https://ijhpm.com
Int J Health Policy Manag 2026;15:9507

OPEN ACCESS

doi 10.34172/ijhpm.9507

IJHPM
International Journal of Health Policy and Management

Viewpoint

# A National Case Study Demonstrates Why Outcomes-Based Payment Systems Fail

**Alexander J. Gajewski[1*], James L. Gajewski[2]**

***Correspondence to:** Alexander J. Gajewski, Email: alexander.gajewski@utsouthwestern.edu

**Copyright:** © 2026 The Author(s); Published by Kerman University of Medical Sciences. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Citation:** Gajewski AJ, Gajewski JL. A national case study demonstrates why outcomes-based payment systems fail. *Int J Health Policy Manag.* 2026;15:9507. doi:10.34172/ijhpm.9507

Received: 2 October 2025; Accepted: 21 July 2026; ePublished: 1 September 2026

## Background

Medicare Part B is the primary payor in the United States (US) for all outpatient care for Americans over the age of 65. As part of a broader movement towards "value-based care", the Medicare Merit-based Incentive Payment System (MIPS) was first implemented in 2019 and grades physicians in four categories: quality measures, cost of care, engagement in self-improvement, and ability to share records with other health systems. Participation in MIPS or one of a few alternative payment models is largely mandatory for those accepting patients with Medicare Part B.[1] Consequently, participation in MIPS exceeded half a million providers in 2022.[2] As opposed to other value-based arrangements and alternative payment models where providers generally take on some risk related to excess costs, MIPS payments are largely based on clinical outcomes – historically about 45% of the composite score is derived from clinical outcomes.[3] MIPS tries to reward higher quality care, as opposed to just efficient care. Individual physician practices are rewarded or penalized based on their MIPS score; however, any attempt to systematically measure and reward quality at the level of the physician's practice is a fool's errand.

We are two US internal medicine physicians, with broad experience ranging from the practice of primary to tertiary care and the design of clinical trials to national health legislation. In this paper, we argue that MIPS is a failed program. We explain several reasons why and provide insights for those interested in the design of outcomes-based payment systems.

## The Case Against Merit-based Incentive Payment System

MIPS is a well-intentioned program, but comes with significant administrative burden. The mean cost per physician to participate in MIPS is $13 000 per year.[4] The cost created an incentive for doctors to join a larger group and diffuse the cost of administrative staff that handle this task.[5] Meanwhile, payouts range from just $60 to $800 dollars annually.[6] Even the most successful practices do not offset the cost. There are no winners and losers, just losers and bigger losers.

Outpatient physician practices associated with health systems or multi-specialty practices, ie, larger practices, tend to score higher than independent, smaller practices.[7] Practices with larger volume are also advantaged in sum payout[6]; however, when practice size and quality of care are compared, bigger does not mean better.[8] When looking at individual outcome metrics, there was no compelling evidence that higher MIPS scores equated to better quality of care. For composite outcomes, the same data suggests that physicians caring for sicker and more socially vulnerable populations were disadvantaged.[3] These findings were replicated in at least one additional study.[9] Simply put, MIPS rewards size, not quality.

These findings are not surprising to those of us practicing internal medicine. Clinical quality is a challenge to measure accurately. Often the treatment for one condition worsens another. For example, rheumatologists provide steroids to treat inflammatory arthritis. Oncologists use steroids to manage nausea from chemotherapy. Steroids worsen blood sugar control, especially in patients with diabetes mellitus. If treatment quality is measured by control of blood sugar, then rheumatologists and oncologists must not be very talented. MIPS appears unable to adjust for natural confounders—chiefly medical and social acuity—when attempting to measure quality of care. Larger practices with greater administrative capacity may be able to devote more resources to coding comorbidities or, so to speak, playing the game.[3,7]

The architects of MIPS are at a crossroads. They could (1) continue trying to reduce the administrative burden, but antithetically adjust more robustly for the underlying social and medical determinants of health, or (2) they could accept that the entire venture has been a mistake. Given the expansion of MIPS "value pathways" reform, they seem to have chosen the former option.[10] This effort too will fail. Let us dissect why.

## Into the Weeds: Why Acuity Adjustment Is So Hard

Episodes of care have a wide variance in cost, indicating a large role of medical and social comorbidities.[11] Any comparison of practice outcomes must include robust adjustment for

Full list of authors' affiliations is available at the end of the article.

patients' baseline acuity, but this task is much easier said than done. Medicare has two medical acuity adjustment methodologies. Hierarchical Condition Categories (HCC) are used to risk adjust Medicare Advantage plan reimbursement. Major Complication and Comorbidity Categories (MCC) are used to adjust traditional Medicare payments for inpatient care. These tools predict mortality and expenditures at a population level,[12,13] but have not been validated at the level of the physician's practice.

