

Identifying undercompensated groups defined by multiple attributes in risk adjustment

Anna Zink,¹ Sherri Rose²

To cite: Zink A, Rose S. Identifying undercompensated groups defined by multiple attributes in risk adjustment. *BMJ Health Care Inform* 2021;**28**:e100414. doi:10.1136/bmjhci-2021-100414

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjhci-2021-100414>).

Received 19 May 2021
Accepted 25 August 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹PhD Candidate in Health Policy, Harvard University, Cambridge, Massachusetts, USA

²Center for Health Policy and Center for Primary Care & Outcomes Research, Stanford University, Stanford, California, USA

Correspondence to

Anna Zink; azink@g.harvard.edu

ABSTRACT

Objective To identify undercompensated groups in plan payment risk adjustment that are defined by multiple attributes with a systematic new approach, improving on the arbitrary and inconsistent nature of existing evaluations.

Methods Extending the concept of variable importance for single attributes, we construct a measure of ‘group importance’ in the random forests algorithm to identify groups with multiple attributes that are undercompensated by current risk adjustment formulas. Using 2016–2018 IBM MarketScan and 2015–2018 Medicare claims and enrolment data, we evaluate two risk adjustment scenarios: the risk adjustment formula used in the individual health insurance Marketplaces and the risk adjustment formula used in Medicare.

Results A number of previously unidentified groups with multiple chronic conditions are undercompensated in the Marketplaces risk adjustment formula, while groups without chronic conditions tend to be overcompensated in the Marketplaces. The magnitude of undercompensation when defining groups with multiple attributes is many times larger than with single attributes. No complex groups were found to be consistently undercompensated or overcompensated in the Medicare risk adjustment formula.

Conclusions Our method is effective at identifying complex undercompensated groups in health plan payment risk adjustment where undercompensation creates incentives for insurers to discriminate against these groups. This work provides policy-makers with new information on potential targets of discrimination in the healthcare system and a path towards more equitable health coverage.

INTRODUCTION

Risk adjustment is an essential tool in regulated health insurance markets. It redistributes health plan payments to insurers with higher cost patients, aiming to decrease the relationship between health and profits to ensure that sicker individuals are not discriminated against by insurers.¹ Risk adjustment for health plan payments has been successful in reducing selection incentives for many individuals, but incentives to discriminate still exist for groups of individuals whose costs are underpredicted by the risk adjustment

Summary

What is already known?

- Risk adjustment for health plan payment is known to undercompensate some groups of enrollees.
- Undercompensation can lead to discrimination against groups through advertising and insurance benefit design choices.
- No standardised method for identifying undercompensated groups exists today.

What does this paper add?

- We present a novel method for identifying marginalised groups defined by multiple attributes, with a focus on chronic conditions in plan payment risk adjustment.
- Applying our new tool in the health insurance Marketplaces and Medicare, we identify previously unknown undercompensated groups.
- This work provides policy-makers a tool for uncovering incentives for selection in insurance markets and a path towards more equitable health coverage.

formula, including those with mental health and substance use disorders.^{2 3}

The ability of insurers to discriminate against some groups of patients has been curbed since the passage of the Affordable Care Act (ACA) in 2010, which prevents insurers from refusing enrolment or changing premium prices based on enrollee health. However, insurers are able to attract more profitable enrollees or discriminate against less profitable enrollees through advertising as well as benefit design choices, such as which providers comprise their provider networks and which drugs are placed in higher cost tiers or subject to drug formulary management tools.^{4–7} Recently, machine learning methods have been used to identify individual conditions that are underpaid by the risk adjustment formula as well as drugs predictive of unprofitability, but these studies did not focus on groups defined by more than one attribute.^{5 8} In Germany and the Netherlands, researchers have incorporated

interactions of variables into the risk adjustment formula based on the partitioning of a single regression tree predicting the residual, but only considered groups defined within the current risk adjustment formula.^{9,10} To date, no systematic method exists for identifying groups at risk for discrimination in the healthcare payment system, or more broadly in the algorithmic fairness literature, especially groups defined with a more complex set of attributes. Even groups defined by a single sensitive attribute are typically prespecified.¹¹

In this paper, we present the first data-driven method for identifying undercompensated groups in health plan payment risk adjustment that are defined by multiple attributes. Extending the concept of variable importance for single attributes or predefined groups, we construct a new measure of ‘group importance’ in the random forests algorithm that relies on multiple attributes.^{12–15} Groups are defined based on demographic and clinical categories susceptible to benefit redesign, namely chronic conditions, age and documented sex. We designed our method for the purposes of studying the risk adjustment formulas used in the individual health insurance Marketplaces created by the ACA as well as Medicare. Our newly proposed method for identifying undercompensated groups overcomes the arbitrary nature of existing evaluations of risk adjustment, providing policy-makers with a tool for uncovering incentives for selection that persist in insurance markets and a path towards more equitable health coverage.

