ASA Comment on the Proposed Regulation for Federal Financial Assistance

Prepared under the guidance of the Scientific and Public Affairs Advisory Committee

July 13, 2026

The American Statistical Association (ASA) — the nation's leading professional association of statisticians and data scientists — appreciates the opportunity to comment on the proposed rule, Regulation for Federal Financial Assistance. Because of our strong concerns about the broad, long-lasting and harmful impacts that the implementation of this rule will have on our nation's research enterprise and, therefore, on our nation's competitiveness and well-being, we join the broader grantmaking stakeholder community in urging that the proposed rule be withdrawn. While we have concerns about numerous provisions in the proposed rule, we recognize the broader community has articulated those concerns well, so we have concentrated on the two provisions where statistical expertise is most distinctive. Each comment is tagged to its section per the docket's instructions. We do not address indirect costs (which OMB has stated it will not consider) or provisions being covered by other organizations. We also note that the ASA executive director, Ron Wasserstein, submitted comments separately on the publication provisions of the rule.

Merit review, "Gold Standard Science," and performance benchmarks (§§200.202, 200.205, 200.329)

The ASA shares the goal of rigorous, reproducible, and trustworthy federally funded science. Our concern is with implementation: how these provisions are operationalized will determine whether they strengthen the federal evidence base or inadvertently narrow it. For the reasons provided below, we strongly recommend that each of these provisions be withdrawn. If the administration chooses to pursue these provisions further, the following comments should be fully addressed, and the resulting provisions should be submitted for public comment. (We do not recommend resubmission of these provisions.) We offer four points.

1. **Consistent, authoritative definitions of reproducibility and replicability are not provided.** Because funding preference under §200.205 will turn on "rigorous, reproducible scholarship," these terms must be defined precisely and applied consistently. Across current federal documents, they are used inconsistently — in places reversing the definitions established by the National Academies (NASEM, 2019) and used by the ASA. If the rule were adopted despite our comments, we recommend using a single standard: reproducibility, defined as obtaining consistent results using the same data and methods, and replicability, defined as obtaining consistent results from new data. Reproducibility is largely achievable through data and code sharing and is a reasonable expectation; replicability is not fully within an investigator's control, and its absence is often a normal feature of scientific progress rather than evidence of waste. Clear, shared definitions will prevent inconsistent or arbitrary application across agencies and reviewers.
2. **Rigor should be judged by fitness to the research question, not by a single design.** Gold Standard status should depend on whether a study's design is appropriate to its question and executed rigorously. Otherwise, if it is operationalized narrowly — for example, as including only fixed-design, pre-specified randomized controlled trials — adaptive and platform trials, Bayesian designs, and the observational and real-world studies that are frequently the most appropriate and most sophisticated tools available could be disadvantaged. Many essential questions cannot be answered by a randomized controlled trial: post-market safety surveillance for rare adverse events, for instance, depends on large real-world datasets rather than randomization. If this rule were adopted despite our comments, we recommend that it explicitly state that methodological rigor is assessed relative to the research question.
3. **Benchmarks should not penalize long-horizon research (§§200.202, 200.329).** Performance "benchmarks," "unit-cost" metrics, and similar measures, if narrowly specified, raise well-understood measurement-validity problems: metrics reward what is easily and quickly measured, systematically favoring projects whose benefits can be specified in advance and penalizing basic and long-term work whose value accrues later. This is a textbook setting for Goodhart's law, in which a measure that becomes a target ceases to be a good measure. We recommend that benchmarks explicitly credit long-term scientific value, data resource creation, and research infrastructure. Modern artificial intelligence illustrates the stakes.

The U.S. National Science Foundation has invested in artificial intelligence research since the early 1960s, laying the technical and conceptual foundations that drive today's AI innovations. Pioneering work supported by NSF includes reinforcement learning, which improves the conversational ability of chatbots and trains self-driving cars; neural networks, which enable computer understanding of human speech and analysis of visual scenes; and large language models, which power generative AI systems like ChatGPT. None of this was foreseeable at the time, and none of it would have scored well against near-term benchmarks.

In late 2023, the Task Force on American Innovation held briefings on Capitol Hill, where a panel of computing and IT experts from across academia, government, and industry argued that recent advances in AI could not have happened without federal investment in areas such as hardware, software, and high-performance computing. Simply put, advances in AI are intrinsically tied to investments in basic research across the federally funded research enterprise.

