The Hilltop Institute
UMBC

Risk Score Specifications and Codebook for The Hilltop Institute's Pre- Models

December 19, 2023

Version 2

Maryland
DEPARTMENT OF HEALTH

Documentation Edit History

Version	Date	Description of Primary Change(s)
MDPCP 1	October 3, 2019	Initial Release
MDPCP 2	January 11, 2020	▪ Added clarification of time lag of estimates; consistency of risk scores over time; reasons for risk; and model performance in production ▪ Updated model coefficients and appendix table ▪ Added List of Tables and Figures
MDPCP 3	June 29, 2020	▪ Added section on nonlinear modeling tests ▪ Updated weighting methodology for environmental risk factors ▪ Added section on new risk factors as of June 2020 ▪ Updated Appendix 1 to reflect additional risk factors
Pre- Models 1	November 7, 2022	▪ Substantially restructured to accommodate new predictive models ▪ Added details about census tract-level environmental risk factors ▪ Updated Pre-AH outcome definition to reflect latest AHRQ PQI specifications
Pre- Models 1.1	April 27, 2023	▪ Updated Pre-HE Model risk score name.
Pre- Models 2	December 19, 2023	▪ Updated Pre-AH outcome definition to reflect 2022 AHRQ PQI specifications ▪ Updated CCW risk factors to reflect the CCW 30 Version (2022) ▪ Updated Appendix 1 to reflect additional risk factors

Suggested Citation: Henderson, M., Goetschius, L.G., Han, F., & Stockwell, I. (2023, December 19). *Risk score specifications and codebook for The Hilltop Institute's Pre- Models (Version 2)*. Baltimore, MD: The Hilltop Institute, UMBC.

Acknowledgements

The Hilltop Institute would like to thank Howard Haft, Chad Perman, Candice Morrison, Rachel Maxwell, and Stephanie Scott from the Maryland Department of Health, Craig Behm and Megan Priolo from CRISP, and Audrey Speter from hMetrix for their comments and support.

Risk Score Specifications and Codebook
for The Hilltop Institute's Pre- Models (Version 2)

Table of Contents

List of Tables and Figures

Tables

Figures

Risk Score Specifications and Codebook
for The Hilltop Institute's Pre- Models (Version 2)

Section 1. Introduction

In 2014, the state of Maryland partnered with the Centers for Medicare and Medicaid Services (CMS) to modernize its unique all-payer rate-setting system for hospital services to improve the overall health of Maryland residents by increasing health care quality and reducing the cost of care. In service of providing better care at lower costs, The Hilltop Institute at UMBC, in partnership with the Maryland Department of Health, has developed predictive risk stratification models to identify patients at high risk for potentially preventable health care utilization that can be used to help target care resources to the patients who need them most.

This document strives to explain the intended use, technical implementation, and model performance of the Hilltop Pre- Models as of November 2022. The Pre- Models are a suite of prediction tools spanning the Pre-AH Model™, Pre-CH Model™, Pre-DC Model™, and Pre-HE Model™. This document will be updated as the models are updated or when new models become operational, and significant changes will be noted in the documentation edit history table and in the text when necessary. This first section of the codebook provides a short introduction; the second section provides a general overview of data sources, training methodology, and scoring methodology; the third section provides specific details on the performance and operations of each model within the Hilltop Pre- Models suite; and the fourth section presents limitations.

As of November 2022, the Pre- Models are operational in two distinct populations:

- Medicare beneficiaries who are attributed to practices participating in the Maryland Primary Care Program (MDPCP). MDPCP is a key element of the Total Cost of Care (TCOC) All-Payer Model, an agreement between the CMS and the state of Maryland. MDPCP is a voluntary program that provides funding and support for the delivery of advanced primary care throughout the state. It allows primary care providers to play an increased role in the prevention and management of chronic disease, as well as in the prevention of unnecessary hospital utilization. As an important part of supporting providers in their care management efforts, the MDPCP provides event risk scores to participating practices of their attributed beneficiaries according to each patient's risk of incurring a model-specific outcome. Patient-level risk scores and reasons for risk are provided to participating medical practices every month via the MDPCP Prediction Tools area on Chesapeake Regional Information System for our Patients (CRISP).
- Medicaid recipients enrolled in the Maryland HealthChoice program. In HealthChoice, managed health care organizations (MCOs) provide services to Medicaid recipients by contracting with a network of licensed and certified health care providers. Patient-level risk scores are provided monthly to the MCOs to help identify patients at high risk for potentially preventable health care utilization.

Section 2. Pre- Models Overview, Data, and Methodology

Intended Use

The Hilltop Pre- Model risk scores are intended to facilitate improved efficiency in the allocation of scarce care coordination resources. Theoretically, if such resources are limited and the patients in a given practice panel differ in the benefit they would obtain through care coordination, then patient outcomes are optimized by focusing those care coordination resources on the patients for whom these resources will generate the most benefit.[1] Hilltop's models are intended to be used to rank attributed beneficiaries in each practice's or MCO's panel based on their risk of experiencing a potentially preventable utilization event in order to assist in the identification and care coordination efforts for those high-risk individuals.

Hilltop conceptualizes benefit, in this context, as the avoidance of a patient-specific adverse event. Many distinct adverse events are possible (ranging from disease onset to institutionalization to death), but for each model (i.e., the Pre-AH Model™), Hilltop treats these events as homogeneous and therefore focuses on patients' *probabilities* of incurring the specified outcome. This forms the theoretical foundation for the Hilltop Pre- Model framework: those individuals with the highest *probability* of incurring a potentially preventable utilization event are likely to benefit the most from advanced primary care services with respect to that outcome. Through the dissemination of risk scores and reasons for risk, Hilltop aims to facilitate the identification of these individuals within each practice or MCO so that they can allocate their care management resources accordingly.

It is crucial that the risk scores are as accurate as possible: ideally, the riskiest individuals as identified by the model have the highest *actual* likelihood of incurring a potentially preventable utilization event, and the individuals identified by the model as lowest risk have the lowest actual likelihood. Due to the nature of the modeling problem—estimating the distribution of risk, rather than binary classification—it is not appropriate to use the traditional Receiver Operator Characteristic curve as a measure of model fit. Instead, the utility of the model is assessed using *concentration curves,* which estimate the share of all avoidable hospital events occurring within the riskiest patients. Concentration curves can indicate, for example, that 50% of all patients who experience an avoidable hospital event are in the top 10% riskiest patients as estimated by the Hilltop Pre-AH Model™. Concentration curves and month-by-month summary scores for the model are presented for all models in the Predictive Power sections for each model and population, below.

[1] There is some evidence to suggest that different patients receive different benefits from care coordination services. Researchers have found that proactive care coordination interventions for patients with a high risk of hospitalization have so far led to reductions in avoidable hospitalizations, ED utilization, and readmissions for the Medicaid population but not the Medicare population (Berkowitz et al., 2018).

Clinical Vignette

In order to illustrate the intended use of the Hilltop Pre- Models, we have created a hypothetical clinical vignette using the Hilltop Pre-AH Model™ risk scores for an MDPCP practice. For the sake of exposition, the patient panel consists of thirteen patients, each of which represents ten similar patients. Table 1 displays the patient panel, along with each patient's (hypothetical) Hilltop Pre-AH Model™ risk score and CMS Risk Tier.

Table 1. Hypothetical Patient Panel

Patient Name	Pre-AH Risk Score (%)	CMS Risk Tier
Patient 1	75%	Complex[2]
Patient 2	15%	Complex
Patient 3	5%	Tier 4
Patient 4	4%	Complex
Patient 5	2%	Tier 3
Patient 6	1%	Tier 3
Patient 7	Less than 1%	Tier 2
Patient 8	Less than 1%	Tier 2
Patient 9	Less than 1%	Tier 1
Patient 10	Less than 1%	Tier 2
Patient 11	Less than 1%	Tier 1
Patient 12	Less than 1%	Tier 1
Patient 13	Less than 1%	Tier 1

Patients in this practice are listed in descending order of risk. Based on the most recently available month of risk factors spanning diagnoses, procedures, medications, utilization, demographics, and geographic information, in conjunction with risk coefficients derived from training data, Patient 1 (or, equivalently, the ten patients represented by Patient 1) has a 75% chance of incurring an avoidable hospital event in the near future.[3] Patient 2 is the next riskiest and has a 15% chance of incurring an avoidable hospital event. Patient 3 is the next riskiest, with a 5% chance. The distribution of risk is highly skewed: the majority of the practice's panel has

[2] It is important to note that while the CMS risk tier is correlated with Hilltop Pre-AH Model™ risk scores, the correlation is not perfect for two reasons: first, CMS risk tiers are based on underlying HCC score, which is conceptually distinct from the Pre-AH risk score. Second, certain groups of patients are automatically assigned to certain CMS risk tiers, which further reduces the correlation between the two measures. In particular, beneficiaries without sufficiently long clinical histories are assigned to CMS risk tier 2, while beneficiaries with "a diagnosis of dementia, substance use disorder, or severe and persistent mental illness" are assigned to the Complex tier, regardless of their HCC score (Center for Medicare & Medicaid Innovation, 2019). These individuals may, in turn, have relatively low (or high) risk of avoidable hospitalizations, meaning that an individual in, for example, the Complex CMS risk tier may have a low Pre-AH risk score. We highlight this point in Table 1 by presenting a non-monotonic relationship between Pre-AH risk score and CMS risk tier.

[3] See below for a more detailed discussion of the Pre-AH Model training and scoring process.

less than 1% chance of incurring an avoidable hospital event, and all but two of the patients have under a 6% event risk.[4]

Distributing available care coordination resources equally to all 130 underlying patients would result in each patient receiving a relatively small portion of available resources. This distribution of resources may be unlikely to have a significant impact on patient outcomes: the low-risk individuals would be low-risk even without the advanced primary care intervention, and the high-risk individuals may require more resource-intensive interventions to experience improvement in outcomes.[5] The Pre-AH Model™ risk scores, used in conjunction with provider clinical guidance, can assist practices with a more efficient and impactful allocation of their care management efforts.

Care Interventions

Hilltop remains agnostic as to the types of interventions that are best suited for the high-risk patients. Many interventions are possible, ranging from medication reconciliation to patient education to scheduling assistance, and patients are likely to respond best to different interventions based on their clinical and social needs. Interested readers should see published best practices in care coordination and care management.[6] Whatever the intervention strategy, Hilltop recommends that care managers and other users of the Hilltop Pre- Model risk scores allocate their effort first to individuals with the highest risk of incurring potentially preventable utilization events in the following month. This risk score is not, however, meant to override the clinical and subject matter expertise of the practice, their care transformation organization (CTO) partners, or the MCOs and should be used in conjunction with the practice's current care coordination protocols.

Risk Factor Overview

The risk factors in each of the Hilltop Pre- Models are derived from comprehensive literature reviews designed to identify risk factors that have been shown, in previously published research, to be statistically associated with the outcome of interest. Initially, Hilltop identified over 190 risk factors for the Pre-AH Model™ risk factor pool based on a literature review conducted in early 2019 (Pelser et al., 2019). Hilltop subsequently expanded the pool of Pre-AH Model™ risk factors in 2020; then, for each new predictive model, Hilltop conducted an additional literature review

[4] While the data for this clinical vignette are hypothetical, the Hilltop Pre-AH Model™ risk scores are, in actuality, even more skewed: the average probability of incurring a future hospitalization is roughly 0.6 percent, while the maximum probability in both the MDPCP and HealthChoice cohorts is greater than 99 percent.

[5] Liaw et al. (2015) conclude that, based on a review of four CMS-funded demonstrations involving care management fees, "to generate savings, resource allocation cannot be homogeneous. Instead, practices must focus more intensely on those at highest risk of utilization" (p. 557). Indeed, this may (partly) explain the varying effectiveness of care management, care coordination, and intensive primary care interventions as documented in the academic literature; many patients have low underlying risk of adverse outcomes, thus obviating the need for intervention, and the few high-risk patients may require significant intervention resources. For summaries of the literature on this subject, see Edwards et al. (2017) and Baker et al. (2018).

[6] See examples at Hong et al. (2014); McCarthy et al. (2015); and Anderson et al. (2015).

to identify risk factors relevant for that particular model (but which were not included in the Pre-AH risk factor pool).

Data Sources

MDPCP

The administrative claims for the risk factors in the MDPCP model are derived from the Claim and Claim Line Feed (CCLF) Medicare Parts A, B, and D claims files. Each month, Hilltop receives 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.[7] Additionally, Hilltop receives beneficiary attribution files and practice rosters each quarter.

Upon receipt of the monthly claims files, Hilltop first performs automated data validity checks in order to assess the integrity of the CCLF data files, followed by a data reduction step that subsets the claims files against the beneficiary attribution file. The resulting files retain the raw claims data that are inputs to the risk factor feature engineering process but discard the claims for individuals that are not in the MDPCP population.

In addition to risk factors based on administrative claims, the models also include risk factors based on publicly available, environmental risk factors. Appendix 2 details the data sources for these risk factors.

Maryland HealthChoice

The administrative claims for the risk factors in the HealthChoice model are derived from the Medicaid Management Information System (MMIS2) eligibility, recipient information, inpatient, outpatient, physician, and pharmacy claim files for the Medicaid beneficiaries. Each month, Hilltop receives new claim files from the Maryland Department of Health.

Upon receipt of the monthly claims files, Hilltop first performs automated data validity checks in order to assess the integrity of the MMIS2 data files, followed by a data reduction step that subsets the claims files against the MCO eligibility files. The resulting files retain the raw claims data that are inputs to the risk factor feature engineering process but discard the claims for individuals that are not in the HealthChoice population.

