IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF COLUMBIA

ALIGNMENT HEALTHCARE INC.,

Plaintiffs,

v.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES, *et al.*,

Defendants.

Case No. 25-cv-0074-CRC

MOTION FOR SUMMARY JUDGMENT

Pursuant to Federal Rule of Civil Procedure 56, plaintiff Alignment Healthcare Inc. (Alignment) moves for an order granting summary judgment in its favor on Counts I through VI of the Complaint, Dkt. 1.

The Court should grant Alignment's motion for summary judgment for all of the reasons given in the memorandum of law filed herewith, and declare that (1) the Tukey Outlier Deletion Rule is arbitrary and capricious as applied to Alignment without appropriate denominator adjustments; (2) CMS's inclusion of "very low reliability" data in composite CAHPS survey measures is arbitrary and capricious and contrary to law; (3) CMS acted arbitrarily and contrary to law by including CAHPS survey data infected by a major administration error in Alignment's 2025 Star Ratings; (4) CMS's allowance for "oversampling" unlawfully and arbitrarily disadvantages smaller plans; (5) CMS has unconstitutionally delegated its regulatory authority to a private, for-profit entity in violation of the non-delegation doctrine; and (6) the independent review entity's final

decisions for Case Nos. 1-13226962526 and 1-12757246876 were arbitrary, capricious, and contrary to law. Consistent with those declarations, the Court should vacate Alignment's 2025 Star Ratings for the contracts affected by CMS's unlawful actions and policies, and remand with instructions for CMS to recalculate the Star Ratings for all of Alignment's impacted contracts.

Dated: April 7, 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 April 7, 2025.

/s/ 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. 25-cv-0074-CRC

MEMORANDUM OF POINTS AND AUTHORITIES 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

Cases—continued

Statutes and regulations

Statutes and regulations—continued

Statutes and regulations—continued

Other authorities

INTRODUCTION

This case concerns the Medicare Advantage and Part D Star Ratings system. It is the seventh challenge in less than a year to CMS's unlawful actions in administering the system. In the prior cases, courts across the country have ruled against CMS on varied grounds, holding that the agency violated its own regulations in determining the 2024 Star Ratings (*SCAN Health Plan v. HHS*, 2024 WL 2815789 (D.D.C. June 3, 2024); *Elevance Health, Inc. v. Becerra*, 736 F. Supp. 3d 1 (D.D.C. 2024)) and handled certain measure-level data for the 2025 Star Ratings contrary to its own guidelines and without adequately considering objections (*UnitedHealthcare Benefits of Texas v. CMS*, 2024 WL 4870771, at *3-*7 (E.D. Tex. Nov. 22, 2024)). In another case, CMS has confessed data errors. *See, e.g.*, Dkt. 25 at ¶ 3, *Centene Corp. v. Becerra*, No. 4:24-cv-01415 (E.D. Mo. Dec. 10, 2024). And the court in *UnitedHealthcare* held further that the agency abdicated its statutory duties by delegating them to a private entity. *See* 2024 WL 4870771, at *7-*9.

In this case, Alignment Healthcare challenges the case-specific application of a recent final rule—the Tukey Outlier Deletion Rule (*Contract Year 2021 Policy and Technical Changes to the Medicare Advantage Program*, 85 Fed. Reg. 33796, 33832-33836 (June 2, 2020))—which CMS began implementing in October 2023 for the 2024 Star Ratings. The Tukey Outlier Deletion Rule requires CMS to remove outliers from its raw data as part of the process of calculating the Star Ratings for a given year, ostensibly to "stabilize" the Star Ratings and "prevent large year-to-year fluctuations." *See id.* at 33833. But, as Alignment explained in a comment letter at the time, the Rule is based on objectively bad data science, and it produces arbitrary and capricious results. Although relatively simple adjustments would fix the Rule as applied to Alignment, CMS has refused.

Alignment further challenges CMS's handling of "very low reliability" data in composite measures underlying the 2025 Star Ratings. CMS regulations require the agency to discard measure-specific results when the data underlying the results fall below the agency's threshold for "very low reliability" data. When scores for a measure are based on survey participants' answers to a single survey question, for example, and the question yields "very low reliability" data for a given plan, CMS excludes the measure from the plan's Star Rating—because a score based primarily on random errors or statistical "noise" cannot be trusted to reflect true performance. But the way in which CMS assesses reliability for composite measures—those that take account of answers to multiple survey questions—is arbitrary and capricious. Instead of evaluating the reliability of the responses to each component question and discarding measure results when one or more component questions has "very low reliability," CMS evaluates reliability only at the measure level, even when there are different respondent subsets for each question. This methodology masks the presence of "very low reliability" components to a composite measure, allowing CMS to include measures that, by regulation, should be excluded.

Beyond application of the Tukey Outlier Deletion Rule and CMS's failure to suppress composite measures with very low reliability, Alignment alleges that the data underlying the company's 2025 Star Ratings is fundamentally flawed.

First, the rate of Spanish-language responses to the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey showed a marked and unexplained decline for two Alignment contracts. Data produced in this lawsuit demonstrates why: Large portions of Alignment's Spanish-speaking enrollees were erroneously sent survey questionnaires in English rather than Spanish. The data show that Alignment properly designated the enrollees as Spanish speakers who should have received Spanish-language surveys. And

CMS's data confirms that these enrollees are Spanish speakers. Yet the CMS-approved survey contractor who administered the surveys for Alignment sent English-language surveys to many of these enrollees. Alignment suspected as much and raised the issue with CMS before the 2025 Star Ratings were finalized. But CMS brushed off the concern, without meaningfully investigating whether Alignment's suspicions—now borne out in the data—were true. The scores for Alignment's CAHPS measures were adversely impacted, reducing its 2025 Star Ratings.

That is not all. CMS also has wholesale outsourced its administrative claim appeal process to a third party known as Maximus. With respect to two appeals involving Alignment, Maximus failed to follow its own policy manual and CMS guidance, again adversely impacting Alignment's 2025 Star Ratings. First, rather than dismissing a particular request for reconsideration (as required under the circumstances), Maximus instead reached the merits of the request and ruled for Alignment. While that may seem a favorable outcome for Alignment, it was not. If the contractor had dismissed the reconsideration request as required, Alignment would have had 100% performance on the "Plan Makes Timely Decisions about Appeals" measure for contract H3443, receiving 5.0 Stars on that highly-weighted measure. But because Maximus instead wrongly resolved the request on its merits, Alignment's measure-level raw score fell to 94.4% (it had 17 of 18 appeals timely resolved), and its measure-level Star Rating thus fell to 3.0 Stars.

Second, Maximus wrongly overturned a claim denial involving an enrollee that had other primary health insurance. Maximus did not disagree that the claim should have been denied as a substantive matter but faulted Alignment for not "coordinating coverage" with the other insurer. Yet Alignment had done exactly that, and CMS's contractor was simply wrong to find otherwise.

The outcome in each of these two cases is arbitrary in its own right. But in addition, they show why application of the Tukey Outlier Deletion Rule here is arbitrary and capricious. Because of the relatively low number of appeals under contract H3443 (a good thing, one would think), Alignment had no opportunity to earn 4.0 Stars; rather, a single incident pushed Alignment from 5.0 Stars straight to 3.0 Stars on two different measure scores. They also show why CMS's delegation of regulatory authority to a contractor is so problematic—Alignment had no meaningful opportunity to seek CMS's independent review of or judgment upon the private entity's incorrect decisions.

Star Ratings are tremendously important to the Medicare Advantage and Part D programs. They provide the public with critical information about a plan's quality, enabling them to compare plans when shopping during the annual enrollment period. They also enable plans to offer richer benefits packages, while keeping costs low for beneficiaries. "CMS is also obligated by statute to offer additional funding to plans with better Star Ratings," and the funds must be used to "lower costs for [Medicare] beneficiaries or to provide them with additional benefits." *SCAN*, 2024 WL 2815789, at *1. This additional funding amounts to tens or hundreds of millions of dollars per contract. *Id.*

CMS's careless and error-prone approach to the Star Ratings does not appropriately reflect their importance to the MA program. Like all stakeholders connected to the program, Alignment counts on CMS to administer Star Ratings in a consistent, transparent, and rational manner. Alignment has been recognized consistently for its high-quality MA plans—a fact well reflected in the company's Star Ratings when the Star Ratings system is working as it should. Even now, 98% of Alignment's members are enrolled in plans rated 4.0 Stars or higher, and the company is one of the very few Medicare Advantage organizations with a 5.0 Star plan.

Notwithstanding its industry-leading performance, Alignment brings this suit because the Star Ratings system no longer can be counted on to produce rational, predictable results. The Star Ratings system cannot work if CMS is allowed to promulgate rules based on bad data science, outsource the bulk of its work to unaccountable private entities, and turn a blind eye to data errors that arbitrarily disadvantage smaller Medicare Advantage organizations like Alignment. Vacatur of Alignment's 2025 Star Ratings, and a remand for recalculation of the Star Ratings consistent with law, is thus in order.

BACKGROUND

I. STATUTORY AND REGULATORY BACKGROUND

A. The Medicare Advantage and Medicare Part D programs

The Medicare program provides federally funded health insurance to Americans who are 65 and older, have received federal disability benefits for at least 24 months, or have end-stage renal disease. *See Becerra v. Empire Health Foundation*, 597 U.S. 424, 428-429 (2022). The Medicare program is administered by CMS on behalf of the Secretary of HHS. 42 U.S.C. § 1395kk(a); *Health Care Financing Administration Reorganization Order*, 42 Fed. Reg. 13262 (Mar. 9, 1977).

Medicare comprises Parts A, B, C, and D. *See Medicare Program; Establishment of the Medicare Advantage Program*, 70 Fed. Reg. 4588, 4589 (Jan. 28, 2005). Part A covers inpatient hospital treatment, and Part B covers outpatient services. Together, Parts A and B are known as "traditional" or "original" Medicare. Traditional Medicare use a fee-for-service payment model. *See* 42 U.S.C. § 1395w-22(a)(1). Under this model, CMS reimburses providers directly for the services they provide to Medicare beneficiaries. *United-Healthcare Insurance v. Becerra*, 16 F.4th 867, 872 (D.C. Cir. 2021).

