UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ALIGNMENT HEALTHCARE INC., Plaintiff, v. U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, *et al.*, Defendants.	Civil Action No. 25-0074 (CRC)

DEFENDANTS' REPLY IN FURTHER SUPPORT OF
DEFENDANTS' CROSS-MOTION FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Statutes, Regulations, and Federal Register

INTRODUCTION

The Star Ratings methodology is the product of years of agency experience and iterative refinement, codified in regulation after notice-and-comment rulemaking. It is designed to apply uniformly across all Medicare Advantage contracts, providing beneficiaries with a meaningful and reliable tool for comparing plan quality. The process reflects a deliberate and methodical effort by the Centers for Medicare & Medicaid Services ("CMS") to promote consistency, accuracy, and fairness. Alignment insists this case is simple—and in one respect, it is: dissatisfied with the Star Ratings assigned to certain of its contracts, Alignment asks this Court to invalidate key components of the methodology for those contracts alone and to require CMS to treat Alignment differently from every other Medicare Advantage plan.

But against the backdrop of a uniform, carefully structured regulatory scheme, Alignment's programmatic objections—including to the use of Tukey outlier deletion methodology, composite measures, survey sample size, and the use of an independent review entity—fail. CMS applies these methodologies consistently when evaluating all contracts because they are grounded in sound policy and are supported by both the regulatory framework and longstanding program experience. What Alignment challenges here are not flaws in the system, but results it finds inconvenient for the contracts at issue. Alignment would prefer different policies applied to particular contracts, because it would prefer a higher rating for those contracts. That is not a valid basis to overturn a lawful and uniformly-applied regulatory framework.

Alignment's contract-specific complaints fare no better. Its concerns regarding the Spanish-language administration of the CAHPS survey and the independent review entity's determinations in two administrative cases were raised through the plan preview process and were appropriately considered by CMS. The Administrative Record reflects that CMS reviewed Alignment's concerns, engaged with Alignment, and reasonably concluded that no changes to

Alignment's Star Ratings were warranted. Alignment's dissatisfaction with CMS's well-reasoned determinations does not render those decisions arbitrary or capricious.

For these reasons, as well as those set forth in Defendants' Opposition to Plaintiff's Motion for Summary Judgment and Cross-Motion for Summary Judgment (ECF No. 12, "Defs. Mot."), this Court should grant Defendants' motion for summary judgment and deny Alignment's motion for summary judgment and its request that the Court order CMS to recalculate Alignment's Star Ratings for contracts H3443, H3815, and H9686 by applying Alignment's preferred policies.

ARGUMENT

I. Alignment's Critique of Tukey Outlier Deletion Fails to Show that CMS's Policy Regarding Minimum Denominator Thresholds is Arbitrary and Capricious.

Alignment argues that it is arbitrary and capricious for CMS to apply Tukey Outlier Deletion in determining Star Ratings for certain measures without adjusting the minimum denominator thresholds so that the contracts at issue could theoretically achieve each possible Star Rating for those measures. Alignment contends that CMS has two choices: abandon Tukey Outlier deletion for its two smaller contracts or exclude those contracts from all measures where small contracts may not be able to earn a measure-level Star Rating at each numerical increment. Pl. Opp'n (ECF No. 14) at 7-11; *see also* Pl. Mem. (ECF No. 11-1) at 29-30. This is a false dichotomy. As discussed below, the choice CMS made regarding the utilization of Tukey Outlier deletion and to evaluate all contracts with the required minimum denominator for a given measure is entirely reasonable, reasonably explained, and supported by record evidence.

A. Cut Point Compression from Tukey Outlier Deletion Is Neither Improper nor a Sign of a Methodological Flaw.

In its reply, Alignment continues to argue that "the removal of outliers is certain to compress measure-level cut-points," Pl. Opp'n (ECF No. 14) at 7, and claims that "CMS does not deny that cut-point compression produces arbitrary and unfair results for contracts with small

measure-level denominators," *id.* at 6. But Alignment's characterization of CMS's position is wrong. *See, e.g.*, Defs. Mot. (ECF No. 12). at 18-25 (explaining in great detail Defendants' response to Plaintiff's argument on this issue). As explained in the government's opening brief, "CMS's application of Tukey outlier deletion is neither arbitrary nor capricious. It reflects a well-reasoned, data-driven effort to enhance the integrity of the Star Ratings system by preventing statistical distortions, ensuring fairness across all contracts, and accurately reflecting relative performance." *Id.* at 24.

Alignment fails to meaningfully engage with the actual relevant question—which is not merely "is there compression?[1]" but instead is whether any compression that results after Tukey outlier deletion undermines the validity of the Star Ratings. The mere fact that cut points may be closer together after Tukey outlier deletion is applied does not, standing alone, signal a methodological flaw. As CMS explained, "[o]utlier deletion may increase or decrease cut point thresholds, depending on the shape of the measure's score distribution," and "closer cut points do not necessarily imply lower reliability or lessen the ability to distinguish between contracts." 88 Fed. Reg. 22,120, 22,297 (Apr. 12, 2023). Rather, the purpose of Tukey outlier deletion is to remove statistical anomalies so that cut points reflect the true performance scores. Any resulting compression is a byproduct of more representative data—not a sign of invalidity. That shift demonstrates the very reason outliers are removed: a single extreme contract can exert undue

