

A breakdown of ACA risk adjustment validation

Applying past experience to adjust future risk transfers since 2017

Cameron Gleed, FSA, MAAA
Jason Karcher, FSA, MAAA
Jason Petroske, FSA, MAAA



Since the inception of the Patient Protection and Affordable Care Act (ACA), the U.S. Department of Health and Human Services (HHS) realized a validation process would be needed to “ensure the accuracy and consistency of the data” underlying risk adjustment.¹

On August 1, 2019, HHS released the results of the first impactful issuer-level ACA risk adjustment data validation (RADV) audit. Many in the market are still digesting what happened and what it means to their businesses going forward.

In this article, we break down the purpose and technical details of ACA RADV. In a follow-up article, we further analyze how the process has broken down and explore issuer considerations in light of the recent results.

Risk adjustment is important, so how can we be confident it’s working?

ACA risk adjustment compensates issuers based on a calculated level of risk—the risk score.² These risk scores can be influenced by medical coding practices via diagnosis codes submitted to external data gathering environment (EDGE) servers. Given the significant marketwide impacts of risk transfers,³ HHS instituted an auditing process to verify the alignment of enrollee diagnosis codes, conditions, and risk scores between EDGE and medical charts.^{4,5}

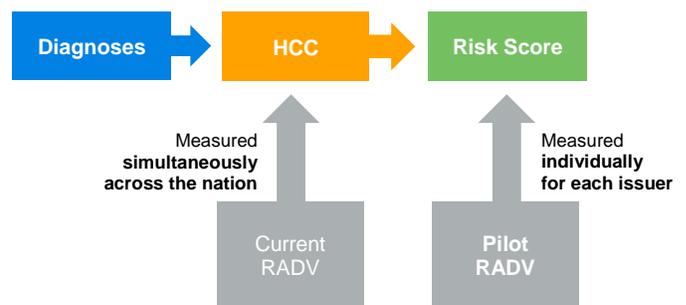
RADV protocols during program pilot years (2015 and 2016) operated quite differently than they do today. In early iterations, HHS focused on issuer risk score accuracy and directly aimed to correct overstated or understated risk scores on EDGE. Validators reviewed medical charts for relevant diagnoses, recalculated risk scores, and then determined whether these new scores were consistent with the EDGE server. Risk score adjustments were independent among market participants, and auditing essentially measured an issuer against itself.⁶ This methodology changed dramatically when RADV went live for the 2017 benefit year.

So how does RADV work now?

The 2019 Notice of Benefit and Payment Parameters (BPPs)⁷ codified a new RADV process with the stated goal of increased stability through fewer transfer adjustments. Under the new guidelines, RADV measures hierarchical condition category (HCC) accuracy relative to other market participants. Issuers with significantly different ratios of medical record HCC frequency and EDGE HCC frequency, compared with the nation, receive risk score adjustments to correct the relative HCC over- or under- identification.

This methodology does not directly assess absolute issuer risk score accuracy but, instead, provides more direct feedback on the coding practices of each issuer and whether those practices are outside the nationwide norm. We capture the essence of the differences in RADV methodologies within the context of a risk score build-up in Figure 1.

FIGURE 1: COMPARISON OF RADV METHODOLOGIES



HHS designed the RADV framework around six essential steps:⁸

1. Sample selection
2. Initial validation audit (IVA)
3. Second validation audit (SVA)
4. Error estimation
5. Appeals
6. Payment adjustments.

RADV

Dictionary

Failure rate

Percentage by which the number of validated HCCs within medical records is different from the number of HCCs reported on EDGE.

HCC failure rate group

Grouping of HCCs with similar average failure rates, stratified into low, medium, and high buckets.

Risk stratum

One of 10 groupings in the IVA sample representing enrollees with similar risk adjustment characteristics.

Outlier issuer

Issuer whose failure rate for at least one of the three HCC failure rate groups is significantly different from the average, as determined by a 95% confidence interval.

Error rate

The combined impact on risk scores after applying outlier failure rates from one or more HCC failure rate groups.

Transfer adjustment

The amount by which RADV affects initial risk transfers.

Cheat Sheet

What is the IVA sample size?

Set at 200 enrollees per issuer, with adjustments for issuers with fewer than 4,000 enrollees.

What is the SVA sample size?

