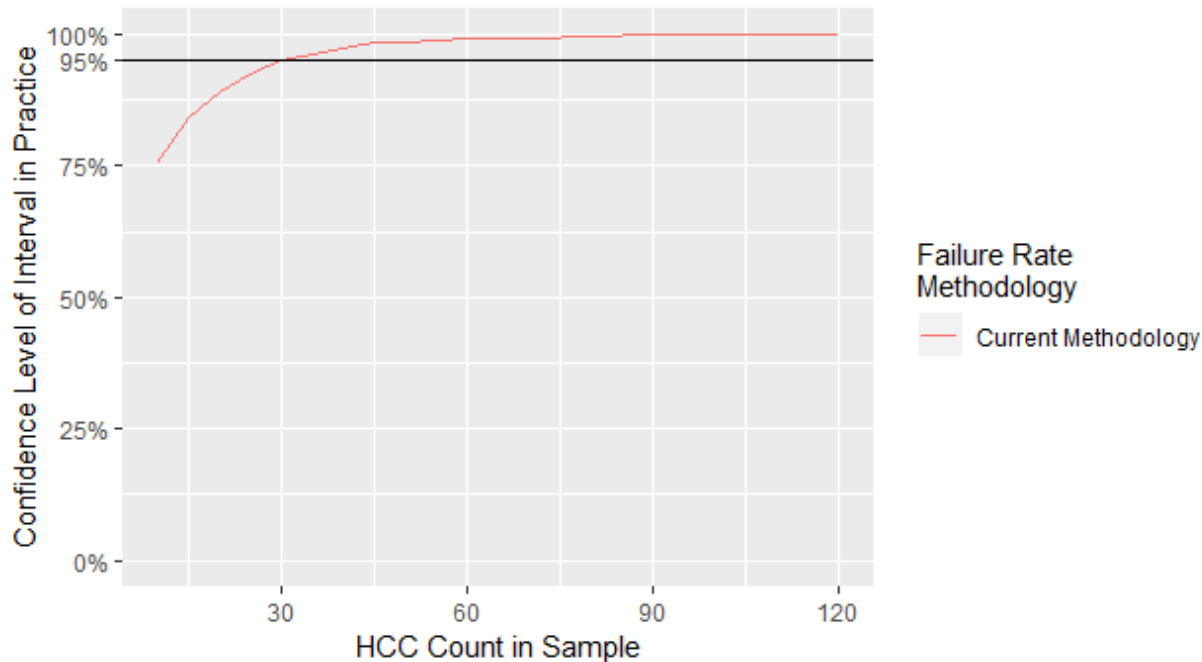
- The current methodology determines an issuer's outlier status based on national, static, confidence intervals common to all issuers
- Issue 1: Current methodology does not adjust for issuer HCC count and may lead to:
 1. Some issuers appearing to be outliers, although their population-level failure rates are indistinguishable from the national average
 2. Some issuers with population-level failure rates very far from the national mean could have sample failure rates that fall within the national confidence interval

Outlier Detection – Issue 1

Confidence Level	Theoretical probability that an issuer whose population-level failure rate for an HCC group is very similar to the national mean will <i>not</i> be found to be an outlier, given that all statistical assumptions about the underlying distribution are upheld.
Practical Confidence Level	Simulated, empirical probability that an issuer whose population-level failure rate for an HCC group is very similar to the national mean will <i>not</i> be found to be an outlier, given possible violations to statistical assumptions about the underlying distribution that may be present in actual HHS-RADV data.

Outlier Detection – Issue 1

Current Methodology: Sample is based on enrollee count, but outlier status is based on HCC counts



- Due to random chance, fewer HCCs may appear in one sample than are expected and necessary to satisfy assumptions of the methodology
- An HCC grouping count of <30 HCCs in a sample reduces the practical confidence level below the 95% theoretical value

Outlier Detection – Issue 1

Options Explored:

1. Establish multiple sets of national confidence intervals based on issuer HCC count
2. Use issuer-specific bootstrapped confidence intervals
3. Use issuer-specific confidence intervals based on the binomial distribution
4. Use issuer-specific confidence intervals based on McNemar's Test
5. Use issuer-specific confidence intervals based on Bayesian Methods
6. Determine outlier status through machine learning methods

Outlier Detection – Issue 2

- The current methodology allows for HCC hierarchies in RA to be combined or split across HCC failure rate groupings in HHS-RADV
- Issue 2: HCC hierarchies in RA and HCC failure rate groupings in HHS-RADV can interact in a number of ways that could lead to lower or higher risk score adjustments than may be warranted by the individual HCC and within-hierarchy risk score changes