In addition to risk factors based on administrative claims, the models also include risk factors based on publicly available environmental risk factors. Appendix 2 details the data sources for these risk factors.

[7] For detailed documentation, see "Maryland Primary Care Program (MDPCP) CRISP Extract" (June 2019).

Condition-Based Risk Factors

A significant portion of Hilltop's risk factor pool is composed of condition-based risk factors: that is, 0/1 variables that indicate—based on an individual's claims history—whether they have been recorded as having diagnoses consistent with a given condition. These condition flags largely rely on diagnostic information from hospital, nursing home, physician, and lab claims in conjunction with Chronic Conditions Data Warehouse (CCW) coding specifications in order to generate beneficiary-level risk factors that represent underlying disease states.[8]

Update December 2023: Updated CCW condition list from the 27 CCW Conditions to 30 CCW Conditions in May 2023 for the MDPCP models and September 2023 for the HealthChoice models.

Utilization-Based Risk Factors

Risk factors from this category cover utilization of certain services (such as vaccinations, lab tests, or J-code procedures), place of service (for example, urgent care or rural health clinic), and provider specialty (for example, endocrinology or oncology). These risk factors also capture information on inpatient and outpatient hospital admissions, ED visits, and nursing home admissions over the past 12 months.

Prescription Drug-Related Risk Factors

Risk factors from this category index utilization of prescription drugs. The coding logic relies on first mapping drug names to National Drug Codes (NDCs) and then identifying those NDCs in pharmacy claims files. In order to capture compound drugs, which are drugs that contain multiple active ingredients, Hilltop relies largely on text-based, "contains"-type searches of the FDA's National Drug Code Directory to map drug names to NDCs.[9] We regularly update the list of NDCs to account for the addition of new NDCs. For the MDPCP population, the pharmacy claims used to create these risk factors are the Medicare Part D claim files. For the HealthChoice population, the pharmacy claims used to create these risk factors are the pharmacy claims and encounters in MMIS2.

Demographic Risk Factors

Risk factors from this category index cover beneficiary demographic characteristics such as age, race, and Medicare or Medicaid eligibility information. For both the MDPCP and HealthChoice populations, the respective beneficiary eligibility files are used to create these risk factors.

---

[8] Additional detail on the CCW condition flag specifications can be found here: https://www2.ccwdata.org/documents/10280/19139421/chr-chronic-condition-algorithms.pdf , https://www.ccwdata.org/documents/10280/19139421/ccw-chronic-condition-algorithms-reference-list.pdf

[9] For example, "Simcor" contains two active substances: Simvastatin and Niacin. This is flagged as a statin because one of its active ingredients is a statin. Source for the FDA NDC directory: https://www.fda.gov/drugs/drug-approvals-and-databases/national-drug-code-directory

Social and Environmental Risk Factors

Social and environmental variables play an important role in health; however, many individual-level demographic and socioeconomic factors are unavailable in administrative claims data (for example, marital status). Consequently, Hilltop developed an extensive database of area risk factors from publicly available data sources (i.e., the percentage of the population aged 15+ that is currently married) that can be linked to an individual's administrative claims using their recorded address to proxy for the unobserved individual-level variables. Other environmental risk factors (e.g., area poverty rate) are intended to capture social determinants of health—the neighborhood conditions in which people live and age that may affect health outcomes. Hilltop created two versions of these variables: one that maps to an individual's ZIP code (ZCTA), and, in October 2021, more granular versions of the variables at the census tract level. See Appendix 2 for more details on the risk factors and how they are linked to claims data.

General Methodology

Each of the Hilltop Pre- Models is a recurrent event modeled using time-dependent covariates. Accordingly, each of the Hilltop Pre- Models is operationalized as a discrete-time survival model that uses the *current* month of risk factors in order to predict avoidable hospitalization/ED visits in the *following* month. The model uses month as a time unit—instead of, for example, weeks or years—to balance the increased model accuracy obtained by a more granular time unit with the increased prediction error due to rare events.

The raw claims data span three years, or 36 person-months for individuals with full coverage. Since the model estimates the risk of incurring an outcome in the *next* month, however, it is not possible to use the most recently available month of risk data in the training model (since the next month's outcomes have not been realized at this point). Therefore, the training data are based on underlying data covering 35 person-months per attributed patient with full coverage. While, in general, a reduction in sample size can adversely impact the statistical precision of the risk factor estimates, lagged predictors are used for three reasons. First, several of the risk factors—such as the count of hospitalizations in the previous 12 months, or the condition flag for diabetes—overlap with the definition of the outcome variables. Consequently, including these risk factors as *contemporaneous* predictors could artificially increase the predictive power of the model. Second, Hilltop believes that using lagged predictors aids in the interpretability of the model. The goal of the Hilltop Pre- Models is to predict future events and using contemporaneous predictors to generate future risk scores requires the assumption that individuals' risk factors do not change in the future. Finally, the use of lagged predictors implies a natural "person-now" data set: the most recent month of risk factors, which is not included in the training data set.

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

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

- $\varphi(t)$ is a function of time at risk
- t is duration of time at risk in months
    - counting start from the first month of available data if the patient is in coverage longer than three years, or
    - counting from the coverage start month if the patient's coverage start is within three years
- β and Ω 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 a given outcome of patient i at time t (i.e., the month following the realization of the risk factors)

The statistical model uses six types of risk factors: diagnostic, pharmacy, procedural, utilization-based, geographic, and demographic. It is important to note that not all risk factors are available for every person-month. Hilltop uses a twelve-month lookback period for most of the time-varying risk factors, implying that an individual with, for example, five months of claims history will have incomplete information in her risk factors: if this individual truly has glaucoma, then it is possible that she will not amass the claims history by month five that meets the qualifications required for a glaucoma flag in our model. Furthermore, while most individuals in the data have addresses that link to the environmental risk factor data set, there are individuals for whom a valid census tract cannot be identified or who have ZIP codes for which there is no equivalent ZCTA, and therefore receive no environmental risk factors.[10] Table 2 presents the risk factor availability, depending on claims history and availability of area-level (ZCTA or census tract) data.

Table 2. Risk Factors by Data Availability

		At Least 12 Months of Claims History	
		Yes	**No**
Availability of Geographic Risk Factors	Yes	Dx/Rx/Proc/Util/Geo/Demo	Geo/Demo
	No	Dx/Rx/Proc/Util/Demo	Demo

Risk factor availability is an issue for the "scoring" step, in which risk scores are assigned to every individual based on the parameter estimates derived in the training step. For example, suppose that the vector of estimated coefficients from the logistic regression is as follows in Table 3.

[10] These individuals appear to use P.O. boxes as their mailing address, which, being point representations, do not have ZCTA areal equivalents.

Table 3. Risk Factor Availability Example 1

Risk Factor	Value for individual i
Asthma Flag	.1
…	
ZIP Code Income	-.00001
…	
Age	.02

These hypothetical risk factor coefficients indicate that, as expected, if an individual meets the clinical criteria for asthma, the risk of the outcome is higher; if he or she lives in a ZIP code with higher income, the risk is lower; and if he or she is older, the risk is higher. The scoring step will apply this vector of coefficients to the "person-now;" that is, the current month for each individual. Individual *i*'s predicted probability of incurring an outcome in the next month, then, will be scored as follows:

$$Risk_i = \frac{e^{.1*Asthma_i + \cdots - .00001*ZIP\ Code\ Income_i + \cdots + .02*Age_i}}{1 + e^{.1*Asthma_i + \cdots - .00001*ZIP\ Code\ Income_i + \cdots + .02*Age_i}}$$

Suppose that these variables (Asthma Flag, ZIP Code Income, and Age) are the only three risk factors in the model. Furthermore, suppose that individual *i* has the following characteristics:

Table 4. Risk Factor Availability Example 2

Risk Factor	Value for individual i
Asthma Flag	1
ZIP Code Income	$55,000
Age	66

This hypothetical individual has asthma, lives in a ZIP code in which the median income is $55,000 and is 72 years old. Then, that individual's risk of an outcome event in the following month is $\frac{e^{(.1*1 - .00001*55{,}000 + .02*66)}}{1 + e^{(.1*1 - .00001*55{,}000 + .02*66)}}$ = 70.47%.

Suppose, however, that this individual is newly eligible for Medicare and does not have sufficient claims history to meet the criteria for an asthma flag (anything under 12 months). In this instance, the individual might truly have asthma as an underlying disease state, but this is not observable. The individual's risk factors, then, are:

Table 5. Risk Factor Availability Example 3

Risk Factor	Value for individual i
Asthma Flag	NOT OBSERVED
ZIP Code Income	$55,000
Age	66

If the model's coefficients are applied only to the risk factors that are *observed*, then this individual's estimated risk is 68.35%. By failing to account for the risk factors that are not present in the model, the risk of incurring the outcome is underestimated for individual *i*.

Hilltop's solution to this issue is to estimate four different regression models for a given outcome based on the risk factors that are available for each group. This allows the risk factors that are present to "compensate," to a certain extent, for the risk factors that are missing due to data availability. For example, suppose that an individual lacks sufficient claims history to generate diagnostic risk factors but does have the following demographic risk factors: age, gender, and race. If gender is correlated with the unobserved diagnostic risk factors (if, for example, female beneficiaries are more likely to experience chronic conditions than male beneficiaries), then the coefficient for the "gender" risk factor will capture this correlation, and thus represent the marginal impact of being female *and* the portion of unobserved diagnostic risk factors that is correlated with gender. Consequently, if female beneficiaries are more likely to experience chronic conditions than male beneficiaries, then the risk factor coefficient for "gender" will be larger in the models without diagnostic risk factors than in the models with diagnostic risk factors. By allowing observed risk factors to capture some of the predictive power of unobserved risk factors, the loss in predictive power due to missing data is minimized. Note that this method is analogous to that used in the CMS HCC Risk Adjustment Model (Centers for Medicare and Medicaid Services, 2018, p. 80).

The four models are trained on the subset of person-months for which all risk factors are complete (that is, person-months with at least 12 months of claims history and a valid geographic linkage), and include the following sets of risk factors (analogous to the four partitions of the person-month sample):

- Model 1: use Rx/Dx/Util/Proc/Geo/Dem risk factors
- Model 2: use Geo/Dem risk factors
- Model 3: use Rx/Dx/Util/Proc/Dem risk factors
- Model 4: use Dem risk factors

Variable selection can improve the performance of predictive models by reducing prediction variance and increasing generalizability (Bagherzadeh-Khiabani et al., 2016; Walter & Tiemeier, 2009). Hilltop performed this in two steps: first, the team selected initial risk factors for the Pre-AH Model™ based on an extensive literature review, which screened over 3,300 articles and ultimately selected 211 published, peer-reviewed papers from which to extract risk factors. This generated a pool of roughly 190 risk factors; each of the additional Pre- Models is based on its own literature which adds risk factors to this baseline pool. Additionally, Hilltop used stepwise selection in the multivariable logistic model to remove insignificant predictors from the model before adding significant predictors.

In the current version of the Hilltop Pre- Models, the risk factors typically enter the model additively: that is, if an individual has both diabetes and heart failure diagnostic flags, then his or

her risk score will reflect the risk coefficient for the diabetes flag, plus that of the heart failure flag. It is possible, however, that there is additional risk due to the fact of the beneficiary having *both* conditions, over and above the sum of the risks of having each condition. We have included such "interaction terms" where indicated by the literature reviews (for example, in the Pre-HE Model, we include a measure of frailty, a 0/1 variable indicating a history of Alzheimer's disease or related dementia, and the interaction of the two).

Hilltop trains each of the Pre- Models on a quarterly basis unless otherwise specified. We will, however, monitor the predictive accuracy of the model and adjust the training schedule as needed.

Validation Model

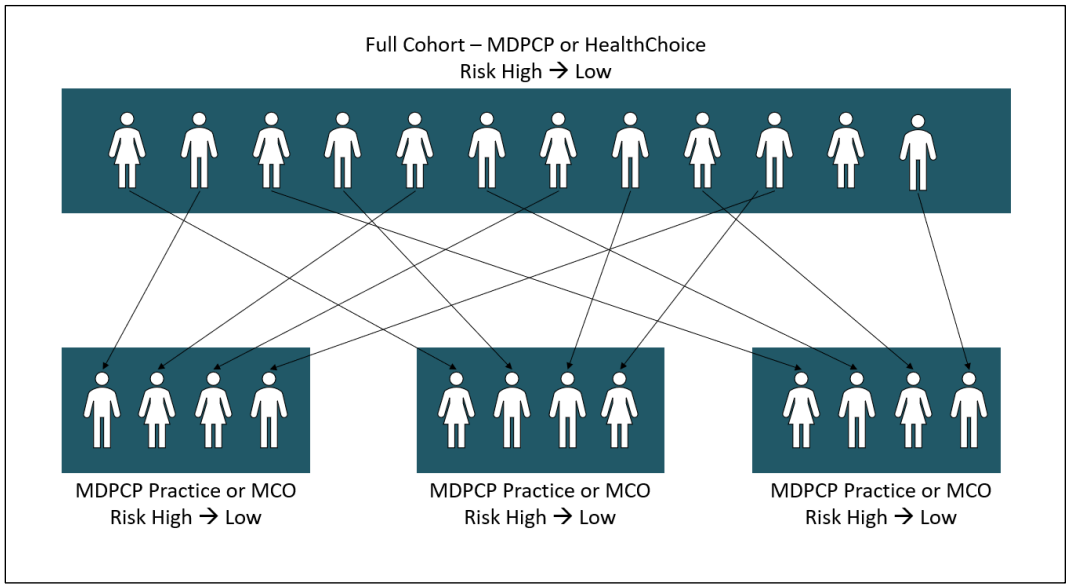
As an additional quality check, we regularly create validation version of the Pre- Models for both the MDPCP and HealthChoice populations using Stata (v16.1). This is to ensure that the risk scores are not tied to, or influenced by, a particular statistical software and to ensure there are no errors in the code used to create the risk factors and outcome variable or estimate the model itself. We monitor the correlation between the production and validation risk scores, as well as the overlap in patients who have the highest risk scores.

