Shawn J. Organ
Organ Law LLP
1330 Dublin Road
Columbus, Ohio 43215

October 7, 2021

Re: *Paramount Advantage v. Ohio Department of Medicaid, et al.*, Franklin County Common Pleas. Case No. 21-CV-04337

Dear Shawn;

This letter constitutes my report on the bias analyses I performed as an expert witness on the Ohio Department of Medicaid (ODM) evaluation process for selecting new Medicaid managed care organizations (MCOs). My goal was to assess whether the process was created and implemented in a way that protected it from bias, as well as to identify any sources of bias that may have impacted the evaluations.

Context

To provide some context, below are three cases from literature on challenges to procurement processes due to the process not being protected from bias, taken from *The Art of Tendering: A Global Due Diligence Guide* (2021).

Consider a case where in its July 2015 decision in *Problem Gambling Foundation of New Zealand v. Attorney General*, the High Court of New Zealand struck down a contract award after finding flaws in the government's evaluation and award process. The dispute dealt with a Request for Proposals (RFP) for treating problem gamblers. The Court found that the government improperly conducted group consensus scoring and lowered the complainant's score based on undisclosed "alignment with strategic plan" criteria during a non-transparent "moderation stage". These evaluation problems were compounded by the government's failure to maintain contemporaneous evaluation records. While its external procurement consultants advised the government to retroactively rectify those records, the Court was unimpressed with the government's after-the-fact efforts to fill in those record-keeping gaps. The Court concluded that the government's "moderation stage" breached the RFP rules since it had the effect of nullifying the results from the prior stages of the government's evaluation process and stressed

that government institutions must maintain lineal integrity throughout an evaluation process to ensure the defensibility of their ultimate contract award decisions.

Consider another, from June 2015, *In the Matter of An Appeal by Dynalife*. A bid dispute panel found significant irregularities in an Alberta Health Services (AHS) request for proposals (RFP) evaluation process. The panel report nullified the results of the evaluation process that had initially resulted in the award of a $3 billion laboratory services contract. The Alberta Minister of Health subsequently announced the cancellation of the contract award. The panel found that AHS failed to clearly document the process details, finding a need for greater transparency in AHS's evaluation, ranking, selection, and award process path. The panel was also highly critical of the apparently subjective "Good Fit Question" used by AHS evaluators (referred to by the complainant as "choosing your own evaluation"). The ruling highlighted the need to avoid any criteria or process rules that appear to introduce an element of subjectivity or arbitrariness into an evaluation process.

In its May 2018 decision in *Lancashire Care NHS Foundation Trust & Anor v. Lancashire County Council*, the England and Wales High Court – Technology and Construction Court ordered a re-evaluation after finding flaws in the original consensus scoring results of the government evaluators. The dispute dealt with a procurement process for public health nursing services conducted by a county council. As the Court noted, the organization's procurement department maintained the evaluation records and moderated the consensus scoring sessions. However, since the group consensus scoring discussions resulted in changes to evaluator scores, the Court raised concerns that the initial evaluator notes were not reliable in documenting the reasons for the final adjusted scores. As the Court noted, the evaluators failed to keep their own records of the reasons behind their changed scores and instead relied on the group notes taken by the moderators. The Court also found that the evaluation group failed to document whether it properly considered the complete list of sub-criteria for each evaluation category during their group discussions since they did not maintain an adequate record of those discussions. As the Court stated, "a procurement in which the contracting authority cannot explain why it awarded the scores which it did fails the most basic standard of transparency." The Court concluded that the "inconsistency in approach in the recording of the moderation of different questions in each tenderer's bid means that it is not possible to identify a structure in the notes which reveals the reasoning process adopted by the panel that led to and explains their consensus scores on a given question." Overall, the Court found that the evaluation notes of the discussions that led to the consensus scores did not provide a full, transparent, or fair summary and therefore did not enable the losing proponent to assert its review rights or enable the reviewing court to exercise is supervisory jurisdiction. The Court concluded that the resulting contract award should be set aside as unlawful.

What about in the US? The Government Accountability Office put out a report in 2017 entitled "GAO Bid Protest Annual Report to Congress for Fiscal Year 2016," which found that the number of sustained protests (rate of protest wins) was 22.56% in 2016. According to the GAO report, "a flawed decision in selecting the contract awardee" due to bias is one of the top reasons protests were upheld in 2016.