[1] Alignment asserts that "Before this Court, CMS disputes (at 20–21) that Tukey outlier deletion leads to cut-point compression." Pl. Opp'n (ECF No. 14) at 7. Not so. As explained in the government's opening brief, CMS acknowledged that: (1) not all measures have outliers; (2) for those that do, outlier removal did not significantly affect most cut points (AR 103–04; *see also* 88 Fed. Reg. 22,120, 22,296 (Apr. 12, 2023); but (3) for some measures, removing outliers does shift the cut points required for each Star Rating, narrowing the bands required for each Star Level. Defs.' Mot. (ECF No. 12) at 19-22. CMS does deny that the fact that there can be compression, standing alone, is evidence of invalidity of the Star Ratings methodology. 88 Fed. Reg. at 22,297.

influence on the determination of cut points used to assign stars to all contracts. AR 0104. Put another way, the relevant question is not whether scores are closer together, but whether that compression obscures or misrepresents differences in performance across contracts.

And CMS has confirmed that the Star Ratings continue to reliably reflect material differences in performance levels after outlier deletion: the "Tukey outlier approach lessens the influence of a few outliers on cut point formation, leading to more reliable and stable thresholds, especially for the 1–2 star cut points." 88 Fed. Reg. at 22,926. As Defendants explained in their opening brief, far from distorting results, the method improves reliability and preserves the Star Ratings' core function—helping beneficiaries distinguish among plans based on meaningful differences in quality. Alignment has not—and cannot—show that, for any measure affected by outlier deletion, the resulting cut points are no longer a reliable metric for contract performance.

B. Alignment's Preferred Policy of Applying an Adjusted Denominator Requirement to Guarantee that Each Contract Can Achieve Each Star Level Is Not Required.

On reply, Alignment again claims that CMS must guarantee every plan has the ability to obtain every star level on every measure in the Star Ratings system. Pl. Opp'n (ECF No. 14) at 10. It contends CMS's "refusal" to adopt this policy is arbitrary and capricious—without offering any explanation beyond the bare assertion that "[i]f it is logically possible for larger plans to [earn all measure-level Star Ratings, 1 through 5], it must be logically possible for all plans to do so." *Id.* at 11. This superficial invocation of "logic" glosses over the implications of Alignment's argument and reflects a misunderstanding of both CMS's methodology and the concept of reliability.

First, CMS does not impose a universal denominator threshold, as Alignment claims. *Id.* at 1. Rather, each measure has its own minimum denominator requirement, set through the technical specification process based on what is necessary to ensure statistically meaningful results. *See*, *e.g.*, Medicare 2025 Part C & D Star Ratings Technical Notes, AR 0982-1055. While

two measures require a minimum denominator of 11, most require substantially more, reflecting CMS's judgment about data sufficiency tailored to each individual measure. *Id*. Alignment cites no authority—other than its own policy preference—for the proposition that CMS must guarantee that every contract could theoretically receive each of the five Star Ratings on every measure. Nothing requires the Court to adopt Alignment's preferred policy, under which some high ratings would be invalidated for contracts that are not challenging their ratings. *See* Defs. Mot. (ECF No. 12) at 24.

Alignment also misconstrues how reliability is assessed in the Star Ratings system. Its claim that "[f]ailure to address cut-point compression by adjusting the minimum denominator requirement . . . makes the Star Ratings unreliable," Pl. Opp'n (ECF No. 14) at 7, misapprehends both the purpose of the denominator threshold and how reliability is achieved. Reliability involves two distinct but related steps: first, ensuring that each contract's score on a given measure (which is used to determine the Star Rating) reflects true performance rather than statistical noise (addressed through minimum denominator thresholds); and second, distinguishing contract performance using a clustering algorithm with mean resampling and Tukey outlier deletion. AR 0961. These are separate and sequential methodological safeguards. The denominator threshold ensures that a contract's measure score is meaningful; the clustering step ensures that the Star Ratings distinguish true differences in quality across contracts. Nothing requires CMS to guarantee access to each rating level for every contract. If two contracts have the same measure score and the scores are valid, it is reasonable for CMS to apply the same Star Rating to both contracts, regardless of differences in the denominators.

Alignment's argument further rests on a misunderstanding of the denominator threshold. Alignment asserts that CMS has already accepted the exclusion of some high-performing smaller

contracts through its use of a minimum denominator threshold for measure scores, and therefore must accept further exclusion for Star Ratings to "mitigate the effects of cut point compression." *See* Pl. Opp'n (ECF No. 14) at 10. But that misstates both the function of the denominator criterion and CMS's rationale. The denominator threshold for measure scores is not a mechanism to distribute Star Ratings evenly across plans; it is a statistical safeguard to ensure that each measure is sufficiently reliable, i.e., that the measure reflects true performance and not random noise. Scores with denominators that fall below the threshold are excluded not due to plan size itself, but because they fail to meet the reliability standard required for valid comparisons.

Once CMS has determined which contracts have met the minimum denominator requirements for a particular measure to help ensure reliable measure scores, it applies a separate methodology to distinguish performance among those contracts—using the clustering algorithm with mean resampling and Tukey outlier deletion—to assign Star Ratings based on actual performance. At this stage, measure scores, not denominator size, determine the distribution of Star Ratings. Alignment's suggestion that CMS must ensure all contracts can earn each star level misrepresents the goal. The standard is not whether every contract has theoretical access to every rating, but whether the Star Ratings reliably reflect true differences in contract performance. CMS treats all plans—large and small—the same under this framework. Alignment's claim that CMS's system "produces irrational differences in the treatment of [Medicare Advantage] contracts," Pl. Opp'n (ECF No. 14) at 9, reflects its continued misunderstanding: CMS's approach yields valid and meaningful ratings precisely because it first confirms score reliability through meeting the measure denominator requirement, and then reliably distinguishes across contracts using consistent, statistically sound methods.