Medicare Part C, also known as Medicare Advantage or MA, uses a different model. *See generally* Medicare Prescription Drug, Improvement, and Modernization Act, Pub. L. No. 108-173, 117 Stat. 2066 (2003) (codified at 42 U.S.C. §§ 1395w-21 to 1395w-28); 70 Fed. Reg. 4588. The program avoids the pitfalls of traditional Medicare and its single-payer, one-size-fits-all approach by offering plans sponsored by private companies called Medicare Advantage organizations, or MAOs. These companies must cover at least the same services as traditional Medicare. 42 U.S.C. § 1395w-22(a). But, to attract enrollees, MA plans typically include additional benefits not covered by traditional Medicare, such as dental and vision benefits. *UnitedHealthcare*, 16 F.4th at 872.

Under this public-private partnership model, MAOs do not receive fee-for-service reimbursements from CMS for the healthcare services their enrollees receive. *See generally* 42 U.S.C. § 1395w-23(a). Instead, they receive a per-enrollee monthly allowance, which they use to provide coverage to the beneficiaries enrolled in the plan. In turn, MAOs pay healthcare providers for the services they provide to MA enrollees. *Id.* § 1395w-23(a)(1); *see UnitedHealthcare*, 16 F.4th at 874.

CMS determines a plan's monthly payment by comparing the plan's "bid" (its risk-adjusted estimated cost of providing Medicare-covered services to a particular patient population) (42 U.S.C. § 1395w-24(a)), to a "benchmark" (the maximum amount the federal government will pay to provide coverage in the plan's service area) (*id.* §§ 1395w-23(b)(1)(B), (n)). If the MAO's bid is below the benchmark, CMS pays the MAO its bid rate, while also returning a specified percentage of the difference between the benchmark and the bid as a "rebate," which the MAO must return to plan participants through additional benefits, lower premiums, or cost sharing. 42 U.S.C. §§ 1395w-23(a)(1)(B)(i), (E); 1395w-24(b)(1)(C). If, in contrast, an MAO's plan bid is at or above the benchmark, the MAO

receives monthly payments at the benchmark rate and must charge enrollees an additional premium to cover the shortfall. *Id.* §§ 1395w-23(a)(1)(B)(ii), 1395w-24(b)(2)(A). *See also MMAPA v. Emanuelli Hernández*, 58 F.4th 5, 8 n.1 (1st Cir. 2023); *Elevance Health, Inc. v. Becerra*, 736 F. Supp. 3d 1, 5-6 (D.D.C. 2024).

In addition to inpatient treatment and outpatient services through traditional Medicare or MA, Medicare beneficiaries may also obtain prescription drug coverage through Medicare Part D. Like the MA program, the Part D prescription drug benefit provides coverage through a public-private partnership with plan sponsors. These plan sponsors may offer Part D coverage through standalone prescription drug plans (PDPs) for individuals enrolled in traditional Medicare, or through drug coverage bundled with an MA plan, known as an MA-PD plan. 42 U.S.C. §§ 1395w-101(a)(1), (3)(C).

B. The Star Ratings system generally

1. Each MAO is required by statute to establish "ongoing quality improvement program[s]." 42 U.S.C. §§ 1395w-22(e)(1), (3). The MA statute prescribes a system of "quality rating[s]" for plans using a "5-star rating system" that is based on the data collected under MAOs' quality improvement programs. *Id.* § 1395w-23(o)(4)(A).

To implement this statutory mandate, and to assist would-be enrollees in choosing the plans best suited to their needs, CMS established the Star Ratings system early in the MA program's existence. Star Ratings purport to measure the quality of health and drug plans and provide comparative information to beneficiaries. *See* 42 C.F.R. §§ 422.160-(b)(1), 423.180(b)(1); *see also Contract Year 2019 Policy & Technical Changes to the Medicare Advantage Program*, 83 Fed. Reg. 16440, 16520 (Apr. 16, 2018).

CMS evaluates plans on a range of quality, compliance, and other "measures," and develops ratings on a five-star scale based on the scores that plans earn on these measures.

See 42 C.F.R. §§ 422.166(a)(4), 423.186(a)(4). A 1.0 Star Rating is the worst rating, and 5.0 Star Rating is the best. *Id.* §§ 422.166(a)(4), (c)(3), (d)(2)(iv), 423.186(a)(4), (c)(3), (d)(2)(iv). The system is intended to reflect the quality and performance of each plan. 42 C.F.R. §§ 422.162(b)(1), 423.182(b)(1); *see also Elevance*, 736 F. Supp. 3d at 5.

To develop the Star Ratings, CMS first determines each plan's numerical score for each of more than three-dozen measures. It then converts these raw scores into "measure-level" Star Ratings. Measure-level ratings use four thresholds, or "cut points," to divide the measure scores into five "whole star increments." 42 C.F.R. §§ 422.166(a)(4), 423.186(a)(4). From the measure-level Star Ratings, CMS calculates Part C and Part D "summary" ratings, which reflect the weighted average of a contract's measure-level Star Ratings (meaning that some measures are given more weight than others). *Id.* §§ 422.166, 423.186. CMS further calculates an overall rating for each MA-PD contract, which reflects the weighted mean of that contract's Part C and Part D measure-level Star Ratings.

2. The Star Ratings system is integral to the MA and Part D programs. It serves three primary functions, each of which requires the ratings to "accurately . . . reflect true performance." 83 Fed. Reg. at 16519.

First, the system is designed to provide Medicare beneficiaries with "comparative information on plan quality and performance," allowing them to make "knowledgeable enrollment and coverage decisions in the Medicare program." 42 C.F.R. §§ 422.160(b)(1), 423.180(b)(1). "The MA and Part D Star Ratings system is designed to provide information to the beneficiary that is a true reflection of the plan's quality and encompasses multiple dimensions of high quality care," with the goal of "inform[ing] plan choice" by beneficiaries. 83 Fed. Reg. at 16520. To this end, CMS maintains the Medicare Plan Finder website, which displays information about available plans, including each plan's Star Rating.

Elevance, 736 F. Supp. 3d at 5. CMS displays Star Ratings in its online and print resources on available MA plans. *See* 42 C.F.R. §§ 422.166(h); 423.186(h). Through the online Medicare Plan Finder tool, CMS displays MA plans to prospective enrollees in order of highest to lowest Star Ratings to guide beneficiaries to higher-rated plans first.

Star Ratings thus influence each plan's position in the marketplace by affecting how prospective enrollees, and the agents and brokers who advise them, perceive the comparative quality of various plans. For instance, MA-only plans with a 5.0 Star Part C summary rating, Part D plans with a 5.0 Star summary rating, and MA-PD contracts with a 5.0 Star overall rating are displayed with a high-performing icon, while a plan that had any combination of Part C or Part D summary ratings of 2.5 Stars or lower in the most recent three consecutive years is marked with a "low performance" icon. *See* 42 C.F.R. §§ 422.166-(h)(1)(ii), 423.186(h)(1)(ii).

Second, the system is designed to help CMS perform "oversight, evaluation, and monitoring of MA and Part D plans" and compliance with regulatory and contract requirements. 83 Fed. Reg. at 16520-16521; *see also* 42 C.F.R. §§ 422.160(b)(3), 423.180(b)(3). For that reason, CMS conditions certain aspects of a plan's status within the MA program on its Star Rating. For example, only plans with a Star Rating of 5.0 Stars may market to and enroll existing beneficiaries outside of open enrollment. 42 C.F.R. § 422.62(b)(15). And CMS may treat low-performing MA plans as "hav[ing] failed to comply with a contract," and thus deny a Medicare Part C or D application, if it "[r]eceived any combination of Part C or D summary ratings of 2.5 or less in both of the two most recent Star Rating periods." *Id.* §§ 422.502(b)(1)(*i*)(D), 423.503(b)(1)(*i*)(D).

The Star Ratings program's *third*, more recent, purpose is to provide "quality ratings on a 5-star rating system" to be used in administering the scheme of additional

payments for high quality MA plans, known as quality bonus payments (QBPs). *Id.* § 422.160(b)(2). These payments were established in 2010 by the Patient Protection and Affordable Care Act (ACA). *See Medicare Program; Proposed Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs for Contract Year 2012 and Other Proposed Changes*, 75 Fed. Reg. 71190, 71218 (Nov. 22, 2010). The ACA provides that an MA plan is entitled to QBPs from CMS depending on the "quality rating" of the plan, which "shall be determined according to a 5-star rating system." 42 U.S.C. § 1395w-23(o)(4)(A). In particular, if an MA plan receives a Star Rating of 4.0 Stars or higher, its benchmark amount is increased, in turn increasing the rebates that CMS will pay by increasing the difference between the plan sponsor's benchmark and its bid. *Id.* § 1395w-23(o)(1), (3)(A). Star Ratings also determine the percentage of the difference that is returned as a rebate. Plans with a 4.5 Star Rating or higher receive 70% of the difference between the benchmark and the bid; plans with a Rating between 3.5 and 4.5 Stars receive a 65% rebate, and plans with a rating under 3.5 Stars receive a 50% rebate. *Id.* § 1395w-24(b)(1)(C)(v); 42 C.F.R. § 422.266(a)(2)(ii).

C. Calculating the Star Ratings measures

Broadly speaking, the measures that underlie plans' Star Ratings fall into two categories. The first category of measures is based on data drawn from a family of patient-experience surveys, known as the Consumer Assessment of Healthcare Providers & Systems (CAHPS) surveys. The CAHPS surveys—which are written by CMS and administered by CMS-approved contractors—ask enrollees (and in some cases their families) about their experiences with, and assessments of, their health care providers and plans. Plans' scores on the CAHPS measures reflect enrollees' responses to the survey questions. The second category—non-CAHPS measures—includes data concerning plan performance reported by

the MAO or generated by CMS or its subcontractors. These data reflect such issues as the plan's success in delivering preventive care like cancer screenings, the health outcomes of plan participants with respect to certain conditions, and the quality of plan administration measured by such factors as the speed with which it pays claims and resolves appeals.