Outlier Detection– Issue 2

<p>HCC Failure Rate Grouping</p>	<p>The high, medium, and low groupings of all HCCs in HHS-RADV based on the individual HCC failure rates. Determines the confidence intervals for the HCC grouping at the national level.</p>
<p>HCC Hierarchy</p>	<p>RA uses HCCs to estimate a risk score for each enrollee in issuer's RA population that is used to calculate the issuer's plan liability risk score that is used in the RA state payment transfer formula. Clinically similar HCCs are placed in a hierarchy and are grouped together in the HHS RA model, and are constrained within-hierarchy to have either the same risk score factor, or to have explicitly increasing risk scores with increases in severity.</p>

Outlier Detection– Issue 2

Examples of Interaction of HCC Hierarchies and HCC Groupings

Examples	Result
Hierarchy of HCCs w/ unequal coefficients; Different HCC failure rate groupings	Adjustment may only capture a part of risk score error of enrollees who have one HCC recoded as another during HHS-RADV
Hierarchy of HCCs w/ unequal coefficients; Same HCC failure rate grouping	Adjustment may not capture any of the risk score error of enrollees who have one HCC recoded as another during HHS-RADV
Hierarchy of HCCs w/ equal coefficients; Different HCC failure rate grouping	Adjustment may reflect a risk score error that is not present when considering that the HCCs in question have the same coefficient
Hierarchy of HCCs w/ equal coefficients; Same HCC failure rate grouping	Adjustment may be unaffected by any recoding between EDGE and audit data

Outlier Detection– Issue 2

Options Explored:

- Assess ordinal-by-ordinal relationships
- Assess the statistical significance of issuer's pre- and post-RADV difference in risk score directly, rather than through an HCC-count-based failure rate metric

Error Rate Calculation

Error Rate Calculation

- Issue 1: Examines alternative adjustment thresholds for calculating error rates for issuers that are just outside of the acceptable range of variation (the “payment cliff” or “leap frog effect”)
- Issue 2: Examines potential approaches to mitigate the impact of HHS-RADV adjustments due to negative error rate issuers with negative failure rates

Error Rate Calculation

<p>Error Rate</p>	<p>The rate at which an outlier issuer’s risk score is adjusted based on HHS-RADV results. If an issuer is identified as a group failure rate outlier in one or more HCC groups, its individual enrollee risk scores are adjusted based on the differences between the issuer’s group failure rate and the national mean HCC group failure rate in every HCC group in which the issuer is identified as an outlier. The issuer’s error rate equals</p> $1 - \frac{\textit{stratum weighted sum of adjusted enrollee risk scores}}{\textit{sum of original EDGE enrollee risk scores}}$
<p>Failure Rate</p>	<p>The rate at which the frequency of HCCs identified through the IVA or SVA differ from the frequency of the HCCs identified on EDGE. Failure rate equals</p> $1 - \frac{\textit{HCC frequency in IVA or SVA}}{\textit{HCC frequency on EDGE}}$ <p>During HHS-RADV error estimation, the HCC failure rate is calculated for each HCC in order to determine national mean HCC groups. Then, each issuer’s failure rate is calculated for each HCC group to determine whether the issuer is a group failure rate outlier, which would lead to a non-zero error rate and an adjustment to the issuer’s enrollee risk scores.</p>

Error Rate Calculation

Outlier

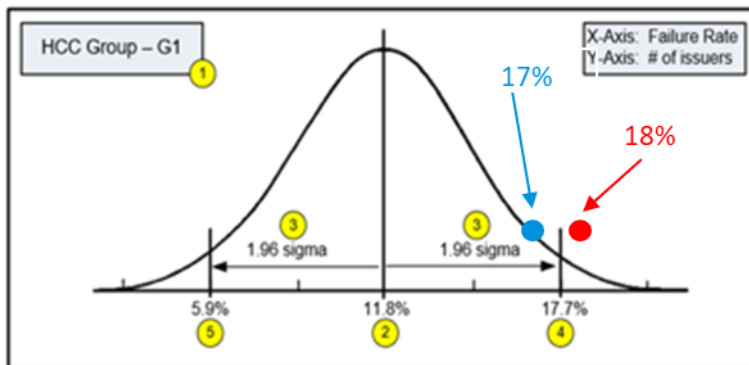
A value that falls outside of an established threshold. In HHS-RADV, a HIOS ID with a failure rate that falls outside of the HCC Group upper or lower boundary is an outlier. A HIOS ID may be identified as an outlier in one, two, or all three HCC Groups.

