Stanford University
Human-Centered
Artificial Intelligence

POLICY BRIEF
NOVEMBER 2022

SPECIAL HEALTHCARE SERIES

Balancing Fairness and Efficiency in Health Plan Payments

Anna Zink, Thomas G. McGuire, and Sherri Rose

IT IS NO SECRET THAT THE HEALTHCARE SYSTEM IS RIFE WITH INEQUITIES—from geography to race to class. Recent advances in data science and algorithmic fairness modeling are empowering researchers to quantify those inequities at unprecedented scale. These AI-driven opportunities may have a significant impact on the pricing of healthcare.

Currently, the U.S. healthcare industry represents nearly one-fifth of overall U.S. economic output. Even the slightest tweaks to this system—and the incentives and costs baked into healthcare pricing and policy—can have potentially trillion-dollar reverberations. This links to AI because health insurance companies use a complex set of calculations to estimate how much money each enrollee will cost in healthcare risk adjustment formulas.

In our paper in the journal *Biometrics*, "Fair Regression for Health Care Spending," we expand on existing work in computer science, statistics, and health economics to propose a new model of fair regression in calculating healthcare costs for risk

KEY TAKEAWAYS

- Health insurance marketplaces use risk adjustment formulas to estimate how much money each enrollee will cost their plan—and to prevent insurers who take on sicker enrollees from suffering catastrophic losses.
- Current risk adjustment methods underpredict spending for specific groups like individuals with mental health and substance use disorders. Health policy researchers are developing new statistical methods to significantly improve this situation.
- Policymakers should familiarize themselves with the latest fairness research and metrics, engage with patient advocacy groups and insurers, support deeper research and exploration of fairness in the health insurance industry, and think beyond technical responses to consider legal and other remedies.

Our work provides a new opportunity for policymakers to realign the healthcare market's incentives in favor of patients.

adjustment formulas. We examine the lack of incentives for insurers to cover undercompensated groups of individuals, such as those diagnosed with a mental health or substance use disorder, and how our method can realign coverage incentives. Our new statistical methods have the potential to improve the healthcare risk framework and the provision of healthcare in the United States.

Two additional publications, "Improving the Performance of Risk Adjustment Systems" in the *American Journal of Health Economics* and "Identifying Undercompensated Groups Defined by Multiple Attributes in Risk Adjustment" in *BMJ Health & Care Informatics*, build on this work. In the former, we examine how reinsurance (which protects insurance firms from high-dollar claims), machine learning methods, and other design elements impact the risk adjustment systems used to pay health plans. In the latter, we present a new way of identifying marginalized groups across multiple characteristics, focused especially on people with chronic conditions.

For policymakers, our work provides a new opportunity to realign the healthcare market's incentives in favor of patients. Policymakers could explore policy interventions that shape companies' incentives around the pricing models they deploy—all of which come with their own trade-offs. Although we refrain from advocating for one specific replacement over any other at this time, policymakers should consider the following set of recommendations:

- Familiarize themselves with the details of these fairness metrics;
- Welcome dialogue and debate with patients' groups, researchers, and insurers as to which of these methods would be best to implement, centering principles of justice;
- Prioritize deeper research and promote exploration of algorithmic fairness in the healthcare industry; and
- Think beyond purely technical responses, such as considering legal improvements to better protect individuals marginalized by the healthcare system.

Introduction

Data-driven decision-making has justifiably come under scrutiny in recent years for its tendency to perpetuate human biases, and few use-cases make the dangers as clear as healthcare. Health insurance companies presently use risk calculations to predict spending at the individual level. Most federally regulated health insurance markets, including Medicare Advantage and the individual health insurance Marketplaces created by the Affordable Care Act, use these risk adjustment methodologies and other statistical methods to estimate "correct" payments to health insurers based on costs that enrollees are expected to incur. Organizations do this by using select demographic information and diagnosis codes from medical claims.

Most commonly, measures of fairness are based on the idea of group fairness—similarity in predicted outcomes or errors for groups. For instance, if health insurance companies on average provide care to people of one race that leads to substantially different health outcomes than people of another race, that violates the concept of group fairness; the same goes for health pricing decisions that systematically make more errors concerning one group than they do with another. The ways to address fairness can be categorized based on where in the algorithm pipeline the fairness intervention occurs: the data preprocessing phase (aiming to "de-bias" data before it is used in an algorithm), the fitting phase (changing how algorithms make decisions to lead to fairer outcomes), and the postprocessing phase (refining algorithm outputs to improve fairness).

Our three papers focus on the challenge of designing insurance risk adjustment systems with both fairness and performance in mind. These goals are not mutually exclusive. For example, risk adjustment systems that have greater error rates on people with chronic health conditions are also unfair.

One common approach to algorithmic fairness in the fitting phase is to add new variables on underrepresented groups into risk adjustment formulas. The idea is to fix undercompensation for those groups by accounting for other factors. However, in practice, such a practice can backfire or create additional problems if the variables are unavailable; if they over- or under-incentivize utilization of healthcare services; or if the risk formulas do not change in light of the variables. Other fairness approaches in the fitting phase include, for example, using separate formulas for different groups. Some organizations already elect for this last option, as adults and children have different health spending patterns.