1. Non-CAHPS measures and the Turkey Outlier Rule. For non-CAHPS measures, the agency converts raw measure scores into measure-level Star Ratings using a clustering methodology. 42 C.F.R. §§ 422.166(a)(2); 423.186(a)(2). The clustering methodology "identifies the natural gaps that exist within the distribution of the scores" and creates five non-overlapping "groups (clusters)" among the scores that correspond to the five whole-star increments in the measure-level Star Rating. *See* CMS, *Medicare 2025 Part C & D Star Ratings Technical Notes* 143 (Oct. 3, 2024) (*2025 Technical Notes*) (AR1095).

Clustering is a statistical technique used to partition a dataset into distinct groups, such that the observations within a group are as similar as possible to each other, and as dissimilar as possible to observations in any other group. *See id.* at 9 (AR961); 42 C.F.R. § 422.162(a); 83 Fed. Reg. at 16525.

Beginning with the 2024 Star Ratings, CMS has implemented the Tukey Outlier Rule. Accordingly, the first step of the clustering methodology to convert non-CAHPS measure scores into measure-level Star Ratings is to remove "Tukey outer-fence outliers" from the industry-wide raw scores for a given measure. 42 C.F.R. §§ 422.166(a)(2)(i), 423.186(a)(2)(i). This is a statistical technique used to "identify and to remove extreme outliers in a dataset." *Elevance*, 736 F. Supp. 3d at 8.

Once outlier deletion is complete, CMS sorts the measure scores into groups industry wide. 42 C.F.R. §§ 422.166(a)(2)(i), 423.186(a)(2)(i). The "dividing lines" between the groups establish the "cut points" for that measure—that is, the numerical values at

which "a score results in a higher or lower star rating." *See SCAN Health Plan v. HHS*, 2024 WL 2815789, at *2 (D.D.C. June 3, 2024); *Elevance*, 736 F. Supp. 3d at 8. A plan's measure-level Star Rating for a given measure depends on where the plan's numerical measure score falls relative to the cut points. *See SCAN*, 2024 WL 2815789, at *2. Because the measure-level Star Ratings are based on whole star increments, minor changes to the cut points can lead to a whole-star drop in a plan's measure-level Star Rating.

2. CAHPS measures and reliability. CMS applies a different methodology for deriving Star Ratings for measures that depend on enrollee responses to the CAHPS surveys. On this front, CMS uses a method that evaluates "the relative percentile distribution" of a plan's score, together with the "significance" and "reliability" of the underlying survey data. It then converts the raw numerical scores into measure-level Star Ratings. 42 C.F.R. §§ 422.166(a)(3); 423.186(a)(3). A measure-level CAHPS Star Rating is thus a function of a plan's raw performance relative to other plans, adjusted for the "significance" and "reliability" of the plan's survey data. *Id.* Because numerical measure scores are converted to measure-level Star Ratings with whole-star increments, as with non-CAHPS measures, even minute fluctuations in CAHPS data can change a contract's base group assignment and lead to a whole-star change in the contract's Star Rating on that measure.

a. Reliability. Reliability measures the consistency of survey scores across different samples. In essence, it asks whether a given CAHPS score reflects true patient experience or instead random fluctuations or "noise" in the data. 42 C.F.R. § 422.162(a). A measure with low reliability does not reflect actual plan performance; differences across plans with low reliability scores are as likely to reflect random data variation as meaningful distinctions between the plans. *Id.*

Reliability values range from 0.0 to 1.0. *Id.* A reliability score of 0.70—the generally accepted cutoff for reliability (*see* Agency for Healthcare Research & Quality, *Fielding the CAHPS Clinician and Group Survey* 6, Doc. No. 2033 (June 12, 2017))—means the result is 30% "noise." If the reliability of a plan's score for a given CAHPS measure is below 0.60, CMS designates the score "very low reliability" and discards the measure-level Star Rating for the plan. *See* 42 C.F.R. § 422.166(a)(3); *2025 Technical Notes* 189 (AR1141).

Six CAHPS measures are "composite" measures that combine multiple question-level scores: *See* AR129, 194, 399-401, 1059-1061, 1063-1065. CMS has explained that, for these composite measures, "answers to questions about the same topic are combined to form composites" for "scoring . . . purposes." *Protocols & Specifications* 7, 73 (AR129, 195). For example, for the "Getting Appointments and Care Quickly" measure, the measure score depends on enrollee responses to two survey questions. *2025 Technical Notes* 63 (AR1015). The composite measure-level score is thus the average of the question-level scores. *Protocols & Specifications* 73-74 (AR195-196).

Although CMS calculates reliability on a question-by-question basis, it evaluates the overall reliability of a measure at the measure level. For a non-composite measure (a measure comprising a single survey question), the two are one-and-the-same. But for composite measures, CMS must determine the *overall* measure reliability based on the reliability of the constituent questions. Using a "weakest link" theory of reliability, that should mean that a composite measure is discarded as "very low reliability" if any one of the questions comprising the measure has "very low reliability." *See Oceana, Inc. v. Raimondo*, 530 F. Supp. 3d 16, 47 (D.D.C. 2021) (Cooper, J.) (use of "underlying data [that] is unreliable necessarily means that [calculations] based on that data would be equally if not more unreliable—garbage in, garbage out"), *aff'd,* 35 F.4th 904 (D.C. Cir. 2022). But that is not

what CMS does. Rather than evaluating whether each component question meets the 0.60 reliability threshold, CMS instead calculates a measure-level reliability score using a statistical model that pools data across the constituent questions, even when those questions are answered by different subsets of beneficiaries and bear different question-level reliability scores. This can produce a composite reliability score above the 0.60 threshold, even when some (or many) of the underlying questions fall below it.

b. Significance. Measure-level CAHPS scores are also adjusted for statistical significance. For MA contracts with greater than 800 CAHPS-eligible enrollees, CMS samples 800 eligible enrollees. For MA contracts with between 600 and 799 eligible enrollees, CMS samples all eligible enrollees. If a contract has fewer than 600 eligible enrollees, it is not required to field the CAHPS surveys. If a contract does not field the surveys, the CAHPS measures do not factor into the contract's Star Ratings. *Protocols & Specifications* 17 (AR139). CMS allows contracts with greater than 800 CAHPS-eligible enrollees to "oversample" at their own expense, provided "there is sufficient eligible enrollee volume to support additional sampl[ing]." *Id.* at 18 (AR140). Some larger contracts—many with tens or hundreds of thousands of enrollees—will oversample by an order of magnitude.

The larger the number of respondents a survey receives, the greater the "significance" that CMS will assign to the contract's measure-level mean score. CMS's allowance of oversampling thus confers a substantial advantage on larger contracts. A large contract may earn a higher measure-level Star Rating than a smaller contract with that same mean score and reliability rank if it oversamples and obtains a greater significance score.

c. Survey languages. CMS writes the CAHPS surveys but does not administer them. It instead directs MAOs to work with any one of a number of approved survey vendors to administer the surveys. These survey vendors transmit the survey responses to CMS and

must comply with CMS's protocols. *Protocols & Specifications* 8 (AR130). As relevant here, CMS has recognized that the language in which the CAHPS surveys are administered has important implications for response rates and data quality. In an October 26, 2023, memorandum that furnished "information on administering the survey in other languages," the agency expressly acknowledged that "[e]nsuring that . . . enrollees have the opportunity to complete the survey in the language with which they are most comfortable provides the most accurate picture of patient experience." AR77. Moreover, "[m]aximizing response rates" ensures "more robust information about patient experience." AR113.

To that end, CMS provides its survey vendors with translations of the surveys and supporting materials, so the surveys can be administered in languages other than English, including Spanish. *See Protocols & Specifications* 32-33, 50-51 (AR154-155, 172-173); *accord* CMS, *MA & PAP CAHPS Survey Vendor Training* 100 (Nov. 2024) (*Training*), perma.cc/Y695-7R2X. Indeed, CMS guidance states expressly that approved contract vendors "must" make "Spanish-language questionnaires available to all Spanish-speaking enrollees." *Protocols & Specifications* 50-51 (AR172-173); *Training* 100. If an MA plan requests it, the vendor must send Spanish-language questionnaires to enrollees identified using language preference data provided by the plan. *Protocols & Specifications* 50-51 (AR172-173). The decision whether to send Spanish-language questionnaires to particular enrollees is the plan's decision, not the vendor's. *Id.* at 50 (AR172).

D. Reconsideration of coverage determinations and IRE review

One of the more important Star Ratings measures concerns the plan's accurate and timely handling of enrollee claim appeals.

1. Each MAO must maintain a "procedure for making [coverage] determinations" (42 U.S.C. § 1395w-22(g)(1)(A)) and "provide for reconsideration" of any adverse

determinations "upon request by the enrollee involved." *Id.* § 1395w-22(g)(2)(A). CMS refers to reconsideration by MAOs of an initial determination to deny coverage as a "level 1 appeal." CMS, *Parts C & D Enrollee Grievances, Organization/Coverage Determinations, and Appeals Guidance* 6 (Nov. 18, 2024) (*Appeals Guidance*). https://perma.cc/5AHF-FKNF. Not only may an enrollee seek reconsideration, but so also may the "physician who is treating" the enrollee and acting on the enrollee's behalf. 42 C.F.R. §§ 422.578, 422.582(a).

The statute provides for a "level 2 appeal" to an independent review entity (IRE). Congress has directed CMS to "contract with an independent, outside entity 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)), which CMS has done by regulation (42 C.F.R. § 422.592(a)). CMS guidance instructs that if an MA plan makes an untimely decision on a level 1 appeal, or if it affirms an adverse coverage determination, the plan must forward the appeal to the IRE for review for a "level 2 appeal." *See Appeals Guidance* 79, 81-83.

CMS contracts with Maximus Federal Services (Maximus), a subsidiary of Maximus, Inc., to serve as the IRE authorized under 42 C.F.R. § 422.590(i) to review MAOs' adverse resolutions of reconsideration requests. *See* CMS, *Reconsideration by Part C Independent Review Entity (IRE)*, (Jan. 7, 2025), https://bit.ly/3ZQRlE9. Maximus is a private, for-profit corporation. Since 1997, its parent has been a publicly traded corporation.