METHODS

Defining groups

Our primary objective is to identify complex groups vulnerable to discrimination in the health insurance market through benefit design. Therefore, in defining groups, we consider what factors are necessary for this type of discrimination. To start, the groups must be actionable for insurers, meaning the insurers must know which groups are profitable or unprofitable and be able to act on this information via the design of provider networks or drug formularies.¹⁶ The losses or gains caused by the group must also be large enough to substantially impact insurers’ revenue. There are a number of factors that could lead to this type of large impact for insurers, including the size of the group, the size of the undercompensation of the group and the persistence of the undercompensation over time.¹ For these reasons, we focus on identifying groups defined by prevalent chronic conditions that require specialist care or drug treatments. In addition to chronic conditions, we consider whether age and documented sex affect the size of undercompensation.

Notation

The canonical risk adjustment formula estimates individual annual spending Y using a vector of input variables X that contains j demographic variables, $D = \{D_1, D_2, \dots, D_j\}$ and k health condition categories,

$H = \{H_1, H_2, \dots, H_k\}$ or $X = \{D, H\}$. The risk adjustment formula is some function f mapping X to Y , $f(X) = Y$. The residual of the risk adjustment formula, $R = \hat{Y} - Y$, measures the underprediction or overprediction for each individual, where Y is predicted spending. We are interested in the set of groups G predictive of the residual R . To define the groups, we select s indicator variables $I = \{I_1, I_2, \dots, I_s\}$ that will comprise the components of the groups in G : a set of chronic conditions, documented sex, and age categories. A group in G is defined as any combination of components in I . This allows us to consider various levels of complex groups. For example, older women with cancer and mental health disorders would be one group and women with mental health disorders another. We examine only binary variables; relaxing this restriction would increase the number of groups substantially.

Estimation

Insurers can adjust their premiums at defined market levels to account for differences in costs in different geographical regions.¹⁷ Because we are interested in undercompensated groups within markets, we want to adjust for market-level differences in spending. Therefore, we start by estimating a slightly modified risk adjustment formula using ordinary least squares regression:

$$Y = \beta X + \text{Market} + \epsilon$$

where, beyond X , we additionally control for geographic market, Market, to account for routine premium adjustments made at the market level.

To find the groups in G most predictive of the residual R , we predict R with the group component variables I using the random forests algorithm. The random forests algorithm grows many decision trees to predict R . Each tree uses a bootstrapped sample from the original data to decrease the variance. This is often referred to as bagging, but unlike other forms of bagging, the random forests algorithm also selects a subset of variables to consider at each split point in the tree so the same highly correlated variables do not appear in every single tree.¹⁸

Tree-based methods are an intuitive way of forming groups because each decision tree splits the sample into a set of mutually exclusive groups. These groups are defined based on the intersections of the input variables used for predicting the outcome. At each split in the tree, a variable is selected that best partitions the variable space in order to reduce variation within each new node. By using the component variables I to predict R , we naturally create groups in G that are interactions of the component variables in I used to split the tree. For each tree, we find the set of groups that are formed in the terminal nodes. We aggregate this information across all trees, recording the number of times a group appears and its mean estimated residual R across trees and years. See figure 1 for an illustration of this process.