Scoring

The four risk models above are trained on the subset of data with at least twelve months of claims history and full environmental data (ZIP code or census tract) data to estimate the vectors of coefficients for the risk factors in each model. Then, using the most recently available month of risk factors (that is, the "person-now" data set), individuals are scored using the model coefficients that correspond to the risk factors available in the person-now data set.

The Hilltop Pre- Models generate risk scores for the entire MDPCP or HealthChoice cohort, but individual practices or MCOs will only receive risk scores for their specific beneficiaries. This has the consequence that, if a practice or MCO contains disproportionately high-risk patients, and another contains disproportionately low-risk patients, then the riskiest patients within each will differ in their *absolute* risk. Figure 1 presents a diagram of this point.

Figure 1. Stratification of Full Pre-AH Cohort by Practices/MCOs



Hilltop scores the Pre- Models monthly. During this process, we create risk factors from raw claims data for the most recent one month of claims history and apply the most recent model coefficients to create risk scores. Patient-level risk scores, risk percentiles, and reasons for risk for the HealthChoice population are deployed via secure file transfer to each MCO and for the MDPCP population are deployed via CRISP.

Note: Due to the dramatic impact of the COVID-19 pandemic on health care utilization, Hilltop temporarily transitioned to training the Pre-AH Model™ on a semi-monthly cadence starting in Summer 2020 to account for altered utilization patterns of health services. Hilltop reverted to its quarterly training cadence in May 2022.

Predictive Performance Metrics

Predicted Probabilities

The output for the all the Pre- Models is a set of probabilities that estimate the patient-specific risk of incurring the model-specific outcome. In general, these events are rare and, consequently, the predicted probabilities are low. Hilltop does not interpret this as a limitation of the risk scores; rather, this reflects the relative rarity of the outcome events. Moreover, the *relative* risk is the key metric that should be used to allocate care resources: no matter the absolute risk of the patient panel, the efficient allocation of care resources requires the identification (and treatment) of the riskiest patients.

Patient-level risk tends to persist across time: that is, high-risk patients tend to remain high-risk from one month to the next, and low-risk patients tend to remain low-risk. This is likely due to two factors. First, to prevent coding idiosyncrasies from introducing noise into the predictions, the majority of risk factors are coded with at least one year of lookback. This has the consequence of making the Pre- Model risk factors relatively stable over time, and 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 in risk scores can occur 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 July 2021, then her risk score will likely increase significantly because of that underlying change. Second, Hilltop estimates new risk factor coefficients every quarter (in the model re-training step). As a result, not only can the underlying risk factors for a given patient change from one month to the next, the *relationship* between that risk factor and the model specific outcome events can also change upon retraining. To continue the previous example, if the risk factor coefficient for heart failure rises after the model is re-trained, the individual's risk score would rise not only because she has a new heart failure risk factor, but also because the heart failure risk factor has risen in predictive importance.

We present the predicted probabilities separately for each model and population (if applicable).

Predictive Power

It is imperative that the accuracy of predictive models be assessed during both model development using holdout data, and in a production environment once the scores have been deployed. "Holdout data" are data that are available at the time of model training *but not used to train the models*; the Pre- Models reserve 20% of all data to use as holdout data for purposes of model assessment. Testing model performance on holdout data constitutes assessing the predictive performance on the model on data that is new to the model (although which is technically available at the time of model training). Assessing model performance in a production environment, however, means that we check the accuracy of scores that were released to MDPCP providers against events that *actually occurred* in the following month. Since this requires knowledge of the "true" events in the month after a given score release, this is only possible several months following the release of a given month of risk scores.

Typically, the discriminatory power of predictive models is summarized using the c-statistic, which is a measure of the area under the Receiver Operating Characteristic (ROC) curve (Steyerberg et al., 2010). 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 non-diseased subject" (Mauguen & Begg, 2016). However, this measure is uninformative regarding model calibration, which is the degree to which estimated risk scores match underlying true risk: it is possible to have a model with good discrimination and poor calibration (Alba et al., 2017). Moreover, the objective of the Hilltop Pre- Models is not binary classification, but instead the estimation of individual-level risks of incurring the model specific outcome event so that care

managers can, by focusing on the riskiest individuals, potentially intervene. To that end, the performance of the Hilltop Pre- Models is assessed using the *concentration curve*.[11]

This measure of model accuracy estimates the cumulative share of all model-specific outcome events incurred by the riskiest patients, where the reader can determine the share of all outcome events occurring for individuals above different risk thresholds. 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 outcome events captured by the riskiest patients on the Y axis. We estimate the percent of outcome events incurred by the top 1% and 10% riskiest patients.

Concentration curves can be summarized by a Gini coefficient, a measure of 0 to 1, that can be interpreted as a measure of risk concentration in the population: the greater the Gini index, the more concentrated is the risk of the model-specific outcome event in a small proportion of persons (Llorca & Delgado-Rodríguez, 2002). A higher Gini coefficient indicates better model fit. To assess whether model performance is improving or declining over time, we estimate monthly concentration curves for the 20% holdout sample of the training data set.

Where possible, we assess predictive power on both holdout and production samples. We present the predictive performance metrics separately for each model and population (if applicable) for each model.

Reason for Risk

As of January 11, 2020, the Hilltop Pre- Models have—in addition to generating individual-level risk scores—also displayed the top actionable risk factors underlying each patient's risk of incurring a future model-specific outcome. The intention of this update was to augment the information provided to practices in order to further facilitate patient-specific advanced primary care. For example, in addition to a risk score of 3.2% for a particular patient, care managers will also be able to see that the patient (for example) meets the clinical criteria for diabetes and heart failure and incurred a claim for insulin within the past year (in descending order of contribution to risk). While that patient may also have had other salient risk factors—for example, meeting the clinical criteria for depression—Hilltop only displays the most predictive, intervene-able risk factors in order to allow care mangers to focus their attention on the most pressing patient needs.[12]

These reasons for risk are based on the underlying risk factor coefficients, which are derived from the training phase of the model. It is important to note that these coefficients do not necessarily have a *causal* interpretation: they only capture the strength of *association* between a

[11] This is very similar to the Lorenz curve, which "is especially useful in the context of disease prevention because it maps out what public health policy investigators need to know. That is, it tells us how much disease burden will occur in any given proportion of the population with risks above a chosen threshold" (Mauguen & Begg, 2016).
[12] Hilltop collected stakeholder feedback from clinical partners in order to ensure that we only displayed those risk factors over which patients, providers, and care managers can exert some control. We did not, for example, include most environmental risk factors, since providers cannot directly assist patients with the management of this factor.

given risk factor and the risk of incurring a future outcome. For example, if the risk factor coefficient for diabetes is positive in a particular model, then that could mean that having diabetes *causes* an increased risk of that model's outcome; however, it could also mean that having diabetes is only *correlated* with some unobserved factor that causes an increased risk of that model's outcome. While these risk factors do not have a strictly causal interpretation, they are intended to provide care managers with a useful starting point from which to address specific patient needs.

In order to operationalize the identification of reasons for risk, the Hilltop team first re-coded select risk factors so that a higher level of a given risk factor is theoretically associated with greater risk of incurring the outcome event for each model. Consider the example of flu vaccinations: there is evidence that influenza and/or pneumococcal vaccinations reduce the risk of hospitalization for various prevention quality indicators (PQIs) in various populations (Furumoto et al., 2008; Hedlund et al., 2003; Nichol et al., 2003). This implies that receipt of a flu vaccination should be *negatively* associated with the risk of incurring an avoidable hospital event. This risk factor, then, was re-coded to be 1 if the individual has *not* received a flu vaccination, and 0 if the individual *has* received a flu vaccination.

The Hilltop research team used two criteria for determining which risk factors to recode. First, we reviewed the existing evidence for the sign and magnitude of risk factors based on the foundational Pre-AH Model™ literature review (Pelser et al., 2019). If there was strong *a priori* empirical evidence that certain risk factors—again, like having had a flu vaccination—have a negative association with the risk of incurring an avoidable hospitalization, then the variable was re-coded accordingly. Second, if the literature review indicated that the impact of a given risk factor on the risk of incurring an avoidable hospital event was ambiguous, then the Andersen Behavioral Model of health services utilization was applied to guide the re-coding logic. The Andersen model posits that health services utilization is a function of predisposing, enabling, and need factors (Andersen, 1995). Consensus was required among members of the Hilltop research team before recoding using this criterion.

While the baseline model contains approximately 200 risk factors, only a subset of these is included in the pool of potential reasons for risk for reasons of statistical interpretation and clinical utility. Most non-binary and non-count risk factors are excluded because these cannot easily be translated into reason for risk contributions for lack of a meaningful reference group. Additionally, based on the feedback from stakeholders, Hilltop excludes risk factors that are not potentially modifiable; that is, for which the effects cannot be meaningfully modified by clinical intervention (e.g., area income). Finally, risk factors that are not positive and statistically significant are also excluded.

Consider the following illustrative example. Suppose that the Pre-AH Model contains only three risk factors: a flag for diabetes, the number of recent avoidable hospitalizations, and a flag for heart failure. In this example, the coefficients for these three risk factors are 0.1, 0.08, and 0.07, respectively. The coefficient for diabetes represents the increase in risk of avoidable hospitalizations associated with having diabetes (relative to not having diabetes), holding all

other factors constant.[13] The coefficient for the number of avoidable hospitalizations reflects the added risk associated with one additional previous avoidable hospitalization, and the coefficient for heart failure reflects the added risk associated with having heart failure (relative to not having heart failure), again holding all other factors constant.

It is important to note that these risk coefficients are marginal effects; that is, the *additional* risk due to, for example, a patient having one additional previous avoidable hospitalization. In order to translate these marginal effects to reason for risk contributions, Hilltop multiplies each marginal estimate by the *level* of that risk factor for each individual. Thus, if an individual has four previous avoidable hospitalizations, then the risk contribution of avoidable hospitalizations is 4*0.08 = 0.32.[14] Crucially, this risk contribution is still interpreted relative to a reference category: in this case, individuals with no history of avoidable hospitalizations. More broadly, the risk contribution should be interpreted relative to individuals *without* that particular risk factor.[15]

Suppose that, in this example, there are four patients in the MDPCP program. Patient 1 has diabetes, no history of avoidable hospitalization, and heart failure. Patient 2 does not have diabetes, has no history of avoidable hospitalization, and has heart failure. Patient 3 has diabetes, four prior avoidable hospitalizations, and does not have heart failure. Finally, patient 4 does not have diabetes, has one previous avoidable hospitalization, and has heart failure. This information is presented in Table 6, below.

Table 6. Hypothetical Reason for Risk Example

Patient ID	Diabetes	Diabetes * Coefficient	# AH	# AH * Coefficient	Heart Failure	Heart Failure * Coefficient
1	1	0.1	0	0.0	1	0.07
2	0	0.0	0	0.0	1	0.07
3	1	0.1	4	0.32	0	0.0
4	0	0.0	1	0.08	1	0.07

In this example, the top reason for risk for Patient 1 is diabetes: this risk factor yields the largest positive contribution (risk factor level * coefficient) among all the risk factors for that individual. For Patient 2, the top reason for risk is heart failure; for patient 3, the top reason for risk is the history of avoidable hospitalizations; and for patient 4, the top reason for risk is the history of avoidable hospitalizations. The second reason for risk is calculated analogously: it is the second highest contribution of (risk factor level * coefficient) for each individual. All other reasons for risk are estimated in a similar fashion.

[13] Since our baseline model is a multivariate logistic regression, the coefficient is technically the marginal impact on log odds of incurring an avoidable hospital event. For the sake of exposition, we label this as "risk."

[14] This assumes that marginal effect is constant across units: that is, that the effect neither grows, nor shrinks, as the level of the risk factor rises. Since the vast majority of the reason for risk factors are binary variables, for which this assumption does not bind, we believe that this is a reasonable simplification.

[15] This motivates the exclusion of continuous (that is, non-binary and non-count) risk factors from the reason for risk pool: there is no natural reference group for these risk factors. For example, there is no meaningful group of people that do not have the "age" risk factor.

Users can also see the contribution of each risk factor category (Condition, Demographic, Pharmacy, Utilization, and Environmental) in percentage terms for the risk models. These are intended to provide a high-level description of the contribution of various types of risk factors that are positive and significant for an individual. The contribution for a given category is calculated as the sum of (risk factor level * coefficient) for all reasons for risk in that category, divided by the sum of (risk factor level * coefficient) for all positive, statistically significant reasons for risk. This is an important point: an individual's *overall* risk is a function of *all* risk factors, including those that are not included as potential reasons for risk. The category contributions, however, are only interpretable relative to the reason for risk factor pool, which is restricted to the operationalizable, modifiable risk factors.[16]

Return to Table of Contents

[16] If an individual has 3.2% overall risk and the Condition category contribution is 50%, then it is not appropriate to conclude that 50% of that individual's risk is due to Condition risk factors. Instead, it is appropriate to conclude that, of the positive, statistically significant, operationalizable, modifiable risk factors for that individual, conditions represent 50% of the total (risk factor level * coefficient).

