IN THE 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.

Case No. 1:25-cv-00074-CRC

MEMORANDUM IN OPPOSITION TO DEFENDANTS' CROSS-MOTION FOR SUMMARY JUDGMENT AND REPLY IN SUPPORT OF PLAINTIFF'S MOTION FOR SUMMARY JUDGMENT

TABLE OF CONTENTS

TABLE OF AUTHORITIES*

Cases

* Authorities marked with an asterisk are those on which we chiefly rely.

Cases—continued

Statutes, Rules, and Regulations

Other Authorities

INTRODUCTION AND SUMMARY OF ARGUMENT

Although the Star Ratings system may be complex and cumbersome, this lawsuit is not. Alignment's claims are simple, and the grounds for relief straightforward.

A. Consider first Alignment's claim that the compression of the Star Ratings cut points resulting from the Tukey Outlier Deletion Rule requires adjustments to the minimum denominator requirement for each measure. The agency has acknowledged that cut-point compression can "decreas[e] . . . variability across contracts" and thus "mak[e] [a] measure unreliable" for purposes of the Star Ratings. 83 Fed. Reg. 16440, 16535 (Apr. 16, 2018). It has further admitted that "minimum sample size and/or denominator requirements" may be necessary "to ensure measure data are reliable" and the Star Ratings system accurately reflects plan performance when cut points are compressed. *See* 87 Fed. Reg. 27704, 27765-27766 (May 9, 2022). Indeed, that is why the agency adopted a universal minimum denominator rule (it requires 11 data points per measure) in the first place.

But CMS has refused to adjust the minimum denominator requirement measure-by-measure to ensure that each contract can achieve each Star Rating on each measure. It does not deny that many contracts are therefore logically precluded from achieving middle-of-the-range Star Ratings on many measures because their denominators are too small—even when a small denominator itself may indicate higher quality (*e.g.*, fewer appeals). The agency's only response is to parrot back (Br. 15, 19-20, 23) the unexplained, unsupported assertion that it offered in the rulemaking: "Closer cut points do not necessarily imply lower reliability or lessen the ability to distinguish between contracts." 88 Fed. Reg. 22120, 22297 (Apr. 12, 2023). That is not enough. CMS must "offer[] 'a satisfactory explanation for its action[s]'" and not "simply ignore 'an important aspect of the problem'" with *ipse dixits*. *Ohio v. EPA*, 603 U.S. 279, 292-293 (2024).

B. Take next Alignment's claim that the CAHPS surveys for contracts H3443 and H3815 were administered unlawfully. On this point, CMS does not dispute what the administrative record now plainly confirms: The CMS-approved survey vendor that administered Alignment's CAHPS surveys provided English-language materials to scores of enrollees designated as Spanish speakers, resulting in lower response rates. This was a clearcut violation of CMS's rules and requirements. *See* Opening Br. 19-20, 37.

CMS's responses are stunning. First, it insists (Br. 31) that it "undertook a careful review" of this issue and "reasonably found" that there were no survey administration errors. That is simply false. In fact, CMS looked only at the *results* of the surveys, which it figured looked close enough and thus warranted no "further validation." AR48. It did not even pull, let alone review, the underlying data to confirm Alignment's concern that Spanish speakers had not received Spanish-language surveys, as required by law. And there is no debating that if it had, it would have uncovered the error.

Second, CMS says (at 33) that any survey administration error that may occur is an issue between the MA plan and its survey vendor; it is not "within CMS's direct oversight or control." That is a red herring. Alignment is not seeking to hold CMS responsible for the survey error or accusing it of inadequate supervision. The point is simply that when a demonstrable error occurs—when an MA plan shows that its CAHPS survey is conducted in a manner that plainly violates CMS's own rules and requirements—the data must be discarded, and the survey results cannot factor into the plan's Star Ratings. To say otherwise would treat plans inconsistently—to subject them to variable rules and procedures. That is a clear violation of the APA: An "unexplained inconsistency" in "agency practice" with respect to similarly situated cases is the hallmark of arbitrary and capricious decision-making. *Encino Motorcars, LLC v. Navarro*, 579 U.S. 211, 222 (2016)).

C. That same principle explains the problem with CMS's consideration of CAHPS survey oversampling. We showed in the opening brief (at 39-40) that oversampling gives an irrational advantage to larger plans that can and do arrange for oversampling; plans that do not oversample will sometimes receive lower Star Ratings *because* they do not oversample, and for no other reason. This case proves the point: On at least two CAHPS measures, contract H3815 earned a higher mean score and the same reliability assessment as other contracts, but it received a lower measure-level Star Rating because Alignment did not engage in oversampling. Yet "[a]gency action is . . . arbitrary and capricious if it 'offer[s] insufficient reasons for treating similar situations differently.'" *California Communities Against Toxics v. EPA*, 928 F.3d 1041, 1057 (D.C. Cir. 2019) (quoting *Transactive Corp. v. United States*, 91 F.3d 232, 237 (D.C. Cir. 1996)).

CMS does not deny that oversampling confers an arbitrary advantage generally, or that Alignment's decision not to oversample for contract H3815 resulted in a lower Star Rating for the contract. Its only response is to say (at 36) that "Alignment could have chosen to oversample and elected not to pursue that option." That, too, is a red herring. An MA plan's decision not to oversample is its own. (Conducting the CAHPS surveys is expensive, and oversampling adds substantial additional cost.) The point is that the Star Ratings system is intended to "accurately . . . reflect true performance" based on objective indicators of plan quality. 83 Fed. Reg. at 16519. And the ability and willingness to engage in oversampling does not *in any way* reflect plan quality. CMS neither denies that basic observation nor attempts to justify its approach to oversampling in light of it.