In support of its argument, Alignment invites the Court to engage with hypotheticals that are both factually inaccurate and legally irrelevant to the contracts at issue. For example, it claims under a hypothetical, regarding the measure *Plan Makes Timely Decisions About Appeals*, that a "plan with just 11 enrollee appeals (reflecting either smaller size, higher enrollee satisfaction with claim handling, or both) that handles just one appeal untimely receives 2.0 Stars," while a plan "with 100 enrollee appeals that handles 10 appeals untimely receives 2.0 Stars," and another plan "with 100 appeals that handles five appeals untimely receives 4.0 Stars." Pl. Opp'n (ECF No. 14) at 9; *see also* Pl. Mot. (ECF No. 11) at 29. But this example misstates how the Star Ratings are actually assigned for this metric under CMS's published methodology. AR 1028. In reality, a contract with 100 appeals and 10 untimely decisions receives 3.0 Stars, not 2.0 Stars according to Alignment's hypothetical; one with 100 appeals and 5 untimely decisions receives 4.0 Stars; and a plan with 11 appeals and one untimely decision also receives 3.0 Stars, not 2.0 Stars. *See* AR 1028. These plans would thus have measure scores of 90%, 95%, and 91%, respectively. The plan scoring 95% justifiably earns a higher Star Rating than the plans scoring 90% and 91%. Under CMS's clustering algorithm with mean resampling and Tukey outlier deletion, those results are entirely reasonable—not arbitrary or irrational. The results reflect reliable, performance-based distinctions—not differences in plan size.

More fundamentally, Alignment's hypothetical is irrelevant. The contract-level data at issue here do not involve any plans with denominators as small as 11; Alignment's relevant contracts had denominators of 18 and 15, respectively, for the appeals measures for contracts H3443 and H9686. Its example, therefore, bears no resemblance to the facts of this case and offers no support for its challenge to CMS's methodology.

Behind Alignment's sweeping assertions about Star Ratings measures generally is the fact that its actual objection concerns only the two appeals measures, which have a minimum denominator of 11. Compl. (ECF No. 1) ¶ 91 ("reliability of the 'Plan Makes Timely Decisions about Appeals' and 'Reviewing Appeals Decisions' measures was compromised for contracts H3443 and H9686"). Despite claiming that CMS's minimum denominator policies broadly disadvantage small plans, Alignment fails to identify any other non-CAHPS measures on which it was evaluated but unable to achieve the full range of star-level scores. *See* Pl. Opp'n (ECF No. 14) at 9. The reality is that this claim is narrowly focused; it applies only to the two appeals measures.

To avoid the consequences of the performance on the measure for two of its contracts, H3443 and H9686, Alignment's request would require this Court to direct CMS to exclude plans with 19 or fewer appeals for the measure Plan Makes Timely Decisions about Appeals. AR 1028-29 (minimum denominator and cut points for the measure Plan Makes Timely Decisions about Appeals). Alignment's contracts just missed this self-serving proposed "cutoff"—Contracts H3443 and H9686 had 18 and 15 appeals, respectively. Compl. (ECF No. 1) ¶¶ 92, 93. Alignment's proposal would amount to an Alignment-shaped exception to an evaluation metric, tailored to its own circumstances. The Court should reject Alignment's request that the Court order CMS to apply its preferred policy to the contracts at issue—advanced solely to excuse underperformance on a measure central to plan accountability. This proposal is untethered to any rational policy goal, and as already explained, would "inject new arbitrariness into the system." Defs. Mot. (ECF No. 13) at 25.

Alignment contends that "[w]hen some plans are mathematically incapable of attaining certain Star Ratings levels, the assigned ratings cannot be said to reflect comparative plan quality and performance." Pl. Opp'n (ECF No. 14) at 11. But again, just because Alignment says it is so

does not mean it's a fact. Alignment has not—and cannot—demonstrate that the "middle range" scores it received are not reflective of its contracts' performance. Alignment cannot dispute that for the contracts at issue, H3443 and H9686, its performance on this measure was notably lower than other contracts. *See* 2025 Star Ratings Data Tables, *available at* https://www.cms.gov/files/zip/2025-star-ratings-data-tables.zip. And it is reasonable that lower performance should be reflected in its Star Ratings.

In its briefing, Alignment seems to suggest that even a single late appeal can have an undue adverse effect on the Star Rating for a measure. But appeals are an important reflection of enrollees' access to care. Any appeal that is not timely submitted to the independent review entity and any decision to deny coverage that is not appropriate can create significant issues for enrollees—regardless whether they are enrolled in a large or small plan. For the enrollees who receive care, plan size is not a relevant distinction.

Alignment simultaneously asks CMS to carve out an exception for its contracts (to have a different minimum denominator) and to "treat like cases alike." Pl. Opp'n (ECF No. 14) at 11 (citing *Consol. Edison Co. of New York, Inc. v. FERC*, 45 F.4th 265, 279 (D.C. Cir. 2022)). But CMS has made a reasonable, data-driven determination that performance on the appeals measures is critical information for beneficiaries so long as there is a minimum denominator of 11—regardless of plan size. Alignment's contracts are not an exception for which the Court should order treatment different from other similar contracts. CMS does "treat like cases alike" by giving the same Star Rating to contracts with valid measure scores in the same range.