When Maximus rules against a provider or beneficiary, the provider or beneficiary has a right to administrative review before an administrative law judge (ALJ) pursuant to 42 U.S.C. § 405(b). But when Maximus rules against an MAO, the plan has no right to administrative review, and the decision is final. Once Maximus has reviewed the MAO's decision regarding a reconsideration request, it makes a "reconsidered determination." 42

C.F.R. § 422.594(a). This "reconsidered determination is final and binding on all parties, unless a party *other than the MA organization* files a request for [an ALJ] hearing" or it is reopened and revised by Maximus. *See id.* § 422.616 (emphasis added).

If an MAO "dismisses a reconsideration request" because it determines that the request is procedurally defective (*id.* § 422.582(f)), the enrollee or the treating physician "has the right to request review of the dismissal by the independent entity" (*id.* § 422.590(i); *see also id.* § 422.592(a)). Upon review, Maximus may dismiss the request for reconsideration, and such dismissal "is binding and not subject to further review" unless a party files a request for a hearing before an ALJ, or Maximus later vacates its own dismissal. *Id.* § 422.592(g). By regulation, however, an MAO again has no right to request an ALJ hearing. *Id.* § 422.600. If Maximus "determines that the [MAO's] dismissal was in error," it must "vacate[] the dismissal and remand[] the case to the plan for reconsideration." *Id.* § 422.592(i). Maximus's decision regarding the MAO's dismissal determination "is binding and not subject to further review." *Id.*

2. Two Star Ratings measures—"Plan Makes Timely Decisions about Appeals" and "Reviewing Appeals Decisions"—track plan performance in relation to level 1 and level 2 appeals. CMS guidance explains that the "Plan Makes Timely Decisions about Appeals" measure "shows how fast a plan sends information for an independent review." *2025 Technical Notes* 75 (AR1027). However, "[d]ismissed appeals . . . are excluded from this measure." *Id.* The "Reviewing Appeals Decisions" measure "shows how often an independent reviewer found the health plan's decision to deny coverage to be reasonable." *Id.* at 76 (AR1028). The metric tracked by this measure is the "[p]ercent of [level 1] appeals where a plan's decision was 'upheld' by the Independent Review Entity (IRE) . . . out of all the plan's appeals . . . that the IRE reviewed." *Id.* at 77 (AR1029).

In a memo CMS issued to MAOs concerning the availability of underlying measure data that plans could use to validate their various measure-level Star Ratings, the agency directs plans to the Maximus website (www.medicareappeal.com) to obtain data for the appeals measures, "Plan Makes Timely Decisions about Appeals" and "Reviewing Appeals Decisions," and more broadly for "[i]nformation regarding the Part C reconsideration process." CMS, *Information to Review Data Used for Medicare Part C and D Star Ratings and Display Measures* 3 (Apr. 11, 2024) (*Information to Review Data*) (AR12). In case MAOs notice discrepancies or have other concerns about the data, the memo instructs them to "bring these issues to the attention of Maximus." *Id.*

E. Plan sponsor participation in Star Ratings (plan preview periods)

Given the importance of Star Ratings to the MA program, and the sensitivity of the system to erroneous or unreliable data, CMS's regulations establish an administrative process through which MAOs and other plan sponsors can review and comment on, and challenge the adequacy of, the agency's preliminary calculations. The regulations call this administrative process "plan preview" periods. "CMS will have plan preview periods before each Star Ratings release during which MA organizations can preview their Star Ratings data in HPMS prior to display on the Medicare Plan Finder." 42 C.F.R. § 422.166(h)(2); *see also id.* § 423.186(h)(2). The plan preview process is the only administrative process available to a plan permitting it to comment on and participate in the Star Ratings process before Star Ratings are finalized and published by CMS.

CMS holds two preview periods. During the first plan preview, CMS "expect[s]" MAOs to "closely review the methodology and their posted numeric data for each measure." 83 Fed. Reg. at 16588. During the second plan preview, CMS will post to its website "any revisions made as a result of the first plan preview," as well as the "preliminary Star

Ratings for each measure, domain, summary score, and overall score." *Id.* CMS again "expect[s]" plan sponsors to "closely review the methodology and their posted data for each measure, as well as their preliminary Star Rating assignments." *Id.*

A core purpose of the plan preview process is data validation. The two plan previews allow "sponsors to review and raise any questions about their own plan's data prior to the public release of data for all plans," so that if there are any errors, "necessary corrections" can be made prior to the Star Ratings being announced to the public. *Id.*

The plan preview process is rooted in CMS's view that, for Star Ratings to be a "true reflection of the quality, performance and experience of the beneficiaries enrolled in MA and Part D contracts," the data and analysis underlying measure scores must be "complete, accurate, and unbiased." *Id.* at 16567. "Data validation is a shared responsibility among CMS, CMS data providers, contractors, and Part C and D sponsors." *Id.* at 16562.

Reflecting the importance of data integrity and accuracy, CMS imposes harsh penalties on MAOs for submitting inaccurate data. For example, CMS will "reduce a contract's measure rating when CMS determines that" the data reported to it "are inaccurate, incomplete, or biased," which can result from "mishandling" and "inappropriate processing" of data. 42 C.F.R. §§ 422.164(g)(1), 423.184(g)(1). For measures based on data that an MAO must submit to CMS, the rating will be reduced to 1 Star when a contract "was not compliant with CMS data validation standards." *Id.* §§ 422.164(g)(1)(ii), 423.184(g)(1)(ii).

II. FACTUAL BACKGROUND

A. Unexplained drops in Spanish-language survey response rates

1. The data that CMS posted during the plan preview periods for the 2025 Star Ratings showed a sharp and unexpected decline in the Spanish-language response rates to the CAHPS survey for two Alignment contracts, H3815 and H3443.

According to CMS's own analysis, the percentage of completed Spanish-language surveys for H3815 dropped from 31% of overall responses in 2023 to 20% in 2024. For the H3443 contract, it fell from 24% of total responses in 2023 to 19% in 2024. AR62, 66. These marked declines in Spanish-language responses occurred against the backdrop of modest *increases* for both contracts in the percentage of CAHPS survey-eligible enrollees who should have received Spanish-language surveys. AR64, 72.

The declines were anomalous and account for the two contracts' otherwise unexplained drops in performance. Notably, Alignment's CAHPS scores for contracts H4961 and H9686—which are highly similar to contracts H3815 and H3443 in terms of provider network structure, member experience, and member demographics—both showed significant *improvements* in overall performance from 2023 to 2024. AR8, 27-28.

For two reasons, the lower response rates among Alignment's Spanish-speaking enrollees materially impacted its 2025 Star Ratings. First, past CAHPS surveys and Alignment's internal 2025 surveys consistently show that its Spanish-speaking members report approximately 10% higher satisfaction compared with English-speaking members. AR72. And contracts H3815 and H3443, which have the highest concentration of Spanish-speaking members among all of Alignment's contracts, were therefore disproportionately affected by the underrepresentation of this population in the responses to the CAHPS survey. AR72. Second, the reliability scores for the "Getting Appointments and Care Quickly" and "Care Coordination" measures for H3815 and H3443 were both within a hair of 0.60. As to each, a single additional response containing an inapt answer could have caused the disqualification of these measures. AR48, 51-53. At the same time, several other CAHPS measures were within reach of the next Star Rating threshold, meaning even marginal changes in measure-level scoring would likely have resulted in higher Star Ratings.

2. Alignment called CMS's attention to the unexplained declines in Spanish-speaking enrollee responses (AR59, 64-65, 73), but CMS declined to conduct a meaningful review of the issue or to adjust the CAHPS measure scores for H3815 and H3443.

CMS dismissively insisted that it is "typical" and "common" in CAHPS and other patient surveys for the proportion of Spanish speakers among respondents to be lower than the proportion among enrollees sampled "due to a lower response rate." AR61. The agency also implicitly faulted Alignment for the drop in Spanish-language responses, asserting (incorrectly) that Alignment had not furnished the company's own Spanish-speaker data to its contracted survey vendor. AR68. At points during the plan preview period, CMS even disclaimed responsibility for implementing language preferences in the administration of the CAHPS surveys. Insisting that it "does not get involved in how survey vendors implement language preference data," CMS declared that "[w]hether or not Spanish-speaking members received surveys in English despite their preference is outside of CMS control" and thus not a factor it has any obligation to address. AR58.

Alignment explained that it *did* share its foreign-language designations with the CMS-approved survey contractor, "providing consistent member preference data year after year, particularly for our Spanish-speaking membership." AR54. The agency then changed tack, implying that Alignment had in fact shared its Spanish-speaker data with the survey vendor by indicating that the vendor had "attested" that it had "used the language preference data shared by the plan." AR53. But CMS declared that there was "no support for the contention that survey sampling or survey administration . . . adversely affected scores in any way," relying only on two assertions.

First, CMS asserted that the survey sample was random and specifically that it was representative with respect to predicted Spanish preference. *Id.*

Second, CMS maintained that the statistical relationship between the percentage of responses that were in Spanish and the percentage of Spanish speakers within the sample, as well as the relationship between the percentage of Spanish-language responses and the percentage of Spanish speakers among CAHPS-eligible enrollees, was "as expected and consistent with the correct administration of the survey." *Id.* CMS provided no evidentiary support for these assertions, simply reporting that they had been "verified." *Id.*

3. CMS's position during the plan preview period is not consistent with the data that the agency has produced in this litigation. That data show that a material number of enrollees designated by Alignment as Spanish speakers nonetheless received surveys in English, contrary to Alignment's language preference designation. Approximately 15% of enrollees designated by Alignment as Spanish speakers—29 out of 200—received surveys in English for contract H3443. For contract H3815, the number is approximately 10%, or 20 out of 201. These data show that CMS did not, in fact, verify that the survey contractor correctly distributed Spanish-language surveys to Alignment's expressly designated Spanish-speaking enrollees.[1]

[1] The administrative record produced by the agency includes CAHPS survey data that has been "deidentified," meaning that any information that can be used to discern the identities of survey respondents has been removed. The agency also has produced the original, non-deidentified data (subject to the Court's protective order), but without conceding that the data is a part of the administrative record before the Court. *See* Dkt. 9-1 ¶ 3. That is an odd position to take. The original, non-deidentified data was plainly before the agency during its decisionmaking process. And it has produced it in the litigation subject to the terms of a protective order. The data is thus properly considered a part of the administrative record. "After all, the administrative record properly consists of the materials before the agency and no more *nor less*." *Oceana, Inc. v. Ross*, 290 F. Supp. 3d 73, 78 (D.D.C. 2018) (Cooper, J.). "If a plaintiff can show that a piece of evidence was before the agency at the time the decision was made," it follows that "that evidence is part of the administrative record" even if not formally designated as such. *Id.* In any event, the deidentified data circumstantially confirms the same approximate results as the non-deidentified data.