D. We showed also that CMS uses a flawed approach to reliability testing for composite measures. In particular, it uses composite reliability scores that at times mask the unreliability of a contract's individual measure-level inputs. *See* Opening Br. 32-36. It is

common sense that use of "underlying data [that] is unreliable necessarily means that" calculations relying on that data are "equally if not more unreliable" because of the "garbage in, garbage out" rule. *Oceana, Inc. v. Ross*, 290 F. Supp. 3d 73, 47 (D.D.C. 2018). It follows that if any of the underlying items in a composite measure have a reliability score below CMS's regulatory threshold of 0.60, the whole measure is unreliable.

CMS rejoins (at 27) that reliability increases with greater response numbers, and "[t]he total number of respondents providing information towards a composite or multi-item measure is usually greater than the number responding to any one item," producing more reliable results. But the academic source that it cites for that proposition is clear that "composite measures can be sensitive to the methodology that has been used" and "may be misused." Pinar Kara et al., *Composite measures of quality of health care: Evidence mapping of methodology and reporting*, 17 PLOS ONE 1, 2 (2022), https://perma.cc/52R2-VL9M. That same academic article cautions that, "if not constructed in a methodologically sound way," the results obtained from composite measures "may not be reliable." *Id.*

And according to the source cited by the authors for that proposition, composite measures should test the same respondent population with respect to correlated, binary data points. *See* Peter C. Austin et al., *Ranking hospital performance based on individual indicators: can we increase reliability by creating composite indicators?*, 19 BMC Medical Research Methodology 1, 8 (2019), https://perma.cc/36ZU-E4QH. But as we explained in the opening brief (at 34-35), that is not what CAHPS composite measures do. Rather, they test distinct respondent populations along non-correlated lines of inquiry—a point that CMS does not deny. In that circumstance, the rule holds: garbage in, garbage out.

E. We demonstrated in the opening brief (at 41-42) that appeal review by Maximus, CMS's "independent review entity" under the MA program, reflects an unconstitutional

delegation of regulatory authority. Maximus's decisions are not subject to supervision or further review of any kind when Maximus determines that coverage should not be denied. Decisions against a MA organization (MAO) are final, binding, and entirely unreviewable; CMS does not retain *any* authority, let alone "ultimate authority," to "approve, disapprove, or modify" Maximus's decisions in such cases. *Alpine Securities Corp. v. FINRA*, 121 F.4th 1314, 1325 (D.C. Cir. 2024).

CMS denies this (at 38), insisting that "[f]or purposes of Star Ratings, if a Medicare Advantage contract believes that an Independent Review Entity determination was incorrect and that it adversely impacted its Star Ratings, it has opportunities to raise those concerns during the Star Ratings plan preview periods." But that is another non-answer. As we showed in the opening brief (at 42), CMS effectively delegates unsupervised authority to resolve Star Ratings objections, too. The facts of this case again prove the point: When Alignment raised concerns about the handling of two appeals, CMS simply "shared" Alignment's objection with Maximus and then simply accepted Maximus's decision that the appeals were properly resolved. *See* AR3-4. The process involved no more supervision than did the original appeal review decision.

F. Finally, the opening brief showed (at 43-45) that Maximus incorrectly resolved two of Alignment's appeals, resulting in lower Star Ratings for two Alignment contracts. CMS rejoins (at 41-42) that Alignment erred in sending case 1-13226962526 to Maximus in the first place and insists (at 43-44) that the reversal of case 1-12757246876 was supported by substantial evidence. Neither of those arguments has merit. Nor does its contention (at 44) that Alignment should have sought administrative reconsideration. Alignment is not seeking to overturn the appeal decisions, but instead merely to exclude them from improper consideration in the Star Ratings calculations, which is a distinct issue.

In the final analysis, CMS's opposition brief confirms rather than refutes the need for immediate judicial relief. CMS does not deny that cut-point compression produces arbitrary and unfair results for contracts with small measure-level denominators; that Alignment's CAHPS surveys were administrated in violation of CMS's regulations concerning survey administration to Spanish-speaking enrollees; or that oversampling arbitrarily manipulates Star Ratings scores in a way entirely unreflective of plan quality. Beyond that, it does not persuasively refute our showing that the agency's approach to composite scores masks the unreliability of question-level data; that Congress and CMS have unlawfully delegated huge swaths of regulatory authority to an unaccountable, profit-driven private entity; or that that entity has mishandled two of Alignment's appeals. Any one of these legal errors, taken alone, would be troubling. Taken together, they paint a picture of shocking and systemic violations of the APA.

Against this background, the Court should deny CMS's cross motion, grant summary judgment to Alignment, and declare that the Tukey Outlier Deletion Rule is arbitrary and capricious as applied to Alignment without appropriate denominator adjustments; CMS acted arbitrarily and capriciously and contrary to law by including in Alignment's 2025 Star Ratings the CAHPS survey data infected by a demonstrated administration error; CMS's allowance for oversampling arbitrarily and capriciously disadvantages plans that do not engage in oversampling; CMS's inclusion of "very low reliability" data in composite CAHPS survey measures is arbitrary, capricious and contrary to law; CMS has unconstitutionally delegated its regulatory authority to Maximus; and Maximus's final decisions for Case Nos. 1-13226962526 and 1-12757246876 were arbitrary, capricious, and contrary to law. It then should vacate Alignment's 2025 Star Ratings and remand with instructions for CMS to recalculate the Ratings for all Alignment's impacted contracts.