Finally, Alignment's requested relief as to this issue—either prohibit the use of Tukey outlier deletion for small contracts or impose a 20-appeal denominator threshold—is not merely a request to "set aside agency action" as "arbitrary, capricious, an abuse of discretion, or otherwise

not in accordance with law" pursuant to 5 U.S.C. § 706(2)(A) as Alignment contends in its Complaint, but rather a claim to "compel agency action" pursuant to 5 U.S.C. § 706(1). Yet, the APA authorizes courts to "compel agency action unlawfully withheld or unreasonably delayed," 5 U.S.C. § 706(1), "only within strict limits," *Anglers Conservation Network v. Pritzker*, 809 F.3d 664, 668, 670 (D.C. Cir. 2016). Courts may compel "*discrete* agency action that [an agency] is *required to take*," *Norton v. S. Utah Wilderness All.*, 542 U.S. 55, 64 (2004) (emphasis in original)—only when a statute or regulation imposes a "ministerial or non-discretionary" duty, *id*. at 63-64. But Alignment identifies no statute or regulation imposing a "specific, unequivocal command," *id.* at 63–64, requiring CMS to take the action it seeks. The relief Alignment requests, therefore, exceeds the bounds of judicial review under the APA and should be denied.

II. Contrary to Alignment's Assertions, There is No Clear Error in the CAHPS Survey Administration for Alignment's Spanish Speaking Enrollees.

Alignment argues that CMS's actions were arbitrary and capricious because, after raising concerns about the Spanish-language survey administration, "CMS refused to meaningfully investigate this issue and denied relief," Pl. Opp'n (ECF No. 14) at 11, and that "there is no longer any dispute that the surveys were improperly administered, *id.* at 13. These contentions lack merit.

While it is true that CMS did not discard the CAHPS survey data based on the response rate among Spanish-speaking enrollees, it is simply incorrect to claim that CMS failed to meaningfully investigate the issue. CMS considered the underlying data, evaluated potential methodological concerns, reviewed the response from Alignment's survey vendor, and concluded that the response patterns did not warrant invalidating the survey results. AR 0046-0073 (email correspondence between Alignment to CMS regarding the Spanish-language survey administration).

This Court need not take CMS's word alone: Alignment itself acknowledged that CMS was undertaking a careful and thorough review of the CAHPS survey issue. After raising its concerns, Alignment wrote to CMS, stating via email, "We appreciate your thorough review of our concerns and your ongoing efforts to address the issues raised. We value CMS's validation that the CAHPS data was handled in accordance with established methodologies and acknowledge that any changes to these processes must follow the formal rulemaking procedure." AR 0056.

CMS took concrete steps to investigate Alignment's concern: it directed its contractor to follow up with Alignment's CAHPS survey vendor, DataStat Inc., to determine whether there were any discrepancies in the Spanish-language survey administration for contracts H3815 and H3443. AR 0053, 1195. DataStat confirmed that it used client-supplied language preference data to identify Spanish-preferring enrollees, and further attested that it followed the Quality Assurance Plan and Technical Specifications. AR 0053. DataStat explained that it had matched Alignment's data to the CMS sample conservatively, avoiding mismatches when there were discrepancies in names or addresses, and also incorporated CMS's own Spanish-language preference flags. AR 1194.

CMS independently verified the integrity of the sampling process, i.e., the process selecting the representative sample of enrollees for Alignment's contracts. It analyzed the predicted probability of Spanish-language preference among the sampled enrollees and found that it closely matched the broader CAHPS-eligible population. AR 0053. CMS further examined the share of sampled enrollees with high predicted Spanish preference, the share who actually responded to the survey, and the proportion who completed it in Spanish. AR 0053, 0061.

CMS identified three key findings from its review: (1) the proportion of sampled enrollees with high predicted Spanish preference closely matched that of the eligible population; (2) as is

common in patient surveys, response rates were somewhat lower among individuals with high predicted Spanish preference; and (3) not all enrollees with high predicted preference chose to complete the survey in Spanish. AR 0061. These are typical patterns in survey administration and did not suggest any error in how the survey was conducted.

In short, CMS did not disregard Alignment's concerns—it undertook a careful, multi-step review, including direct vendor verification and independent statistical analysis. As CMS explained, "[O]ur analyses have established that the CAHPS sample represented Spanish-preferring members, used the language preference information you provided, and resulted in Spanish-preferring members choosing to respond in Spanish and rates that were high and higher than average. No further validation is needed." AR 0048. Alignment may be dissatisfied with the outcome, but that does not render CMS's process arbitrary or capricious.

Alignment then argues that even if CMS did conduct a review (it did), the CAHPS survey "was, indeed, not administered consistent with the agency's own rules and requirements," "there is no longer any dispute that the surveys were improperly administered," and CMS must therefore "discard the CAHPS survey data in calculating the Star Ratings for the affected contracts." Pl. Opp'n (ECF No. 14) at 13. Once again, Alignment mischaracterizes the government's position and its premises are incorrect. As discussed below, the record does not conclusively demonstrate that the surveys were improperly administered, and Alignment's inferences certainly provide no valid basis for discarding data.

