

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

Public Domain Mark 1.0

<https://creativecommons.org/publicdomain/mark/1.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

**Please provide feedback**

Please support the ScholarWorks@UMBC repository by emailing [scholarworks-group@umbc.edu](mailto:scholarworks-group@umbc.edu) and telling us what having access to this work means to you and why it's important to you. Thank you.

# Predicting avoidable hospital events in Maryland

Morgan Henderson PhD<sup>1</sup>  | Fei Han PhD<sup>1</sup> | Chad Perman MPP<sup>2</sup> |  
Howard Haft MD<sup>2</sup> | Ian Stockwell PhD<sup>1</sup>

<sup>1</sup>The Hilltop Institute, University of Maryland, Baltimore County (UMBC), Baltimore, Maryland, USA

<sup>2</sup>Maryland Primary Care Program, Maryland Department of Health, Baltimore, Maryland, USA

## Correspondence

Morgan Henderson, The Hilltop Institute, University of Maryland, Baltimore County (UMBC), 1000 Hilltop Circle, Baltimore, MD 21250, USA.  
Email: mhenderson@hilltop.umbc.edu

## Funding information

Maryland Department of Health

## Abstract

**Objective:** To develop and validate a prediction model of avoidable hospital events among Medicare fee-for-service (FFS) beneficiaries in Maryland.

**Data sources:** Medicare FFS claims from Maryland from 2017 to 2020 and other publicly available ZIP code-level data sets.

**Study design:** Multivariable logistic regression models were used to estimate the relationship between a variety of risk factors and future avoidable hospital events. The predictive power of the resulting risk scores was gauged using a concentration curve.

**Data collection/extraction methods:** One hundred and ninety-eight individual- and ZIP code-level risk factors were used to create an analytic person-month data set of over 11.6 million person-month observations.

**Principal findings:** We included 198 risk factors for the model based on the results of a targeted literature review, both at the individual and neighborhood levels. These risk factors span six domains as follows: diagnoses, pharmacy utilization, procedure history, prior utilization, social determinants of health, and demographic information. Feature selection retained 73 highly statistically significant risk factors ( $p < 0.0012$ ) in the primary model. Risk scores were estimated for each individual in the cohort, and, for scores released in April 2020, the top 10% riskiest individuals in the cohort account for 48.7% of avoidable hospital events in the following month. These scores significantly outperform the Centers for Medicare & Medicaid Services hierarchical condition category risk scores in terms of predictive power.

**Conclusions:** A risk prediction model based on standard administrative claims data can identify individuals at risk of incurring a future avoidable hospital event with good accuracy.

## KEYWORDS

forecasting, hospitalization, Medicare, models, risk assessment, statistical

## What is known on this topic

- Risk prediction models are increasingly used by providers and payers in order to allocate care resources.
- These models tend to be proprietary and can incorporate systematic biases.
- Individuals at the highest risk of a particular adverse outcome are not necessarily the same individuals that will benefit the most from a particular intervention.

**What this study adds**

- Seventy-three risk factors are strongly predictive of future avoidable hospital events among Medicare fee-for-service beneficiaries in Maryland in our primary model from June 2020.
- The resulting individual-level risk scores capture significant variation in the underlying risk of incurring a future avoidable hospital event.

**1 | INTRODUCTION**

On January 1, 2019, Maryland implemented the Total Cost of Care (TCOC) Model, a one-of-its-kind program intended to transform care and contain costs in the state of Maryland. A key element of the TCOC model is the Maryland Primary Care Program (MDPCP), which provides funding and support to primary care providers for the delivery of advanced primary care throughout the state. This program is set to run for 8 years, provided Maryland meets the terms of the Model.<sup>1</sup>

In order to support providers in their care management efforts, the MDPCP provides participating practices with patient-specific risk scores intended to capture each beneficiary's risk of incurring a potentially avoidable hospitalization or emergency department (ED) visit. The Hilltop Institute at the University of Maryland, Baltimore County (UMBC) has developed the predictive models needed to operationalize these risk scores. These patient-level risk scores are provided to participating practices on a monthly basis via the Chesapeake Regional Information System for our Patients (CRISP) and were first released on October 11, 2019, to 375 practices covering 211,461 attributed Medicare fee-for-service (FFS) beneficiaries in Maryland. The population of attributed beneficiaries and participating practices grew to 329,412 and 476, respectively, as MDPCP entered its second year in January 2020, and 392,241 and 525 as of January 2021. This pool of attributed beneficiaries represents almost 43% of all Medicare FFS beneficiaries in Maryland.<sup>2</sup>