Moreover, the data show that enrollees who were (1) identified by Alignment as having a preference to receive the survey in Spanish but (2) nonetheless received the survey in English were by far the least likely to complete the survey. In fact, the response rate was less than one half of the response rate of Spanish speakers who received Spanish-language surveys, and less than one third the response rate of English speakers who received English-language surveys. Alignment's internal quality-control programs show that its Spanish-speaking members receiving surveys in Spanish report notably higher satisfaction compared with English-speaking members. AR72.

B. Very low reliability answers in composite measures

As noted, the score for the "Getting Appointments and Care Quickly" measure is based on two survey questions. *See 2025 Technical Notes* 63 (AR1015). For contracts H3815 and H3443, one of the two underlying questions for each had a "very low reliability" rating. Yet, the measure was not disqualified for these contracts because the pooled responses to the two questions produced a reliability score of 0.604931, just above the 0.60 "very low reliability" threshold. Similarly, the "Care Coordination" measure is based on enrollee responses to six questions. *2025 Technical Notes* 68-69 (AR1020-1021). For contracts H3815 and H3443, four of these six questions were "very low reliability," yet the combined responses produced a reliability rating of 0.60605.

C. Maximus's mishandling of two Alignment appeals

1. CMS has published a guidance document concerning the process for MAO reconsideration and IRE review, which instructs that a "non-contract provider" may "request a reconsideration (*i.e.*, a level-1 appeal) for a denied claim only if the non-contract provider completes a Waiver of Liability (WOL) statement, which provides that the non-contract provider will not bill the enrollee regardless of the outcome of the appeal." *Appeals*

Guidance 60. This Waiver of Liability "must be filed with the appeal." *Id.* In cases where a non-contract provider submits a request for reconsideration of a denied claim but fails to submit the required WOL statement, CMS directs the MA plan that receives the request to "make attempts to secure the missing documentation." But, if such attempts prove fruitless, the plan should "dismiss[] the request." *Id.* at 76.

Maximus uses a manual that sets forth "the procedures for the coordination of [MA plans] with Maximus in the processing of IRE level reconsiderations." Maximus, *Medicare Health Plan Reconsideration Process Manual* 6 (Sept. 2024) (*Reconsideration Process Manual*), https://perma.cc/7346-PRMA. Maximus stresses that where the manual uses mandatory language, "complete compliance" by the MA plan "is expected." *Id.* In line with CMS's guidance, the Reconsideration Process Manual specifies that a "non-contract provider" may become a party to a request for an MAO to reconsider a denial of coverage only if that non-contract provider "has executed a *Waiver of Liability* form," which serves to "ensure that the enrollee will not be held financially liable if the provider loses the appeal." *Id.* at 21. If an MA plan receives a reconsideration request without the WOL, then the "reconsideration review should not begin" (*id.* at 24), and the plan "*must* dismiss the request and issue a Notice of Dismissal of Appeal Request" (*id.* at 21 (emphasis added)).

In reviewing Case 1-13226962526, Maximus failed to abide the mandatory language of its own manual and CMS guidance. This case began when Alignment's delegated entity for claims processing, Arizona Priority, received a claim for coverage from a non-contract provider, Arizona Diagnostic Radiology Group LLC. The claim was for a lower extremity study, with pain in the right leg as the stated primary diagnosis. After three unsuccessful attempts to obtain medical records to support the claim, Arizona Priority denied coverage. AR19, 39. The denial notice issued to the provider included details on the

timeline to submit a request for reconsideration and explained that any such request must be submitted with a WOL form.

Arizona Diagnostic Radiology submitted a request for reconsideration to Arizona Priority, attaching medical records to support the claim, but without attaching a WOL. AR19, 39. 180 days from the claim denial date, Arizona Diagnostic Radiology filed their request for reconsideration with Alignment, again without a signed WOL. AR20, 39. Despite repeated efforts to obtain a WOL, Arizona Diagnostic Radiology never submitted the form. Consistent with CMS guidance and the Reconsideration Process Manual, Alignment therefore dismissed the request.

Because it was untimely, Alignment forwarded the case to Maximus. Instead of dismissing the reconsideration request, as its manual and CMS guidance both require, Maximus affirmed Alignment's denial of coverage on its merits. AR20, 39. Had Maximus dismissed the request for reconsideration owing to the lack of a WOL, the appeal would have been excluded from the "Plan Makes Timely Decisions about Appeals" measure. *2025 Technical Notes* 75. But because Maximus entered a determination on the merits instead, Alignment was penalized for making an untimely appeal decision, and the measure-level score for H3443 fell from 5.0 Stars to 3.0 Stars.

2. Alignment received a $254 claim from contracted provider Sonora Quest Laboratories for outpatient laboratory services. Alignment denied the claim, explaining that the enrollee had other primary health insurance coverage at the time of the claim receipt. AR9, 17-18, 39. The enrollee filed an appeal (Case 1-12757246876), stating that she no longer had other primary health insurance coverage and that her other policy had terminated before receiving care. AR17, 40. During the appeal review, Alignment confirmed using CMS's enrollment data that the enrollee had other primary health insurance coverage as of the

appeal date. The enrollee later confirmed that to be true. The same day, Alignment thus upheld the denial. AR17-18, 40.

When Alignment submitted the upheld appeal to Maximus, the IRE overturned the plan's denial of coverage, on the (mistaken) ground that Alignment had failed to coordinate coverage with the primary insurer. In fact, as explained by Maximus in its own IRE response, an MA plan can coordinate benefits with a primary insurer by, among other things, directing the provider to contact the primary insurer. AR18, 40. Alignment had done exactly that. In the remittance advice to the contracted provider, Alignment noted that the claim was "missing prior insurance carriers EOB," which under Alignment's provider manual is a direction to the provider that the enrollee had another primary insurer on CMS enrollment records. AR9. The remittance advice that Alignment sent to the provider thus instructed it to submit the claim to the primary insurance first. AR18, 40.

Had Maximus upheld the denial of Case 1-12757246876, contract H3443's score for "Reviewing Appeals Decisions" would have been 5.0 Stars rather than 3.0 Stars. Alternatively, if CMS had adjusted the denominator for "Reviewing Appeals Decisions" to ensure every plan could achieve each Star Rating for the measure, contract H3443's score for "Reviewing Appeals Decisions" would have been 4.0 Stars rather than 3.0 Stars.

ARGUMENT

Alignment is entitled to summary judgment because the CMS actions it challenges fall far short of the standards prescribed by the Administrative Procedure Act (APA).

First, the Tukey Outlier Deletion Rule is arbitrary and capricious as applied to Alignment because it produces such compressed cut points that it has become numerically impossible for Alignment to earn 4.0 Stars on some measures; with a single error, Alignment's contracts will jump straight from 5.0 Stars to 3.0 Stars.

Second, CMS's handling of "very low reliability" data for composite CAHPS survey measures is arbitrary and capricious. When a single measure comprises more than one survey question, and when the data for one of the survey questions is "very low reliability," the entire measure score must be thrown out.

Third, CMS did not adequately respond to the concerns raised during the plan preview periods that the vendor for the CAHPS survey sent English-language surveys to Spanish-speaking enrollees, resulting in a significant error in survey administration. The data disclosed in this lawsuit confirms Alignment's concern and demonstrates that CMS did not meaningfully evaluate the objection.

Fourth, CMS's allowance for "oversampling" arbitrarily and capriciously advantages larger plans. In Alignment's case, plans with objectively *lower* performance on two different measures were awarded *higher* measure-level Star Ratings simply because they paid to survey a larger number of enrollees, producing higher "significance" rankings.

Fifth, CMS has unconstitutionally delegated its regulatory authority to Maximus, a private, for-profit entity in violation of the non-delegation doctrine. Even when Congress authorizes delegations of this kind (as it has here), government officials must retain the ultimate authority to approve, disapprove, or modify the private entity's actions and decisions. Here, Maximus acts without any such supervision.

Sixth, Maximus mishandled two different appeals. As to one appeal, it ignored its own practice manual and CMS guidance, refusing to dismiss a request for reconsideration that lacked the necessary waiver of liability. And as to another appeal, it upheld Alignment's decision rather than dismissing the appeal as procedurally improper.

Any one of these errors would independently require vacatur of Alignment's 2025 Star Ratings and a remand for recalculation. Together, they may that outcome inevitable.

I. APPLICATION OF THE TUKEY OUTLIER DELETION RULE WITHOUT ADJUSTING THE MINIMUM DENOMINATOR REQUIREMENTS IS ARBITRARY AND CAPRICIOUS AS APPLIED TO ALIGNMENT

We begin with the Tukey Outlier Deletion Rule, as applied to Alignment's 2025 Star Ratings. The APA directs a reviewing court to "hold unlawful and set aside" arbitrary and capricious agency action. 5 U.S.C. § 706(2)(A). Agency action is arbitrary and capricious unless it is "reasonable and reasonably explained." *FCC v. Prometheus Radio Project*, 592 U.S. 414, 423 (2021). To meet this standard, an agency must "offer[] a satisfactory explanation for its action" and "cannot simply ignore '[any] important aspect of the problem.'" *Ohio v. EPA*, 603 U.S. 279, 294 (2024) (quoting *Motor Vehicle Manufacturers Association of United States, Inc. v. State Farm Mutual Automobile Insurance Co.*, 463 U.S. 29, 43 (1983)). The APA thus requires a reviewing court to "ensur[e] the agency has engaged in 'reasoned decisionmaking' within [the] boundaries" of its delegated authority." *Loper Bright Enterprises v. Raimondo*, 603 U.S. 369 (2024) (quoting *Michigan v. EPA*, 576 U.S. 743, 750 (2015)). CMS's application of the Tukey Outlier Deletion Rule in calculating the cut points for non-CAHPS measures fails this standard.