Error Rate Calculation – Issue 1

- Current Methodology: Group Adjustment Factor (GAF) is the difference between issuer's group failure rate and the national weighted mean failure rate

$$\text{GAF} = (\text{Issuer GFR} - \text{Weighted Mean GFR})$$

- Two (2) issuers may have very similar failure rates, but will be impacted very differently depending on their outlier status:



- Issuer is not an outlier, and will not receive an error rate
- Issuer will have a 6.2% adjustment factor for all HCCs in group G1, and will receive an error rate

Error Rate Calculation – Issue 1

Options Explored:

1. Original Error Estimation Methodology
2. Only Adjust to Confidence Intervals
3. Only Adjust for Positive Error Rate Outliers
4. Sliding Scale Adjustment Options

Error Rate Calculation – Issue 1

Individual Market Risk Pools 2018 Risk Adjustment

Metrics	Current Error Rate Methodology	Original Error Rate Methodology
Total RADV Payment Transfer Amounts	\$329,819,454	\$2,018,305,677
Percent RADV Payment Transfers Over Total Transfers Before RADV	8.23%	50.36%
Issuer's Average Absolute Transfer over Premium	0.89%	5.27%
Member Weighted Risk Score with RADV	1.553	1.448
Risk Score % Change	0.35%	-6.87%
% Billable Member Months by issuers with Adjusted Risk Scores	15.3%	70.5%
# State Market Risk Pools with RADV Adjustments	18	44
# Issuers with Adjusted Risk Scores	28	190
# Issuers with Adjusted RA Transfers	127	237
% of Issuers with Adjusted RA Transfers	49.2%	91.9%

Small Group Market Risk Pools 2018 Risk Adjustment

Metrics	Current Error Rate Methodology	Original Error Rate Methodology
Total RADV Payment Transfer Amounts	\$346,330,506	\$1,407,927,984
Percent RADV Payment Transfers Over Total Transfers Before RADV	29.81%	121.17%
Issuer's Average Absolute Transfer over Premium	1.26%	5.39%
Member Weighted Risk Score with RADV	1.279	1.176
Risk Score % Change	0.68%	-8.01%
% Billable Member Months by issuers with Adjusted Risk Scores	22.1%	86.2%
# State Market Risk Pools with RADV Adjustments	31	49
# Issuers with Adjusted Risk Scores	78	379
# Issuers with Adjusted RA Transfers	329	473
% of Issuers with Adjusted RA Transfers	69.6%	100.0%

Error Rate Calculation – Issue 1

Only Adjust to Confidence Intervals Example

- Based on 2017 Benefit Year HHS-RADV Results, an issuer with a 70 percent failure rate in the high HCC group would be considered an outlier under the current methodology
 - This issuer's failure rate is more than 4 standard deviations away from the national mean, well beyond the 1.96 standard deviations required to be determined to have outlier status

Group Adjustment Factor Calculation Difference:

- Current Methodology: 70 percent – 26.2 percent = **43.8 percent**
- Confidence Interval Methodology: 70 percent – 47.1 percent = **22.9 percent**

Error Rate Calculation – Issue 1

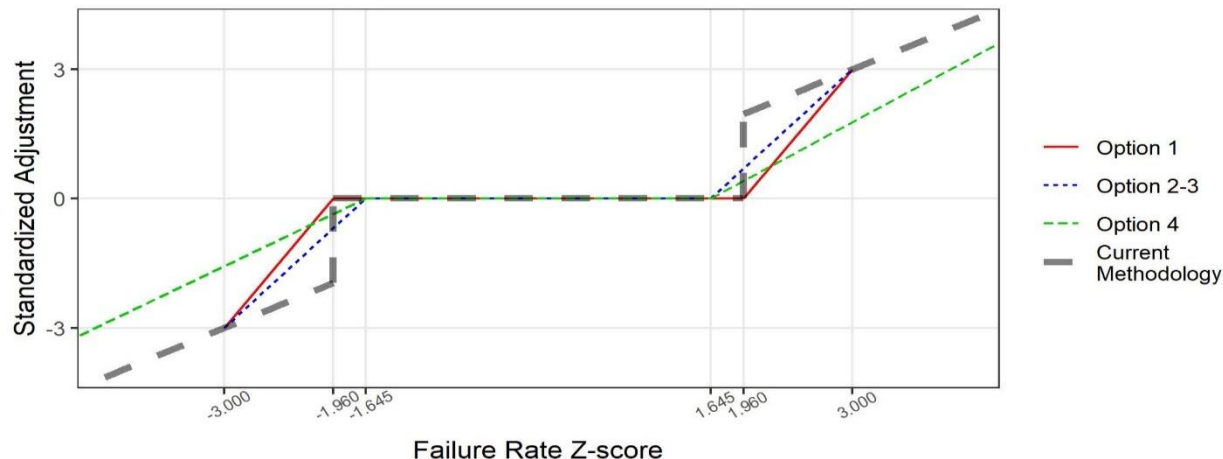
Only Adjust for Positive Error Rate Outliers