The models represent a significant advance in the use of predictive modeling in a publicly funded health system. To our knowledge, this is the first instance of a state adopting a near-to-real-time predictive risk scoring model for use in the allocation of scarce care coordination resources with the goal of improving patient care. While the landscape of predictive models in health services is emergent,<sup>3</sup> fragmented,<sup>4</sup> and proprietary,<sup>5</sup> this use case demonstrates the viability of the development and dissemination of event risk modeling in publicly funded health systems.

This article will discuss the theoretical underpinnings, development, performance, validation, and limitations of these risk scores as of July 2020. However, where possible, we also refer to more recent developments.

**2 | THEORETICAL UNDERPINNING**

Our risk scores are intended to facilitate the primary care transformation process by improving efficiency in the allocation of scarce care

coordination resources. If such resources are limited and the patients in a given practice differ in the benefit they would obtain through receipt of care coordination, then patient outcomes are optimized by focusing the limited care coordination resources on the patients for whom these resources will generate the most benefit.

It is important to note that it is not necessarily the case that the individuals at highest risk of incurring a particular outcome are the same as those individuals who will benefit the most from a particular intervention intended to prevent that outcome. The identification of individuals at highest risk of an event is a prediction problem; the question of which individuals will benefit the most from a particular intervention is, however, a question of causal inference. In order to answer this causal question, it would be necessary to consider the counterfactual: that is, to compare outcomes between an individual who received that intervention and a similar individual who did not receive that intervention. Machine learning prediction models are optimized to predict an outcome, not perform causal inference, and causal inference methods are not optimal for prediction.<sup>6,7</sup>

Many researchers simplify this distinction by assuming that the individuals at the highest risk of incurring an event would also benefit the most from an intervention designed to prevent that event.<sup>5</sup> However, it is possible that for many high-risk individuals, the event in question is not preventable: for example, recent research has demonstrated that a coordinated care program targeted toward “superutilizers” did not lead to a reduction in hospital admissions.<sup>8</sup> To that end, while we do not attempt to identify the individuals who would benefit the most from an intervention, we carefully selected the outcome for this prediction model in order to avoid conflating prediction and causality.

Potentially avoidable hospitalizations/ED visits are those incurred for medical conditions or diagnoses “for which timely and effective outpatient care can help to reduce the risks of hospitalization by either preventing the onset of an illness or condition, controlling an acute episodic illness or condition, or managing a chronic disease or condition.”<sup>9</sup> By focusing on *avoidable* hospital events, we are predicting events which, by definition, can be prevented; therefore, the individuals with the highest probability of an avoidable hospital event (should) have the highest potential benefit from an intervention that leads to the prevention of that event. Moreover, while certain models use future health care costs as a proxy for future health care need, we focus on *event* risk in order to avoid the possibility of introducing bias into our predictions due to differential access to health care services. Through the dissemination of these risk scores, we aim to help practices identify these high-risk individuals, so that providers can allocate their care management resources accordingly.

### 3 | DATA

Our predictive models are primarily based on medical, hospital, and pharmacy claims for Medicare FFS beneficiaries in Maryland. Each month, we receive 36 months of part A claims, part A revenue centers, part A procedure codes, part A diagnosis codes, part B claim lines, part B durable medical equipment claims, part D claims, and patient demographic information (which also includes eligibility information) from CMS. The data span the 36 most recent months up to the previous calendar month: for example, the data we received in July 2020 cover the period of July 2017 to June 2020. We also receive beneficiary attribution files and practice rosters each quarter. Additionally, in order to operationalize and include environmental risk factors, we use publicly available data from a variety of sources. We provide detail on these risk factors and data sources in the supplement, and additional detail can be found in the publicly available model documentation.<sup>10</sup>

As of July 2020, the resulting data comprise 350,404 individuals across 473 practices. These individuals incurred approximately 3.0 million part A claims, 54.7 million part B claim lines, and 19.5 million part D claims in the 3-year period from July 2017 to June 2020.