Section 3. Pre- Models Operations and Performance

Pre-AH: Model Operations

The Hilltop Pre-AH Model™ is a risk prediction model that uses a variety of risk factors derived from administrative claims and publicly available social and environmental data to estimate the probability that a given patient incurs an avoidable hospital event in the following month. It was initially developed by The Hilltop Institute, in conjunction with the Maryland Department of Health, to support the care management efforts of primary care providers enrolled in MDPCP. Given the MDPCP's emphasis on the reduction of unneeded utilization, the Hilltop Pre-AH Model™ focuses on *potentially avoidable* hospitalization or ED visits.[17] These events, by definition, are more likely to be prevented with targeted, outpatient care efforts than all-cause, general hospitalizations and ED visits.

The Hilltop Pre-AH Model™ risk scores were first deployed for the MDPCP population in October 2019. Patient-level risk scores and reasons for risk are provided to participating practices monthly for their attributed beneficiaries via the MDPCP Prediction Tools area on CRISP. These risk scores were originally referred to as the "Likelihood of Avoidable Hospital Event" (LAH) scores; at the time of this writing, they are known as the "Avoidable Hospital Events (Pre-AH)" scores.

Beginning in May 2021, a second version of the Hilltop Pre-AH Model™ was deployed for the MCOs that are part of the Maryland Medicaid HealthChoice program. This version of the model uses the same risk factors but is trained and scored for Medicaid recipients enrolled in the HealthChoice program. These patient-level risk scores are provided to MCOs monthly for their enrollees via secure file transfer.

Data Sources

The Hilltop Pre-AH Model™ relies largely on data from administrative claims data, supplemented with various publicly available environmental data sets used to generate the environmental risk factors. The Pre-AH model is estimated for both the MDPCP and the Maryland Health Choice populations. For more detail about the sources for administrative claims used in each population, see the sub-section above labeled "Data Sources" in Section 2.

Using SAS 9.4, Hilltop creates the model using risk factors identified in the literature review.[18] Hilltop also created a validation version model using Stata 16.1 as an additional quality control

[17] 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" (Billings et al., 1993). This measure is discussed in greater detail in Section 3.2.1.

[18] Certain risk factors identified in the literature review were not ultimately operationalizable in Phase 1 of the Hilltop Pre-AH Model™. We will incorporate additional risk factors in future iterations of the model.

check and to ensure that the risk scores are not dependent on a single statistical package. Appendix 1 describes the risk factors for the Hilltop Pre-AH Model™ in greater detail.

Risk Factors

Literature Review

Based on a comprehensive literature review, Hilltop identified and operationalized approximately 190 risk factors to be included in the risk model (Pelser et al., 2019). Hilltop added several more risk factors in June 2020, and this overall pool of risk factors forms the foundation of the Hilltop Pre- Models. While some of these risk factors are eliminated in the variable selection step, all risk factors are included in the pool of *potential* risk factors to be used in the model. A high-level description of risk factors is provided in the sections below. For a description of each risk factor, along with data source and sample statistics, see Appendix 1.

Condition-Based Risk Factors

A significant portion of the risk factor pool is composed of condition-based risk factors: that is, 0/1 variables that indicate, based on an individual's claims history, whether they have been recorded as having diagnoses consistent with a given condition. These condition flags largely rely on diagnostic information from hospital, nursing home, physician, and lab claims in conjunction with CCW coding specifications in order to generate beneficiary-level risk factors that represent underlying disease states.[19]

Utilization-Based Risk Factors

These risk factors describe utilization of certain services (such as vaccinations, lab tests, or J-code procedures), place of services (for example, urgent care or rural health clinic), and provider specialty (for example, endocrinology or oncology). Hilltop also created risk factors to capture a beneficiary's primary care utilization and continuity of care.

Prescription Drug-Related Risk Factors

Prescription drug-related risk factors index utilization of drugs identified in the literature review as potential risk factors for potentially avoidable hospital events.

Beneficiary Demographics-Related Risk Factors

Information from the beneficiary demographics files, such as date of birth, race, and sex, are used to create potential risk factors for avoidable hospital events. Additionally, Hilltop geocodes

[19] Additional detail on the CCW condition flag specifications can be found here: https://www2.ccwdata.org/documents/10280/19139421/chr-chronic-condition-algorithms.pdf, https://www.ccwdata.org/documents/10280/19139421/ccw-chronic-condition-algorithms-reference-list.pdf

the address listed in beneficiary demographic data to allow the social and environmental risk factors to be linked with a person's administrative claims.

Environmental Risk Factors

Several of the risk factors Hilltop identified during the literature review were individual-level demographic and socioeconomic factors that are unavailable in administrative claims data (for example, marital status). Additionally, the literature review identified several social determinants of health factors that increase an individual's risk for avoidable hospitalizations. These risk factors were created using publicly available data sets, such as the American Community Survey (ACS), CMS Provider data, and others for each ZIP code tabulation area (ZCTA) and census tract in the United States (see Appendix 2 for more detail).

Risk Factor Updates

As part of the ongoing development process, Hilltop makes improvements or additions to the pool of risk factors.

- June 2020: Hilltop added eight new risk factors to the model: an indicator for frailty; an indicator for original Medicare eligibility due to a non-age-related reason; an indicator for DME use within the past year; the number of ED visits in the past six months; an indicator for sickle cell anemia; area-level pollution level; area-level walkability; and area-level pharmacy density.
- October 2021: Hilltop developed an automated geocoding pipeline to identify each beneficiary's census block of residence where possible. This allowed us to use more granular versions of the environmental risk factors (census tract-level) that are posited to more accurately describe an individual's' proximal environment. The census tract versions of the variables are currently only used for MDPCP population (see Appendix 2 for more detail).

Outcome: Avoidable Hospitalizations and ED Visits

The outcome measure in the Hilltop Pre-AH Model™ is a 0/1 indicator variable denoting whether an individual incurred an avoidable hospitalization or ED visit in a given month. To construct this measure, Hilltop relies on technical definitions provided by the Agency for Healthcare Research and Quality (AHRQ) as part of its PQI measures.[20] Diagnosis codes from administrative claims are used to flag the following conditions, which are the basis for the composite outcome variable:[21]

- PQI #1: Diabetes Short-Term Complications
- PQI #3: Diabetes Long-Term Complications

[20] For more information, see https://www.qualityindicators.ahrq.gov/modules/pqi_resources.aspx.
[21] Specifically, Hilltop defines these claims as those with a claim type of either 60 or 61 (indicating an inpatient claim) or a claim type of 40 (indicating an outpatient claim) and revenue center codes of 0450-0459 and 0981. Source: https://www.resdac.org/articles/how-identify-hospital-claims-emergency-room-visits-medicare-claims-data.

- PQI #5: COPD or Asthma in Older Adults
- PQI #7: Hypertension
- PQI #8: Heart Failure
- PQI #11: Bacterial Pneumonia
- PQI #12: Urinary Tract Infection
- PQI #14: Uncontrolled diabetes
- PQI #15: Asthma in Younger Adults
- PQI #16: Lower-Extremity Amputation among Patients with Diabetes

This is implemented in the model as an indicator variable at the person-month level. If an individual incurs at least one avoidable hospitalization or ED visit in a given month, then that person receives a value of 1 for this variable—and 0 otherwise.

Update August 2021: Hilltop updated the avoidable hospitalization event definition to match the 2020 PQI definition from the AHRQ. Previously, we used the 2018 version, which had included PQI#10: Dehydration.

Update July 2022: Hilltop updated the avoidable hospitalization event definition to match the 2021 PQI definition from the AHRQ. This version has the same PQI indicators; however, changes were made to the exclusion diagnosis codes for the Bacterial/Community Pneumonia and Urinary Tract Infection indicators.

Update December 2023: Hilltop updated the avoidable hospitalization event definition to match the 2022 PQI definition from the AHRQ. This version has the same PQI indicators; however, changes were made to the following indicators: Hypertension, Heart Failure, Community Acquired Pneumonia, and Urinary Tract Infection. We made this update in March 2023 for the MDPCP models and in September 2023 for the HealthChoice models.

Pre-AH: MDPCP Model Performance

This section presents the details of the Hilltop Pre-AH Model™ performance in the MDPCP population.

Differentiation from CMS HCC Risk Scores

It is important to note that the Hilltop Preventive Predictive Model risk scores are conceptually distinct from the CMS Hierarchical Condition Category (HCC) risk scores that are currently presented in CRISP. The Hilltop risk scores use risk factors based on diagnoses, procedures, medications, utilization, demographics, and geographic factors in order to produce a practice-specific ranking of patient risk in the near future. The CMS HCC risk scores are based on a model that uses diagnosis codes and a limited set of demographic information from a base year in order to predict *expenditures* over the following year. There is likely to be some overlap among

individuals who incur potentially preventable utilization and individuals who experience high medical spending, but the overlap is unlikely to be complete.[22] High medical expenditure can reflect multiple factors ranging from moderate utilization of high-cost procedures, high utilization of moderate-cost procedures, underlying morbidity, or geographic differences in treatment or referral practices.

Moreover, the theoretical interpretation of each risk score differs substantially. The CMS HCC risk score was developed as a capitated payment risk adjustment methodology for Medicare Advantage participants in order to "address [the] issue of risk selection and to compensate Medicare Advantage health plans for accepting the risk of enrolling beneficiaries of varying health statuses" (Centers for Medicare and Medicaid Services, 2018, pp. 9–10). Additionally, "the underlying risk assessment is designed to accurately explain the variation at the group level, not at the individual level, because risk adjustment is applied to large groups" (Centers for Medicare and Medicaid Services, 2018, pp. 9–10). Note that "risk" for the CMS HCC risk model refers to *actuarial* risk: this model seeks to predict average expenditures over large groups of individuals. In contrast, the Hilltop risk scores are designed to estimate, as closely as possible, event risk: for example, an *individual's* risk of an avoidable hospital event in the following month.

There are also differences in the time horizons of each risk score. CMS HCC "final risk scores are generally available 16-18 months after the close of the base year. For example, 2017 risk scores (based on 2016 diagnoses) will be available in the spring of 2018"(Center for Medicare and Medicaid Innovation, 2017, p. 26). The Hilltop risk scores, however, are updated monthly and use patient-level risk factor information current to the most recently available month of claims in order to generate risk scores. This is a strength of the Hilltop models because these risk scores reflect the underlying patient condition with a lag of only, at most, three months.[23] Finally, by definition, avoidable hospital events are preventable through timely primary care and so, in principle, the identification and management of individuals at high risk of incurring potentially preventable health care utilization may result in the avoidance of that particular utilization event. High medical expenditures, however, may reflect underlying morbidities that would necessitate utilization *regardless* of primary care intervention.

Coefficients

Table 7 presents risk factor coefficient estimates for Model 1 for the training performed in September 2023. Model 1 includes all six types of risk factors: diagnostic, pharmacy, procedural, utilization-based, geographic, and demographic. The risk factors in this table are those that were included in the final model. All other risk factors were eliminated in the variable selection step due to insufficient predictive power. Note that the risk factor coefficients are presented as odds

[22] Internal testing shows a limited degree of substitutability between the two sets of risk scores. Specifically, we find that the Hilltop Pre-AH Model™ outperforms the CMS HCC risk score in predicting avoidable hospitalization in the following month. Both concentration curves are presented below.

[23] This lag is related to the unavoidable delay in obtaining and processing administrative claims data. For example, claims data delivered to Hilltop in late October 2021 reflect utilization through mid-September 2021. We discuss this point further in the "Limitations" section below.

ratios. Odds ratios can be interpreted in terms of a multiplicative impact: for example, an odds ratio of 1.05 indicates that if that risk factor were to increase by one unit, then the risk of incurring an avoidable hospitalization would increase by 5%.

