Predicting Future Utilization Using Self-Reported Health and Health Conditions in a Longitudinal Cohort Study: Implications for Health Insurance Decision Support

Abigail R. Barker PhD[1,2] , **Karen E. Joynt Maddox MD MPH[2,3]** , **Ellen Peters PhD MS[4]** , **Kristine Huang BA[3]**, and **Mary C. Politi PhD[5]**

Abstract

Decision support techniques and online algorithms aim to help individuals predict costs and facilitate their choice of health insurance coverage. Self-reported health status (SHS), whereby patients rate their own health, could improve cost-prediction estimates without requiring individuals to share personal health information or know about undiagnosed conditions. We compared the predictive accuracy of several models: (1) SHS only, (2) a "basic" model adding health-related variables, and (3) a "full" model adding measures of healthcare access. The Medical Expenditure Panel Survey was used to predict 2015 health expenditures from 2014 data. Relative performance was assessed by comparing adjusted-R^2 values and by reporting the predictive accuracy of the models for a new cohort (2015–2016 data). In the SHS-only model, those with better SHS were less likely to incur expenditures. However, after accounting for health variables, those with better SHS were more likely to incur expenses. In the full model, SHS was no longer predictive of incurring expenses. Variables indicating better access to care were associated with higher likelihood of spending and higher spending. The full model (R^2 = 0.290) performed slightly better than the basic model (R^2 = 0.240), but neither performed well at the upper tail of the cost distribution. While our SHS-based models perform well in the aggregate, predicting population-level risk well, they are not sufficiently accurate to guide individuals' insurance shopping decisions in all cases. Policies that rely heavily on health insurance consumers making individually optimal choices cannot assume that decision tools can accurately anticipate high costs.

Keywords

health status, self-report, healthcare utilization, health expenditures, insurance, health, decision support techniques, cohort studies, policy

[1]Brown School, Washington University in St. Louis, St. Louis, MO, USA
[2]Center for Health Economics and Policy, Institute for Public Health, Washington University in St. Louis, St. Louis, MO, USA
[3]Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA
[4]Center for Science Communication Research, School of Journalism and Communication, University of Oregon, Eugene, OR, USA
[5]Division of Public Health Sciences, Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

Corresponding Author:
Abigail R. Barker, Washington University in St. Louis, 1 Brookings Drive – MSC 1196-0251-46, St. Louis, MO, USA.
Email: arbarker@wustl.edu

> **What do we already know about this topic?**
>   Decision support tools can improve health insurance consumers' knowledge and confidence; many online tools offer a degree of customization based upon demographic and health information.
> **How does your research contribute to the field?**
>   This research asks whether such tools can be customized sufficiently to make accurate predictions about utilization and cost that could then guide specific plan selection in a given health insurance market.
> **What are your research's implications toward theory, practice, or policy?**
>   Policies that rely heavily on health insurance consumers making individually optimal choices cannot assume that decision tools can accurately anticipate high costs.

Introduction

Over the past 40 years, health insurance has become an increasingly sophisticated product that plays the dual roles of financial protection and means of access to healthcare services. At the same time, publicly subsidized insurance has begun mimicking the private market as it relies on the notion of informed and savvy consumer shopping behavior to drive efficiency. Many non-elderly American adults are asked to select an insurance plan from an employer-sponsored or individual marketplace, and adults over the age of 65 can choose from an array of Medicare supplemental and Medicare Advantage plans to optimize coverage and access to care. Those anticipating significant care needs may prefer plans with more comprehensive coverage at a higher up-front cost, whereas those anticipating needing little care may prefer less insurance coverage, or a higher-deductible plan.[1] Once this high-level choice is made, more granular choices involving smaller variations in plan design must be weighed against small differences in premiums.

Prior research has found that individuals commonly make mistakes when faced with such decisions. In a study of Medicare beneficiaries selecting Part D plans, people systematically placed too much weight on premiums rather than out-of-pocket costs, resulting in suboptimal choices.[2] In a study of private employer-sponsored insurance selection, people commonly chose plans that left them with higher overall spending than other available options, and lower-income consumers fared worse; the authors found these effects to be driven by a lack of understanding of how health insurance works.[3]

Moreover, future healthcare utilization and its associated costs are often unpredictable at the individual level, especially when one is faced with a new or worsening health condition, as a recent analysis of spending in the last year of life confirmed.[4] Yet anticipating future costs is essential to making the optimal choice for health coverage—especially for the decision about whether to enroll in a high-deductible plan.[5] Resources such as decision aids, which in general aim to inform and help individuals faced with complex or unfamiliar decisions, may include online algorithms that help individuals predict future utilization and associated costs.[6] However, to generate cost estimates, these tools often require individuals to provide demographics, socioeconomic information, and medical comorbidities or prescription drug lists.[7] For some people, these questions feel intrusive.[8] For others, who may have new or undiagnosed conditions—possibly due to prior lack of access to insurance—the tools' results may be inaccurate and missing key information needed to anticipate future expenses.