The hyperparameters of the random forests algorithm are important for defining group criteria (in addition

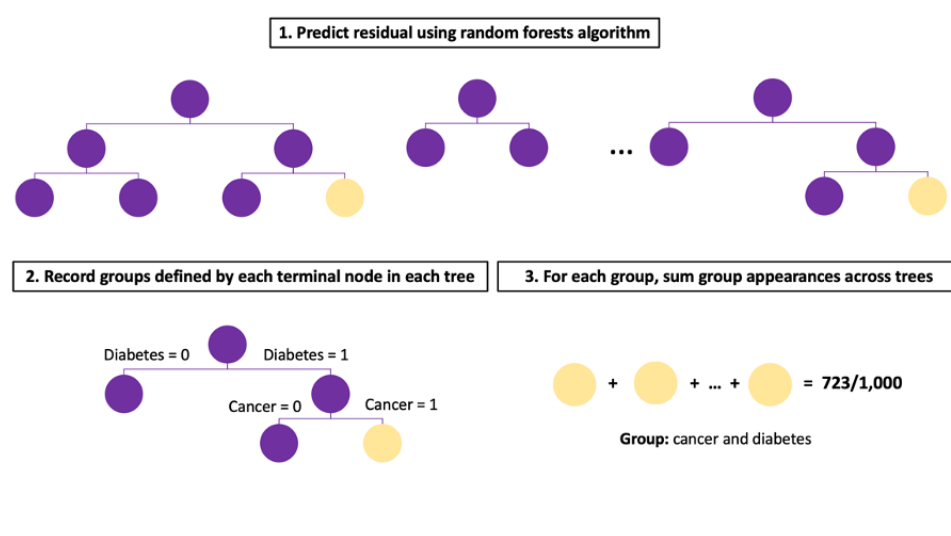


Figure 1 Illustration of random forests for group importance.

to algorithmic performance). Through hyperparameter specification, we set the minimise size of groups by requiring a minimum size for terminal nodes in the decision tree. We also control the complexity of the groups by limiting the depth of the tree and the number of terminal nodes. The deeper the tree grows, the more attributes the groups will contain. Recall that typically only a subset of randomly selected variables is considered at each split. By selecting a smaller number of variables at each split, we can decorrelate not only variables across trees but also groups, ensuring that the same groups are not favoured by every tree. And finally, the more trees we grow, the more stable our results will be, but this can be computationally costly and after a certain number of trees only trivially improves stability.¹⁹

Marketplaces risk adjustment

In 2014, the Department of Health and Human Services (HHS) developed a risk adjustment formula to redistribute insurer payments in the individual and small group Marketplaces established under the ACA.²⁰ Plans with lower-risk enrollees subsidise plans with higher-risk enrollees through a budget neutral risk adjustment transfer programme regulated by the Centers for Medicare & Medicaid Services (CMS).²¹ Premiums are set within market rating areas defined by counties or Metropolitan Statistical Areas (MSAs) depending on the state.¹⁷ The Marketplaces risk adjustment formula is built using the IBM MarketScan Research Databases, which contain claims on over 23 million individuals receiving employer-sponsored private health insurance in the USA.²²

We estimated individual annual spending using the 2019 Marketplaces risk adjustment formula inputs plus geographical fixed effects at the MSA level with the following specification:

$$Y = \text{Age} \times \text{Sex} + \text{HCC} + \text{MSA} + \epsilon,$$

where Age \times Sex are age and documented sex categories and HCC are Hierarchical Condition Categories

(HCC) mapped from diagnosis codes. We used MSAs as a proxy for premium rating areas in all states because we did not have county information.

Data

To evaluate the Marketplaces risk adjustment formula, we sampled adults (aged 21–64) insured for a full year in either 2016, 2017 or 2018 from the MarketScan data. We calculated individual annual spending as the total spending across the individual's outpatient, inpatient, and carrier claims in the given year. The traditional risk adjustment formula predicts insurer spending, but for simplicity we used total spending, which included patient cost sharing. Age, documented sex and MSA information was obtained from the enrolment file. HCC variables were constructed by mapping International Classification of Diseases 10th Revision (ICD-10) diagnosis codes recorded in the inpatient, outpatient, and carrier claims using the 2019 HHS-HCC software. We considered age categories (21–29, 30–39, 40–49, 50–59, 60–64), documented sex and 12 chronic condition indicators (arthritis, asthma and other respiratory conditions, cancer, diabetes, heart disease, hypertension, kidney disease, hyperlipidaemia, mental health and substance use disorders, nervous system conditions, osteoporosis and viral infections) as component variables to define groups in *G*. These 12 conditions were selected because they are common chronic conditions monitored by CMS.²³ The chronic conditions indicators were created using individual and multilevel diagnosis-based categories from the Agency for Healthcare Research and Quality Clinical Classification Software (see online supplemental table S1 for mapping). These categories incorporate more ICD-10 diagnosis codes than HCCs, and therefore, allow for identification of health states not captured by HCCs.

