Stanford University Human-Centered Artificial Intelligence
Stanford | Stanford Health Policy

Policy Brief
HAI Policy & Society
September 2025

Increasing Fairness in Medicare Payment Algorithms

Marissa B. Reitsma, Thomas G. McGuire, and Sherri Rose

RISK ADJUSTMENT IS A CORE COMPONENT OF CAPITATED HEALTH INSURANCE SYSTEMS. Medicare Advantage plans, the private plan alternative to traditional Medicare, are paid using capitation, which means that these plans are paid a prospectively determined amount dependent on a particular beneficiary's demographics and past diagnoses. In contrast, traditional Medicare uses a fee-for-service approach, which administers payments based directly on the services delivered. In capitated systems, risk adjustment adapts prospective payments to account for differences in expected costs of care and thereby mitigate so-called selection incentives that lead healthcare plan providers to attract profitable (less costly) enrollees and avoid unprofitable (more costly) enrollees.

Medicare Advantage enrolls more than half of all Medicare beneficiaries — including a disproportionate share of minority populations — and accounts for more than $450 billion in spending annually. Choices on the design of Medicare's risk adjustment algorithm can therefore have major impacts on healthcare spending, access, and outcomes. The current approach to risk adjustment for Medicare predicts beneficiary-level health spending as a function of their demographic characteristics and diagnosed health conditions. The prediction is based on an analysis of historical fee-for-service Medicare data, which can perpetuate disparities in access, utilization, and spending

Key Takeaways

Medicare Advantage health insurance plans account for more than $450 billion in annual spending and enroll a high share of beneficiaries from minoritized racial and ethnic groups. Improving the risk adjustment algorithm that determines payments for Medicare Advantage plans could therefore have substantial impacts on healthcare spending, access, and outcomes for minoritized populations.

We introduce two algorithms to improve fairness in Medicare Advantage plan payments: constrained regression and post-processing. We evaluate their impacts using enrollment and claims data from more than 4.3 million beneficiaries.

Both algorithms achieve fair spending targets, which reflect more equitable payment levels without sacrificing predictive performance. Constrained regression, which alters payments for health conditions, has the potential for more widespread health equity benefits in comparison to post-processing.

Policymakers should consider payment system reform that includes algorithmic changes to avoid reinforcing spending disparities and achieve more equitable spending. Such reforms must be accompanied by additional policy measures, as algorithmic changes alone cannot eliminate health disparities.

Choices on the design of Medicare's risk adjustment algorithm can have major impacts on healthcare spending, access, and outcomes.

when applied to Medicare Advantage beneficiaries. Exploring how these risk adjustment algorithms can achieve fairness goals for multiple minoritized racial and ethnic groups remains an understudied area.

In our paper "Algorithms to Improve Fairness in Medicare Risk Adjustment," we seek to close this important evidence gap by developing risk adjustment algorithms that can promote fairer spending for minoritized racial and ethnic groups. We analyzed a random sample of Medicare fee-for-service beneficiaries to assess existing levels of net compensation, which is the difference between predicted and observed spending, by racial and ethnic group. We proposed a basic measure of healthcare spending disparity that informed potential fair spending targets and then developed and evaluated two algorithms to achieve fair spending targets.

Policymakers can use the findings from our study to better understand how modifications to the risk adjustment algorithm can achieve greater fairness in health plan payments without sacrificing overall predictive performance. We made our open source code publicly available so that future research can build upon our work to drive further progress in improving healthcare payment systems.

Introduction

Medicare spending accounts for 14% of the federal budget and Medicare Advantage accounts for more than half of Medicare spending. Medicare Advantage plans receive risk-adjusted payments for each beneficiary they enroll. Currently, risk adjustment is based on a least squares regression, which generates spending predictions from observed data (i.e., beneficiaries' demographic characteristics and clinical conditions). Prior research has examined approaches to improve the risk adjustment algorithm by increasing the accuracy of spending predictions, mitigating opportunities for upcoding (i.e., where providers document more severe conditions to increase payments), and reducing the potential for favorable selection (i.e., where insurers aim to attract profitable beneficiaries and avoid enrolling unprofitable beneficiaries).