ARGUMENT

A. CMS acted arbitrarily and capriciously by refusing to adjust minimum denominators to ensure each contract can earn each rating for each measure

Alignment challenges CMS's application of the Tukey Outlier Deletion Rule to its 2025 Star Ratings for non-CAHPS measures without adjusting the minimum denominator thresholds. To derive measure-level Star Ratings for non-CAHPS measures, CMS breaks industry-wide scores into five discrete clusters, taking the "dividing lines" between the clusters as the "cut points" establishing the numerical thresholds at which a measure "results in a higher or lower star rating." *SCAN Health Plan v. HHS*, 2024 WL 2815789, at *2 (D.D.C. June 3, 2024); *see* CMS, *Medicare 2025 Part C & D Star Ratings Technical Notes* 143 (Oct. 3, 2024) (*2025 Technical Notes*) (AR1095). Under the Tukey Outlier Deletion Rule, before CMS partitions the industry-wide scores into clusters, it first identifies and deletes "extreme outliers" from the dataset to prevent them from "influencing cut point determinations." 85 Fed. Reg. 33796, 33833 (June 2, 2020).

As Alignment explained in its opening brief (at 28-33), removal of outliers is certain to compress measure-level cut-points. As a result, some smaller contracts with relatively few data points will be logically incapable of achieving each rating on the measure-level Star Ratings. Failure to address cut-point compression by adjusting the minimum denominator requirement—that is, failure to exclude from a given measure any plan that has too few data points to achieve each Star Rating for the measure—makes the Star Ratings unreliable as an indicator of plan quality and performance.

Before this Court, CMS disputes (at 20-21) that Tukey outlier deletion leads to cut-point compression, noting that when the agency analyzed the effects of outlier deletion during the rulemaking, it found that this statistical method had limited impact on cut points

for the majority of measures. *See* AR103-104; 88 Fed. Reg. at 22296. But the absence of cut-point compression for CMS's cherry-picked examples reflects only that the data for those measures ordinarily have few (or no) outliers. *See* 88 Fed. Reg. at 22296. Where a measure's dataset *does* include outlier scores that must be removed under the Tukey Rule, however, the deletion will necessarily compress the cut points by removing outliers that would otherwise draw the cut points further apart. This is simple mathematics: When scores registering especially "low or high performance" are removed (*id*. at 22297), the remaining scores will exhibit a narrower range of values, which in turn will mean that the cut points dividing score clusters *within* this narrower range will be closer to each other than they would be with the outliers present.

CMS did not truly dispute this commonsense point during the rulemaking. When commenters raised concerns that outlier deletion would "move [cut points] closer together, decreasing reliability," CMS accepted that outlier deletion would result in cut-point compression. *Id*. The agency's only response was the *ipse dixit* that "[c]loser cut points do not necessarily imply lower reliability or lessen the ability to distinguish between contracts." *Id.* Then, in its email correspondence with Alignment during the plan preview periods, CMS parroted back that same unexplained assertion, insisting that cut-point compression does not "necessarily" mean "lower reliability" or lesser "ability to distinguish between contracts." AR103-104. While CMS repeatedly quotes this same language in its brief before the Court (at 15, 19-20, 23), CMS has *never once* offered any reasoning or evidence to support its unexplained assertion.

Wholly apart from the rulemaking itself, Alignment demonstrated that, with outliers removed, the spread from one cut point to the next does indeed become compressed. Opening Br. 29-31. Thus, once the Tukey Rule is applied, even very slight differences in

underlying measure scores can produce huge swings in the Star Ratings awarded for that measure. For smaller plans with low denominators, this is especially problematic because (again) it becomes mathematically impossible for the plan to earn Star Ratings in the middle range. *See* AR104-105; Opening Br. 29. In this way, the cut-point compression associated with outlier deletion undermines what CMS itself has called the "primary goal in setting cut points," namely "to disaggregate the distribution of scores into discrete categories such that each grouping accurately reflects true performance." 88 Fed. Reg. at 22295.

Consistent with Alignment's comments in the rulemaking process, we showed in the opening brief (at 29-30) how Tukey outlier deletion thus "lower[s] reliability" and reduces the Star Ratings system's "ability to distinguish between contracts." 88 Fed. Reg. at 22297. Under the hypothetical described on page 29 of the opening brief, for example, a plan with 100 enrollee appeals that handles 10 appeals untimely receives 2.0 Stars on the "Plan Makes Timely Decisions about Appeals" measure. Another plan with 100 appeals that handles 5 appeals untimely receives 4.0 Stars. But a third 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. It is logically incapable of earning 4.0 Stars or 3.0 Stars.

A system that compares small-denominator plans incapable of earning each Star level with larger plans that are factually eligible to earn the mid-range Star Ratings introduces irrational differences in the treatment of MA contracts. *See California Communities*, 928 F.3d at 1057 (it is "arbitrary and capricious" to treat "similar situations differently" unless the agency explains itself). By doing so, it undercuts the ability of the Star Ratings system to meaningfully compare and distinguish contract performance. 88 Fed. Reg. at 22297. CMS offers no response to this demonstration, again simply repeating

(at Br. 15, 19-20, 23) its unadorned assertion that cut-point compression doesn't "necessarily" produce "lower reliability" results.