The policy approach that relies on individuals optimizing their health insurance decisions may be strengthened if decision aids can improve accuracy while remaining simple enough for widespread use. One potential way to do this is via the concept of self-reported health status (SHS).[9] SHS, which asks patients to rate their own health on a scale, often from poor to excellent, is a straightforward method for assessing individuals' physical symptoms, emotional well-being, and functional status.[10] SHS plays a significant role in predicting total healthcare spending.[11–13] Some state-based tools even use it to suggest insurance plans.[7] SHS does not require individuals to share personal health information or know about potentially undiagnosed conditions; thus it could provide helpful input to tools estimating future healthcare spending.[10,14] However, it is a subjective measure that may be context-dependent, and there is little research regarding its ability to predict individual utilization and costs.

A closely related question when predicting utilization and costs is the degree to which access to care impacts the estimates and the resulting decision aid output; recent studies suggest that differential access for racial minorities may lead to racial bias in risk prediction algorithms.[15] Using a dataset that specifically includes information on whether an individual has access to healthcare (in a variety of forms) can test whether and how access modifies cost estimates. This information may be helpful to understand the risk of using cost prediction tools in certain settings (such as Health Insurance Marketplaces) in which racial minorities disproportionately gain access to new coverage.

In this study, we therefore aimed to: (1) determine the degree of predictive accuracy that can be achieved by SHS

alone; (2) compare these findings to SHS plus a full complement of typical health-related variables to predict future utilization and costs; and (3) examine how direct "healthcare access" variables, and additional indirect access variables such as socioeconomic characteristics, relate to predictions of utilization and spending.

Methods

Data

Data are from the 2014–2016 Medical Expenditure Panel Survey (MEPS), a set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of healthcare and health insurance coverage.[16] Two recent panels from the MEPS were analyzed: the model was estimated on MEPS Panel 19 (2014–2015) and assessed for accuracy using Panel 20 (2015–2016). Each unique panel contains information on about 16,000 respondents over a two-year period; through a restricted access agreement, we were able to supplement MEPS with other county and state data.

Predictor Variables

Our primary predictor of interest was SHS, for which individuals described their overall health in qualitative terms (poor, fair, good, very good, or excellent). We transformed these responses such that 0 represented poor health, 1 indicated fair health, 2 indicated good health, 3 indicated very good health, and 4 indicated excellent health. An analogous variable included in the MEPS survey, self-reported mental health status, was coded identically. We considered various model specifications for this variable, including categorical, linear, and log-linear. As model performance was similar, including the magnitude of predictive errors, we present the linear form as the easiest to interpret.

Respondents' health-related demographics included age and sex, as well as a variable for older age (greater than 64). Indicator variables for 11 chronic and acute conditions asked by MEPS (arthritis, asthma, back/joint pain, cancer, chronic bronchitis, diabetes, emphysema, high cholesterol, heart condition, hypertension, and stroke) were used separately and also combined into a count variable for total number of chronic conditions. A variable capturing the degree to which the respondent's pain level over the past 4 weeks caused limitations in their ability to work was also included, with responses ranging from 1 (indicating "not at all") to 5 (indicating "extremely"). Current smoking status was also used as a health-related indicator.

To measure access to care, we assessed respondents' geographic, cost-related, and personal access. First, to analyze geographic access, we used Urban Influence Codes[17] to categorize counties as metropolitan, micropolitan (ie, rural

counties containing small cities), and non-core (ie, rural counties that are non-micropolitan). We merged MEPS data with county-level indicators of healthcare supply from the Area Health Resources File (AHRF), 2015–16.[18] Counts of primary care providers and specialists served as proxies for availability of care. Second, to analyze cost-related access, we included Medicaid expansion status by year (with states which expanded during the study period coded as expansion states only in the appropriate years), as well as income level and source of health insurance coverage. We did not control for any finer details of insurance coverage because our goal was to determine the expected utilization for a given individual independent of any specific plan features. Finally, to analyze respondents' personal access, we included a MEPS variable asking whether the respondent had a "usual source" of healthcare. Other variables in this category were self-reported mental health status, marital status, and racial and Hispanic identities. Those with mental health conditions are less likely to seek care.[19,20] Marital status is often associated with increased utilization and better health, especially for men,[21,22] whereas racial disparities may be contributing differentially to access due to discrimination.[23]