HCC and MCC scores do not account for the social determinants of health. These factors are critical to patient care. To give an example, patients living in rural regions may not have access to home care delivery providers. Patients without family care givers may need longer hospitalizations to avoid readmission. Methods to account for social determinants in the cost of care are in development.[14] The most prominent method used in the United States may be the Area Deprivation Index, which is factored into Medicare Advantage payments. Like MCC and HCC, Area Deprivation Index is a population-level tool, not validated at the level of the physician's practice.[15] MIPS factors in HCC scores to adjust for medical complexity and Medicaid eligibility as a proxy for social risk. This approach is reasonable on paper, but as established, falls dramatically short.[3,9] Clinicians know it is a mistake to assume population-level statistics will predict an individual patients' course. Similarly, it is a mistake to assume population-level acuity adjustment models maintain validity at the level of an individual practice.

Current acuity adjustment models cannot fix the MIPS program, and future risk adjustment tools are not going to save it either. Acuity adjustment is reliant on information stored in the electronic health record. This data is always incomplete,[16] subject to input error and biases, and often inaccurate, even when assessing straightforward yes/no metrics.[17] MIPS is trying compare outcomes, and that goal requires a quality of data found only in prospective clinical trials. Trials are, of course, very expensive.[18]

More to the point, clinical trials have impartial judges. Measures of quality that are created in research settings but transferred into payment systems will be gamed, especially if providers are asked to measure their own outcomes.[19] This notion is known as Goodhart's law, or the idea that any observed statistical relationship will break down when used for policy purposes.[20] In other words, "When you use a measurement as a target, it stops being a useful measurement."

Here we have described three fundamental reasons MIPS failed. (1) Risk adjustment models are not validated at the practice level. (2) Comparison of outcomes between practices requires higher quality data than is available in electronic health records. (3) Outcomes measures will be gamed if tied to payment. All three reasons are not MIPS-specific; they generalize to any attempt to tie payment to population metrics. Perhaps all three challenges could be overcome, but no one has done it yet. Crucially, a fix for all three issues would undoubtably increase the cost of data collection and further burden providers. Additional payment system complexity would likely advantage well-resourced players over independent practices. The case of MIPS is a lesson in how not to use complexity.

## Conclusion

Proponents of the MIPS program laudably aim to give US citizens better value for every dollar spent on healthcare. The unfortunate reality is that the program imposes steep demands on practices and lacks the capacity to assess practices fairly. Every policy comes with tradeoffs. We believe the harms of the MIPS program outweigh any benefit and the effort being allocated to reform MIPS would be better spent on different projects altogether. The challenges facing the MIPS program generalize to all payment systems hoping to reward clinical outcomes at the individual-practice level. The US may be stuck with our program, but we hope others can avoid the same fate.

## Acknowledgements

The authors acknowledge the lasting influence of the late Drs. John Goodman and Samuel Silver. Their dedication to medicine and health policy continues to inspire our work. The authors also acknowledge Len Lichtenfeld for decades of mentorship and ongoing contribution to the field.

### Disclosure of artificial intelligence (AI) use

Not applicable.

### Ethical issues

Not applicable.

### Conflicts of interest

Authors declare that they have no conflicts of interest.

### Authors' contributions

Conceptualization: Alexander J. Gajewski and James L. Gajewski.
Data curation: Alexander J. Gajewski.
Formal analysis: Alexander J. Gajewski.
Methodology: Alexander J. Gajewski.
Supervision: James L. Gajewski.
Writing–original draft: Alexander J. Gajewski and James L. Gajewski.
Writing–review & editing: Alexander J. Gajewski and James L. Gajewski.

### Authors' affiliations

[1]Department of Internal Medicine, UT Southwestern Medical Center, Dallas, TX, USA. [2]Independent Researcher, Fort Lewis, WA, USA.