The agency's dismissal of these concerns also violates CMS's obligation to "offer[] 'a satisfactory explanation for its action[s]'" and not to "simply ignore 'an important aspect of the problem.'" *Ohio*, 603 U.S. at 292-293 (quoting *Motor Vehicle Manufacturers Association of United States, Inc. v. State Farm Mutual Automobile Insurance Co.*, 463 U.S. 29, 43 (1983)). It is no answer to say, as CMS does (at 24), that there is no specific statutory "guarantee that each contract can achieve each star level." Our argument is that the absence of such a guarantee renders the Star Ratings system arbitrary and capricious for small-denominator contracts, not that it is inconsistent with the statutory text. Again, cut-point compression that is not addressed with an adjusted denominator requirement produces Star Ratings that do not accurately reflect and differentiate the quality and performance of smaller plans. *See* 5 U.S.C. § 706(2)(A).

CMS responds (at 25) that increasing denominator thresholds to mitigate the effects of cut-point compression would be unacceptable because it would "eras[e] valid high-performing results from smaller contracts," thereby penalizing certain high-performing contracts simply because of their small size. But that outcome is both unavoidable and evidently acceptable, given that CMS already implements a uniform minimum denominator requirement of 11. And it does so for the same reason that adjusting denominator thresholds to account for the effects of the Tukey Rule is necessary here: "Star Ratings measures have minimum sample size and/or denominator requirements *to ensure measure data are reliable*." 87 Fed. Reg. at 27765-27766 (emphasis added); *see also* AR105 (CMS recognizing that minimum denominator requirements serve to "obtain a sufficient level of reliability").

As CMS already has recognized, minimum denominator requirements are necessary to obtain "sufficient," not necessarily maximum, reliability. AR105. Implementation of the Tukey Rule without corresponding adjustments in the minimum denominator requirements does not ensure "sufficient" reliability in the measure-level Star Ratings. Instead, it penalizes smaller plans with smaller denominators by making it impossible for them to achieve middle-of-the-range Star Ratings. When some plans are mathematically incapable of attaining certain Star Ratings levels, the assigned ratings cannot be said to reflect comparative plan quality and performance.

Moreover, to say that some plans may achieve 4.0 Stars or 3.0 Stars, but other plans may not, simply because they are smaller, is to violate the bedrock principle of administrative law that an agency must "treat like cases alike." *Consolidated Edison Company of New York, Inc. v. FERC*, 45 F.4th 265, 279 (D.C. Cir. 2022) (quoting *Westar Energy, Inc. v. FERC*, 473 F.3d 1239, 1241 (D.C. Cir. 2007)). If it is logically possible for larger plans to earn 4.0 Stars or 3.0 Stars, it must be logically possible for *all* plans to do so. The Court should therefore vacate Alignment's 2025 Star Ratings for each contract with a measure-level denominator too small for the contract to earn each of the five Star Ratings on that measure, including at minimum contracts H3443 and H9686. It should order recalculation on remand, omitting the impacted measures.

B. Alignment's CAHPS measure scores should be set aside in light of the clear error in survey administration

As Alignment showed in the opening brief (at 19-22), its anticipated CAHPS-survey response rate for Spanish speakers was not achieved in 2024. When Alignment alerted CMS during the plan preview periods to the unexpected and unexplained decline in Spanish-language responses—and its suspicion that its survey vendor may have sent English-

language surveys to Spanish-speaking enrollees—CMS refused to meaningfully investigate this issue and denied relief. That was arbitrary and capricious.

Even now, CMS continues to insist (at 31) that it "undertook a careful review" and "reasonably found" on the basis of that review that there were no survey administration errors that undermined the reliability of the CAHPS survey data for contracts H3815 and H3443. But the record reveals the opposite. The agency's purported "review" involved scanning the survey *results* (which it thought looked good enough) and reviewing an attestation from Alignment's 2024 CAHPS survey vendor (DataStat, Inc.) that it had followed the proper procedure and "used the language preference data shared by [Alignment]." AR53. CMS thus did not really "review" anything—it simply asked the survey vendor to self-certify that its administration of the CAHPS survey complied with the required procedures. The record demonstrates that CMS did not attempt to verify independently that DataStat's attestation was accurate.

In fact, the record shows that CMS affirmatively *refused* to conduct an independent review, concluding that the survey results for contracts H3815 and H3443 were close enough to what CMS expected to see that "[n]o further validation [was] needed." AR48. Specifically, CMS determined that the sample was representative in terms of predicted Spanish preference because the proportion of sampled enrollees with predicted Spanish preference tracked the proportion in the broader CAHPS-eligible enrollee population. AR61. CMS also assumed that the proportion of Spanish-speaking respondents was in line with expectations given what is "common for patient surveys" (*id.*) even though it was lower compared to the proportion among sampled enrollees for Alignment's contracts. And while CMS itself noted that some sampled Spanish-speaking enrollees did not respond in Spanish, it dismissed this observation as "typical." *Id.*

Setting aside whether this was a sufficient response to Alignment's concerns (it assuredly was not), there is no longer any dispute that the surveys *were* improperly administered, as Alignment all along suspected. As we demonstrated in the opening brief (at 22), data that was uniquely in CMS's possession shows that DataStat sent English-language questionnaires to scores of sampled enrollees whom Alignment had designated as Spanish-speaking. This was a clearcut violation of CMS's rules for CAHPS survey administration. *See* CMS, *Quality Assurance Protocols & Technical Specifications* 50 (Nov. 2023) (*Protocols & Specifications*) (AR172); *see also* AR173 (a CMS-approved survey vendor must, "at the request of the contract," send Spanish-language questionnaires to enrollees identified as having a "language preference" for Spanish).