Primary Outcome: Health Expenditures

Annual expenditures were calculated for each respondent. Actual utilization and cost data are reported by individual households and their members, and this information is supplemented by data from their medical providers (doctors, hospitals, pharmacies, etc.,) to characterize the individual's total spending from all sources (private and public insurance, employer, patient, etc.). Expenditures were log-transformed for all positive values due to skewness in the distribution of the data.

Analysis

We first described patient characteristics by SHS group. We then created sequential models for health expenditures. Year 2 expenditures were predicted based upon year 1 characteristics. In the derivation sample, the model used Panel 19's year 1 data (2014) to predict year 2 expenditures (2015). The estimates that fit the relationship best were then applied to Panel 20 data (the validation sample), to predict 2016 expenditures based upon 2015 data. To assess model fit, the predictions were compared to actual 2016 expenditures in the second year of Panel 20. Because a large number of zero expenditures existed, we used a two-stage Heckman model,[24] in which stage 1 estimated the probability that expenses would be incurred at all, and stage 2 estimated the relationship between the explanatory variables and year 2 expenditures, conditional on year 2 expenditures being positive. To predict from the model, the stage 1 and 2 estimates from Panel 19 data were combined to produce an unconditional estimate of Panel 20 year 2 expenditures, given Panel 20 year

1 data. The SAS procedure QLIM, with the HECKIT option, was used to estimate a binary probit selection model with a second-stage OLS regression.

To gauge the value of adding additional variables to the predictive model, we compared (a) a bare-bones model in which SHS alone predicted expenditures; (b) a "basic" model in which SHS and the health-related variables described above were used; and (c) a "full" model which used all variables that were statistically significant, which specifically included several measures of healthcare access as described above. Because prior work found an interactive effect between age and SHS,[9] we tested various specifications. We assessed the relative performance of the three models both by comparing adjusted-R^2 values and root mean squared errors (MSEs) and by reporting on the predictive accuracy of the model using a new cohort of people. For the latter, we were particularly interested in how well the full model could predict extremely high costs compared to the performance of the basic model, given that the decision to purchase health insurance is often about mitigating the risk of unexpectedly high costs.

All analyses were conducted using SAS Enterprise Guide Version 7.15 (Cary, NC). We considered a two-tailed P-value of less than 0.05 to be statistically significant. This study was determined not to meet the definition of human subjects research by the Washington University Office of Human Research Protection Office due to the de-identified nature of the data.

Results

Sample Characteristics

Our sample consisted of 11,204 respondents, weighted using MEPS survey weights to represent the U.S. population (Table 1). Of those reporting excellent health, 52.7% were male, while 47.4% of those reporting poor health were male. The mean age for those reporting excellent health was 42.1 years, compared to a mean age of 56.0 years for those reporting poor health. The proportion of adults over age 65 was more than twice as high (28.1%) in the poor health category than in the excellent health category (13.2%). Those in excellent health reported very few (0.9 on average) chronic conditions compared to those in poor health, who reported an average of 4.5 conditions. People in excellent health typically reported no limitations due to pain over the past 4 weeks, while the average pain limitation score for those in poor health was 3.3 (where 3 is "moderate" and 4 is "quite a bit"). Smoking was 4 times as prevalent (32.4% vs. 8.1%) among those in poor health compared to those in excellent health.

In terms of socioeconomic and access indicators, people in excellent health were more educated, earned higher incomes, and were more likely to have private insurance, live in metropolitan areas, and live in a Medicaid expansion state. More than half (54.9%) of those reporting poor health had a

high school diploma or less compared to only 31.8% of those reporting excellent health. Among those in poor health, only 7.2% were uninsured, while 53.0% were covered by a public source, compared to 11.7% of those in excellent health being uninsured and only 11.6% with public coverage. A greater share of respondents with poor, rather than excellent, health were from non-metropolitan areas, and especially from rural non-core counties. Similarly, a greater share of respondents who reported poor health (87.3%) vs. those who reported excellent health (69.0%) had a usual source of care. Finally, 60.7% of those in excellent health lived in a Medicaid expansion state, whereas only 56.8% of those in poor health lived in an expansion state.