### References

1. The Medicare Access and CHIP Reauthorization Act of 2015.
2. Centers for Medicare and Medicaid Services. PY 2022 Clinician Public Reporting: Overall MIPS Performance. 2024; Authors' analysis counting the number of unique national provider identifiers in the dataset. https://data.cms.gov/provider-data/dataset/a174-a962.
3. Bond AM, Schpero WL, Casalino LP, Zhang M, Khullar D. Association between individual primary care physician Merit-based Incentive Payment System score and measures of process and patient outcomes. *JAMA*. 2022;328(21):2136-2146. doi:10.1001/jama.2022.20619
4. Khullar D, Bond AM, O'Donnell EM, Qian Y, Gans DN, Casalino LP. Time and financial costs for physician practices to participate in the Medicare Merit-based Incentive Payment System: a qualitative study. *JAMA Health Forum*. 2021;2(5):e210527. doi:10.1001/jamahealthforum.2021.0527
5. Berdahl CT EM, Ryan G, Needleman J, Nuckols TK. Primary care physicians in the Merit-based Incentive Payment System (MIPS): a qualitative investigation of participants' experiences, self-reported practice changes, and suggestions for program administrators. *J Gen Intern Med*. 2019;34(10):2275-2281. doi:10.1007/s11606-019-05207-z
6. Lin M-Y, Zhang Z, Carey K, Gidwani R, Hanchate AD. Merit-Based Incentive Payment System: longitudinal performance and uneven rewards for safety-net providers over 5 years. *Health Aff Sch*. 2025;3(6):qxaf105.

doi:10.1093/haschl/qxaf105
7. Johnston KJ, Wiemken TL, Hockenberry JM, Figueroa JF, Joynt Maddox KE. Association of clinician health system affiliation with outpatient performance ratings in the Medicare Merit-based Incentive Payment System. *JAMA.* 2020;324(10):984-992. doi:10.1001/jama.2020.13136
8. Gaynor M, Town RJ. *The Impact of Hospital Consolidation – Update*. Robert Wood Johnson Foundation;2012.
9. Johnston KJ, Hockenberry JM, Wadhera RK, Maddox KEJ. Clinicians with high socially at-risk caseloads received reduced Merit-based Incentive Payment System Scores. *Health Aff (Millwood).* 2020;39(9):1504-1512. doi:10.1377/hlthaff.2020.00350
10. Center for Medicare and Medicaid Services. MIPS Value Pathways. https://qpp.cms.gov/mips/mips-value-pathways. Accessed August 10, 2025.
11. Lee VS, Kawamoto K, Hess R, et al. Implementation of a value-driven outcomes program to identify high variability in clinical costs and outcomes and association with reduced cost and improved quality. *JAMA.* 2016;316(10):1061-1072. doi:10.1001/jama.2016.12226
12. Li P, Kim MM, Doshi JA. Comparison of the performance of the CMS Hierarchical Condition Category (CMS-HCC) risk adjuster with the Charlson and Elixhauser comorbidity measures in predicting mortality. *BMC Health Serv Res.* 2010;10:245. doi:10.1186/1472-6963-10-245
13. Kim J, Ock M, Oh IH, et al. Comparison of diagnosis-based risk adjustment methods for episode-based costs to apply in efficiency measurement. *BMC Health Serv Res.* 2023;23(1):1334. doi:10.1186/s12913-023-10282-4
14. Huffstetler AN, Phillips RL Jr. Payment structures that support social care integration with clinical care: social deprivation indices and novel payment models. *Am J Prev Med.* 2019;57(6 Suppl 1):S82-S88. doi:10.1016/j.amepre.2019.07.011
15. Limburg A, Rehkopf DH, Gladish N, Phillips RL, Udalova V. Validating 8 area-based measures of social risk for predicting health and mortality. *JAMA Health Forum.* 2025;6(8):e252669. doi:10.1001/jamahealthforum.2025.2669
16. Haneuse S, Arterburn D, Daniels MJ. Assessing missing data assumptions in EHR-based studies: a complex and underappreciated task. *JAMA Network Open.* 2021;4(2):e210184-e210184. doi:10.1001/jamanetworkopen.2021.0184
17. Morris NA, Couperus C, Jasani G, Day L, Stultz C, Tran QK. Discrepancies between retrospective review of "real-time" electronic health record documentation and prospective observer documentation of in-hospital cardiac arrest quality metrics in an academic cardiac intensive care unit. *J Clin Med.* 2023;12(22):7102. doi:10.3390/jcm12227102
18. Moore TJ, Heyward J, Anderson G, Alexander GC. Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015-2017: a cross-sectional study. *BMJ Open.* 2020;10(6):e038863. doi:10.1136/bmjopen-2020-038863
19. Sukul D, Hoffman GJ, Nuliyalu U, et al. Association Between Medicare Policy Reforms and Changes in Hospitalized Medicare Beneficiaries' Severity of Illness. *JAMA Network Open.* 2019;2(5):e193290. doi:10.1001/jamanetworkopen.2019.3290
20. Goodhart C. Problems of Monetary Management: The UK Experience. *Papers in Monetary Economics.* 1975;1:1-20.