Table 7. Pre-AH MDPCP Risk Model Odds Ratios for Model 1

Risk Factor	Odds Ratio
Prior hospitalization discharge status - other	1.985
CCW indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis	1.55
CCW indicator for heart failure and non-ischemic heart disease	1.532
Beneficiary race - Black	1.442
CCW indicator for intellectual disabilities and related conditions	1.433
Number of avoidable hospitalizations	1.409
CCW indicator for hypertension	1.386
Indicator for original Medicare eligibility for a non-age-related cause	1.358
Prior hospitalization admission type - emergency	1.355
Beneficiary race - Hispanic	1.346
Indicator for retinopathy	1.324
Indicator for urinary tract infection	1.319
Indicator for previous conservative diabetic wound procedure	1.308
Indicator for insulin use	1.282
CCW indicator for tobacco use	1.274
Indicator for hospice enrollment	1.272
Indicator for durable medical equipment (DME) use	1.265
Prior hospitalization discharge status - home	1.222
Indicator for problems with care provider dependency	1.205
Indicator for gastroparesis	1.203
Indicator for arrhythmia	1.198
CCW indicator for lung cancer	1.186
Indicator for fluid and electrolyte imbalance	1.155
CCW indicator for chronic kidney disease	1.154
Indicator for oral corticosteroid use	1.152
Indicator for dual eligibility with Medicaid	1.152
Indicator for frailty	1.148
CCW indicator for diabetes	1.148
Prior hospitalization admission type - urgent	1.145
Beneficiary race - White	1.138
Indicator for diabetes with complications	1.138
Indicator for albuminuria	1.134
CCW indicator for asthma	1.125
CCW indicator for ischemic heart disease	1.118
CCW indicator for Parkinson's Disease or Secondary Parkinsonism	1.118
CCW indicator for atrial fibrillation and flutter	1.106

Risk Factor	Odds Ratio
Number of emergency department visits within the past 6 months	1.101
Indicator for oral antibiotic use	1.1
Indicator for respiratory infection	1.096
Beneficiary gender - female	1.083
CCW indicator for pressure and chronic ulcers	1.082
CCW indicator for peripheral vascular disease	1.082
Indicator for cerebrovascular disease	1.078
CCW indicator for anxiety disorders	1.078
Indicator for beta blocker use	1.075
Indicator for pulmonary circulatory disorder	1.074
CCW indicator for pneumonia, all-cause	1.069
CCW indicator for anemia	1.065
Indicator for provider administered drug	1.063
Indicator for neuropathy	1.062
Indicator for no vaccination (flu or pneumonia)	1.058
Number of urgent care visits	1.051
Rurality index	1.029
Age	1.022
Percent with less than high school education, ages 65+	1.003
Number of outpatient visits	1.002
Median household income	1
Total health spending	1
Continuity of primary care - Duration	.998
Percent foreign born	.997
Number of primary care visits	.997
Number of prior admissions	.979
CCW indicator for glaucoma	.952
CCW indicator for hyperlipidemia	.948
CCW indicator for obesity	.945
CCW indicator for osteoporosis with or without pathological fracture	.945
Indicator for sepsis	.941
CCW indicator for rheumatoid arthritis/osteoarthritis	.926
Indicator for rivaroxaban use	.919
Indicator anti-diabetes medication use	.91
Indicator for prior surgery	.888
CCW indicator for cataracts	.877
CCW indicator for viral hepatitis	.867
CCW indicator for hip/pelvic fracture	.862
Indicator for protein-calorie malnutrition	.857
Indicator for statin use	.828
CCW indicator for alcohol use disorders	.783

Risk factor coefficient estimates will change as the model is re-trained. Risk factor coefficients for other models are available upon request.

Predicted Probabilities

The outcome for the all the Pre- Models is a set of probabilities that estimate the patient-specific risk of incurring the model-specific outcome. In general, these events are rare and, consequently, the predicted probabilities are low. Table 8 presents summary statistics of a recent month of Pre-AH Model™ risk scores in the MDPCP patient population.

Table 8. Summary Statistics for Pre-AH Scores in MDPCP Population

Score Date	Model Date	Cohort Size	Events per Month	Average Score	N > 1% Risk	Monthly Correlation
Oct 2022	July 2022	361,192	2512	.0042	25,726	0.966

Hilltop does not interpret this as a limitation of the risk scores; rather, this reflects the relative rarity of the outcome events. Moreover, the *relative* risk is the key metric that should be used to allocate care resources: no matter the absolute risk of the patient panel, the efficient allocation of care resources requires the identification (and treatment) of the riskiest patients.

Predictive Power

Holdout Data Testing

Figure 2 on the following page shows how we use the Pre-AH Model™ scores to estimate a concentration curve (teal line) for holdout data from May 2022. We find that the top 10% riskiest patients account for approximately 53% of all avoidable hospital events in the following month, and the top 20% riskiest patients account for over two-thirds of all avoidable hospitalizations. We performed the same exercise using the CMS HCC risk score for the July 2022 MDPCP attributed beneficiary cohort. This is represented by the orange line in the figure. The top 10% riskiest patients account for approximately 33% of all avoidable hospitalizations in the following month, and the top 20% riskiest patients accounted for approximately 50% of all avoidable hospitalizations. Given a baseline of approximately 2,500 avoidable hospital events per month, this implies that if care managers were to rely solely on the CMS HCC risk score and focus on the riskiest 10% of the cohort, then they would fail to identify 350 avoidable hospital events (relative to the number that would be identified using the Hilltop Pre-AH Model™ risk scores).

Figure 2. Pre-AH MDPCP Concentration Curves as of May 2022

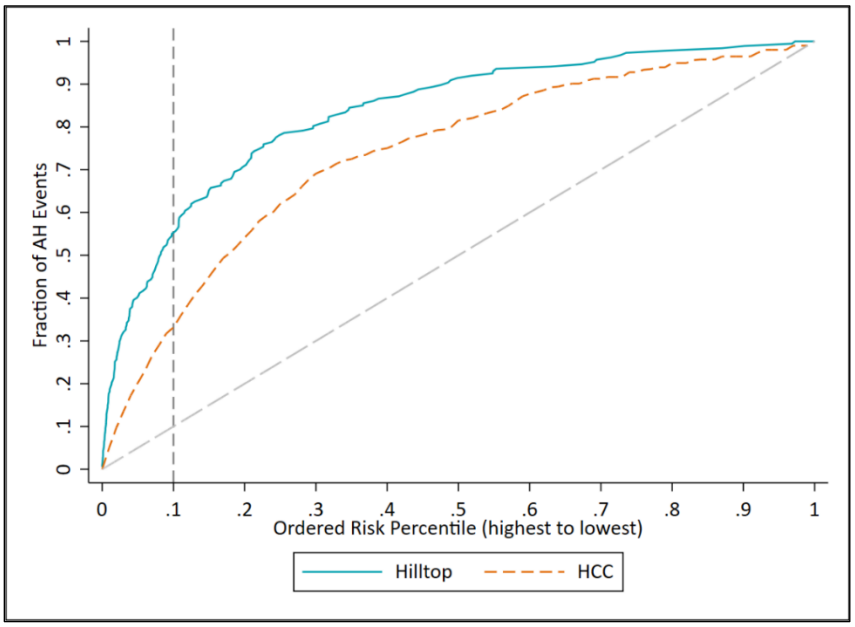
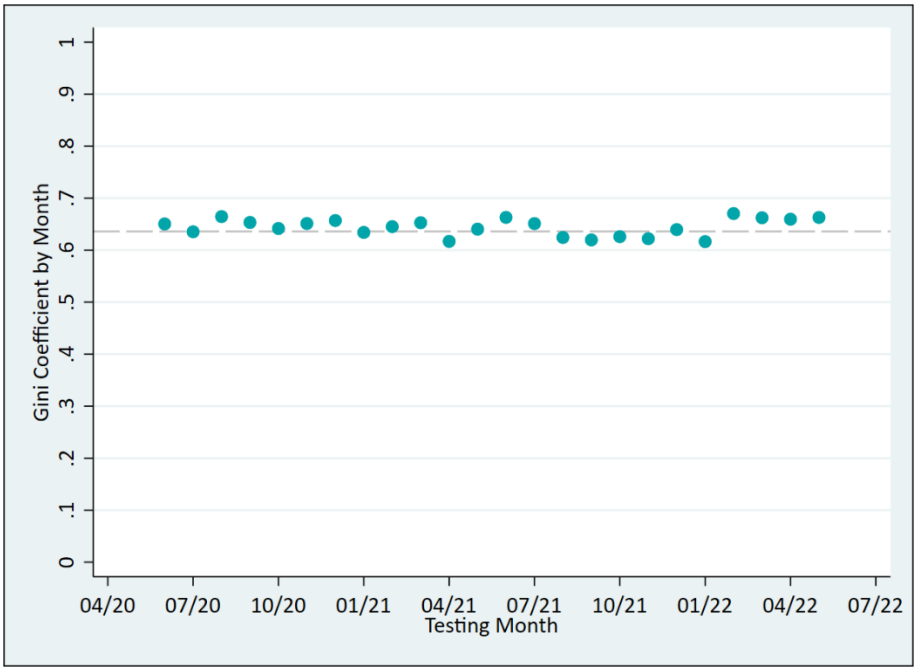


Figure 3 plots the Gini coefficients from 23 months based on the concentration curves estimated on the 20% holdout sample from each month.

Figure 3. Pre-AH MDPCP Gini Scores by Month



These scores indicate that model performance is steady over time.

Production Data Testing

Post-deployment model evaluation is a crucial component of the predictive model lifecycle. The first Pre-AH Model™ risk scores were released to participating providers on October 11, 2019, and have since been updated monthly. Hilltop monitors the accuracy of the Pre-AH Model™ predictions in a production environment by comparing the risk scores released in a given month with the *true outcomes* that occur in the following month.

For the scores released between December 2021 and May 2022, we find that the top 10% riskiest patients accounted for between approximately 46% and 50% of all avoidable hospital events over the following month (until the next score release). See Table 9.

Table 9. Predictive Performance of MDPCP Pre-AH Model™ Scores by Month

Score Month	Evaluation Month	Top 1% of Patients, by Risk	Top 10% of Patients, by Risk
December 2021	January 2022	13.5%	47.7%
January 2022	February 2022	12.8%	49.3%
February 2022	March 2022	11.3%	46.3%
March 2022	April 2022	11.6%	47.7%
April 2022	May 2022	12.9%	49.7%
May 2022	June 2022	13.4%	46.4%

We will continue to monitor model field performance as additional data updates are received.

Pre-AH: HealthChoice Model Performance

This section presents the details of the Hilltop Pre-AH Model™ performance in the HealthChoice population.

Coefficients

Table 10 shows risk factor coefficient estimates for Model 1 for the training performed in August 2023. Model 1 includes all six types of risk factors: diagnostic, pharmacy, procedural, utilization-based, geographic, and demographic. The risk factors in this table are those that were included in the final model. All other risk factors were eliminated in the variable selection step due to insufficient predictive power. Note that the risk factor coefficients are presented as odds ratios. Odds ratios can be interpreted in terms of a multiplicative impact: for example, an odds ratio of 1.05 indicates that if that risk factor were to increase by one unit, then the risk of incurring an avoidable hospitalization would increase by 5%.

Table 10. Pre-AH HealthChoice Risk Model Odds Ratios for Model 1

Risk Factor	Odds Ratio
CCW indicator for asthma	2.635
Indicator for insulin use	1.768
CCW indicator for diabetes	1.668
Indicator for urinary tract infection	1.624

Risk Factor	Odds Ratio
Number of avoidable hospitalizations	1.555
Indicator for dual eligibility with Medicaid	1.518
CCW indicator for pressure and chronic ulcers	1.517
Indicator for diabetes with complications	1.511
Indicator for oral corticosteroid use	1.461
CCW indicator for heart failure and non-ischemic heart disease	1.409
CCW indicator for hypertension	1.377
CCW indicator for tobacco use	1.374
Discontinuity of primary care - Index	1.354
Beneficiary race - Black	1.321
Indicator for fluid and electrolyte imbalance	1.321
Indicator for respiratory infection	1.3
Indicator for anti-diabetes medication use	1.287
CCW indicator for benign prostatic hyperplasia	1.262
Indicator for arrhythmia	1.251
Indicator for gastroparesis	1.24
CCW indicator for learning disabilities	1.238
Indicator for leukotriene receptor modifier use	1.23
Indicator for retinopathy	1.224
Located in partial county primary care shortage area	1.22
Indicator for pneumonia	1.213
Beneficiary race - Unknown	1.212
Discontinuity of primary care - Proportion	1.207
CCW indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis	1.194
CCW indicator for drug use disorders	1.177
Indicator for oral antibiotic use	1.175
Beneficiary gender - female	1.164
Indicator for provider administered drug	1.119
CCW indicator for chronic kidney disease	1.115
Indicator for no vaccination (flu or pneumonia)	1.112
CCW indicator for atrial fibrillation and flutter	1.11
CCW indicator for peripheral vascular disease	1.103
Beneficiary race - Hispanic	1.098
CCW indicator for depression, bipolar, and other depressive mood disorders	1.066
Indicator for presence of a for-profit hospital	1.052
Number of emergency department visits within the past 6 months	1.045
Percent speak Spanish, aged 65+	1.011
Number of urgent care visits	1.009
Percent in poverty	1.006
Number of medications	1.004

Risk Factor	Odds Ratio
Percent single mothers	1.001
Percent non-white, ages 65+	1.001
Continuity of primary care - Duration	.999
Number of specialist visits	.996
Percent married	.994
Number of primary care visits	.991
Percent foreign born	.99
Indicator for no federally qualified health center	.962
CCW indicator for anxiety disorders	.95
CCW indicator for fibromyalgia, chronic pain and fatigue	.939
CCW indicator for liver disease, cirrhosis and other liver conditions (except viral hepatitis)	.938
CCW indicator for migraine and chronic headache	.937
CCW indicator for obesity	.932
CCW indicator for ADHD, conduct disorders, and hyperkinetic syndrome	.914
CCW indicator for viral hepatitis	.912
Number of HbA1c tests	.909
Number of prior admissions	.908
Indicator for beta blocker use	.901
Indicator for neuropathy	.879
CCW indicator for hyperlipidemia	.871
CCW indicator for rheumatoid arthritis/osteoarthritis	.863
CCW indicator for HIV/AIDS	.862
Number of lab tests	.858
CCW indicator for epilepsy	.858
Located in whole county mental health care shortage area	.844
Indicator for problems with employment and unemployment	.839
CCW indicator for other developmental delays	.835
CCW indicator for breast cancer	.816
Located in partial county mental health care shortage area	.811
Prior hospitalization admission type - elective	.744
Prior hospitalization admission type - urgent	.717
Indicator for sickle cell anemia	.669
Prior hospitalization discharge status - none	.637

It is important to note that risk factor coefficient estimates will change as the model is re-trained. Risk factor coefficients for other models are available upon request.

Predicted Probabilities

The outcome for all the Pre- Models is a set of probabilities that estimate the patient-specific risk of incurring the model-specific outcome. Table 11 presents summary statistics of a recent month of Pre-AH Model™ risk scores in the HealthChoice patient population.

