IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS
FORT WORTH DIVISION

ELEVANCE HEALTH, INC.; COMMUNITY INSURANCE COMPANY; FREEDOM HEALTH, INC.; GROUP RETIREE HEALTH SOLUTIONS, INC.; WELLPOINT INSURANCE COMPANY; and WELLPOINT TEXAS, INC., Plaintiffs, v. ROBERT F. KENNEDY, JR., in his official capacity as Secretary of Health and Human Services; and STEPHANIE CARLTON, in her official capacity as Acting Administrator of the Centers for Medicare & Medicaid Services, Defendants.[1]	Civil Action No. 4:24-cv-01064-P

DEFENDANTS' CROSS-MOTION FOR SUMMARY JUDGMENT

Defendants, Robert F. Kennedy, Jr., in his official capacity as Secretary of Health and Human Services, and Stephanie Carlton, in her official capacity as Acting Administrator of Centers for Medicare and Medicaid Services, move this Court for entry of summary judgment in their favor on all Plaintiffs' claims, pursuant to Federal Rule of Civil Procedure 56.

[1] Xavier Becerra has been substituted with Robert F. Kennedy, Jr. as Secretary of the United States Department of Health and Human Services, and Chiquita Brooks-LaSure has been substituted with Stephanie Carlton as Acting Administrator of the Centers for Medicare & Medicaid Services, pursuant to Federal Rule of Civil Procedure 25(d).

This response is accompanied by a consolidated brief that contains the contents required by Federal Rule of Civil Procedure 56(a) and Local Civil Rule 56.3(a).

Respectfully submitted,

CHAD E. MEACHAM
ACTING UNITED STATES ATTORNEY

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney
Texas Bar No. 24007419
Burnett Plaza, Suite 1700
801 Cherry Street, Unit #4
Fort Worth, Texas 76102-6882
Telephone: 817.252.5230
Fax: 817.252.5458
Andrea.Hyatt@usdoj.gov

Attorneys for Defendants

Certificate of Service

On March 21, 2025, I electronically submitted the foregoing document with the clerk of court for the U.S. District Court, Northern District of Texas, using the electronic case filing system of the court. I hereby certify that I have served all parties electronically or by another manner authorized by Federal Rule of Civil Procedure 5(b)(2).

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS
FORT WORTH DIVISION

ELEVANCE HEALTH, INC.; COMMUNITY INSURANCE COMPANY; FREEDOM HEALTH, INC.; GROUP RETIREE HEALTH SOLUTIONS, INC.; WELLPOINT INSURANCE COMPANY; and WELLPOINT TEXAS, INC.,

Plaintiffs,

v.

ROBERT F. KENNEDY, JR., in his official capacity as Secretary of Health and Human Services; and STEPHANIE CARLTON, in her official capacity as Acting Administrator of the Centers for Medicare & Medicaid Services,

Defendants.[1]

Civil Action No. 4:24-cv-01064-P

DEFENDANTS' CONSOLIDATED BRIEF IN SUPPORT OF THEIR RESPONSE TO PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT AND CROSS-MOTION FOR SUMMARY JUDGMENT

CHAD E. MEACHAM
ACTING UNITED STATES ATTORNEY

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney
Texas Bar No. 24007419
Burnett Plaza, Suite 1700
801 Cherry Street, Unit #4
Fort Worth, Texas 76102-6882
Telephone: 817.252.5230
Fax: 817.252.5458
Andrea.Hyatt@usdoj.gov

Attorneys for Defendants

[1] Xavier Becerra has been substituted with Robert F. Kennedy, Jr. as Secretary of the United States Department of Health and Human Services, and Chiquita Brooks-LaSure has been substituted with Stephanie Carlton as Acting Administrator of the Centers for Medicare & Medicaid Services, pursuant to Federal Rule of Civil Procedure 25(d).

Table of Contents

Table of Authorities

Cases

Statutes

Statutes (con.)

Regulations

Regulations (con.)

Regulations (con.)

Regulatory

Rules

Other Authorities

I. Summary

Congress directed the Centers for Medicare & Medicaid Services (CMS) to calculate Medicare Advantage Star Ratings on a one- through five-star scale in half-star increments, which allows Medicare beneficiaries to comparison shop among hundreds of private health insurance plans. Following its regulations, CMS calculated two of Plaintiff Elevance's Medicare Advantage contracts' overall ratings to be 3.749565 and 3.220943. In accordance with applicable regulations, technical guidance, and the traditional rules of rounding, which require rounding to the nearest half star, CMS rounded Elevance's scores to 3.5 and 3.0 stars, respectively. Had the scores come out to 3.75 and 3.25, respectively, or more, CMS would have rounded to the next half-star increment. This lawsuit is Elevance's effort to nudge their scores over the line into the next half-star category. As part of this pursuit, Elevance asks this Court to overlook CMS's regulations and endorse invalid statistical methods. Elevance advocates for rule changes to the calculation of Medicare Advantage Star Ratings that Elevance did not and would never advance as neutral, statistically sound principles before it knew its overall scores. Elevance's dubious, post hoc efforts do not satisfy its legal burden of proving that any of the four rules it challenges are arbitrary and capricious.

First, Elevance contends that case-mix adjusting survey data—that is, adjusting contracts' survey results for age, education, health status, and income—is not supported by CMS's regulations. It is. The applicable regulations authorize case-mix adjustment.

Second, Elevance contends that when CMS calculates the "national average" used in certain survey measures, CMS is required to average contract-level scores without regard to how many enrollees are in each contract. Some measure-level Star Ratings that factor into the overall Star Rating a contract receives are based on whether that contract's average survey measure score

is statistically significantly lower or higher than the "national average" survey measure score. In calculating this national average, Elevance contends that CMS is required to calculate the country's contract-level average. The regulations do not require, and CMS does not calculate, the average of the scores this way. The number of people covered by each contract varies significantly contract to contract. As a result, CMS calculates a true national average that captures the average beneficiary experience—not the contract's experience—by weighting the contract scores according to enrollment and using those scores to calculate the national average. CMS's method is reasonable and entirely consistent with the regulation.

Third, Elevance asks this Court to require CMS to double round overall scores, even though the regulation only requires CMS to round once. Because 3.749565 is closer to 3.5 stars than it is to 4 stars, CMS rounded one time to 3.5 stars. Elevance asks this Court to require that CMS double round its measure score first up to 3.75, then up again to 4. This kind of double rounding is at odds with CMS's regulations, guidance, and the traditional rules of rounding. Relatedly, Elevance appears to believe that because overall scores are displayed in CMS publications out to the sixth decimal, that CMS rounds to the sixth decimal when calculating the overall Star Rating. There is no basis for this belief. CMS rounds once to the half star when calculating the overall rating. On the basis of its false premise, Elevance advances a number of arguments related to the supposed imprecision of CMS's calculation methodology. This Court should reject them not only because they are based on Elevance's false premises, but also because CMS's Star Ratings calculation methodology is in fact adequately precise.

Finally, Elevance contends that using the same seed—the initiating number sequence used to create the random groups—year over year makes mean resampling less random and more predictable. This is false. Mean resampling is a statistical reshuffling and random grouping of

data aimed at stabilizing results year over year. It is simply a standardized process that does not dictate results. CMS's 8-6-7-5-3-0-9 seed has different and random effects in different years because it is applied to different data from different contracts every year.

Every year, it is a near certainty that some contracts, like Elevance's, will receive scores that just miss the cut-off for a higher star increment. If this Court grants Elevance the relief it seeks, it opens its doors to yearly, statistically dubious arguments of Medicare Advantage organizations seeking to nudge their scores up into the next half-star category. The Court should reject these efforts at the outset.

II. Background

A. Statutory and Regulatory Background

Title XVIII of the Social Security Act, 42 U.S.C. § 1395—1395*lll* ("Medicare statute"), establishes the Medicare program, a federally funded and administered health insurance program for eligible elderly and disabled persons and certain individuals with end stage renal disease. *See* 42 U.S.C. § 1395c. The Secretary administers the Medicare program through the Centers for Medicare & Medicaid Services ("CMS"), a component agency of the United States Department of Health and Human Services.

The Medicare program is divided into four major components:

1. Part A, the hospital insurance benefit program, provides health insurance coverage for certain inpatient hospital care, post-hospital care in a skilled facility, post-hospital home care services, and other related services. *See* 42 U.S.C. §§ 1395c, 1395d.
2. Part B, the supplemental medical insurance benefit program, generally pays for a percentage of certain medical and other health services, including physician services, supplemental to the benefits provided by Part A. *See* 42 U.S.C. §§ 1395j, 1395k, 1395*l*.

3. Under Part C, the Medicare Advantage program, a Medicare beneficiary can elect to receive his or her Medicare benefits through a public or private healthcare plan. *See* 42 U.S.C. §§ 1395w-21 —1395w-29.
4. Finally, Part D is the voluntary prescription drug benefit program. *Id.* §§ 1395w-101—1395w-154.

Under Part C's Medicare Advantage program, the federal government pays insurers to provide the coverage that participating beneficiaries would otherwise receive through Parts A and B (sometimes known, collectively, as "traditional" Medicare). *Id.* § 1395w-22(a). These insurers, known as Medicare Advantage Organizations ("MAOs"), contract to provide coverage in a particular geographic area. Beneficiaries can then choose among the plans available where they reside. *Id.* § 1395w-21(b). MAOs receive a predetermined sum for providing coverage to each beneficiary, based in part on the demographic and health characteristics of that beneficiary. *Id.* § 1395w-23(a)(1)(A), (C).

To calculate payments to MAOs, CMS first determines its "benchmark," based on the per-capita cost of covering Medicare beneficiaries under Parts A and B in the relevant geographic area. *Id.* § 1395w-23(n); 42 C.F.R. § 422.258. Each MAO then submits a "bid," telling CMS what payment the MAO will accept to cover a beneficiary with an average risk profile in that area. 42 C.F.R. § 422.254. If the insurer's bid is less than the benchmark, the bid becomes its "base payment"—the amount it is paid for covering a beneficiary of average risk—and the insurer receives a portion of the difference between its bid and the benchmark as a "rebate" that the MAO can use to fund supplemental benefits for beneficiaries or reduce plan premiums. 42 U.S.C. § 1395w-24(b)(1)(C); 42 C.F.R. § 422.260. If the MAO's bid is greater than the benchmark, then

the benchmark becomes its base payment, and the insurer must charge beneficiaries a premium to make up the difference. *See* 42 U.S.C. §§ 1395w-23(a)(1)(B)(ii), 1395w-24(b)(2)(A).

Star Ratings are a means by which CMS measures the quality of Medicare Advantage plans and Part D Prescription Drug Plans. Plans are rated on a scale of one to five "stars" in half-star increments based on Medicare Advantage and Part D data collected by CMS. 42 U.S.C. § 1395w-23(o)(4)(A); *see also* 42 U.S.C. § 1395w-22(e)(3). Star Ratings reflect the experiences of beneficiaries in these plans and assist beneficiaries in finding the best plans for their needs. CMS, *Advance Notice of Methodological Changes for Calendar Year (CY) 2026 for Medicare Advantage (MA) Capitation Rates & Part C & Part D Payment Policies* 109 (Jan. 10, 2025) https://www.cms.gov/files/document/2026-advance-notice.pdf [https://perma.cc/KWB8-VLWK].