Starts with a 12-enrollee subsample and expands, as needed, until either IVA and SVA conclusions agree or the SVA reviews (potentially) the full sample.

What is reviewed?

Demographic and enrollment data in issuer source systems; medical records from providers.

What does the validation measure?

Failure rates, or the number of HCCs supported in medical records versus the number of HCCs reported on the EDGE server.

Which issuers receive adjustments?

Issuers with a failure rate in an HCC failure rate group falling outside a 95% confidence interval—in other words, issuers whose documented coding practices for HCCs in a group are different from the average practices across the nation.

What are error rates based on?

The HCC component of risk scores.

What are error rates applied to?

The total risk score.

When is the adjustment?

The following benefit year.

When are payments collected?

Four years after the RADV benefit year (e.g., in 2021 for 2017 RADV).

What about exiting issuers?

Risk scores (and transfers) are adjusted for the current benefit year but, starting with the 2018 RADV, only when this results in payments from the exiting issuer(s) to other issuers.

Will issuers with 0% error rates be adjusted?

Because risk-score-based transfers sum to zero in each risk adjustment pool, an issuer with a 0% error rate may still experience transfer adjustments if the error rate for another issuer in that risk adjustment pool is nonzero.

SELECTING SAMPLE ENROLLEES

RADV relies on a 200-enrollee sample⁹ at the HIOS issuer ID level.¹⁰ Before sample selection, HHS separates an issuer's enrollees into 10 risk strata—nine (low, medium, and high risk for each of the adult, child, and infant risk score models) for enrollees with at least one HCC and a 10th for enrollees with no HCCs. HHS uses an allocation method¹¹ to determine the size of each stratum to achieve the full 200 sample¹² and then randomly places enrollees in each until the stratum sample size is reached. HHS does review the RADV sample to ensure it is representative of the total population before releasing it to the issuer.

THE INITIAL VALIDATION AUDIT

The issuer engages an eligible IVA vendor to audit its EDGE data. While the process includes an evaluation of demographic and enrollment data,¹³ the IVA vendor focuses on HCC support within medical records.

THE SECOND VALIDATION AUDIT

HHS engages an SVA vendor, which functions as an independent validation of the IVA results. Starting with a 12-enrollee subsample,¹⁴ if risk scores independently calculated from the IVA and SVA agree closely,¹⁵ then HHS uses the IVA results for all sampled enrollees throughout the remainder of the RADV process. If the initial IVA and SVA results disagree, then the subsample is gradually expanded up to 100 enrollees until statistical agreement is reached.

If at this point the IVA and SVA results still do not agree, the situation becomes more complex,¹⁶ but HHS will use the SVA sample in lieu of the IVA sample. Therefore, the remainder of the RADV evaluation could be based on as few as the 100 SVA-reviewed enrollee subsample instead of the full sample.

ERROR ESTIMATION

HHS estimates error rates after a three-phase process.

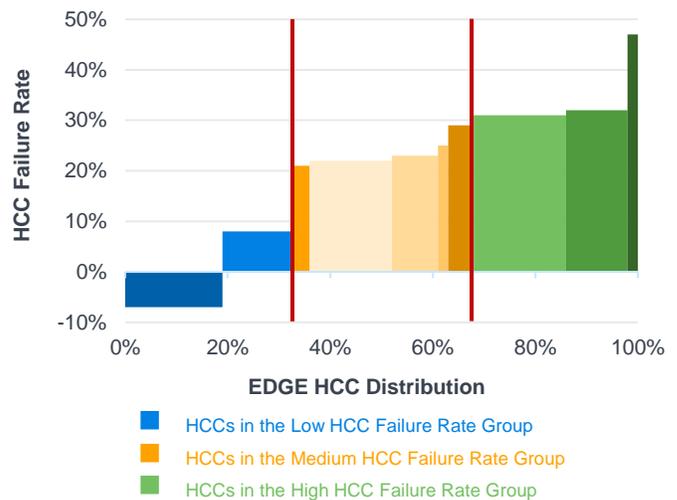
Phase 1 – Determining HCC failure rate groups

Using the validation results, HHS assigns a “failure rate” to each HCC by comparing the frequency of that HCC in the audits with the frequency of that HCC in EDGE using the following formula:¹⁷

$$\text{HCC Failure Rate} = 1 - \frac{\text{Validated Medical Record HCC count}}{\text{EDGE HCC count}}$$