Table 11. Summary Statistics for Pre-AH Scores in the HealthChoice Population

Score Date	Model Date	Cohort Size	Events per Month	Average Score	N > 1% Risk	Monthly Correlation
Aug 2022	May 2022	1,506,700	2,812	.00222	32,411	0.969

In general, these events are rare and, consequently, the predicted probabilities are low. Hilltop does not interpret this as a limitation of the risk scores; rather, this reflects the relative rarity of the outcome events. Moreover, the *relative* risk is the key metric that should be used to allocate care resources: no matter the absolute risk of the patient panel, the efficient allocation of care resources requires the identification (and treatment) of the riskiest patients.

Predictive Power

Holdout Data Testing

Figure 4 shows the concentration curve for these scores on April 2022 holdout data. We find that the top 10% riskiest patients account for approximately 57% of all avoidable hospital events in the following month. Because the CMS HCC scores are not available for the HealthChoice data, we were not able to compare their performance with the Hilltop Pre-AH Model™ performance in this population.

Figure 4. Pre-AH HealthChoice Concentration Curves as of April 2022

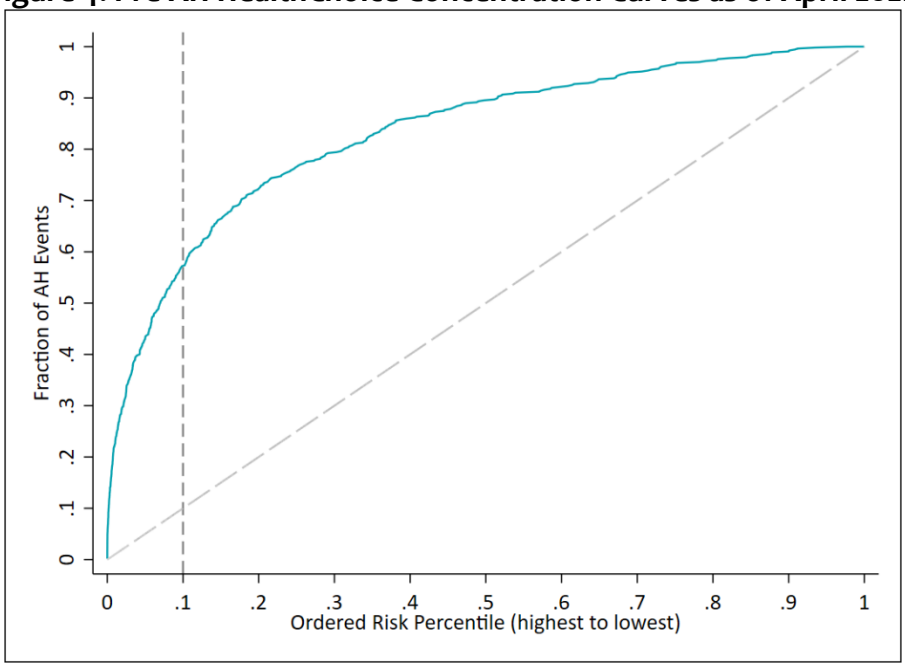
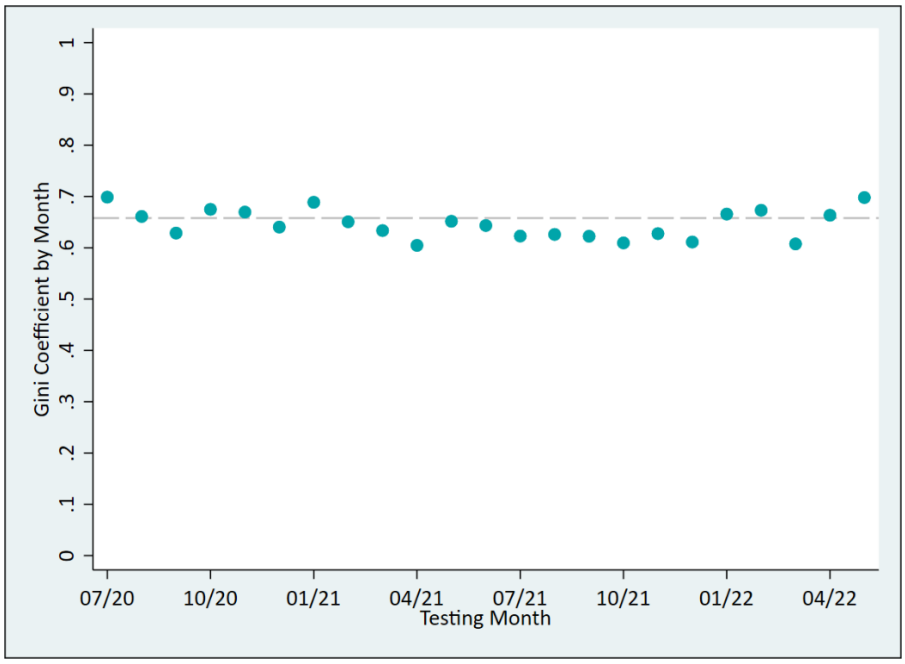


Figure 5 plots the Gini coefficients from 23 months of data based on the concentration curves estimated on the 20% holdout sample from each month. The final month of holdout data was not used due to incomplete information on avoidable hospital events in that month. These scores indicate that model performance is steady over time.

Figure 5. Pre-AH HealthChoice Gini Scores by Month



Production Data Testing

Post-deployment model evaluation is a crucial component of the predictive model lifecycle. The first Pre-AH Model™ risk scores were estimated for the HealthChoice population in April 2021 and have since been regularly updated. Hilltop monitors the accuracy of the Pre-AH Model™ predictions in a production environment by comparing the risk scores released in a given month with the *true outcomes* that occur in the following month.

For the scores released between December 2021 and May 2022, we find that the top 10% riskiest patients accounted for between approximately 54% and 60% of all avoidable hospital events over the following month (until the next score release). See Table 12.

Table 12. Predictive Performance of HealthChoice Pre-AH Model™ Scores by Month

Score Month	Evaluation Month	Top 1% of Patients, by Risk	Top 10% of Patients, by Risk
December 2021	January 2022	24.9%	59.6%
January 2022	February 2022	23.7%	58.2%
February 2022	March 2022	21.2%	55.1%
March 2022	April 2022	22.8%	55.7%
April 2022	May 2022	21.3%	57.2%
May 2022	June 2022	22.4%	54.4%

Return to Table of Contents

Pre-CH: Model Operations

This predictive model is an extension of the Hilltop Pre-AH Model™ that is designed to estimate individual-level risk of hospitalization due to COVID-19 for Medicaid enrollees across the state of Maryland. This predictive model is intended to help health care providers prospectively identify individuals at risk of hospitalization for *the current and future pandemics*. With the tool, it may be possible for them to design suitable proactive interventions to try to reduce these individuals' risk of hospitalization because health care providers can identify these individuals before they reach the hospital. Moreover, such an evidence-based forward-triage mechanism—particularly if implemented via telehealth—could help control the spread of COVID-19 through reduced hospital-based exposure and alleviate excess demand on critical acute care infrastructure.

This project was originally funded by the University of Maryland, Baltimore, Institute for Clinical and Translational Research (ICTR) through the Accelerated Translational Incubator Pilot (ATIP) Grant Program in October 2019 (awarded to Dr. Fei Han). Risk scores were deployed for the HealthChoice population starting in May 2021. These risk scores have not been deployed for the MDPCP population.

Data Sources

The Hilltop Pre-CH Model™ is built on the Hilltop Pre-AH Model™ and therefore uses the same administrative claims data, supplemented with various publicly available environmental data sets used to generate the environmental risk factors and data from the Maryland Department of Health to quantify monthly COVID-19 prevalence. The Hilltop Pre-CH Model™ was estimated for the MDPCP populations during development and then re-trained using the Maryland Health Choice population for deployment.

Hilltop uses SAS 9.4 to create the production version of the model and Stata 16.1 to create a validation version as an additional quality check. The additional risk factors developed for the Hilltop Pre-CH Model™ are described below and in greater detail in Appendix 1.

Risk Factors

Built on the Hilltop Pre-AH Model™, the Hilltop Pre-CH Model™ uses all the Pre-AH risk factors plus an additional 23 created from hospital and nursing home-related claims, physician and lab-related claims, prescription drug-related claims, publicly available environmental data, and area-based COVID-19 prevalence data from the Maryland Department of Health.

Literature Review

Based on the literature review, Hilltop identified and operationalized 23 risk factors to be included in the risk model in addition to the Pre-AH risk factors. While some of these risk factors are eliminated in the variable selection step, this process is data-driven, and all risk factors are included in the pool of *potential* risk factors to be used in the model. For a description of each risk factor, along with data source and sample statistics, see Appendix 1.

Facility-Related Risk Factors

Facility-related risk factors cover information on admissions over the past 12 months; nursing home stays over the past 12 months; and certain procedures. These claims are used to construct the COVID-19-related hospitalization event outcome, as well as the diagnostic condition flags. These condition flags rely on diagnostic information from hospital, nursing home, physician, and lab claims in conjunction with CCW coding specifications to generate beneficiary-level risk factors that represent underlying disease states that increase risk for COVID-19 hospitalization.[24]

Provider-Related Risk Factors

Provider-related risk factors cover utilization of certain services (such as vaccinations, lab tests, or J-code procedures), place of service (for example, urgent care or rural health clinic), and provider specialty (for example, endocrinology or oncology). As above, the physician and lab-related claims are used to construct the COVID-19-related hospitalization event outcome, as well as the diagnostic condition flags.

Prescription Drug-Related Risk Factors

Prescription drug-related risk factors index utilization of drugs identified in Hilltop's literature review as potential risk factors for COVID-19-related hospital events. The only prescription risk factor added to the list of Pre-AH prescription risk factors for the Pre-CH model was an indicator for immunosuppressive drug use.

Beneficiary Demographics-Related Risk Factors

Information from the beneficiary demographics files, such as date of birth, race, and sex, are used to create potential risk factors for hospitalization due to COVID-19. We geocode the address listed in beneficiary demographic data to allow the social and environmental risk factors to be linked with a person's administrative claims. These data were not used to create any additional risk factors for the Pre-CH Model™.

Environmental Risk Factors

Several of the risk factors Hilltop identified during the literature review were individual-level demographic and socioeconomic factors that are unavailable in administrative claims data (for example, 60+ minute commute). Additionally, the literature review identified several social determinants of health factors that increase an individual's risk for COVID-19 hospitalizations. These risk factors are linked to claims data based on the beneficiary's ZIP code. These additional environmental risk factors are not currently available at the census tract-level. See Appendix 2 for more detail.

[24] Additional detail on the CCW condition flag specifications can be found here: https://www2.ccwdata.org/documents/10280/19139421/chr-chronic-condition-algorithms.pdf, https://www.ccwdata.org/documents/10280/19139421/ccw-chronic-condition-algorithms-reference-list.pdf

COVID-19 Prevalence Data

Information on the monthly prevalence of COVID-19 cases in the state of Maryland (grouped by ZCTA) and the change in prevalence from the previous month come from the Maryland Department of Health. These risk factors are updated monthly and linked to claims data based on the beneficiary's ZIP code.

Outcome: COVID-19 Hospitalizations

The outcome measure in the Hilltop Pre-CH Model™ is a 0/1 indicator variable denoting whether an individual incurred a COVID-19-related hospitalization in a given month. To construct this measure, Hilltop uses ICD-10 diagnosis codes from inpatient claims to flag the following conditions, which are the basis for the composite COVID-19 hospitalization flag:[25]

- COVID-19 (U07.1)
- Other coronavirus as the cause of diseases classified elsewhere (B97.29)

This is implemented in the model as an indicator variable at the person-month level. If an individual incurs at least one COVID-19 inpatient hospital visit in a given month, then that person receives a value of 1 for this variable—and 0 otherwise.

Pre-CH: HealthChoice Model Performance

This section presents the details of the Hilltop Pre-CH Model™ performance in the HealthChoice population using data from the period of March 2020 to March 2022.

Coefficients

Table 13 presents risk factor coefficient estimates for Model 1 for the training performed in August 2023. Model 1 includes all six types of risk factors: diagnostic, pharmacy, procedural, utilization-based, geographic, and demographic. The risk factors in this table are those that were included in the final model. All other risk factors were eliminated in the variable selection step due to insufficient predictive power. Note that the risk factor coefficients are presented as odds ratios. Odds ratios can be interpreted in terms of a multiplicative impact: for example, an odds ratio of 1.05 indicates that if that risk factor were to increase by one unit, then the risk of incurring a COVID-19-related hospitalization would increase by 5%.

[25] Specifically, Hilltop defines these claims as those with a claim type of either 60 or 61 (indicating an inpatient claim) or a claim type of 40 (indicating an outpatient claim) and revenue center codes of 0450-0459 and 0981. Source: https://www.resdac.org/articles/how-identify-hospital-claims-emergency-room-visits-medicare-claims-data. To the extent that claims for observation stays are coded in this manner in the CCLF Medicare claims, observation stays are included in this outcome.