Implementation

We sampled 1 million individuals for each sample year and ran separate ordinary least squares regressions to

Table 1 Marketplaces sample characteristics (in per cents)

Variables	2016	2017	2018
Documented sex, female	52.3	52.2	51.6
Age			
21–29	17.9	17.9	18.0
30–39	20.4	20.6	21.0
40–49	23.7	23.7	23.7
50–59	27.2	26.9	26.6
60–65	10.8	10.8	10.7
Arthritis	4.5	4.5	4.5
Asthma	10.6	10.7	10.8
Cancer	7.1	7.0	6.8
Diabetes	8.9	9.0	8.9
Heart disease	9.1	9.1	9.3
Hypertension	14.1	13.9	13.7
Kidney disease	0.6	0.6	0.6
Lipid disorders	10.2	9.7	9.5
Mental health	11.1	11.7	12.6
Nervous	0.7	0.7	0.7
Osteoporosis	0.6	0.6	0.6
Viral infections	0.4	0.4	0.4
N	1 000 000	1 000 000	1 000 000

predict annual spending. We then deployed the random forests algorithm to predict the residual for each sample year. We identified groups under different size and complexity hyperparameters holding the number of trees (1000) and the number of component variables selected at each branch (10) constant. This resulted in four hyperparameter settings: (1) minimum size=100 and maximum nodes=8, (2) minimum size=100 and maximum nodes=64, (3) minimum size=10 000 and maximum nodes=8 and (4) minimum size=10 000 and maximum nodes=64. The random forests algorithm was run in R using the random-Forest package.²⁴ To ensure that identified groups persisted in all three sample years with some frequency across trees, we limited the results to groups that appeared in at least 1% of the trees in each sample year. Predicted versus observed residuals for each identified group were compared to assess the accuracy of our predictions.

RESULTS

Table 1 summarises the Marketplaces sample characteristics for each sample year. About 52% of the sample was documented as female. The sample was fairly evenly distributed across the five age categories with more enrollees in their 50s than any other age group (26.6%–27.2%). Of the 12 chronic conditions, hypertension was the most common, documented in about 14% of the sample. Osteoporosis, chronic kidney disease and chronic viral infections were the least common, appearing in <1% of the

Table 2 Marketplaces average residual (US\$) by chronic condition

Chronic condition	2016		2017		2018	
	No	Yes	No	Yes	No	Yes
Arthritis	263	-5611	293	-6213	283	-5957
Asthma	325	-2751	335	-2783	458	-3785
Cancer	163	-2120	165	-2200	189	-2583
Diabetes	23	-236	19	-188	47	-481
Heart disease	393	-3924	383	-3813	502	-4903
Hypertension	263	-1596	255	-1583	315	-1985
Kidney disease	18	-2838	20	-3126	28	-4596
Lipid disorders	121	-1065	113	-1044	98	-930
Mental health	312	-2499	330	-2492	398	-2770
Nervous	33	-4775	30	-4382	36	-5229
Osteoporosis	12	-1963	15	-2593	10	-1760
Viral infections	2	-419	6	-1508	5	-1276

sample. For each sample year, we observed an average undercompensation of around US\$1500–US\$5000 for most of the selected chronic conditions. Diabetes was the least undercompensated group with an average undercompensation of a few hundred dollars (see **table 2**). Average annual spending ranged from US\$6500 to US\$7000 in the study period.

In general, all four hyperparameter settings yielded similar results. Individuals with multiple chronic conditions, in particular enrollees with some combination of asthma, heart disease, arthritis and mental health and substance use disorders, tended to be undercompensated whereas individuals with no chronic conditions were overcompensated. In **figure 2**, we present the top undercompensated and overcompensated groups (measured by the average residual across the three sample years) limiting the group size to 10 000 and maximum nodes to 8. Predicted undercompensation was substantially larger than overcompensation. The groups we identified with multiple chronic conditions were undercompensated, on average, by at least US\$10 000 and up to US\$29 600, whereas overcompensated groups were all overcompensated by less than US\$1000. Also, the predicted residuals for groups defined by multiple attributes were larger than the average residuals for each condition individually. For example, individuals documented as having a combination of asthma, heart disease and mental health and substance use disorders were estimated to be undercompensated by up to US\$12 000 on average, which is more than the sum of their observed average residuals.