CMS has released Star Ratings for Medicare Advantage contracts since 2008. *See* Contract Year 2019 Policy & Technical Changes to the Medicare Advantage Program, 83 Fed. Reg. 16,440, 16,520 (Apr. 16, 2018). In 2018, CMS adopted the regulatory framework for the Star Ratings and since then has used rulemaking to adopt changes in the methodology and add new measures. *Id.*; *see also* 42 C.F.R. § 422.164(c), (d). The 2018 final rule describes the purpose of the Star Ratings system: it "is designed to provide information to the beneficiary that is a true reflection of the plan's quality and encompasses multiple dimensions of high quality care." 83 Fed. Reg. at 16,520.

Star Ratings are assigned to each individual contract held by an MAO. The overall Star Ratings are based on a 5-star scale, set in half-star increments, with 1 star being the lowest rating and 5 stars being the highest. *See* 42 U.S.C. §§ 1395w-23(o)(4)(A), 1395w-24(b)(1)(C)(v); 42 C.F.R. §§ 422.162(b), 422.166(h)(1)(ii).

Star Ratings affect payments to MAOs in two main ways. First, Medicare Advantage plans that earn a rating of four stars or higher qualify for Medicare Advantage Quality Bonus Payments

in the form of an increased benchmark for the contract year following the ratings year (e.g., the 2025 Star Ratings can increase the Medicare Advantage bidding benchmarks for contract year 2026). 42 U.S.C. § 1395w-23(o)(1) (increasing, for qualifying plans, the applicable percentage that calculates the benchmark); § 1395w-23(o)(3)(A)(i) (a qualifying plan is one that earns a rating of four stars or higher). Higher benchmarks in turn can allow a Medicare Advantage plan to increase its bid, receive higher rebates, or lower premiums. *See id.* § 1395w-24(b)(1)(C); 42 C.F.R. § 422.260.

Second, Star Ratings affect the level of rebate received by plans that bid below their benchmarks for the contract year following the ratings year (e.g., the 2025 Star Ratings are used to set plans' rebate percentages for contract year 2026). Plans that earn a rating of four-and-a-half stars or higher receive a rebate of seventy percent of the difference between their bid and the benchmark, while plans that earn three-and-a-half or four stars receive a rebate of sixty-five percent of that difference, and plans that earn less than three-and-a-half stars are eligible for a rebate of fifty percent of that difference. 42 U.S.C. § 1395w-24(b)(1)(C)(v) (listing the "final applicable rebate percentage[s]" by rating); 42 C.F.R. § 422.266(a)(2)(ii) (same).

CMS publishes the Star Ratings each October for the upcoming year at the contract level, with each plan offered under that contract assigned the contract's rating. *See* 42 C.F.R. §§ 422.162(b), 422.166, 423.182(b), 423.186. For example, CMS published the 2025 Star Ratings, in October 2024. CMS, *Fact Sheet – 2025 Medicare Advantage and Part D Star Ratings* (Oct. 10, 2024) https://www.cms.gov/newsroom/fact-sheets/2025-medicare-advantage-and-part-d-star-ratings [https://perma.cc/8TLH-G7ZL].

1. Measure-Level Star Ratings Calculation Methodology

To calculate overall Star Ratings, CMS scores Medicare Advantage contracts on approximately 30 to 40 quality measures, depending on whether the plan is Medicare Advantage-only or also includes Part D coverage. Administrative Record ("A.R.") 13, ECF No. 23-1 ("Technical Guidance"). These measures relate to five broad categories: (1) outcomes; (2) intermediate outcomes; (3) patient experience; (4) access; and (5) process. A.R. 9. CMS uses a variety of data, including administrative and medical record review data collected as part of the Healthcare Effectiveness Data and Information Set ("Healthcare Effectiveness Data" or "HEDIS") and survey-based data from the Health Outcomes Survey and from the Consumer Assessment of Healthcare Providers and Systems ("CAHPS"). 83 Fed. Reg. at 16,520, 16,525. The 2025 Star Ratings are calculated in late 2024 using data primarily from measurement year 2023. A.R. 21. These measure-level scores are also expressed in "stars" but are awarded in whole-star increments, not half stars like the overall Star Ratings. 42 C.F.R. § 422.166(a)(4).

CMS's regulations have incorporated its Technical Notes into its operative regulations. *See* 42 C.F.R. § 422.164(a), 423.184(c) ("CMS lists the measures used for a particular Star Rating each year in the Technical Notes or similar guidance document with publication of the Star Ratings."). The regulations require CMS to, in advance of a measurement period, announce potential new measures and solicit feedback. §§ 422.164(c)(2), 423.184(c)(2). Subsequently, CMS is required to propose and finalize new measures through rulemaking. §§ 422.164(c)(2), 423.184(c)(2). "New measures added to the Part C Star Ratings program will be on the display page on *www.cms.gov* for a minimum of 2 years prior to becoming a Star Ratings measure." §§ 422.164(c)(3); 423.184(c)(3) (same for Part D). If CMS finds reliability or validity issues with

the measure specification, it will remain on display longer than two years. §§ 422.164(c)(4), 423.184(c)(4).

a. CAHPS Measures

Since 1998, CMS has conducted the Medicare Advantage CAHPS surveys annually with a sample of people with Medicare, currently enrolled in a Medicare Advantage contract for six months or longer, and who live in the United States. CMS, *Medicare CAHPS Fact Sheet* 1 (Mar. 2024) https://perma.cc/4E8A-C8VT. For 2025, nine of the 40 unique quality measures used CAHPS data as their primary data source. A.R. 125. For example, the quality measure, "Ease of Getting Prescriptions Filled When Using the Plan" is one of the nine measures that relies on CAHPS data. A.R. 96. Medicare Advantage enrollees were asked some formulation of the question, "In the last six months, how often was it easy to use your prescription drug plan to get the medicines your doctor prescribed?" A.R. 96. The score for this measure uses the mean of the distribution of responses converted to a scale from 0 to 100, and the score is the percentage of the best possible score each contract earned. A.R. 96.

1. Case-Mix Adjustment

CMS's regulations authorize CMS to apply a case-mix adjustment to CAHPS quality measures. The regulations explain what the term means: "Case-mix adjustment means an adjustment to the measure score made prior to the score being converted into a Star Rating to take into account certain enrollee characteristics that are not under the control of the plan." 42 C.F.R. § 422.162(a). The regulation continues, "[f]or example age, education, chronic medical conditions, and functional health status that may be related to the enrollee's survey responses." *Id.* Most, but not all of CAHPS measures are case-mix adjusted. *See* A.R. 36-113. The regulations explicitly contemplate that at least some quality measures will be case-mix adjusted. Section

422.166(f)(2)(ii) requires that "[i]n determining the [Categorical Adjustment Index] values, a measure will be excluded from adjustment if the measure" "is already case-mix adjusted for socioeconomic status." §§ 422.166(f)(2)(ii)(A), 423.186(f)(2)(ii). And the 2018 final rule expects that CAHPS measures could be case-mix adjusted: "CAHPS measure specification, including case-mix adjustment, is described in the Technical Notes." 83 Fed. Reg. at 16,537.

Consistent with regulations, 2025 technical guidance explains that CAHPS measures are case-mix adjusted to "take into account the mix of enrollees." A.R. 114. The case-mix variables include age, education, general health status, and various measures of income. A.R. 114-15. As an example, contracts with higher proportions of beneficiaries who are in the 75-79 age range will be adjusted downward on this measure to compensate for the positive response tendency of their respondents. A.R. 114. CAHPS measure case-mix adjustments are calculated each year with current data and may be positive or negative. CMS makes case-mix adjustment data available online. *See* CMS, Medicare Advantage and Prescription Drug Plan CAHPS Survey: Scoring and Star Ratings (Nov. 7, 2024) https://ma-pdpcahps.org/en/scoring-and-star-ratings; A.R. 809, 824. Additionally, CMS provided non-public data associated with the specific measures on which Elevance's contracts were evaluated. Pls.' App. 41.

2. CAHPS Measure Star Ratings Calculation

The regulation describes the method for calculating the raw CAHPS survey data into measure-level stars. 42 C.F.R. §§ 422.166(a)(3), 423.186(a)(3). The method CMS uses to calculate the CAHPS measure-level Star Ratings is called "relative distribution and significance testing." 42 C.F.R. §§ 422.166(a)(3), 423.186(a)(3). This method "combines evaluating the relative percentile distribution with significance testing and accounts for the reliability of scores produced from survey data." 42 C.F.R. §§ 422.166(a)(3), 423.186(a)(3). "[N]o measure Star

Rating is produced if the reliability of a CAHPS measure is less than 0.60." 42 C.F.R. §§ 422.166(a)(3), 423.186(a)(3).

For example, to obtain 5 stars, a contract's CAHPS measure score needs to be ranked at least at the 80th percentile and be statistically significantly higher than the national average CAHPS measure score, as well as either not have low reliability or be more than one standard error above the 80th percentile. 42 C.F.R. §§ 422.166(a)(3)(v), 423.186(a)(3)(v); A.R. 18. To obtain 1 star, a contract's CAHPS measure score needs to be ranked below the 15th percentile and be statistically significantly lower than the national average CAHPS measure score, as well as either not have low reliability or be more than one standard error below the 15th percentile. 42 C.F.R. §§ 422.166(a)(3)(i), 423.186(a)(3)(i).

To calculate CAHPS measure scores in accordance with the above-described specifications, CMS must calculate the national average. To calculate the national average for each CAHPS measures, CMS weights the contract scores by the survey-eligible contract enrollment assessed at the time of sample design, and then averages them. Given that the number of enrollees covered by each MAO contract varies significantly contract to contract, CMS calculates the national average to account for the number of enrollees in each contract to create a fair comparison of these customer-level satisfaction or patient experience of care scores. CMS, *Summary of Analyses for Reporting, MA & PDP CAHPS* 2 (Aug. 2024) https://perma.cc/E626-FQ2N; A.R. 9.

b. Non-CAHPS Measure Star Ratings Calculation

The measures predicated on the data other than the CAHPS survey data are collectively called the "non-CAHPS measures." These measures come from four data sources, including data from health and drug plans such as the HEDIS, administrative data, data collected from CMS contractors, and non-CAHPS survey data. A.R. 13. Where the regulations require relative

distribution and significance testing for CAHPS measures, they require the clustering algorithm for non-CAHPS measures. 42 C.F.R. § 422.166(a)(2); A.R. 17. CMS applies the clustering algorithm to the measure's numeric value scores from all contracts. A.R. 17. Conceptually, the clustering algorithm identifies the "gaps" among the scores and creates four demarcations or "cut points" resulting in the creation of five levels (one for each Star Rating). A.R. 17. The scores in the same Star Rating level are as similar as possible; the scores in different Star Rating levels are as different as possible. A.R. 17. Star Rating levels 1 through 5 are assigned with 1 being the worst and 5 being the best. A.R. 17.

CMS regulations require mean resampling to achieve a random separation of the applicable contracts into groups. 42 C.F.R. §§ 422.162(a), 423.182(a). Mean resampling "refers to a technique where measure-specific scores for the current year's Star Ratings are randomly separated into 10 equal-sized groups." § 422.162(a). The contracts are divided randomly into ten groups. *Id.* CMS runs the clustering algorithm ten times, each time leaving out one of the ten groups of randomly assigned contracts. At the end, each measure has ten sets of calculated cut points. *Id.* CMS takes the average of the ten sets of cut points, the goal being to improve the stability of cut points and reduce year-over-year variation and sensitivity in cut points to individual contract scores.