CMS attempts to excuse its failure to meaningfully investigate the suspected errors in survey administration by shifting the blame to DataStat—and indeed to Alignment itself. CMS insists (at 33) that whether a survey vendor sends enrollees the CAHPS survey in a language consistent with the plan's language designations concerns the plan's "coordination with its own survey vendor" and is not a matter "that falls within CMS's direct oversight or control." But that misses the point. When an MAO like Alignment plausibly asserts that that its CAHPS surveys were administered in a manner inconsistent with CMS's own rules and requirements, the agency has an obligation to confirm whether or not the concern is well founded. And when data solely in CMS's possession shows beyond all dispute that the survey was, indeed, *not* administered consistent with the agency's own rules and requirements, it must discard the CAHPS survey data in calculating the Star Ratings for the affected contracts. To do otherwise is to administer the Star Ratings system using different rules for similarly situated entities, again declining to "treat like cases alike." *Consolidated Edison*, 45 F.4th at 279.

CMS is required by the APA to uniformly apply the rule that "Spanish language questionnaires must be made available to all Spanish-speaking enrollees." *Protocols & Specifications* 50 (AR172); *see also* AR173. As the opening brief explained (at 38), when an MAO like Alignment brings a serious survey administration error to the CMS's attention, it does not suffice for the agency simply to say "oh well, not our fault" and count the survey results anyway. Nor does it suffice for the agency to claim that it verified that the survey was properly administered when, as a matter of simple and irrefutable fact, *it was not*. CMS never directly addresses these very basic and commonsense points.

Because DataStat did not administer the CAHPS survey for contracts H3443 and H3815 in a manner consistent with Alignment's language designations or CMS rules and requirements, the response rates and survey scores for contracts H3443 and H3815 were impacted. AR54-55, 72-73. CMS violated its obligation to address the substantive problems with the survey by discarding the results. The 2025 Star Ratings for H3443 and H3815 therefore must be set aside and remanded for recalculation without considering the defective CAHPS survey results.

C. Allowing oversampling for the CAHPS survey arbitrarily and capriciously disadvantages small plans

There is yet another flaw in the calculation of the Star Ratings for contract H3815: Alignment's decision not to oversample the CAHPS survey for that contract, which has nothing to do with plan quality, arbitrarily resulted in a lower Star Rating.

While the standard sample size for the CAHPS survey is 800 enrollees for each contract (AR78), CMS allows contracts to "request an increase in sample size" for the CAHPS survey—a practice known as "oversampling" (*Protocols & Specifications* 18

(AR140)). Contracts are allowed to oversample, at their own expense, provided "there is sufficient eligible enrollee volume to support additional sampl[ing]." *Id.*

As we explained (Opening Br. 39-40), a contract that oversamples gains an advantage because oversampling affects the significance testing methodology that CMS uses to derive measure-level Star Ratings from the raw scores of CAHPS measures. *See* 42 C.F.R. § 422.166(a)(3); AR962. Under the significance-testing component of the Star Ratings measure calculations, a contract can receive a higher Star Rating if its measure score has greater "statistical significance" relative to other plans. AR962. Since the statistical significance of a contract's score increases with the number of survey respondents, oversampling allows a contract to boost the significance of its CAHPS scores to achieve higher Star Ratings overall. More simply put: The practice of oversampling inflates the Star Ratings of contracts that are willing and able to oversample, even though oversampling bears no relation to plan quality or performance. Disadvantaging plans that choose not to oversample, or that lack sufficient enrollee volume to oversample, amounts to "treating [relevantly] similar situations differently," which is "arbitrary and capricious" unless the agency can offer a sufficient explanation. *See California Communities*, 928 F.3d at 1057.

CMS offers no explanation—literally none—to justify the role of oversampling in the Star Ratings system. It instead notes (at 36) simply that the "option to oversample is available to any contract that wishes to obtain more granular data for quality improvement purposes." True enough. But that once again misses the point, which is that oversampling is optional and not required; and those plans with the enrollees and financial resources to engage in oversampling obtain an arbitrary advantage in the Star Ratings calculations, an advantage having nothing to do with plan quality. "Agency action is . . . arbitrary and

capricious if it 'offer[s] insufficient reasons for treating similar situations differently.'" *California Communities*, 928 F.3d at 1057 (quoting *Transactive Corp.*, 91 F.3d at 237).

Perhaps aware of the indefensibility of allowing oversampling, CMS attempts to reframe Alignment's claim as a complaint of procedural unfairness. There was no procedural unfairness, CMS asserts (at 37), because Alignment could have oversampled contract H3815 but chose not to. That misapprehends Alignment's claim, which is not that oversampling is procedurally defective. Rather, Alignment contends that oversampling distorts CMS's significance metric, such that the calculation of Star Ratings for CAHPS measures favors contracts based on factors such as their number of enrollees and their decision to oversample, which have nothing to do with quality and performance. Agency action is arbitrary and capricious not only if it is based on unreliable data, but also if it "frustrate[s] the policy that Congress sought to implement." *Mylan Laboratories Ltd. v. FDA*, 910 F. Supp. 2d 299 (D.D.C. 2012) (quoting *Beaty v. FDA*, 853 F. Supp. 2d 30, 41 (D.D.C. 2012)). That is the case here. That Alignment could have obtained the same arbitrary advantage as other plans, but did not, is just another bright red herring.

If the Court agrees that the CAHPS survey for contract H3815 must be discarded in light of the demonstrated Spanish-language administration error, it need not address this issue. But if, against the evidence, the Court upholds CMS's decision to consider the misadministered CAHPS surveys, it should at least remand with instructions to recalculate the Star Ratings for contract H3815 without regard for significance testing.