- The two-sided outlier identification, and the resulting adjustments to outlier issuer risk scores that have significantly better-than-average or poorer-than-average data validation results is to ensure that HHS-RADV makes adjustments for identified, material risk differences between what issuers submitted to the EDGE servers and what was validated by the issuer's medical records
- Adjusting for negative error rate outliers ensures that issuers who are coding well are able to recoup funds that might have been lost in the absence of data validation when its competitors are coding badly
- Retains the “payment cliff” for positive error rate outliers

Error Rate Calculation – Issue 1

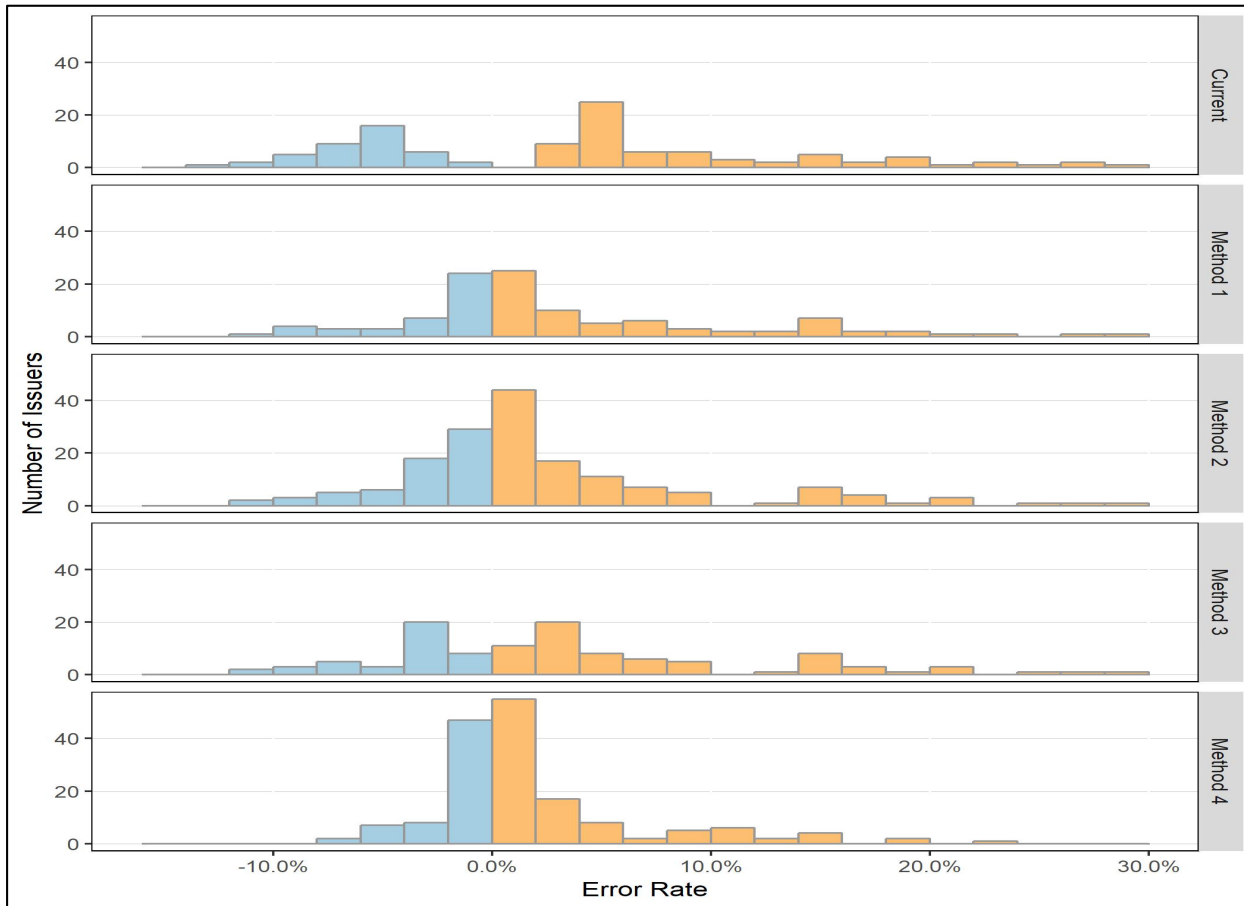
Sliding Scale Adjustment: linearly adjust between: 1) A failure rate value that occurs at the edge of the confidence interval; and 2) The group mean failure rate. The adjustment would take the following form: $A = a \times FR + b$, where the coefficients a (the slope) and b (intercept) would be calculated based on the empirical HHS-RADV failure rate results for each HCC group