The seed—the sequence of numbers, here 8-6-7-5-3-0-9—randomly determines which of the ten groups each contract is placed into. The seed is the starting point for mean resampling. A seed that can be easily remembered and reused is necessary to replicate results of a randomization process. Pls.' App. 54. Consequently, CMS uses the seed 8-6-7-5-3-0-9, a reference to the Tommy Tutone song. CMS uses the same seed every year. Because the group of contracts differ from year to year, CMS receives random, non-predictable outcomes. Defs.' App. 2 ¶ 3 (Goldstein

Decl. ¶ 3).[2] The value of the seed can change the cut points because it changes the way the contracts are grouped into the 10 groups. *Id.* However, any particular seed value is equally likely to result in upward and downward movement in the cut points across the set of measures compared to all of the other seeds that could have been chosen. *Id.* There is no way to choose in advance a particular seed that will benefit or harm contracts relative to another seed. *Id.*

To test its calculation of non-CAHPS measures, CMS's Star Ratings contractor, RAND, ran 1,000 simulations using 1,000 randomly generated seeds for mean resampling for non-CAHPS measure cut points and recalculated the overall Star Ratings across all contracts for each of the 1,000 simulations. Defs.' App. 3 ¶ 4 (Goldstein Decl. ¶ 4.) The analyses demonstrate strong stability in the Star Ratings methodology with respect to the choice of random seed. *Id.* On average across contracts, 92.4% of simulations/seeds resulted in the same 2025 overall Star Ratings as the 2025 overall Star Ratings calculated using the original seed. *Id.* This result means that changing the seed used in the Star Ratings methodology would not change the 2025 overall Star Ratings for 92.4% of contracts on average across simulation seeds. *Id.* The most common overall Star Rating assigned to each contract across the 1,000 simulations was the same as the 2025 overall Star Rating for 92% of contracts. *Id.* Only 45 contracts (8%) had a different most common overall Star Rating across the 1,000 simulations compared to their 2025 overall Star Rating. Twenty-four contracts had a most common overall Star Rating across the 1,000 simulations that was a half star lower than their 2025 overall Star Rating. *Id.* Twenty-one contracts had a most common overall Star Rating across the 1,000 simulations that was a half star higher than their 2025 overall Star Rating. *Id.*

[2] In compliance with Local Rule 56.6(a), this motion is supported by a separately filed appendix, Defendants' Appendix in Support of Their Summary Judgment Briefing. Citations to "Defs.' App." refer to that appendix.

2. Overall Star Ratings Calculation Methodology

CMS calculates summary and overall ratings[3] using the 40 unique quality measures. The overall rating for a contract is calculated using the average of the Part C and Part D measure Star Ratings. 42 C.F.R. §§ 422.166(d)(1), 423.186(d)(1); A.R. 20. The average is weighted based on measure type because not all measures are equally important. CMS assigns the highest weight to the improvement measures,[4] followed by patient experience, complaints and access measures, then outcome and intermediate outcome measures, and finally process measures. *See* 42 C.F.R. § 422.166(e); A.R. 20. New measures are assigned the same weight as process measures for the first year in the Star Ratings. 42 C.F.R. §§ 422.166(e)(1)(v), 423.166(e)(2).

Two adjustments are made to the results of the summary and overall calculations described above: reward factor and categorical adjustment index (CAI). 42 C.F.R. §§ 422.166(f)(1), 422.166(f)(2), 423.186(f)(1), (f)(2); A.R. 21-23. First, to reward consistently high performance, CMS uses both the mean and the variance of the measure stars to differentiate contracts for overall ratings. If a contract has both high and stable relative performance, a reward factor is added to the contract's ratings. The Reward Factor is 0.0, 0.1, 0.2, 0.3, or 0.4 added to the weighted average star rating. A.R. 21, 23. Second, the overall ratings include the CAI, which is added to or subtracted from a contract's summary and overall ratings. 42 C.F.R. § 422.166(f)(2); A.R. 21, 23-

[3] This brief uses the phrase "overall ratings" to refer to both summary and overall ratings. Technically, they are different ratings. The Part C and Part D summary ratings are calculated by taking a weighted average of the measure stars for Parts C and D, respectively. A.R. 20. For Medicare Advantage Prescription Drug plans to receive an overall rating, the contract must have stars assigned to both the Part C and Part D summary ratings. Plans that do not only receive a summary rating.

[4] Both the Part C and Part D improvement measures are based on a comparison of a contract's current and prior year measure scores. A.R. 13. The ultimate improvement measure score is a complicated combination of the net improvement for process measures, for outcome and intermediate outcome measures, and access measures divided by all of the eligible measures in each of those measure categories. A.R. 134-35.

30. The CAI adjusts for the average within-contract disparity in performance associated with the percentage of beneficiaries who receive a low-income subsidy, are dual eligible (meaning eligible for both Medicare and Medicaid), or have a disability status. 42 C.F.R. § 422.166(f)(2); A.R. 23-24. Some measures are included in the CAI adjustment, and some are not. Section 422.166(f)(2)(ii)(A) requires that "[i]n determining the CAI values, a measure will be excluded from adjustment if the measure" "is already case-mix adjusted for socioeconomic status."

Standard rounding rules are applied to convert the results of the final summary and overall ratings calculations into the publicly reported Star Ratings. A.R. 21. Regulations establish that "the overall rating is on a 1- to 5-star scale ranging from 1 (worst rating) to 5 (best rating) in half-increments using traditional rounding rules." 42 C.F.R. § 422.166(d)(2)(iv); A.R. 21. "Traditional rounding rules mean that the last digit in a value will be rounded." 42 C.F.R. § 422.162(a). The regulation further instructs that if rounding to a whole number, for example, CMS will look to the digit in the first decimal place; if it is lower than five, the digit in the first decimal place should be deleted. *Id.* If it is 5 or greater, that digit in the first decimal place should be deleted and the digit to the left of the decimal point should be increased by 1. *Id.*

"Overall ratings are calculated with at least six digits of precision after the decimal whenever the data allow it." A.R. 22. Overall scores are displayed in the technical guidance and elsewhere with six decimal places, and they are calculated to at least the sixth decimal places so they can be rounded once to the half star. Defs.' App. 2 ¶ 2 (Goldstein Decl. ¶ 2).

B. Procedural History

Plaintiffs Elevance Health and direct or indirect subsidiaries of Elevance Health, Community Insurance Company, Freedom Health Inc., Group Retiree Health Solutions, Inc., Wellpoint Insurance Company, Wellpoint Texas, Inc. (together "Elevance"), filed an amended

complaint on December 23, 2024. Am. Compl., ECF No. 22. Defendants filed the certified index and administrative record on December 27, 2024. A.R., ECF No. 23, 23-1. On February 14, 2025, Elevance filed its motion for summary judgment. Pls.' Mem. in Support of Mot. Summ. J. ("Pls.' Br."), ECF No. 32. Elevance contends that four contracts, H3655, H6078, H5427, and H8849 were harmed by CMS's allegedly improper Star Ratings calculation.[5] Pls.' Br. 21-22. But Elevance only argues that its preferred calculation methodology would have resulted in an overall half-star increase for contracts H3655 and H6078. Pls.' Br. 21-22. Elevance's overall scores for contracts H3655 and H6078 were 3.749565 and 3.220943, respectively. App. in Supp. of Pls.' Mot. for Summ. J. ("Pls.' App.") 11, ECF No. 33. For the other contracts, Elevance admits that the overall Star Rating would be the same. Pls.' Br. 21-22.

III. Legal Standards

A. Legal Standard for Summary Judgment

"Summary judgment is [the] appropriate procedure for resolving a challenge to a federal agency's administrative decision when review is based upon the administrative record, even though the Court does not employ the standard of review set forth in the rule governing summary judgment motions." *Larson v. Geren*, No. SA-08-CA-722, 2010 WL 11542078, at *4 (W.D. Tex. Apr. 14, 2010) (internal quotation marks omitted), *aff'd*, 432 F. App'x 356 (5th Cir. 2011). The Fifth Circuit has "consistently upheld, without comment, the use of summary judgment as a mechanism for review of agency decisions." *Girling Health Care, Inc. v. Shalala*, 85 F.3d 211, 214 (5th Cir. 1996).

As this Court has noted, summary judgment is proper when there is no genuine dispute about the material facts and judgment can be rendered as a matter of law:

[5] Elevance appears to have abandoned Wellpoint Texas, Inc.'s claims because it does not include its H2593 contract as one of the allegedly harmed contracts. Pls.' Br. 21-22.

> Summary judgment is proper where "there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law." Fed. R. Civ. P. 56(a). A dispute is "genuine" if, based on the evidence, "a reasonable jury could return a verdict for the nonmoving party." *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 248, 106 S. Ct. 2505, 91 L.Ed.2d 202 (1986). A fact is "material" if it would affect a case's outcome. *Id.* Generally, the "substantive law will identify which facts are material" and "[f]actual disputes that are irrelevant or unnecessary will not be counted." *Id.* In assessing if summary judgment is warranted, the Court "view[s] all evidence in the light most favorable to the nonmoving party and draw[s] all reasonable inferences in that party's favor." *Cunningham v. Circle 8 Crane Servs., LLC*, 64 F.4th 597, 600 (5th Cir. 2023).

Am. Hosp. Ass'n v. Becerra, 738 F. Supp. 3d 780, 790 (N.D. Tex. 2024) (Pittman, J.).

B. Legal Standard for an APA Challenge of an Agency's Decision

"Judicial review has the function of determining whether the administrative action is consistent with the law—that and no more." *Girling*, 432 F. App'x at 215 (quoting 10A Charles Alan Wright, Arthur R. Miller & Mary Kay Kane, *Federal Practice and Procedure: Civil 2d* § 2733 (1983)). In this action challenging CMS's Star Ratings determinations, judicial review is governed by the scope of review in § 706 of the Administrative Procedure Act ("APA"). 5 U.S.C. § 706. A court's review of agency action alleged to be arbitrary and capricious is deferential to the agency:

> The scope of review under the "arbitrary and capricious" standard is narrow and a court is not to substitute its judgment for that of the agency. Nevertheless, the agency must examine the relevant data and articulate a satisfactory explanation for its action including a "rational connection between the facts found and the choice made."

Motor Vehicle Mfrs. Ass'n v. State Farm Mutual Auto Ins. Co., 463 U.S. 29, 43 (1983). The *State Farm* standard is satisfied if the final agency finding is supported by "such relevant evidence as a reasonable mind might accept as adequate to support a conclusion." *Consolo v. Fed. Mar. Comm'n*, 383 U.S. 607, 619–20 (1966) (internal quotation marks and citation omitted). The standard "is something less than the weight of the evidence, and the possibility of drawing two

inconsistent conclusions from the evidence does not prevent an administrative agency's finding from being supported by substantial evidence." *Id.* at 620.

IV. Argument and Authorities

A. CMS Acted Reasonably and in Accordance with Law by Applying a Case-Mix Adjustment to Certain CAHPS Measures.

CMS case-mix adjusted the CAHPS measures Elevance challenges here in accordance with CMS regulations. CMS's regulations allow for case-mix adjustments in the calculation of measure scores:

> Case-mix adjustment means an adjustment to the measure score made prior to the score being converted into a Star Rating to take into account certain enrollee characteristics that are not under the control of the plan. For example age, education, chronic medical conditions, and functional health status that may be related to the enrollee's survey response.