Table 13. Pre-CH HealthChoice Risk Model Odds Ratios for Model 1

Risk Factor	Odds Ratio
Indicator for sickle cell anemia	5.068
Prior hospitalization discharge status - other	4.524
Prior hospitalization discharge status - transferred to inpatient care	3.516
Number of previous COVID hospitalizations	1.856
Indicator for having diabetes and being over 40 years old.	1.841
CCW indicator for intellectual disabilities and related conditions	1.74
CCW indicator for leukemias and lymphomas	1.705
Indicator for insulin use	1.535
Indicator for fluid and electrolyte imbalance	1.51
CCW indicator for obesity	1.424
Indicator for arrhythmia	1.413
Chronic Renal Insufficiency/ESRD	1.392
CCW indicator for anemia	1.387
CCW indicator for drug use disorders	1.258
Indicator for no vaccination (flu or pneumonia)	1.25
Indicator for problems with housing and economic conditions	1.243
Indicator for protein-calorie malnutrition	1.238
CCW indicator for asthma	1.237
CCW indicator for diabetes	1.219
Indicator for urinary tract infection	1.215
Indicator for oral antibiotic use	1.214
Indicator for prior readmission	1.208
CCW indicator for schizophrenia and other psychotic disorders	1.207
CCW indicator for heart failure and non-ischemic heart disease	1.192
CCW indicator for epilepsy	1.186
Indicator for sepsis	1.16
CCW indicator for hypertension	1.158
Indicator for oral corticosteroid use	1.155
Indicator for sleep apnea	1.155
Immunosuppressive disease	1.148
Located in partial county mental health care shortage area	1.121
Beneficiary race - Black	1.12
Number of prior admissions	1.055
Percent aged 0-4	1.034
Age	1.032
Number of emergency department visits within the past 6 months	1.019
Percent of units with 0 or 1 bedrooms - ZCTA	1.008
Prior admission length of stay	1.003
Air pollution level	1.001

Risk Factor	Odds Ratio
Continuity of primary care - Duration	.997
Number of medications	.995
Percent foreign born	.992
Vitamin D deficiency	.867
Indicator for lifestyle problems	.851
CCW indicator for rheumatoid arthritis/osteoarthritis	.841
Indicator for durable medical equipment (DME) use	.839
CCW indicator for hyperlipidemia	.828
Pure Hypercholesterolemia	.796
Prior hospitalization admission type - urgent	.749
Indicator for prior nursing home stay	.738
Prior hospitalization discharge status - none	.648
Prior hospitalization admission type - elective	.643

It is important to note that risk factor coefficient estimates will change as the model is re-trained. Risk factor coefficients for other models are available upon request.

Predicted Probabilities

The outcome of the Pre-CH Model is a set of probabilities that estimate the patient-specific risk of incurring a COVID-19 inpatient hospitalization event in the following months. Table 14 presents summary statistics of a recent month of Hilltop Pre-CH Model™ risk scores in the HealthChoice patient population.

Table 14. Summary Statistics for Pre-CH Scores in the HealthChoice Population

Score Date	Model Date	Cohort Size	Events per Month	Average Score	N > 1% Risk	Monthly Correlation
Aug 2022	Mar 2022	1,506,700	305	.00059	5,403	0.961

In general, these events are very rare; consequently, the predicted probabilities are very low. Like the Pre-AH risk scores, Hilltop does not interpret this as a limitation of the risk scores; rather, this reflects the relative rarity of the events. Moreover, the *relative* risk is the key metric that should be used to allocate care resources: no matter the absolute risk of the patient panel, the efficient allocation of care resources requires the identification (and treatment) of the riskiest patients.

Predictive Power

Validation Data Testing

Figure 6 estimates the concentration curve for these scores on actual COVID-19 hospital events as of December 2021.

Figure 6. Pre-CH HealthChoice Concentration Curves as of December 2021

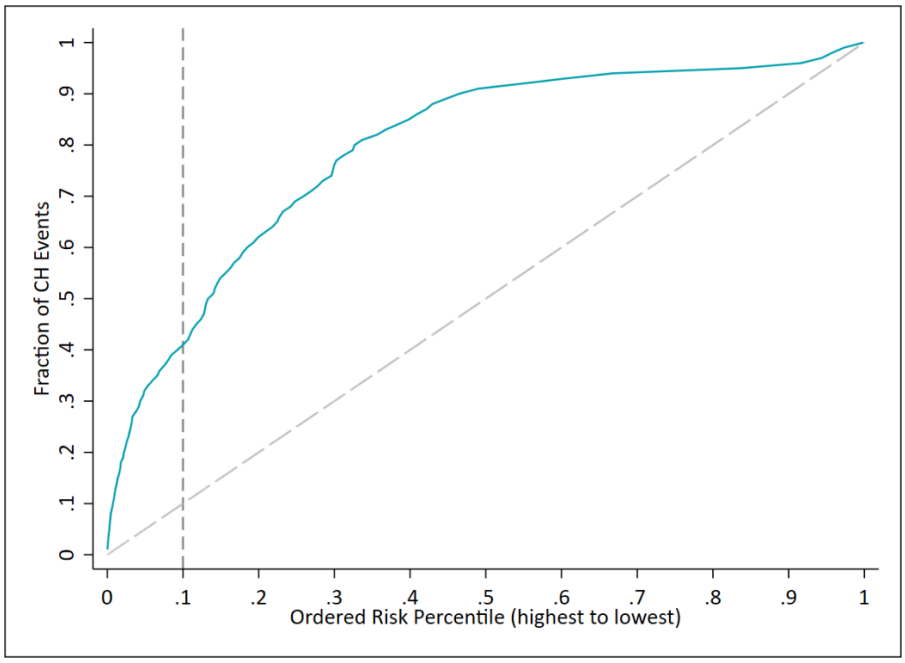
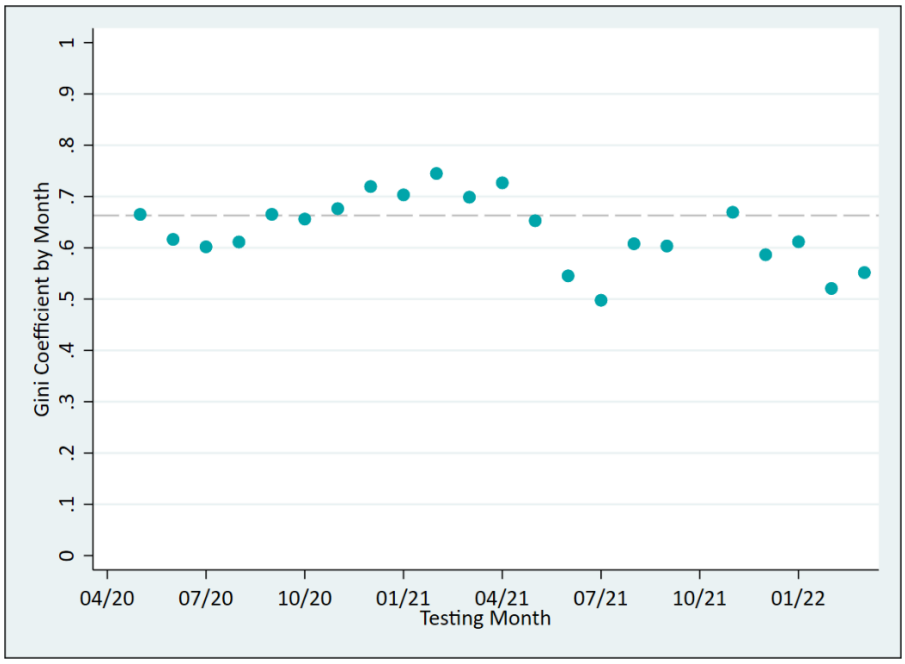


Figure 7 plots the Gini coefficients from 19 months of data based on the concentration curves estimated on the 20% holdout sample from each month. The final month of holdout data was not used due to incomplete information on COVID-19 hospital events in that month.

Figure 7. Pre-CH HealthChoice Gini Scores by Month



Production Data Testing

Post-deployment model evaluation is a crucial component of the predictive model lifecycle. Table 15, below, assesses the accuracy of the Pre-CH risk scores in a production environment from July 2021 to January 2022. Due to the relative rarity of COVID-19 hospitalizations, model performance is variable over time. We find that the top 10% riskiest patients accounted for between 37.9% and 52.7% of all COVID-19 hospitalization over the following month (until the next score release).

Table 15. Predictive Performance of HealthChoice Pre-CH Model Scores by Month

Model	Evaluation Month	Top 10% of Patients, by Risk
y21m04	July 2021	42.53%
y21m08	September 2021	37.86%
y21m08	October 2021	46.40%
y21m10	November 2021	47.96%
y21m12	December 2021	45.79%
y21m12	January 2022	52.72%

Return to Table of Contents

Pre-DC: Model Operations

The Hilltop Institute's Pre-DC Model™ is designed to facilitate the active management of type 2 diabetes by estimating individuals' risk of incurring inpatient admissions or ED visits for severe diabetes complications. Both the Pre-AH Model™ and the Pre-DC Model™ include diabetes complications in the outcome that is predicted; however, the predicted outcome differs significantly across the two models, and the resulting risk scores are statistically distinct.[26] Hilltop estimates risk scores and reasons for risk for all patients of MDPCP-attributed practices every month to help care teams proactively identify high-risk individuals and thus target care management resources accordingly. This model was initially developed by The Hilltop Institute, in conjunction with the Maryland Department of Health, to support the care management efforts of primary care providers enrolled in MDPCP in alignment with the State's Statewide Integrated Health Improvement Strategy (SIHIS) goal of reducing the public health costs of diabetes.

The Hilltop Pre-DC Model™ risk scores were first deployed for the MDPCP population in October 2022. Patient-level risk scores are provided to participating practices monthly for their attributed beneficiaries via CRISP and are known as the "Severe Diabetes Complications (Pre-DC)" scores.

Data Sources

The Hilltop Pre-DC Model™ is built on the Hilltop Pre-AH Model™ and therefore uses the same administrative claims data, supplemented with various publicly available environmental data sets used to generate the environmental risk factors.

Hilltop uses SAS 9.4 to create the production version of the model and Stata 16.1 to create a validation version as an additional quality check. The risk factors for the Hilltop Pre-DC Model™ are briefly described below and in greater detail in Appendix 1.

Risk Factors

The Pre-DC Model™ is built on the Hilltop Pre-AH Model™ and thus uses all the risk factors from the Pre-AH Model™. However, there are 18 additional risk factors created from facility-related claims, provider-related claims, and prescription drug-related claims. While some of these risk factors are eliminated in the variable selection step, this process is data-driven, and all risk factors are included in the pool of *potential* risk factors to be used in the model. A high-level description of risk factors, as well as the process for identifying them, is provided in the sections below. For a description of each risk factor, along with data source and sample statistics, see Appendix 1.

[26] For additional information see https://health.maryland.gov/mdpcp/Documents/PreDC_PreAH_Outcome_Distinction_Final.pdf

Literature Review

As part of the development process for its type 2 diabetes complications predictive model, Hilltop conducted a literature review to identify potential risk factors for inclusion in the model. This is a crucial element of model development: including high-quality risk factors as predictors can improve model performance, transparency, and interpretability. The review was intended to survey the existing literature and locate risk factors for which there is statistical evidence of association with type 2 diabetes complications. In early 2022, the research team searched PubMed to identify published literature that identifies risk factors for hospitalization for type 2 diabetes complications.[27] This review proceeded in three phases: a title screen, an abstract screen, and a full-text review. All records were reviewed by two independent reviewers on the research team. Any disagreements were reconciled through additional reviewer discussion.

We identified 107 articles that met the search criteria and conducted title and abstract screens on this pool of results. This process yielded 35 papers for full-text review. In the risk factor extraction process, we excluded as candidate risk factors those that were similar in substance to those already in the Pre-AH risk factor library. We then grouped similar remaining risk factors. The risk factor extraction yielded 18 unduplicated risk factors that have been shown to be highly predictive of type 2 diabetes complications.

Facility-Related Risk Factors

Facility-related risk factors cover information on admissions over the past 12 months; nursing home stays over the past 12 months; and certain procedures. These claims are used to construct the severe type 2 diabetes complication event outcomes as well as diagnostic condition flags. These condition flags rely on diagnostic information from hospital, nursing home, physician, and lab claims to generate beneficiary-level risk factors that represent underlying disease states that increase risk for severe complications from type 2 diabetes. Ten risk factors that use these claims were added for the Pre-DC Model™.

Provider-Related Risk Factors

Provider-related risk factors cover utilization of certain services (such as vaccinations, lab tests, or J-code procedures), place of service (for example, urgent care or rural health clinic), and provider specialty (for example, endocrinology or oncology). Physician and lab-related claims are used to construct diagnostic and procedural condition flags, such as for the risk factors indexing history of diabetic complications. Nine risk factors that use these claims were added for the Pre-DC Model™.

Prescription Drug-Related Risk Factors

Prescription drug-related risk factors index utilization of drugs identified in Hilltop's literature review as potential risk factors for severe complications from type 2 diabetes. Seven risk factors

[27] We used the following search strings: "(Diabetes Mellitus, Type 2/complications*) AND (Machine Learning)"; "(Diabetes Mellitus, Type 2/complications*) AND (Predict*) AND (Administrative data)"

that use these claims were added for the Pre-DC Model™, including the total out-of-pocket spending on all prescriptions over the previous 12 months.

Beneficiary Demographics-Related Risk Factors

Information from the beneficiary demographics files, such as date of birth, race, and sex, are used to create potential risk factors for severe complications from type 2 diabetes events. We geocode the address listed in beneficiary demographic data to allow the social and environmental risk factors to be linked with a person's administrative claims. One risk factor was added for the Pre-DC Model™ that uses beneficiary demographic data.