### 4 | MODEL

#### 4.1 | Risk factors

The first step in the development process for the models was a targeted literature review with the goal of locating peer-reviewed academic journal articles that identified risk factors for potentially avoidable hospital events, thus providing a basis for risk factor extraction and feature creation. Using inclusion and exclusion criteria designed to reflect the MDPCP patient population, the research team screened over 3300 articles in both a primary and secondary literature search, ultimately selecting 211 articles for risk factor extraction.<sup>11</sup>

Based on a review of these articles, the research team initially coded 190 risk factors for inclusion in the model. These risk factors covered six domains: diagnoses, pharmacy utilization, procedure history, prior utilization, social determinants of health, and demographic information. For diagnosis-based risk factors, we relied on Chronic Conditions Data Warehouse (CCW) coding specifications in order to generate beneficiary-level risk factors that represent underlying disease states.<sup>12,13</sup> We later added eight additional risk factors to the model on the basis of stakeholder feedback and additional literature review, for a total of 198 risk factors.<sup>10</sup> We have included additional information on our variable definitions in Table S2.

#### 4.2 | Statistical methodology

In order to model future avoidable hospital events, we created a person-month panel dataset spanning 35 months in which risk factors are updated every month comprising approximately 11.6 million

person-month observations. Then, we used a multivariable logistic regression model that uses current values of procedural, diagnostic, utilization-based, pharmacy, demographic, and environmental risk factors to predict the likelihood that an individual incurs an avoidable hospitalization or ED visit in the *following* month. The outcome is a 0/1 indicator variable denoting whether an individual incurred an avoidable hospitalization or ED visit in a given month. In order to construct this measure, we relied on 2018 technical definitions provided by the Agency for Healthcare Research and Quality (AHRQ) as part of its prevention quality indicator (PQI) measures. This is a composite measure capturing diagnoses for diabetes complications, chronic obstructive pulmonary disease (COPD) or asthma, hypertension, heart failure, dehydration, bacterial pneumonia, and urinary tract infections.<sup>14</sup>

The statistical model is trained on an 80% sample of our analytical person-month data set (sampled at the person level). The functional form of the statistical model is:

$$\log\left(\frac{p_i(t)}{1-p_i(t)}\right) = \varphi(t) + \beta X_i(t-1) + \Omega V_i + t_i.$$

In this model,  $\varphi(t)$  is a cubic spline function of time at risk;  $\beta$  and  $\Omega$  are the vectors of model parameters to be determined by training data;  $X_i(t-1)$  is a vector of patient  $i$ 's time-dependent features in the previous month;  $V_i$  is a vector of patient  $i$ 's time-independent features;  $p_i(t)$  is the probability of avoidable hospitalization or ED visit of patient  $i$  at time  $t$  (i.e., the month following the realization of the risk factors); and  $t_i$  is time-at-risk in months. The time-at-risk variable measures the number of months that an individual has been in the 36-month rolling window and is included because we do not have a balanced 36-month panel. Because individuals can join the sample at any point in the 36-month window (e.g., when they age into Medicare, or when they move into the state), while we have the full 36 months of data for most individuals, this is not true for all individuals in our sample. The inclusion of this variable is intended to capture both the increasing reliability of covariates, given the accumulation of experience and any factors that might affect the probability of avoidable hospital events that are correlated with number of months that an individual has been in the sample. We use stepwise variable selection in order to retain only the statistically significant predictors (but overall risk scores do not meaningfully change when using all risk factors to generate risk scores).

It is not the case that all risk factors are available for every person-month. We use a 12-month lookback period for most of the time-varying risk factors, implying that an individual with, for example, 5 months of claims history will have incomplete information in her risk factors: if this individual truly has chronic kidney disease (CKD), then it is possible that she will not amass the claims history by month 5 that meets the qualifications required for a CKD flag in our model. Furthermore, while most individuals in the data have valid ZIP codes that link to the environmental risk factor data set, certain have ZIP codes for which there is no equivalent

ZIP Code Tabulation Area (ZCTA) and, therefore, receive no environmental risk factors.

