**UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS**

ELEVANCE HEALTH, INC., et al.,	
Plaintiffs,	
v.	Case No. 4:24-cv-01064
ROBERT F. KENNEDY, JR., in his official capacity as Secretary of Health and Human Services, U.S. Department of Health and Human Services,	**ORAL HEARING REQUESTED**
and	
MEHMET OZ, in his official capacity as Administrator, Centers for Medicare and Medicaid Services,	
Defendants.	

**PLAINTIFFS' RESPONSE TO DEFENDANTS' CROSS-MOTION
FOR SUMMARY JUDGMENT AND REPLY IN SUPPORT
OF PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT**

Plaintiffs, Elevance Health, Inc. f/k/a Anthem Inc. ("Elevance"), along with its affiliated entities Community Insurance Company; Freedom Health, Inc.; Group Retiree Health Solutions; Wellpoint Insurance Company; Wellpoint Texas, Inc. (the "Health Plan Plaintiffs," and collectively with Elevance, the "Plaintiffs"), file their Response in Opposition to Defendants' Cross-Motion for Summary Judgment and Reply in support of Plaintiffs' Motion For Summary Judgment. ECF No. 38; ECF No. 31.

This Response and Reply is accompanied by a consolidated brief that contains the contents required by Local Civil Rule 56.4(a).

Dated: April 11, 2025   Respectfully submitted,

**ELEVANCE HEALTH, INC. and the
HEALTH PLAN PLAINTIFFS**

By: */s/ Lesley C. Reynolds*

Lesley C. Reynolds (N.D. Tx. Bar No. 487580DC)
Lara E. Parkin (*pro hac vice*) (D.C. Bar No. 475974)
David A. Bender (*pro hac vice*) (D.C. Bar No. 1030503)
REED SMITH LLP
1301 K Street, N.W.
Suite 1000 – East Tower
Washington, D.C. 20005
Ph: (202) 414-9200
Fax: (202) 414-9299
lreynolds@reedsmith.com
lparkin@reedsmith.com
dbender@reedsmith.com

Martin J. Bishop (*pro hac vice*) (Tx. Bar No. 24086915)
Steven D. Hamilton (*pro hac vice*) (Ill. Bar No. 6289663)
REED SMITH LLP
10 South Wacker Drive
40th Floor
Chicago, IL 60606
Ph: (312) 207-1000
Fax: (312) 207-6400
mbishop@reedsmith.com
shamilton@reedsmith.com

Scott T. Williams (Tx. Bar No. 00791937)
REED SMITH LLP
2850 N. Hardwood Street
Suite 1500
Dallas, TX 75201
Ph: (469) 680-2200
Fax: (469) 680-4299
scott.williams@reedsmith.com

Meredith W. Knudsen (Tx. Bar No. 24088617)
Klayton S. Hiland (Tx. Bar No. 24116516)

2

KELLY HART & HALLMAN LLP
201 Main Street, Suite 2500
Fort Worth, Texas 76102
Ph: (817) 332-2500
Fax: (817) 878-9280
meredith.knudsen@kellyhart.com
klayton.hiland@kellyhart.com

Counsel for Plaintiffs

CERTIFICATE OF SERVICE

I hereby certify that on April 11, 2025, I electronically filed the foregoing document with

the Clerk of the Court using the CM/ECF system, which will send notification of this filing to the

attorneys of record and all registered participants.

/s/ Meredith W. Knudsen
Meredith W. Knudsen

3

**UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS**

ELEVANCE HEALTH, INC., et al., Plaintiffs, v. **ROBERT F. KENNEDY, JR.**, in his official capacity as Secretary of Health and Human Services, U.S. Department of Health and Human Services, and **MEHMET OZ**, in his official capacity as Administrator, Centers for Medicare and Medicaid Services, Defendants.	Case No. 4:24-cv-01064 **ORAL HEARING REQUESTED**

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

TABLE OF CONTENTS

TABLE OF AUTHORITIES

INTRODUCTION

Medicare Advantage Star Ratings determine whether Medicare Advantage organizations ("MAOs") like Plaintiffs and scores of other plans receive billions of dollars in quality bonus and other payments. Those payments must be used to directly benefit Medicare beneficiaries through improved benefits and reduced premiums. Notwithstanding that the stakes involve billions of dollars used to improve beneficiary experiences, Defendants have implemented sub-regulatory guidance and calculated Star Ratings in ways that are divorced from the plain regulatory text and fail to consider significant issues. In Plaintiffs' Opening Brief, Plaintiffs clearly explained the ways that Defendants' calculation of Plaintiffs' Star Ratings were contrary to the plain regulatory text governing the calculation of Star Ratings and arbitrary and capricious. In Defendants' Consolidated Brief in Support of their Response to Plaintiffs' Motion for Summary Judgment and Cross-Motion for Summary Judgment (hereinafter "Defendants' Brief"), Defendants engage in "doublespeak" arguments that are not supported by the Administrative Record and amount to nothing more than *post hoc* rationalizations of "trust us, we did it right."

In summary, Defendants' conduct is contrary to the applicable regulatory text and arbitrary and capricious for the following reasons:

First, **Defendants applied a case-mix adjustment contrary to the regulatory requirements for calculating Star Ratings**. Defendants violated the plain text of 42 C.F.R. § 422.166(a)(3) by applying a "case-mix adjustment" to the CAHPS data used to calculate Star Ratings, causing several individual measures in Plaintiffs' contracts to be improperly decreased and two contracts, H3655 and H6078, to receive incorrect overall Star Ratings. The plain text of this regulation does not permit this case-mix adjustment to the CAHPS-based measure scores. In their response, Defendants do not dispute that the regulatory text does not contemplate a case-mix adjustment to CAHPS-based measure scores, but then argue they have discretion to do so in their

1

guidance. However, Defendants must act in accordance with their own regulations and have failed to do so.

Second, **Defendants used a national weighted average contrary to the regulatory requirements for calculating Star Ratings.** The applicable regulations require Defendants to calculate measure-specific Star Ratings by comparing Plaintiffs' contract to the "national average" contract score. *See* 42 C.F.R. § 422.166(a)(3)(i)–(v). Instead of using the "national average" as the regulation requires, Defendants concede they used a "national *weighted* average" and Defendants do not dispute that these are fundamentally different concepts. Despite their concession that they are using a weighted average contrary to the plain regulatory text, Defendants attempt to justify their position by arguing it "makes more sense." But neither Defendants' sub-regulatory guidance nor *post hoc* rationalizations can override the plain regulatory text.

Third, **Defendants acted contrary to the regulatory requirements and arbitrarily and capriciously in calculating and rounding H3655 to six decimals**. Defendants calculated the Final Summary Score for Plaintiffs' contract H3655 to be 3.749565 and then applied sub-regulatory guidance to conclude that the Overall Star Rating was 3.5 Stars. However, Defendants violated their own regulation by not applying traditional rounding rules, as defined. In their response, Defendants engage in *post hoc* rationalizations without any support in the Administrative Record in an attempt to justify their failure to follow the plain regulatory text, but those justifications contradict what Defendants actually do, as evidenced by their own sub-regulatory guidance. Furthermore, Defendants acted arbitrarily and capriciously by calculating Star Ratings based upon approximately 40 different measures that use data fraught with imprecision and error yet purporting to calculate the Star Ratings to the millionth-place decimal. In the process, Defendants completely ignore this imprecision and error. Defendants do not meaningfully respond

2

to this, other than pointing to a fleeting sentence in a 2018 rulemaking that does not demonstrate they considered these important issues in calculating Plaintiffs' Star Ratings.

Fourth, **year after year, Defendants violated regulatory requirements by using the same number, the refrain from the popular song, "867-5309/Jenny" by Tommy Tutone as the seed for its mean resampling process (hereinafter the "'Jenny' seed")**. Applicable regulations require, however, that mean resampling be random. The "seed" is a number that controls the starting point of the random sequence of numbers needed to perform mean resampling. Defendants have conceded that by using the same seed, or starting point, year after year they necessarily rely on outside factors to ensure the randomness required by the regulation. Randomness is not ensured by their methodology alone. Moreover, Defendants offer no authoritative support or any evidence of their reasoning in the Administrative Record for their use of a seed that may violate the requirement of randomness, and are left with only *post hoc* rationalizations to defend their approach. Thus, Defendants' use of the Jenny seed year after year is contrary to law, arbitrary, and capricious.

At bottom, Defendants have arbitrarily and capriciously devised a complex Star Ratings process that is almost entirely borne out in sub-regulatory guidance that ignores the plain regulatory text. Accordingly, Plaintiffs respectfully submit that their motion for summary judgment should be granted and Defendants' cross-motion for summary judgment should be denied.

<div align="center">

ARGUMENT

</div>

I.   LEGAL STANDARD OF REVIEW

Defendants contend that the Court's review of agency action is rooted in deference to the agency. *See* ECF No. 38 ("Defs.' Br.") at 16. However, the arbitrary-and-capricious standard is "not toothless" and has "serious bite." *See Texas v. E.P.A.*, 91 F.4th 280, 291 (5th Cir. 2024)

<div align="center">

3

</div>

(hereafter, "*Texas*") (citations omitted); *Louisiana v. U.S. Dep't of Energy*, 90 F.4th 461, 471 (5th Cir. 2024). Put simply, the Administrative Procedures Act ("APA") requires a court to hold unlawful and set aside arbitrary and capricious agency action "premised on reasoning that fails to account for 'relevant factors' or evinces a 'clear error of judgment.'" *See State v. Biden*, 10 F.4th 538, 552 (5th Cir. 2021).

Similarly, the APA requires courts to set aside agency actions that are contrary to law. *See id.*; *see also* 5 U.S.C. § 706(2). Although Defendants wholly omit the standard for assessing whether agency action is contrary to law, the Fifth Circuit is clear that "[i]n taking final action, an agency must comply with its own regulations." *See Texas*, 91 F.4th at 291 (citing *Gulf States Mfrs., Inc. v. NLRB*, 579 F.2d 1298, 1308 (5th Cir. 1978) ("[T]he failure of an agency to follow its regulations renders its decision invalid.")). Further, "[i]t is elementary that an agency must adhere to its own rules and regulations" and "ad hoc departures… cannot be sanctioned." *Texas v. E.P.A.*, 829 F.3d 405, 430 (5th Cir. 2016) (quoting *Reuters Ltd. v. FCC*, 781 F.2d 946, 950 (D.C. Cir. 1986)).

Importantly, in assessing whether agency action is contrary to applicable regulations (and therefore also arbitrary and capricious), "[w]hen a regulation is unambiguous courts ***owe no deference*** to the agency's interpretation of it and simply apply the regulation's plain meaning." *See Dominion Ambulance, LLC v. Azar*, 968 F.3d 429, 434 (5th Cir. 2020) (emphasis added); *see also ExxonMobil Pipeline Co. v. U.S. Dep't of Transp.*, 867 F.3d 564, 573 (5th Cir. 2017) ("If the regulation is unambiguous, we will not defer to the agency's interpretation."); *Cargill v. Garland*, 57 F.4th 447, 464 (5th Cir. 2023) (holding that statute was unambiguous therefore no deference to agency applied).

Finally, the agency must support its actions through the Administrative Record.[1] An agency

cannot invent *post hoc* justifications for its decision in court and outside of the administrative

record; the grounds upon which an administrative action must be judged are only "those upon

which the [administrative] record discloses its action was based." *See Louisiana*, 90 F.4th at 469

(quoting *SEC v. Chenery Corp.*, 318 U.S. 80, 87 (1943)). Where the agency fails to do so, courts

must disregard any *post hoc* rationalizations for agency action. *See Luminant Generation Co., LLC

v. E.P.A.*, 675 F.3d 917, 925 (5th Cir. 2012); *Nat'l Ass'n of Mfrs. v. SEC*, 105 F.4th 802, 814 (5th

Cir. 2024) ("[I]n reviewing an agency's action, we may consider only the reasoning articulated by

the agency itself; we cannot consider *post hoc* rationalizations.") (internal quotations omitted).

II. DEFENDANTS ACTED CONTRARY TO THE LAW IN THEIR CALCULATION OF CAHPS MEASURE SCORES BY APPLYING A CASE-MIX ADJUSTMENT.

A. Defendants Concede that the Plain Text of the Applicable Regulation (42 C.F.R. § 422.166(a)(3)) Does Not Contemplate Applying a Case-Mix Adjustment to Calculate CAHPS-Based Measure Scores.

Plaintiffs and Defendants agree that 42 C.F.R. § 422.166(a)(3) dictates how CAHPS-based

measures scores are calculated for any contract. That regulation expressly provides that, to

calculate measure scores for CAHPS-based individual measures, Defendants use the relative

distribution and significance testing methodology to determine whether any CAHPS-based

measure receives a score of 1, 2, 3, 4, or 5 Stars. 42 C.F.R. § 422.166(a)(3). In conducting that

---

[1] Defendants filed the Administrative Record, including A.R. 1-844, on December 27, 2024. *See* ECF No. 23. Defendants subsequently produced non-publicly available data to Plaintiffs as part of the Administrative Record, including A.R. 845-1680, pursuant to the Protective Order entered by the Court on February 4, 2025. *See* ECF No. 28. Plaintiffs identified Defendants' supplemental production with a placeholder in Plaintiffs' Appendix in Support of Plaintiffs' Motion for Summary Judgment, filed on February 14, 2025. *See* ECF No. 33 (designating "App. 224" for "De-identified contract-level data for cut point calculations"). To date, Defendants have not filed with the Court the supplemental Administrative Record to include A.R. 845-1680 as produced to Plaintiffs under the Protective Order. Defendants' entire production to Plaintiffs, as supplemented, constitutes the final Administrative Record in this case.

calculation, the regulation sets forth a specific formula to be followed for each Star Rating. As an

example, the regulation expressly lays out the following formula for any CAHPS-based measure

score to receive a rating of 5 Stars:

> (v)   A [measure score for a] contract is assigned 5 stars if both of the following criteria in [subparagraphs (A) and (B)] of this section are met plus at least one of the criteria in [subparagraphs (C) or (D)] of this section is met:
>
>> (A)   Its average CAHPS measure score is at or above the 80th percentile; and
>>
>> (B)   Its average CAHPS measure score is statistically significantly higher than the national average CAHPS measure score;
>>
>> (C)   The reliability is not low; or
>>
>> (D)   Its average CAHPS measure score is more than one standard error above the 80th percentile.

42 C.F.R. § 422.166(a)(v). Essentially, this same formula is followed with minor variation to assign

1, 2, 3, or 4 Stars for any CAHPS-based measure. Nowhere does the plain regulatory text allow

Defendants to use "case-mix adjustments"—a defined term in the regulation—when determining

measure scores for CAHPS-based measures.

Indeed, as Defendants recognize, the applicable regulations define the term "case-mix

adjustment" as follows:

> 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 responses.

42 C.F.R. § 422.162(a) (emphasis added). However, the definitional section is not the controlling

regulatory provision for calculating Star Scores, as that instruction is found in 42 C.F.R. §

422.166(a)(3). The plain language in 42 C.F.R. § 422.166(a)(3) does not contemplate that the

measure score is case mix adjusted prior to when the Star Rating is assigned for the measure—

rather, the plain language merely states that a measure score is assigned a particular Star Rating if the "average CAHPS measure score" meets certain requirements and, if applicable, the reliability is not low. *See* 42 C.F.R. § 422.166(a)(3).

Although the defined term "case-mix adjustment" **is not used** in 42 C.F.R. § 422.166(a)(3), that term **is used** in connection with two elements of a contract's overall Star Rating—*i.e.*, in 42 C.F.R. § 422.166(f)(2) relating to the Categorical Adjustment Index and 42 C.F.R. § 422.166(f)(3) relating to Health Equity Index. The Categorical Adjustment Index is an adjustment made to the contract's overall Star Rating to account for the number of low income and disabled beneficiaries. *See* A.R. 23–24, Plaintiffs' Supplemental Appendix ("Suppl. App.") 312–13, 2025 Technical Notes.[2] In determining the appropriate categorical adjustment to make, 42 C.F.R. § 422.166(f)(2)(ii)(A) contemplates that a case mix adjustment may be performed to measures to calculate the Categorical Adjustment Index. Likewise, the Health Equity Index (which will go into effect starting with 2027 Star Ratings) contemplates that contract Star Ratings will be adjusted for a "health equity index." *See* 42 C.F.R. § 422.166(f)(3). This regulation also contemplates that the Health Equity Index will be based upon measures that are case mix adjusted. Thus, Defendants know how to apply that defined term in the regulations—as they did for the Categorical Adjustment Index and the Health Equity Index—and they simply did not do so with respect to the calculation of CAHPS-based measure scores in 42 C.F.R. § 422.166(a)(3).

Congress, too, knows how to authorize case-mix adjustments, as it has done so in other areas of the Medicare statute. *See, e.g.*, PROSPECTIVE PAYMENT FOR HOME HEALTH SERVICES, 42

---

[2] "App. ___ " refers to a page in Plaintiffs' Appendix in Support of Plaintiffs' Motion for Summary Judgment, filed on February 14, 2025. *See* ECF No. 33. "Suppl. App. ___ " refers to a page in Plaintiffs' Supplemental Appendix in support of Plaintiffs' Response to Defendants' Cross-Motion for Summary Judgment and Reply in support of Plaintiffs' Motion for Summary Judgment, filed contemporaneously with Plaintiffs' Response to Defendants' Cross-Motion for Summary Judgment and Reply in Support of Plaintiffs' Motion for Summary Judgment on April 11, 2025.

U.S.C. § 1395fff(b)(3)(B)(iv) ("*Adjustment for case-mix changes*. Insofar as the Secretary determines that the adjustments under paragraph (4)(A)(i) for a previous fiscal year or year (or estimates that such adjustments for a future fiscal year or year) did (or are likely to) result in a change in aggregate payments under this subsection during the fiscal year or year that are a result of changes in the coding or classification of different units of services that *do not reflect real changes in case-mix, the Secretary may adjust the standard prospective payment amount (or amounts)* under paragraph (3) for subsequent fiscal years or years so as to eliminate the effect of such coding or classification changes.") (emphasis added); *see also* END STAGE RENAL DISEASE PROGRAM, 42 U.S.C. § 1395rr(b)(12) ("The *Secretary shall establish a basic case-mix adjusted prospective payment system* for dialysis services furnished by providers of services and renal dialysis facilities in a year to individuals in a facility and to such individuals at home. *The case-mix under such system shall be for a limited number of patient characteristics*.") (emphasis added). The applicable statute authorizing Star Ratings and corresponding quality bonus determinations does not mention case-mix adjusting scores. *See* 42 U.S.C. § 1395w-23(O).

Defendants spill significant ink attempting to distinguish other instances in which Congress and Defendants specifically refer to case-mix adjustments, in various statutes and regulations. *See* Defs.' Br. at 19–21. But the distinctions that Defendants attempt to draw miss the point: Congress and Defendants know exactly how to invoke case-mix adjustments and they do so when they deem it appropriate. Yet, neither Congress nor Defendants authorized case-mix adjustments for calculating Star ratings for CAHPS-based measures. This is particularly striking given that Defendants have defined the term "case mix adjustment" within the very same regulatory scheme and applied it elsewhere. Set against the backdrop of statutes and regulations that specifically invoke "case-mix adjustments," the absence of any such invocation for CAHPS-based measures

carries legal significance. *See United States v. Wong Kim Bo*, 472 F.2d 720, 722 (5th Cir. 1972) ("[W]here Congress has carefully employed a term in one place and excluded it in another, it should not be implied where excluded."); *Smith v. Comm'r of Soc. Sec.*, 482 F.3d 873, 876 (6th Cir. 2007) ("When an agency includes a requirement in only one section of a regulation, we presume the exclusion from the remainder of the regulation to be intentional.").

Thus, it is undisputed that the plain text of the applicable regulation (42 C.F.R. § 422.166(a)(3)) does not contemplate a case-mix adjustment. But Defendants applied a case-mix adjustment to calculate Plaintiffs' CAHPS-based measure scores. This action is contrary to law (and, by definition, also arbitrary and capricious) and had the catastrophic impact of driving down Plaintiffs' contract scores and preventing Plaintiffs from obtaining millions in quality bonus payments. *See Erie Blvd. Hydropower, LP v. FERC*, 878 F.3d 258, 269 (D.C. Cir. 2017) ("[I]f an agency action fails to comply with its regulations, that action may be set aside as arbitrary and capricious."). Ostensibly in recognition of the lack of *express* regulatory support, as set forth in the next two sections, Defendants justify applying a case-mix adjustment to Plaintiffs' CAHPS-measure scores by pointing to sub-regulatory guidance to justify its deviation from the plain language. But Defendants lack legal authority to case-mix adjust CAHPS-based measure scores and cannot do so through sub-regulatory fiat—whether in a preamble or in written agency guidance.