Relationship Between Health and Expenditures

In the aggregate, SHS was associated with unadjusted expenditures (Table 1). In the SHS-only model, we found that better SHS was associated with a lower probability that expenditures would be incurred as well as a lower value of logged expenditures when incurred. The adjusted R^2 and root MSE for this model were 0.070 and 1.731, respectively, (Table 2). The basic model, which added age, sex, smoking status, and a count of the number of chronic conditions, yielded an adjusted R^2 of 0.240 and root MSE of 1.544, suggesting improved prediction. Under this model, however, those with better SHS were more likely to incur expenses (with each level of health adding 0.052 to the likelihood in the basic model), although the magnitude of those expenses was predicted to be lower (−0.105).

The full model, which added race and socioeconomic variables, as well as access variables, yielded an adjusted R^2 of 0.290 and root MSE of 1.507. For the most part, variables indicative of better access to care were associated with a higher likelihood of spending as well as higher spending. For example, having a usual source of care, having higher income, and having higher educational attainment were associated with a higher probability of expenses being incurred as well as higher-magnitude logged expenses. Being a racial minority was associated with a lower likelihood of expenses and lower spending. In this full model, the coefficients for SHS changed markedly. In fact, SHS was no longer predictive of the probability of expenses being incurred (ie, stage 1). Further, the SHS coefficient on the magnitude of expenses was negative, and almost twice the magnitude, in the full model (−0.179) compared to the basic model (−0.105).

Out-Of-Sample Model Performance

After applying the basic and full models to Panel 20 MEPS data, the full model performed slightly better at the upper tail of the cost distribution. Figure 1 displays the median and extreme (99%, 95%, and 5%) values of the differences between predicted expenditures and the actual expenditures observed in the data by SHS. For example, the median

Table 1. Selected Descriptive Characteristics of Weighted MEPS Data, 2014–2015.

Self-Reported Health Status	Excellent (N=2904)	Very Good (N=3227)	Good (N=3059)	Fair (N=1560)	Poor (N=454)
Health-Related Variables					
Sex					
Male	52.7%	46.7%	47.6%	47.2%	47.4%
Female	47.3%	53.3%	52.4%	52.8%	52.6%
Age					
Average age	42.1	46.4	49.6	52.8	56.0
Percent over 65	13.2%	17.3%	23.2%	26.6%	28.1%
Chronic conditions					
Average number	0.9	1.5	2.2	3.2	4.5
Min, Max	0, 7	0, 9	0, 8	0, 10	0, 10
Percent with none	56.1%	36.1%	22.7%	10.9%	4.9%
Pain limitation (last 4 wks)					
Average score (1=none, 5=extreme)	1.0	1.1	1.4	2.1	3.3
Smoking (currently)					
Percent saying yes	8.1%	15.1%	18.2%	23.4%	32.4%
Socioeconomic and access variables					
Income distribution					
Under 100% FPL	7.9%	9.8%	14.4%	24.7%	29.2%
100–250% FPL	24.4%	22.5%	28.9%	35.4%	39.0%
250–400% FPL	20.9%	20.4%	21.2%	17.9%	17.2%
Above 400% FPL	46.9%	47.3%	35.5%	21.9%	14.6%
Educational attainment					
High school or less	31.8%	33.9%	45.7%	57.1%	54.9%
Some college or associate's degree	29.7%	33.5%	30.7%	28.6%	33.6%
Bachelor's degree	23.4%	20.3%	15.0%	8.8%	7.8%
Post-graduate degree	15.1%	12.3%	8.6%	5.5%	3.7%
Race					
Percent non-white	35.9%	31.9%	38.6%	46.0%	31.3%
Percent non-Hispanic	85.7%	87.7%	82.6%	77.3%	84.7%
Insurance status					
Private	76.7%	74.4%	64.7%	49.6%	39.8%
Public	11.6%	15.7%	23.9%	36.9%	53.0%
Uninsured	11.7%	9.9%	11.4%	13.5%	7.2%
County of residence					
Metropolitan	88.8%	85.6%	82.7%	85.4%	82.0%
Micropolitan	7.4%	9.5%	11.9%	7.8%	9.0%
Rural, non-micro	3.9%	4.9%	5.4%	6.8%	9.0%
Has a usual source of care					
Percent saying yes	69.0%	75.6%	79.6%	82.3%	87.3%
Medicaid expansion					
Resides in an expansion state in 2014	60.7%	61.4%	59.7%	56.3%	56.8%
Cost/utilization variable					
Total healthcare expenditures, 2014$	$3017	$3813	$6228	$10,323	$21,470

individual in fair health had expenditures that were $5.48 lower than the basic model predicts, and $3.71 lower than the full model predicts. At the upper tail, where a well-informed insurance choice becomes critical, the basic model underpredicted the 99th percentile costs by $886, while the full model underpredicted by $832. This difference of $54 was the largest difference across the five SHS groups. Moreover, the full model was actually less accurate than the basic model at the 99th percentile for those in poor health. Thus, while the full model performed slightly better, on average, than the basic model, neither model performed well at the upper tail of the distribution.