Public funding has advanced core capabilities such as machine learning, neural networks, computer vision, and natural-language processing, which the private sector then developed into the AI systems we use today. The triumph of the market is the outcome of investments made in the commons: the private sector is now harvesting returns on public investments made decades ago, back when that research would have looked unpromising by any near-term measure. Criteria that reward only pre-specifiable, near-term outcomes would have cut off the funding that made this possible.

This is precisely the vision Vannevar Bush articulated in his 1945 report to President Truman, *Science: The Endless Frontier*, which argued that government must fund basic scientific research because private industry, focused on near-term products, would systematically underinvest in the open-ended discovery work that industry itself would later depend on. Bush's framework became the blueprint for the postwar federal research system, including the National Science Foundation itself (Zachary, 1999). The lesson of AI's own history is that this blueprint still holds: the discoveries that seem least justifiable by short-term metrics are often the ones that matter most decades later. Continuing to build American scientific strength means preserving the patient, long-horizon funding that Bush argued no market alone would ever supply.

4. **Scientific — including statistical — expertise should inform merit decisions (§200.205(b)–(d)).** Whatever the locus of final authority, determinations of scientific merit and methodological soundness require scientific expertise. Statisticians are essential to that judgment, and their absence is a recurring and avoidable source of error in study design and interpretation. We recommend that merit and methodological soundness continue to be evaluated by qualified scientific experts, with statistical expertise explicitly represented, regardless of who holds final funding authority.

This rushed rule process aside, and given the importance of greater engagement with the scientific community, the ASA would welcome the opportunity to help define fieldable rigor

standards (analysis plans, pre-registration where appropriate, code and data availability, and uncertainty quantification) and valid, gaming-resistant performance benchmarks.

Discretionary mid-award termination

Regarding §200.340's authorization to terminate awards that "no longer effectuate program goals or agency priorities," we believe, based on extensive experience and expertise and in alignment with the rule's own interest in preventing waste, that terminating active data collection mid-stream frequently destroys value out of proportion to the funds saved.

1. **Interrupting data collection destroys disproportionate value.** The cost of stopping a study mid-stream is largely invisible in budget terms but severe in scientific terms: it breaks time series, introduces attrition and survivorship bias, forfeits statistical power, and strands already-collected data that yields value only at completion. A five-year cohort terminated at year three is not sixty percent of a five-year cohort; it is a cross-sectional sample with an irregular follow-up interval, of far less analytic value than its cost would suggest.
2. **Incomplete studies are not merely smaller — they can mislead.** A trial halted mid-enrollment for administrative rather than pre-specified reasons can produce uninterpretable or actively misleading results that nonetheless influence clinical practice. Additionally, participants may have been exposed to risk for no offsetting knowledge gain. Midstream termination of human-subjects research can also conflict with research ethics and clinical best-practice obligations.
3. **Termination can create the very waste the rule seeks to prevent.** Because prior investment becomes largely unrecoverable, mid-award termination can be a waste-creation mechanism, contrary to the rule's anti-waste rationale. For regulated trials, it does not even end financial obligations — it shifts them: investigational new drug (IND) safety reporting, IRB-mandated wind-down, and participant notification continue regardless of federal funding, reallocating costs onto institutions and, in some cases, participants.
4. **Long-horizon studies are public infrastructure.** Much of the most consequential federally funded research functions as public infrastructure: the government funds long-term, high-value work that is not commercially profitable to sustain, and private innovation builds on the resulting evidence base. The Framingham Heart Study, funded to study cardiovascular disease, produced the understanding of risk factors that underlie modern prevention and, ultimately, therapies such as statins — benefits unforeseeable at launch and realized only through sustained, uninterrupted follow-up. A single well-designed longitudinal study is often far more economically efficient than a series of disconnected short ones. Criteria that reward only pre-specifiable benefits systematically undervalue this *option value* — the discoveries that become possible only as long-term data accumulate. Cooperative-agreement infrastructure, such as the Clinical and Translational Science Awards, which fund shared biostatistics and

research-design cores, compounds the risk: terminating or under-benchmarking these platforms cascades across the many individual studies they enable.

5. **Recommendation: a structured wind-down standard.** We recommend that any termination of an award involving active human-subjects contact, an active IND, or ongoing longitudinal data or biospecimen collection trigger a structured wind-down that includes a minimum analytical-completion window — not merely data preservation — so that partial data remains scientifically usable and prior public investment is not forfeited. This gives agencies a concrete, operationalizable alternative to abrupt termination that serves the rule's efficiency goals.

Questions or comments may be directed to ASA Director of Science Policy Steve Pierson: spierson@amstat.org.