Given that the risk scores are in operational use, it is imperative that each individual in a scoring cohort receives a score, no matter how (in)complete their risk factors. Moreover, it is also imperative to avoid the underestimation of person-level risk based on data availability: it would be inaccurate to code an individual as not having CKD (as above) when, in fact, they did not have sufficient claims history to qualify for a CKD flag. Doing so would impose the assumption that that individual does not have CKD when, in fact, the true value is unknown. To avoid this issue, we split the scoring sample into four separate partitions, depending on data availability: group 1 consists of individuals with 12 or more months of claims experience and a valid ZIP code, and the scoring sample includes all risk factors. Group 2 consists of individuals with fewer than 12 months of claims availability but with a valid ZIP code, and the scoring sample has data on only geographic and demographic risk factors. Group 3 consists of individuals with 12 or more months of claims experience without a valid ZIP code, and the scoring sample includes all but the geographic risk factors. Finally, group 4 consists of individuals with fewer than 12 months of experience and without a valid ZIP code, and the scoring sample includes only the demographic risk factors.

We train the corresponding models on the same analytic dataset: those individuals with at least 12 months of experience in the sample and with valid ZIP codes. For group 1, we include all risk factors; for group 2, we include only geographic and demographic risk factors; for group 3, we include all except geographic risk factors; and for group 4, we include only demographic risk factors. Thus, the risk factors that are present (e.g., demographic factors) are allowed to compensate for the risk factors that are not present (e.g., condition flags due to lack of experience in the sample) for a given partition.

In general, we retrain the model once per quarter to generate updated risk factor coefficients. However, during the COVID-19 pandemic, we retrained the model monthly so that the resulting risk scores reflected, as much as possible, the changing conditions in the health services landscape.

### 4.3 | Scoring

The four risk models are trained as described above in order to estimate the vectors of coefficients for the risk factors in each model. Then, using the most recently available month of risk factors (i.e., the “person-now” data set), the scoring sample is partitioned into the four groups based on data availability, and individuals are scored using the corresponding model coefficients applied to their risk factors in the person-now data set. For the risk scores created in July 2020 and released in August 2020, there were 347,710 individuals in the scoring cohort: 340,984 in group 1 (98.1%); 5234 in group 2 (1.51%); 1473 in group 3 (0.42%); and 19 in group 4 (0.00%). Individual probabilities of incurring an avoidable hospital event are then calculated based on the logistic regression coefficients using the standard transformation.

## 5 | RESULTS

### 5.1 | Coefficients

As of July 2020, the most recent model re-training had occurred in June 2020, using an analytical dataset spanning June 2017–May 2020 and consisting of 11,653,740 person-months and 346,580 unique individuals. The model was trained on an 80% sample of this analytical dataset. All 198 risk factors were candidates for inclusion in the group 1 model, and risk factors were selected for inclusion and retention using stepwise selection, with a cutoff value based on the Bayesian Information Criterion (BIC). In the June 2020 model training, 73 risk factors survived the variable selection and entered into the final scoring model for group 1. All included risk factors, across all models, have *p* values of 0.0012 or lower.

Table 1 presents the top 30 of these risk factors grouped by risk factor domain in our group 1 model. While the coefficients do not have a causal interpretation, they largely accord with intuition: for example, individuals with a history of COPD are at significantly higher risk of incurring an avoidable hospital event (a composite event which includes hospitalizations for COPD) in the following month than individuals without COPD. Full group 1 model coefficients, and coefficients from other models, are available in Tables S3–S6.

### 5.2 | Predicted probabilities

The outcome of the models is a set of probabilities that estimate the patient-specific likelihood of incurring an avoidable hospital event in the following months. In general, these events are rare: for example, in December 2019, approximately 0.8% of MDCPC-attributed beneficiaries experienced an avoidable hospital event. As such, the predicted probabilities are low: for the 325,923 patients scored in February 2020, only 2.6% of the patient population had risk scores above 2%, and only 47 patients had risk scores above 50%.

Patient-level risk tends to persist across time: that is, high-risk patients tend to remain high-risk from 1 month to the next, and low-risk patients tend to remain low-risk. For example, risk scores from January 2020 to February 2020 display a correlation of 0.967; from December 2019 to January 2020, the correlation is 0.968. This is likely due to two factors. First, in order to prevent coding idiosyncrasies from introducing noise into the predictions, all risk factors are coded with at least 1 year of lookback. This has the consequence of making the risk factors relatively stable over time, thus smoothing out variation in the risk scores. Second, it is likely that true, underlying patient risk is also persistent: if some patients tend to have high (or low) risk for structural reasons, then the risk scores should also be relatively stable across time.