A positive failure rate means fewer instances of an identified HCC in the medical records while a negative failure rate indicates more instances of an identified HCC in the medical records. HHS then ranks all HCCs by their nationwide failure rates and creates three groups (low, medium, and high) such that approximately one-third of EDGE HCCs fall into each group. We illustrate the delineation of the HCC failure rate groups in Figure 2:

FIGURE 2: ILLUSTRATION OF FAILURE RATE GROUP CREATION



In this example using just 10 HCCs, two HCCs fall into the low failure rate group (blue), five into the medium (orange), and three into the high (green). Validators find the first blue HCC more frequently in medical records than on EDGE (designated by a negative failure rate), while the last green HCC shows up nearly twice as often on EDGE (designated by a positive failure rate of almost 50%). HHS's goal, relating to Figure 2, is roughly equal widths for the three different color groups (blue vs. orange vs. green), based on HCC prevalence on EDGE, as opposed to evenly dividing the number of HCCs into each of the three categories.

Phase 2 – Determining outlier issuers

After ranking HCCs and assigning failure rate groups, HHS determines whether an issuer's coding practices are inconsistent with national average practices. HHS first determines issuer-level and national average failure rates for each failure rate group and then calculates the failure rate standard deviation¹⁸ and a two-sided 95% confidence interval.¹⁹ This confidence interval estimates the range within which the average population (i.e., all ACA enrollees, not just the 200-enrollee validation sample) group failure rate should lie 95% of the time. Issuers with a failure rate outside the confidence interval likely do not have national average coding practices, and RADV flags them as outliers.²⁰

Phase 3 – Determining issuer adjustments

An outlier issuer in any failure rate group receives an adjustment to its HCC risk scores within that group, which increases as the issuer's failure rate is further from the boundary of the confidence interval. It measures the issuer's percentage of unsupported HCCs relative to the nation in that failure rate group.

$$\begin{aligned} \text{HCC failure rate group adjustment} \\ &= \text{Issuer HCC group failure rate} \\ &\quad - \text{national HCC group failure rate} \end{aligned}$$

A positive number means the issuer has a greater rate of unsupported HCCs, while a negative number means the issuer has a lower rate of unsupported HCCs.

HHS next computes an enrollee-level adjusted risk score.²¹ This risk score can be likened to an estimate of the true risk score if the issuer's coding were more like the national average.

Figures 3 and 4 illustrate this calculation for an age 45 male enrolled for a full year in a silver 87% cost-sharing reduction plan.

FIGURE 3: ILLUSTRATIVE CALCULATION OF RADV ENROLLEE RISK SCORE ADJUSTMENT

HCC	HCC GROUP	EDGE HCC RISK SCORE	ISSUER HCC GROUP ADJUSTMENT	ADJUSTED EDGE HCC RISK SCORE
1	Medium	0.330	0.0%	0.330
12	High	2.451	45.0%	1.348
37	High	0.930	45.0%	0.512
135	Medium	5.206	0.0%	5.206
251	Low	25.158	-20.0%	30.190
HCC RISK SCORE		34.075		37.585
ENROLLEE RISK SCORE ADJUSTMENT				1.103

In our example, the adjusted risk score is greater than the EDGE risk score, so this enrollee contributes an overall negative failure rate (and, as such, suggests the enrollee has fewer²² HCCs incorrectly coded than the national average). HHS then uses this adjustment factor to calculate an adjusted EDGE risk score.²³

FIGURE 4: ILLUSTRATIVE OF RISK SCORE AND ADJUSTED RISK SCORE CALCULATION

COMPONENT	VALUE	ANNOTATION
Demographic Factor	0.245	(1)
Enrollment Duration Factor	0.000	(2)
Diagnosis Factor†	34.075	(3)
Standard Risk Score	34.320	(4) = (1) + (2) + (3)
CSR Induced Utilization PLRS Factor	1.120	(5)
EDGE Risk Score	38.438	(6) = (4) * (5)
RADV Adjustment	1.103	(7)
Adjusted EDGE Risk Score	42.398	(8) = (6) * (7)

† The diagnosis factor reflects HCCs, prescription drug condition categories (RXCs), enrollment duration factors, and any interaction or severity terms indicated based on EDGE data.