Alignment's argument turns on its unsupported conclusion that the record demonstrates that data "uniquely in CMS's possession shows that DataStat sent English-language questionnaires to scores of sampled enrollees whom Alignment had designated as Spanish-speaking." Pl. Opp'n (ECF No. 14) at 13; *see also id*. (Alignment alleges there was "data solely in CMS's possession

shows beyond all dispute that the survey was … not administered consistent with the agency's own rules."). But the relevant data here—the Plan's enrollee-level Spanish language preference data—was not in CMS's possession. That is precisely why CMS permits (but does not require) Medicare Advantage Organizations to submit their own language preference data to the survey vendor. AR 0172-73. If CMS had the data, there would be no need for plans to supply it. This is Alignment's own data, which it elected to provide to its vendor, DataStat, for survey administration purposes, not to CMS. Together, Alignment and its chosen survey vendor devise how they wish to approach the survey administration.

Alignment's claim of survey error boils down to this: based on the language preference data it provided to its survey vendor and information in the record, Alignment infers that some enrollees it designated to DataStat as Spanish-speaking did not receive the initial survey pre-notification mailing in Spanish. From this, Alignment argues that the survey administration did not comply with CMS's protocol requiring that "Spanish language questionnaires must be made available to all Spanish-speaking enrollees." Pl. Opp'n (ECF No. 14) at 14 (citing AR 0172). But Alignment has not proven that this requirement was violated.

The protocol does not mandate that every Spanish-speaking enrollee automatically receive the initial mailing in Spanish. AR 0172-73 It requires that the Spanish-language version of the CAHPS survey be made available. AR 0172. That requirement was met. For enrollees who received the prenotification letter in English, those letters included instructions in Spanish for requesting a Spanish-language mail survey. AR 0173. For the web-based version, respondents could select Spanish directly when accessing the survey. AR 0172. And as Alignment's own survey vendor, DataStat, confirmed, Spanish-language versions were made available to all of Alignment's sampled enrollees. AR 0053. Alignment's argument is not about access to Spanish-

language materials—they were available to all sampled enrollees—but rather about whether a subset of individuals received the Spanish version automatically. Because that's not what CMS protocol requires, Alignment fails to show that CAHPS survey administration violated it. AR 0172.

Alignment argues that because it agreed with its survey vendor to send the Spanish language surveys to the data set Alignment identified, there was a survey error if the survey vendor did not follow its preferences. Pl. Opp'n (ECF No. 14) 13-14. But even the data cited by Alignment does not prove that there was an issue with DataStat's administration that resulted in error that invalidated the results. That some individuals designated as Spanish-speaking by Alignment did not receive the pre-notification letter initially in Spanish, is not ipso facto proof of error that would invalidate the survey. CMS provides the selected sample of enrollees to the survey vendor chosen by the plan; the vendor then matches the plan's language preference data (if provided by the plan) to those sampled individuals. Because Alignment does not know in advance which enrollees will be sampled, it must provide general enrollee files for the plan. AR 0174. The vendor must then attempt to reconcile those files with the CMS sample (*id*.)—a process that can result in imperfect matches. As DataStat explained, it used a conservative approach to this matching process, including that the plan's the name and address match the selected survey. AR 0053, 1194. There are thus plausible, non-error-based reasons why a small subset of enrollees may not have received the initial mailing in Spanish, such as if Alignment was not able to match Alignment's plan supplied data with the enrollee sample. AR 1194. But importantly, the record confirms that DataStat certified that it used Alignment's data consistent with CMS's protocols, which require that Spanish-language materials were made *available* to all sampled enrollees. AR 0053. Alignment infers and speculates that the survey was improperly administered in violation of CMS

requirements and protocols, and that CMS failed to properly investigate and remedy the violations, but the record does not prove its position.

Ultimately, Alignment offers no evidence that any enrollee was denied the opportunity to complete the survey in Spanish. Nor does it explain how any supposed mismatch between its own Spanish-language preference data and the CMS sample undermines the reliability of the survey results, particularly when CMS's review confirmed that the sample was representative with respect to predicted Spanish-language preference, and that Spanish-preferring enrollees did, in fact, respond to the survey in Spanish at rates higher than average, and there was no evidence that Spanish surveys were not made available to Spanish-speaking enrollees. AR 0048, 0053, 0061.

In short, the record supports the reasonableness of CMS's actions regarding survey administration. The record does not demonstrate a violation of CMS protocol or denial of access to Spanish-language materials, and there is no basis for discarding the survey results.

III. Unable to Show that Contract H3815 Was Harmed Because It was "Small," Alignment Shifts Its Argument Regarding Oversampling.

Alignment alleged in its Complaint, and again in its opening brief, that CMS's methodology for calculating CAHPS measure-level Star Ratings "arbitrarily" disadvantages smaller plans, asserting that it was "unable to oversample" for its contract H3815 while larger contracts were permitted to do so. Compl. (ECF No. 1) ¶¶ 153-55; Pl. Mot. (ECF No. 11) at 39-40. But as explained in Defendant's summary judgment motion, contract H3815—encompassing over 91,000 enrollees—was eligible to oversample and simply chose not to. Def. Mot. (ECF No. 13) at 36. Now, Alignment appears to abandon its original claim that this contract was too "small" or "unable" to oversample, pivoting to the theory that it was simply the plan's choice not to do so and that it should not be adversely affected by its own survey administration choice. Pl. Opp'n (ECF No. 14) at 14. That shifting position only underscores the weakness of its argument.