However, large month-to-month changes can occur in risk scores for two reasons. First, using a given set of risk factors coefficients, any changes in underlying risk factors will lead to changes in patients' predicted risk. For example, if an attributed beneficiary meets the conditions for heart failure beginning in December 2019, then her risk

**TABLE 1** Top 30 risk factor coefficients for model 1

Risk factor	Odds ratio
<b>Diagnosis risk factors</b>	
Indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis	1.640
Indicator for retinopathy	1.597
Indicator for heart failure	1.421
Indicator for urinary tract infection	1.334
Indicator for hypertension	1.267
Indicator for tobacco use	1.214
Indicator for problems with care provider dependency	1.212
Indicator for arrhythmia	1.211
Indicator for fluid and electrolyte imbalance	1.210
Indicator for chronic kidney disease	1.194
Indicator for metastatic cancer	1.190
Indicator for intellectual disabilities and related conditions	1.189
Indicator for ischemic heart disease	1.160
Indicator for asthma	1.150
Indicator for diabetes with complications	1.144
Indicator for pressure and chronic ulcers	1.138
<b>Prior utilization risk factors</b>	
Prior hospitalization discharge status—other	2.117
Indicator for hospice enrollment	1.677
Prior hospitalization admission type—emergency	1.398
Number of previous avoidable hospitalizations	1.375
Discontinuity of primary care—proportion	1.303
Prior hospitalization admission type—urgent	1.255
Indicator for durable medical equipment (DME) use	1.213
<b>Pharmacy risk factors</b>	
Indicator for insulin use	1.257
Indicator for no statin use	1.145
<b>Demographic risk factors</b>	
Beneficiary race—Black	1.604
Beneficiary race—Hispanic	1.436
Indicator for original Medicare eligibility for a non-age related cause	1.408
Indicator for dual eligibility with Medicaid	1.257
Beneficiary race—White	1.183

*Note:* These represent the top 30 largest coefficients from the model training from June 2020, sorted descending by coefficient within risk factor domain. The coefficients are from Model 1. All have *p* values of 0.0012 or lower. Coefficients are presented as odds ratios for ease of interpretation. We do not present the model intercept, time-at-risk, or cubic spline terms here; full coefficients are available for Models 1–4 in Tables S3–S6.

*Source:* Authors' analysis of Medicare FFS claims.

score will rise significantly from November 2019 to January 2020 because of that underlying change. Second, we estimate new risk factor coefficients with each model re-training. As a result, not only can

the underlying risk factors for a given patient change from 1 month to the next, but the relationship between that risk factor and the risk of avoidable hospital events can also change upon retraining.

### 5.3 | Prediction validation

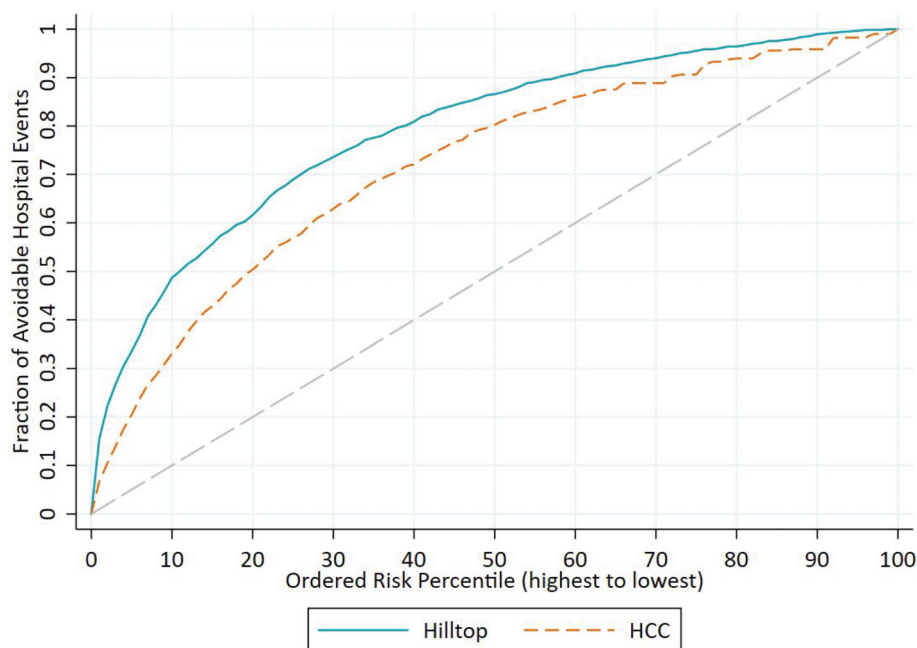
Postdeployment model evaluation is a crucial component of the predictive model lifecycle. While we gauge model performance on hold-out samples during model training, we also track how well the model performs in a production environment: that is, how well the risk scores predict actual avoidable hospital events in the MDPCP population. The first round of risk scores were released to participating providers on October 11, 2019, and subsequent rounds have been released on the Friday of the first full week of each month. In general, by approximately 3 months following each release, we have sufficient claims experience to compare the risk scores with actual experience.