When we allowed for smaller (minimum group size=100) and more complex groups (maximum nodes=64) we saw groups with more conditions appear in the top 10 (**figure 3**). Namely, we identified more groups with asthma, cancer, hypertension, chronic kidney disease and mental health and substance use disorders. Interestingly, we also found that most top overcompensated groups were older, but relatively healthy individuals. In general, tuning the maximum node

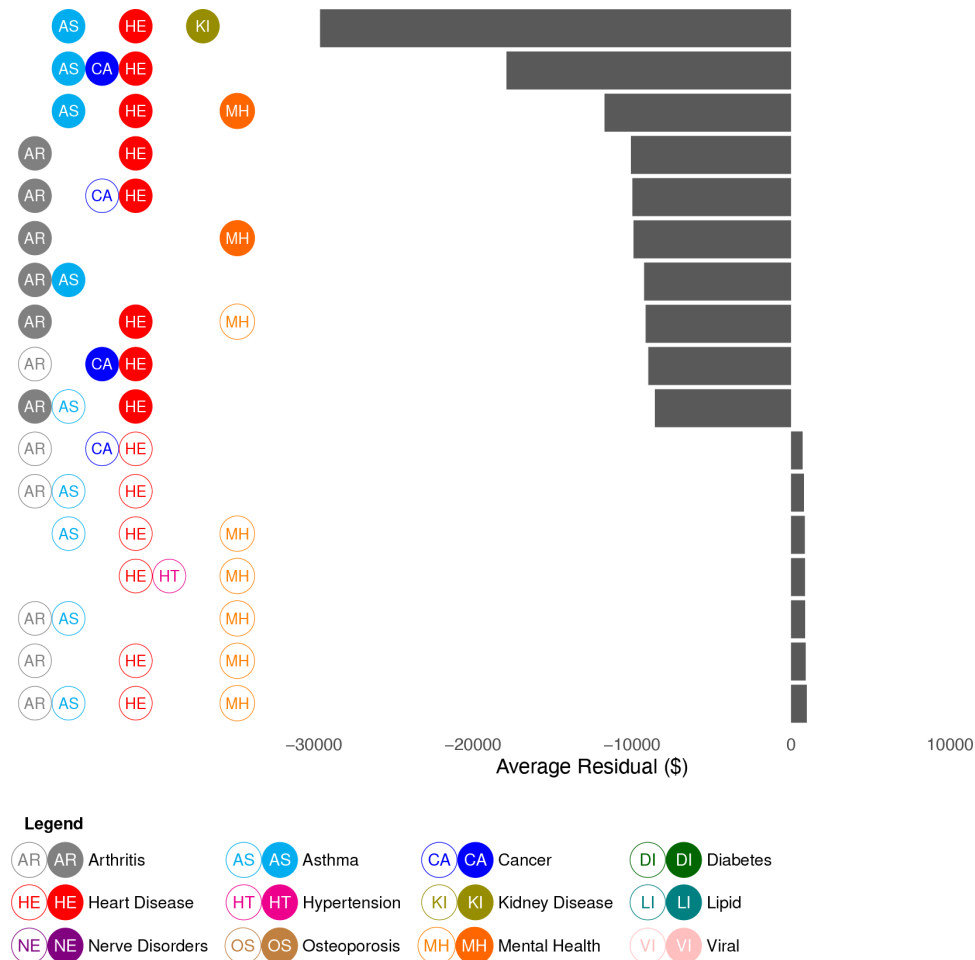


Figure 2 Top undercompensated and overcompensated groups in the Marketplaces risk adjustment (minimum node size: 10 000, maximum nodes: 8). Unfilled circles indicate the lack of a condition. Only seven overcompensated groups were identified in this setting, after removing groups that appeared in less than 1% of trees across the 3 years.

size, which limits the depth of the trees, was more important for determining group complexity than limiting the size of the groups. Some chronic conditions, such as chronic viral infections and disorders of lipid metabolism as well as documented sex, rarely if ever appeared as components in our identified groups.

When we disaggregated the results by year, we observed the same pattern of undercompensation for enrollees with multiple chronic conditions and overcompensation for those without, but there was some temporal variation in the types of conditions that were identified (online supplemental figure S1). For example, in 2016, 5 out of the 10 top undercompensated groups were identified as having a hereditary and generative nervous system in addition to other chronic conditions, but this attribute did not appear in later years.