To justify the Tukey outlier deletion in the calculation of the cut points for non-CAHPS measure-level Star Ratings, CMS explained that it would enhance the "stability" and "predictability" of the Star Ratings, "while maintaining the intent of the cut point methodology to accurately reflect true performance." 85 Fed. Reg. at 33832; *see also Contract Year 2024 Policy and Technical Changes to the Medicare Advantage Program*, 88 Fed. Reg. 22120, 22295 (Apr. 12, 2023). "[E]xtreme outliers," CMS asserted, were unduly "influencing cut point determinations." 85 Fed. Reg. at 33833. The agency's stated "objective" in introducing Tukey outlier deletion was thus to "stabilize cut points and prevent large year-to-year fluctuations in cut points caused by the scores of a few contracts." *Id.*

But an analysis performed by Alignment demonstrated that implementing the Tukey method undermine the reliability of the measure-level Star Ratings for non-CAHPS measures. When outliers are removed, the spread from one measure-level cut point to the next is greatly compressed, meaning that even slight changes in data could produce major swings in a contract's Star Rating for that measure. *See* AR104-105.

This compression of the spread in raw scores from one cut point to the next is arbitrary and capricious as applied to contracts with small score denominators. The denominator is the number of measure-level data points from which the agency derives a measure-level raw score. For some measures, a small denominator can itself be an indication of quality—for example, a measure that evaluates the timeliness of appeals may have a small denominator if there are relatively few claim appeals in the first place.

Having few denominators (which is more likely for smaller plans) leads to reliability and fairness issues under the Tukey Rule and the compressed cut points that it produces. For example, if a 5.0 Star score for Measure A corresponds with a raw score of 98%-100, a 4.0 Star score with a raw score of 95-98%, and a 3.0 Star score with a raw score of 92-95%, a plan with a denominator of 11 data points cannot earn either 4.0 or 3.0 Stars. If it has a perfect score (11/11, or 100%), it will earn 5.0 Stars. But if it has a single adverse data point (10/11, or 91%), it will drop to 2.0 Stars. With the highly compressed cut points just described—similar to many real cut points after Tukey outlier deletion—it is mathematically impossible for a plan with 11 data points to obtain 4.0 or 3.0 Stars. Just a single error moves it to 2.0 Stars.

CMS had two options to correct for the compression of measure-level cut points. The first is to eliminate Tukey outlier deletion, which is the source of the compression—an option to agency has declined. The second is to retain Tukey outlier deletion, but to apply it

for each measure only to those contracts with a denominator large enough to ensure that it is mathematically possible to earn each Star level—1, 2, 3, 4, and 5. *See* AR105. According to this second approach, Measure A would not factor into the Star Ratings for any contract with a Measure A denominator below 25. It would apply only to contracts with a denominator of at least 25, ensuring that the contract could earn 4.0 Stars if it had one adverse response and 3.0 Stars if it had two adverse responses.

In its comment letter on the final rule (RMAR2159-2161)[2] and again during the plan preview period (AR104-105), Alignment illustrated this point. It warned that adopting the Tukey method would be arbitrary and capricious without adjustments to each measure's minimum denominator criterion. RMAR2159-2161; accord AR104-105. For measure-level Star Ratings to remain reliable, rational, and fair, it explained, CMS must adjust the minimum denominator requirement for each non-CAHPS measure to ensure that each contract can achieve each Star Rating—1, 2, 3, 4, and 5. RMAR2159-2161.

CMS acknowledged in earlier rulemakings that compression of cut points in small sample sizes can undermine reliability, noting that "minimum sample size and/or denominator requirements" are one method for "prevent[ing] unreliable data from impacting Star Ratings" and "ensur[ing] measure data are reliable." *Contract Year 2023 Policy and Technical Changes to the Medicare Advantage and Medicare Prescription Drug Benefit Programs*, 87 Fed. Reg. 27704, 27766 (May 9, 2022). CMS similarly recognized that score compression can undermine a measure's overall reliability, noting that when the scores for a given measure exhibit "decreas[ed] . . . variability across contracts," this "mak[es] the measure unreliable." 83 Fed. Reg. at 16535. Despite Alignment's comments and CMS's own recog-

[2] "RMAR" citations are citations to the administrative rulemaking record.

nition of the fairness and reliability problems caused by cut-point compression and small sample sizes, the agency declined either to adjust its minimum denominator requirement or to abandon Tukey outlier deletion. In so doing, CMS "simply ignore[d] an important aspect of the problem." *Ohio*, 603 U.S. at 294.

Compounding this defect is the fact that the agency's position was not "reasonably explained." *Prometheus*, 592 U.S. at 423. In the preamble to the final rule, CMS noted the concerns raised by commenters that outlier deletion will "move [cut points] closer together, decreasing reliability," but offered a perfunctory response. 88 Fed. Reg. at 22296-22297. The agency's rejection of commenters' views was the conclusory declaration that "[c]loser cut points do not necessarily imply lower reliability" and that "[l]essening the influence of outliers on cut point formation" through Tukey outlier deletion "leads to more reliable . . . cut points." *Id.* at 22297. That is all.

"For an agency's decisionmaking to be rational, it must respond to significant points raised during the public comment period." *Allied Local and Regional Manufacturers Caucus v. EPA*, 215 F.3d 61, 80 (D.C. Cir. 2000); *accord Perez v. Mortgage Bankers Association*, 575 U.S. 92, 96 (2015). But "an agency's ipse dixit cannot substitute for reasoned decisionmaking." *Music Choice v. Copyright Royalty Board*, 970 F.3d 418, 429 (D.C. Cir. 2020). CMS's bald assertion that the denominator issue is not a problem is thus insufficient as a response. *See Connecticut v. U.S. Department of the Interior*, 363 F. Supp. 3d 45, 61 (D.D.C. 2019) ("[C]onclusory statements will not do; an agency's statement must be one of *reasoning*."). That is to say, an agency "act[s] arbitrarily and capriciously by failing to seriously respond to comments and proposed alternatives submitted by [commenters] regarding perceived problems" with a proposed rule. *American College of Emergency Physicians v. Price*, 264 F. Supp. 3d 89, 94 (D.D.C. 2017). Just so here.

As applied to contracts with denominators too small for the contract to earn each of the five Star Ratings, the Tukey Outlier Deletion Rule is arbitrary and capricious. 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 for contracts H3443 and H9686. It should order recalculation on remand, omitting the impacted measures.

II. CMS'S USE OF QUESTIONS WITH "VERY LOW RELIABILITY" IN COMPOSITE MEASURES IS ARBITRARY AND CAPRICIOUS

The arbitrary-and-capricious standard of review requires the Court to "determine whether the agency examined the relevant data and articulated a satisfactory explanation for its action including a rational connection between the facts found and the choice made." *NRDC v. Coit*, 597 F. Supp. 3d 73, 85 (D.D.C. 2022) (Cooper, J.) (cleaned up, and quoting *Motor Vehicle Manufacturers Association v. State Farm Mutual Automobile Insurance*, 463 U.S. 29, 43 (1983)). Underlying this general rule is the commonsense observation that the data on which an agency bases its decision must be reliable. 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).

For just that reason, CMS regulations and guidance direct that when the reliability of a plan's score on a CAHPS measure falls below 0.60, the measure must be discarded as "very low reliability." 42 C.F.R. § 422.166(a)(3); *2025 Technical Notes* 189 (AR1141). Measures with "very low reliability" do not figure in contracts' Star Ratings because they are as highly likely to reflect random data fluctuations rather than "accurately . . . reflect true performance." 83 Fed. Reg. at 16519. When a plan's measure-level score is "very low reliability," it means that the agency lacks sufficient confidence that the difference in the

plan's score from other plan's scores reflect "real differences in quality" as opposed to just random "noise" in the data. 42 C.F.R. § 422.162(a).

CMS's method for computing reliability of multi-question CAHPS survey measures does not comply with these regulations. For the 2025 Star Ratings, six CAHPS measures were "composite" measures: "Getting Needed Care" (C19), "Getting Appointments and Care Quickly" (C20), "Customer Service" (C21), "Care Coordination" (C24) and "Getting Needed Prescription Drugs" (D06). CMS explained in a guidance document on the CAHPS survey that in composite measures, "answers to questions about the same topic are combined to form composites" for "scoring . . . purposes." *Protocols & Specifications* 7, 73 (AR129, 195). That is to say, a plan's raw question-level results are averaged (simply, or according to a weighted formula) into a single composite score. Some of the questions combined in these composite measures are "screener questions" that filter out certain enrollees who are factually ineligible to answer the question, thus ensuring that "only those enrollees for whom the item is relevant" will provide responses. *Id.* at 7, 27 (AR129, 149).

It stands to reason that if any of the inputs into the final, composite score are "very low reliability," then so too will be the output. A composite measure cannot be reliable if any one of the essential inputs to the measure—the data drawn from enrollees' answers to each of the component questions—is unreliable. As this Court has previously explained, use of "underlying data [that] is unreliable necessarily means that" calculations relying on that data are "equally if not more unreliable—garbage in, garbage out." *Oceana*, 530 F. Supp. 3d at 47.

CMS's use of "very low reliability" data in composite CAHPS measures offends this commonsense rule. The agency does not base its determination whether to discard a composite measure as unreliable upon the reliability of each survey question (each input) taken

individually. Instead, it calculates a measure-level reliability score using a pooled statistical model that aggregates the reliability of all component questions—even when one or more of those questions fall below the "very low reliability" threshold. This approach allows the agency to mask "very low reliability" (statistically invalid) data by blending it with other, distinct inputs. This methodology is not rationally defensible, as it simply obscures the flaws in the underlying data.