42 C.F.R. § 422.162(a); *see also* §§ 422.162(a), 422.164, 422.166 (providing definitions applicable to calculating Star Ratings, including definitions that instruct how to add, update, remove, and calculate measures). The regulations allow, but do not mandate, that CAHPS measures be case-mix adjusted because the regulations themselves do not prescribe the measures CMS adds or the sources of data those measures use. *See* 42 C.F.R. § 422.164(c). Instead, the regulations require CMS to announce potential new measures and solicit feedback before proposing and finalizing them through rulemaking, § 422.164(c)(2), and to follow notice-and-comment rulemaking for substantive updates, § 422.164(d). The regulations are together nonprescriptive; they confer significant discretion upon CMS to engage in an iterative process to develop measures that are "nationally endorsed," "align with the private sector," "appropriate to measure," and "reflect performance specific to the Medicare program." 42 C.F.R. § 422.164(c)(1). And it is through this iterative measure development process that CMS determines whether a particular measure is to be case-mix adjusted.

To determine whether a particular measure is case-mix adjusted, as the regulations provide, CMS lists the measures used for each particular Star Rating year "in the Technical Notes or similar guidance document with publication of the Star Ratings." 42 C.F.R. §§ 422.164(a), 423.184(c). When CMS adds new case-mix adjusted measures through rulemaking, it specifies that "more specific identification of a measure's . . . case-mix adjustment" will be provided in the Medicare Part C & D Star Ratings Technical Notes as required by § 422.164(a). *See* Medicare Program; Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Program for Contract Year 2024, 89 Fed. Reg. 30,448, 30,636-37 (Apr. 23, 2024) (specifying measures applicable to the 2027 Star Ratings); *see also* Medicare Program; Contract Year 2024 Policy and Technical Changes, 88 Fed. Reg. 22,120, 22,270-71 (Apr. 12, 2023) (same for 2026 Star Ratings); 83 Fed. Reg. at 16,537 ("[R]esponses are also case-mix adjusted to account for certain respondent characteristics not under the control of the health or drug plan such as age, education, dual eligible status and other variables."). Following this regulatory process, CMS implemented the case-mix adjustments to the CAHPS measures that Elevance challenges.

CMS rulemakings expressly authorize case-mix adjusting for the CAHPS measures on which Elevance was evaluated for the 2025 Star Rating year. As part of the 2018 rulemaking, CMS reproposed and finalized all of its existing case-mix adjusted measures, including eight CAHPS measures that were case-mix adjusted and on which Elevance was evaluated for the 2025 Star Ratings year:

1. Getting Needed Care;
2. Getting Appointments and Care Quickly;
3. Customer Service;
4. Rating of Health Care Quality;

5. Rating of Health Plan,
6. Care Coordination,
7. Rating of Drug Plan; and
8. Getting Needed Prescription Drugs.

83 Fed. Reg. at 16,549; A.R. 70, 71, 72, 74, 75-76, 76-77, 87, 88. In that final rule, CMS explained that CAHPS measures are case-mix adjusted: "For CAHPS measures, contracts are first classified into base groups by comparisons to percentile cut points defined by the current-year distribution of *case-mix adjusted* contract means." 83 Fed. Reg. at 16,568 (emphasis added). In response to a comment in the 2018 final rule requesting more insight into the relationship between case-mix adjusting and CAHPS measures, CMS indicated that it "provides a detailed explanation of the CAHPS methodology including case-mix adjustment in the annual Star Ratings Technical Notes, in CAHPS plan reports provided to each contract each year, and on the MA and PDP CAHPS web page (*https://www.mapdpcahps.org*)." 83 Fed. Reg. at 16,555. CMS stated further that, "CMS also provides survey vendors all of the necessary data to perform case-mix adjustment validation." *Id.* In short, given this regulatory scheme, the eight CAHPS measures at issue were appropriately case-mix adjusted.

Elevance, however, contends that CMS has no authority to case-mix adjust CAHPS measures. As support for its contention, Elevance points to the regulations regarding the Categorical Adjustment Index (CAI) and the health equity index containing the words "case-mix adjusted" to claim that "CMS knows exactly how to authorize case-mix adjustments in other areas of the Star Ratings regulations." Pls.' Br. 20-21 (citing to sections 422.166(f)(2) and 422.166(f)(3)). Those cited regulations, however, do not authorize case-mix adjustments per se; they simply describe how to account for measures that have *already* been case-mix adjusted as the

CAI and health equity index calculations occur *after* the calculation of scores for individual measures, such as CAHPS measures. *See* 42 C.F.R. § 422.166(f) (entitled "Completing the Part C summary and overall ratings calculations"). The health equity index regulation lays out how to account for "[m]easures that are case-mix adjusted in the Star Ratings" in calculating the health equity index. *Id.* § 422.166(f)(3)(i)(A). The CAI regulation states that individual measure scores that are "already case-mix adjusted" are excluded from the CAI. *Id.* § 422.166(f)(2)(ii)(A) ("In determining the CAI values, a measure will be excluded from adjustment if the measure . . . is already case-mix adjusted for socioeconomic status.").

Indeed, the rule implementing the CAI makes clear that CAHPS measures specifically are excluded from the CAI calculation under this regulation because they are already case-mix adjusted for socioeconomic status. *See* 83 Fed. Reg. at 16581 ("Measures would be excluded as candidates for adjustment if the measures are already case-mix adjusted for [socioeconomic status] (for example, CAHPS and HOS outcome measures.")). The applicable CAI rule, therefore, assumes that CAHPS measures *are* case-mix adjusted—far from showing that CAHPS measures cannot be case-mix adjusted. Said another way, if case-mix adjusting CAHPS measures violates CMS's regulations as Elevance contends, the CAI regulations would not have to exclude them from the CAI calculation on the basis that they are "already case-mix adjusted."

Elevance's argument that references to case-mix adjustments in unrelated statutes is likewise unavailing. CMS's use of case-mix adjustments in Star Ratings differs substantially from the uses Elevance cites from other statutes, which do not relate to Medicare Advantage—let alone the Star Ratings—or the iterative process established by CMS's regulations to develop quality measures. *See* Pls.' Br. 20. For example, 42 U.S.C. § 1395rr(b)(12) states that "[t]he Secretary shall establish a basic case-mix adjusted prospective payment system for dialysis services." Such an across-the-board requirement to use case mix in the End Stage Renal Disease Program does not apply to Star Ratings measures, most of which are not case-mix adjusted. While 42 U.S.C. § 1395fff(b)(3)(B)(iv) permits the Secretary to make some "case-mix changes" in payments under the prospective payment system for home health services, that statutory provision only allows such changes in extremely limited circumstances. *See* § 1395fff(b)(3)(B)(iv). By design, such case-mix prescriptions do not exist under the Star Ratings regulatory scheme devised for flexible measure creation and design, nor do any of these statutes bear on the Star Ratings system in any way.

Finally, in support of its contention that there is "no regulatory support" for CMS case-mix adjusting CAHPS measures for two of its contracts, H3655 and H6078, Elevance wrongly contends that "CMS's sub-regulatory guidance *adds* the case-mix adjustment," Pls.' Br. 19 (emphasis added). Elevance's suggestion that CMS establishes case-mix adjustments in its technical guidance separate and apart from the regulations is wrong. As explained above, CMS regulations provide that sub-regulatory guidance is where details about case-mix adjustments are found, which is all part of the overall regulatory scheme applicable to Star Ratings. *See* 42 C.F.R. § 422.164(a) ("CMS lists the measures used for a particular Star Rating each year in the Technical Notes").

CMS has been case-mix adjusting CAHPS measures since the initiation of CAHPS in 1998. CMS, *MA & PDP CAHPS Variables Used as Case-Mix Adjustors 1998-2024* (July 30, 2024) https://ma-pdpcahps.org/globalassets/ma-pdp/scoring-and-star-ratings/2024/case-mix-variables.pdf [https://perma.cc/W8WS-4T5U]. And to CMS's knowledge, no MAO has suggested that case-mix adjustments are not authorized by CMS's regulations. Tellingly, Elevance does not challenge case-mix adjusting—that is, taking into account differences in the characteristics of enrollees across contracts that may potentially impact survey responses—as unreasonable. Case-mix adjusting is not a policy to which Elevance objected or would have objected before Elevance knew its overall contract scores.

This Court should reject Elevance's contention that there is "no regulatory support" for CMS case-mix adjusting measures scores.[6] Case-mix adjusting CAHPS measures is amply supported by CMS's rules and regulations.

B. CMS Acted Reasonably and in Accordance with Law When It Calculated the National Average for CAHPS Measure Scores.

As an initial matter, Elevance has waived its challenge to CMS's calculation of the national average because it did not raise this challenge in its Amended Complaint. *See* Am. Compl. 20-23, ECF No. 22 ("Claims for Relief"). Elevance cannot move for summary judgment on claims it did not raise in its Amended Complaint, and this Court cannot grant relief that Elevance did not request. *See Johnson v. Thibodaux City*, 887 F.3d 726, 736 (5th Cir. 2018) (reasoning that the only claims that are relevant at the summary-judgment stage are those the plaintiff pleaded).

[6] While Elevance initially alleged that CMS "hid[] case-mix adjustment data and prevent[ed] plans from validating calculations for their Star Ratings," Am. Compl. ¶ 59, Elevance appears to have abandoned this claim, perhaps because CMS has provided the relevant data. *See* Pls.' App. 46.

Even if Elevance had not waived this challenge, Elevance's claim fails on the merits. When calculating overall Star Ratings, CMS evaluates contracts against the national average CAHPS measure score in accordance with its regulations. CMS's regulations provide that the Star Rating a contract receives on a given CAHPS measure is in part based on whether that contract's "average CAHPS measure score is statistically significantly lower than the national average CAHPS measure score" or "statistically significantly higher than the national average CAHPS measure score." 42 C.F.R. § 422.166(a)(3)(i)-(v). Elevance contends that CMS guidance requires contracts to be compared to a "weighted national average," which Elevance contends is at odds with CMS's regulations. Pls.' Br. 22. This is wrong. CMS's guidance requires CMS to assign measure stars based in part on the "statistical significance of the difference of the contract mean from the national mean," not the national contract mean. A.R. 161. Elevance's arguments are premised on a fundamental misunderstanding about the requirements of § 422.166(a)(3), which only requires CMS to compare a contract's CAHPS measure score to the "national average." This is exactly what CMS did.

To calculate the national average of CAHPS measure scores, CMS takes the contract-level scores for each CAHPS measure, weights those scores by beneficiary enrollment, and then averages those scores. CMS, *Summary of Analyses for Reporting, MA & PDP CAHPS* 2 (Aug. 2024) https://perma.cc/E626-FQ2N. CAHPS scores are based on "surveys that ask [Medicare beneficiaries with Medicare Advantage Plans] to evaluate the interpersonal aspects of health care" where "consumers and patients are the best or only source of information," 42 C.F.R. § 422.162(a) (definition of CAHPS). Given that the number of consumers covered by each MAO contract varies significantly contract to contract, CMS calculates the national average to account for the number of enrollees in each contract to create a fair comparison of these customer-level satisfaction scores.

Weighting the contract to account for differences in enrollment when calculating the national average of CAHPS measure scores is reasonable and entirely consistent with what the regulation requires from the data available to CMS.