B.   **Defendants' First Argument That They Have Discretion to Determine Whether or Not to Case Mix Adjust Measures Is Incorrect and Concedes That Defendants Are Violating the Plain Language of 42 C.F.R. § 422.166(a)(3).**

In Defendants' Brief, Defendants argue that the "regulations allow for case mix-adjustments in the calculation of measure scores" by pointing to the regulatory definition of that term. *See* Defs.' Br. at 17 (citing 42 C.F.R. § 422.162(a)). Defendants suggest that the existence of this definition allows them to case-mix adjust seemingly any measure score within the Medicare

Advantage universe. But the mere existence of a regulatory definition for a term does not confer authority on the agency to implement the defined concept ad hoc when the applicable regulations do not even use the term—particularly when other calculations do use the defined term. *See* 42 C.F.R. § 422.166(f)(2), (3). And here, conspicuously absent from Defendants' argument is any reference to a regulation authorizing the agency to case-mix adjust CAHPS-based measure scores. None exists. This is fatal to Defendants' position.

Defendants next cite to an entirely different regulation (42 C.F.R. § 422.164) to argue that "[t]he 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." *Id.* (citing 42 C.F.R. § 422.164(c)). For support, Defendants state that "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)." Defs.' Br. at 17. In other words, Defendants argue that because they are permitted to propose new measures and remove prior measures from the Star Ratings calculation, they are given *carte blanche* to determine whether a measure is case-mix adjusted. *Id.* Defendants' reasoning is fundamentally flawed.

When or how Defendants announce new measures is a red herring and has no relevance to this lawsuit. What *is* relevant is that Defendants case-mix adjusted Plaintiffs' CAHPS-based measure scores without legal authorization. And Defendants have not identified a single law that sanctions their overreach. Instead, Defendants wade into a discussion of a so-called "iterative process" under 42 C.F.R. § 422.164 through which they announce new measures that they intend to include in Star Ratings. Defendants argue that this "iterative process" somehow implicitly grants them *carte blanche* to "determine[] whether a particular measure is to be case-mix adjusted." *See*

10

Defs.' Br. at 17. When or how Defendants "announce[] new measures and solicit[] feedback" under 42 C.F.R. § 422.164 is of no moment to whether Defendants are authorized by the applicable regulation in 42 C.F.R. § 422.166(a)(3) to calculate measure-level Star scores with a case-mix adjustment. Defendants' "iterative process," wherein new measures are revealed, does not trump the express mandates set forth in 42 C.F.R. § 422.166(a)(3) as to how to calculate a measure-level Star rating for CAHPS-based measures. In other words, while Defendants do have authority to identify new measures, if those measures are CAHPS-based, then the Star score assigned to each measure must still follow the prescriptive process set forth in 42 C.F.R. § 422.166(a)(3).

Further, the Administrative Record fails to include any discussion of the impact on the ad hoc case-mix adjustment vis-à-vis the standard error adjustments outlined in the regulation. There is also no discussion of any implications on the weighting of any racial, age, educational bias or other demographic scores utilized by CMS to "rationalize" these survey measures scores. Such a lack of basic information underscoring these case-mix adjustments fails to satisfy even the most basic threshold for adequately considering application of such a significant action by the agency to take outside of a regulation. *See generally Motor Vehicle Mfrs. Ass'n of U.S., Inc. v. State Farm Mut. Auto. Ins. Co.*, 463 U.S. 29 (1983) (holding an agency's failure to consider an important aspect of the problem and relevant factors is arbitrary and capricious).

Undeterred, however, Defendants continue to argue that "Elevance wrongly contends that 'CMS's sub-regulatory guidance *adds* the case-mix adjustment[.]'" Defs.' Br. at 21 (emphasis in original). According to Defendants, "Elevance's suggestion that CMS establishes case-mix adjustments in its technical guidance separate and apart from the regulations is wrong." *Id.* But the applicable regulatory text in 42 C.F.R. § 422.166(a)(3) speaks for itself and does not authorize case-mix adjustments for CAHPS-based measures. Therefore, Defendants acted arbitrarily and

11

capriciously by case-mix adjusting CAHPS-based measure scores—injuring Plaintiffs by tens of millions of dollars—on the basis of sub-regulatory guidance inconsistent with applicable law. *See Texas*, 91 F.4th at 291 ("[i]n taking final action, an agency must comply with its own regulations"); *Gulf States Mfrs.*, 579 F.2d at 1308 ("[T]he failure of an agency to follow its regulations renders its decision invalid."); *Erie Blvd. Hydropower, LP*, 878 F.3d at 269 ("[I]f an agency action fails to comply with its regulations, that action may be set aside as arbitrary and capricious.").

C.   Defendants' Second Argument That the Preamble to 2018 Rulemaking Authorizes Defendants to Case-Mix Adjust CAHPS-Based Measures is Legally Incorrect, and Similar Arguments Have Repeatedly Been Rejected by Courts (Including With Respect to Star Ratings).

Unable to rely upon the express regulatory text to case-mix adjust CAHPS-based measures, Defendants fall back on an argument previously rejected by several courts. Specifically, after acknowledging that the applicable regulatory text does not authorize a case-mix adjustment, Defendants cite to the preamble to the 2018 final rule that implemented 42 C.F.R. § 422.166, in which Defendants responded to a comment by stating that they "'provide[] a detailed explanation of the CAHPS methodology including case-mix adjustment in the annual Star Ratings Technical Notes. . . .'" *See* Defs.' Br. at 19 (quoting 83 Fed. Reg. at 16,555 (April 16, 2018)). In other words, adding to their prior argument that "[t]he regulations allow, but do not mandate, that CAHPS measures be case-mix adjusted," Defendants contend the "2018 final rule *expects* that CAHPS measures *could be* case-mix adjusted." *Id*. at 8 (emphasis added).[3] Defendants' argument—which Defendants have tried over the years to advance—is legally wrong as numerous courts have held.

---

[3] Defendants themselves waffle in how they characterize the significance of the rulemakings to which they cite. On the one hand, Defendants claim that their rulemakings "*expressly authorize*" case-mix adjustments for CAHPS measures. *See* Defs.' Br. at 18 (emphasis added). On the other, Defendants contend the 2018 final rule (to which Defendants referred in the same paragraph as the preceding quotation) "*expects that* CAHPS measures could be case-mix adjusted." *See* Defs.' Br. at 8 (emphasis added). Defendants' characterizations of the same rulemaking are inconsistent.

Specifically, the gist of Defendants' argument is that its cherry-picked commentary from the 2018 rulemaking preamble somehow constitutes the legal authorization for case-mix adjusting CAHPS-based measures. However, courts across the country have held—time and again—that preamble language does not have the force and effect of law, nor can it be used to contradict plain regulatory text. That is because "a preamble does not create law; that is what a regulation's text is for." *See Tex. Children's Hosp. v. Burwell*, 76 F. Supp. 3d 224, 237 (D.D.C. 2014). Indeed, the "'real dividing point' between the portions of a final rule with and without legal force is designation for 'publication in the Code of Federal Regulations.'" *AT&T Corp. v. FCC*, 449 U.S. App. D.C. 106, 112 (D.C. Cir. 2020) (quoting *Brock v. Cathedral Bluffs Shale Oil Co.*, 796 F.2d 533, 539 (D.C. Cir. 1986)); *see also All. for Hippocratic Med. v. U.S. F.D.A.* 78 F.4th 210, 264 (5th Cir. 2023) ("[W]e do not use preambles to expand the meaning of clear regulatory text.") (Ho, J., concurring); *Air Prods. & Chems., Inc. v. GSA*, 700 F. Supp. 3d 487 (N.D. Tex. 2023) ("[P]reambles do not 'limit or expand the scope of the operative' text.") (internal quotations omitted); *Texas v. HHS*, No. 6:24-cv-348, 2025 WL 818155, at *9 (E.D. Tex. Mar. 13, 2025) (holding that preamble cannot be read to conflict with the language of the regulation) (quoting *Peabody Twentymile Mining, LLC v. Sec'y of Lab.*, 931 F.3d 992, 998 (10th Cir. 2019) ("[W]hile the preamble can inform the interpretation of a regulation, it is not binding and cannot be read to conflict with the language of the regulation itself.") (citations omitted)).

Indeed, this is not the first time Defendants have attempted to take refuge in language from a regulatory preamble that was inconsistent with their own regulations on calculating Star Ratings.

Furthermore, none of the cited statements from the 2018 rulemaking preamble expressly authorize case-mix adjustments for CAHPS-based measures. But to the extent that the preamble language could be read to "***expect***" (in the agency's parlance) or "presuppose" that CAHPS-based measures could be case-mix adjusted, Defendants still lack the necessary statutory or regulatory predicate for case-mix adjusting CAHPS-based measures. No such legal foundation exists.

Just last year, two separate federal courts rejected Defendants' attempt to do just that. *See Elevance Health, Inc. v. Becerra*, 736 F. Supp. 3d 1, 23–25 (D.D.C. June 7, 2024) (Moss, J.); *see also Scan Health Plan v. HHS*, Case No. 23-cv-03910, 2024 WL 2815789, at *6–7 (D.D.C. June 3, 2024) (Nichols, J.). In those cases, Defendants attempted to rely upon the preamble to the rulemaking revising 42 C.F.R. § 422.166(a)(2) to argue that they could calculate measure-level Star scores in a way that was not permitted by the regulatory text. However, in both cases, the courts rejected Defendants' reliance upon the preamble because the regulation did not permit what Defendants attempted to do. *See Elevance Health, Inc.*, 736 F. Supp. 3d at 23–24; *see also Scan Health Plan*, No. 23-cv-03910, 2024 WL 2815789 at *6.

As in the cases referenced above, here, the issue is simple. The preamble language that Defendants cite and the accompanying language from their sub-regulatory guidance is not consistent with the applicable regulatory requirements to calculate CAHPS-based measure scores in 42 C.F.R. § 422.166(a)(3). That regulation does not authorize "case-mix adjustments" with respect to calculating the measure scores for CAHPS-based measures, *see* 42 C.F.R. § 422.166(a)(3), and Defendants cannot rely on a sentence or two in the preamble to expand or modify the regulatory text. This is no small issue. Indeed, by case-mix adjusting CAHPS-based measures, Defendants have harmed Plaintiffs by hundreds of millions of dollars. Such a substantive dimension of Defendants' methodology for calculating Star Ratings must be authorized by statute or regulation, not buried within stray remarks in a regulatory preamble. Accordingly, Defendants' continued reliance on the preamble is incorrect as a matter of law.

D.   Defendants' Final Argument That They Have Previously Case-Mix Adjusted CAHPS Measures Without Being Sued is Unavailing.

In a last-ditch attempt to justify its actions that are contrary to law, Defendants point out that they have applied case-mix adjustments to CAHPS measure scores without facing a prior legal

challenge. *See* Defs.' Br. at 21. They then argue that "to CMS's knowledge, no MAO has suggested that case-mix adjustments are not authorized by CMS's regulations" and contend that "[c]ase-mix adjusting is not a policy to which Elevance objected or would have objected before Elevance knew its overall contract scores." *Id.* But the fact that Defendants have acted arbitrarily and capriciously in the past without facing legal action is irrelevant, particularly given that Plaintiffs were harmed this year. Notably, Defendants do not challenge Plaintiffs' case as untimely, because they cannot. *See Corner Post, Inc. v. Bd. of Governors of the Fed. Rsrv. Sys.*, 603 U.S. 799, 825 (2024) (holding that "[a]n APA claim does not accrue for purposes of [28 U.S.C.] § 2401(a)'s 6-year statute of limitations until the plaintiff is injured by final agency action"). And the fact that Plaintiffs have now sued Defendants on this issue when Defendants' unlawful actions caused them harm underscores Plaintiffs' credibility. Plaintiffs have not raised frivolous lawsuits in the past. Rather, Plaintiffs sued Defendants with respect to their 2025 Star Ratings because those improperly calculated ratings caused Plaintiffs to suffer a cognizable injury.

III.   DEFENDANTS IMPROPERLY CALCULATED CAHPS-BASED MEASURE SCORES USING A NATIONAL *WEIGHTED* AVERAGE; 42 C.F.R. § 422.166(A)(3) REQUIRES THE USE OF THE AVERAGE WITHOUT ANY WEIGHTING.

Defendants have violated the plain text of 42 C.F.R. § 422.166(a)(3) in a second way when calculating Plaintiffs' CAHPS-based measure scores. Specifically, under that regulation, when calculating the CAHPS-based measure scores, Defendants must compare the contract's actual CAHPS-based measure scores to the "national average" measure score to determine if it is "statistically significantly different" from that national average. While the concepts of "statistically significantly different" can be complex, Defendants' violation is straight-forward. Namely, instead of comparing the contract's measure scores to the "national average" measure score, as the regulation requires, Defendants concede that they created sub-regulatory guidance that requires the use of the national *weighted* average, which is an entirely different concept all together.

15

A.   Plaintiffs Did Not Waive this Challenge.

As an initial matter, Defendants cite to *Johnson v. Thibodaux City*, 887 F.3d 726, 736 (5th Cir. 2018) and make the conclusory two-sentence argument that Plaintiffs "waived" their challenge on this ground. Defs.' Br. at 22. Defendants gave short shrift to this argument because it is baseless. Under the applicable notice pleading requirements under Rule 8, Plaintiffs' Amended Complaint more than provides fair notice of Plaintiffs' claims related to Defendants' calculation of their CAHPS measure scores, including by referencing the national average (or mean). *See* ECF No. 22, Am. Compl. at ¶¶ 31, 34, 40, 59–60, 64–70 (alleging that Defendants improperly calculated Plaintiffs' CAHPS-based measure scores); *see also Arredondo v. Schlumberger Ltd.*, 583 F. Supp. 3d 783, 800 (W.D. Tex. 2022) ("plaintiffs are not required to use 'magic words' to state a claim")*; Montasir Hossain v. Boeing Co.*, No. 3:23-cv-2494-S, 2025 WL 242246, at *4 (N.D. Tex. Jan. 17, 2025) (holding the defendant was on notice of plaintiff's intent to assert disparate treatment and failure to accommodate theories based on its Title VII claim and facts in the complaint); *De la Hoya v. Coldwell Banker Mex., Inc.*, 125 Fed. Appx. 533, 537 (5th Cir. 2005) (holding that plaintiff's pleading put defendants on notice of plaintiff's intention to pursue any available theory of corporate liability). Furthermore, Defendants do not and cannot assert they did not have adequate notice of Plaintiffs' claim or were otherwise prejudiced by Plaintiffs' claim, because Plaintiffs briefed this claim in their motion for summary judgment and Defendants fully evaluated and responded to Plaintiffs' claim. *See* Defs.' Br. at 22–25.

Defendants' reliance on *Johnson*, 887 F.3d at 736, is wholly misplaced. In *Johnson,* the Fifth Circuit held that plaintiffs "may not ***defeat*** summary judgment by raising a theory… that was never pleaded or tried…" *See Johnson*, 887 F.3d at 736 (emphasis added). In other words, a plaintiff generally may not raise a legal claim for the first time in response to a defendant's motion for summary judgment. Plaintiffs not only raised their claim in the Amended Complaint but also

in their motion for summary judgment. *See* Pls.' Br. at 22–25. Thus, *Johnson* is inapplicable, and Plaintiffs did not waive their challenge to Defendants' calculation and use of a national *weighted* average to calculate CAHPS-based measures.

B.   Defendants Concede That 42 C.F.R. § 422.166(a)(3) Requires the Comparison of Plaintiffs' Contract Measure Scores to the "National Average," and That They Use a National *Weighted* Average Contrary to the Plain Language.

Defendants' substantive response fares no better. Specifically, Defendants agree that 42 C.F.R. § 422.166(a)(3) is the applicable regulation and that it plainly states that, when calculating the CAHPS-based measure scores, Defendants must compare the contract's actual CAHPS-based measure scores to the "national average CAHPS measure score" to determine if the contract's measures are "statistically significantly different" from that national average. *See* Defs.' Br. at 23; *see also* 42 C.F.R. § 422.166(a)(3)(i)(B), 422.166(a)(3)(ii)(C), 422.166(a)(3)(iii)(C), 422.166(a)(3)(iv)(C) & 422.166(a)(3)(v)(B) (relating to assigning 1, 2, 3, 4, or 5 Stars to a CAHPS-based measure score). Defendants further concede that their sub-regulatory guidance calculates CAHPS-based measures scores using the national *weighted* average. Defs.' Br. at 23-24. Specifically, Defendants explain that "[t]o 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." Defs.' Br. at 23 (emphasis added); *see also id.* at 10 ("To calculate CAHPS measure scores . . . . CMS weights the contract scores by the . . . enrollment assessed at the time of sample design, and then averages them.").[4]

In other words, Defendants admit that instead of calculating the actual national average as the regulation mandates, Defendants use the weighted measure scores for each contract to calculate

---

[4] At the same time, in direct conflict with their statements on how they calculate the national *weighted* average, Defendants inexplicably assert that "CMS does not calculate weighted national averages." *See* Defs.' Br. at 24. This is doublespeak.

a weighted national average. Defendants justify using a weighted average to create a "fair comparison" between contracts and their measure level scores and contend this is "reasonable and entirely consistent with what the regulation requires . . . " Defs.' Br. at 23–24. Defendants baldly assert without any support in the Administrative Record that using the actual "national average"—as required by the regulation—"would make little sense" because it would not account for the size of certain contracts. *Id.* at 25. In other words, Defendants posit that following the plain language of the regulation "would not make sense" because it is not how Defendants conduct the calculation per their guidance. *See id.*

Ultimately, Defendants do not and cannot in good faith dispute that an "average" and "weighted average" are two different things. While 42 C.F.R. § 422.166(a)(3) does not define "national average," courts "give undefined words their ordinary, contemporary, common meaning." *See Contender Farms LLP. v. U.S. Dep't of Agric.*, 779 F.3d 258, 269 (5th Cir. 2015) (internal quotations omitted). Courts "assume that the plain meaning of those words is the same as their ordinary (i.e., 'common' or 'natural') meaning, which is often derived from dictionaries." *See Vitol, Inc. v. United States*, 30 F.4th 248, 253–54 (5th Cir. 2022). Cambridge Dictionary defines "average" as "the result you get by adding two or more amounts together and dividing by the total number of amounts."[5] Similarly, Merriam-Webster provides that "average" equals an "arithmetic mean," and defines "arithmetic mean" as "a value that is computed by dividing the sum of a set of terms by the number of terms."[6] Plaintiffs' expert, J. Mark Abernathy, similarly

---

[5] *See Average*, Cambridge University Press & Assessment, DICTIONARY.CAMBRIDGE.ORG, https://dictionary.cambridge.org/dictionary/english/average (last visited Apr. 8, 2025).

[6] *See Average*, Merriam-Webster, Inc., MERRIAM-WEBSTER.COM, https://www.merriam-webster.com/dictionary/average (last visited Apr. 8, 2025); *Arithmetic Mean*, Merriam-Webster, Inc., MERRIAM-WEBSTER.COM https://www.merriam-webster.com/dictionary/arithmetic%20mean (last visited Apr. 8, 2025).

explained that the "national average without any weighting" is "the simple average of all contract values… where each contract is weighted equally," consistent with the aforementioned dictionary definitions of "average." *See* Pls.' Br. at 22–23 (citing App. 11–12, Declaration of J. Mark Abernathy ¶ 22 ("Abernathy Decl.")). Using a weighted average—as Defendants concede they do pursuant to their sub-regulatory guidance—is inconsistent with this plain text. Accordingly, Defendants' calculation of the CAHPS measure scores is contrary to the plain language of the regulation. *See Dominion*, 968 F.3d at 434 ("When a regulation is unambiguous courts owe no deference to the agency's interpretation of it and simply apply the regulation's plain meaning."); *ExxonMobil*, 867 F.3d at 573 ("If the regulation is unambiguous, we will not defer to the agency's interpretation."); *Diamond Servs. Corp. v. Curtin Mar. Corp.*, 99 F.4th 722, 731 (5th Cir. 2024) (explaining that a regulation must be ambiguous for a court to defer to the agency's interpretation); *see also Thomas Jefferson Univ. v. Shalala*, 512 U.S. 504, 512 (1994) (holding a court cannot accept an agency's interpretation that is contrary to the regulation's plain language); *Perez v. Mortg. Bankers Ass'n*, 575 U.S. 92, 104–05 (2015) (holding an agency's interpretation is substantively invalid when it conflicts with the text of the regulation the agency purported to interpret); *Medica Ins. Co v. Becerra*, No. 1:22-cv-1440, 2023 WL 6314571, at *10-14 (D.D.C. Sept. 28, 2023) (holding that CMS's calculation violated the plain meaning of the regulation and the agency misinterpreted the regulation governing the calculation formula).

Finally, presumably because Defendants are not following the plain language of 42 C.F.R. § 422.166(a)(3), Defendants engage in linguistic gymnastics to mischaracterize Plaintiffs' argument. *See* Defs.' Br. at 24–25. According to Defendants, Plaintiffs "appear[] to contend that CMS must use the national *contract*-level average as the national average" and this "misconstru[es] CMS's regulation." *Id.* at 24. In reality, as set forth in Plaintiffs' Opening Brief

19

and as explained in the Declaration of J. Mark Abernathy, when the regulatory strictures are followed and the CAHPS-measure scores for Plaintiffs' contracts are calculated using the "national average," as opposed to using a weighted average as Defendants did, two of Plaintiffs' contracts are impacted. *See* Pls.' Br. at 2; *see also* App. 11, Abernathy Decl. ¶ 20. Plaintiffs did nothing more than apply the requisite formula using the appropriate national average. Defendants' sleight of hand does not change that outcome.