However, few prior studies have examined fair regression methods, which optimize for both overall and group-level performance. An important finding from previous studies is that adding marginalized group indicators as predictors in the risk adjustment algorithm can reinforce data-embedded inequities in spending between populations, necessitating alternative approaches. Past literature has not specifically examined algorithms to achieve fairness goals across multiple minoritized racial and ethnic groups in Medicare.

This research gap matters for several reasons. First, a greater percentage of Black, Hispanic, and Asian/Pacific Islander beneficiaries, compared to non-Hispanic white beneficiaries, are enrolled in Medicare Advantage plans. Second, the population aged 65 years and older, which is the largest Medicare-eligible population, is projected to become more racially and ethnically diverse in the coming decades. Third, although Medicare eligibility reduces racial and ethnic disparities in insurance coverage, disparities persist in healthcare access, utilization, spending, and outcomes. Historical fee-for-service spending data, which are used to estimate risk scores and determine payments, embed many of these long-standing disparities.

Approach

To address this research gap, we analyzed a 20% random sample of Medicare fee-for-service beneficiaries, reviewing enrollment and claims data between January 1, 2017, and December 31, 2020. We focused our study on beneficiaries who were aged 65 years and older, continuously enrolled in Medicare Parts A and B, not enrolled in Medicare Part C (i.e., Medicare Advantage), and not dual-eligible for Medicaid. Among beneficiaries meeting our eligibility criteria, we created three analytic cohorts that spanned two consecutive years each (2017–2018, 2018–2019, and 2019–2020). Consistent with the existing approach to Medicare risk adjustment, demographic and diagnosis data from the earlier year was used to predict spending in the subsequent year.

Next, we extracted demographic characteristics (i.e., age, documented sex, race/ethnicity, county, and original reason for Medicare eligibility) from beneficiary enrollment files and extracted diagnosis codes and total annual payments by Medicare from billing claims files. We mapped detailed diagnosis codes to aggregate condition categories (e.g., diabetes without complications) that are used for risk adjustment. Using these data points, we established a baseline regression that approximated the least squares regression used by Medicare to estimate risk scores.

We then developed two algorithms, constrained regression and post-processing, that achieve potential fair spending targets. For our analysis, we explored two sets of fair spending targets, which illustrate potential approaches to incentivize more equitable spending. The first, "disparities-based targets," uses a regression, fit using data from non-Hispanic white beneficiaries, to predict counterfactual spending levels for beneficiaries from minoritized racial and ethnic groups. This approach aims to capture additional spending by minoritized beneficiaries if they had the same healthcare access and utilization as non-Hispanic white beneficiaries. The second, "five percent targets," increases payments for minoritized beneficiaries by five percent of the mean spending across all eligible beneficiaries, akin to the five percent bonus paid to Medicare Advantage plans with ratings of at least four out of five stars.

Research Findings

Both algorithms that we developed achieved fair spending targets without sacrificing predictive performance compared to our baseline.

We analyzed data from 4.3 million beneficiaries meeting our eligibility criteria, of whom 86% were non-Hispanic white, 6% Black, 3% Hispanic, 2% Asian/Pacific Islander, 1% part of an Additional Group, and fewer than 1% American Indian/Alaska Native. Mean Medicare spending was $8,345 per year, median spending was $2,421 per year, and the mean age of beneficiaries was 75 years.

Consistent with prior literature, we found that predicted spending from our baseline regression was lower than observed spending for American Indian/Alaska Native and non-Hispanic white beneficiaries. On the other hand, predicted spending was higher than observed spending for beneficiaries from other racial and ethnic groups, including Black, Hispanic, and Asian/Pacific Islander groups. Observed spending was lowest among these minoritized groups, likely as a result of healthcare access barriers that lead to lower than appropriate healthcare utilization.

We evaluated the overall performance of our two proposed algorithms against the baseline regression and found that there were no meaningful differences in how well the algorithms explain variations in spending. While the baseline regression achieves a payment system fit of 12.7%, our post-processing and constrained regression algorithms also achieved fits of 12.7% and 12.6%, respectively. The meaningful difference between our algorithms and the baseline regression is that our algorithms increased payments for minoritized beneficiaries. Achieving disparities-based fair spending targets added $300 to $518 in additional payments per beneficiary from minoritized racial and ethnic groups, while achieving five percent targets added $417 per beneficiary across all minoritized groups. From a health equity perspective, increasing payments for marginalized groups can help reduce inequalities in healthcare access and outcomes.