An example demonstrates the point. Imagine a two-question composite measure where Question 1 has a reliability of 0.40 and Question 2 has a reliability of 0.65. The first question would not qualify for inclusion if evaluated for reliability independently, and the second question just barely passes muster. Imagine further that Question 2 is answered by a larger and distinct subset of respondents—say, those with a more common medical condition than those prompted to answer Question 1. In CMS's aggregate measure-level reliability calculation, the larger sample for Question 2 can push the overall reliability for the measure above 0.60. But that makes no sense. Combining the statistically unreliable first input with the second input does not produce a reliable measure—it simply conceals the unreliability of the first input. This is especially problematic because CMS typically weights both questions equally when calculating the measure's *performance* score. In other words, the statistically invalid first input is given the same influence in the calculation of the composite measure-level Star Rating as the more reliable second input.

This is an especially irrational approach because the constituent questions are often answered by different populations of respondents, often of very different sizes. For instance, some questions are "screener" questions that serve a gating function (*e.g.*, "answer questions 5 and 6 only if your answer to question 4 is 'yes'"). Other times, the questions are conditional, depending on the respondent's experiences or medical conditions (*e.g.*, "if

you have visited your doctor for X condition . . ."). Each question draws responses from a distinct subpopulation. Yet, in calculating a composite reliability score, CMS treats the aggregated responses as though they came from a single, uniform group. Producing a single reliability score by blending the responses from statistically distinct and non-overlapping populations is methodologically unsound and only compounds the problem.

According to CMS's own data standards (42 C.F.R. § 422.166(c)(3)), the presence of one or more "very low reliability" inputs should disqualify the entire composite measure. Instead, CMS ignores both the unreliability of the input and the mismatch in underlying populations, using a statistically indefensible methodology to produce an artificial reliability score as a composite level, serving only to mask statistically invalid inputs.

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. CMS's method for evaluating the reliability of composite measures obscures cases in which measure scores are tainted by "very low reliability" inputs. Given the centrality of the accuracy and reliability of the underlying data to the Star Ratings system as a whole (83 Fed. Reg. at 16567), CMS's inclusion of objectively unreliable data in composite measures is arbitrary and capricious.

CMS's use of a flawed methodology had widespread adverse consequences for Alignment's Star Ratings. For contract H3815, CMS included four composite CAHPS survey measures that contained multiple inputs with "very low reliability": Getting Appointments and Care Quickly (1 out of 2 questions), Customer Service (2 out of 3 questions), Care Coordination (4 out of 6 questions), and Getting Needed Prescription Drugs (1 out of 3 questions). AR25-26, 84-85, 88-89. Similarly, for H3443, which received a 4.5-Star

rating overall, three composite measures were impacted: Getting Care Quickly (1 of 2 questions), Care Coordination (4 of 6 questions), and Getting Needed Prescription Drugs (1 of 3 questions). AR23-24, 45, 86-87. For H9686, the similar pattern holds: Getting Care Quickly (1 of 2 questions), Care Coordination (4 of 6 questions), and Getting Needed Prescription Drugs (2 of 3 questions).

In each case, CMS's methodology allowed these measures to escape disqualification on reliability grounds—even though their individual component inputs clearly fell below the 0.60 "very low reliability" threshold. If CMS had applied its rules in a rational and consistent manner, these measures would have been excluded. Instead, these statistically unreliable measures were used to calculate—and thus distorted—the affected contracts' Star Ratings. At a minimum, the Court should set aside the impacted CAHPS measures for H3815, H3443, and H9686 and remand with direction to recalculate the contracts' Star Ratings without the impacted measures. AR31-32, 94-95.

III. CMS DID NOT ADEQUATELY ADDRESS THE DROPS IN CAHPS SURVEY RESPONSES FROM SPANISH SPEAKERS

A "fundamental norm of administrative procedure requires an agency to 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)). "Unexplained inconsistency" in the treatment of similarly situated regulated entities "is a reason for holding [agency action] to be" arbitrary and capricious. 32 Wright & Miller, *Federal Practice and Procedure* § 8248, at 431 (2006) (citing *National Cable & Telecommunications Association v. Brand X Internet Services*, 545 U.S. 967, 981 (2005)).

CMS guidance states that approved contract vendors "must" make "Spanish-language questionnaires available to all Spanish-speaking enrollees" (*Protocols & Specifica-*

tions 50-51 (AR172-173)), and the agency must require compliance uniformly for all plans. Here, if errors by CMS or its approved survey vendor resulted in sending English-language surveys to Alignment's Spanish-speaking enrollees, and the error adversely impacted the contracts' CAHPS measure scores, the agency has unlawfully relieved the contractor of compliance with agency guidelines and subjected Alignment to different treatment.

That is what the record demonstrates. As we noted in the Background section of this brief (at 19-20), data disclosed by CMS in this lawsuit demonstrate that a material number of enrollees designated by Alignment as Spanish speakers received English-language surveys, contrary to Alignment's language preference designation. For contract H3443, 29 of 200 Spanish-speaking enrollees received surveys in English. For contract H3815, 20 out of 201 did. These data, which are undisputable and clear, show that CMS simply was incorrect when it stated that the data were "consistent with the correct administration of the survey" and that it had "verified" correct administration. AR53.

The same data show that enrollees who were (1) identified by Alignment as having a preference to receive the survey in Spanish but (2) nonetheless received the survey in English were the least likely to complete the survey. Moreover, Alignment's internal quality-control data suggest that its Spanish-speaking enrollees receiving surveys in Spanish report higher satisfaction compared with English-speaking members. AR72. Thus, the survey administration error here artificially depressed the response rates from those members most likely to report favorably on Alignment's plans. *See* AR54-55.

In refusing to correct this problem, CMS arbitrarily treated Alignment differently from other MAOs by grounding Alignment's Star Ratings on CAHPS survey data based on English-language surveys sent to Spanish-speaking enrollees.

It is no answer to say, as CMS did in the plan preview period, that the agency "does not get involved in how survey vendors implement language preference data," which is an issue "outside of CMS control." AR58. As a starting point, CMS's approved survey vendors are *required* by CMS's own guidelines to make Spanish-language questionnaires available "to all Spanish-speaking enrollees" in the "mail . . . administration" of the CAHPS survey. *Protocols & Specifications* 50-51 (AR172-173); *Training* 100. Those guidelines would be meaningless if, when a vendor fails to follow them, the agency can simply disavow any obligation to enforce the guidelines on the ground that it "does not get involved" with such matters. Ensuring that its approved survey vendors comply with its guidelines for CAHPS survey administration is plainly the agency's job.

In any event, the underlying reason for the survey administration error is irrelevant, whether it is CMS's fault or the vendor's. The point is simply that *an error occurred*. And when an MAO like Alignment brings a glaring and material survey administration error to the agency's attention, it assuredly does not suffice for CMS effectively to say "oh well" and count the survey results nonetheless. Neither does it suffice for the agency to claim, incorrectly, that it has verified that the survey was properly administered.

Again under CMS's own guidelines, the CMS-approved survey vendor must, "at the request of the contract," send Spanish-language questionnaires to enrollees identified using "language preference data received from the contract." *Protocols & Specifications* 50-51 (AR173); *Training* 101. The decision to send Spanish-language surveys to specific enrollees flagged as Spanish speakers is not optional for the vendor; the only other option would be to send Spanish-language surveys to *all* respondents. *Protocols & Specifications* 50 (AR172). Because the vendor did not administer the CAHPS survey for contracts H3443 and H3815 in a manner consistent with CMS guidelines or Alignment's designations, the

response rates and survey scores for contracts H3443 and H3815 were artificially depressed. AR54-55, 72-73. Moreover, 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.

IV. CMS'S METHODOLOGY FOR CALCULATING STAR RATINGS FOR CAHPS MEASURES ARBITRARILY FAVORS LARGER CONTRACTS

"Agency 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 has not provided any justification for adopting a significance testing method that favors or disfavors contracts based on the irrelevant factor of their number of enrollees. Here, CMS's significance-testing methodology to calculate the measure-level Star Ratings for CAHPS measures arbitrarily disadvantages smaller plans that have fewer enrollees and are unable to oversample. Because the agency has not given a rational explanation for treating similar plans differently, its allowance for oversampling must be declared unlawful.

The arbitrariness of CMS's approach to oversampling is clear in the facts of this case. On at least two CAHPS measures, contract H3815 earned a *higher* mean score than, and the same reliability assessment as, other contracts—but it received a *lower* measure-level Star Rating because Alignment did not engage in oversampling. Contract H3815 received a lower measure-level rating for "Rating of Drug Plan" than did contracts H1036, H1889, and H5386, despite that it received a higher raw score than those other contracts. It also received a lower measure-level Star Rating on "Rating of Health Plan" than did

contract H1036, despite, again, that it had a higher mean score on the measure. On both measures, H3815 received 4.0 Stars while the oversampled contracts received 5.0 Stars—despite that H3815 achieved a higher mean score and had the same reliability.

These other plans achieved a higher Star Rating on these two measures because they engaged in oversampling, obtaining between about 1200 and 3000 responses to their CAHPS surveys and thus a higher "significance" score. That is the only difference, and it has no bearing whatsoever on the quality of the underlying plan. Simply put, the significance testing step in CMS's methodology for calculating measure-level Star Ratings for CAHPS measures arbitrarily rewarded these other plans with higher Star Ratings simply because they are larger and are permitted to engage in oversampling. CMS allows contracts with greater than 800 CAHPS-eligible enrollees to oversample if "there is sufficient eligible enrollee volume to support additional sampl[ing]" (*Protocols & Specifications* 18 (AR140)), which in practice means that only contracts with multiples more than 800 CAHPS survey-eligible enrollees are able to oversample.

The ability and willingness to engage in oversampling does not indicate higher quality. And if CMS did not use a significance testing method that arbitrarily favors larger contracts, Alignment's H3815 contract would have received 5.0 Stars on the "Rating of Drug Plan" and "Rating of Health Plan" measures.