D. CMS's policy of assessing reliability for composite measures at the measure level is arbitrary and capricious

Some CAHPS measures are composite measures, which combine survey items into a single measure. *See* Opening Br. 33. Specifically, "answers to questions about the same

topic are combined to form composites" for "scoring . . . purposes." *Protocols & Specifications* 7, 73 (AR129, 195). For "[m]any" of the composite measures, some of the component items are "screener questions" that serve to filter the respondent population, "allow[ing] only those enrollees for whom the item is relevant" to provide responses to the subsequent questions. *Id.* at 7, 27 (AR129, 149). CMS's policy is to assess the reliability of composite measures at the measure-level, without regard to question-level reliability.

As the opening brief explained (at 32-36), CMS's composite reliability calculations mask unreliable question-level results, allowing composite measures with "very low reliability" (*i.e.*, with reliability scores below 0.60) to be considered in the Star Ratings system when CMS regulations require such measures to be suppressed. *See* 42 C.F.R. §§ 422.166(a)(3), 422.186(a)(3); *2025 Technical Notes* 189 (AR1141). It is fundamental, however, that an agency's calculations "are only as reputable as the inputs upon which they rely to produce their" results. *Mississippi v. EPA*, 744 F.3d 1334, 1352 (D.C. Cir. 2013). Simply put, where any of the underlying question-level data in a composite measure have a reliability score of less than 0.60, the reliability of the measure cannot be 0.60 or more. A composite measure is only as reliable as its least reliable input.

CMS argues (at 27) that assessing the reliability of composite measures at the measure-level alone complies with its own regulatory definition of "very low reliability" measures because the requirement concerns measure-level reliability only, not the "standalone reliability" of a composite measure's constituent questions. *See* 42 C.F.R. §§ 422.166(a)(3), 422.186(a)(3) ("[N]o measure Star Rating is produced if the reliability of a CAHPS measure is less than 0.60."). The agency insists (at 27) that the regulation is silent as to whether a composite measure must be suppressed if the reliability of its component questions falls below 0.60. But CMS once again misunderstands the point,

which is that the reliability of a composite measure—"the fraction of the variation among the observed measure values that is due to real differences in quality ('signal') rather than random variation ('noise')" (42 C.F.R. § 422.162(a))—cannot be greater than the reliability of its component parts.

CMS disagrees (at 26), insisting that, by "aggregating responses across multiple related items," the composite Star Ratings measures can achieve "higher reliability than their component parts." Although that may be so in some circumstances, it is not so here. That follows from the very sources that CMS cites in its brief (at 27). CMS's academic authority states in the sentence following the one cited by CMS that using composite measures "come with limitations," including that "[d]ifferences and relationships between individual indicators may be masked and information regarding specific aspect(s) of performance can be lost." Kara et al., *Composite measures*, *supra*, at 2. As one of the sources cited by the authors explains, moreover, combining indicators in a composite measure improves reliability (sometimes called "rankability") only when the prompts are binary (i.e., "yes" or "no" questions) and produce correlated answers. Austin et al., *Ranking hospital performance*, *supra*, at 8. But "combining binary outcomes that are negatively correlated, uncorrelated or only weakly correlated, into an ordinal outcome *decreases* rankability." *Id.* (emphasis added).

That is the case here, because the CAHPS survey prompts are often continual and not binary, and the results are logically *un*correlated. *See* Opening Br. 34-35. For example, the CAHPS "Clinician & Group Survey," which includes six "Care Coordination" measure questions, comprises 33 questions in total. *See* CAHPS Clinician & Group Survey, https://perma.cc/ELX3-BUXK. Among those 33 questions, most are not binary; some have as many as five, six, seven—even ten—possible answers. What's more, nine of the 33

questions are "screener" questions that bear no substantive relation to the remaining 24 questions and therefore cannot bear a meaningful correlation. The presence of the screener questions also means that the respondent populations across all 33 questions are *by definition* variable. This simply is not a case in which a constant population is answering a series of correlated yes-or-no questions. And where the underlying questions in a composite measure draw responses from distinct subpopulations and produce non-correlated answers—*as here*—CMS's own sources confirm that aggregation *decreases* rather than increases reliability. Austin et al., *Ranking hospital performance*, *supra*, at 8; Kara, *Composite measures*, *supra*, at 2.

Generating a single reliability score by blending the responses from statistically distinct and non-overlapping populations is methodologically unsound and simply masks the presence of unreliable inputs. Simply put: garbage in, garbage out.

E. Congress's and CMS's delegation of authority to Maximus to review appeals is an unconstitutional delegation

As we explained in the opening brief (at 15-16), MAOs must, by statute, establish procedures to "provide for reconsideration" of adverse coverage determinations "upon request by the enrollee involved." 42 U.S.C. § 1395w-22(g). Congress has directed CMS to contract with an independent review entity (here, Maximus) to "review and resolve in a timely manner reconsiderations that affirm denial of coverage, in whole or in part." 42 U.S.C. § 1395w-22(g)(4).

We showed (Opening Br. 41-42) that Congress's and CMS's delegation of regulatory authority to Maximus is a plain violation of the non-delegation doctrine. "Federal law-makers cannot delegate regulatory authority to a private entity," which is "delegation in its most obnoxious form.'" *Association of American Railroads v. U.S. Department of Trans-*

portation, 721 F.3d 666, 670 (D.C. Cir. 2013), *vacated and remanded on other grounds*, 575 U.S. 43 (2015) (quoting *Carter v. Carter Coal Co.*, 298 U.S. 238, 311 (1936)). Thus, for the "delegation of governmental authority to a private entity to be constitutional, the private entity must act only 'as an aid' to an accountable government agency that retains the ultimate authority to 'approve, disapprove, or modify' the private entity's actions and decisions on delegated matters." *Alpine Securities*, 121 F.4th at 1325 (quoting *Sunshine Anthracite Coal Co. v. Adkins*, 310 U.S. 381, 388, 399 (1940)) (cleaned up).