In online supplemental text, we present results from the other setting we considered: identifying undercompensated groups in the Medicare risk adjustment formula used by CMS to set capitation rates for Medicare Advantage plans. We found no groups that were consistently undercompensated or overcompensated across the three sample years.

DISCUSSION

In this research, we presented a new approach for identifying undercompensated groups defined by multiple attributes, extending the random forests algorithm to determine group importance without requiring groups to be prespecified. We implemented our group importance algorithm on the Marketplaces and Medicare risk adjustment formulas. The Marketplaces formula favoured healthier individuals compared with those with multiple chronic conditions. In particular, groups that had some combination of asthma, heart disease, arthritis and mental health and substance use disorder were largely undercompensated. These results expand on and re-enforce earlier work exploring issues of undercompensation in risk adjustment for people with multiple chronic conditions.²⁵ Undercompensation in the Medicare risk adjustment formula was smaller in magnitude compared with the Marketplaces formula and no groups were persistently undercompensated or overcompensated across the three sample years. This suggests that the Medicare risk adjustment formula may induce fewer incentives to discriminate with respect to the twelve selected conditions.

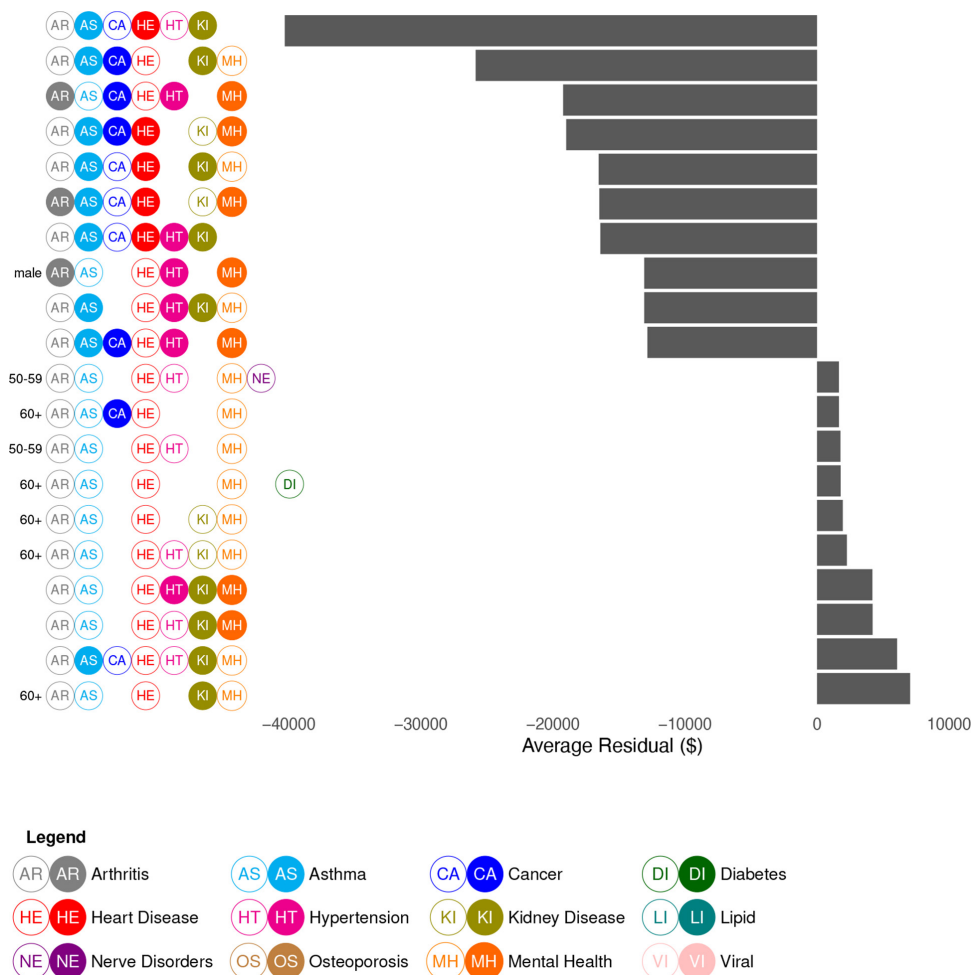


Figure 3 Top undercompensated and overcompensated groups in the Marketplaces risk adjustment (minimum node size: 100, maximum nodes: 64). Unfilled circles indicate the lack of a condition.