Discussion

This study found that SHS alone was not a strong predictor of medical expenditures on an individual level, despite strong correlations between better health and lower

Table 2. Coefficients of Basic vs. Full Models of Healthcare Spending.

Coefficient	SHS-Only Model	Basic Model	Full Model
Stage 1 (Probability of Expenses Being Incurred)			
Intercept	1.219***	−0.545***	−0.662***
Health related			
SHS (0=poor, 4=excellent)	−0.160***	0.052***	N.S.
Age		0.015***	0.010***
Sex		0.914***	0.785***
Age * Sex		−0.011***	−0.009***
Number of conditions		0.322***	0.260***
Smoker, currently		−0.198***	−0.119**
Access related			
Has usual source of care			0.486***
Self-reported mental health status			−0.052***
In Medicaid expansion state			0.078**
Has public insurance			0.559***
Has private insurance			0.524***
Has high income (>400% FPL)			0.188***
Educational attainment level			0.186***
Lives in rural county			−0.203**
Non-white			−0.331***
Stage 2 (magnitude of logged expenses given that expenses are incurred)			
Intercept	0.113***	1.651***	0.515*
Inverse Mills ratio	18.629**	−0.414	0.488**
Health related			
SHS (0=poor, 4=excellent)	−1.650***	−0.105***	−0.179***
Age		0.016***	0.017***
Senior citizen		0.241***	0.201***
Sex		0.660***	0.975***
Age * Sex		−0.008***	−0.012***
Number of conditions		0.243***	0.281***
Experienced pain in last 4 weeks		0.097***	0.111***
Access related			
Has usual source of care			0.355***
Has public insurance			0.708***
Has private insurance			0.642***
Has high income (>400% FPL)			0.212***
Educational attainment level			0.117***
Non-white			−0.493***
Adjusted R2 for model	0.070	0.240	0.290
Root mean squared error	1.731	1.544	1.507

Notes: Blank cells indicate that a variable was not included in the model. N.S. indicates that a variable was not statistically significant and was excluded. Significance levels are indicated by * (P<0.10), ** (P<0.05), and *** (P<0.01).

expenditures at the aggregate level. The addition of a small number of key health-related variables such as age, sex, and number of chronic conditions, however, led to significant improvements in the amount of variation explained by the model. The addition of access-related variables such as possession of health insurance and a usual source of healthcare, most of which were associated with higher expenditures, improved accuracy marginally. Neither model performed well in predicting the upper tail of the cost distribution where individuals incur high costs in a given year.

The basic model, which included variables directly related to health, explained significantly more variation in the data than the SHS-only model could. The model predicted future expenditures based upon previous data, performing well for a large majority of the sample. However, it did poorly at predicting expenditures at the highest end of the distribution, as did the SHS-only model.