How should we think about addressing bias in the government procurement process from a judicial perspective? According to an article on May 27, 2021 by Dr. Christopher R. Yukins in *The Regulatory Review*, in the past, U.S. courts have often treated bid protests as a means of remedying bidders' injuries. But a recent study sponsored by the Administrative Conference of the United States (ACUS) and endorsed by the U.S. Congress has helped to frame a new narrative. Namely, bid challenges are risk-reduction devices that lend governments early notice of system failures. The study is called "Stepping Stones to Reform: Making Agency-Level Bid Protests Effective for Agencies and Bidders by Building on Best Practices From Across the Federal Government." Thereby, if a protester is really a whistleblower rather than an injured party seeking monetary redress, a judicial stay of the procurement pending the outcome of the protest becomes more important. Controversial as a stay might be, a stay preserves the protesting contractor's chance at award—the linchpin to the protester's willingness to challenge its government customer. By treating bid challenges as risk-reduction measures that ultimately can benefit the government, ACUS clarifies a fundamentally different approach to bid protests—one grounded in reducing risk in public procurement.

What is Bias?

To understand whether bias was indeed present in the 2021 ODM MCO procurement process, first, we need to define what bias means. Bias refers to using inappropriate criteria or information in decision making. That might mean deliberate bias, meaning a preconceived, prejudiced perspective intended to weigh for or against certain choices. However, bias is often implicit, where the decision maker does not realize they are making biased judgments. In fact, they might be trying their best to make the most accurate possible decision, but factors related to the structure of the process and context of their evaluations lead to undeniably biased conclusions. This is why to prevent bias, any process must be structured to address both deliberate and unintentional bias.

As I describe in two of my books, *Never Go With Your Gut: How Pioneering Leaders Make the Best Decisions and Avoid Business Disasters* (Career Press, 2019), and *The Blindspots Between Us: How to Overcome Unconscious Cognitive Bias and Build Better Relationships* (New Harbinger, 2020), our brains are unfortunately not adapted for the modern environment. Instead, they're adapted for the ancestral savanna environment, when we lived as hunter-gatherers in small tribes of 50-150 people. Our decision-making tendencies are powerfully impacted by the more primitive, emotional and intuitive part of our brains that ensured survival in our ancestral environment. This quick, automatic reaction of our emotions represents the autopilot system of thinking, also known as System 1, one of the two systems of thinking in our brains. It makes good decisions most of the time but also regularly makes certain systematic thinking errors that scholars refer to as cognitive biases. The other thinking system, known as the intentional system or System 2, is deliberate and reflective. It takes effort to turn on but it can catch and override the thinking errors committed by our autopilots. This way, we can address the systematic mistakes made by our brains in our workplace relationships and other areas of life. Keep in mind that the autopilot and intentional systems are only approximations of more complex processes, but for our decision making in real life, this systems-level approach is very useful in helping us address

bias in our evaluations and choices. It's very possible to address such errors by using debiasing techniques, if we build them into our systems and processes.

2021 ODM MCO Procurement Process

Protections From Bias

What would it have meant to protect the ODM MCO procurement process from bias?

A basic protection is having clearly-established, thoroughly-defined criteria for evaluation that are made available in advance to the applicants. After all, if you want a process that is transparent, unbiased, and fair, and gets the best possible applications, you will want to help applicants do their best to address the criteria.

Otherwise, you are liable to fall into cognitive biases such as anchoring bias, where you are anchored to the initial information you have about a topic. That initial information would cause an evaluators to develop their own idiosyncratic interpretations of applications, rather than following set criteria. Another dangerous cognitive bias here is the confirmation bias, where you look for information that confirms your beliefs and ignore information that does not. Thus, evaluators would look for information that confirms their beliefs about what the criteria should be, rather than question this information. A third bias is called availability bias, where we look for information that is most easily available in our memory. Having clear criteria that reminds us of additional information that we need to address helps address this problem. A fourth serious cognitive bias here is the overconfidence bias, our tendency to be way too confident about our decisions. Having clear criteria helps us avoid overconfidence by causing us to go through a variety of prompts to help us be less overconfident about our own judgments. A fifth bias is the egocentric bias, where our brains tend to rely too strongly on our own perspective. The presence of clear criteria causes us instead to orient toward an external, objective perspective. A sixth bias is called the false consensus effect, where we overestimate the extent to which others agree with us. This is illustrated by evaluators falsely believing that they share the same definitions of vague and broad criteria that is held by other evaluators, applicants, and stakeholders in general.