Elevance appears to contend that CMS must use the national *contract*-level average as the national average, which would increase overall Star Ratings for two of its contracts, H3655 and H6078. Pls.' Br. 24-25. Elevance errs in misconstruing CMS's regulation. The regulation directs CMS to only calculate a "national average CAHPS measure score," which CMS did. Elevance's assertion that CMS must calculate a national average of contract-level scores is simply wrong.[7]

Elevance asserts that the technical guidance requires CMS to use the "'weighted' national average of all contract scores" contrary to the "national average" language in the regulation. Pls.' Br. 22. It does not. As established, CMS does not calculate weighted national averages. Elevance asserts that, contrary to the "national average" language in the regulation, "CMS's sub-regulatory guidance requires CMS to use the 'weighted' national average *of all contract scores*." Pls.' Br. 22 (emphasis added). Nothing in that guidance suggests that CMS weights the national average of all contract scores; it instead weights *contract-level CAHPS scores* to account for enrollment to calculate the "national average" required in the regulation. Elevance's reading of the regulation effectively adds in the word "contract" such that it would say: "statistically significantly higher [or lower] than the national average CAHPS measure *contract* score." *See* Pls.' Br. 22. But the word "contract" appears nowhere in the regulation. 42 C.F.R. § 422.166(a)(3)(i)-(v).

[7] Elevance's declarant, Mark Abernathy, concludes that CMS's "improper" use of the weighted national average negatively impacted two of its contracts, and he uses the national average of contract-level CAHPS measure scores in his calculation, which he terms the "non-weighted average." Pls.' App. 14. Because the regulations do not provide for a national average of CAHPS measure based on the average of contract scores, Abernathy's calculations are not compelling.

Using the national average of all CAHPS measure contract scores without accounting for the individual enrollment in those contracts—as Elevance seeks to do here, *see* Pls.' Br. 22—would make little sense. If CMS were to take the national average of CAHPS scores at the contract level, contracts with 2,000 enrollees would influence the national average as much as contracts with 500,000 enrollees. In this kind of average, scores in very large contracts would be underweighted and scores in very small contracts would be overweighted. An apt analogy would be if, in response to a request to compute a national average, you summed the average for each state, added those averages together, and divided by 50. Citizens in states like Wyoming would have larger per capita representation than citizens in California. If CAHPS measure scores were calculated this way, it would not capture the true national average of CAHPS measure scores as the regulation intended.

Elevance cites to a series of regulations that use the phrase "weighted mean" and "weighted average" to demonstrate that CMS "know[s] how to specifically require the use of a weighted average in regulatory text." Pls.' Br. 23. CMS agrees; it deliberately chose not to include the phrase "weighted average" in its regulation. To reflect the average beneficiary experience, CMS finds the weights contract-level scores for each CAHPS measure to reach the national average. It does not weight the national average. Tellingly, Elevance cites no regulatory definitions where CMS used the phrase "weighted national average," or which otherwise use the term "weighted" in conjunction with a national average. As in this situation, weighting is often a necessary step to achieve an accurate national average, but national averages are not themselves weighted.[8]

[8] Elevance repeatedly misrelies on two cases: *United States v. Lauderdale Cnty.*, 914 F.3d 960 (5th Cir. 2019) and *Petteway v. Galveston Cnty.*, 111 F.4th 596, 633 (5th Cir. 2024). Pls.' Br. 19, 23. Those cases are inapposite because they pertain to statutory interpretation. *Galveston Cnty.*, 111 F.4th at 599 (identifying the issue in the case as how to interpret the Voting Rights Act); *Lauderdale Cnty.*, 914 F.3d at

Pursuant to its regulations, CMS weights contract-level scores by enrollment to reach the national average of these scores.

C. CMS Acted in Accordance with the Statute, Regulations, and Traditional Rounding Rules by Rounding Once to the Nearest Half-Star Increment.

Once the measure-level scores are calculated and adjusted according to the regulations, CMS calculates the overall numerical score used to determine a contract's overall Star Ratings to the nearest half-star increment. The numerical scores are generally calculated with at least six digits of precision and publicly displayed to the sixth decimal place. A.R. 30-31. CMS then takes that numerical score and rounds once to the nearest half-star increment (0.5) as required. *See* 42 C.F.R. §§ 422.162(a), 422.166(d)(2)(iv); 42 U.S.C. § 1395w-23(o)(4)(A) ("The quality rating for a plan shall be determined according to a 5-star rating system."); *see also* § 1395w-24(b)(1)(C)(v) (indicating a 5-star rating system on half-star increments).

This is precisely what CMS did here. The unrounded overall ratings score to the sixth decimal place for Elevance's H3655 contract is 3.749565. CMS initially calculated this score—like all other MAO's scores—without rounding. Consistent with the statute, regulations, and traditional rounding rules, CMS then rounded Elevance's score once to the nearest half-star increment of 3.5. Put another way, because 3.749565 is closer to 3.5 stars than it is to 4 stars, CMS rounded to 3.5 stars. Elevance, however, asks this Court to require that CMS double round its measure score, first to 3.75, then to 4.0 for no other reason than it would benefit Elevance to do so in this instance. There is no basis for the Court to require that CMS calculate overall ratings this way, and Elevance's argument should be rejected.

---

961 ("This case presents a question of statutory interpretation."). Elevance is not making an argument premised on statutory interpretation.

1. CMS Does Not Round to the Sixth Decimal.

As an initial matter, Elevance is factually wrong that CMS rounded Elevance's overall numerical score to the sixth decimal place. Pls.' Br. 26. CMS did not round Elevance's overall rating to the sixth decimal, nor does CMS's technical guidance create a policy of rounding to the sixth decimal place.

Contrary to Elevance's assertion, when CMS *calculates* overall ratings with "at least six digits of precision," that does not mean it *rounds* overall rating scores to six digits of precision. In making this claim, Elevance conflates CMS's policy concerning "Calculation Precision" and its policy concerning "Rounding Rules for Measure Scores." A.R. 30-31. These policies are separate, distinct, and in accordance with the statute and applicable regulations. CMS's calculation precision policy can be found under the heading "Calculation Precision" in the Medicare 2025 Part C & D Star Ratings Technical Notes. *See* A.R. 30. This heading is followed by a heading at the same hierarchical level, "Rounding Rules for Summary and Overall Ratings." A.R. 30-31. Under the "Calculation Precision" heading, CMS explains that "[t[he improvement measures, summary, and overall ratings are calculated with *at least* six digits of precision after the decimal whenever the data allow it." A.R. 30 (emphasis added). CMS explains the rounding rules for overall ratings under the heading "Rounding Rules for Summary and Overall Ratings," not in the section of the guidance related to calculation precision. *See* A.R. 31.

Tellingly, Elevance does not point to anything that supports its assertion that CMS rounds to six digits of precision. Instead, Elevance appears to infer—from the six-digit decimals set out in one page of technical guidance, A.R. 31, and CMS's calculation precision policy—that CMS only calculates overall scores to six decimal places and that, in doing so, CMS rounds to the sixth decimal place. But CMS could not be clearer that it calculates overall rating scores with *at least*

six digits of precision. A.R. 30. When CMS calculates overall scores, those are scores are calculated out to at least six decimal places such that the overall score can be rounded once to the half star. In rounding to the half star, the second, hundredth decimal place determines the rounding. In order to round once to the half star, it is necessary to carry enough decimal places throughout each step of the overall score calculation such that the second, hundredth decimal place is not rounded. CMS calculates each step out to the sixth decimal place because nothing that happens at sixth decimal place ever impacts the second, hundredth decimal place. Practically, this is no different than if the calculations were completed with infinite decimal places but is done *only* for the purposes of computational efficiency. Defs.' App. 2 ¶ 2 (Goldstein Decl. ¶ 2). When CMS displays the unrounded overall rating the health plan management system for contracts to view and when CMS lists the rounding rules in the technical notes, it uses the first six decimal places *only* for the purposes of display. Defs.' App. 2 ¶ 2 (Goldstein Decl. ¶ 2). It simply would not be practical for CMS to display near-infinite decimal places in these documents because they would not fit on the page. Using six decimal places for display purposes does not indicate rounding in the calculation of the overall rating, and nowhere in regulation or in technical guidance does CMS state that it rounds to six decimal places.

2. Elevance Asks This Court to Require CMS to Double Round to Reach a Higher Star Rating—a Request at Odds with Its Regulations and Traditional Rounding Rules.

CMS rounds MAOs' overall numerical scores to determine overall Star Ratings only once because its regulation requires it to round only once: "The overall rating is on a 1- to 5-star scale ranging from 1 (worst rating) to 5 (best rating) in half-increments using traditional rounding rules." 42 C.F.R. § 422.166(d)(2)(iv). Under the regulations, "[t]raditional rounding rules mean that the last digit in a value will be rounded." 42 C.F.R. § 422.162(a). The regulation further instructs that

if rounding to a whole number, for example, CMS will look to the digit in the first decimal place; if it is lower than five, the digit in the first decimal place should be deleted. *Id.* If it is five or greater, that digit in the first decimal place should be deleted and the digit to the left of the decimal point should be increased by 1. *Id.* CMS's regulation is unambiguous.

Applying traditional rounding rules to overall Star Ratings, CMS correctly determined Elevance's contract H3655 Star Rating to be 3.5. In specifying that CMS ranks plans on a 1- to 5-star scale ranging from 1 (worst rating) to 5 (best rating) *in half-increments*, Congress permitted CMS to round to the nearest half-star, or 0.5. *See* 42 U.S.C. § 1395w-23(o)(4)(A) ("The quality rating for a plan shall be determined according to a 5-star rating system."); § 1395w-24(b)(1)(C)(v) (indicating a 5-star rating system on half-star increments). Following its regulation adopting traditional rounding rules, CMS looked not to the digit in the first decimal place, but to the digit in the second decimal place. For Elevance's contract H3655 score of 3.749565, CMS looked to the digit 4. Finding that 4 is lower than 5, CMS rounded 3.74 to 3.5, not 4.0. In the margin is a table that provides a visual representation of CMS's rounding policy. Because 3.749565 is between 3.250000 and 3.750000, CMS rounded Elevance's score to 3.5.

Codified Star Ratings Methodology
(Traditional Rounding Rules)

Raw Summary / Overall Score	Final Summary / Overall Rating
≥ 0.000000 and < 0.250000	0
≥ 0.250000 and < 0.750000	0.5
≥ 0.750000 and < 1.250000	1.0
≥ 1.250000 and < 1.750000	1.5
≥ 1.750000 and < 2.250000	2.0
≥ 2.250000 and < 2.750000	2.5
≥ 2.750000 and < 3.250000	3.0
≥ 3.250000 and < 3.750000	3.5
≥ 3.750000 and < 4.250000	4.0
≥ 4.250000 and < 4.750000	4.5
≥ 4.750000 and ≤ 5.000000	5.0

Elevance asks the Court to require that CMS double round, in violation of CMS's regulations. On the basis of the table replicated in the margin and the Rounding Rules for

Summary and Overall Ratings heading in the 2025 Technical Guidance, A.R. 31, Elevance appears to assert that "CMS's technical guidance explains that you should look to the second (or hundredth place) decimal to reach the midpoint," and that consequently, using CMS's traditional rounding rules as well as "basic math principles," "you would look to the decimal to the immediate right of the second decimal (i.e., the third or thousandth place decimal) and round there." Pls.' Br. 26. This is the improper step in Elevance's analysis. The second decimal place only tells you whether you are at the midpoint or not. There is no "reach[ing] the midpoint." Either your score is above, below, or equal to the midpoint. If above or equal to the midpoint, CMS rounds up to the nearest half-star increment. If below, CMS rounds down to the nearest half-star increment. That is the *only* rounding that CMS does. Elevance's approach would require CMS to round twice, not once as the regulations provide. Elevance is adding an additional rounding step because that is the only way it can get its score to be closer to the next higher half-star increment.