After controlling for health needs through basic health-related variables, having high levels of access—geographic, cost-related, and personal—in a given year was associated with higher costs the following year. However, the "value
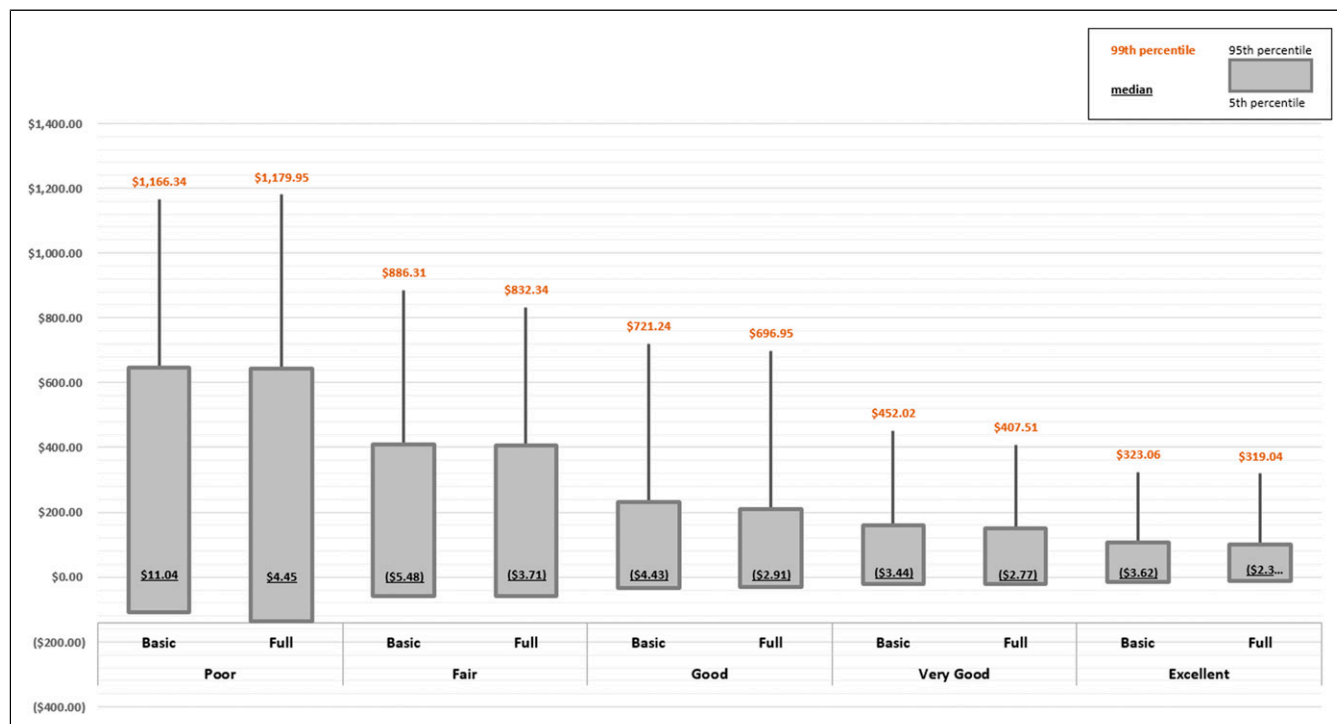
Figure 1. Predictive Accuracy (Actual—Predicted Expenditures) of Basic vs. Full Models Across Five Health status levels.

added" in terms of modeling accuracy was quite modest, suggesting that a person's basic health-related data and SHS stand in, to a large extent, for these other variables. In particular, the findings are consistent with the notion that SHS at a given point in time is already a product of the access a person has had in the past, and is likely very similar to their current access measures. Insurance status and type, income level, geographic proximity to care, and urban/rural status can impact healthcare access.[25] Barriers to access are sometimes present for racial and ethnic minorities,[26] and for those of lower educational attainment.[25] Because the dependent variable is in log form, it is not straightforward to interpret the coefficients themselves, but the signs and other relationships help illustrate the relative importance of the key health-related variables.

Among the access variables included as significant in the final model, we note that many relate directly to the ability to access healthcare (rural geography, residence in a Medicaid expansion state, and having health insurance, high income, and a usual source of care) in a logistical or economic sense, and that others relate only indirectly (race, educational status). We found that better health status was more strongly predictive of lower spending when we controlled for all geographic, cost-related, and personal access variables. The negative coefficient on non-white race (-0.493) reinforces extensive literature on lower access for racial minorities, since we are controlling for health need through SHS and number of health conditions. Estimates of the public and private insurance variables indicate, unsurprisingly, that both

are associated with higher expenditures, compared to having no insurance.

Even if access variables had added significantly to the predictive power of the model, it is likely inappropriate to include them for purposes of guiding insurance plan selection if they stem from inequitable allocation of resources rather than underlying health need. The implication would be that utilization should be estimated under the assumption that such inequities will persist. Similarly, a model that controlled for more specific details of health insurance coverage, while potentially being more accurate, would risk, in a decision aid application, steering consumers to high-deductible health plans that could end up influencing them to avoid seeking necessary care.

Ultimately, we found that all three models are far from accurate when it comes to predicting very high outlier expenditures, in the sense that the size of the error can be large enough that a higher up-front premium to obtain a lower deductible would have been a good decision for the individual. Predicating a specific insurance plan choice on such a prediction would likely be a suboptimal choice for such individuals. The slight differences between them were not meaningful in dollar terms, and much additional sensitive and private information was required to generate the full model estimates. It is possible that predicting particularly high costs may be improved with more complex modeling, for example, machine-learning techniques such as LASSO that test for higher-order relationships among variables, or by including variables capturing functional status, frailty, or other high-cost conditions with greater specificity than

SHS. A study with a different model specification in which insurance choice data are examined to predict the chance of high, outlier spending, defined *ex ante* as an amount that would cause a high-deductible health plan to be a bad decision *ex post*, could be useful in quantifying the error rate. Alternatively, examination of particularly high costs may reveal that a large proportion are essentially unpredictable (eg, an accident, an unexpected cancer diagnosis).