Our results raise the question of how to address incentives in insurance markets, especially for people with multiple health conditions. There are numerous approaches for improving risk adjustment performance aimed at removing selection incentives in regulated insurance markets. The Marketplaces aim to partly address undercompensation for enrollees with two chronic health conditions, where one condition is severe, by including interactions in the risk adjustment formula.²⁰ Given our results, additional interactions for those with multiple chronic conditions could be beneficial. However, in Germany and the Netherlands, researchers found mixed results when they included the interactions found in a single regression tree predicting the residual, sometimes leading to negative consequences for other groups.^{9,10} While not yet used in practice, machine learning methods have also been proposed as an alternative to predict spending and may be better able to capture non-linear spending trends for enrollees with multiple chronic conditions.²⁶ Another plausible option could be to enforce improved fit for selected groups through constrained or penalised regression, incorporating group fairness directly into the loss function where negative consequences for other groups were not observed.^{27,28} Solutions to discrimination via insurance benefit design may lie outside of the risk adjustment formula itself and be better addressed

by, for example, additional legal remedies, such as requiring adequate coverage of treatments and services by insurers for the set of conditions we identified.

This work has several limitations. We developed a methodology that focused on groups with the largest magnitude average residual that appeared in all three sample years. The groups were composed of conditions that we considered actionable for insurers, but we did not consider a priori whether certain combinations of conditions (or lack of conditions) would be more actionable for insurers compared with others. In our algorithm, we set the size and number of terminal nodes in order to restrict the size and complexity of a group, but future work could develop more sophisticated algorithms that aim to empirically examine whether specific combinations of chronic conditions are actionable for insurers within the decision rules of the tree rather than post hoc.

The techniques we presented are relevant to other insurance markets relying on risk adjustment. They can be used by developers and evaluators of risk adjustment formulas to identify and assess performance across many different groups defined by multiple attributes. In this paper, we focused on 12 chronic conditions, but our method allows for the inclusion of additional chronic conditions and other variables

appropriate for identifying groups, subject to computational constraints and ethical factors. Variable choices will be limited by data availability: while countries like the USA and the Netherlands have rich data sources to construct and assess their risk adjustment formulas, not all countries have access to such detailed data.¹

Our group importance method may be relevant in a wide range of applications beyond risk adjustment, although we caution researchers to carefully consider the context before attempting to identify potentially marginalised groups. In some settings, identifying groups could actively cause harm, for example, may involve collecting or amplifying stigmatising information. In other cases, our tool may be useful in mitigating ongoing harms. Machine learning predictions for clinical outcomes have been found to be less accurate for groups defined by age, race or other attributes, contributing to health disparities.²⁹ Our new method could help ensure that algorithms deployed in such settings remedy inequities for currently unidentified marginalised groups. We recommend researchers create a social impact statement and follow an ethical pipeline for building algorithms when considering adapting our tool to any setting.^{29,30}

Acknowledgements The authors thanks Thomas G. McGuire, Richard van Kleef, and participants at the 2020 Risk Adjustment Network Meeting for helpful comments.

Contributors Both authors contributed to the conceptualisation of the work. SR supervised the work. AZ performed the analysis and drafted the manuscript. Both authors reviewed and edited the manuscript.

Funding This work was supported by National Institute of Health Director's New Innovator Award grant number DP2-MD012722 and the National Science Foundation Graduate Research Fellowship Programme.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available. Under the requisite data use agreements, data for the analyses are not publicly available. We created a simulated data set and provide our analysis code on GitHub (<https://github.com/zinka88/Identifying-Groups>).