1. Create the sliding scale adjustment from +/- 1.96 to 3 standard deviations.
2. Create a sliding scale adjustment from +/-1.645 to 3 standard deviations
3. Create a sliding scale adjustment from +/- 1.645 and 3 standard deviations and only apply it to issuers between +/-1.96 to 3 standard deviations
4. Create a sliding scale adjustment starting +/-1.645 to the maximum failure rate z score



Error Rate Calculation – Issue 1

Comparing the Distribution of Estimated Error Rates Between the Sliding Scale Options



Error Rate Calculation – Issue 2

Current Methodology: does not distinguish between low failure rates due to accurate data submission and those that are depressed through found HCCs

Option Explored: When an issuer with negative failure rates is determined to be negative error rate outlier, constrain the issuer's failure rate to 0 in the GAF calculation

Example: A negative outlier issuer with a -15 percent failure rate for the low HCC grouping

- Current Methodology:
 - -15 percent (Outlier Issuer's Failure Rate) – 4.8 percent (Weighted HCC Group Mean) = **-19.8 percent GAF**
- White Paper Methodology:
 - 0 (Issuer's Constrained Failure Rate) – 4.8 (Weighted HCC Group Mean) = **-4.8 percent GAF**

Error Rate Calculation

The negative failure rate constraint option:

- Ensures that negative error rate issuers are rewarded for high validation rates while mitigating incentives for under-reporting on EDGE
- Is easy to implement under current error estimation methodology as a temporary measure

APPLICATION OF HHS-RADV RESULTS

APPLICATION OF HHS-RADV RESULTS

With the exception of exiting issuers, HHS currently uses an issuer's HHS-RADV error rate from the prior year to adjust the issuer's risk score in the current transfer year

Option Explored: Apply HHS-RADV results to the same RA benefit year PLRS and transfers (i.e., 2021 HHS-RADV results applied to 2021 RA PLRS and transfers)

- Help maintain actuarial soundness if an issuer's risk profile or enrollment changes substantially from year to year
- Has potential to provide stability for issuers and help them better predict the impact of HHS-RADV results
 - Eliminate the exiting issuer policy and new issuer policy
 - Limit the number of state market risk pools that are adjusted based on one year of HHS-RADV results

APPLICATION OF HHS-RADV RESULTS

Average Error Rate Option: Calculate an average value between 2020 and 2021 HHS-RADV error rates and apply this average error rate to 2021 RA PLRS and transfers

RA Transfer Option: Separately calculate 2020 and 2021 HHS-RADV adjustments and then calculate the difference between these values using these three steps:

- a. Calculate 2020 benefit year HHS-RADV transfer adjustments to 2021 RA transfers and 2021 HHS-RADV transfer adjustments to 2021 RA transfers separately;
- b. Calculate the difference between each of these values and the unadjusted 2021 risk adjustment transfers; and
- c. Add these differences together to arrive at the total HHS-RADV modification to the 2021 benefit year RA transfers

Combined PLRS Option: Separately calculate and apply 2020 and 2021 HHS-RADV risk score adjustments using these three steps:

- a. Apply 2020 HHS-RADV risk score adjustments to 2021 RA PLRS;
- b. Apply 2021 HHS-RADV risk score adjustments to the adjusted 2021 PLRS (reflecting the 2020 benefit year HHS-RADV results); and
- c. Apply the final adjusted PLRSs (reflecting both the 2020 and 2021 HHS-RADV results) to adjust 2021 benefit year RA transfers

Next Steps

RADV White Paper Comments

- Comments are due:

January 6, 2020

- Comments must be submitted:

- With a “December 2019 HHS-RADV White Paper”
Subject Line

- To: CCIIOACARADDataValidation@cms.hhs.gov

Questions