Another basic protection is having a blinded process where the evaluators don't know the identity of those they are evaluating. This does not mean that the applicants need to hide their capabilities, just not mention their name in the applications. Certainly, evaluators might try to guess, but the fact that there were 3 Ohio-based applicants and 8 national applicants makes it difficult to do so.

Biases relevant here include the belief bias, where our beliefs about the outcomes sway our process of decision making. Thereby, if we believe certain applicants should or should not be awarded a contract, we will follow through on that belief in our decision making, often without realizing it. A related pair of biases is the halo effect and the horns effect. The halo effect refers to the fact that when we like a certain characteristic of an option, we will tend to have too-favorable views of the option as a whole without realizing it; the horns effect refers to how when we have a negative perception of one characteristics of an option, we will tend to have too-

negative views of the option as a whole. If we know the identity of certain applicants, we will then let our impressions of their characteristics sway the outcomes. Some of same biases already discussed above apply to this basic protection as well. Anchoring is an example: if we know the identity of applicants, we will be anchored to our initial impressions and information about them. Confirmation bias will cause evaluators to look for information confirming their pre-existing beliefs. Availability bias will lead evaluators to look for information that matches what they have available in their memory.

A third basic protection is avoiding discussing anything that's not an outlier in the consensus meeting, in other words what's called "enhanced consensus scoring," where only scores that lie outside of the pre-established range of evaluation are discussed and reassessed. Consensus group meetings are notoriously flawed mechanisms for decision making.

Such meetings are very vulnerable to biases such as groupthink, where the desire for consensus among a group of people – especially those who are familiar with each other and where there are people with more or less power – leads to a biased outcome due to unintentional (or sometimes intentional) peer pressure effects. A related bias is the bandwagon effect, where over time more and more group members start to align with the opinion of the group. Another bias that is seriously threatening is called the authority bias. If there is a source of authority, such as a facilitator, this authority may powerfully sway the evaluators and lead to a biased outcome. That biased outcome may result from the authority having potential conflicts of interest, and falling into self-serving bias, the predilection to assess ambiguous data as aligning with your own interest. But it might also result from the authority having unintended biases. One problematic unintended bias is the mere exposure effect: the more familiar an external authority is with an applicant, the more the source of authority trusts and likes the applicant, unless there is a serious conflict between the authority and the applicant.

A fourth basic protection from bias is having clear evaluation records. That includes individual evaluators justifying thoroughly the reasons for giving a score on the individual assessment component. Having written justifications forces evaluators to thoroughly consider their own reasoning and defend it against potential observers, which helps minimize the biases described above. That also includes a thorough recording of the deliberations at a consensus meeting. Again, such recording leads both evaluators and facilitators to consider potential external observers and thus minimize biases.

A fifth basic protection is having diverse external evaluators, rather than evaluators who: 1) belong to the same organization; 2) especially not those who belong organization making the request for applications; 3) and you want to avoid at all costs evaluators who were involved in formulating the questions.

If all evaluators belong to the same organization, you fall into the common source bias, where all evaluators have the same perspectives induced by a shared organizational culture. Another dangerous bias is the conformity bias, where all tend to conform to the same vision and perspectives: good for organizational cohesion, bad for unbiased decision making. A third dangerous problem is that all belonging to the same organization makes groupthink much more likely. All evaluators working at the organization making the request for proposals results in the

threat of belief bias, where a shared belief about what the organization wants from the process leads to a biased evaluation of applicants – especially if unblinded. Evaluators being involved in formulating the questions results in certain specific evaluators having undue authority due being able to interpret for others what the question "really means," thus resulting in deviations from unbiased decision making due to the authority bias.

A seventh basic protection is ensuring that no key stakeholders have any material conflicts of interest with any of the applicants. The potential for self-serving bias and belief bias are evident.