CMS agrees with Alignment that the Star Ratings system is designed to "accurately . . . reflect true performance." *Id.* at 3 (citing 83 Fed. Reg. 16,440, 16,519 (Apr. 16, 2018)). And that's why CMS undertook an evidence-based approach to determining required minimum sample size for the CAHPS survey: CMS determined that for contracts with at least 800 enrollees, a minimum sample size of 800 enrollees ensures statistically valid survey results. This threshold is rooted in established statistical principles and longstanding programmatic standards, designed to promote fairness, comparability, and reliability across all Medicare Advantage contracts. *See Frequently Asked Questions About the CAHPS® Program and Surveys, available at* https://www.ahrq.gov/cahps/faq; *see also* Orr, N., *Development, Methodology, and Adaptation of the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Patient Experience Survey*, at 445 (2022), *available at* https://link.springer.com/article/10.1007/s10742-022-00277-9 (establishing the 800-case standard "based on analyses of reliability from previous years"). This uniform baseline applies across the board and provides a consistent foundation for calculating Star Ratings.

As laid out in the government's opening brief, oversampling is one of several choices a contract may make in administering its survey. Def. Mot. (ECF No. 13) at 34-37. With a larger sample size, a contract may receive more survey responses, and thus the reliability of the CAHPS results may improve statistical confidence in the scores. But oversampling does not guarantee higher Star Ratings. Increased reliability is a neutral feature of the methodology—it can either help or hurt a contract, depending on actual performance. A high-performing plan may benefit from

oversampling, while a low-performing one may see its lower scores reinforced. Significance testing rewards reliable measurement[2] of beneficiary experience, not the act of oversampling itself.

Alignment was eligible to oversample and opted not to. Its dissatisfaction is simply results oriented. That is evident from its requested relief: Alignment asks this Court to "at least remand with instructions to recalculate the Star Ratings for contract H3815 without regard for significance testing." Pl. Opp'n (ECF No. 14) at 16. Alignment does not seek the removal of significance testing for any of its other contracts, including ones challenged in this litigation on other grounds. Its selective challenge underscores the results-oriented complaint. And its requested remedy to calculate Alignment's Star Ratings for one contract without regard to significance testing is asking this Court to change the codified Star Ratings methodology for Alignment only and only for one contract. Alignment's dissatisfaction with its performance on one contract does not render CMS's use of significance testing or its treatment of oversampling arbitrary or capricious.

IV. Alignment Has Not Shown that Assessing Reliability at the Composite Level Is Arbitrary and Capricious.

In its opening brief, Alignment challenged how CMS determines the reliability of composite measures (measures that combine related survey items that address a common aspect of a beneficiary's experience), alleging that CMS's method of how it computes the reliability of these measures does not comply with CMS's regulations. Pl. Mot. (ECF No. 11) at 32-36. And as explained in the government's opening brief, Alignment's argument misreads the regulation: contrary to Alignment's position, because the provision applies to the measure as a whole—not to each individual survey question within a composite measure—the regulation says nothing about

[2] Star Ratings for CAHPS measures take into consideration where a score on a CAHPS measure falls in the distribution of all contract scores, the reliability of the score, and whether the score is significantly different from the national mean score on the measure. *See e.g.*, AR 200-01 (Significance Testing, Reliability and Star Assignment).

discarding the data derived from individual questions within a composite measure based on their standalone reliability. *See* 42 C.F.R. §§ 422.166(a)(3), 422.186(a)(3). That threshold, applied to the measure as a whole, reflects an appropriate balance between excluding measures that are too noisy to be meaningful and retaining measures that reflect genuine representation of a contract's performance.

In its reply, Alignment argues that "CMS's policy of assessing reliability for composite measures at the measure level is arbitrary and capricious." Pl. Opp'n (ECF No. 14) at 16. Alignment asserts that "the reliability of a composite measure . . . cannot be greater than the reliability of its component parts." *Id.* at 18 (citing 42 C.F.R. § 422.162(a)). That assertion is incorrect and reflects a fundamental misunderstanding of how the composite measures at issue here function. There is no rule—statistical or regulatory—that prohibits a composite measure as a whole from having higher reliability than its individual components, as Alignment implicitly concedes. *See id.* Here, the challenged composite measures, by design, aggregate related and correlated components to improve overall reliability. This approach is grounded in accepted statistical principles and is not, as Alignment contends, *id.* at 17, a method to "mask" unreliable data. Alignment's assertion is not grounded in either regulation or statistical best practices and should be rejected. CMS's approach is grounded in sound statistical theory and is not arbitrary or capricious.

The academic sources cited by Alignment do not support its conclusion. While Alignment is correct that the academic literature is clear that composite measures "come with limitations," the limitations do not apply here. *See* Kara, P., *Composite measures of quality of health care: Evidence mapping of methodology and reporting*, 17 PLOS ONE 1, 2 (2022), https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0268320. Sources caution that

composite measures can mask differences between items, but CMS addresses this concern through careful design and topic-based grouping of related items within each CAHPS composite. *See* CAHPS Health Plan Survey, https://www.ahrq.gov/cahps/surveys-guidance/hp/index.html. CMS clusters questions related to related survey items focused on a particular topic such as "Getting Appointments and Care Quickly." [3] Alignment does not and cannot explain, much less show, why these individual questions that form these particular measures are not related. The CAHPS composite items are strongly correlated and designed to measure different dimensions of a single construct (e.g., timely access to care), precisely the condition under which compositing increases—not decreases—reliability.