Instead of rounding to whole numbers, CMS rounds to half-stars, or to 0.5 increments. *See* 42 C.F.R. § 422.266(a)(2)(ii); *see also* 42 U.S.C. § 1395w-24(b)(1)(C)(v). Applying Elevance's wished-for policy to its overall H3655 score, Elevance is asking this Court to require CMS to double round 3.749565; that is, to round twice to reach the next benchmark. CMS rounded 3.749565 to 3.5 instead of up to 4 in accordance with its technical guidance. Because 3.749565 is closer to 3.5 than it is to 4, CMS rounded down. Put another way, because 3.749565 is less than 3.750000, CMS rounded down. Elevance is asking CMS to instead first round the third decimal place, 9, up so that the 4 in the second decimal place becomes 5, resulting in 3.75. Elevance says CMS should then round a second time—3.75 should be rounded to 4. Consistent with its regulations and traditional rounding rules, CMS rounds once. To round twice would violate CMS's regulations and traditional rounding rules.

By asking this Court to require CMS to double round, Elevance is effectively asking this Court to write a new regulation, which only provides for rounding once. To see how, consider the two tables below:

Codified Star Ratings Methodology (Traditional Rounding Rules)

Raw Summary / Overall Score	Final Summary / Overall Rating
≥ 0.000000 and < 0.250000	0
≥ 0.250000 and < 0.750000	0.5
≥ 0.750000 and < 1.250000	1.0
≥ 1.250000 and < 1.750000	1.5
≥ 1.750000 and < 2.250000	2.0
≥ 2.250000 and < 2.750000	2.5
≥ 2.750000 and < 3.250000	3.0
≥ 3.250000 and < 3.750000	3.5
≥ 3.750000 and < 4.250000	4.0
≥ 4.250000 and < 4.750000	4.5
≥ 4.750000 and ≤ 5.000000	5.0

Elevance Double Rounding

Raw Summary / Overall Score	Final Summary / Overall Rating
≥ 0.000000 and < 0.245000	0
≥ 0.245000 and < 0.745000	0.5
≥ 0.745000 and < 1.245000	1.0
≥ 1.245000 and < 1.745000	1.5
≥ 1.745000 and < 2.245000	2.0
≥ 2.245000 and < 2.745000	2.5
≥ 2.745000 and < 3.245000	3.0
≥ 3.245000 and < 3.745000	3.5
≥ 3.745000 and < 4.245000	4.0
≥ 4.245000 and < 4.745000	4.5
≥ 4.745000 and ≤ 5.000000	5.0

In both tables, an overall score is rounded by determining which range it fits into. In the first table, if a score is between 3.250000 and 3.750000, for example, the overall rating would be 3.5. In the second table, if a score is between 3.245000 and 3.745000, the overall rating would be 3.5. The first table is a simple illustration of the ordinary rounding rules CMS relies upon to create Star Ratings. A.R. 31. The second is the double rounding scheme Elevance has devised and asserts CMS's regulations require. Which one looks more arbitrary? The second table exposes the weakness in Elevance's argument here: the rule changes for which it advocates are not neutral, statistically sound principles that Elevance did, or ever would, propose before it knew its overall scores. Elevance's only goal is to push its own scores over the line into the next half-star category. This Court should reject Elevance's attempt to re-write the rules.

3. By Requiring CMS to Calculate Star Ratings in Half-Star Increments—to 0.50—CMS Only Needs to Use a Precision of 0.5.

Elevance next contends that "[c]alculating and rounding a contract's Final Summary Score to the sixth decimal is arbitrary and capricious because it ignores the imprecision of the methodology used by CMS to determine a plan's Final Summary Score." Pls.' Br. 28. That argument fails for two reasons.

First, as established above, CMS does not round to the sixth decimal. *See supra* Part IV.C.1. CMS calculates out to at least six decimal places in order to round once to the half star. *See supra* Part IV.C.1. All of Elevance's arguments concerning precision are premised on this basic misconception about how CMS calculates Star Ratings. So all of Elevance's arguments about precision are invalid because of that shared, incorrect premise.

In addition, Elevance's arguments pertaining to precision fail because its expert's opinions are fundamentally flawed. Elevance contends that CMS uses a precision of 0.000001 to calculate the overall Star Ratings. Pls.' Br. 28-29. But, as established, CMS rounds once to the half star (consistent with statute and regulation) to determine the overall rating. *See supra* Part IV.C.2. This means the precision CMS uses in determining the overall rating is 0.5. Given the regulatory requirement that CMS ranks plans on a 1- to 5-star scale ranging from 1 (worst rating) to 5 (best rating) in half-star increments, CMS only need to use a precision of 0.5. Nowhere do the statute, regulations, or guidance require CMS to account for statistical uncertainty to a greater decimal than 0.5.

Elevance relies on Dr. Paul Diver's analysis to support its contention that CMS's calculation of non-CAHPS measures fails to account for imprecision, meaning the likelihood that a calculation methodology can change due to randomness and chance alone instead of the metrics CMS really intends the Star Ratings to reflect. Pls.' Br. 29. Dr. Diver reaches the conclusion that

"[e]valuating the Final Summary Score at the sixth decimal place when determining an Overall Star Rating can be improperly interpreted as analyzing the Overall Star Rating at a higher level of statistical precision than what is presented based upon the estimation methodology used by CMS." Pls.' App. 26. He points to the difference between 3.749565 and 3.750000—0.000435—and concludes that the difference between 3.5 and 4.0 stars should not hinge on such a miniscule number that is, according to his calculations, "smaller than the average statistical uncertainty due to random chance inherent in CMS's methodology." Pls.' App. 3.

Dr. Diver's analysis is flawed. Dr. Diver's point makes intuitive sense, but only if you forget that CMS doesn't round to the second decimal place (i.e., the hundredth place). *See supra* Part IV.C.1. Rather, CMS rounds once to the first decimal place (i.e., the tenth place) to arrive at the nearest half star, to 0.5. *Id.* Dr. Diver errs when looking to the difference between 3.749565 and 3.750000. This is because CMS does not round to 3.750000. Again, CMS only rounds once, not twice. *See supra* Part IV.C.2. Under ordinary rounding rules, Dr. Diver should subtract 3.5 from 3.749565 to get 0.249565 (and because it is less than .25, it indicates that CMS should round down). Under Elevance's double rounding rules, the rounding error would constitute the leap from 3.749565 to 3.75, then to 4.0, which would result in an increased rounding error of 0.250435. More importantly, the rounding error is much higher than 0.01, which Dr. Diver concludes is the amount of "statistical uncertainty due to random chance associated with the calculation of the Final Summary Score." Pls.' Br. 29; Pls.' App. 26. The difference for Elevance's H3655 contract between a 3.5 Star Rating and a 4.0 Star Rating is not 0.000435—it's .249565. Put another way, the precision CMS is using when determining overall ratings is 0.5, which is larger than 0.01. Based on Elevance's own expert, CMS could round the overall Star Ratings using a precision of 0.01 and this would be sufficient to address the statistical uncertainty Dr. Diver identifies. CMS's

use of a precision of 0.5 is even less sensitive to the randomness and chance alone. Bottom line: Elevance's criticism of how CMS accounts for precision is misplaced for two reasons: (1) it's based on the false assumption that CMS rounds to the sixth decimal; and (2) it's based on its expert's fundamentally flawed opinions.

a. Dr. Diver's Mean Resampling Simulation Methodology Is Flawed.

As part of his analysis to determine variability in final summary scores due to random chance, Dr, Diver undertook a simulation. Pls.' App. 40. In performing this simulation, Dr. Diver made two critical errors that impact the calculation of the overall Star Ratings and undermine his conclusions about the extent to which non-CAHPS measures fail to account for imprecision. First, Dr. Diver failed to recalculate reward factor thresholds. *See supra* p. 13 (describing the reward factor threshold calculation). Second, Dr. Diver failed to recalculate the Part C and Part D improvement measure scores and Star Ratings to account for the measure-level hold harmless provision. These oversights make his precision analysis inaccurate and unreliable.

When undertaking his simulations, Dr. Diver acknowledges that although using a different seed has the potential to result in different cut points and therefore different measure-level stars for some contracts on some measures, it is also necessary to recalculate the reward factor thresholds in each simulation. Pls.' App. 53. Dr. Diver explains that he was unable to recalculate the reward factor because "[t]he data provided by CMS in response to this suit were de-identified, notably for the non-Elevance contracts." Pls.' App. 53. In other words, because CMS did not identify the data with specific contracts for contracts that were not Elevance's, Dr. Diver evaluated "each contract's respective simulation performance against the official published Reward Factor thresholds." Pls.' App. 53. This is a fundamental flaw. For each contract's Final Overall Star Rating that were not Elevance's, the reward factors used by Dr. Diver bear no relation to the

measure-specific Star Ratings. The conclusions Dr. Diver draws from his simulations are based on incorrect analyses and cannot be trusted as an accurate representation of the impact of changing the seed on the overall Star Ratings.

Dr. Diver's analysis also fails to sufficiently take into account the nuances of CMS's calculation of improvement measures. Both the Part C and Part D improvement measures are based on a comparison of a contract's current and prior year measure scores. A.R. 13. Part of what determines whether there has been net improvement in each of the process, outcome, and access measure categories is whether a contractor has achieved statistically significant improvement or decline on an attainment measure. A.R. 134. CMS recognized, however, that sometimes contracts will demonstrate statistically significant decline on an attainment measure in situations where those contracts receive 5 stars during both the current contract year and prior contract year. A.R. 134. Consequently, CMS will "hold harmless" such measures, and they "will be counted as showing no significant change." A.R. 134. There is no evidence that Dr. Diver recalculated the Part C and Part D improvement scores to account for the potentiality that 5-star measure-level achievement on measures for the current year and previous year should render applicable measures held harmless.

b. CMS Conducted Its Own Larger, More Accurate Simulation.

Like Dr. Diver, CMS conducted its own simulation to test the stability of the Star Ratings calculation methodology with respect to the utilization of random seeds. Instead of 100 unique seeds, CMS used 1,000. CMS's Star Ratings contractor RAND ran 1,000 simulations using 1,000 randomly generated seeds for mean resampling for non-CAHPS measure cut points and recalculated the overall Star Ratings across all contracts for each of the 1,000 simulations. Defs.' App. 3, ¶ 4 (Goldstein Decl. ¶ 4). Unlike Dr. Diver, RAND fully recalculated the ratings for each

of these 1,000 simulations including cut points, measure star assignments, Part C and D improvement measure calculations, and reward factor calculations so they could understand the impact of using different seeds on the overall Star Rating. *Id.* Overall Star Ratings from each simulation were compared to the 2025 overall Star Ratings that used the seed 8-6-7-5-3-0-9. *Id.* RAND also calculated the most common overall Star Rating assigned to each contract across the 1,000 simulations/seeds and compared it to each contract's 2025 overall Star Rating. *Id.*

The analyses demonstrate strong stability in the Star Ratings methodology with respect to the choice of random seed. *Id.* On average across MA-PD contracts, 92.4% of simulations/seeds resulted in the same 2025 overall Star Ratings as the 2025 overall Star Ratings calculated using the original seed. *Id.* This means that changing the seed used in the Star Ratings methodology would not change the 2025 overall Star Ratings for 92.4% of contracts on average across simulation seeds. *Id.* The most common overall Star Rating assigned to each contract across the 1,000 simulations was the same as the 2025 overall Star Rating for 92% of contracts. *Id.* Only 45 contracts (8%) had a different most common overall Star Rating across the 1,000 simulations compared to their 2025 overall Star Rating: 24 contracts had a most common overall Star Rating across the 1,000 simulations that was a half star lower than their 2025 overall Star Rating, and 21 contracts had a most common overall Star Rating across the 1,000 simulations that was a half star higher than their 2025 overall Star Rating. *Id.*

The simulation supported the notion that CMS's calculation of Elevance's contracts was sufficiently accurate and precise. Elevance's contract H3655 received 3.5 stars for its 2025 overall Star Rating. In CMS's simulations, for Elevance's H3655 contract, the contract with an overall score of 3.749565, 85% (or 850 simulations) resulted in an overall Star Rating of 3.5 stars and 15% (or 150 simulations) resulted in an overall Star Rating of 4.0 stars. Defs.' App. 3, ¶ 5

(Goldstein Decl. ¶ 5). Therefore, H3655's most common overall Star Rating across the simulations is 3.5 stars. *Id.*

In sum, the evidence does not support Elevance's arguments that non-CAHPS measures are imprecise.

c. CMS Considered Imprecision in Its CAHPS Methodology.