We test the predictive performance of the risk scores in a production environment by estimating the predictive power of the risk model on *actual* avoidable hospital events in the month immediately following the release of the risk scores. Good model performance would be indicated by the assignment of high-risk scores to individuals who actually do incur avoidable hospital events in the following month.

Traditionally, the discriminatory power of predictive models has been summarized using the C-statistic, which is a measure of the area under the receiver operating characteristic (ROC) curve.<sup>15</sup> The ROC curve plots the true-positive rate against the false-positive rate for binary classifiers using successive cutoff thresholds and “measures the probability that a randomly selected diseased subject has a higher predicted risk than a randomly selected nondiseased subject.”<sup>16</sup> However, the objective of the models is not binary classification but instead the estimation of individual-level risks of incurring an avoidable hospital event so that care managers can, by focusing on the riskiest individuals, intervene to prevent the most likely avoidable hospitalizations. To that end, the performance of the models is assessed using the concentration curve.

The concentration curve presents the cumulative share of all avoidable hospital events incurred by the riskiest X% of patients. In order to estimate the concentration curve, the patient cohort is ordered from most to least risky (in terms of predicted risk) on the X axis, and the fraction of total avoidable hospital events captured by the riskiest patients on the Y axis. Figure 1, below, presents the concentration curve for the risk scores released in April 2020 (the latest score release at the time of writing for which we had one complete month of outcomes). We find that the predictive performance of our model (the “Hilltop” model) is good: the riskiest 10% of patients in the MDPCP population incurred 48.7% of all avoidable hospital events in the month following the release of the scores, and the riskiest 20% of patients capture 61.6% of all avoidable hospital events. The risk scores in other months, not shown, display a similar level of predictive power: the top 10% riskiest patients capture 45%–50% of all avoidable hospital events in the coming month. Moreover, as of March





**FIGURE 1** Concentration curve for model risk scores, April 2020. This figure displays concentration curves for risk scores released on April 10, 2020, based on avoidable hospital events that occurred over the following month. Each point on the curve represents the fraction of total avoidable hospital events incurred by individuals in that risk percentile and above: for example, the point where  $x = 10$  indicates that top 10% riskiest individuals (per the Hilltop model) incurred almost 49% of all avoidable hospital events in the following month. The solid curve indicates the performance of the risk scores released on April 10, 2020, at predicting true avoidable hospital events in the month following release. The dashed line represents the performance of the Centers for Medicare & Medicaid Services (CMS) Hierarchical Condition Category risk scores at predicting actual avoidable hospital events over that same period.

Source: Authors' analysis of Medicare FFS claims [Color figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

2021, we have not observed any systematic changes to model performance during the COVID-19 pandemic relative to the prepandemic period.

It is worth noting that these risk scores significantly out-perform an alternative risk score that is available within CRISP: CMS's HCC risk scores. HCC scores use information on comorbidities and demographics in order to predict future individual-level Medicare spending.<sup>17</sup> We used this measure of risk in order to predict avoidable hospital events over the same period and found that, in contrast to our models, the CMS-HCC risk scores identify 33.1% of true avoidable hospital events in the top 10% riskiest beneficiaries and 50.4% of true avoidable hospital events in the top 20% riskiest beneficiaries. While the CMS-HCC scores are not calibrated to predict avoidable hospital events, we believe that this performance differential speaks to a wider point: event risk is distinct from financial risk. Models that predict event risk estimate the probability that a particular event will occur; models that predict financial risk estimate the cost that patients are likely to incur. While the two may be correlated in certain circumstances, in general, financial risk and event risk are not interchangeable, and recent research has demonstrated that models that predict patient costs tend to underpredict risk for Black patients relative to White patients.<sup>5</sup> To the extent that racial bias exists in access to care, the cost of care may be a misleading proxy for true health status.