We find that constrained regression, which increases payments for health conditions that are overrepresented among beneficiaries from minoritized groups, can achieve more widespread equity impacts.

Collectively, our results suggest that the two algorithms developed in our study could be applied to the Medicare Advantage plan payment system to achieve fairer spending outcomes with minimal impact to overall risk adjustment performance.

The key difference between our constrained regression algorithm and our post-processing algorithm is that constrained regression changes payments at the level of health conditions, whereas post-processing changes payments at the level of individual beneficiaries based on their race and ethnicity. From this difference in implementation, we find that constrained regression, which increases payments for health conditions that are overrepresented among beneficiaries from minoritized groups, can achieve more widespread equity impacts. Since social and structural forces cluster certain health conditions among marginalized groups, including but not limited to groups defined by race and ethnicity, changing payments at the health condition level can yield broader impacts.

Policy Discussion

Changes to Medicare's risk adjustment algorithm have the potential to boost fairness for minoritized populations without loss of overall predictive performance. Medicare Advantage's reach, with respect to both increasing general enrollment rates and achieving persistent high rates of enrollment of minoritized beneficiaries, creates an opportunity for widespread impact through equity-enhancing payment system reform.

Policymakers should take away that greater fairness in plan payments can yield more equitable, increased spending for minoritized beneficiaries. Increased payments are critical ingredients for downstream reductions in inequalities in healthcare access and outcomes. The two algorithms developed in our study achieve fair spending targets without sacrificing predictive performance. Comparing the two proposed algorithms, constrained regression achieves more widespread changes by updating payments for health conditions, regardless of the race or ethnicity of the beneficiary. Nonetheless, both algorithms are feasible extensions of the current regression approach used for Medicare risk adjustment, positioning them as improvements that are transparent, interpretable, and implementable.

Both algorithms are feasible extensions of the current regression approach used for Medicare risk adjustment, positioning them as improvements that are transparent, interpretable, and implementable.

Critically, algorithmic changes alone cannot eliminate health disparities. To achieve intended goals, payment reform needs to include additional policy measures to mitigate the potential for strategic responses by insurers that could ultimately reduce the impacts of payment changes on beneficiaries' healthcare access and outcomes. Additionally, policymakers should engage in broad consultation and additional research to establish fair spending targets, examine potential consequences for groups not included in fairness objectives, and understand the impacts of differential rates of diagnosis prior to implementation.

Healthcare disparities are the result of many long-standing and interconnected social and structural forces. Algorithmic changes to payment policy are one promising tool for ensuring more equitable spending, and policymakers and researchers alike should work toward making those changes happen.

Reference: The original article is accessible at Marissa B. Reitsma, Thomas G. McGuire, and Sherri Rose, "**Algorithms to Improve Fairness in Medicare Risk Adjustment,**" JAMA Health Forum, 6 (August 2025): 8, https://jamanetwork.com/journals/jama-health-forum/fullarticle/2837998.

Stanford University's Institute for Human-Centered Artificial Intelligence (HAI) applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices. The views expressed in this policy brief reflect the views of the authors. For further information, please contact **HAI-Policy@stanford.edu**.

Marissa B. Reitsma is an assistant professor of health policy at the Stanford University School of Medicine and a faculty affiliate at the Stanford Institute for Human-Centered Artificial Intelligence (HAI).

Thomas G. McGuire is a professor emeritus of health economics at Harvard Medical School.

Sherri Rose is a professor of health policy and director of the Health Policy Data Science Lab at Stanford University and a faculty affiliate at Stanford HAI.

Stanford University
Human-Centered
Artificial Intelligence

Stanford HAI: 353 Jane Stanford Way, Stanford CA 94305-5008

T 650.725.4537 **F** 650.123.4567 **E** HAI-Policy@stanford.edu **hai.stanford.edu**