IV.   **DEFENDANTS ACTED CONTRARY TO LAW AND ARBITRARILY AND CAPRICIOUSLY BY CALCULATING AND ROUNDING THE STAR RATING FOR H3655 TO THE SIXTH (OR MILLIONTH) DECIMAL PLACE.**

Plaintiffs argued in their Opening Brief that by rounding contract H3655's numeric score of 3.749565 at the sixth (*i.e.*, *millionth*) decimal place, Defendants assigned the contract 3.5 Stars instead of 4 Stars, causing Plaintiffs to lose approximately $375 million in quality bonus payments to which they are entitled. As Plaintiffs pointed out, Defendants' rounding methodology is (a) contrary to its own regulations, which require Defendants to use "traditional rounding rules" (a defined term), and (b) arbitrary and capricious because it fails to account for imprecision and statistical variability in Defendants' calculations. Defendants' *post hoc* rationalizations are not supported by the Administrative Record. Even if they were, Defendants' arguments reveal they are contrary to the plain language of the regulation and arbitrary and capricious.

A.   **The Administrative Record Does Not Support Defendants' Actions.**

Defendants' *post hoc* rationalizations made outside of the Administrative Record and in the declaration of its "expert" to support the agency's decisions are inappropriate and should be disregarded by the Court. An agency may not advance arguments before the court "without first presenting them in the administrative record." *Louisiana*, 90 F.4th at 469 (finding that the Supreme Court has repeatedly reaffirmed the prohibition of convenient litigating positions and *post hoc*

20

rationalizations); *see also Luminant Generation Co., LLC*, 675 F.3d at 925 ("We must disregard any *post hoc* rationalizations of the [agency]'s action and evaluate it solely on the basis of the agency's stated rationale at the time of its decision."). When an administrative record is inadequate, a court may obtain affidavits from the agency but "the new materials should be merely explanatory of the original record and should contain no new rationalizations." *See Kirwa v U.S. Dep't of Defense*, 285 F. Supp. 3d 257, 269 (D.D.C. 2018). At bottom, APA review "involves more than rubberstamping action based on bare declarations from the agency amounting to 'trust us, we had good… reasons for what we did." *Id.* at 270 (citations omitted).

Defendants rely extensively on the Declaration of Elizabeth Goldstein ("Ms. Goldstein") to support their cross-motion for summary judgment. Indeed, Ms. Goldstein's declaration is the only document included in the Defendants' appendix. According to her declaration, Ms. Goldstein is the Director of the Division of Consumer Assessment and Plan Performance, Medicare Drug Benefit and C&D Data Group at CMS. *See* Defs.' App. 1. Unlike the declaration of Plaintiffs' experts, Ms. Goldstein's declaration fails to detail any of her qualifications or expertise in the various subject-matter areas at issue in this litigation (e.g., survey research methodology and analysis and statistics). *Compare* App. 225-236, Curriculum Vitae of J. Mark Abernathy; App. 237–41, Curriculum Vitae of Paul Diver, Ph.D.; App. 242–311, Curriculum Vitae of Paul J. Lavrakas, Ph.D. *with* Defs.' App. 1–4. In line with Defendants' overall response to Plaintiffs' claims, Ms. Goldstein's declaration also fails to cite to any portions of the Administrative Record on which she relies—if any—or otherwise disclose any pertinent literature or evidence to support the opinions and conclusions therein. *See generally* Defs.' App. 1-4.

Setting Ms. Goldstein's expert qualifications aside, Defendants may not merely rely on Ms. Goldstein's bare declarations or other *post hoc* rationalizations to explain or support

Defendants' actions and decisions. Simply, the grounds upon which an agency's action must be judged are those upon which the Administrative Record discloses its action was based. *See Louisiana*, 90 F.4th at 469 (internal quotations omitted). Otherwise, the agency and agency's counsel could attempt to fashion explanations and justifications for the agency's arbitrary and capricious actions out of whole cloth. Ms. Goldstein herself certified the completeness and accuracy of the Administrative Record in this case, *see* ECF. No. 23, yet the Defendants' brief (and Ms. Goldstein's declaration) is devoid of Administrative Record support. At bottom, Defendants fail to show through the Administrative Record that the agency complied with applicable statutory and regulatory requirements in calculating Plaintiffs' measure and overall Star Ratings. Thus, the agency's calculation of Plaintiffs' Star Ratings is contrary to law and arbitrary and capricious on this basis.

> **B.   Defendants' Arguments Confirm that They Did Not Follow Traditional Rounding Rules In Calculating Plaintiffs' Overall Star Rating.**

>> ***i.   Defendants' arguments confirm that they did not round Plaintiffs' Overall Star Rating for H3655 using traditional rounding rules as mandated by 42 C.F.R. § 422.166(d).***

Defendants agree with Plaintiffs on three critical things. First, Defendants agree that the applicable regulations in 42 C.F.R. 422.166(d) require a contract's Overall Star Rating to be determined "on a 1- to 5-star scale ranging from 1 (worst rating) to 5 (best rating) in half-star increments **using traditional rounding rules**." 42 C.F.R. § 422.166(d)(2)(iv) (emphasis added); *see also* Defs.' Br. at 26. Second, Defendants agree that the applicable regulations define "traditional rounding rules" as follows:

> Traditional rounding rules mean that the last digit in a value will be rounded. If rounding to a whole number, look at the digit in the first decimal place. **If the digit in the first decimal place is 0, 1, 2, 3, or 4, then the value should be rounded down by deleting the digit in the first decimal place. If the digit in the first decimal place is**

**5 or greater, then the value should be rounded up by 1 and the
digit in the first decimal place deleted**.

42 C.F.R. § 422.162(a) (emphasis added); *see also* Defs.' Br. at 28–29. Third, and finally,

Defendants agree that, in order to appropriately round a contract to the nearest half-star, a midpoint

between two half Star scores must be used. As Defendants state, a contract's score is either "above,

below, or equal to the midpoint" and if "above or equal to the midpoint, [the contract should be]

round[ed] up to the nearest half-star increment" and if "below, [the contract should be] round[ed]

down to the nearest half-star increment." Defs.' Br. at 30; *see also* A.R. 31, App. 101, 2025

Technical Notes. In other words, CMS's technical guidance, consistent with the regulatory

definition, explains that you should look to the second (or hundredth place) decimal to determine

if the contract is equal to, above, or below the midpoint.

Ultimately, the dispositive question is what does 3.749565 round to using the definition of

"traditional rounding rules?" Applying that definition, one looks to the digit to the left of the

rounding point (which the parties agree is 3.75 to round to the nearest half star). Thus, we look at

the 9. Since that digit is 5 or greater, then the value should be rounded up by 1 (*i.e.*, to a 10) and

the 9 should be deleted. Stated more simply, 3.749 rounds to 3.75. This is strictly compliant with

the definition of traditional rounding rules. Further, as Defendants concede in their brief, because

3.75 is "above or equal to the midpoint, [the contract should be] round[ed] up to the nearest half-

star increment." Defs.' Br. at 30; *see also* A.R. 31, App. 101, 2025 Technical Notes.

Conversely, Defendants argue that they merely calculate out a contract's score to the sixth

decimal but actually round at the hundredth decimal. *See* Defs.' Br. at 27–28. Defendants'

argument about rounding at the second decimal is not supported by citation to the Administrative

Record because there is no such support. Rather, it is evident that Defendants are engaging in *post

hoc* rationalizations to attempt to meet the regulatory requirements of using traditional rounding

23

rules. What *is* supported in the Administrative Record shows that Defendants do not actually round the Overall Star Rating at all using traditional rounding rules. Rather, as Defendants' Technical Notes make clear, Defendants calculate out each contract's Overall Star Rating to six decimals and then round at that sixth decimal.

For instance, the 2025 Technical Notes expressly state the following with respect to "rounding rules" for Overall Star Ratings:

<figure>

Rounding Rules for Summary and Overall Ratings

The results of the summary and overall calculations are rounded to the nearest half star (i.e., 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0). Table 22 summarizes the rounding rules for converting the Part C and D summary and overall ratings into the publicly reported Star Ratings.

Table 22: Rounding Rules for Summary and Overall Ratings

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

For example, a summary or overall rating of 3.749999 rounds down to a rating of 3.5, and a rating of 3.750000 rounds up to rating of 4. That is, a score would need to be at least halfway between 3.5 and 4 (having a minimum value of 3.750000) in order to obtain the higher rating of 4.

</figure>

A.R. 31, App. 101, 2025 Technical Notes (emphasis added). Nowhere does this support Defendants' position before this Court that they are rounding at the second decimal. To the contrary, this clearly demonstrates that Defendants are not applying traditional rounding rules at all, but rather "converting" numbers rounded to the sixth decimal to an Overall Star Rating.

Notably, Defendants apply traditional rounding rules to make *other* calculations. Indeed, one only needs to look to the immediate prior section in the 2025 Technical Notes, in which

Defendants expressly explain that they apply traditional rounding rules for rounding *measure scores*. A.R. 30–31, App. 100–101, 2025 Technical Notes. As that section states:

> Measure scores are rounded using traditional rounding rules. These are standard "round to nearest" rules prior to cut point analysis. To obtain a value with the specified level of precision, the single digit following the level of precision will be rounded. If the digit to be rounded is 0, 1, 2, 3 or 4, the value is rounded down, with no adjustment to the preceding digit. If the digit to be rounded is 5, 6, 7, 8 or 9, the value is rounded up, and a value of one is added to the preceding digit. After rounding, all digits after the specified level of precision are removed. If rounding to a whole number, the digit to be rounded is in the first decimal place. If the digit in the first decimal place is below 5, then after rounding the whole number remains unchanged and fractional parts of the number are deleted. If the digit in the first decimal place is 5 or greater, then the whole number is rounded up by adding a value of 1 and fractional parts of the number are deleted.

Id. This dichotomy between what Defendants do for the Overall Star Ratings (*i.e.*, use six decimals to "convert" the score to a Star Rating) versus the measure scores (*i.e.*, apply traditional rounding rules) exposes the fact that Defendants do not use traditional rounding rules for the Overall Star Rating. As a result, Defendants' methodology is contrary to law because it is inconsistent with the regulatory requirement of determining Star Ratings in half-star increments by applying traditional rounding rules. And while Defendants' methodology creates the *appearance* of precision, as explained below, it is fraught with imprecision and errors, which its arbitrary rounding methodology overlooks and only exacerbates.

ii. Defendants' unsubstantiated argument that it does not round to the sixth decimal is inconsistent with the evidence in the Administrative Record.

In an apparent attempt to deflect from the fact that the Administrative Record proves Defendants did not follow traditional rounding rules, Defendants argue that they do not round to the sixth decimal and rather simply calculate numbers out to the sixth decimal. Defs.' Br. at 27. Specifically, Defendants argue at length that CMS "could not be clearer that it calculates overall

rating scores with *at least* six digits of precision" but that "[w]hen CMS displays the unrounded

overall rating [sic] 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.' Br. at 28 (citing Defs.' App. 2, Goldstein Decl. ¶ 2) (emphasis in original).

According to Defendants, the agency truncates the numerical score at the sixth decimal for "display

purposes" and "[u]sing six decimal places for display purposes does not indicate rounding in the

calculation of the overall star rating, and nowhere in regulation or in technical guidance does CMS

state that it rounds to six decimal places." *Id.* at 28.

Not only does this argument miss the point as the dispositive issue is whether Defendants

followed the applicable regulatory requirements by applying traditional rounding rules (and as set

forth above, Defendants did not), it is not supported in the Administrative Record. In fact,

Defendants' argument actually is contrary to the evidence in the Administrative Record.

Specifically, the contract "Score Card" that Defendants provided to Elevance for contract H3655

shows that, notwithstanding Defendants' protestations to the contrary, Defendants do, in fact,

round the contract's numerical scores to the sixth decimal place. A.R. 838, Suppl. App. 314. As

explained in the Supplemental Declaration of Paul Diver, Ph.D., in calculating the weighted

average of measure specific Star Ratings, Defendants round that value to the sixth decimal place.

See Suppl. App. 325, Supplemental Declaration of Paul Diver, Ph.D. ("Diver Suppl. Decl.") ¶ 16.

Defendants do this through the "FIXED" function in Microsoft Excel, which is plainly apparent in

the Excel file. *See id.* As explained in Microsoft's support page, the FIXED function in Excel

"[r]ounds a number to the specified number of decimals . . . and returns the result as text." Suppl.

App. 327, Diver Suppl. Decl. ¶ 17 (citing Suppl. App. 317, Microsoft Support, *FIXED function*,

https://support.microsoft.com/en-us/office/fixed-function-ffd5723c-324c-45e9-8b96-e41be2a8274a (last visited April 9, 2025) .

Defendants then use that score rounded to the sixth decimal to calculate the Overall Star Rating, causing that number to also be rounded to the sixth decimal. *See* Suppl. App. 328-32, Diver Suppl. Decl. ¶¶ 19–22. Thus, contrary to Defendants' assertions, the Overall Star Rating is not simply truncated at the sixth decimal for display purposes, but rather that number is rounded to the sixth decimal place. *See id.* ¶ 22. And as Dr. Diver shows, if one were to remove the FIXED function in the Score Card for H3655, the Overall Star Score numerical number would be 3.7495648235—which Defendants clearly rounded to 3.749565 when reporting it. *See* Suppl. App. 332-34, Diver Suppl. Decl. ¶¶ 24–25; *see also* A.R. 838, Suppl. App. 314. Accordingly, Defendants' argument is not only astray of the dispositive issue but also contrary to the evidence in the Administrative Record. Furthermore, the fact that Defendants do not understand their own methodology demonstrates that they are acting in an arbitrary and capricious manner.

iii. *Defendants' argument that Plaintiffs seek to "double round" is predicated upon an incorrect premise and not relevant.*

Defendants' next argument fares no better. *See* Defs.' Br. at 28–31. Defendants begin by admonishing Plaintiffs for asking the Court to require the agency to "double round" and argue that the "regulation requires it to round only once." Defs.' Br. at 28. Defendants spend the next three pages attempting to persuade the Court that its methodology is superior to Plaintiffs' because, according to the agency, "CMS rounds once" and "Elevance's approach would require CMS to round twice." *See* Defs.' Br. at 30. On that basis, Defendants accuse Plaintiffs of "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" and even assert that "round[ing] twice would violate CMS's regulations and traditional rounding rules." *Id.* Defendants' argument is, once again, unavailing.

27

As an initial matter, Defendants' argument collapses under its own weight because it is premised on the agency's misunderstanding of its own Excel formula and its mistaken belief that it truncates (but does not round) at the sixth decimal place. *See* Suppl. App. 325-32, Diver Suppl. Decl. ¶¶ 16–22. Thus, Defendants round to the sixth decimal using the FIXED function in Excel and then round that number to a half-star increment (again, in a way that is contrary to the regulatory requirement to use traditional rounding rules). In other words, Defendants are guilty of the very conduct they complain is contrary to the regulatory requirements. *See* Defs.' Br. at 30 ("To round twice would violate CMS's regulations and traditional rounding rules.").

Moreover, at the end of the day, this "double rounding" argument is a red herring. Specifically, the applicable regulation states that "[t]he 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). This regulation simply mandates the use of traditional rounding rules but does not foreclose CMS from applying its own policy to determine whether the rounded number is above, below, or equal to the midpoint of .75 or .25. Defs.' Br. at 30 (if a contract's score is "above or equal to the midpoint, [the contract should be] round[ed] up to the nearest half-star increment" and if a contract's score is "below [that midpoint], [the contract should be] round[ed] down to the nearest half-star increment."); *see also* A.R. 31, App. 101, 2025 Technical Notes. This is not an unreasonable application of the regulation, and when applying the traditional rounding rules in that context, the only possible conclusion is that 3.749565 round to 3.75, which is 4 Stars. *See supra* § IV.B.i. Accordingly, Defendants' argument regarding so-called "double rounding" is predicated upon a flawed premise and, in any event, is of no consequence or merit.

28

     C.   **Defendants Acted Arbitrarily and Capriciously in Failing to Account for Imprecision and Statistical Variability for Both Non-CAHPS and CAHPS Measures in Calculating Plaintiffs' Star Ratings.**

           i.   *Non-CAHPS Measures: Defendants fail to account for imprecision in calculating non-CAHPS measure scores and then using those imprecise numbers to calculate the Final Summary Score and Overall Star Rating.*

                    a.   *Defendants do not dispute that the calculation of non-CAHPS measures creates statistical variability.*

As set forth in Plaintiffs' Opening Brief, "[courts] must set aside any action premised on reasoning that fails to account for 'relevant factors' or evinces 'a clear error of judgment.'" *Univ. of Tex. M.D. Anderson Cancer Ctr. v. HHS*, 985 F.3d 472, 475 (5th Cir. 2021) (quoting *Marsh v. Or. Nat. Res. Council*, 490 U.S. 360, 378 (1989)). With respect to the calculation of non-CAHPS measure scores, Defendants acted arbitrarily and capriciously by calculating and rounding the Final Summary Score for H3655 to the sixth decimal (3.749565) because CMS failed to account for the imprecision inherently created by statistical variance in the methodology used to calculate those scores. Fundamental to this argument is the undisputed fact that Defendants calculate the Final Summary Score to "at least six digits of precision after the decimal whenever the data allow it." A.R. 30, App. 100, 2025 Technical Notes. Defendants then use that number calculated out to the sixth decimal to "round" to the half-star. As the 2025 Technical Notes state, "a summary or overall rating of 3.749999 rounds down to a rating of 3.5, and a rating of 3.750000 rounds up to rating of 4. That is, a score would need to be at least halfway between 3.5 and 4 (having a minimum value of 3.750000) in order to obtain the higher rating of 4." A.R. 31, App. 101, 2025 Technical Notes.

As further set forth in Plaintiffs' Opening Brief, the methodology for calculating non-CAHPS measure scores is mean resampling with hierarchical clustering. *See* Pls.' Br. at 12–13; *see also* 42 C.F.R. §§ 422.166(a)(2)(i), 423.186(a)(2)(i). Furthermore, Dr. Paul Diver demonstrated

that a measure's Star Rating can change due to nothing more than random chance alone. *See* Pls.'

Br. at 29–30. As Dr. Diver concluded, calculating and rounding a plan's Final Summary Score to

the second decimal place, as opposed to the sixth, is better reflective of the uncertainty which exists

due to random chance in Defendants' methodology to determine an Overall Star Rating for each

contract and would remove Defendants' current illusion of precision under the current approach.

Id. at 29 (citing App. 26–27, 50, Diver Decl. ¶¶ 8(iii)–(iv), 9, 10, 76). Defendants, however,

calculated and rounded the Final Summary Score to the sixth decimal, and in doing so, completely

ignore the imprecision inherent in the statistical methodology. This is arbitrary and capricious.

b.   *Defendants' reliance on the mystery RAND simulation is improper and, in any event, proves Plaintiffs' argument.*

Defendants agree there is imprecision in the mean resampling with hierarchical clustering

methodology used to calculate the measure scores for non-CAHPS based measures and do not

deny that it would be arbitrary and capricious to not account for this imprecision. Instead,

Defendants again rely upon Ms. Goldstein's Declaration to argue that CMS's vendor, RAND,

"conducted its own larger, more accurate simulation." Defs.' Br. at 35–37. According to

Defendants—citing nothing more than two paragraphs in Ms. Goldstein's Declaration—"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." *Id.* at 35. Based upon this mystery RAND

analysis, Defendants conclude that there is "strong stability" in the calculation of Star Ratings and

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

Defs.' Br. at 36–37.

As an initial matter, Defendants' entire argument should be ignored because it is predicated

upon nothing more than unsubstantiated information that is not in the Administrative Record or

even included in Ms. Goldstein's Declaration. Indeed, the only information before this Court is Ms. Goldstein's two-paragraph summary of that work. Notably, however, Ms. Goldstein is not being offered and has not been qualified as an expert on statistical principles and cannot be used as a proxy for RAND. In other words, whatever mystery work that RAND performed, it is not properly part of the Administrative Record and is nothing more than rank hearsay—*i.e.*, purported testimony about conclusions reached by RAND but being offered by Defendants' lay declarant, Ms. Goldstein. *See* Fed. R. Evid. 801(c) ("'Hearsay' means a statement that (1) the declarant does not make while testifying at the current trial or hearing; and (2) a party offers in evidence to prove the truth of the matter asserted in the statement.").

But even assuming that the RAND analysis could be relied upon, it actually supports Plaintiffs' position. Specifically, RAND allegedly made several critical findings that unequivocally demonstrate that there is unaccounted for imprecision in the calculation of Star Ratings.