A eighth basic protection is a statistical "sanity check" on the difference between individual evaluator scores and consensus meeting scores. Are there any statistical aberrations that might suggest to a reasonable outside observer the possibility of a biased outcome? The standard best practice of accepted variance is $P > 0.05$, meaning that there's a less than 5% chance that the results of the meeting led to biased outcomes.

A ninth basic protection is a performance "sanity check." That involves looking at the ranking of applicants and, for incumbents, comparing the ranks to their performance evaluations in the past few years. After all, anyone can put whatever they want on the application, but the proof is in the pudding: past performance is the best indicator of future results.

The final, tenth basic protection is to build in deliberate defenses against cognitive biases leading to biased outcomes. That involves, first, developing an understanding of cognitive biases and debiasing techniques. Then, it involves evaluating the various sources of bias, for instance the ones described above, and taking intentional steps to address them effectively.

A Fatally Flawed, Biased Process

Was the 2021 ODM MCO procurement process protected against bias? Unfortunately, the evidence suggests the opposite, pointing to the public interest need to help address the risk to public procurement in Ohio represented by this fundamentally flawed and biased process. ODM failed to address a number of obvious sources of bias in awarding over $20 billion in Medicaid funding. ODM hired a well-known, highly-credible consulting firm, Mercer, to manage the process, and it is not credible to believe that Mercer does not know how to mitigate bias. Given that this contract is the biggest sum ever awarded by the State of Ohio, if the courts permit this fatally flawed process to stand, it provides a precedent that threatens to normalize biased procurement as a standard practice of the State of Ohio, paving the way to a deterioration of trust in our public institutions and even opening the door for corruption to flourish.

Consider the first basic protection: clearly-established, thoroughly-defined criteria for evaluation that are made available in advance to the applicants. The criteria, as defined both for the applicants and for the evaluators, were "methods of approach," "capability," and "experience," in that order of importance. These criteria are very broad and vague, and it is not credible that any external observer would agree on what they mean. Moreover, internal stakeholders disagree what they mean. Maureen Corcoran, the Director of ODM, discussed in testimony for the Ohio Senate Finance Committee on April 15, 2021, that "experience" was a category that would favor

incumbents because “they have knowledge of Ohio, they have experience with our population.” That’s an understandable definition. In the notes about a call on 6/8, there was a discussion about whether to add experience to a question, but that was met with pushback because of the desire to “maintain an even playing field.” Again, this definition of experience implies favoring incumbents. However, Jim Tassie, ODM Deputy Director, Project Management and Procurement Implementation, who sat in on the consensus meetings and answered questions without doing evaluations himself, gives a somewhat different answer when asked to define “experience”: on pg 155 of his deposition, he says it means that experience means “they have done it before.” This definition is also understandable, but in no way references previous knowledge of Ohio or experience with the Ohio population, since having “done it before” might mean anywhere across the US and being question-specific rather than focusing on the state. Mary Applegate, an evaluator and Medical Director of ODM, defines experience similarly to Tassie, “have you done it before.” Shaun Bracely, an evaluator and Chief OhioRISE Operations Compliance and Oversight, goes in the exact opposite direction. He says in his deposition testimony on pg 136-7 that he would give more points for experience if insurance providers operated outside of Ohio, rather than just in Ohio. When looking at individual applicant results, several had listed “Experience providing comparable service, populations served, and size to that of the Ohio Medicaid Program.” However, none of the Ohio-based applicants had that description of a strength, even though both CareSource and Paramount actually provided these very exact services to the Ohio Medicaid Program. The same confused, changing, and even contradictory definitions apply to the other categories of “capabilities” and “method of approach.”

This confusion is the very definition of failing to protect from bias. There is no realistic way to check for bias, intentional or unintentional, because criteria definitions are so supremely unclear. More than that: they are contradictory, such as “experience” meaning for some to be “experience” in Ohio and for others “outside of Ohio.” Moreover, while the rating weight was supposed to be “method of approach,” “capabilities,” and “experience,” in that order, who knows how to weight them appropriately? Is it 50% for “method of approach,” 26% for “capabilities,” and 24% “experience?” Is it 97% for “method of approach,” 2% for “capabilities,” and 1% “experience?” We don’t know. This is exacerbated by the fact that some questions reversed that order, such as question 22, which was supposed to be scored on the following weighting according to the Master Evaluation Guide: “method of approach,” “experience,” and “capabilities,” in that order. That’s another area rife for confusion: should evaluators use what they were told in the training or what the prompt indicates? Any evaluator – or facilitator – who has any ulterior motive of deliberate bias, or even unintentional bias, can fully and thoroughly justify their biased evaluations through their idiosyncratic definition of the criteria.