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

- McGuire TG, van Kleef R, eds. *Risk adjustment, risk sharing and premium regulation in health insurance markets*. Elsevier, 2018.
- McGuire TG, Newhouse JP, Normand S-L, et al. Assessing incentives for service-level selection in private health insurance exchanges. *J Health Econ* 2014;35:47–63.
- Montz E, Layton T, Busch AB, et al. Risk-adjustment simulation: plans may have incentives to distort mental health and substance use coverage. *Health Aff* 2016;35:1022–8.
- Shepard M. *Hospital network competition and adverse selection: evidence from the Massachusetts health insurance exchange*. National Bureau of Economic Research, 2016.
- Rose S, Bergquist SL, Layton TJ. Computational health economics for identification of unprofitable health care enrollees. *Biostatistics* 2017;18:682–94.
- Carey C, Change T. Technological change and risk adjustment: benefit design incentives in medicare part D. *Am Econ J Econ Policy* 2017;9:38–73.
- Aizawa N, Kim YS. Advertising and risk selection in health insurance markets. *Am Econ Rev* 2018;108:828–67.
- Rose S. Robust machine learning variable importance analyses of medical conditions for health care spending. *Health Serv Res* 2018;53:3836–54.
- Buchner F, Wasem J, Schillo S. Regression trees identify relevant interactions: can this improve the predictive performance of risk adjustment? *Health Econ* 2017;26:74–85.
- van Veen SHCM, van Kleef RC, van de Ven WPMM, et al. Exploring the predictive power of interaction terms in a sophisticated risk equalization model using regression trees. *Health Econ* 2018;27:e1–12.
- Chouldechova A, G'Sell M. Fairer and more accurate, but for whom? *arXiv:170700046 [cs, stat]*. 2017. Available: <http://arxiv.org/abs/1707.00046> [Accessed 10 Jul 2019].
- Grömping U. Variable importance assessment in regression: linear regression versus random forest. *Am Stat* 2009;63:308–19.
- Loupe G, Wehenkel L, Suter A. Understanding variable importances in forests of randomized trees. In: Burges CJC, Bottou L, Welling M, eds. *Advances in neural information processing systems 26*. Curran Associates, Inc, 2013: 431–9.
- Wehenkel M, Suter A, Bastin C, et al. Random forests based group importance scores and their statistical interpretation: application for Alzheimer's disease. *Front Neurosci* 2018;12:411.
- Fisher A, Rudin C, Dominici F. All models are wrong, but many are useful: learning a variable's importance by studying an entire class of prediction models simultaneously. *J Mach Learn Res* 2019;20:1–81 <https://arxiv.org/abs/1801.01489v5>
- Ellis RP, McGuire TG. Predictability and predictiveness in health care spending. *J Health Econ* 2007;26:25–48.
- CMS. Market rating reforms. Available: <https://www.cms.gov/ccio/programs-and-initiatives/health-insurance-market-reforms/state-gra> [Accessed 14 Dec 2020].
- Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- Probst P, Boulesteix A-L. To tune or not to tune the number of trees in random forest? *J Mach Learn Res* 2017;18 <https://arxiv.org/abs/1705.05654>
- Kautter J, Pope GC, Ingber M. The HHS-HCC risk adjustment model for individual and small group markets under the Affordable care act. *Medicare Medicaid Res Rev* 2014;4. doi:10.5600/mmrr.004.03.a03. [Epub ahead of print: 09 05 2014] https://www.cms.gov/mmrr/articles/a2014/mmrr2014_004_03_a03.html
- Semanskee A, Claxton G. *Explaining health care reform: risk adjustment, reinsurance, and risk corridors*. 17. KFF, 2016. <https://www.kff.org/health-reform/issue-brief/explaining-health-care-reform-risk-adjustment-reinsurance-and-risk-corridors/>
- IBM. Market scan research databases - databases, 2020. Available: <https://www.ibm.com/products/marketscan-research-databases/databases> [Accessed 16 Sep 2020].
- CMS. Chronic conditions. Available: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CC_Main [Accessed 25 Jun 2021].
- Liaw A, Weiner M. Classification and regression by random forest. *R News* 2002;2:18–22.
- Pope GC, Kautter J, Ingber MJ, Freeman S, Sekar R, Newhart C. Evaluation of the CMS-HCC risk adjustment model, final report; 2011.
- Rose S. A machine learning framework for plan payment risk adjustment. *Health Serv Res* 2016;51:2358–74.
- Zink A, Rose S. Fair regression for health care spending. *Biometrics* 2020;76:973–82.
- McGuire T, Rose S, Zink A. Improving the performance of risk adjustment systems constrained regressions, reinsurance, and variable selection. *Am J Health Econ* 2021.
- Chen IY, Pierson E, Rose S, et al. Ethical machine learning in healthcare. *Annu Rev Biomed Data Sci* 2021;4:123–44.
- Diakopoulos N, Friedler S, Arenas M, et al. Principles for accountable algorithms and a social impact statement for algorithms :: FAT ML. Available: <https://www.fatml.org/resources/principles-for-accountable-algorithms> [Accessed 30 Dec 2020].