First, RAND purportedly concluded that Plaintiffs' H3655 contract would have received 4 Stars in 15% of the simulated calculations. Defs.' Br. at 36; *see also* Defs.' App. 3, Goldstein Decl. ¶ 5. This alone proves Plaintiffs' point—based upon nothing more than random chance alone, H3655 would have received 4 Stars at least 15% of the time. To put a fine point on this: Plaintiffs' ability to qualify for hundreds of millions of dollars in quality bonus payments came down to pure chance—yet, Defendants do not account for that in any way in calculating Star Ratings. This is the epitome of arbitrary and capricious agency action.

Second, RAND allegedly concluded that for 21 contracts across the country, the most common resulting Star Rating resulting from RAND's simulations was *higher* than the Star Rating those contracts received. *See* Defs.' Br. at 36; Defs.' App. 3, Goldstein Decl. ¶ 4. In other words,

the RAND simulations prove that at least 21 other contracts received a lower Star Rating than the

most common Star Rating. Yet, again, Defendants simply do not account for this imprecision. The

bottom line is that billions of dollars are at stake with Medicare Advantage Star Ratings. Yet,

Defendants calculate and assign Star Ratings based upon nothing more than the roll of the dice.

Failing to account for this imprecision and leaving each contract subject to the whimsy of chance

is the very definition of arbitrary and capricious. *See Wildearth Guardians v. Bucknall*, No. CV

23-10-M-DLC, 2024 WL 4711128, at *9 (D. Mont. Nov. 7, 2024) (holding that the agency's use

of unreliable data was arbitrary and noting that arbitrary means "existing or coming about

seemingly at random or by chance.") (internal quotations omitted).

> ### c.   Defendants' argument that they "only need[] to use a precision of 0.5" is another attempt to side-step the pertinent issue.

Despite demonstrating that Defendants fail to take into account the imprecision created by

random chance alone when calculating Final Summary Scores to the sixth decimal, Defendants

attempt to argue that they "only need[] to use a precision of 0.5" because they round to the half

star. Defs.' Br. at 32–33. In other words, it appears Defendants are arguing that by allegedly

rounding at the second decimal, they do account for that imprecision.

As an initial matter, as set forth in § IV.B.i., Defendants' argument about rounding at the

hundredth decimal is not supported by a single cite to the Administrative Record because there is

no such support. Rather, as the 2025 Technical Notes make clear, Defendants are not applying

traditional rounding rules at all, but rather "converting" numbers calculated to the sixth decimal to

an Overall Star Rating. In doing so, Plaintiffs unquestionably rely upon the Final Summary Score

as calculated out to the sixth decimal. *See, e.g.,* A.R. 31, App. 101, 2025 Technical Notes ("[f]or

example, a summary or overall rating of 3.749999 rounds down to a rating of 3.5, and a rating of

3.750000 rounds up to a rating of 4. That is, a score would need to be at least halfway between 3.5

and 4 (having a minimum value of 3.750000) to obtain the higher rating of 4."). No matter how many *post hoc* rationalizations Defendants attempt to conjure up, the Administrative Record is clear as to what Defendants actually do—*i.e.*, assess whether the Final Summary Score as calculated to six decimals is "at least halfway between 3.5 and 4." *Id.*

d.   *Defendants' attempt to criticize Dr. Diver's analysis is misplaced and, in any event, irrelevant.*

Finally, Defendants attack Plaintiffs' arguments regarding the failure to account for imprecision by attempting to criticize Dr. Diver's simulation on the bases that he allegedly did not recalculate: (i) reward factor thresholds as part of his simulations, and (ii) Part C and Part D improvement measure scores and Star Ratings to account for the measure-level hold harmless provision. *See* Defs.' Br. at 34-35. But these critiques are red herrings because they rely on data that Defendants have not made publicly available. *See* Suppl. App. 323*,* Diver Suppl. Decl. ¶ 8. Indeed, the measure value data Defendants provided in response to this lawsuit were de-identified in a manner that *prevented* the re-calculation of reward factor thresholds. *See id.*

Moreover, Dr. Diver's simulations did not recalculate Part C and Part D improvement measure scores and Star Ratings to account for the measure-level hold harmless provision because doing so requires possession of the full, non-publicly available measure value data for the prior year (2024), which Defendants do not make public and did not include in the Administrative Record. *See* Suppl. App. 323, Diver Suppl. Decl. ¶ 8. Further, both years' measure value data would need to be made identifiable to reveal which measure value pertained to which contract. *See id*. But in any case, these criticisms are beside the point because they do not in any way invalidate the conclusions reached by Dr. Diver—that random chance alone has a significant impact on the Star Rating assigned to any given contract. And as explained above, Defendants' simulation

33

actually *confirms* Dr. Diver's overarching opinion. *See* Suppl. App. 323-24, Diver Suppl. Decl. ¶¶ 9–11.

> ii.   **CAHPS Measures: Defendants fail to account for the inherent imprecision in survey responses by calculating Plaintiffs' Final Summary Score to the sixth decimal.**

Plaintiffs' Opening Brief explains how Defendants fail to account for inherent imprecision in CAHPS survey data when calculating and rounding the Final Summary Score to the sixth decimal and then relying upon that purportedly precise number to determine the Overall Star Rating. *See* Pls.' Br. at 30–33. The source of that imprecision includes nonresponse bias, sampling error, and measurement error, and compromise the accuracy and reliability of the agency's Star Rating calculations. *Id.* The agency's failure to consider and address these sources of imprecision is arbitrary and capricious, especially considering that these CAHPS surveys can determine whether a plan receives quality bonus payments and rebates that improve member benefits.

Defendants do not dispute that failing to account for this survey imprecision would be arbitrary and capricious. Nor can they in good faith, given that such imprecision has the capability to significantly vary the results and lead to swings of millions of dollars. Instead, Defendants attempt to explain how they have accounted for such imprecision. In doing so, however, Defendants demonstrate they fundamentally misunderstand surveys and misstate the Administrative Record.

> a.   **Defendants failed to consider nonresponse bias (as opposed to simple nonresponse).**

Defendants do not dispute that failure to account for nonresponse bias would be arbitrary and capricious, nor could they given that the Federal Office of Management and Budget ("OMB") issued survey standards for federal agencies in 2006 that require any federal survey that does not expect to attain at least an 80% response rate to conduct an investigation into the presence of

34

nonresponse bias for that survey. App. 63–64, Lavrakas Decl. ¶ 19 (citing, App. 211–12, OFFICE OF MGMT. AND BUDGET, STANDARDS AND GUIDELINES FOR STAT. SURVS. 8 (2006)). Indeed, the Federal Committee on Statistical Methodology has described nonresponse bias as "one of the main threats to data quality in federal surveys," and in 2023 issued best practices for nonresponse bias reporting. *See id*. (citing App. 213–14, FEDERAL COMM. ON STAT. METHODOLOGY, BEST PRACTICES FOR NONRESPONSE BIAS REPORTING 3 (2023)).

Defendants completely ignore the fact that they do not comply with the survey standards issued by OMB for all federal agencies. Indeed, Defendants do not dispute that they failed to conduct any nonresponse bias investigation for the CAHPS surveys for Medicare Advantage generally, let alone for the CAHPS surveys that led to the 2025 Star Ratings consistent with these federal standards. This alone is contrary to federal requirements and arbitrary and capricious. Defendants misleadingly argue that "Dr. Lavrakas' chief criticism of CAHPS data is that the survey response rate is below 80%" and contend that "CMS increased response rates by shortening the survey and improved phone contact information." Defs.' Br. at 38. But Dr. Lavrakas makes clear that response rate is not relevant, but rather the failure to adhere to the federal survey standards of conducting a nonresponse bias analysis when response rates are expected to be below 80%—which is without question the case with CAHPS surveys. App. 63–64, Lavrakas Decl. ¶ 19.

After ignoring the federal requirements for surveys conducted by federal agencies, Defendants assert multiple arguments that are unavailing to attempt to explain how they have properly considered nonresponse bias.

First, Defendants argue that in response to comments raised in connection with the 2018 rulemaking, Defendants addressed concerns regarding low CAHPS survey response rates. Defs.' Br. at 37-38 (citing 83 Fed. Reg. at 16,555 (April 16, 2018)). Defendants argue that CMS sought

to "improve response rates," such as improving phone contact information and shortening the CAHPS survey. *See* Defs.' Br. at 37. However, Defendants demonstrate that they fundamentally misunderstand the difference between low response rates and nonresponse bias—which are entirely separate issues. Dr. Lavrakas is one of the world's most renowned survey methodology experts. As he explained, the response rate is simply the number of people who respond to a survey, and any attempts to increase response dates do not address nonresponse bias. App. 63–68, Lavrakas Decl. ¶¶ 19–26. Nonresponse bias occurs when there are persons in an initial sample who fail to respond, and the nonrespondents would have provided materially different data about what is being measured versus the data that the respondents provided. *Id.*

The Administrative Record is entirely devoid of any evidence that Defendants considered or investigated nonresponse bias in CAHPS surveys generally, let alone for the CAHPS surveys that were used to determine Plaintiffs' 2025 Star Ratings—because there is none. As explained in the Supplemental Declaration of Dr. Lavrakas, nonresponse bias investigations are the only way that survey researchers can determine if there is non-ignorable nonresponse bias that is likely to be present in the final data. *See* Suppl. App. 346-47, Supplemental Declaration of Paul Lavrakas, Ph.D. ("Lavrakas Suppl. Decl.") ¶¶ 12–13. And only through these investigations can researchers determine, with confidence, which respondent characteristics should be adjusted to reduce or eliminate nonresponse bias. *Id.* This is why OMB mandates such investigations. CMS evidently does not have a "reliable and valid method to adjust CAHPS data for nonresponse biases that may be present in the data" used to create Star Ratings. Suppl. App. 347, Lavrakas Suppl. Decl.¶ 14.

Second, Defendants cite to a single sentence in the 2018 Final Rule where CMS stated that "'research specific to patient experience, CAHPS, and MA and PDP CAHPS surveys finds no evidence nonresponse bias affects comparison of case-mix adjusted scores between contracts or

36

other similar reporting units.'" Defs.' Br. at 38 (citing 83 Fed. Reg. at 16,555 (April 16, 2018)).

However, as Dr. Lavrakas initially explained, two articles (Elliot et al., 2005; Elliot et al., 2009) briefly addressed the topic of nonresponse bias in *hospital* CAHPS data, but hospital CAHPS data relates to the surveying of hospital care and is different from the CAHPS data at issue for MAOs. *See* App. 65-66, Lavrakas Decl. ¶ 21. Furthermore, those researchers did not report carrying out any nonresponse bias investigations in a CAHPS survey because they apparently did not assess what data nonrespondents would have produced. *Id*. Accordingly, this single sentence in the 2018 Final Rule does not demonstrate that CMS considered nonresponse bias in the Medicare Advantage CAHPS survey data that led to the 2025 Star Ratings, as federally mandated.

Third, Defendants argue that they address nonresponse bias through case-mix adjustment, *see* Defs.' Br. at 38, but case-mix adjustment is similarly made to correct for nonresponse, not nonresponse bias. *See* Suppl. App. 344-46, Lavrakas Suppl. Decl. ¶¶ 8–9; *see also* App. 65–66, Lavrakas Decl. ¶ 21. In fact, the agency may be unnecessarily adjusting for respondent characteristics that are not causing biases and do not require adjustment. *See* Suppl. App. 346, Lavrakas Suppl. Decl. ¶ 11. By adjusting for unnecessary characteristics, the agency may actually decrease the accuracy of Star Ratings. *Id.*

Defendants misleadingly argue that "[i]n 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 [] CAHPS surveys.'" Defs.' Br. at 38 (citing Cynthia Rae Bland *et al*., *Challenges Facing CAHPS Surveys and Opportunities for Modernization*, RTI INT'L, 1, 9 (2022)) (emphasis added). However, contrary to Defendants' assertion, CMS did not cite that publication in the 2018 Final Rule—nor could they, since it was published in 2022. Moreover, this publication is not part of the

37

Administrative Record and there is absolutely no evidence in the Administrative Record (or in any location from what Plaintiffs can find) that CMS has ever relied upon it when conducting the CAHPS surveys generally or with respect to the 2025 Star Ratings CAHPS surveys. For instance, a search of "Bland" and "CAHPS" in the Federal Register does not yield any results citing to this publication. At bottom, this is nothing more than a wayward cite to a publication available on the internet, without any connection to Defendants relying upon it, let alone conducting the mandatory nonresponse bias investigation in the CAHPS surveys that led to the 2025 Star Ratings.

Although Defendants' citation to the RTI publication should be ignored, even if considered, it is clear that Defendants misleadingly cite to this to suggest that case-mix adjustment eliminates nonresponse bias in Medicare Advantage CAHPS surveys. It does not. Rather, that publication relates to patient-mix adjustments for healthcare facilities.[7] In addition, the publication analyzes nonresponse rates in CAHPS surveys and not nonresponse bias. While it stated in a single unsubstantiated sentence that a "after the patient-mix adjustment process is completed, the survey results no longer have nonresponse bias, which has been shown in some CAHPS surveys," *see id.*, this single fleeting sentence, which is not tied to any research or investigation, is contrary to the article's more robust discussion recognizing that weighting responses (such as case-mix adjustments) can reduce nonresponse bias but it cannot eliminate it altogether. *See* Suppl. App. 345-46, Lavrakas Suppl. Decl. ¶ 9. Finally, this single sentence does not demonstrate that Defendants actually conducted a nonresponse bias investigation as mandated by OMB or that nonresponse bias in the Medicare Advantage CAHPS surveys that led to the 2025 Star Ratings have been accounted for.

---

[7] *See* RTI International, *Challenges facing CAHPS surveys and opportunities for modernization*, https://www.rti.org/rti-press-publication/challenges-facing-cahps-surveys-opportunities-modernization (last visited April 9, 2025).

Ultimately, Defendants' arguments show that they fundamentally do not understand survey methodology and have failed to account for nonresponse bias in the 2025 CAHPS surveys, which is undisputedly arbitrary and capricious.

b.   Defendants failed to account for sampling and measurement error.

Defendants have also failed to adequately account for sampling and measurement error in calculating the Final Summary Score and Overall Star Ratings for Plaintiffs' contracts. Plaintiffs, relying on the declaration of Dr. Lavrakas, initially explained that confidence intervals represent the range of uncertainty that exists in a survey and are routinely reported in survey-related statistical analyses. *See* Pls.' Br. at 32. Although Defendants concede that "all sample surveys have sampling error," *see* Defs.' Br. at 40, Defendants purport that "CMS calculated its final summary score and overall ratings using methods that are *more* accurate and unbiased than confidence intervals" and therefore "CMS does not use confidence intervals." Defs.' Br. at 39–40 (emphasis in original). Focusing exclusively on the declaration of Ms. Goldstein, Defendants claim that CMS relies only on the mean, "[t]he most accurate, unbiased measure of a contract's performance . . . ." Defs.' Br. at 40. Defendants' position demonstrates an apparent lack of understanding of confidence intervals and their use in statistical analysis of survey data—including CAHPS data— and shows why the agency's calculation of Plaintiff's Star Ratings is arbitrary and capricious.

As a preliminary matter, Defendants' "expert" declaration deserves no weight to refute Dr. Lavrakas's opinions. As discussed *supra*, Ms. Goldstein is the Director of the Division of Consumer Assessment and Plan Performance, Medicare Drug Benefit and C&D Data Group at CMS. *See* Defs.' App. 1. Unlike the declaration of Dr. Lavrakas, Ms. Goldstein's declaration does not discuss any of her qualifications or expertise in survey research methodology or analysis. Ms. Goldstein's declaration also fails to cite to any portions of the Administrative Record on which

39

she relies, if any, or otherwise disclose any pertinent literature or evidence to support her opinions. The Court should not credit Ms. Goldstein's bare, unsupported opinions.

In any event, Defendants' claim that CMS's reliance on the actual mean without consideration of confidence intervals is somehow a superior (*i.e.*, more accurate and unbiased) way to compare contracts' performances does not make sense. As Dr. Lavrakas explains, confidence intervals are central to understanding sampling error and survey accuracy such that there are no post-survey adjustments—like case-mix adjustments—that eliminate the need to consider confidence intervals in analyzing survey findings. *See* Suppl. App. 347-48, Lavrakas Suppl. Decl. ¶¶ 15–17. Therefore, CMS's claim that use of weighted, case-mix adjusted means somehow improves the comparability of scores and properly addresses sampling error is wrong because the size of the margin of sampling error and resulting confidence intervals of CAHPS surveys are such "that they make invalid any comparisons of means…" *See id.* ¶ 17. Indeed, "it is statistically incorrect to draw conclusions about survey-based differences" in the way CMS has done for Plaintiffs' plans. *See id.*

Defendants further claim that "CMS is sufficiently transparent" regarding information about its sampling design because CMS describes "CAHPS sample selection and eligibility criteria and sample preparation." *See* Defs.' Br. at 41. But Defendants do not refute that CMS has not made available any technical details and key information about the CAHPS sampling design to allow Plaintiffs (or the public) to determine the precision of CAHPS data. *See* Pls.' Br. at 32; *see also* Suppl. App. 343, Lavrakas Suppl. Decl. ¶ 5 (CMS does not report anything about the "design effects," also known as *deffs*, or other CAHPS sampling design metrics such that it would be impossible for an observer to know the precision (*i.e.*, the amount of error in the form of imprecision/variance) associated with a given CAHPS survey).

Finally, Defendants fail to show through the Administrative Record that the agency accounts for question reliability and measurement error in CAHPS survey data. *See* Pls.' Br. at 33. CAHPS survey data—including data for the questionnaire items used in calculating Star Ratings— "are imperfect in regards to their accuracy (i.e., their validity and reliability)." *See* Suppl. App. 348, Lavrakas Suppl. Decl. ¶ 18. Even though many CAHPS measures may have strong reliability, all are imperfect and therefore their usage to calculate Star Ratings result in Star Ratings that contain errors that are larger than one millionth of a decimal place. *See* Suppl. App. 349, Lavrakas Suppl. Decl. ¶¶ 19–20. Defendants do not respond directly to this claim, and generally rely on CMS's Quality Assurance Protocols & Technical Specifications to argue that CMS "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." *See* Defs.' Br. at 42. But none of these facts about the CAHPS survey itself demonstrate whether (or how) the agency addresses the inherent imprecision and measurement errors in CAHPS survey data, which are used to calculate Star Ratings.

The bottom line is that when Defendants calculated Plaintiffs' Star Ratings out to the sixth decimal and then determined the Final Summary Score and corresponding Overall Star Rating on that purported level of precision, they acted arbitrary and capriciously because there is fundamental and inherent imprecision in CAHPS survey data that Defendants have not accounted for.

V.   DEFENDANTS' USE OF THE "JENNY" SEED YEAR AFTER YEAR VIOLATES THE REGULATORY MANDATE TO ENSURE RANDOMIZATION.

Defendants correctly explain that mean resampling is a technique where measure-specific scores for the current year's Star Ratings are *randomly* separated into groups to determine the cut points between 1–5 whole star ratings for all non-CAHPS measures. *See* Defs.' Br. at 43; 42 C.F.R.

§ 422.162(a) (emphasis added). The "seed" is a number that controls the starting point of the random sequence of numbers needed to perform mean resampling and allows for future replication of the randomization process. *See* A.R. 153, App. 181, 2025 Technical Notes. Defendants' chosen technique for mean resampling and cut point determination, however—namely the use of the "Jenny" seed year after year—requires the fortuitous alignment of outside factors to ensure the randomness required by the regulation. Randomness is not ensured by their methodology alone. Indeed, in order to ensure randomness as required, Defendants admit that their chosen technique relies upon outside factors over which they have no control. *See* Defs.' Br. at 43.

A.   **Defendants Erroneously Claim That Their Repeated Use of the "Jenny" Seed is a Statistical "Best Practice."**

Defendants concede that in their process, the seed is used to randomly sort the contracts into groups at the beginning of the mean resampling process. *Id*. Defendants claim, without citation to a single authoritative source, that it is "the *best practice* . . . to use a number that is well known, like a number from pop culture, hence, the reference to the Tommy Tutone song, 'Jenny.'" *Id*. (emphasis added). This assertion, while absurd on its face, is simply not supported by generally accepted statistical principles. Suppl. App. 340–41, Diver Suppl. Decl. ¶¶ 43–44. Defendants' only attempt to support this baseless assertion is a complete mischaracterization of the Declaration of Dr. Paul Diver. Defendants claim that "a memorable seed is necessary to replicate results of a randomization process," and cite directly to Dr. Diver's Declaration for that assertion. *See* Defs.' Br. at 43. But Dr. Diver says no such thing.

Instead, consistent with CMS's own Technical Notes, Dr. Diver instructs that "[s]pecifying a seed value allows for future replication of a randomization process." App. 54, Diver Decl. ¶ 87; Defs.' Br. at 43. He says nothing about whether the seed needs to be "memorable" because the seed does not need to be "memorable." Suppl. App. 340–41, Diver Suppl. Decl. ¶ 43. The mean

42

resampling process is documented by Defendants in the yearly Technical Notes, and the seed is recorded in that documentation for future use. *See* A.R. 153, App. 181, 2025 Technical Notes. Nobody needs to remember the seed from year to year, because it is written down.