## 6 | DISCUSSION

### 6.1 | Limitations

In general, these risk scores have four limitations. First, there is a lag of approximately 2 months from the most recent claims to the release of the scores. For example, claims that arrive in late July 2020 cover utilization through mid-June 2020, and we use these data to calculate risk scores released in August 2020. This is an unavoidable delay due to the nature of the administrative data we use for the risk model, but we are reassured that the production performance of the model is strong.

Second, the Medicare claims data do not include information on vital statistics—for example, blood pressure or lab results—meaning that we are unable to incorporate those clinical risk factors into the models. It is likely that development of clinical risk factors would improve the predictive power of the model, although researchers have documented only relatively modest improvements in model performance for claims-and-clinical models relative to claims-only prediction models in various settings.<sup>18,19</sup>

Third, in order to control for environmental factors that may affect patients' probabilities of incurring avoidable hospital events, we include a rich set of ZIP code-level covariates derived from publicly available sources. These data have two main limitations: first, they are

relatively coarse. Maryland has 468 ZCTAs, each containing, on average, roughly 13,000 Maryland residents. To the extent that risky individuals tend to live in the same ZIP codes, then ZIP code-level risk factors offer little predictive power. Second, the data are static: the environmental risk factors for a given attributed beneficiary do not change over time. This is largely due to data availability, as the publicly available data sources are only refreshed periodically.

Finally, while we focus on predicting *avoidable* hospital events in order to avoid conflating a prediction problem with causal inference, it is possible that some interventions are more effective than others at preventing avoidable hospital events depending on the patient's circumstances. For example, periodic medication reconciliation might be most useful for an individual for whom polypharmacy is the most salient risk factor, while postdischarge monitoring might be most effective for an individual for whom a recent hospitalization is the greatest risk factor. While we are currently unable to track which particular care management interventions practices actually employ upon receiving the risk scores, we believe that a more thorough understanding of the efficacy of particular interventions is an important avenue for future research.

## 6.2 | Lessons learned

We learned several lessons in the development and implementation of the models that may be of use to states or organizations developing large-scale predictive models intended for widespread, day-to-day use.

First, earning the trust of key stakeholders (e.g., providers and state officials) was crucial to the implementation and adoption of this model. A key element of earning this trust was model explicability: it was necessary to demonstrate *why* certain risk factors were included in the model, and *how* the model operated. This motivated both our literature review and use of transparent prediction methods.

Second, we incorporated significant redundancy into the model development in order to prevent coding errors that could potentially result in the misallocation of scarce care coordination resources away from individuals that are in significant need. To that end, we developed two separate codebases maintained by two separate researchers, and only release scores each month once sufficient model concordance is reached.

Third, we communicated often with stakeholders: multiple training sessions were held with providers, care transformation organization representatives, and care managers, and multiple versions of documentation (long, short, and FAQ) were produced and circulated. Moreover, we also solicited feedback during model development since we sought to provide the information that would be of greatest utility to end-users of the risk scores. Additionally, we responded to stakeholder feedback by adding a feature in the January 2020 release that displays, for a given attributed beneficiary, his or her top risk factors underlying the risk score.

Finally, we monitor the performance of the model closely. We have two versions of the model that were developed independently,

and each version of the model conducts periodic retraining and monthly scoring. We compare the models' coefficients following each re-training, and while they are typically very similar, any discrepancies lead to further investigation. Additionally, as noted above, we monitor the performance of our production model at predicting actual avoidable hospital events, with the idea that any significant change in model performance in the production environment also warrants investigation and adjustment.

## 7 | CONCLUSION

As part of the MDPCP, The Hilltop Institute developed and deployed a set of predictive models of avoidable hospital events. Risk scores are generated using a discrete time survival model, and risk factors are based on published research and derived from Medicare claims and publicly available data sets. As of July 2020, the scores are sent monthly via CRISP to providers for almost 350,000 Medicare FFS beneficiaries in approximately 475 practices and are intended to assist providers with the efficient allocation of scarce care coordination resources. Scores have been disseminated since October 2019, and, to our knowledge, this is the first instance of a state disseminating risk scores directly to providers for the purposes of care coordination. The risk scores show good predictive power in a production environment and superior performance to the CMS HCC model risk scores, and early qualitative feedback suggests that care managers find the scores to be clinically useful.