That is not how CMS uses Maximus. First, when Maximus resolves an MAO's reconsideration decision and rules against the MAO, it does so entirely independently of CMS. The plan has no right at all to administrative review by the agency—Maximus's decision is final and binding. *See* 42 C.F.R. §§ 422.600, 422.616. Since there is no ultimate agency oversight over Maximus when it rules against an MAO, the independent review entity (IRE) review scheme cannot be squared with the private nondelegation doctrine.

This unconstitutional scheme of independent review affects the Star Ratings program because two measures track how plans perform in complying with the scheme. The "Plan Makes Timely Decisions about Appeals" measure tracks "how fast a plan sends information for an independent review." *2025 Technical Notes* 75 (AR1027). That is, it accounts for the timeliness with which plans forward their decisions in enrollee claim appeals for further review by Maximus. And the "Reviewing Appeals Decisions" measure tracks "how often an independent reviewer found the health plan's decision to deny coverage to be reasonable." *2025 Technical Notes* 76 (AR1028).

Maximus's handling of Case 1-13226962526 and Case 1-12757246876, and consequently the 2025 Star Ratings awarded to Alignment's contract H3443 for the "Plan Makes Timely Decisions about Appeals" and the "Reviewing Appeals Decisions"

measures, were adversely affected by the unconstitutional delegation. Alignment accordingly sought vacatur of Alignment's 2025 Star Ratings for contract H3443 and remand to recalculate after appropriate agency review of Maximus's decisions.

CMS now argues (at 37-38) that, "[f]or purposes of Star Ratings," Maximus's decisions are in fact subject to ultimate agency oversight because plans are able to challenge them through the plan preview periods. This "post-decision review and supervision" supposedly demonstrates that CMS retained "final authority." *Id.*

As an initial matter, that rejoinder is non-responsive to Maximus's adverse determination for Case 1-12757246876. Even if, contrary to fact, CMS used Maximus only "as an aid" in the administration of the Star Ratings system (*Alpine Securities*, 121 F.4th at 1325), that would not cure the underlying unconstitutionality of the appeal decision in that case. But beyond that, it is simply wrong to say that CMS uses Maximus only as an aid; on the contrary, the agency refers matters to Maximus and effectively gives it the final word.

The administrative record in this case lays the facts bare: When Alignment raised its concerns about Cases 1-13226962526 and 1-12757246876, CMS said it would "go[] back to Maximus" to resolve the issue. AR3. CMS did not then take Maximus's response as mere advice and conduct an independent review of Maximus's decision. Instead, CMS simply "shared" Alignment's concerns with Maximus and accepted Maximus's subsequent "confirm[ation]" that its initial decisions were correct. AR3; AR4.

Courts have recognized that, for an agency to retain ultimate authority over the input or advice of a private party, it must *actually* conduct a meaningful and independent review of its own. It is not enough for the agency, in theory, to have the authority to conduct independent review, even while it only "reflexively rubber stamp[s] [the] work product

prepared by" the private entity. *Consumers' Research v. FCC*, 109 F.4th 743, 770 (5th Cir. 2024) (cleaned up). Otherwise, the agency's final review would be a meaningless formality. Here, there is no evidence in the administrative record that CMS "independently perform[ed] its reviewing, analytical and judgmental functions" in reviewing Maximus's decisions. *Id.* Thus, even if the plan preview periods could in theory confer supervisory authority over Maximus's decisionmaking, CMS did not use any such authority here. It instead simply rubber-stamped Maximus's decisions. CMS thus unconstitutionally delegated governmental authority to Maximus.

F. Alignment's appeals measure scores were erroneously calculated

1. Under guidance issued by CMS, if a non-contracted provider requests reconsideration of a denied claim, it must complete a Waiver of Liability (WOL) statement. CMS, *Parts C & D Enrollee Grievances, Organization/Coverage Determinations, and Appeals Guidance* 60 (Nov. 18, 2024) (*Appeals Guidance*), https://perma.cc/5AHF-FKNF. If the provider does not provide a WOL, the plan must dismiss the request. *Id.* at 76; *see also* Maximus, *Medicare Health Plan Reconsideration Process Manual* 21 (Sept. 2024) (*Reconsideration Process Manual*), https://perma.cc/7346-PRMA.

In Case 1-13226962526, a non-contract provider sought reconsideration from Alignment of a claim denial. But it did not attach a WOL statement. AR19, 39. Alignment dismissed the request for reconsideration in accordance with guidance issued by CMS and Maximus. Alignment then forwarded the case to Maximus.

CMS now argues (at 41) that Alignment should not have sent the case to Maximus, since CMS guidance states that "[t]he rule requiring that a Part C case be automatically sent to the IRE if the plan upholds a denial on the merits of the request does not apply in the case of a dismissal of a [reconsideration request]." *Appeals Guidance* 77. A plan should

forward the case file to Maximus for a dismissal only if the enrollee requests IRE review of the plan's dismissal. CMS asserts (at 41-42) that, because Alignment forwarded the case file to Maximus, which is ordinarily the procedure when a plan affirms a claim denial on the merits, it was "reasonable" for Maximus to "treat the case as if Alignment had affirmed its initial decision" on the merits instead of dismissing it. Accordingly, CMS argues, the 30-day deadline set forth in 42 C.F.R. § 422.590(a)(2) applied: if an MAO makes a reconsidered determination that affirms the denial of an enrollee's claim, it must send the case file to the IRE no later than 30 days from the date it receives the request for reconsideration. Since Alignment forwarded the case file outside of this 30-day period, the appeal was untimely and negatively affected its Star Rating for the "Plan Makes Timely Decisions About Appeals" measure.