Alignment's additional claim that the CAHPS composite items are answered by "distinct subpopulations" that generate "non-correlated answers," Pl. Opp'n (ECF No. 14) at 19, is factually incorrect. CAHPS surveys are administered to a random sample of eligible enrollees, and the same respondents are asked to complete all survey questions. *See* AR 0139 (Medicare Advantage and Prescription Drug Plan (MA & PDP) CAHPS® Survey Quality Assurance Protocols & Technical Specification). While certain questions may not apply to every respondent, or someone may prefer not to respond to a particular question resulting in occasional nonresponse, it is the same population for all questions on the CAHPS survey. These are not "statistically distinct and non-overlapping populations," contrary to Alignment's characterization. Pl. Opp'n (ECF No. 14) at 19.

Alignment's assertion that CMS's methodology reflects "garbage in, garbage out" is not only baseless but reflects a misunderstanding of both the data and the analytic framework. CMS's

[3] For example, for "Getting Appointment and Care Quickly" the following survey questions make up this composite measure: "In the last 6 months, when you needed care right away, how often did you get care as soon as you needed?"; "In the last 6 months, how often did you get an appointment for a check-up or routine care as soon as you needed?" AR 0399

composite measures are built from rigorously designed and tested items, drawn from consistent respondent populations, and analyzed using standard and validated statistical techniques. Alignment's attempt to discredit this well-established methodology lacks both factual and analytical support and should be rejected.

V. There Was No Unconstitutional Delegation to an Independent Review Entity

For the Star Ratings measures "Plan Makes Timely Decisions about Appeals" and "Reviewing Appeals Decisions," CMS collects data from an Independent Review Entity (here, Maximus), as the measures concern whether Alignment provides information to Maximus in a timely manner and how often an independent reviewer found the health plan's decision to deny coverage to be reasonable. To the extent the contract has concerns with the underlying data for the measures (that is, the decisions made by Maximus), the contract can raise those concerns to CMS for its own independent review as part of the plan preview process, part of the process for determining Star Ratings. That was exactly how the process played for Alignment. During the Star Rating's plan preview process, Alignment raised concerns about two cases decided by Maximus that it believed were handled incorrectly, or wrongly decided, and CMS reviewed its concerns, reviewed the decisions by Maximus, and ultimately decided to uphold Maximus's determinations and utilize the data in the relevant measures. Because CMS provided this level of review to the Independent Review Entity decisions as it pertains to Alignment's Star Ratings, Alignment's claims that the decisions run afoul of the non-delegation doctrine fail to withstand scrutiny.

An Independent Review Entity, like Maximus, is entitled to "operate as an aid" to agencies executing statutory responsibilities, so long as the agency retains final authority. *See Sunshine Anthracite Coal Co. v. Adkins*, 310 U.S. 381, 388, 399 (1940). Contrary to Alignment's assertion otherwise, Pl. Opp'n (ECF No. 14) 19-22, CMS did not abdicate its responsibility by merely accepting Maximus's determinations without review; rather, it reviewed Alignment's objections

and found no basis to overturn the Independent Review Entity's conclusions. Alignment raised concerns regarding two cases (Case 1-13226962526 and Case 1-12757246876), and the impact those cases had on the 2025 Star Ratings awarded to Alignment's contract H3443 for the "Plan Makes Timely Decisions about Appeals" and the "Reviewing Appeals Decisions" measures. *See generally* AR 0001-20. CMS notified Alignment that it would "look *again* at these cases since [CMS] want[s] to make sure they are accurate for the second plan preview." AR 0003 (emphasis added). Alignment itself acknowledged CMS's review, noting its appreciation of CMS's "efforts in investigating the appeals data[.]" AR 0001. After several email exchanges between CMS and Alignment, CMS ultimately determined that Maximus had correctly handled the two cases at issue, and that the Star Ratings appropriately reflected those determinations. AR 0001 (wherein CMS provided its rationale for "keeping the data as is for the []cases"). Alignment cannot explain, other than platitudes that CMS's decision amounts to a "rubber stamp" of Maximus's determinations, why this review and resulting determination by CMS that the Maximus decisions should be included in the data shows anything other than that CMS exercised its own judgment, and did not act with "blind deference" to the findings of the Independent Review Entity. Pl. Opp'n (ECF No. 14) at 21.

Alignment's only support for the idea that there was not a sufficient level of review is Alignment's assertion that CMS also provided Alignment with Maximus's description of the issues with the appeals. *Id.* But the Administrative Record supports that CMS/HHS retained the ability to review, modify, or otherwise exercise its discretion regarding Maximus's decision, in other words CMS retained ultimate authority. AR 0001-20. When Alignment raised its concerns, CMS did not tell Alignment that it was unable to review Maximus's determinations or that it could

not disturb Maximus's findings. To the contrary, the Administrative Record demonstrates that CMS was determined to ensure that Maximus got it right.

That satisfies any alleged concerns with the delegation of responsibilities to an Independent Review Entity for the purposes of the Star Ratings. *See Hight v. U. S. Dep't of Homeland Sec.,* No. 23-5273, 2025 WL 1272034, at *10 (D.C. Cir. May 2, 2025) (no private nondelegation problem when agency adopted recommendation of private entity not to certify pilot even when private entity engaged in "fact gathering" because agency "separately conducted its own review of the records"); *Nat'l Ass'n of Psychiatric Treatment Ctrs. for Child. v. Mendez*, 857 F. Supp. 85, 91 (D.D.C. 1994) (holding "nothing improper" with agency delegation to private contractor of "inspection and utilization review services" related to provider certification including individual adjudications of individual cases of appropriateness of medical care in particular cases when agency "makes all final decisions regarding certification").