Again, this Court need not reach Elevance's arguments concerning precision because they are premised on basic misconceptions about how CMS calculates Star Ratings. However, CMS's Star Ratings calculation methodology sufficiently accounts for nonresponse bias, sampling error, and measurement error.

i. CMS's Star Rating Account for Nonresponse Bias.

Elevance contends that CMS's alleged failure "to investigate, much less account for, nonresponse bias in connection with the CAHPS surveys" was arbitrary and capricious. Pls.' Br. 30. Not so. CMS considered nonresponse bias when issuing the 2018 final rule. CMS summarized comments received: "A commenter stated that the CAHPS survey is long, and a couple commenters expressed concern about low response rates." 83 Fed. Reg. at 16,555. CMS summarized actions it took to improve response rates: "CMS shortened the [Medicare Advantage] CAHPS survey in 2017 by removing questions and measures not used in Star Ratings, and we also improved phone contact information." *Id.* It continued, "As a result of CMS's continuing efforts to improve response rates, overall [Medicare Advantage] and [Prescription Drug Plan] CAHPS response rates increased from 2016 to 2017, despite national trends of declining response rates for most other surveys." These national trend results are consistent with the declining national trends described in 2020 report conducted by the Nonresponse Bias Subcommittee of the U.S. Federal Committee on Statistical Methodology upon which Elevance's expert, Dr. Lavrakas, relies. *See*

Pls.' App. 65 (relying on Peter Miller et al., *A Systematic Review of Nonresponse Bias Studies in Federally Sponsored Surveys* 2 (Federal Committee on Statistical Methodology 2020) ("Over the last two decades, survey response rates have been steadily falling, with more accelerated declines reported in recent years.")). Dr. Lavrakas's chief criticism of CAHPS data is that the survey response rate is below 80%. Pls.' App. 64. But CMS increased response rates by shortening the survey and improved phone contact information. 83 Fed. Reg. at 16,555. Dr. Lavrakas does not account for the response rate improvement.

In its 2018 final rule, CMS continued, "Moreover, research specific to patient experience, CAHPS, and [Medicare Advantage] and [Prescription Drug Plans] CAHPS surveys finds no evidence nonresponse bias affects comparison of case-mix adjusted scores between contracts or other similar reporting units," citing several articles. 83 Fed. Reg. at 16,555. Dr. Lavrakas speculates, without citing any evidence, that case-mix adjusting "may be *reducing the accuracy* of CAHPS data." Pls.' App. 67. His unsupported speculation is at odds with the evidence CMS cites in its 2018 rulemaking. As described above, case-mix adjustment recognizes that patients in some subgroups have a propensity to respond to care experience questions in a way that is significantly different from other groups, even after controlling for other factors. 83 Fed. Reg. at 16,555. In analyses to which CMS cited in its 2018 rulemaking, researchers found that "[a]fter the [case]-mix adjustment process is completed, the survey results no longer have nonresponse bias, which has been shown in some [non-Medicare Advantage and Prescription Drug Plans] CAHPS surveys." Cynthia Rae Bland et al., *Challenges Facing CAHPS Surveys and Opportunities for Modernization* 9 (RTI Press 2022). The researchers concluded that case-mix adjustments "should be used in conjunction with survey methods that increase response from underrepresented groups"

and that "[i]ncreasing response from lagging subgroups can improve overall response rates and data quality." *Id.*

In sum, the record shows that CMS adequately considered and accounted for nonresponse bias in CAHPS results. The law requires that "the agency must examine the relevant data and articulate a satisfactory explanation for its action." *State Farm*, 463 U.S. at 43. That's what CMS did here: it examined the problem of nonresponses bias identified by commenters in 2018. It responded by shortening the CAHPS survey and improving phone contact information. 83 Fed. Reg. at 16,555. Next, CMS analyzed the problem of nonresponse bias by pointing to the efficacy of case-mix adjustments. CMS has provided a "rational connection between the facts found and the choice made." *State Farm*, 463 U.S. at 43. This Court should not substitute its own judgment on the issue of nonresponse bias in CAHPS survey with that of CMS's.

ii. CMS's Star Ratings Account for Sampling and Measurement Error.

Elevance asserts that CMS has not "accounted for confidence intervals when using CAHPS data to calculate the Final Summary Score and Overall Star Rating." Pls.' Br. 32. But this assertion is hard to follow. That's because CMS calculated its final summary score and overall ratings using methods that are *more* accurate and unbiased than confidence internals. So Elevance misses the mark by contending that CMS failed "to disclose any information about its sampling design" such that Elevance is prevented "from determining an appropriate confidence interval associated with the initial sample used for the CAHPS survey." *Id.* at 32. It is unclear what the basis of Elevance's arbitrary and capricious claim is—Dr. Lavrakas asserts only that CMS's lack of transparency about its sampling design is "extremely peculiar." *Id*. CMS, however, is transparent. CMS has made its CAHPS sampling design public.

Elevance does correctly assert that "all sample surveys have sampling error." *Id.* Sometimes surveys are inaccurate because the survey sample is unrepresentative of the general population. CMS's Star Ratings calculation methodology, however, accounts for possible sampling error. CMS accounts for the CAHPS sample design appropriately in scoring, and its scoring and star-assignment procedures appropriately and accurately measure contract performance. The most accurate, unbiased measure of a contract's performance is the mean, not confidence intervals, which are only a descriptive tool. Defs.' App. 3-4 ¶ 6 (Goldstein Decl. ¶ 6). CMS uses weighted, case-mix adjusted means, which further improve the accuracy and comparability of scores. *Id.* CMS does not use confidence intervals. Consider a 95% confidence interval. This is a range of values within which there is a 95% probability that the true population parameter—the mean—lies. Instead of relying on confidence intervals, CMS relies on the actual mean, the most accurate, unbiased measure of a contract's performance.

Moreover, CMS's CAHPS scoring methodology accounts for uncertainty in a principled manner. The test of statistical significance versus the overall mean is part of the star assignment methodology. Consider the Star Ratings calculation methodology set out in 42 C.F.R. § 422.166. The standard error is considered when the measure score is below the 15th percentile (in base group 1), significantly below average, and has low reliability. In this case, 1 star will be assigned if and only if the measure score is at least 1 standard error below the unrounded cut point between base groups 1 and 2. Similarly, when the measure score is at or above the 80th percentile (in base group 5), significantly above average, and has low reliability, 5 stars would be assigned if and only if the measure score is at least 1 standard error above the unrounded cut point between base groups 4 and 5. § 422.166(a)(3); A.R. 151-55. CMS's Star Ratings calculation methodology sufficiently accounts for sampling error and uncertainty by relying on means, weighting and case-mix

adjusting those means, and accounting for uncertainty in its calculation methodology by taking into consideration statistical significance and direction of the difference of the contract mean from the national mean, statistical reliability of the contract score, and the standard error of the mean contract score for a given measure.

Elevance contends that CMS "fails to disclose any information about its sampling design." Pls.' Br. 32. This is incorrect. CMS extensively describes "the process that will be used by CMS for selecting the sample for the 2025 [Medicare Advantage] & Prescription Drug Plan] CAHPS Survey." CMS, *Medicare Advantage & Prescription Drug Plan (MA & PDP CAHPS) CAHPS Survey: Quality Assurance Protocols & Technical Specifications*, Version 15.0, 19-22 (Nov. 2024) https://ma-pdpcahps.org/globalassets/ma-pdp/quality-assurance/2025/ma--pdp-cahps-qapts-v15.0_updated.pdf [https://perma.cc/3JBD-UYVZ]. CMS describes CAHPS sample selection and eligibility criteria and sample preparation. *Id.* CMS is sufficiently transparent. The basis premises of Elevance's arbitrary and capricious claim are simply incorrect.

In its final salvo, Elevance makes clear again that all of the alleged errors in CMS's CAHPS methodology create precision problems. Elevance isn't claiming that CMS's methodology is insufficiently precise as a general matter. This problem only exists, according to Elevance, because CMS calculates overall ratings scores and corresponding overall Star Ratings "to the extreme precision of six decimals." Pls.' Br. 33. As explained already, this is simply false. Those arguments are not compelling because they are premised on a misunderstanding about how CMS calculates and rounds as addressed above. *See supra* Part IV.C.

Elevance cites to a case—the only case it cites to in this section of its brief—in which the Eighth Circuit found a United States Forest Service (USFS) survey statistically deficient. *See* Pls.' Br. 33. (citing *Friends of the Boundary Waters Wilderness v. Bosworth*, 437 F.3d 815 (8th Cir.

2006)). As Elevance states, the Eighth Circuit affirmed the district court's decision that an agency acted arbitrarily and capriciously in relying on survey results without properly considering "[s]ample size, potential for bias, interviewing techniques" and other factors. *Friends of the Boundary Waters*, 437 F.3d at 826-27. But the survey data that the USFS relied on was completed by thirteen respondents who were similar in demographic characteristics and lived in the same geographic area, were aware of the survey's purpose, included only two questions in total (the second of which was only answered by five out of the thirteen survey respondents), and the content of responses varied greatly by individual. *Id.* at 824-25. This is much different than CMS's CAHPS survey process.

For the CAHPS survey, CMS publishes its Quality Assurance Protocols & Technical Specifications every year—which exceeds 800 pages. CMS, *Medicare Advantage & Prescription Drug Plan (MA & PDP CAHPS) CAHPS Survey: Quality Assurance Protocols & Technical Specifications*, Version 15.0 (Nov. 2024) https://ma-pdpcahps.org/globalassets/ma-pdp/quality-assurance/2025/ma--pdp-cahps-qapts-v15.0_updated.pdf [https://perma.cc/3JBD-UYVZ]. Next year will be CMS's sixteenth such publication. This document explains how the survey is administered, by whom, to whom, how survey data are reported, the number of questions asked, the questions asked, the roles and responsibilities of all parties, vendor training requirements, sampling (see above), technical support, data collection protocol, and more. CMS even has a plan for how to factor CAHPS survey data into Star Ratings when there is a major disaster, and another plan for when contracts are doubly affected by a major disaster. A.R. 15. The CAHPS survey "[s]ample size, potential for bias, and interviewing techniques" far exceed those features pertaining to the survey the Eighth Circuit found deficient in *Friends of the Boundary Waters*. 437 F.3d at 826-27. Simply put, the USFS survey and CAHPS survey are not alike.

d. CMS's Use of the Same Seed Year-Over-Year Does Not Make Its Mean Resampling Less Random or Non-Random.