To state that argument is to refute it. If there was no obligation for Alignment to forward the case file to Maximus *at all*, there was necessarily also no obligation to forward the case file within 30 days. This was therefore not a case of an untimely appeal even on CMS's reasoning.

And in any event, Alignment's decision to forward an appeal it did not need to forward cannot excuse Maximus's mishandling of the appeal. CMS regulations are crystal clear that Maximus should not have issued a decision on the merits but rather should have dismissed the appeal. *See* 42 C.F.R. § 422.592(d)(2). Had Maximus followed the applicable regulation, the case would not have figured in the denominator of the "Plan Makes Timely Decisions About Appeals" measure. *2025 Technical Notes* 75 (AR1027) (defining the measure denominator to encompass "upheld, overturned, partially overturned appeals and appeals not evaluated by the IRE because plan agreed to cover"). Accordingly, Case 1-13226962526 should not have impacted Alignment's Star Ratings.

2. Case 1-12757246876 involved an appeal from Alignment's affirmance of a claim denial because the enrollee had active coverage with a primary insurance carrier at the time the claim was made. On review of Alignment's redetermination decision, Maximus reversed the denial of coverage, on the specific ground that Alignment had not coordinated coverage with the primary insurer. AR40. Maximus outlined three ways that plans can coordinate coverage with the primary insurer: (1) sending the bill to the primary insurer; (2) asking the provider to bill the primary insurer; (3) paying the provider and then collecting payment from the primary insurer. *Id.*

As we explained in the opening brief (at 44-45), Alignment had coordinated coverage under option (2), since in its remittance advice to the provider, it had included the code "UDMEOB—Missing Prior Insurance Carrier(s) EOB." AR2, 9, 18. That code was an instruction to the provider to bill the primary insurer. AR 18, 40. CMS makes much of the fact that Alignment has since "implemented changes" to "better clarify" that the stated code is a notification that the member "has other primary insurance" and that the provider should "submit [the] claim to the primary insurance first." AR18. But the fact that Alignment has taken steps to further *clarify* the meaning of the UDMEOB code does not show that the code had a different meaning. On the contrary, it reinforces that the code all along was an instruction to the provider to bill the primary insurer.

3. Finally, CMS (at 44) faults Alignment for not administratively challenging Maximus's appeal resolutions. To be sure, 42 C.F.R. § 422.616(a)-(b) provides that a decision made by the IRE "may be reopened and revised by" Maximus, "at the instigation of any party," including the MAO. And a request for reopening must be made within 180 calendar days. *Appeals Guidance* 106. Because Alignment did not seek reopening within that period, CMS contends (at 44) that Alignment's challenges are "untimely."

There is no authority requiring contracts to seek reopening before Maximus as a condition for challenging the inclusion of erroneous appeal resolutions in the Star Ratings calculations. Nor would any such requirement make sense. Alignment is not seeking to overturn Maximus's decision for purposes of recouping payments. Indeed, Alignment paid the claim at issue in Case 1-12757246876 and has made clear since raising the issue in the plan preview periods that it is not seeking return of the payment. AR2. Instead, it is requesting that contract H3443's 2025 *Star Ratings* be recalculated without penalty for Maximus's erroneous appeal decisions. That Alignment did not timely make use of the procedure for reopening appeals to dispute claim payment is irrelevant to that claim.

CONCLUSION

The Court should grant Alignment's motion for summary judgment and deny CMS's cross-motion for summary judgment, issuing the declarations supported by the arguments made in the opening brief and this reply and remanding to the agency for recalculation of the 2025 Star Ratings for the affected MA contracts.

Dated: May 5, 2025

Respectfully submitted,

/s/ *Michael B. Kimberly*

Michael B. Kimberly (991549)
Caleb H. Yong (1780922)
McDermott Will & Emery LLP
500 North Capitol Street N.W.
Washington, D.C. 20001
(202) 756-8901
mkimberly@mwe.com
cyong@mwe.com

Counsel for Plaintiff

CERTIFICATE OF SERVICE

Undersigned counsel certifies that a true and correct copy of this document was served via CM/ECF on all counsel of record pursuant to the Federal Rules of Civil Procedure on May 5, 2025.

/s/ Michael B. Kimberly

Michael B. Kimberly

IN THE 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.

Case No. 1:25-cv-00074-CRC

[PROPOSED] ORDER

Having considered Defendants' Cross-Motion for Summary Judgment and the memorandum filed in support thereof, the memorandum filed by plaintiff Alignment Healthcare Inc. in opposition thereto, and the reply filed by Defendants in support thereof, the Court hereby **ORDERS** that Defendants' Cross-Motion for Summary Judgment is **DENIED**.

SO ORDERED.

Dated: _____

_____

HON. CHRISTOPHER R. COOPER

United States District Judge

APPENDIX

The following are the names and addresses of the attorneys entitled to be notified of the entry of the proposed order:

Michael B. Kimberly
McDermott Will & Emery LLP
500 N. Capital Street, NW
Washington, DC 20001
Counsel for Plaintiff

M. Jared Littman
U.S. Attorney's Office
601 D Street NW
Washington, DC 20530
Counsel for Defendants