Ultimately, Alignment is dissatisfied that CMS agreed with Maximus's determinations in the two appeals at issue and that the corresponding data is reflected in the Star Ratings. But it has not shown—and cannot show—that CMS failed to retain final authority over the Independent Review Entity as it pertains to Star Ratings.

VI. Maximus Correctly Decided the Specific Appeals at Issue.

Case 1-13226962526: Alignment contends that Case 1-13226962526 should be excluded from the calculation for "Plan Makes Timely Decisions About Appeals." The measure "Plan Makes Timely Decisions About Appeals" assesses if a plan sends case files to the Independent Review Entity, Maximus, for review in a timely manner. *See* AR 0020, 0038; *see also* 42 C.F.R.

§§ 422.582(i), 422.590(b)(2).[4] Here, Alignment failed to process a case file within the required 60-day timeframe. AR 0038 (explaining that Alignment "failed to process this case appropriately"). Rather than accept responsibility for this untimely submission, Alignment blames Maximus for not dismissing the case and argues that the case should not be considered in the measure.

In its reply, Alignment claims that because it should not have sent the case file to Maximus at all, it was also not required to do so within the deadline. Pl. Opp'n (ECF No. 14) at 11. But by submitting the case to Maximus, Alignment itself treated the case as though it had affirmed its initial decision on the merits. *See* Defs. Mot. (ECF No. 13) at 42. Maximus reviewed the case, affirmed the plan's determination, and appropriately noted the procedural deficiency of untimeliness in sending the file to Maximus. AR 0019–20. Alignment's failure to comply with the 60-day deadline—despite initiating Independent Review Entity review—is solely responsible for the resulting negative mark under the appeals timeliness measure. This was not, as Alignment suggests, an error in CMS's reasoning, but a straightforward consequence of Alignment's own actions under the governing metric.

Case 1-12757246876: Alignment contends that the record lacks substantial evidence to support Maximus's decision reversing its denial of coverage for outpatient lab services. Pl. Mot.

[4] In its motion, the government incorrectly cited 42 C.F.R. § 422.590(a)(2), instead of the correct provision, § 422.590(b)(2) as relevant for Case 1-13226962526. The applicable subsection depends on the basis of the underlying appeal, which in this case was a claim denial. Under § 422.590(b)(2), the plan had 60 days—not 30—to transmit the case file. That distinction does not alter the analysis, as Alignment still failed to meet the 60-day deadline (the relevant date that Alignment received notification of the providers appeal was Feb 8, 2023; Alignment forwarded the appeal to Maximus on September 6, 2023).

(ECF No. 11) at 44. But as explained in the government's opening brief, the record amply supports Maximus's determination. *See* Defs. Mot. (ECF No. 13) at 43–44.

In its reply, Alignment continues to rely on the same flawed argument it advanced in its opening brief: that the inclusion of denial code "MISSING PRIMARY INSURANCE CARRIER(S) [EOB]" on a remittance advice form shows that it attempted coordination of benefits. Pl. Mot. (ECF No. 11) at 44-45; Pl. Opp'n (ECF No. 14) at 24. But Alignment cites no authority, regulation, or evidence suggesting that inclusion of this code alone satisfies the obligation to coordinate coverage. *See, e.g.,* AR 0002, 0018, 0038. It does not demonstrate or even allege that it contacted the enrollee's primary insurer or obtained a denial from that insurer. In fact, the record shows that Alignment acknowledged the insufficiency of the code and represented that it had changed its procedures to improve communication going forward. AR 0018.

This is more than enough to meet the substantial evidence standard. A reasonable mind could conclude from the record that Alignment did not adequately coordinate with the enrollee's other insurer, and no more is required. Alignment's reply brief treads no new ground, and for the same reasons set forth in the government's opening brief, its challenge fails.

VII. If Required, Remand Is the Sole Appropriate Remedy

Even if this Court were to find that Alignment's Star Ratings for the specified contracts were somehow defective, the proper remedy would be to set aside the agency's action and remand to the agency. As the D.C. Circuit has held, if "the record before the agency does not support the agency action" the appropriate remedy is for the court "to remand to the agency for additional investigation or explanation," except in "exceptional circumstances." *Cty. of Los Angeles v. Shalala*, 192 F.3d 1005, 1023 (D.C. Cir. 1999) (quoting *Fla. Power & Light Co. v. Lorion*, 470 U.S. 729, 744 (1975)). No exceptional circumstances exist in this case. Thus, if the Court grants

summary judgment to Plaintiff, the proper remedy is to remand to the Secretary for further proceedings consistent with the Court's opinion.

CONCLUSION

For the foregoing reasons, and for those previously discussed in Defendants' cross-motion for summary judgment, Defendants respectfully request that the Court enter summary judgment in their favor.

Dated: May 12, 2025

Respectfully submitted,

EDWARD R. MARTIN, JR., D.C. Bar #481866
United States Attorney

By: */s/ M. Jared Littman*
M. JARED LITTMAN
Assistant United States Attorney
601 D Street, NW
Washington, DC 20530
(202) 252-2523

Attorneys for the United States of America