Defendants have published generally accepted statistical practices for their contractors to use in other statistical applications on behalf of Defendants, such as creating a sample of claims for auditing or extrapolating an overpayment from a sample to a universe of claims. *See* Suppl. App. 315–16, CTRS. FOR MEDICARE & MEDICAID SERVS., Pub. 100-08, MEDICARE PROGRAM INTEGRITY MANUAL, Ch. 8 (2024) ("MPIM"). Although not directly applicable here, the MPIM is nonetheless instructive regarding what Defendants have determined are generally accepted statistical practices. The MPIM clearly requires Medicare contractors to document all steps of a statistical process so that it can be replicated—including the "starting point" or "known seed value." Suppl. App. 316, MPIM § 8.4.4.2 (emphasis added). The MPIM does not require that the seed be "memorable." *See* Suppl. App. 340-341, Diver Suppl. Decl. ¶¶ 43-44.

B.   Defendants Concede That Their Mean Resampling Methodology Does Not Ensure Randomness, as Required by the Regulation.

Defendants agree with Dr. Diver that "if the list of consolidated 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." Defs.' Br. at 43; App. 54. In other words, Defendants have conceded that their methodology does not, on its own, ensure randomness that is required by the regulation. Defs.' Br. at 44.[8]

---

[8] Moreover, the only thing Defendants have established is that their non-random process harms even more contracts than just Plaintiffs, and this random error was statistically significant for Elevance's contract H3655. *See* Defs.' App. 3, Goldstein Decl. ¶¶ 4–5 (conceding 21 contracts would have received a higher score by random chance alone, 45 contracts could have received different scores, either higher or lower, and Elevance's contract H3655 would receive a higher score 15% of the time). And yet again Defendants refuse to provide any showing in the

Instead, to ensure randomness, Defendants rely on the assumption that the list of contracts

for which mean resampling is performed does not remain the same from year to year. While it is

true that contracts likely change from year to year, this is not guaranteed. Despite the possibility,

however slight, that contracts may remain the same from one year to the next, Defendants cite to

their assumption as though it were an immutable fact, relying only on Ms. Goldstein's declaration

for support. *Id.*; Defs.' App. 2. Variability introduced from outside actors and external factors

should not be relied upon to guarantee the required randomness. Accordingly, Defendants acted

arbitrarily by setting up a methodology that does not guarantee randomness on its face when the

regulation requires randomness, especially because all Defendants have to do to rectify this issue

is to specify and record a new seed each year.

> **C.   Defendants Lack Authoritative Support, Other Than Dr. Diver's Report, Which is Fatal to Their Position.**

Defendants have not contested that Dr. Diver is an expert in statistical methods, and he is

the only statistical expert who has offered an opinion in this litigation. He holds a Ph.D. in Statistics

from the University of Virginia, a Master's Degree in Economics from the University of Virginia,

a Master's Degree in Mathematics and Statistics from Georgetown, a Bachelor's Degree in

Mathematics from Georgetown, and has worked evaluating and applying statistical methodologies

for over 15 years. *See* App. 27. His opinions come from a place of deep experience and high-level

education.

Defendants have not offered a competing statistical expert to support their positions, they

have not included any authoritative evidence in the Administrative Record to support their

reasoning, and Defendants do not claim that Ms. Goldstein—despite her 25 years at CMS—is a

---

Administrative Record that they have considered the random implications and effects they will
have on MAOs—neglecting their obligations to engage in reasoned decision-making. *See
generally State Farm Mut. Auto. Ins. Co.,* 463 U.S. 29 (1983).

statistical expert. In her Declaration—the only declaration proffered by Defendants—Ms. Goldstein similarly fails to cite to any expert or authoritative sources, or expertise of her own, for her assertions. Instead, she baldly states her opinions without citation or support and expects this Court to simply accept her assertions as true.

Indeed, Defendants have included no authoritative support for their decision to use the Jenny seed year after year in the Administrative Record, which makes it impossible for the Court to determine whether Defendants' choice is the rational outcome of the consideration of all relevant factors. *Lloyd v. Illinois Reg'l Transp. Auth.*, 548 F. Supp. 575, 590 (N.D. Ill. 1982) ("In assessing whether the agency's action was arbitrary and capricious, the court must test the action against the full administrative record that was before the [agency official].") (internal citation omitted).

Ultimately, Defendants are left to provide nothing but *post hoc* rationalizations to justify their Medicare Advantage Star Ratings. They cannot demonstrate through the Administrative Record that their methodology fulfills their regulatory requirement of randomness. Left with nothing else to defend the agency's approach, Defendants' use of *post* hoc rationalizations must be rejected by this court. *See Luminant Generation Co., LLC*, 675 F.3d at 925 (where agency fails to support its action through the administrative record, courts must disregard any *post hoc* rationalizations for agency action); *Nat'l Ass'n of Mfrs.*, 105 F.4th at 814 ("[I]n reviewing an agency's action, we may consider only the reasoning articulated by the agency itself; we cannot consider *post hoc* rationalizations.") (internal quotations omitted).

Accordingly, Defendants' use of the "Jenny" seed every year was contrary to law and, by definition, arbitrary and capricious.

45

CONCLUSION

Defendants have acted contrary to law and arbitrarily and capriciously for the reasons set

forth herein. Therefore, Plaintiffs are entitled to judgment on their claims.

Dated: April 11, 2025   Respectfully submitted,

                                         ELEVANCE HEALTH, INC. and the
                                         HEALTH PLAN PLAINTIFFS

                                         By: */s/ Lesley R. Reynolds*

                                         Lesley C. Reynolds (N.D. Tx. Bar No. 487580DC)
                                         Lara E. Parkin (*pro hac vice*) (D.C. Bar No. 475974)
                                         David A. Bender (*pro hac vice*) (D.C. Bar No. 1030503)
                                         REED SMITH LLP
                                         1301 K Street, N.W.
                                         Suite 1000 – East Tower
                                         Washington, D.C. 20005
                                         Ph: (202) 414-9200
                                         Fax: (202) 414-9299
                                         lreynolds@reedsmith.com
                                         lparkin@reedsmith.com
                                         dbender@reedsmith.com

                                         Martin J. Bishop (*pro hac vice*) (Tx. Bar No. 24086915)
                                         Steven D. Hamilton (*pro hac vice*) (Ill. Bar No. 6289663)
                                         REED SMITH LLP
                                         10 South Wacker Drive
                                         40th Floor
                                         Chicago, IL 60606
                                         Ph: (312) 207-1000
                                         Fax: (312) 207-6400
                                         mbishop@reedsmith.com
                                         shamilton@reedsmith.com

                                         Scott T. Williams (Tx. Bar No. 00791937)
                                         REED SMITH LLP
                                         2850 N. Hardwood Street
                                         Suite 1500
                                         Dallas, TX 75201
                                         Ph: (469) 680-2200
                                         Fax: (469) 680-4299
                                         scott.williams@reedsmith.com

Meredith W. Knudsen (Tx. Bar No. 24088617)
Klayton S. Hiland (Tx. Bar No. 24116516)
KELLY HART & HALLMAN LLP
201 Main Street, Suite 2500
Fort Worth, Texas 76102
Ph: (817) 332-2500
Fax: (817) 878-9280
meredith.knudsen@kellyhart.com
klayton.hiland@kellyhart.com

Counsel for Plaintiffs

CERTIFICATE OF SERVICE

I hereby certify that on April 11, 2025, I electronically filed the foregoing document with

the Clerk of the Court using the CM/ECF system, which will send notification of this filing to the

attorneys of record and all registered participants.

/s/ Meredith W. Knudsen
Meredith W. Knudsen

47

UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS

ELEVANCE HEALTH, INC., et al., Plaintiffs, v. **ROBERT F. KENNEDY, JR.,** in his official capacity as Secretary of Health and Human Services, U.S. Department of Health and Human Services, and **MEHMET OZ**, in his official capacity as Administrator, Centers for Medicare and Medicaid Services, Defendants.	Case No. 4:24-cv-01064 **ORAL HEARING REQUESTED**

**SUPPLEMENTAL APPENDIX IN SUPPORT OF PLAINTIFFS' RESPONSE TO
DEFENDANTS' CROSS-MOTION FOR SUMMARY JUDGMENT AND
REPLY IN SUPPORT OF PLAINTIFFS' MOTION FOR SUMMARY JUDGMENT**

Contents

<u>Document Titles</u>	<u>Page Numbers</u>
A.R. 23-24 Ctr. For Medicare and Medicaid Services, Medicare 2025 Part C & D Star Ratings Technical Notes ("2025 Technical Notes") (*Specific Excerpts Only*)	312-313
A.R. 838	314

CTRS. FOR MEDICARE & MEDICAID SERVS., Pub. 100-08, MEDICARE PROGRAM INTEGRITY MANUAL, Ch. 8 (2024) ("MPIM"). (*Specific Excerpts Only*)	315-316
Microsoft Support, *FIXED function*, https://support.microsoft.com/en-us/office/fixed-function-ffd5723c-324c-45e9-8b96-e41be2a8274a (last visited April 9, 2025).	317-318
Supplemental Declaration of Paul Diver, Ph.D.	319-341
Supplemental Declaration of Paul J. Lavrakas, Ph.D.	342-349

Dated: April 11, 2025

Respectfully submitted,

**ELEVANCE HEALTH, INC. and the
HEALTH PLAN PLAINTIFFS**

By: */s/ Lesley C. Reynolds*

Lesley C. Reynolds (N.D. Tx. Bar No. 487580DC)
Lara E. Parkin (*pro hac vice*) (D.C. Bar No. 475974)
David A. Bender (*pro hac vice*) (D.C. Bar No. 1030503)
REED SMITH LLP
1301 K Street, N.W.
Suite 1000 – East Tower
Washington, D.C. 20005
Ph: (202) 414-9200
Fax: (202) 414-9299
lreynolds@reedsmith.com
lparkin@reedsmith.com
dbender@reedsmith.com

Martin J. Bishop (*pro hac vice*) (Tx. Bar No. 24086915)
Steven D. Hamilton (*pro hac vice*) (Ill. Bar No. 6289663)
REED SMITH LLP
10 South Wacker Drive
40th Floor
Chicago, IL 60606
Ph: (312) 207-1000
Fax: (312) 207-6400
mbishop@reedsmith.com
shamilton@reedsmith.com

Scott T. Williams (Tx. Bar No. 00791937)
REED SMITH LLP
2850 N. Hardwood Street
Suite 1500
Dallas, TX 75201
Ph: (469) 680-2200
Fax: (469) 680-4299
scott.williams@reedsmith.com

Meredith W. Knudsen (Tx. Bar No. 24088617)
Klayton S. Hiland (Tx. Bar No. 24116516)
KELLY HART & HALLMAN LLP
201 Main Street, Suite 2500
Fort Worth, Texas 76102
Ph: (817) 332-2500

Fax: (817) 878-9280
meredith.knudsen@kellyhart.com
klayton.hiland@kellyhart.com

Counsel for Plaintiffs

CERTIFICATE OF SERVICE

I hereby certify that on April 11, 2025, I electronically filed the foregoing document with

the Clerk of the Court using the CM/ECF system, which will send notification of this filing to the

attorneys of record and all registered participants.

/s/ *Meredith W. Knudsen*
Meredith W. Knudsen

- Subtract the summary or overall star from each performance measure's star; square the results; and multiply each squared result by the corresponding individual performance measure weight.
- Sum these results; call this 'SUMWX.'
- Set n equal to the number of individual performance measures available for the given contract.
- Set W equal to the sum of the weights assigned to the n individual performance measures available for the given contract.
- The weighted variance for the given contract is calculated as: n * SUMWX / (W * (n-1)). For the complete formula, please see Attachment H: Calculation of Weighted Star Rating and Variance Estimates.

- Categorize the variance into three categories:
  - low (0 to < 30th percentile),
  - medium (≥ 30th to < 70th percentile) and
  - high (≥ 70th percentile)

- Develop the reward factor as follows:
  - r-Factor = 0.4 (for contract w/ low variance & high mean (mean ≥ 85th percentile))
  - r-Factor = 0.3 (for contract w/ medium variance & high mean (mean ≥ 85th percentile))
  - r-Factor = 0.2 (for contract w/ low variance & relatively high mean (mean ≥ 65th & < 85th percentile))
  - r-Factor = 0.1 (for contract w/ medium variance & relatively high mean (mean ≥ 65th & < 85th percentile))
  - r-Factor = 0.0 (for all other contracts)

Tables 8 and 9 show the final threshold values used in reward factor calculations for the 2025 Star Ratings.

Table 8: Performance Summary Thresholds

Improvement	Percentile	Part C Rating	Part D Rating (MA-PD)	Part D Rating (PDP)	Overall Rating
With	65th	3.703125	3.666667	3.535714	3.646465
With	85th	4.014493	4.000000	4.035714	3.949495
Without	65th	3.707692	3.718750	3.687500	3.662921
Without	85th	4.044118	4.062500	4.173913	3.977528

Table 9: Variance Thresholds

Improvement	Percentile	Part C Rating	Part D Rating (MA-PD)	Part D Rating (PDP)	Overall Rating
With	30th	0.820452	0.742679	0.847865	0.828220
With	70th	1.275376	1.268610	1.533170	1.240423
Without	30th	0.807024	0.654297	0.717578	0.795388
Without	70th	1.256410	1.210645	1.508203	1.216635

Categorical Adjustment Index (CAI)

CMS has implemented an analytical adjustment called the Categorical Adjustment Index (CAI). The CAI is a factor that is added to or subtracted from a contract's Overall and/or Summary Star Ratings to adjust for the

Elevance Health, Inc., et al. v. Becerra, et al., No. 4:24-cv-01064-P (N.D. Tex.)
A.R.23

App. 312

average within-contract disparity in performance for Low Income Subsidy/Dual Eligible (LIS/DE) beneficiaries and disabled beneficiaries. The CAI value (factor) depends on the contract's percentage of beneficiaries with LIS/DE and the contract's percentage of beneficiaries with disabled status. These adjustments are performed both with and without the improvement measures included. The value of the CAI varies by the contract's percentage of beneficiaries with LIS/DE and disability status.

The CAI values use data collected for the 2024 Star Ratings. To calculate the CAI, case-mix adjustment is applied to all clinical Star Rating measure scores that are not adjusted for SES using a beneficiary-level logistic regression model with contract fixed effects and beneficiary-level indicators of LIS/DE and disability status, similar to the approach currently used to adjust CAHPS patient experience measures. However, unlike CAHPS case-mix adjustment, the only adjusters are LIS/DE and disability status. Adjusted measure scores are then converted to measure stars using the 2024 rating year measure cutoffs and used to calculate Adjusted Overall and Summary Star Ratings. Unadjusted Overall and Summary Star Ratings are also determined for each contract.

The 2024 measures used in the 2025 CAI adjustment calculations are:

- Breast Cancer Screening (Part C)
- Colorectal Cancer Screening (Part C)
- Annual Flu Vaccine (Part C)
- Monitoring Physical Activity (Part C)
- Osteoporosis Management in Women who had a Fracture (Part C)
- Diabetes Care – Eye Exam (Part C)
- Diabetes Care – Blood Sugar Controlled (Part C)
- Controlling Blood Pressure (Part C)
- Reducing the Risk of Falling (Part C)
- Improving Bladder Control (Part C)
- Medication Reconciliation Post-Discharge (Part C)
- Plan All-Cause Readmissions (Part C)
- Statin Therapy for Patients with Cardiovascular Disease (Part C)
- Transitions of Care (Part C)
- Follow-up after Emergency Department Visit for People with Multiple High-Risk Chronic Conditions (Part C)
- Medication Adherence for Diabetes Medication (Part D)
- Medication Adherence for Hypertension (RAS antagonists) (Part D)
- Medication Adherence for Cholesterol (Statins) (Part D)
- MTM Program Completion Rate for CMR (Part D)
- Statin Use in Patients with Diabetes (SUPD) (Part D)

To determine the value of the CAI, contracts are first divided into an initial set of categories based on the combination of a contract's LIS/DE and disability percentages. For the adjustment for the overall and summary ratings for MA-Only and MA-PD contracts, the initial groups are formed by the ten groups of LIS/DE and quintiles of disability, thus resulting in 50 initial categories. For PDPs, the initial groups are formed using quartiles for both LIS/DE and disability. The mean differences between the Adjusted Overall or Summary Star Rating and the corresponding Unadjusted Star Rating for contracts in each initial category are determined and examined. The initial categories are collapsed to form final adjustment groups. The CAI values are the mean differences between the Adjusted Overall or Summary Star Rating and the corresponding Unadjusted Star Rating for contracts within each final adjustment group. Separate CAI values are computed for the overall and summary

Native File Attachment to AR829

"H3655_2025_SR_Calculations_2024_1 0_16"

Medicare Program Integrity Manual
Chapter 8 – Administrative Actions and Sanctions and Statistical Sampling for Overpayment Estimation

Table of Contents
(Rev. 12515, Issued: 02-22-24)

Transmittals for Chapter 8

environment, must be designed so that overpayment amounts can be accurately reflected on the provider's cost report. Therefore, sampling units must coincide with an estimation methodology designed specifically for that type of provider/supplier to ensure that the results can be placed at the appropriate points on the cost report. The sample may be either claim-based or composed of specific line items. For example, home health cost reports are determined in units of "visits" for disciplines 1 through 6 and "lower of costs or charges" for drugs, supplies, etc. If claims are paid under cost report, the services reviewed and how those units link to the provider/supplier's cost report must be known. The contractor shall follow the instructions contained in section 8.4 et seq., but use the projection methodologies provided in Pub. 100-08, Exhibits 9 through 12, for the appropriate provider type. Pub. 100-08, Exhibits 9 through 12, are to be used only for claims not paid under PPS.

8.4.3.2.3 - The Sampling Frame
(Rev. 11962; Issued: 04-21-23; Effective: 05-22-23; Implementation: 05-22-23)

The sampling frame is the set of all the possible sampling units from which the sample is selected. As examples, the frame may be a list of all beneficiaries receiving items from a selected supplier, a list of all claims for which fully or partially favorable determinations have been issued, or a list of all the line items for specific items or services for which fully or partially favorable determinations have been issued.
.
8.4.4 - Sample Selection
(Rev. 377, Issued: 05-27-11, Effective: 06-28-11, Implementation: 06-28-11)

8.4.4.1 - Sampling Methodology
(Rev. 906; Issued: 09-26-19; Effective: 01-02-19; Implementation: 01-02-19)

The contractor shall identify the sampling methodology to be followed. There are various ways a probability sample can be generated. The most appropriate method will depend on factors such as, the target universe and the resources available for sampling. (Refer to section 8.4.1.5 of this chapter regarding consultations with a statistical expert to determine the appropriate methodology.)

8.4.4.2 - Random Number Selection
(Rev. 906; Issued: 09-26-19; Effective: 01-02-19; Implementation: 01-02-19)

The contractor shall identify the source of the random numbers used to select the individual sampling units, if used. The contractor shall also document the program and its algorithm or table, when available, that is used; this documentation becomes part of the record of the sampling and must be available for review. The contractor shall document any starting point if using a random number table or drawing a systematic sample. In addition, the contractor shall document the known seed value if a computer algorithm is used. The contractor shall document all steps taken in the random selection process exactly as done to ensure that the necessary information is available for anyone attempting to replicate the sample selection.

Microsoft | Support   Microsoft 365   Office   Products ⌄   Devices ⌄   Account & billing ⌄   More ⌄   [Buy Microsoft 365]   All Microsoft ⌄   🔍   👤

FIXED function

▸ *Applies To*

Effortlessly highlight, filter, and sort data with Copilot in Excel   ✕

We're giving you a free trial of Copilot Pro.

[Activate now ›]

This article describes the formula syntax and usage of the **FIXED** function in Microsoft Excel.

Description

Rounds a number to the specified number of decimals, formats the number in decimal format using a period and commas, and returns the result as text.

Syntax

FIXED(number, [decimals], [no_commas])

The FIXED function syntax has the following arguments:

- **Number**   Required. The number you want to round and convert to text.

- **Decimals**   Optional. The number of digits to the right of the decimal point.

- **No_commas**   Optional. A logical value that, if TRUE, prevents FIXED from including commas in the returned text.

Remarks

- Numbers in Microsoft Excel can never have more than 15 significant digits, but decimals can be as large as 127.

- If decimals is negative, number is rounded to the left of the decimal point.

- If you omit decimals, it is assumed to be 2.

- If no_commas is FALSE or omitted, then the returned text includes commas as usual.

- The major difference between formatting a cell containing a number by using a command (On the **Home** tab, in the **Number** group, click the arrow next to **Number**, and then click **Number**.) and formatting a number directly with the FIXED function is that FIXED converts its result to text. A number formatted with the **Cells** command is still a number.

Example

Copy the example data in the following table, and paste it in cell A1 of a new Excel worksheet. For formulas to show results, select them, press F2, and then press Enter. If you need to, you can adjust the column widths to see all the data.