CMS regulations expressly require mean resampling to achieve a random separation of the applicable contracts into groups. *See* App. 216; 42 C.F.R. § 422.162(a) ("Mean resampling refers to a technique where measure-specific scores for the current year's Star Ratings are randomly separated into 10 equal-sized groups."). Again, mean resampling is used to determine the cut points—the demarcations between 1-5 whole star ratings—for all non-CAHPS measures. A.R. 17. The contracts are divided randomly into ten groups. A.R. 17. The clustering algorithm is run ten times, leaving out one of the ten groups of randomly assigned contracts, each time. At the end, each measure has ten sets of calculated cut points. A.R. 17. CMS takes the average of the ten cut point sets, the goal being to improve the stability of cut points and reduce year over year variation and sensitivity in cut points to individual contract scores.

The seed—the initial sequence of numbers, here 8-6-7-5-3-0-9—randomly determines which of the ten groups each contract is placed into. It is simply the process and does not dictate the results. The seed is the starting point for mean resampling. The best practice is to use a number that is well known, like a number from pop culture, hence, the reference to the Tommy Tutone song, "Jenny." A memorable seed is necessary to replicate results of a randomization process. Pls.' App. 54. Dr. Diver correctly explains that "[s]pecifying the seed allows a researcher to exactly replicate which specific plans are grouped together, *all else equal*." *Id.* (emphasis added). Dr. Driver continues, correctly, "if the list of considered plans remained the same from one year to the next, and the same seed is used each year, the plan grouping in CMS's clustering methodology would be effectively pre-determined year over year." *Id.* From these true premises, Elevance makes a leap. It says, "repeated use of the 'Jenny seed' results in *non-random* and *potentially predictable* outcomes from one year to the next, defying the regulatory requirement

that mean resampling be random." Pls.' Br. 35. But Elevance fails to appreciate the caveats in Dr. Diver's analysis. All else is not equal. The list of considered contracts *does not* remain the same from year to year. CMS uses the same seed every year. But the 8-6-7-5-3-0-9 seed has different effects in different years because it is applied to different data from different contracts. Defs.' App. 2 ¶ 3 (Goldstein Decl. ¶ 3). It is as simple as that.

Recall that CMS ran 1,000 simulations using 1,000 randomly generated seeds for mean resampling and recalculated the overall Star Ratings across all contracts for each of the 1,000 simulations. *See supra* Part IV.C.3(b). CMS's analyses demonstrate strong stability in the Star Ratings methodology with respect to the choice of random seed because, on average across MA-PD contracts, 92.4% of simulation seeds resulted in the same 2025 overall Star Ratings as the 2025 overall Star Ratings calculated using the original 8-6-7-5-3-0-9 seed. *See supra* Part IV.C.3(b). This means that changing the seed used in the Star Ratings methodology would not change the 2025 overall Star Ratings for 92.4% of contracts on average across simulation seeds. CMS's analysis supports the conclusion that choosing different seeds is not necessary to ensure the randomness that mean resampling at § 422.162(a) requires. CMS did not act contrary to law in using the 8-6-7-5-3-0-9 seed year after year.

V. Conclusion

For the reasons explained above, CMS's actions were not arbitrary and capricious in violation of the APA. Therefore, the Court should grant Defendants' cross-motion for summary judgment and deny Plaintiffs' motion for summary judgment.

Respectfully submitted,

CHAD E. MEACHAM
ACTING UNITED STATES ATTORNEY

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney
Texas Bar No. 24007419
Burnett Plaza, Suite 1700
801 Cherry Street, Unit #4
Fort Worth, Texas 76102-6882
Telephone: 817.252.5230
Fax: 817.252.5458
Andrea.Hyatt@usdoj.gov

Attorneys for Defendants

Certificate of Service

On March 21, 2025, I electronically submitted the foregoing document with the clerk of court for the U.S. District Court, Northern District of Texas, using the electronic case filing system of the court. I hereby certify that I have served all parties electronically or by another manner authorized by Federal Rule of Civil Procedure 5(b)(2).

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS
FORT WORTH DIVISION

ELEVANCE HEALTH, INC.; COMMUNITY INSURANCE COMPANY; FREEDOM HEALTH, INC.; GROUP RETIREE HEALTH SOLUTIONS, INC.; WELLPOINT INSURANCE COMPANY; and WELLPOINT TEXAS, INC., Plaintiffs, v. ROBERT F. KENNEDY, JR., in his official capacity as Secretary of Health and Human Services; and STEPHANIE CARLTON, in her official capacity as Acting Administrator of the Centers for Medicare & Medicaid Services, Defendants.[1]	Civil Action No. 4:24-cv-01064-P

DEFENDANTS' APPENDIX
IN SUPPORT OF THEIR SUMMARY-JUDGMENT BRIEFING

[1] Xavier Becerra has been substituted with Robert F. Kennedy, Jr. as Secretary of the United States Department of Health and Human Services, and Chiquita Brooks-LaSure has been substituted with Stephanie Carlton as Acting Administrator of the Centers for Medicare & Medicaid Services, pursuant to Federal Rule of Civil Procedure 25(d).

Respectfully submitted,

CHAD E. MEACHAM
ACTING UNITED STATES ATTORNEY

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney
Texas Bar No. 24007419
Burnett Plaza, Suite 1700
801 Cherry Street, Unit #4
Fort Worth, Texas 76102-6882
Telephone: 817.252.5230
Fax: 817.252.5458
Andrea.Hyatt@usdoj.gov

Attorneys for Defendants

Certificate of Service

On March 21, 2025, I electronically submitted the foregoing document with the clerk of court for the U.S. District Court, Northern District of Texas, using the electronic case filing system of the court. I hereby certify that I have served all parties electronically or by another manner authorized by Federal Rule of Civil Procedure 5(b)(2).

/s/ Andrea Hyatt
Andrea Hyatt
Assistant United States Attorney

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS
FORT WORTH DIVISION

ELEVANCE HEALTH, INC.; COMMUNITY INSURANCE COMPANY; FREEDOM HEALTH, INC.; GROUP RETIREE HEALTH SOLUTIONS, INC.; WELLPOINT INSURANCE COMPANY; and WELLPOINT TEXAS, INC., Plaintiffs, v. ROBERT F. KENNEDY, JR., in his official capacity as Secretary of Health and Human Services; and STEPHANIE CARLTON, in her official capacity as Acting Administrator of the Centers for Medicare & Medicaid Services, Defendants.[1]	Civil Action No. 4:24-cv-01064-P

DECLARATION OF ELIZABETH GOLDSTEIN

I, Elizabeth Goldstein, declare pursuant to 28 U.S.C. § 1746 as follows:

1. I am the Director, Division of Consumer Assessment and Plan Performance, Medicare Drug Benefit and C & D Data Group, Center for Medicare, Centers for Medicare & Medicaid Services ("CMS"), United States Department of Health and Human Services. I have held this position since October 2000. In my role, I oversee and administer the calculation of Star Ratings for Medicare Advantage and Medicare Part D Plans. The statements made in this

[1] Xavier Becerra has been substituted with Robert F. Kennedy, Jr. as Secretary of the United States Department of Health and Human Services, and Chiquita Brooks-LaSure has been substituted with Stephanie Carlton as Acting Administrator of the Centers for Medicare & Medicaid Services, pursuant to Federal Rule of Civil Procedure 25(d).

declaration are based on my personal knowledge, information contained in agency files, and information furnished to me in the course of my official duties.

2. The rounding rules that translate overall Star Rating score ranges to the final rounded overall Star Rating are displayed in CMS's Medicare 2025 Part C & D Star Ratings Technical Notes. See A.R. 31. Contracts' overall scores are also displayed in the Health Plan Management System for contracts to view. When CMS calculates overall scores, those scores are calculated out to at least six decimal places such that the overall score can be rounded once to the half star. In rounding to the half star, the second, hundredth decimal place determines the rounding. In order to round once to the half star, it is necessary to carry enough decimal places throughout each step of the overall score calculation such that the second, hundredth decimal place is not rounded. CMS calculates each step out to the sixth decimal place because nothing that happens at the sixth decimal place ever impacts the second, hundredth decimal place. Practically, this is no different than if the calculations were completed with infinite decimal places but is computationally more efficient. When those overall scores are displayed in CMS's technical guidance or for contracts to view, CMS displays only the six decimal places.

3. When calculating non-CAHPS measure Star Ratings, CMS's regulations require that mean resampling be used to achieve a random separation of applicable contracts into ten equal-sized groups. Those ten groups are random. To create those random groups, CMS uses a seed. The seed is a sequence of numbers that serves as the starting point for mean resampling. Every year, CMS uses the same seed: 8-6-7-5-3-0-9. While the seed is the same, because the set of contracts subject to mean resampling differs from year to year, the ten groups generated by mean resampling are random and non-predictable.

4. To test its calculation of non-CAHPS measures, CMS's Star Ratings contractor, RAND, ran 1,000 simulations using 1,000 randomly generated seeds for mean resampling for non-CAHPS measure cut points and recalculated the overall Star Ratings across all contracts for each of the 1,000 simulations. The analyses demonstrate strong stability in the Star Ratings mean resampling methodology with respect to the choice of random seed. On average across contracts, 92.4% of simulations/seeds resulted in the same 2025 overall Star Ratings as the 2025 overall Star Ratings calculated using the original seed. This result means that changing the seed used in the Star Ratings methodology would not change the 2025 overall Star Ratings for 92.4% of contracts on average across simulation seeds. The most common overall Star Rating assigned to each contract across the 1,000 simulations was the same as the 2025 overall Star Rating for 92% of contracts. Only 45 contracts (8%) had a different most common overall Star Rating across the 1,000 simulations compared to their 2025 overall Star Rating. Twenty-four contracts had a most common overall Star Rating across the 1,000 simulations that was a half star lower than their 2025 overall Star Rating. Twenty-one contracts had a most common overall Star Rating across the 1,000 simulations that was a half star higher than their 2025 overall Star Rating.

5. Elevance's contract H3655 received 3.5 stars for its 2025 overall Star Rating. In CMS's simulations for Elevance's H3655 contract, the contract with an overall score of 3.749565, 85% (or 850 simulations) resulted in an overall Star Rating of 3.5 stars and 15% (or 150 simulations) resulted in an overall Star Rating of 4.0 stars. Therefore, H3655's most common overall Star Rating across the simulations is 3.5 stars.

6. All surveys have sampling error. The most accurate, unbiased measure of a contract's performance is the mean. CMS uses weighted, case-mix adjusted means for most CAHPS measures which further improve the accuracy and comparability of scores compared to

simple means. Confidence intervals are a descriptive tool, but the upper confidence limit is a positively biased estimate of a contract's performance (and the lower confidence limit is a negatively biased estimate). On average, the mean will neither overestimate nor underestimate a contract's true performance. Thus, CMS's approach for measuring performance through the CAHPS survey results in accurate, unbiased assignment of contract scores to base groups. An approach using the upper confidence limit or the lower confidence limit would be biased and less accurate.

In accordance with 28 U.S.C. § 1746, I hereby declare under penalty of perjury that the foregoing is true and correct to the best of my knowledge.

Executed this 21st day of March, 2025, in Baltimore, Maryland.

Elizabeth H. Goldstein -S

Digitally signed by Elizabeth H. Goldstein -S
Date: 2025.03.21 07:32:15 -04'00'

Elizabeth Goldstein