Limitations

We included variables in the basic model which we judged that people would be willing to share in the context of health insurance decision aids (all were variables directly connected to health while avoiding questions on race, income, and other potentially sensitive subjects). However, validating the acceptability of this list of questions was beyond the scope of the current study.

We used the most recently available MEPS panels at the time the analyses were conducted in February 2020, but more recent data would now be available through a new restricted-access request. Our primary goal was to predict year 2 utilization based upon year 1 information, and we do not anticipate these trends to vary greatly due to the age of the data.

Conclusion

SHS was not an entirely satisfactory predictor of individual medical expenditures, even when augmented with other variables likely to be available in a decision-aid context. The most significant variation in plan benefit designs within a given insurance market is the contrast between standard and high-deductible plans, and the errors in our model were of a magnitude that would sometimes produce errors in recommending one of these options over the other. Decision aids may be quite useful as health insurance literacy interventions, which should focus on communicating the *concepts* of low-probability, high-cost risks, and how different types of insurance can mitigate them.[9,27,28] While we find that our SHS-based models perform well in the aggregate, predicting population-level risk well, we also caution that they are not sufficiently accurate to guide individuals' insurance shopping decisions in all cases.

Policies that rely heavily on health insurance consumers making individually optimal choices cannot assume that decision tools can accurately anticipate high costs. Only 1% of Health Insurance Marketplace consumers failing to buy comprehensive coverage when they "should" have done represents more than 100,000 people making a costly mistake—or finding themselves in a situation in which they cannot afford needed care. This analysis calls into question the policy approach of relying heavily on a model in which savvy health insurance consumers assess risks and tradeoffs, weighing expected cost and other factors to arrive at an optimal choice. Insurance exists in part to protect against financial duress, that is, to help individuals with outlier costs, and it appears that no simple decision aid can contribute meaningfully to this assessment of outlier risk, even as it may educate consumers as to the nature of risks and tradeoffs. However, modern health insurance is also about facilitating access to care. More work should explore how shopping across plans which vary in their quality, networks, and other features (perhaps holding cost constant as a function of income) could improve individual- and societal-level health outcomes.

ORCID iDs

Abigail R. Barker https://orcid.org/0000-0002-0826-5156
Karen E. Joynt Maddox https://orcid.org/0000-0003-3779-8653
Ellen Peters https://orcid.org/0000-0003-0702-6169
Mary Politi https://orcid.org/0000-0001-9103-6495

References

1. Barnes AJ, Hanoch Y, Rice T. Can Plan Recommendations Improve the Coverage Decisions of Vulnerable Populations in Health Insurance Marketplaces? *PLoS One*. 2016;11(3): e0151095. doi:10.1371/journal.pone.0151095.
2. Abaluck J, Gruber J. Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *Am Econ Rev*. 2011;101(4):1180-1210. doi:10.1257/aer.101.4.1180.
3. Bhargava S, Loewenstein G, Sydnor J. Choose to Lose: Health Plan Choices from a Menu with Dominated Option*. *Q J Econ*. 2017;132(3):1319-1372. doi:10.1093/qje/qjx011.
4. Einav L, Finkelstein A, Mullainathan S, Obermeyer Z. Predictive Modeling of U.S. Health Care Spending in Late Life.

Science. 2018;360(6396):1462-1465. doi:10.1126/science.aar5045.

5. Barnes AJ, Hanoch Y, Rice T. Determinants of Coverage Decisions in Health Insurance Marketplaces: Consumers' Decision-Making Abilities and the Amount of Information in Their Choice Environment. *Health Serv Res.* 2015;50(1):58-80. doi:10.1111/1475-6773.12181.

6. Housten AJ, Furtado K, Kaphingst KA, Kebodeaux C, McBride T, Cusanno B, et al. Stakeholders' perceptions of ways to support decisions about health insurance marketplace enrollment: a qualitative study. *BMC Health Serv Res.* 2016; 16(1):634. doi:10.1186/s12913-016-1890-8.