Data

App. 317

1 / 3

4/10/2025 1:24:40 PM

-1234.567

44.332

Formula	Description	Result
=FIXED(A2, 1)	Rounds the number in A2 one digit to the right of the decimal point.	1,234.6
=FIXED(A2, -1)	Rounds the number in A2 one digit to the left of the decimal point.	1,230
=FIXED(A3, -1, TRUE)	Rounds the number in A3 one digit to the left of the decimal point, without commas (the TRUE argument).	-1230
=FIXED(A4)	Rounds the number in A4 two digits to the left of the decimal point.	44.33

Need more help?

How can we help you? →

Want more options?

⊕ **Discover**   ⊘ **Community**
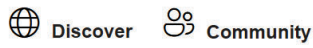
Explore subscription benefits, browse training courses, learn how to secure your device, and more.

Microsoft 365 subscription benefits

Microsoft 365 training

Microsoft security

Accessibility center

Was this information helpful?   Yes   No

What's new	Microsoft Store	Education	Business	Developer & IT	Company
Surface Pro	Account profile	Microsoft in education	Microsoft Cloud	Azure	Careers
Surface Laptop	Download Center	Devices for education	Microsoft Security	Microsoft Developer	About Microsoft
Surface Laptop Studio 2	Microsoft Store support	Microsoft Teams for Education	Dynamics 365	Microsoft Learn	Company news
Surface Laptop Go 3	Returns	Microsoft 365 Education	Microsoft 365	Explore ISV Success	Privacy at Microsoft
Microsoft Copilot	Order tracking	How to buy for your school	Microsoft Power Platform	Microsoft Tech Community	Investors
AI in Windows	Certified Refurbished	Educator training and development	Microsoft Teams	Azure Marketplace	Diversity and inclusion
Explore Microsoft products	Microsoft Store Promise		Microsoft 365 Copilot	AppSource	Accessibility

App. 318

2 / 3

4/10/2025 1:24:40 PM

UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS

ELEVANCE HEALTH, INC., et al., Plaintiffs, v. **ROBERT F. KENNEDY, JR.**, in his official capacity as Secretary of Health and Human Services, U.S. Department of Health and Human Services, and **MEHMET OZ**, in his official capacity as Administrator, Centers for Medicare and Medicaid Services, Defendants.	Case No. 4:24-cv-01064 Hon. Mark T. Pittman

SUPPLEMENTAL DECLARATION OF

PAUL DIVER, PH.D.

April 11, 2025

∴ BRG

App. 319

Table of Contents

1   Introduction and Scope

I, Paul Diver, Ph.D., declare the following to be true and correct:

1.   I am over twenty-one years of age, of sound mind, and fully competent to make this declaration.

2.   I am a Director with Berkeley Research Group, LLC ("BRG"). I was retained by Reed Smith LLP ("Counsel") on behalf of Elevance Health Inc. ("Elevance Health") and its affiliated entities ("Elevance Health") to provide my statistical opinions on certain aspects of the Centers for Medicare and Medicaid Services' ("CMS") calculation of the 2025 Medicare Advantage Star Ratings ("Star Ratings").

3.   I previously submitted a declaration (the "Diver Declaration") in this matter on February 14th, 2025, in which I have included relevant background, my qualifications, and my opinions.[1]

4.   I have reviewed Defendants' Consolidated Brief in Support of Their Response to Plaintiffs' Motion for Summary Judgment and Cross-Motion for Summary Judgment ("Defendants' Consolidated Brief") [2] as well as the Declaration of Elizabeth Goldstein ("Goldstein Declaration") submitted in support of Defendants' summary-judgment briefing.[3] This report provides a reply and rebuttal.

2   Summary of Opinions

5.   Based on my review of the Defendants' Consolidated Brief and the Goldstein Declaration, it is my statistical opinion that:

> a.   CMS's own simulation analysis fails to refute – and indeed confirms – one of the overarching findings of the Diver Declaration simulations: Overall Star Ratings are subject to fluctuation due to random chance alone. CMS did not dispute this but rather provides evidence in support of this fact. Its

---

[1] *See* App. 21-54, Declaration of Paul Diver, Ph.D.; App. 237-241, Curriculum Vitae of Paul Diver, Ph.D.

[2] *See* ECF No. 38, Defendants' Brief ("Defs.' Br.").

[3] *See* ECF No. 39, Defendants' Appendix ("Defs' App.") 1-4.

criticisms of the Diver Declaration simulations do not change this fact, and moreover can all be traced back to the fact that CMS did not make public the data necessary to address the issues it raises.

b. CMS is incorrect when it states that it "does not round to the sixth decimal." The scorecards used to calculate the Final Summary Scores, made available in the Administrative Record,[4] make use of the "FIXED" Excel function to round an intermediate weighted average of measure star ratings – and thus all subsequent scores - to the sixth decimal places.

c. CMS's discussion of precision is inherently flawed. It confuses the difference between fixed thresholds (e.g., 3.250000 and 3.750000) with the (im)precision of estimates subject to random chance. CMS's analysis is misleading as it conflates these two concepts.

d. Contrary to CMS's assertion, using a well-known seed – "like a number from pop culture"[5] – is not a best practice requirement for selecting an initialization seed.

3 CMS's Own Simulation Analysis Fails to Refute – and Indeed Confirms – the General Findings of the Diver Declaration

6.   At a high level, one of the overarching themes offered through the Diver Declaration is that a Medicare Advantage Organization's Final Summary Score and therefore Overall Star Rating are subject to fluctuation due to random chance alone. CMS does not refute this opinion but rather the Defendants' Consolidated Brief instead confirms this to be the case.

7.   To assess if random chance alone could cause fluctuations in a contract's Final Summary Score and Overall Star Rating, I performed a simulation exercise where I allowed for different initialization seeds (i.e., those other than CMS's 8-6-7-5-3-0-9 "Jenny Seed"). In the Defendants'

---

[4] *See for example* A.R. 838, Supp. App. 314, Scorecard for Elevance Contract H3655.

[5] *See* Defs.' Br. at 43.

Consolidated Brief, CMS suggests that the simulations summarized in the Diver Declaration are flawed for the following two reasons:

(1)   The Diver Declaration simulations did not recalculate reward factor thresholds as part of his simulations;[6] and

(2)   The Diver Declaration simulations did not recalculate Part C and Part D improvement measure scores and Star Ratings to account for the measure-level hold harmless provision.[7]

8.   CMS's discussion of what it believes to be "errors" in the Diver Declaration simulations are a complete red herring for several reasons. Principally, the so-called "flaws" posited by CMS are due to the fact that the necessary data were not made available by CMS. Indeed, as to CMS's first criticism, as noted in the Diver Declaration, the measure value data provided by CMS in response to this suit were de-identified in such a manner which *prevented* the re-calculation of the reward factor thresholds. As to CMS's second criticism, that the Diver Declaration simulations did not recalculate Part C and Part D improvement measure scores and Star Ratings to account for the measure-level hold harmless provision is also a result of data not made available by CMS. To account for the hold harmless provision depends on possessing the full, non-publicly available measure value data for the prior year, 2024 – which CMS also did not provide. Further, both years measure value data would need to be made identifiable to reveal which measure value pertained to which contract.

9.   But in any case, CMS **offers its own simulation** analysis, which purports to correct for the criticisms CMS levies at my own simulations, and the findings reveal that, indeed, the Final Summary Score and subsequent Final Overall Star Rating **are subject to fluctuation due to random chance alone.** In other words, CMS's simulation actually confirms the overarching opinion that I offered in my original declaration.

---

[6] *See* Defs.' Br. at 34.

[7] *See* Defs.' Br. at 35.

10.   Moreover, CMS's simulation shows that 21 contracts received a 2025 Overall Star Rating **lower** than their respective "most common overall Star rating."[8]

11.   Critically, CMS concedes that for **15%** of its own simulation runs, Elevance's H3655 contract received a higher Final Overall Star Rating of a 4.0. This is notable because CMS's own 15% finding exceeds the commonly used 5% or 10% statistical significance thresholds used in statistical hypothesis tests when evaluating how likely a scenario is to occur due to random chance alone.

12.   The Diver Declaration states that the variability in the Final Summary Score due to random chance is generally on the order of 0.01. Despite running its own simulation analysis, CMS does not provide in response its own counter-measurement of variability in Final Summary Score due to random chance. CMS is notably quiet on this front.

13.   In short, CMS's criticisms of the simulations in the Diver Declaration are moot because (1) the data were not available from CMS to proactively address them, and (2) because CMS offers its own simulation results which confirm the overarching finding of the Diver Declaration – i.e., that Final Summary Scores and Overall Star Ratings are subject to change due to random chance alone.

4 CMS Is Incorrect When It States that It "Does Not Round to the Sixth Decimal"

14.   In Defendants' Consolidated Brief, CMS states that it did not round Elevance's overall numerical scores to the sixth decimal place.[9] CMS goes on to state that "[t]ellingly, Elevance does not point to anything that supports its assertion that CMS rounds to six digits of precision." However, the information within the record shows that CMS is incorrect in its statement that it does not round to the sixth decimal.

---

[8] Defs.' Br. at 35-36. Note CMS's use of the phrase "most common"; CMS is further acknowledging that there can be multiple Star Ratings due to fluctuations arising from random chance alone. Additionally, CMS tellingly also does not provide a more general discussion on how many contracts in their simulations are associated with multiple Star Ratings due to random chance. Instead, it limits its discussion just to the number of contracts whose *most common* Star Rating is different than its published one.

[9] Defs.' Br. at 27-28.

15.   As noted in the Diver Declaration, Elevance's overall numerical score is derived by calculating a weighted average of measure specific star ratings and then adjusting that average with a reward factor and a CAI value. For example, CMS calculates a weighted average for Elevance Contract H3655 (i.e., 3.647059) to which it adds a Reward Factor of 0.1 and a CAI Value of 0.002506 to derive a Finally Summary Score of 3.749565.

16.   CMS calculates the weighted average of measure specific star ratings by dividing the sum of weighted measure star ratings by the sum of their associated weights **and then rounding the resulting ratio value to the sixth decimal place**. CMS does this through the use of the "FIXED" function in Microsoft Excel as evidenced by the contract "Score Card" provided by CMS to Elevance for Contract H3655 (*see* Figure 1 from A.R. 838, Suppl. App. 314).

Figure 1: CMS Rounds the Weighted Average of Measure Specific Star Ratings to Six Decimals Using the "FIXED" Function in Microsoft Excel (A.R. 838, Suppl. App. 314) (emphasis added)

=NUMBERVALUE(FIXED((O48/N48),6))

P48 f_x =NUMBERVALUE(FIXED((O48/N48),6))

	Domain	Primary Data Source	Quality Measure	Score	Star	Weight	Weight * star	x bar	diff	diff squared	multiply by measure weight	Weight	Weight * star	x bar	diff	diff squared	multiply by measure weight
	Contract: H3655	Contract Type: Local & Regional CCP with SNP						Calculation Without Improvement						Calculation With Improvement			
	Contract Name: COMMUNITY INSURANCE COMPANY																
Part C Measures																	
1 – Staying Healthy: Screening, Tests, and Vaccines	HEDIS	C01: Breast Cancer Screening		71	3	1	3	3.608696	-0.608696	0.370511	0.370511	1	3	3.647059	-0.647059	0.418685	0.418685
	HEDIS	C02: Colorectal Cancer Screening		78	4	1	4	3.608696	0.391304	0.153119	0.153119	1	4	3.647059	0.352941	0.124567	0.124567
	CAHPS	C03: Annual Flu Vaccine		71	4	1	4	3.608696	0.391304	0.153119	0.153119	1	4	3.647059	0.352941	0.124567	0.124567
	HEDIS / HOS	C04: Monitoring Physical Activity		44	2	1	2	3.608696	-1.608696	2.587903	2.587903	1	2	3.647059	-1.647059	2.712803	2.712803
2 – Managing Chronic (Long Term) Conditions	Plan Reporting	C05: Special Needs Plan (SNP) Care Management		59	2	1	2	3.608696	-1.608696	2.587903	2.587903	1	2	3.647059	-1.647059	2.712803	2.712803
	HEDIS	C06: Care for Older Adults – Medication Review		99	5	1	5	3.608696	1.391304	1.935727	1.935727	1	5	3.647059	1.352941	1.830449	1.830449
	HEDIS	C07: Care for Older Adults – Pain Assessment		98	5	1	5	3.608696	1.391304	1.935727	1.935727	1	5	3.647059	1.352941	1.830449	1.830449
	HEDIS	C08: Osteoporosis Management in Women who had a Fracture		30	2	1	2	3.608696	-1.608696	2.587903	2.587903	1	2	3.647059	-1.647059	2.712803	2.712803
	HEDIS	C09: Diabetes Care – Eye Exam		72	3	1	3	3.608696	-0.608696	0.370511	0.370511	1	3	3.647059	-0.647059	0.418685	0.418685
	HEDIS	C10: Diabetes Care – Blood Sugar Controlled		83	3	3	9	3.608696	-0.608696	0.370511	1.111532	3	9	3.647059	-0.647059	0.418685	1.256056
	HEDIS	C11: Controlling Blood Pressure		72	2	3	6	3.608696	-1.608696	2.587903	7.763708	3	6	3.647059	-1.647059	2.712803	8.138410
	HEDIS / HOS	C12: Reducing the Risk of Falling		54	2	1	2	3.608696	-1.608696	2.587903	2.587903	1	2	3.647059	-1.647059	2.712803	2.712803
	HEDIS / HOS	C13: Improving Bladder Control		40	2	1	2	3.608696	-1.608696	2.587903	2.587903	1	2	3.647059	-1.647059	2.712803	2.712803
	HEDIS	C14: Medication Reconciliation Post-Discharge		86	4	1	4	3.608696	0.391304	0.153119	0.153119	1	4	3.647059	0.352941	0.124567	0.124567
	HEDIS	C15: Plan All-Cause Readmissions		12	3	3	9	3.608696	-0.608696	0.370511	1.111532	3	9	3.647059	-0.647059	0.418685	1.256056
	HEDIS	C16: Statin Therapy for Patients with Cardiovascular Disease		86	3	1	3	3.608696	-0.608696	0.370511	0.370511	1	3	3.647059	-0.647059	0.418685	0.418685
	HEDIS	C17: Transitions of Care		59	3	1	3	3.608696	-0.608696	0.370511	0.370511	1	3	3.647059	-0.647059	0.418685	0.418685
	HEDIS	C18: Follow-up after Emergency Department Visit for People with Multiple High-Risk Chronic Conditions		67	4	1	4	3.608696	0.391304	0.153119	0.153119	1	4	3.647059	0.352941	0.124567	0.124567
3 – Member Experience with Health Plan	CAHPS	C19: Getting Needed Care		83	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
	CAHPS	C20: Getting Appointments and Care Quickly		85	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
	CAHPS	C21: Customer Service		89	3	4	12	3.608696	-0.608696	0.370511	1.482043	4	12	3.647059	-0.647059	0.418685	1.674741
	CAHPS	C22: Rating of Health Care Quality		88	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
	CAHPS	C23: Rating of Health Plan		88	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
	CAHPS	C24: Care Coordination		87	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
4 – Member Complaints and Improvement in the Health Plan's Performance	CTM	C25: Complaints about the Health Plan		0.12	5	4	20	3.608696	1.391304	1.935727	7.742907	4	20	3.647059	1.352941	1.830449	7.321797
	MBDSS	C26: Members Choosing to Leave the Plan		15	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
	Star Ratings	C27: Health Plan Quality Improvement	Medical		4			Not used in this Calculation				5	20	3.647059	0.352941	0.124567	0.622837
5 – Health Plan Customer Service	IRE	C28: Plan Makes Timely Decisions about Appeals		97	4	4	16	3.608696	0.391304	0.153119	0.612475	4	16	3.647059	0.352941	0.124567	0.498269
	IRE	C29: Reviewing Appeals Decisions		99	5	4	20	3.608696	1.391304	1.935727	7.742907	4	20	3.647059	1.352941	1.830449	7.321797
	Call Center	C30: Call Center – Foreign Language Interpreter and TTY Av.		100	5	4	20	3.608696	1.391304	1.935727	7.742907	4	20	3.647059	1.352941	1.830449	7.321797
Part D Measures																	
1 – Drug Plan Customer	Call Center	D01: Call Center – Foreign Language Interpreter and TTY Av.		100	5	4	20	3.608696	1.391304	1.935727	7.742907	4	20	3.647059	1.352941	1.830449	7.321797
2 – Member Complaints and Improvement in the Drug Plan's Performance	CTM	D02 Complaints about the Drug Plan		0.12	5			Counted in Part C						Counted in Part C			
	MBDSS	D03: Members Choosing to Leave the Plan		15	4			Counted in Part C						Counted in Part C			
	Star Ratings	D04: Drug Plan Quality Improvement	Medical		4			Not used in this Calculation				5	20	3.647059	0.352941	0.124567	0.622837
3 – Member Experience with Drug Plan	CAHPS	D05: Rating of Drug Plan		86	2	4	8	3.608696	-1.608696	2.587903	10.351611	4	8	3.647059	-1.647059	2.712803	10.851213
	CAHPS	D06: Getting Needed Prescription Drugs		89	3	4	12	3.608696	-0.608696	0.370511	1.482043	4	12	3.647059	-0.647059	0.418685	1.674741
4 – Drug Pricing and Patient Safety	PDE & MPF Pricing Files	D07: MPF Price Accuracy		98	3	1	3	3.608696	-0.608696	0.370511	0.370511	1	3	3.647059	-0.647059	0.418685	0.418685
	PDE data	D08: Medication Adherence for Diabetes Medications		85	3	3	9	3.608696	-0.608696	0.370511	1.111532	3	9	3.647059	-0.647059	0.418685	1.256056
	PDE data	D09: Medication Adherence for Hypertension (RAS antagoni		89	3	3	9	3.608696	-0.608696	0.370511	1.111532	3	9	3.647059	-0.647059	0.418685	1.256056
	PDE data	D10: Medication Adherence for Cholesterol (Statins)		88	3	3	9	3.608696	-0.608696	0.370511	1.111532	3	9	3.647059	-0.647059	0.418685	1.256056
	Part D Plan Reporting	D11: MTM Program Completion Rate for CMR		89	4	1	4	3.608696	0.391304	0.153119	0.153119	1	4	3.647059	0.352941	0.124567	0.124567
	PDE data	D12: Statin Use in Persons with Diabetes (SUPD)		85	2	1	2	3.608696	-1.608696	2.587903	2.587903	1	2	3.647059	-1.647059	2.712803	2.712803
Rated Like		Local & Regional CCP with SNP needs at least 19 of 38 measures				92	332	3.608696			83.913043	102	372	**3.647059**			85.294118
MA-PD	2022 Major Disaster %	0				Sum of weights	Sum of weights * stars	Calculated Summary Mean			Sum of weighted squared diffs	Sum of weights	Sum of weights * stars	Calculated Summary Mean			Sum of weighted squared diffs
	2023 Major Disaster %	0															
New Measure(s)	With	With															
Improvement	Without	With															
# Measures Needed	19	19															
# Measures Scored	38	40				38 eligible		0.936750 Calculated Variance				40 eligible		0.857658 Calculated Variance			
Variance Category	med	med															
Reward Factor	0	0.1															
Interim Summary	3.608696	3.747059															
CAI Value	0.002506	0.002506															
Final Summary	3.611202	3.749565															
Overall Rating	3.5	3.5															
Final Overall Rating	3.5																

17.   As noted on Microsoft's own support page, the FIXED function in Excel "[r]ounds a number to the specified number of decimals…and returns the result as text." It does not purport to truncate the digits at the specified number of decimals for reporting or displaying purposes.

Figure 2: Microsoft's Description of the FIXED Function[10]

FIXED function

▶ *Applies To*

This article describes the formula syntax and usage of the **FIXED** function in Microsoft Excel.

Description

==Rounds a number to the specified number of decimals,== formats the number in decimal format using a period and commas, and returns the result as text.

Syntax

FIXED(number, [decimals], [no_commas])

The FIXED function syntax has the following arguments:

- **Number**   Required. The number you want to round and convert to text.
- **Decimals**   Optional. The number of digits to the right of the decimal point.
- **No_commas**   Optional. A logical value that, if TRUE, prevents FIXED from including commas in the returned text.

18.   The syntax for the function is such that a user provides the function with a number to be rounded to a specified number of decimal places. As an example provided by Microsoft, passing the number "1234.567" to the FIXED function and specifying that it should be rounded to one digit to the right of the decimal point, results in the output "1,234.6" (*see* **Figure 3**).

---

[10] Microsoft Support, *FIXED function*, https://support.microsoft.com/en-us/office/fixed-function-ffd5723c-324c-45e9-8b96-e41be2a8274a (last visited April 9, 2025) (Suppl. App. 317) (emphasis added).

Figure 3: Example of Functionality of the FIXED Function[11]

Example

Copy the example data in the following table, and paste it in cell A1 of a new Excel worksheet. For formulas to show results, select them, press F2, and then press Enter. If you need to, you can adjust the column widths to see all the data.