From this point, HHS computes an average EDGE risk score and adjusted EDGE risk score for each of the 10 risk strata and composites these to the issuer level using the number of enrollees in each stratum as the weight, as shown in Figure 5.

FIGURE 5: ILLUSTRATION OF COMPOSITE SAMPLE RISK SCORE CALCULATION

STRATUM	ISSUER ENROLLMENT	AVERAGE EDGE RISK SCORE	AVERAGE ADJUSTED EDGE RISK SCORE
Adult Low	4,500	2.500	2.468
Adult Medium	4,000	15.257	15.288
Adult High	2,600	44.125	44.743
Child Low	1,500	3.750	3.791
Child Medium	1,400	10.250	10.291
Child High	1,300	27.891	28.226
Infant Low	200	5.250	5.324
Infant Medium	150	27.652	27.320
Infant High	300	235.411	234.234
No HCCs	84,050	0.300	0.300
ISSUER TOTAL	100,000	3.443	3.460

Finally, HHS calculates the issuer-level error rate that adjusts next year's risk scores:

$$Error\ Rate = 1 - \frac{Average\ Adjusted\ EDGE\ Risk\ Score}{Average\ EDGE\ Risk\ Score}$$

In our example, the adjusted EDGE risk score is greater than the EDGE risk score based on current year enrollment, indicating a negative error rate and a favorable risk score adjustment in the following year. Note that the transfer adjustments arising directly from outlier issuers will also give rise to transfer adjustments for non-outlier issuers, because the risk adjustment calculation is zero-sum across each risk adjustment pool.

APPEALS AND PAYMENT ADJUSTMENTS

After publishing draft transfer adjustments, HHS grants issuers a window to appeal certain elements of the RADV process (perhaps the most notable exclusion is appeal of the IVA results). Once closed, issuers bear responsibility for the reported transfers and all transfer adjustments.

The final 2020 BPPs²⁴ outlined minor RADV procedure changes, including one regarding issuers exiting the market after the audit year.²⁵ The most significant change extended the RADV timeline to grant issuers more time to prepare for RADV payments and associated reporting. Instead of applying 2017 benefit year RADV adjustments during the standard 2018 benefit year risk adjustment payment cycle (in calendar year 2019 as envisioned in the 2019 NBPPs), HHS prolonged the appeals process and delayed payment until calendar year 2021—with amounts included in 2021 minimum loss ratio (MLR) reporting year filings during the second half of 2022. While this extension may ease the cash flow of issuers paying a transfer adjustment, it will delay the receipt for issuers on the receiving end and introduce a new set of challenges to 2021 rate development and MLR.²⁶

What comes next?

RADV error rates can be significant and widespread in the market.

Widespread: Because adjustments apply to the issuer's risk score (and, by extension, affect the state average risk score), any adjustment to one issuer impacts all issuers in the state. Thus, in theory, while only 5% of issuers should be outliers in each HCC failure rate group, the number of issuers reasonably expected to feel the results of RADV is significantly higher.

Significant: Error rates tend to be large due to the current widths of the confidence intervals and the adjustment back to the average failure rate within an HCC failure rate group. Transfer adjustments have, likewise, proven to be significant. An ostensibly small error rate, such as 3%, actually represents a significant transfer adjustment (because it is applied essentially as a percentage of premium)²⁷ and could eliminate expected profit margins unexpectedly.

To help avoid, or at least understand, adverse RADV impacts, the technical background presented here is essential. In follow-up articles, we explore early takeaways from the initial round of 2017 benefit year RADV transfer adjustments and illustrate what these results suggest for participating issuers and the ACA market as a whole.

Acknowledgments

The authors would like to thank Scott Jones and Erik Huth for their review. Additionally, the authors would like to acknowledge the contributions of Peter Fielek, Annie Man, Jeff Milton-Hall, and Alan Vandagriff to the technical analyses and for their interpretations of the RADV process.

CAVEATS AND LIMITATIONS

Guidelines issued by the American Academy of Actuaries require actuaries to include their professional qualifications in actuarial communications. The authors Cameron Gleed, Jason Karcher, and Jason Petroske are actuaries with Milliman Inc. The authors are members of the American Academy of Actuaries and meet the Qualification Standards of the American Academy of Actuaries to render the actuarial analyses herein.