The second basic protection against bias is missing. Given everyone knows who the applicants are, anyone – evaluator and facilitator alike – can sway the process based on their intentional or unintentional bias.

In fact, there is reason to believe bias may play a role.

Paramount had a series of conflicts with ODM in 2019 over Medicaid patient assignments. All the evaluators came from ODM. That provides a clear reason to lower the scores of Paramount intentionally, not because of the criteria but because of factors external to the evaluation process.

It might also result in unintentional lowering due to the horns effect. Did that intentional or unintentional lowering happen? We can't know without reference to clear criteria that would be able to demonstrate bias or not.

Similarly, according to an October 6, 2021 article in the *Ohio Capital Journal* entitled "Ohio's Medicaid director owns the stock of some major contractors, but won't say how much," Corcoran owns some UnitedHealth stocks. Given all evaluators are from ODM, there is a natural unintentional halo effect threat of them wanting to please their boss, and perhaps even an intentional bias if Corcoran deliberately asked some to rate UnitedHealth higher. We just can't know without reference to clear criteria that would be able to demonstrate bias or not.

The third basic protection is missing of avoiding discussing anything that's not an outlier in the consensus meeting, in other words what's called "enhanced consensus scoring," where only scores that lie outside of the pre-established range of evaluation are discussed and reassessed. In fact, Kendallyn Markman, an evaluator, Bureau Chief of Clinical Analytics at ODM, said that Mercer failed to follow the process described at the training. According to the "Evaluation Committee Training" document, in the consensus meeting, if different ratings were assigned, only then would the facilitator ask members to explain rationale. Yet Markman reports, on pg 103 of her deposition, that even if everybody got the same score, the Mercer facilitators would still lead everyone into a discussion of strengths and weaknesses, and then reassess everyone's score to come to a consensus. Is it any surprise, then, that in question 32, all evaluators rated Paramount's application as a 3, but the final consensus score was a 2?

Mercer had a powerful opportunity to bias the scoring. Evaluator Bracely describes in his deposition testimony on pg 116 that Mercer staff "had the scoring guide. And if our comments didn't necessarily match what the scoring said, then they would probe us to say, you know, your comments sound more like this rating. You know, is there something else you would like to say or add, or do you – do your comments still really support the rating that you've given it? And so they would prompt us to re-evaluate if necessary." That's a perfect description of conformity to the external source of authority, a fundamentally powerful authority bias. Is it any wonder that Paramount's scores, on 16 out of 32 questions, were inexplicably lowered during the consensus meeting in such a way that the statistical report by Dr. William Nots finds incredibly unlikely to happen by random chance, a less than 1% chance probability?

Did Mercer have reasons to sway the scoring? Certainly. Mercer's prospective clients – and potentially current ones – include nationwide insurance giants such as UnitedHealth and others. If these insurance giants did not make it through the application process, they would be less likely to hire Mercer for consulting services.

Mercer's scoring bias may be unintentional as well. The mere exposure effect – unless guarded against – would lead Mercer to evaluate nationwide insurance giants more favorably than regional bidders, to which Mercer staff would have much less exposure.

The fourth basic protection is not present, of clear evaluation records. Clear records of the consensus meeting would have enabled us to assess whether or not Mercer's opportunity to sway

the consensus meeting actually swayed the consensus meeting. Without these records, we only have evidence of clear opportunity for bias and clear motivation for bias, intentional and not.

The fifth basic protection is not present. All the evaluators belong to ODM, instead of having independent external evaluators, such as Medicaid staff from neighboring states or academics. They are the ones who solicited the application. Several of them were involved in formulating the questions, such as Applegate and Beatty, per the document "Summary of Comments: Draft 1 MCO RFA Questions to ODM – notes from 6/1 meeting."