Data
1234.567
-1234.567
44.332

Formula	Description	Result
=FIXED(A2, 1)	Rounds the number in A2 one digit to the right of the decimal point.	1,234.6

19.   In calculating the weighted average of measure specific star ratings, CMS passes the ratio of the sum of the weighted measure star ratings over the sum of the weights (i.e., 372/102) as the number it wishes to round using the FIXED function (*see* **Figure 1 and Figure 4**). In this function, CMS specifies that it wants to round this value to the sixth decimal place to the right of the decimal point by specifying the number "6" in the FIXED function.

---

[11] Microsoft Support, *FIXED function*, https://support.microsoft.com/en-us/office/fixed-function-ffd5723c-324c-45e9-8b96-e41be2a8274a (last visited April 9, 2025) (Suppl. App. 318) (emphasis added).

Figure 4: CMS's Rounding of the Weighted Average of Measure Specific Star Ratings to the Sixth Decimal Place (A.R. 838, Suppl. App. 314) (emphasis added)
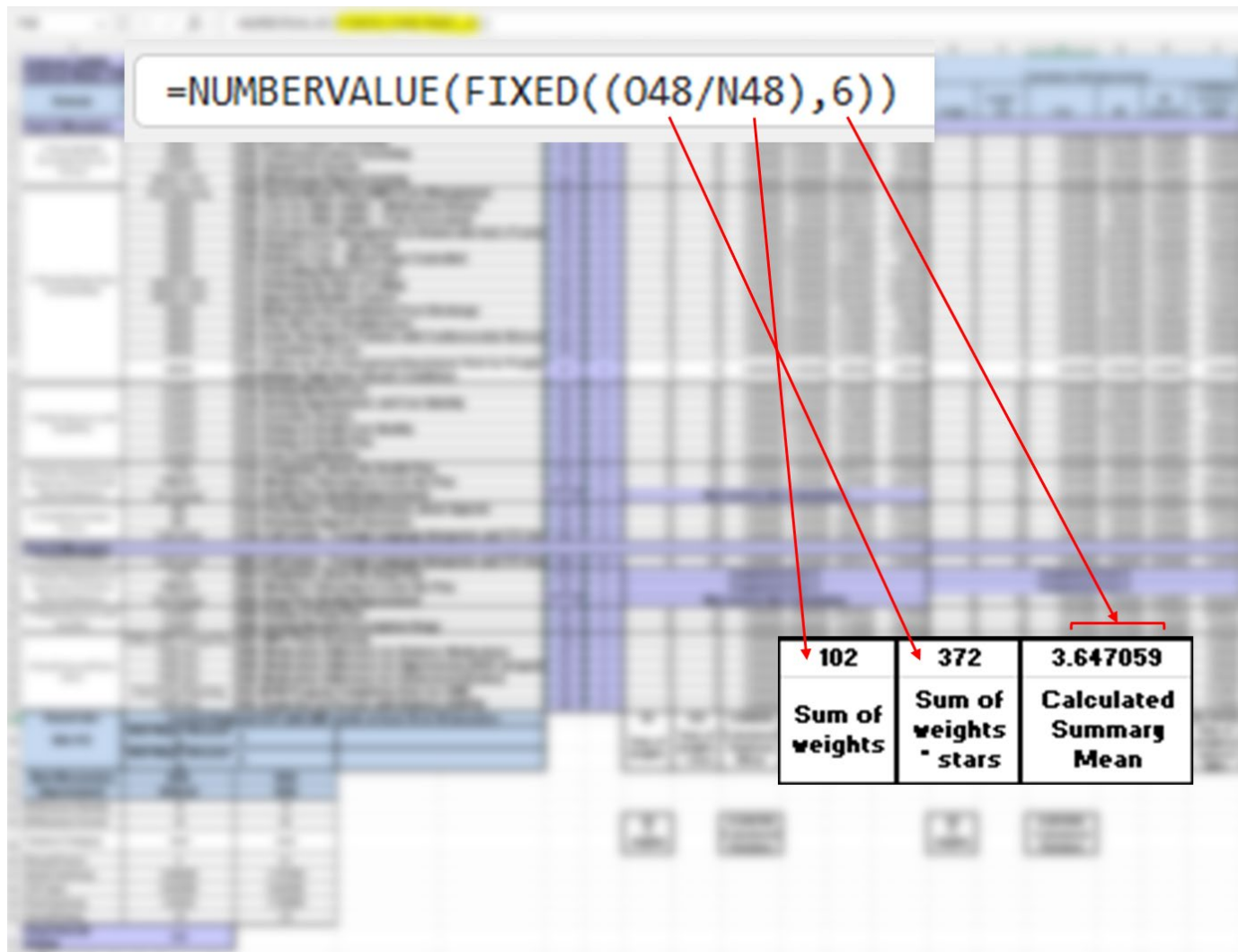
20.   As noted, the FIXED function converts this value to a "text" format. CMS uses the "NUMBERVALUE" function to convert this text formatted value to a numerical format for use in its later calculations. The application of this "NUMBERVALUE" function does **not** restore values beyond the sixth decimal place.

21.   The effect of the application of the FIXED function can be seen simply by *removing* it. To illustrate, allowing Excel to display beyond the sixth decimal place shows the effect of CMS's use of the FIXED function. When using the FIXED function as CMS does but allowing Excel to display 10 digits to the right of the decimal place, the calculated value is "3.6470590000." In other words, the sixth decimal place is rounded to a "9" and then 0's follow for every decimal place past the sixth.

Figure 5: Rounded Figure Displayed Out to 10 Decimal Places (A.R. 838, Suppl. App. 314) (emphasis added)

App. 331
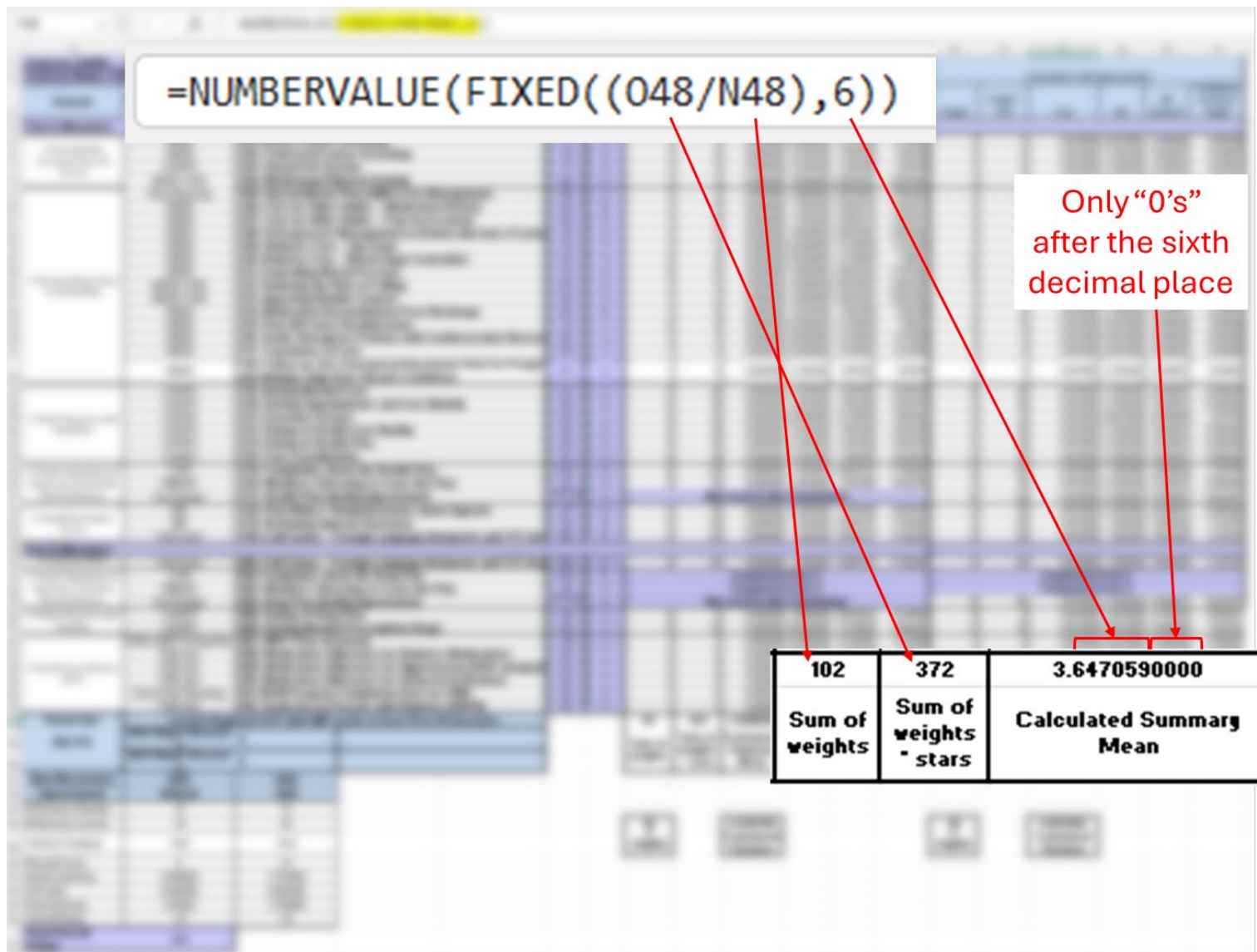
22.   As a result, the Interim and Final Summary Scores which are calculated using this rounded score are also rounded to the sixth decimal.
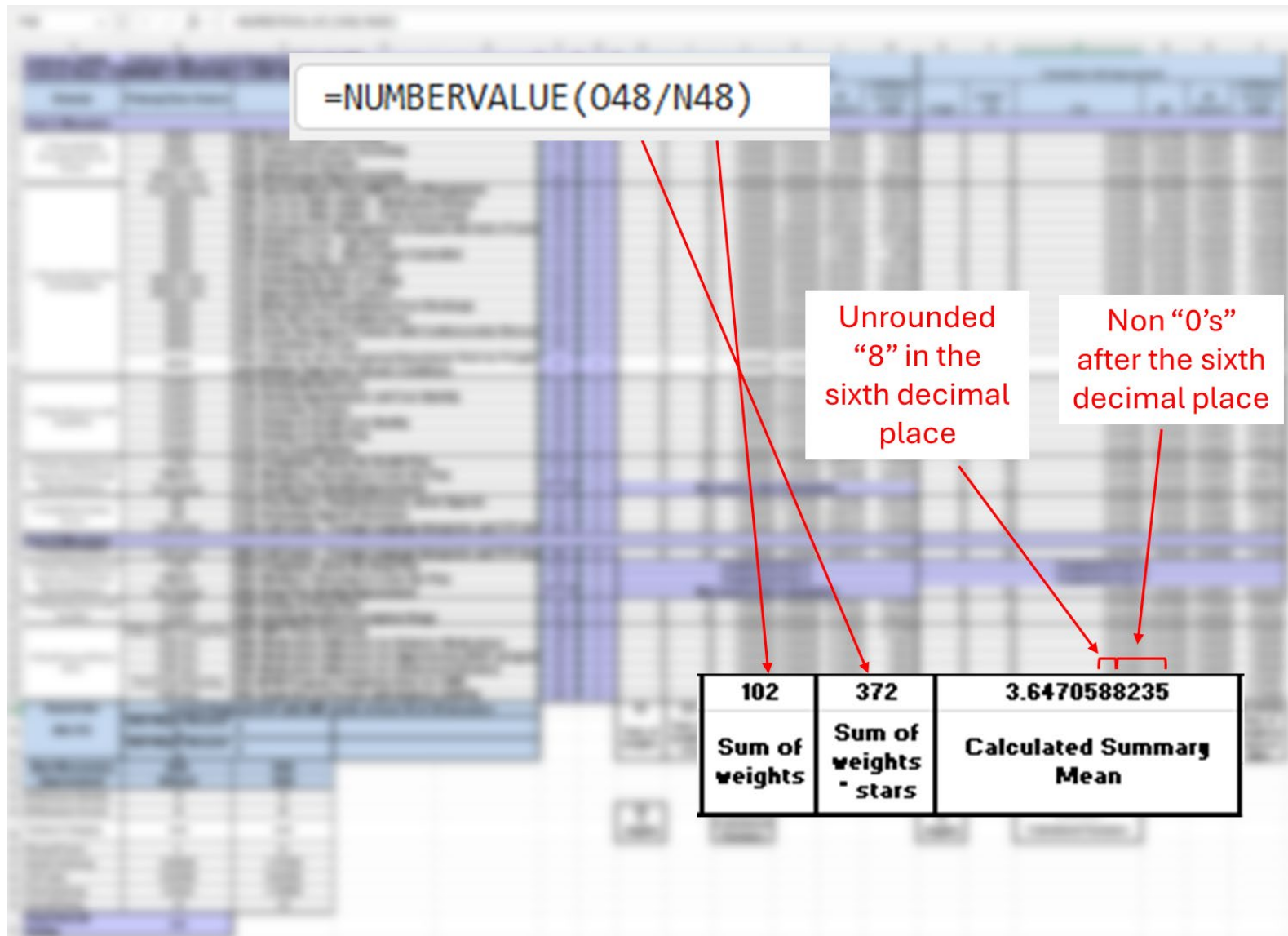
Figure 6:

# Measures Needed	19	19
# Measures Scored	38	40
Variance Category	med	med
Reward Factor	0	0.1
Interim Summary	3.608696	3.7470590000
CAI Value	0.002506	0.002506
Final Summary	3.611202	3.7495650000
Overall Rating	3.5	3.5

23.   CMS then takes the rounded Final Summary Score (i.e., 3.749565) and compares that to its star categorization thresholds (e.g., 3.750000) to determine its final Overall Star Rating (i.e., 3.5 per CMS's rules).

24.   In contrast, removing the FIXED function when calculating the weighted average reveals a Calculated Summary Mean number with an "8" in the sixth decimal place – in other words, a number no longer rounded to a "9" in the sixth decimal place – followed by non-zero numbers in subsequent decimal places (*contrast* **Figure 5** vs. **Figure 7)**.

Figure 7: Removing the FIXED Rounding Function Results in a Number No Longer Rounded to the Sixth Decimal Place (A.R. 838, Suppl. App. 314) (emphasis added)

25.   This weighted average carries through to the Interim Summary and Final Summary Scores as well, notably resulting in a Final Summary Score which is no longer rounded to the sixth decimal.

Figure 8

New Measure(s) Improvement	With Without	With With
# Measures Needed	19	19
# Measures Scored	38	40
Variance Category	med	med
Reward Factor	0	0.1
Interim Summary	3.608696	3.7470588235
CAI Value	0.002506	0.002506
Final Summary	3.611202	3.7495648235
Overall Rating	3.5	3.5

26.   The above also demonstrates that CMS is rounding at the sixth decimal and is *not* simply reporting out to the sixth decimal as suggested by the Goldstein Declaration when it mentions that:

> *"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...When those overall scores are displayed in CMS's technical guidance or for contracts to view, CMS displays only the six decimal places".* [12]

27.   A simple example demonstrates this issue where a theoretical Final Summary value lands on 3.7499999 (unrounded and reported to seven decimal places). Per CMS's "Codified Star Ratings Methodology" shared in Defendants' Consolidated Brief[13] (see **Figure 9**), this Final Summary value would result in a 3.5 Star Overall Rating because it is less than 3.750000; however, we have shown that CMS would display and report this unrounded Final Summary value as 3.750000 within the plan scorecard. Per the formula within CMS's plan scorecard files, this value would then get categorized as a 4.0 Star Overall Rating. Therefore, what happens at the sixth

[12] Defs.' App. ℙ 2.
[13] Defs.' Br. at 31.

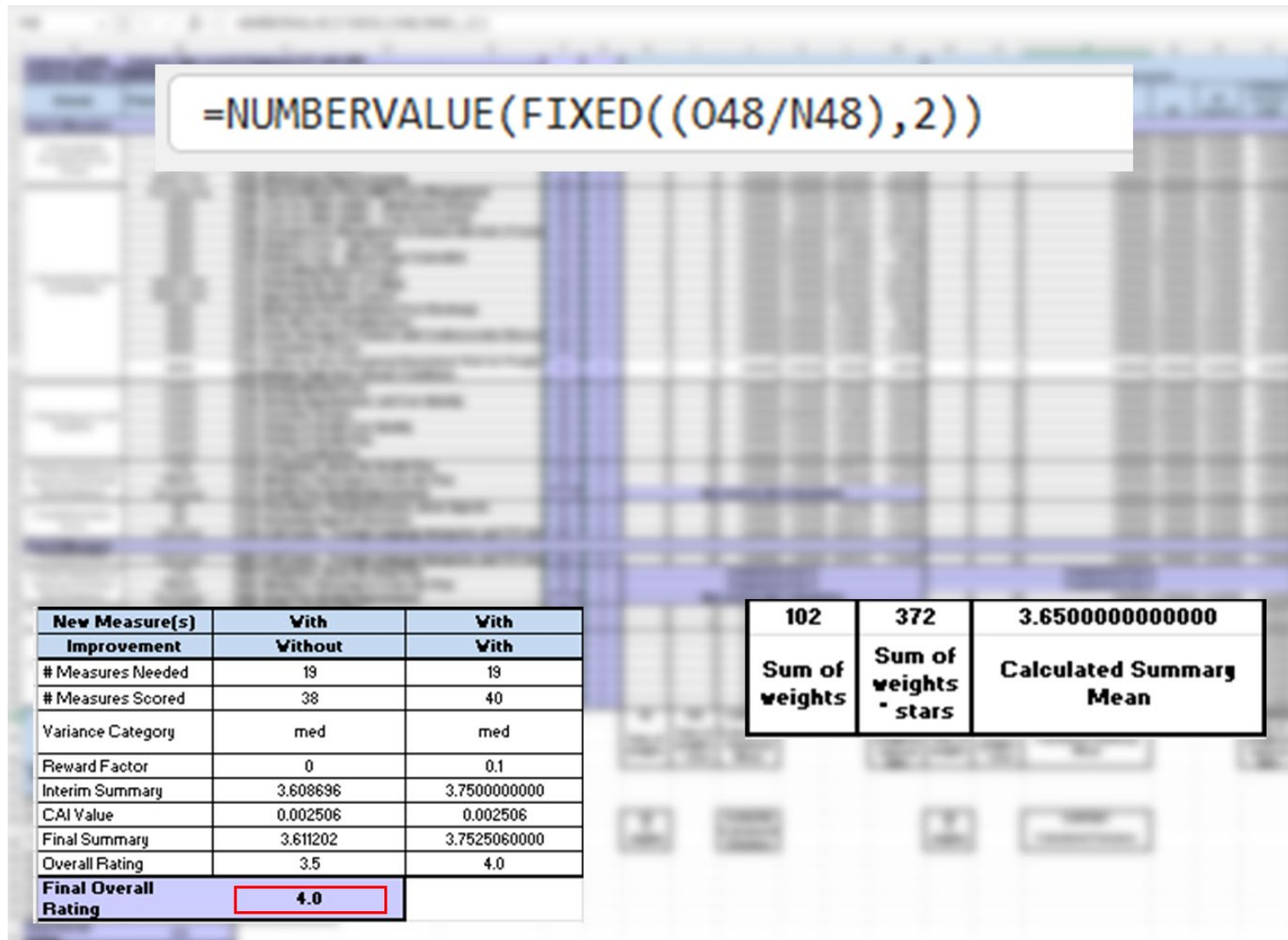decimal place clearly impacts the second, hundredth decimal place using CMS's current rounding method.

Figure 9

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

As a final exemplar – if one were to leave CMS's use of the "FIXED" function as it is currently – but change the specified rounding decimal places from 6 to 2 – the Elevance's final Overall Star Rating changes from a 3.5 to a 4.0 (*see* **Figure 10**).

Figure 10: Keeping CMS's Fixed Function But Rounding to the Second Decimal Place Instead of the Sixth Results in a Higher Overall Star Rating of a 4.0 for Elevance Contract H3655 (A.R. 838, Suppl. App. 314) (emphasis added)

New Measure(s) Improvement	With Without	With With
# Measures Needed	19	19
# Measures Scored	38	40
Variance Category	med	med
Reward Factor	0	0.1
Interim Summary	3.608696	3.7500000000
CAI Value	0.002506	0.002506
Final Summary	3.611202	3.7525060000
Overall Rating	3.5	4.0
Final Overall Rating	4.0	

=NUMBERVALUE(FIXED((O48/N48),2))

102	372	3.6500000000000
Sum of weights	Sum of weights · stars	Calculated Summary Mean

App. 336

5   CMS's Discussion of Precision Is Inherently Flawed

28.   CMS's discussion of precision also is flawed. Its criticism of the Diver Declaration conflates the width of an interval between two fixed points (e.g., the difference between the fixed thresholds 3.25 and 4.25 is 0.50) with that of the (im)precision of an estimate subject to fluctuation due to random chance (e.g., the uncertainty due to random chance in a contract's Final Summary Score).