In preparing this article, they relied upon the following federal publications:

- The Premium Stabilization rule
- The HHS Notice of Benefit and Payment Parameters for 2014 through 2020

- The 2018 Benefit Year RADV Protocols for PPACA HHS Risk Adjustment Data Validation, Version 7.0 dated June 25, 2019, and published on the Registration for Technical Assistance Portal (REGTAP) July 5, 2019, and the 2016 Benefit Year RADV Protocols for PPACA HHS Risk Adjustment Data Validation Version 14.0 dated and published on October 20, 2017, on REGTAP
- Published RADV Reports, including the August 1, 2019, Summary Report of 2017 Benefit Year Risk Adjustment Data Validation Adjustments to Risk Adjustment Transfers, the May 31, 2019, 2017 Benefit Year HHS Risk Adjustment Data Validation Results, and the 2016 pilot year RADV summary report
- 2018 Benefit Year Risk Adjustment Updated HHS-Developed Risk Adjustment Model Algorithm "Do It Yourself (DIY)" Software published on April 4, 2019

Any changes to RADV regulations or guidance in future rulemaking or as a result of legislation or litigation may impact the results discussed.

Any distribution of this article should be in its entirety. Milliman does not intend to benefit, or create a legal duty to, any third-party recipient of this article.

Differences between the theory discussed in this article and actual results depends on the extent to which future experience conforms to the assumptions made for this analysis. It is certain that actual experience will not conform exactly to the assumptions used in this analysis. Actual impacts will vary from the theoretical impacts for a variety of reasons, and issuers subject to RADV should monitor their results and take corrective action when necessary.

We are not lawyers and, therefore, cannot provide legal advice. Readers are advised to confer with counsel before use of the information herein.