## ACKNOWLEDGMENTS

The authors would like to thank Chris Koller and Cynthia Woodcock for valuable feedback. The Maryland Department of Health provided funding for this research.

## ORCID

Morgan Henderson  <https://orcid.org/0000-0002-0869-5738>

## REFERENCES

- Center for Medicare and Medicaid Innovation. Maryland total cost of care model: Maryland Primary Care Program request for applications. Accessed August 7, 2020. <https://innovation.cms.gov/files/x/mdtcocm-rfa.pdf>
- Kaiser Family Foundation. State health facts. Accessed April 7, 2021. <https://www.kff.org/medicare/state-indicator/total-medicare-beneficiaries/>
- Emanuel EJ, Wachter RM. Artificial intelligence in health care: will the value match the hype? *JAMA*. 2019;321(23):2281-2282.
- Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. *Npj Digit Med*. 2019;2(1):1-3.
- 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.
- Kleinberg J, Ludwig J, Mullainathan S, Obermeyer Z. Prediction policy problems. *Am Econ Rev*. 2015;105(5):491-495.
- Mullainathan S, Spiess J. Machine learning: an applied econometric approach. *J Econ Perspect*. 2017;31(2):87-106.



8. Finkelstein A, Zhou A, Taubman S, Doyle J. Health care hotspotting—a randomized, controlled trial. *N Engl J Med*. 2020;382(2):152-162.
9. Billings J, Zeitel L, Lukomnik J, Carey TS, Blank AE, Newman L. Impact of socioeconomic status on hospital use in New York City. *Health Aff*. 1993;12(1):162-173.
10. Henderson M, Han F, Stockwell I. *Maryland Primary Care Program (MDPCP) Pre-AH Risk Score Specifications and Codebook (Version 3)*. The Hilltop Institute, UMBC; 2020.
11. Pelsner C, Henderson M, Stockwell I. *Risk Factors for Potentially Avoidable Hospital Admissions and Emergency Department Visits: A Literature Review*. The Hilltop Institute, UMBC; 2019.
12. Center for Medicare & Medicaid Services. CMS Chronic Condition Warehouse (CCW): CCW condition algorithms. Accessed August 7, 2020. <https://www.ccwdata.org/documents/10280/19139421/ccw-chronic-condition-algorithms.pdf>
13. Center for Medicare & Medicaid Services. CMS Chronic Conditions Data Warehouse (CCW): CCW chronic conditions reference list. Accessed August 7, 2020. <https://www.ccwdata.org/documents/10280/19139421/ccw-chronic-condition-algorithms-reference-list.pdf>
14. Agency for healthcare research and quality prevention quality indicators technical specifications. Accessed August 7, 2020. [https://www.qualityindicators.ahrq.gov/Modules/PQI\\_TechSpec\\_ICD10\\_v2020.aspx](https://www.qualityindicators.ahrq.gov/Modules/PQI_TechSpec_ICD10_v2020.aspx)
15. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*. 2010;21(1):128.
16. Mauguen A, Begg CB. Using the Lorenz curve to characterize risk predictiveness and etiologic heterogeneity. *Epidemiology*. 2016;27(4):531.
17. Center for Medicare & Medicaid Services. Report to congress: risk adjustment in medicare advantage. December 2018. Accessed August 7, 2020. <https://www.cms.gov/Medicare/Health-Plans/MedicareAdvSpecRateStats/Downloads/RTC-Dec2018.pdf>
18. Hammill BG, Curtis LH, Fonarow GC, et al. Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circ Cardiovasc Qual Outcomes*. 2011;4(1):60-67.
19. Morawski K, Dvorkis Y, Monsen CB. Predicting hospitalizations from electronic health record data. *Am J Manag Care*. 2020;26(1):e7-e13.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Henderson M, Han F, Perman C, Haft H, Stockwell I. Predicting avoidable hospital events in Maryland. *Health Serv Res*. 2022;57(1):192-199. doi:10.1111/1475-6773.13891