In our first paper, we expand on work in health economics, statistics, and computer science to develop new algorithmic penalized and constrained approaches to fairness in health insurance spending. In our study, we use the IBM MarketScan Research Databases to understand which estimators are best at reducing undercompensation for enrollees. This data is of great policy significance because the federal government used these databases to develop the individual healthcare Marketplaces' risk adjustment formulas.

In our second paper, we examine how constrained regressions, reinsurance, and machine learning methods can be used to improve health plan payments. This work builds on our prior studies and the risk adjustment literature, but our paper is the first to study how these tools can work together. Our focus is on the risk adjustment formula used in the Marketplaces, and we present modifications that improve individual and group fit while also reducing the number of variables used, which prevents gaming by insurers.

In our third paper, we focus on both the Marketplaces and Medicare risk adjustment formulas. We introduce a method for identifying undercompensated groups defined by multiple attributes in health plan payments. In particular, we focus on chronic conditions, age, and documented sex attributes.

Critically, recent work in the healthcare policy space argues that a purely statistical analysis of these issues sidelines policy considerations. In that vein, our work presents estimators and comparisons that provide options for balancing trade-offs—all while considering how legal and other remedies could pair with technical responses to insurance cost and payment problems.

Research Outcomes

We found that risk adjustment formulas underestimate the average amount of spending for enrollees with mental health and substance use disorders by around $2,000. In short, insurance providers spend less than they should on enrollees with mental health and substance use disorders.

The best improvements in fairness for the enrollees came from two methods: average constrained regression and our new covariance constrained regression. Both these approaches increased fairness for enrollees with mental health and substance use disorders—reducing the average undercompensation by 98 percent compared to other methods. These methods only lowered overall model performance by 4 percent. This small decrease is likely tolerable to policymakers. All changes to the risk adjustment formulas we explored had different fairness and overall performance trade-offs.

Our second study found that a combination of removing drug and certain health condition variables, using 1 percent of funds for reinsurance, and introducing fairness constraints on the loss function for four undercompensated chronic illnesses greatly improved overall fit and fairness. This approach also improved group fit for most undercompensated groups not included in the loss function.

Finally, our third study aimed to identify previously unknown complex undercompensated groups defined by multiple attributes. After considering age, documented sex, and 12 chronic health indicators, we found many groups with multiple chronic conditions that were undercompensated by at least $10,000 and up to $29,600. For instance, enrollees with asthma, heart disease, and mental health and substance use disorders were undercompensated by about $12,000.

We found many groups with multiple chronic conditions that were undercompensated by at least $10,000 and up to $29,600.

Policy Discussion

The patients at the center of our studies are marginalized in the healthcare system and many problems in the system persist. Health plan enrollees with multiple chronic conditions already face challenges in maintaining access to care. Patients with undercompensated health conditions also represent a considerable portion of the U.S. population—approximately 20 percent of the country's population have a mental health or substance use disorder.

Policymakers should use their convening power to bring together industry and patient advocacy groups to parse out the consequences of new fairness metrics.

Policymakers should remain steadfast in their support for health policy research and make necessary digital investments to explore scalable, fairness-driven risk solutions.

These new methods' fairness gains and high overall performance make a strong case for reorienting the risk adjustment process. For many of our new approaches, improvements in fairness were larger than the subsequent decreases in overall model performance—arguably a net gain. Selecting which method might be "best" will rely on a subjective matrix of perspectives from policymakers, researchers, patient advocacy groups, and insurers about how best to balance group fairness against overall method performance. Central to this objective is the belief that developing an ethical framework requires a diverse, inclusive, patient-centered policy conversation.

Looking ahead to policy implementation, adopting any of these new risk adjustment formulas will require analyzing data from millions of enrollees. Doing so may require investment in the research infrastructure and computational resources available to health policy researchers. Policymakers should remain steadfast in their support for health policy research and make necessary digital investments to explore scalable, fairness-driven risk solutions. Finally, policymakers should remember to look beyond purely technical responses. For instance, discrimination stemming from insurance benefit formulas might necessitate legal remedies.

It's not always easy to quantify the trade-offs embedded in risk algorithms, like model performance and measures of fairness. But the large volumes of data generated by the U.S. health system can enable some aspects of these analyses—in this case, understanding healthcare spending of marginalized groups of people. Policymakers must do what they can to promote these technological advances and ensure that the resulting benefits are distributed as widely and equitably as possible.

Policymakers should remember to look beyond purely technical responses.

The first article, "**Fair Regression for Healthcare Spending**," can be accessed at: https://onlinelibrary.wiley.com/doi/10.1111/biom.13206

The second article, "**Improving the Performance of Risk Adjustment Systems**," can be accessed at: https://www.journals.uchicago.edu/doi/abs/10.1086/716199

The third article, "**Identifying Undercompensated Groups Defined by Multiple Attributes in Risk Adjustment**," can be accessed at: https://informatics.bmj.com/content/28/1/e100414

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**.

Anna Zink is a Principal Researcher at the Center for Applied Artificial Intelligence at the University of Chicago Booth School of Business.

Thomas G. McGuire is a Professor of Health Economics Emeritus in the Department of Health Care Policy at Harvard Medical School.

Sherri Rose is a Professor at Stanford University in the Center for Health Policy and Department of Health Policy. She is also Co-Director of the Health Policy Data Science Lab.

Stanford University
Human-Centered
Artificial Intelligence