CONTACT

Cameron Gleed
cameron.gleed@milliman.com

Jason Karcher
jason.karcher@milliman.com

Jason Petroske
jason.petroske@milliman.com

Endnotes

- ¹ Federal Register (March 11, 2014). Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2015. Retrieved October 18, 2019, from <https://www.federalregister.gov/documents/2014/03/11/2014-05052/patient-protection-and-affordable-care-act-hhs-notice-of-benefit-and-payment-parameters-for-2015#p-133>.
- ² Starting with the 2018 benefit year, risk adjustment incorporated prescription drug condition categories (RXC) into RADV, which are handled by a separate process without statistical elements.
- ³ Jason Petroske, Brandy Millen, Peter Fielek, Alan Vandagriff, and Lisa Mattie explore the ACA's risk adjustment program in more detail in a four-part series, available at <http://www.milliman.com/insight/2017/ACA-risk-adjustment-management-Going-all-out/>.
- ⁴ Federal Register (March 23, 2012). Patient Protection and Affordable Care Act; Standards Related to Reinsurance, Risk Corridors and Risk Adjustment. Retrieved October 18, 2019, from <https://www.federalregister.gov/documents/2012/03/23/2012-6594/patient-protection-and-affordable-care-act-standards-related-to-reinsurance-risk-corridors-and-risk#p-212>.
- ⁵ Federal Register (March 11, 2013). Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2014. Retrieved October 18, 2019, from <https://www.federalregister.gov/documents/2013/03/11/2013-04902/patient-protection-and-affordable-care-act-hhs-notice-of-benefit-and-payment-parameters-for-2014#h-50>.
- ⁶ A side effect of this independence is that all issuers could be adjusted—or none.
- ⁷ Federal Register (April 17, 2018). Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2019. Retrieved October 18, 2019, from <https://www.federalregister.gov/documents/2018/04/17/2018-07355/patient-protection-and-affordable-care-act-hhs-notice-of-benefit-and-payment-parameters-for-2019>.
- ⁸ This structure was first outlined in the 2014 BPPs. See: Federal Register (March 11, 2013). Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2014, available at <https://www.federalregister.gov/documents/2013/03/11/2013-04902/patient-protection-and-affordable-care-act-hhs-notice-of-benefit-and-payment-parameters-for-2014#h-50>.
- ⁹ HHS reduces the sample for issuers with fewer than 4,000 enrollees by a factor, with a minimum size of 50 sampled enrollees. For example, if Issuer A has 3,000 enrollees, its sample size is:
- $$200 \times (3,000 - 200) / 3,000 = 187$$
- Issuers with fewer than 50 enrollees but still subject to RADV (which is unlikely) will have all enrollees validated.
- ¹⁰ HIOS issuer IDs are unique by entity and state but not by market within a state. For example, an issuer operating a PPO and HMO entity in three states could have up to six HIOS IDs, though this count would not necessarily be affected by whether the issuer offered only small group coverage or both individual and small group. This distinction implies issuer coding patterns are unlikely to vary by market. For the remainder of the paper, we use "issuer" to refer to a unique HIOS Issuer ID.
- ¹¹ RADV utilizes Neyman allocation, which, by definition, minimizes the sample variance when a fixed population is segmented into distinct groups, as is the case here. Practically speaking, this creates the smallest confidence intervals.
- ¹² A minimum size constraint is applied to the no-HCC group, which otherwise has the smallest risk score variance and could be under-sampled for issuers who under-code diagnoses.
- ¹³ An issuer with significant errors in EDGE-reported demographic and enrollment information may be penalized if such errors are significant and pervasive.
- ¹⁴ SVA subsamples prioritize enrollees with supporting medical records.
- ¹⁵ Using a pairwise means test at a 95% confidence level. Note that risk scores are calculated using EDGE demographic, enrollment, and RXC information.
- ¹⁶ HHS performs a bootstrap resampling analysis to identify the potential standard error of the 100-enrollee SVA results. If HHS deems the standard error to be too large, then the SVA sample expands to the full IVA sample.
- ¹⁷ When determining issuer error rates, the calculation uses the final validated sample—IVA or SVA. There is an element of circularity when the SVA sample is used because the bootstrap resampling analysis, in essence, requires replicating all three phases of the error estimation process when determining which sample size to use. It is also worth emphasizing RADV focuses on HCC validation, and the factors leading to an HCC match between EDGE and the medical records may be completely different.
- ¹⁸ Each issuer's contribution is weighted by the number of EDGE diagnoses in that HCC failure rate group. This tends to lend additional weight to group failure rates of issuers with more HCCs reported on EDGE, regardless of whether the frequency is driven by actual prevalence or over-coding. One consequence of incidence-based weighting is possible underrepresentation of issuers using a 100-enrollee SVA sample.
- ¹⁹ The current version of the RADV protocols uses 1.96 standard deviations without any comment as to the justification for this selection. Earlier versions reference a two-sided t-test but still use 1.96 standard deviations, which is only valid for a sufficiently large sample. As such, the confidence interval determined may be somewhat lower than the 95% threshold indicated in the methodology.
- ²⁰ The methodology does not make an explicit statement but implies the group failure rate distribution is normal (or else large enough to assume it fits a t-distribution for a sufficiently large population).
- ²¹ Our description of the RADV process from this point is somewhat different from, but mathematically equivalent to, the process outlined in the BPPs and the RADV protocols.
- ²² In this sense, fewer refers to the weighted average impact of the mis-coded HCCs, rather than an absolute number of HCCs. As such, it is influenced by the magnitude of outlier HCC risk scores in addition to the directionality of each.
- ²³ The RADV protocols and the 2019 BPPs differ in how they describe the enrollee EDGE risk score. The BPPs clearly state the EDGE risk score "is the risk score for EDGE HCCs," while the protocols indicate it "include(s) all EDGE risk score components." If the BPP interpretation is correct, the no-HCC risk group will not contribute to the error rate (because its HCC risk score is, by definition, zero), and the error rate would only reflect the impact of RADV on individuals with HCCs. Because the RADV descriptions in the BPPs are not included in the Code of Federal Regulations and, therefore, represent guidance only, it is reasonable to assume the RADV protocol methodology applies.
- ²⁴ Federal Register (April 25, 2019). Patient Protection and Affordable Care Act; HHS Notice of Benefit and Payment Parameters for 2020. Retrieved October 18, 2019, from <https://www.federalregister.gov/documents/2019/04/25/2019-08017/patient-protection-and-affordable-care-act-hhs-notice-of-benefit-and-payment-parameters-for-2020>.
- ²⁵ Starting with the 2018 benefit year, exiting issuers with negative error rates will not receive RADV payouts.
- ²⁶ Issuers affected (whether by their own error rates or that of another issuer in the markets in which they participate) may have to book RADV assets or liabilities that will not be reconciled until 2021, and this may impact risk-based capital and other statutory financial reporting.
- ²⁷ Because a 3% error rate could raise the state average risk score as well as the issuer's risk score, the impact may be smaller depending on the size of the issuer.