Accordingly, Alignment's 2025 Star Ratings for H3815 should be set aside as arbitrary and capricious. The Court should remand the matter to the agency with directions to recalculate Alignment's 2025 Star Ratings for that contract without consideration of statistical significance thresholds that favor oversampled contracts.

V. CMS HAS UNCONSTITUTIONALLY DELEGATED GOVERNMENTAL AUTHORITY TO MAXIMUS, A PRIVATE ENTITY

The legal problems with the 2025 Star Ratings run yet deeper, because CMS has unlawfully delegated its regulatory authority to Maximus. The law is settled that "[f]ederal lawmakers 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 Transportation*, 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)). "When it comes to [delegations to] private entities, . . . there is not even a fig leaf of constitutional justification." *Id.*, 575 U.S. at 62 (Alito, J., concurring). "Private entities are not vested with . . . the 'executive Power,' which belongs to the President." *Id.* Yet CMS treats Maximus as if *were* vested with executive power—and in practice, it is.

The baseline principles are familiar: While private entities may "help a government agency make its regulatory decisions," "private parties must be limited to an advisory or subordinate role." *Association of American Railroads*, 721 F.3d at 671, 673. Accordingly, "[f]or a 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," and government officials must "retain[] the ultimate authority to 'approve, disapprove, or modify' the private entity's actions and decisions on delegated matters." *Alpine Securities Corp. v. Financial Industry Regulatory Authority*, 121 F.4th 1314, 1325 (D.C. Cir. 2024) (quoting *Sunshine Anthracite Coal Co. v. Adkins*, 310 U.S. 381, 388, 399 (1940)) (cleaned up). The private entity must remain subject to the "authority and surveillance" of the government. *Association of American Railroads*, 721 F.3d at 546 (quoting *Adkins*, 310 U.S. at 399).

Here, however, CMS—at Congress's direction—has delegated unchecked, unsupervised authority to Maximus to perform independent reviews of MAOs' adverse coverage determinations. Maximus's decisions are subject to further review via a Section 405(b) ALJ hearing only if it is requested by the enrollee or the enrollee's healthcare provider in the event Maximus upholds a *denial* coverage. But its decisions are not subject to further review of any kind if Maximus determines that coverage should *not* be denied. That is to say, decisions by Maximus *against* an MAO's 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*, 121 F.4th at 1325 (quoting *Adkins*, 310 U.S. at 399).

CMS also delegates authority to Maximus to resolve issues raised by MAOs regarding the Star Ratings calculations and underlying data for the "Plan Makes Timely Decisions about Appeals" and the "Reviewing Appeals Decisions" measures. CMS directs MAOs to the Maximus website and a Maximus email address for resolution of any questions about the data, data discrepancies, or other concerns. AR12. It is then up to Maximus, again without meaningful CMS supervision, to exercise executive judgment to resolve issues.

The IRE review process and Maximus's role in resolving disputes regarding the calculation of Star Ratings are both straightforwardly unlawful delegations of executive authority to a private entity. *Association of American Railroads*, 721 F.3d at 671, 673.

The Court should declare as much and set aside Alignment's 2025 Star Ratings for contract H3443, which was adversely affected by the delegation. Having done so, it should remand the matter to CMS with directions to recalculate the 2025 Star Ratings for contract H3443, with full and non-delegated agency review of the "Plan Makes Timely Decisions about Appeals" and the "Reviewing Appeals Decisions" measures.

VI. MAXIMUS'S ACTION ON TWO OF ALIGNMENT'S APPEALS WAS CONTRARY TO ITS AND CMS'S RULES AND GUIDELINES

Not only is CMS's unsupervised employment of Maximus a violation of the non-delegation doctrine, but Maximus has committed at least two substantial errors concerning Alignment contracts, resulting in unlawful reductions in Alignment's 2025 Star Ratings.

A. Maximus's refusal to dismiss the reconsideration request in Case 1-13226962526 was arbitrary and capricious

First, Maximus's refusal to dismiss the reconsideration request in Case 1-13226962526, rather than deciding the case in Alignment's favor, was arbitrary and capricious. "An agency may not of course depart from prior policy without explanation." *ANR Pipeline Co. v. FERC*, 205 F.3d 403, 407 (D.C. Cir. 2000). And, again, "a fundamental norm of administrative procedure requires an agency to treat like cases alike." *Consolidated Edison*, 45 F.4th at 279.

CMS guidance concerning requests for reconsideration and IRE review instruct that a request for reconsideration must be dismissed if not accompanied by a WOL, or Waiver of Liability. *Appeals Guidance* 76. Maximus's Reconsideration Process Manual contains the same instruction. *Reconsideration Process Manual* 21.

Case 1-13226962526 involved a decision by Alignment's plan administrator for contract H3443 to dismiss a request for reconsideration. The dismissal was made because the request for reconsideration was not accompanied by a WOL, despite numerous requests for the same. AR19-20, 39. Alignment forwarded the matter to Maximus for a level two appeal. By CMS's and its own guidelines, Maximus should have dismissed the request for failure to include a WOL, as had Alignment.

But Maximus did not dismiss the ensuing request for reconsideration, as it has in other similar cases and as all applicable rules require given the absence of the WOL.

Maximus instead issued a reconsidered determination *on the merits*, affirming the underlying adverse coverage determination. Maximus did not provide any explanation for deviating from the normal policy applied in similar cases.

The decision to resolve Case 1-13226962526 on its merits, rather than dismissing it on the procedural ground identified, meant that it was included in CMS's calculation of the "Plan Makes Timely Decisions about Appeals" measure for contract H3443. Inclusion of Case 1-13226962526 in that measure resulted in a lower Star Rating for the contract.

The 2025 Star Ratings for contract H3443 thus should be set aside. The Court should remand the matter to the agency with directions to recalculate H3443's 2025 Star Ratings, excluding Case 1-13226962526 from the calculation of the measure-level score for the "Plan Makes Timely Decisions about Appeals" measure.

B. Maximus wrongly reversed the denial of Case 1-12757246876

Second, Maximus's reversal of the denial of coverage in Case 1-12757246876 on the grounds that Alignment failed to coordinate coverage with the enrollee's primary insurer was "unsupported by substantial evidence." 5 U.S.C. § 706(2)(E).

On this score, Alignment denied a claim for outpatient laboratory services on the ground that the enrollee had other primary health insurance coverage. AR9, 17-18, 39. Alignment upheld its decision in a level one appeal (AR17-18, 40) and forwarded the case to Maximus for a level two appeal. Maximus, overturned the denial of coverage on the erroneous ground that Alignment had failed to "coordinate coverage" with the primary insurer. But as Maximus has recognized, an MA plan may coordinate coverage with a primary insurer by, among other things, directing the provider to contact the primary insurer. AR18, 40. Alignment did so. In the remittance advice to the contracted provider, Alignment noted that the claim was "missing prior insurance carriers EOB," which under Alignment's

provider manual is a direction to the provider that the enrollee had another primary insurer on CMS enrollment records. AR9. The remittance advice that Alignment sent to the provider thus instructed it to submit the claim to the primary insurance first. AR18, 40.

Inclusion of Case 1-12757246876 as an overturned denial in calculating the "Reviewing Appeals Decisions" measure for contract H3443 resulted in a 3.0 measure-level Star Rating, rather than a 5.0 Star Rating. The 2025 Star Ratings for contract H3443 thus should be set aside. The Court should remand the matter to the agency with directions to recalculate the contract's 2025 Star Ratings, including Case 1-12757246876 as "upheld" in the calculation of the score for the "Reviewing Appeals Decisions" measure.

CONCLUSION

The Court should grant Alignment's motion for summary judgment and declare that (1) the Tukey Outlier Deletion Rule is arbitrary and capricious as applied to Alignment without appropriate denominator adjustments; (2) CMS's inclusion of "very low reliability" data in composite CAHPS survey measures is arbitrary and capricious and contrary to law; (3) CMS acted arbitrarily and capriciously and contrary to law by including CAHPS survey data infected by a major administration error in Alignment's 2025 Star Ratings; (4) CMS's allowance for "oversampling" arbitrarily and capriciously disadvantages smaller plans; (5) CMS has unconstitutionally delegated its regulatory authority to Maximus, a private, for-profit entity in violation of the non-delegation doctrine; and (6) Maximus's final decisions for Case Nos. 1-13226962526 and 1-12757246876 were arbitrary, capricious, and contrary to law.

Against that backdrop, the Court should vacate Alignment's 2025 Star Ratings and remand with instructions for CMS to recalculate the Ratings for all Alignment's impacted contracts, consistent with the Court's findings and holdings.

Dated: April 7, 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 April 7, 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. 25-cv-0074-CRC

[PROPOSED] ORDER

Having considered plaintiff Alignment Healthcare Inc.'s (Alignment's) Motion for Summary Judgment and the Memorandum of Points and Authorities filed in support thereof, together with any opposition filed in response and any reply filed in the support, the Court hereby **ORDERS** that Alignment's Motion for Summary Judgment is **GRANTED**.

The Court **DECLARES** that (1) the Tukey Outlier Deletion Rule is arbitrary and capricious as applied to Alignment without appropriate denominator adjustments; (2) CMS's inclusion of "very low reliability" data in composite CAHPS survey measures is arbitrary and capricious and contrary to law; (3) CMS acted unlawfully and arbitrarily by including CAHPS survey data infected by a major administration error in Alignment's 2025 Star Ratings; (4) CMS's allowance for "oversampling" unlawfully and arbitrarily disadvantages smaller plans; (5) CMS has unconstitutionally delegated its regulatory authority to a private, for-profit entity in violation of the non-delegation doctrine; and (6) the independent review entity's final decisions for Case Nos. 1-13226962526 and 1-12757246876 were arbitrary, capricious, and contrary to law.

The Court further **ORDERS** that Alignment's 2025 Star Ratings for all of Alignment's adversely impacted contracts are **VACATED**, and that this matter is **REMANDED** to CMS to recalculate the 2025 Star Ratings and quality bonus payments for all adversely impacted contracts, consistent with the Court's findings and holdings.

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