29.   It is helpful to take a step back and consider at a high level what CMS does to assign an Overall Star Rating to each contract:

      i.   Calculate a weighted average of performance measures;

      ii.   Adjust the weighted average based on several factors to form a Final Summary Score;

      iii.   Compare the Final Summary Score to a series of fixed thresholds to determine an Overall Star Rating.

30.   CMS states that:

> *"Either your score is above, below, or equal to [a fixed threshold]."*[14]

31.   Problematically, this statement can only be universally true if the effect of random chance is ignored. Both the Diver Declaration and CMS's own simulations confirm this. The work contained therein demonstrates that elements in all three of the above steps are subject to fluctuation due to random chance alone:

Figure 11

      i.   Calculate a **weighted average** of performance measures;

      ii.   Adjust the **weighted average** based on several **factors** to form a Final Summary Score;

      iii.   Compare the **Final Summary Score** to a series of fixed thresholds to determine an **Overall Star Rating**

      Red box indicates an element in CMS's Star Rating determination methodology subject to random chance

---

[14] Defs.' Br. at 30.

32.   CMS does not dispute this. As such, under one initialization seed a Final Summary Score could be below a threshold, but under another initialization seed the same contract's Final Summary score could be above that same threshold.

33.   The Diver Declaration provides a discussion of precision as the degree to which Final Summary Scores can fluctuate due to random chance alone under CMS's methodology. In other words, the Diver Declaration provides a discussion on how much a contract's Final Summary Score could change due to nothing other than arbitrary, random chance. This is consistent with the definition of precision provided in the Diver Declaration as well as in materials published by CMS.[15]

34.   The Diver Declaration provides an assessment that this variability is on the order of 0.01. CMS could have easily offered a similar assessment based upon its own simulations but conspicuously did not.

35.   In contrast, CMS discusses precision through the lens that the difference between two fixed points not subject to random chance is 0.5 (e.g., the difference between a star rating of 3.5 and 4 is 0.5 and equivalently the difference between star rating determination thresholds – e.g., 3.250000 and 3.750000 – is also 0.5).[16] CMS suggests that since this difference between two fixed points is larger than 0.01 that the points concerning statistical uncertainty in the Diver Declaration are rendered moot.[17] CMS's own discussion however reveals either a fundamental failure to grasp the concept of statistical uncertainty or a misleading discussion of the concept.

36.   For example, CMS states:

> *"[Dr. Diver] points to the difference between 3.749565 and 3.75 – 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…*

---

[15] *See* Ctrs. for Medicare & Medicaid Servs., *HHS Risk Adjustment Data Validation (HHS-RADV) White Paper* (Dec. 6, 2019), *https://www.cms.gov/files/document/2019-hhs-risk-adjustment-data-validation-hhs-radv-white-paper.pdf*.

[16] Defs.' Br. at 32.

[17] Defs.' Br. at 33-34.

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.' 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."[18]

37.   There are several flaws in CMS's analysis here.

38.   First, in most simple terms – CMS establishes a fixed threshold for assigning a contract an Overall Star Rating of a 4.0. That threshold is a *Final Summary Score value* of a 3.750000. The difference between Elevance's observed Final Summary Score and this threshold is 0.000435. In other words, had Elevance's observed Final Summary Score been 0.000435 larger, CMS would have assigned contract H3655 an Overall Star Rating of a 4.0. CMS's analysis mistakenly confuses this issue by taking the difference between a *Star Rating* and a Final Summary Score. To put it another way, CMS performs an apples-to-oranges calculation. However, this flaw pales in comparison to its most substantive issue.

39.   CMS states that the difference between Elevance's observed Final Summary Score (3.749565) and a Star Rating of 4.0 is 0.250435, which is greater than 0.250000. Similarly, CMS states that the difference between Elevance's observed Final Summary Score and a Star Rating of 3.5 is 0.249565, which is less than 0.250000. This is reason enough for CMS to say Elevance should receive a 3.5. However, this rationale *ignores the random chance* inherent in CMS's methodology.

40.   A different initialization seed (i.e., a different roll of random dice) – changing nothing about Elevance H3655's actual performance – could have resulted in a Final Summary Score greater than 3.750000. For illustration, assume under this alternate (non-"8-6-7-5-3-0-9") initialization seed, H3655 receives a Final Summary Score of 3.7**5**9565 – a difference of 0.01 from

---

[18] Defs.' Br. at 33.

its observed score due to random chance alone. Now, CMS's calculated supposed "rounding error" differences are essentially flipped: the difference between this alternate score and 4.0 is 0.240435 while the difference from 3.5 is 0.259565.

41.   In other words, CMS's discussion of the "rounding error" calculated between an Overall *Star Rating* and Final Summary Score is misleading and conceptually flawed. What CMS terms "rounding error" is itself subject to random chance.

42.   It is also worth noting that rounding the Final Summary Score to the second decimal instead of the sixth prior to assigning an Overall Star Rating does **not** increase the width between relevant fixed thresholds but rather maintains the current difference (i.e., the difference between 3.75 and 4.25 is still 0.5). However, it does decrease the probability that a plan is harmed by not considering the effect of random chance in CMS's methodology.

6   Contrary to CMS's Assertion, Using a Well Known Seed – "Like a Number from Pop Culture" – Is Not a Best Practice Requirement for Selecting an Initialization Seed

43.   As noted in the Diver Declaration, CMS uses the same initialization seed "8-6-7-5-3-0-9" each year in its mean resampling methodology. CMS claims that "a memorable seed is necessary to replicate results of a randomization process."[19] It states that "[t]he best practice is to use a number that is well known, like a number from pop culture, hence [CMS's] reference to the Tommy Tutone song, 'Jenny.'"[20] In my experience as a statistician, I am aware of no such best practice that suggests a "memorable" seed like one from pop culture be used. In contrast, best practice is typically simply to ensure that the seed that is used is properly documented so that the otherwise random process can be replicated. Indeed to this end, CMS's own Medicare Program Integrity Manual provides guidance to this effect.   It requires the documentation of the

---

[19] *See* Defs.' Br at 43.

[20] *See* Defs.' Br. at 43.

initialization seed but notably makes no mention of inherent "memorability" nor more specifically any mention of using a pop culture reference: [21]

> *"The contractor shall identify the source of the random numbers used to select the individual sampling units, if used. The contractor shall also document the program and its algorithm or table, when available, that is used; this documentation becomes part of the record of the sampling and must be available for review. The contractor shall document any starting point if using a random number table or drawing a systematic sample. **In addition, the contractor shall document the known seed value if a computer algorithm is used.** The contractor shall document all steps taken in the random selection process exactly as done to ensure that the necessary information is available for anyone attempting to replicate the sample selection."*

44.   Not only is there no mention of selecting a "memorable" seed, but the guidance to properly document the source of the seed number carries some implication that there is no expectation that the seed number be inherently "memorable."

Pursuant to 28 U.S.C. § 1746, I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct. Executed on April 11, 2025 in Washington, D.C.

_____

Paul Diver, Ph.D.

---

[21] *See* App. 315-316, Medicare Program Integrity Manual, Ch. 8, Section 8.4.4.2 (emphasis added), available at: https://www.cms.gov/regulations-and-guidance/guidance/manuals/downloads/pim83c08.pdf.

UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS

ELEVANCE HEALTH, INC., et al., Plaintiffs, v. **ROBERT F. KENNEDY, JR.**, in his official capacity as Secretary of Health and Human Services, U.S. Department of Health and Human Services, and **MEHMET OZ**, in his official capacity as Administrator, Centers for Medicare and Medicaid Services, Defendants.	Case No. 4:24-cv-01064 Hon. Mark T. Pittman

SUPPLEMENTAL DECLARATION OF PAUL J. LAVRAKAS, PH.D.

I, Paul J. Lavrakas, Ph.D., declare the following to be true and correct:

1.   I am over 21 years of age, of sound mind, and fully qualified and competent to make this declaration.

I.   Error in Survey Research

2.   There are several types of major sources of error that can reduce the accuracy (validity and reliability) of the data that a survey produces. And in evaluating the presence of errors in the CAHPS dataset, I identified three types of error as being particularly troublesome: (a) Sampling Error (Imprecision), (b) Nonresponse Error (Bias); and (c) Measurement Error (Imprecision and Bias).

3.   This surrebuttal will address each of these forms of error in CAHPS data (and in the CMS's usage of CAHPS data) and the misunderstandings about each demonstrated by the Defendants' filing and the supporting Declaration of Ms. Elizabeth Goldstein.

4.   It is of note that in her Declaration, Ms. Goldstein includes one paragraph that purports to address sampling error in surveys. Ms. Goldstein then states that confidence intervals in surveys are a "descriptive tool," that the "most accurate, unbiased measure of a contract's performance is the mean," and "using the upper confidence limit and lower confidence limit would be biased and less accurate." Ms. Goldstein makes these statements without quoting a single expert source or providing any evidence that she possesses any expertise in survey research methods. In addition. Ms. Goldstein provides no evidence that she has been trained in the conceptualization, conduct, interpretation and/or evaluation of sample surveys.

5.   Ms. Goldstein further makes these statements without disclosing any information about the sampling designs of the CAHPS data quality (for instance, the "design effects" also known as *deffs*)[1]. Without sampling design metrics, such as *deffs*, being calculated and transparently disclosed, it is impossible for an observer to know the precision (i.e., the amount of error in the form of imprecision/variance) associated with a given CAHPS survey or know basic information about the size of the confidence intervals around a given mean or percentage.

II.   Nonresponse vs. Nonresponse Bias

6.   CMS' brief asserts that my principal concern regarding CAHPS surveys is that the response rates are less than 80%. That is incorrect. As I wrote in my original declaration and address further below, my principal concern is that CMS does not address Nonresponse Bias at all

---

[1]   Gabler, S., Ganninger, M., Hader, S., and Munnich, R. (2008). Design Effects. In P.J. Lavrakas editor, *Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Publishing, pp. 193-197.

in the specific CAHPS surveys. CMS addresses some aspects of nonresponse, but not the bias (error) that nonresponse can, and often does, cause.

7.   Nonresponse and Nonresponse Bias are not the same thing, and CMS's documents and statistical procedures and survey methods pertaining to CAHPS data do not accurately differentiate between the two. *Nonresponse* (in particular "unit" nonresponse vs. "item" nonresponse) refers to occasions when an individual who is sampled to participate in a given survey fails to provide any of the data the survey is seeking to gather. Nowadays, this occurs in essentially all surveys. In contrast, *Nonresponse Bias* refers to instances when nonrespondents to a given survey, as a group, would have provided materially different data for a given survey measure than was provided by respondents, as a group. Whereas Nonresponse is a necessary condition for Nonresponse Bias, it is not a sufficient condition (cf. Groves, 2006). Therefore, a survey may have Nonresponse associated with the recruitment of its respondents, but that does not guarantee that Nonresponse Bias will result from that Nonresponse (cf. Groves, 2006; Groves and Peytcheva, 2008, Bland et al. 2022[2]). And as reported by Groves and Peytcheva (2008), nonresponse rates are relatively poor predictors (i.e., weak correlates) of Nonresponse Bias. Furthermore, if nonrespondents are a random subset of those who were initially sampled to participate in a survey, then there will be no Nonresponse Bias due to nonrespondents.

8.   CMS's brief (as supported by Ms. Goldstein's declaration) does not appear to understand the difference between Nonresponse and Nonresponse Bias, and the important implications of this difference. They report that statistical adjustments (e.g., their case-mix

---

[2]   Bland, C., Zuckerbranun, S., Lines, L.M/ Kenyon, A., Hinsdale-Shouse, M. et al. (2022). Challenges Facing CAHPS Surveys and Opportunities for Modernization. RTI Press: Occasional Paper ISSN 2378-7996.; Groves, R.M. (2006). "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly*, 70(50), 646-675; Groves R.M. and Peytcheva, E. (2008). The Impact of Nonresponse Rates on Nonresponse Bias: A Meta-Analysis. *Public Opinion Quarterly*, 72(2) 167–189.

3

adjustment) are made to correct for *nonresponse* when they instead should be correcting for *Nonresponse Biases*.

9.   CMS cites to a 2022 Research Triangle Institute (RTI) article (Bland et al., 2022, p. iv.) that reviewed the methods used to conduct the CAHPS studies and the manner in which CMS put those data to use. That article recommended CMS "to review and implement [our recommended] innovations to the CAHPS surveys and their dissemination of the results" (Bland et al., 2022, p. iv.). And as part of the recommended innovations were "Address nonresponse through survey and statistical methods" (Bland et al. 2022, p. 7) and " Analyze Nonresponse Bias and conduct adjustments where bias is present" (Bland et al. 2022, p. 8). Of note, the article is actually focused on decreasing nonresponse rates and adjusting for nonresponse (rather than adjusting for nonresponse bias) while referencing at times nonresponse bias. Bland et al. go on to opine that, "Nonresponse analysis is an important part of identifying subgroups that may be left behind, and nonresponse bias analysis is a tool that identifies whether the missing responses are causing bias." (p.8). Those RTI authors further explain, "[w]hen subgroups of the population are missed, there is an increased risk to data quality because of nonresponse bias. Researchers can address some of the nonresponse bias with statistical solutions such as weighting. However, it is possible to reach a point where weights are not enough to make up for a lack of specific respondents." (p.8). In other words, weighting responses under certain conditions may help address bias, but it does not eliminate it when there is nonresponse. Furthermore, as Groves (2006) notes, nonresponse bias exists at the level of an individual survey question, not at the level of the entire questionnaire or subsets of questions. Therefore, any adjustment such as the case-mix one that CMS uses that is a survey-wide adjustment is wholly unsuited to make proper adjustments at the level of each of the individual CAHPS survey items that are used to help create the STARS

4

Ratings. Thus, applying the CMS's case-mix adjustment to an entire CAHPS dataset is wholly contrary to best practices in survey research.

10.   And more importantly, as the federal Office of Management and Budget ("OMB") mandates for federal agencies conducting surveys, an investigation into Nonresponse Bias is necessary to understand if it exists and how to help account for it. Without such an investigation, one cannot properly address Nonresponse Bias, whether by weighting the responses or through other means.

11.   Failing to investigate and adjust for Nonresponse Bias creates additional problems for the accuracy of CAHPS data. First, CMS may be failing to adjust for all the Nonresponse Biases that may occur. And second, CMS may unnecessarily adjust for respondent characteristics that do not require adjustments if they are not causing Nonresponse Biases. And worse yet, if CMS is needlessly adjusting/weighting for unnecessary characteristics, then they are likely increasing the *deffs* in their findings thereby reducing the effective sample sizes of their CAHPS surveys by increasing their variances.

12.   As I noted in my original declaration, CMS does not report that they have carried out any Nonresponse Bias investigations (cf. Montaquila and Olsen, 2012) for CAHPS surveying as has been required by the OMB for federal surveys since 2006. It is only through the use of Nonresponse Bias investigations that survey researchers can determine if there is likely to be non-ignorable Nonresponse Biases in their final data. And it is only through these investigations that researchers will know with confidence what are the respondent characteristics (e.g., Age? Sex? Education? Race? Income? etc.) that should be adjusted for, in order to try to reduce or eliminate the Nonresponse Biases.

13.   Thus, to correctly deal with Nonresponse Bias, CMS would need to stop adjusting for Nonresponse, carry out Nonresponse Bias investigations, and make adjustments for respondent characteristics that are found to be associated with the Nonresponse Biases.

14.   Since CMS does not report carrying out any Nonresponse Bias investigations as required by the OMB it has no reliable and valid method to adjust CAHPS data for Nonresponse Biases that may be present in the data used to create its Star Ratings.

III.   Why Confidence Intervals (CIs) are Central to Understanding Sampling Error and Survey Accuracy

15.   Sampling Error, in the form of imprecision/variance caused in all sample surveys due to sampling, occurs by mere chance because only a sample of the target population is studied. Thus, merely by chance, the initially selected sample and the final sample that provided data may be unrepresentative of the target population and the extent of this uncertainty is quantified in the form of Confidence Intervals when the survey's sample is a probability-based one. [3]

16.   CMS's brief and Ms. Goldstein's declaration does not appear to understand Confidence Intervals, their usage, and their value to decision-makers in survey research. In survey research, including the CAHPS surveys, it is widely recognized that there are no post-survey adjustments, including the CMS's case-mix adjustment, that eliminate the need to consider confidence intervals or to "make them go away" whilst one is making a decision about a survey's findings.

---

[3]   "A probability sample can provide a point estimate of an unknown population parameter [e.g., proportion of the target population that is satisfied with their health insurance provider] and the standard error of that point estimate. This information can be used to construct a confidence interval to give an estimated range of values around the point estimate [found by a CAHPS survey] that is likely to include the unknown population parameter." Vaish, A. (2008). Confidence Interval. In P.J. Lavrakas editor, *Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Publishing, pp. 126-130.

6

17.   The size of the margin of sampling error, and resulting Confidence Intervals, in all CAHPS surveys are such that they make invalid any comparisons of means or percentages smaller than several integers in the size of the difference, and thus it is statistically incorrect to draw conclusions about survey-based differences that are far less than one integer in size, let alone differences that are as small as CMS does for some Elevance Health plans. In particular, no CAHPS survey data can validly be used to conclude that 3.749565 is reliably different from 3.750000.

IV.   Question Reliability and Measurement Error in CAHPS Data

18.   It is widely understood in survey research that the data gathered (including data for the questionnaire items used in creating the Stars Ratings) are imperfect in regards to their accuracy (i.e., less than perfect validity and reliability). They may be "good" quality survey questions, but they are *fallible and imperfect measures* of the concepts that CAHPS is trying to measure.

19.   In survey research, the Total Survey Error framework (Groves et el., 2014; Lavrakas, 2013) that I described in my original declaration (see Section II, paragraphs 11 & 12). Figure 2 in that declaration shows that Measurement Error occurs in all surveys. This form of error can, and most often does, occur because of multiple factors, including *questionnaire-related error*[4] and *respondent-related error*.[5]   CMS does not address this in its response. And there is no evidence presented indicating that CMS does that when they use the CAHPS data.

20.   Regarding the imperfect reliability of the data gathered by the CAHPS and used by CMS to create the Star Ratings there are several articles that address these matters that have been
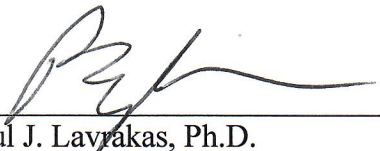
---

[4]   Basson, D. (2008). Questionnaire-Related Error. In P.J. Lavrakas editor, *Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Publishing, pp. 660-663.

[5]   Dykema, J., Blizt, S. & Stevenson, J. (2008), Respondent-Related Error. In P.J. Lavrakas editor, *Encyclopedia of Survey Research Methods*. Thousand Oaks CA: Sage Publishing, pp.745-748.

published by researchers using CAHPS data. In publications using CAHPS data, researchers have reported on the reliability of CAHPS measures and have *never reported that a measure is perfectly reliable*. Granted many CAHPS measures have strong reliability, but all are imperfect (i.e., all are in error) to some extent (see Table 4, Hargraves et al. (2003)[6]; p. 689, Hays et al. (2014)[7]; pp 6-7, Dyer et al.(2012)[8]; Table 3, Maldonado, et al. (2012)[9]; Table 2 in Roberts et al., (2019)[10]; and Figure 2 in Orr et al., (2023)[11]).

Pursuant to 28 U.S.C. § 1746, I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed on April 11, 2025, in Evanston, Illinois.

Paul J. Lavrakas, Ph.D.

---

[6]   Hargraves, J. L., Hays, R.D., Cleary, P.D. (2003). Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 Adult Core Survey. Health Services Research, 38(6), 1509-1528.

[7]   Hays, R.D., Berman, L.J., Kanter, M.H., Hugh, M., Oglesby, R.R., et al. (2014). Evaluating the Psychometric Properties of the CAHPS Patient-Centered Medical Home Survey. *Clinical Theraputics*, 36(5), 689-696.

[8]   Dyer, N. Sorra, J.S., Smith, S.A., Cleary, P. & Hays, R. (2012). Psychometric Properties of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group Adult Visit Survey. *Medical Care,* 50(Suppl): S28–S34. doi:10.1097/MLR.0b013e31826cbc0d.

[9]   Maldonado, R.W., Carle, A., Weidmer, B., Hurtado, M., Ngo-Metsger, Q., et al. (2012). The Consumer Assessment of Healthcare Providers and Systems (CAHPS) Cultural Competence (CC) Item Set. *Medical Care*, 50(9 0 2), 1-22.

[10]   Roberts, B.W., Yao, J., Bosire, J., Mazzarelli, A., and Trzeciak, S. (2019). Development and Validation of a Tool to Measure Patient Assessment of Clinical Compassion. *JAMA Open Network,* 2(5), doi:10.1001/jamanetworkopen.2019.3976.

[11]   Orr, N. Zaslavsky, A.M., Hays, R.D., Cleary, P.D., Haviland, A.M., Brown, J.A., Dembosky, J.W., Martino, S.C., Gaillot, S., and Elliott, N.M. (2023). Development, methodology, and adaptation of the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) patient experience survey, 2007–2019. *Health Services and Outcomes Research Methodology,* 23(1), 1-20.

8