The seventh basic protection is not there, given financial interest by Corcoran and Mercer's staff.

The eighth basic protection statistical "sanity check" fails on the merits. Paramount's lowered scores in the consensus meeting fail basic probability likelihoods: in the most generous interpretation, there is a less than 1% chance that the extent of the lowering happened by chance.

The ninth basic protection of a performance "sanity check" also fails on the merits. In 2020, CareSource and Paramount tied for top performance in the Quality 360, with UnitedHealth last. The same thing happened in 2019. Yet mysteriously, UnitedHealth ended up being top of the rankings by far, while Paramount did not make it in, and CareSource barely squeaked in.

What about the final protection of defending against bias? Well, Applegate describes in her deposition testimony on pg 73 how, for awarding $20 billion in Medicaid funding, "removing bias would be another thing that's important," along with consistency. Applegate goes on to add "this process was designed to be as objective as possible." Tassie, on pg 190 in his deposition testimony, says that "the entire [procurement] process was designed to prevent bias." ODM required evaluators to disclose potential bias and participate in a quiet period. It even asked PHPG to evaluate the process and affirm protection against bias. Unfortunately, the PHPG letter seemed to express confusion about the process. For example, on page 6, it states that the consensus process was used to prevent the potential for an individual evaluator to act as an outlier. Yet that's not how the process worked: it did not focus on addressing outliers. Instead, the process went through every question to come to a "consensus" where 16 of 32 questions resulted in scores being lowered for Paramount below in a very statistically unlikely manner. Perhaps PHPG's examination of the documents did not reflect how the process worked in practice, no matter how it was supposed to work, making its evaluation of the process irrelevant for actual assessment of bias.

But perhaps Mercer took steps in the meeting to protect against cognitive biases? Well, when Katie Falls, one of the two Mercer staff who led the consensus meeting and the lead principal on the Mercer contract with ODM, was asked about her expertise, she described herself in her deposition testimony as an expert and someone experienced in "facilitation of the evaluation related to the evaluation components" on pg. 58 and on pg. 94-95, describes her experience in "working on evaluation committees and different processes in which we have to deal with identifying bias." Yet when asked whether she had any formal training in procurement evaluation, she said she did not on pg 60. When asked what is "confirmation bias" – perhaps the most famous cognitive bias – she said "I do not know what you mean by that" on page 92. When asked whether she has any training in the area of bias, she indicated that she did not on pg 94.

Given her self-acknowledged lack of understanding of training in bias or even a basic understanding of fundamental concepts such as "confirmation bias," how could Falls effectively protect the procurement process from either intentional or unintentional bias? Perhaps her collaborator on the consensus meeting, Mercer staff member Deidre Abbott, had more expertise in addressing bias? Unfortunately not. In her video testimony on October 5, 2021, at around minutes 40-44, Abbott indicates that she neither consulted experts in bias, nor has she taken bias training. When she was asked about what anchoring bias refers to, she indicated she does not, despite anchoring bias being one of the most dangerous threats to successful procurement processes. Similarly, asked about other cognitive biases, she indicated she could not define any of the ones she heard. Moreover, she indicated that Mercer has no policy of going back to the procurement process and checking it to see if any potential bias crept into the process.

Conclusion

In conclusion, we see that, regardless of ODM and Mercer's claims that the $20 billion Medicaid procurement process was protected against bias, the reality is quite different. Mercer staff who represented themselves as "experts" at dealing with and identifying bias have no training in addressing bias, failed to consult experts on bias, and don't even know basic facts about bias in procurement. The ODM and Mercer process failed to provide basic and simple protections against bias, resulting in a process rife with potential bias, both intentional and unintentional. Both ODM and Mercer staff had ulterior motives to subvert Paramount's application. But we do not even need deliberate bias: failing to protect against unintentional bias would do it.

Indeed, sanity checks of both statistical outcomes and performance measures demonstrate clearly biased outcomes. We have a situation where the structure of the process enabled bias, and where the outcome demonstrates clear bias. That strongly indicates a need for the courts to address the fundamental failures of public procurement in Ohio before it becomes a precedent that normalizes such flaws going forward.

Sincerely,
Gleb Tsipursky, PhD
CEO, Disaster Avoidance Experts

Gleb Tsipursky