7. *Checkbook Plan Comparison Tool 2021.* Healthplanratings.org. https://www.healthplanratings.org/hie/. Published 2021. Accessed November 5, 2021.

8. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *J Med Internet Res.* 2013;15(4): e85. doi:10.2196/jmir.1933.

9. Politi MC, Shacham E, Barker AR, George N, Mir N, Philpott S, et al. A Comparison Between Subjective and Objective Methods of Predicting Health Care Expenses to Support Consumers' Health Insurance Plan Choice. *MDM Policy & Practice.* 2018; 3(1):238146831878109. doi:10.1177/2381468318781093.

10. Hunt SM, McEwen J, McKenna SP. Measuring Health Status: A New Tool for Clinicians and Epidemiologists. *J Roy Coll Gen Pract.* 1985;35(273):185-188.

11. Wherry LR, Burns ME, Leininger LJ. Using Self-Reported Health Measures to Predict High-Need Cases among Medicaid-Eligible Adults. *Health Serv Res.* 2014;49(suppl 2):2147-2172. doi:10.1111/1475-6773.12222.

12. Boscardin CK, Gonzales R, Bradley KL, Raven MC. Predicting Cost of Care Using Self-Reported Health Status Data. *BMC Health Serv Res.* 2015;15:406. doi:10.1186/s12913-015-1063-1.

13. Cunningham PJ. Predicting High-Cost Privately Insured Patients Based on Self-Reported Health and Utilization data. *Am J Manag Care.* 2017;23(7):e215-e222.

14. Kaplan G, Baron-Epel O. What Lies Behind the Subjective Evaluation of Health Status? *Soc Sci Med.* 2003;56(8): 1669-1676. doi:10.1016/s0277-9536(02)00179-x.

15. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science.* 2019;366(6464):447-453. doi:10.1126/science.aax2342.

16. U.S. Department of Health & Human Services. *Medical Expenditure Panel Survey.* Rockville, Maryland: Agency for Healthcare Research and Quality; 2021. https://www.meps.ahrq.gov/mepsweb/. Accessed November 5, 2021.

17. United States Department of Agriculture Economic ResearchUrban Service. *Influence Codes. Kansas City, MO: USDA's Economic Research Service; 2021.* https://www.ers.usda.gov/data-products/urban-influence-codes/. Accessed November 5, 2021. KansasUSDA's Economic Research Service City, MO:

18. Health Resources and Services Administration. *Data from: Area Health Resources Files*; 2019.

19. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to Mental Health Treatment: Results from the National Comorbidity Survey Replication. *Psychol Med.* 2011; 41(8):1751-1761. doi:10.1017/s0033291710002291.

20. Mechanic D. Removing Barriers To Care Among Persons With Psychiatric Symptoms. *Health Aff* 2002;21(3):137-147. doi:10.1377/hlthaff.21.3.137.

21. Hu Y, Goldman N. Mortality Differentials by Marital Status: An International Comparison. *Demography.* 1990;27(2):233-250.

22. Ben-Shlomo Y, Smith GD, Shipley M, Marmot MG. Magnitude and Causes of Mortality Differences between Married and Unmarried men. *Journal of Epidemiology and Community Health.* 1993;47(3):200-205. doi:10.1136/jech.47.3.200.

23. Smedley B, Stith A, Nelson A. , Institute of Medicine (US). Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, D.C.: National Academies Press; 2003.

24. Heckman JJ. Sample Selection Bias as a Specification Error. *Econometrica.* 1979;47(1):153-161. doi:10.2307/1912352.

25. Office of Disease Prevention and Health Promotion. *Access to Health Services.* Washington, D.C: U.S. Department of Health and Human Services. https://www.healthypeople.gov/2020/topics-objectives/topic/Access-to-Health-Services.

26. Agency for Healthcare Research and Quality. *Access and Disparities in Access to Health Care.* Rockville, MD: Agency for Healthcare Research and Quality. https://www.ahrq.gov/research/findings/nhqrdr/nhqdr15/access.html (2016).

27. Politi MC, Grant RL, George NP, et al. Improving Cancer Patients' Insurance Choices (I Can PIC): A Randomized Trial of a Personalized Health Insurance Decision Aid. *Oncologist* 2020 Jul;25(7):609-619. doi:10.1634/theoncologist.2019-0703.

28. Politi MC, Kuzemchak MD, Liu J, Barker AR, Peters E, Ubel PA, et al. Show Me My Health Plans. *MDM Policy & Practice.* 2016; 1(1):238146831667999. doi:10.1177/2381468316679998.