Environmental Risk Factors

The last category of risk factors indexes social and environmental risk. These are risk factors identified during the literature review were individual-level demographic and socioeconomic factors that are unavailable in administrative claims data (e.g., food insecurity). Hilltop did not identify any social/environmental risk factors that were highly predictive of severe type 2 diabetes complications in previous research that were not already included in the existing pool of social/environmental features. See Appendices 1 and 2 for more detail.

Outcome: Severe Type 2 Diabetes Complication

Severe complication of type 2 diabetes (1/0) is defined as an inpatient hospitalization or ED visit in a person-month with one or more of the following ICD-10 diagnosis codes (in any position on the claim) associated with severe complications of diabetes as defined by the Diabetes Complication Severity Index (DCSI):[28]

Retinopathy

Retinal detachments and breaks: H33.x
Type 2 diabetes mellitus with severe non-proliferative diabetic retinopathy: E11.34xx
Type 2 diabetes mellitus with proliferative diabetic retinopathy: E11.35xx
Blindness and low vision: H54.x
Vitreous hemorrhage: H43.1x

Nephropathy

Type 2 diabetes mellitus with chronic kidney disease (stage 4 or 5): E11.22, N18.4, N18.5
Type 2 diabetes mellitus with end stage renal disease: E11.22, N18.6
Unspecified kidney failure: N19

Cerebrovascular Complications

Nontraumatic intracerebral hemorrhage: I61.x
Cerebral infarction: I63.x
Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction: I65.x
Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction: I66.x

[28] Centers for Disease Control and Prevention, 2020; Chang et al., 2012; Glasheen et al., 2017

Acute cerebrovascular insufficiency: I67.81

Cardiovascular Complications
Acute myocardial infarction (STEMI, NSTEMI): I21.x
Subsequent acute myocardial infarction (STEMI, NSTEMI): I22.x
Complications from acute myocardial infarction (STEMI, NSTEMI): I23.x
Old myocardial infarction: I25.2
Atrial fibrillation and flutter: I48.x
Cardiac arrest: I46.x
Paroxysmal tachycardia: I47.x
Other cardiac arrythmia: I49.x
Heart failures: I50x
Atherosclerosis of native arteries of the extremities with ulceration/gangrene: I70.25x, I70.26x
Aortic aneurysm/dissection: I71.x

Peripheral Vascular Disease
Gas gangrene: A48.0
Embolism and thrombosis of arteries of the lower extremities: I74.3
Non-pressure chronic ulcer of limb, not elsewhere classified: L97.x
Type 2 diabetes with diabetic peripheral angiopathy, with gangrene: E11.52
Gangrene, not elsewhere classified: I.96

Metabolic Complications
Type 2 diabetes mellitus with hyperosmolarity, with coma: E1101
Type 2 diabetes mellitus with ketoacidosis, with coma: E1111
Type 2 diabetes mellitus with hypoglycemia, with coma: E11641

The DSCI scores complications from 0 (no abnormality) to 2 (severe abnormality). Only complications with a score of 2 are included in the event definition for "severe complications of type 2 diabetes" (Young et al., 2008).

What Makes the Pre-DC Outcome Different from the Pre-AH Outcome?

Both the Pre-AH Model and the Pre-DC Model outcomes include diabetes complications; however, they are conceptually and statistically distinct. The Pre-AH outcome is a composite of 10 conditions that are determined to be potentially preventable with high-quality outpatient care by the AHRQ.[29] These PQIs are intended to serve as a high-level check of primary/outpatient care access in a community and to help organizations identify potentially unmet needs in their communities. The Pre-AH outcome indexes, among other non-diabetes-related conditions, uncontrolled diabetes complications as well as complications from type 1 diabetes, type 2 diabetes, and other forms of diabetes (e.g., gestational diabetes). In comparison, the Pre-DC outcome is based on the DCSI, which is designed to quantify the severity of diabetes complications based on risk for adverse medical outcomes including future medical needs, high

[29] For more information, see https://www.qualityindicators.ahrq.gov/modules/pqi_resources.aspx

treatment costs, hospitalizations, and mortality (described in more detail above). Although the DCSI can measure non-severe and severe complications from all forms of diabetes, we focused on severe complications related to type 2 diabetes only.

In order to confirm that the Pre-DC outcome was statistically distinct from the Pre-AH outcome, we quantified the overlap in the ICD-10 diagnosis codes included in both outcomes and calculated the correlation between the Pre-AH and Pre-DC outcome frequency and risk scores in the MDPCP scoring data. See Table 16.

Table 16. Coding Differences in Pre-AH and Pre-DC Model Outcomes

Breakdown of ICD-10 Diagnosis Codes in AH & DC Outcomes	
Overlapping	*N*=36
Unique to Pre-AH Outcome	*N*=491
Unique to Pre-DC Outcome	*N*=244

Hilltop also determined that, as of July 2022, only 12.46% of the avoidable hospital events in the MDPCP scoring data were related to PQIs indexing diabetes or its complications (i.e., PQI#1, PQI#3, PQI#14, PQI#16). The most prevalent PQIs in the MDPCP population were PQI#14: Urinary Tract Infections (25.43%) and PQI#5: COPD or Asthma in Older Adults (18.47%).

For additional detail on the differences in these outcomes, please see the standalone document entitled "What's the Difference between the Pre-DC and Pre-AH Models?".

Pre-DC: MDPCP Model Performance

Coefficients

Table 17 presents risk factor coefficient estimates for Model 1 for the training performed in September 2023. Model 1 includes all six types of risk factors: diagnostic, pharmacy, procedural, utilization-based, geographic, and demographic. The risk factors in this table are those that were included in the final model. All other risk factors were eliminated in the variable selection step due to insufficient predictive power. Note that the risk factor coefficients are presented as odds ratios. Odds ratios can be interpreted in terms of a multiplicative impact: for example, an odds ratio of 1.05 indicates that if that risk factor were to increase by one unit, then the risk of incurring severe complication of type 2 diabetes would increase by 5%.

Table 17. Pre-DC MDPCP Risk Model Odds Ratios for Model 1

Risk Factor	Odds Ratio
Prior hospitalization discharge status - other	4.97
Prior hospitalization discharge status - transferred to inpatient care	2.306
Indicator for hospice enrollment	2.213
Indicator for sickle cell anemia	2.065
DCSI Score – Cardiovascular	1.735
CCW indicator for atrial fibrillation and flutter	1.684

Risk Factor	Odds Ratio
CCW indicator for heart failure and non-ischemic heart disease	1.484
CCW indicator for sensory (blindness and visual) impairment	1.398
Chronic Renal Insufficiency/ESRD	1.366
CCW indicator for hypertension	1.34
Prior hospitalization admission type - emergency	1.297
Indicator for original Medicare eligibility for a non-age-related cause	1.262
Prior hospitalization admission type - urgent	1.26
CCW indicator for spina bifida and other congenital anomalies of the nervous system	1.205
CCW indicator for intellectual disabilities and related conditions	1.204
Number of Previous Severe Type 2 Diabetes Complications	1.194
Prior hospitalization discharge status - transferred to post-acute care	1.188
Indicator for metastatic cancer	1.176
CCW indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis	1.168
Beneficiary race - Black	1.154
DCSI Score – Nephropathy	1.151
CCW indicator for lung cancer	1.146
Indicator for frailty	1.145
Indicator for fluid and electrolyte imbalance	1.142
CCW indicator for leukemias and lymphomas	1.137
Indicator for arrhythmia	1.137
CCW indicator for anemia	1.135
Indicator for beta blocker use	1.129
Beneficiary race - White	1.125
Indicator for gastroparesis	1.117
Indicator for insulin use	1.114
CCW indicator for tobacco use	1.107
DCSI Score - Peripheral Vascular Disease	1.103
CCW indicator for Parkinson's Disease or Secondary Parkinsonism	1.099
Indicator for previous conservative diabetic wound procedure	1.094
Indicator for dual eligibility with Medicaid	1.094
Indicator for cerebrovascular disease	1.093
Indicator for oral corticosteroid use	1.091
Number of emergency department visits within the past 6 months	1.091
Discontinuity of primary care - Index	1.089
Indicator for albuminuria	1.086
CCW indicator for pressure and chronic ulcers	1.075
CCW indicator for diabetes	1.072
CCW indicator for chronic kidney disease	1.068
Indicator for rheumatoid arthritis/collagen vascular disease	1.067

Risk Factor	Odds Ratio
Indicator for durable medical equipment (DME) use	1.064
DCSI Score – Retinopathy	1.064
Indicator for provider administered drug	1.061
Indicator for warfarin use	1.06
Indicator for problems with care provider dependency	1.055
Indicator for urinary tract infection	1.053
Indicator for oncologist visit	1.051
Indicator for no vaccination (flu or pneumonia)	1.051
Indicator for rivaroxaban use	1.048
CCW indicator for ischemic heart disease	1.046
Indicator for use of Anticoagulants	1.044
CCW indicator for depression, bipolar, and other depressive mood disorders	1.038
DCSI Score – Cerebrovascular	1.038
Number of avoidable hospitalizations	1.034
DCSI Score – Neuropathy	1.032
Indicator for pulmonary circulatory disorder	1.032
CCW indicator for obesity	1.028
Age	1.021
Number of urgent care visits	1.018
Prior admission length of stay	1.006
Number of outpatient visits	1.003
National ranking of deprivation	1.002
Percent with less than high school education, ages 65+	1.001
Total health spending	1.000
Diabetes Duration	1.000
Part D OOP spending	1.000
Continuity of primary care - Duration	.999
Percent aged 65 and over	.998
Percent foreign born	.998
Number of primary care visits	.997
Number of lab tests	.987
Number of HbA1c tests	.975
CCW indicator for benign prostatic hyperplasia	.958
Number of prior admissions	.953
Indicator for losartan use	.952
CCW indicator for glaucoma	.952
Number of heart-related procedures	.951
CCW indicator for prostate cancer	.948
Indicator for prior surgery	.947
Indicator for anti-diabetes medication use	.946
CCW indicator for osteoporosis with or without pathological fracture	.945

Risk Factor	Odds Ratio
Beneficiary gender - female	.936
CCW indicator for schizophrenia and other psychotic disorders	.927
CCW indicator for hip/pelvic fracture	.926
CCW indicator for cataracts	.917
CCW indicator for hyperlipidemia	.913
CCW indicator for ADHD, conduct disorders, and hyperkinetic syndrome	.89
Indicator for statin use	.863
Indicator for prior nursing home stay	.826
Located in whole county primary care shortage area	.747

Predicted Probabilities

The outcome of the Pre-DC Model is a set of probabilities that estimate the patient-specific risk for an inpatient hospitalization or ED visit for a severe complication of type 2 diabetes in the next month across six domains of complications: ophthalmic/retinopathy; nephropathy; cerebrovascular; cardiovascular; peripheral vascular; and metabolic. See Table 18.

Table 18. Summary Statistics for Pre-DC Scores in the MDPCP Population

Score Date	Model Date	Cohort Size	Events per Month	Average Score	N > 1% Risk	Monthly Correlation
Oct 2022	July 2022	361,192	7,139	.01303	97,038	NA

In general, the predicted probabilities are relatively low; Hilltop interprets this as a reflection of the relative rarity of the outcome event. The key metric that should be used to allocate care resources is the relative risk: no matter the absolute risk of the patient panel, the efficient allocation of care resources requires the identification (and treatment) of the riskiest patients. At the time of writing, we are unable to calculate the month-over-month stability of risk scores because only one month of Pre-DC Model risk scores has been released; future versions of this document will update this field.

Predictive Power

Validation Data Testing

Figure 8 shows the concentration curve for these scores on severe diabetes complication events as of May 2022. We find that the top 10% riskiest patients account for approximately 61% of all severe diabetes complication events in the following month.

Figure 8. Pre-DC MDPCP Concentration Curves as of May 2022

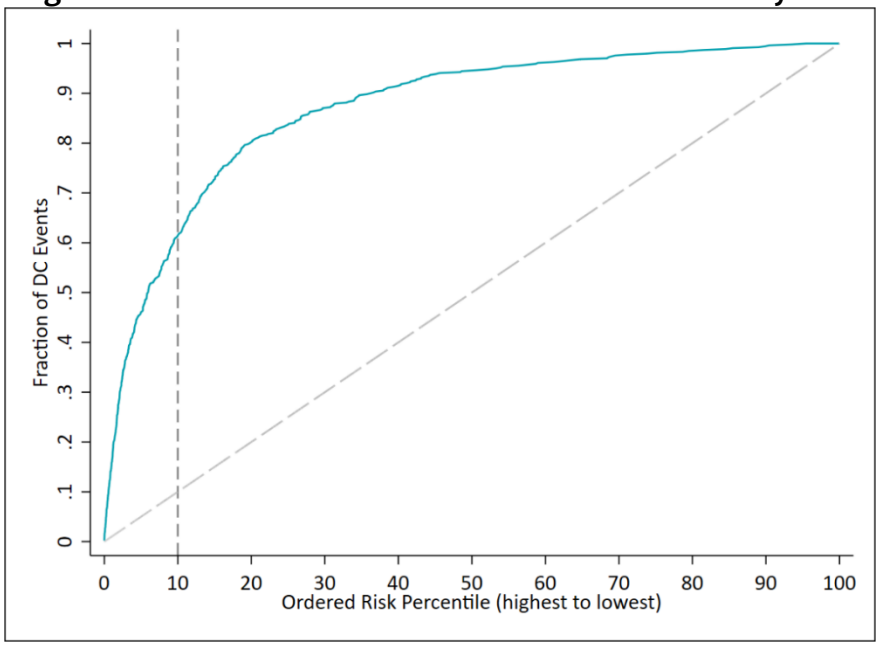
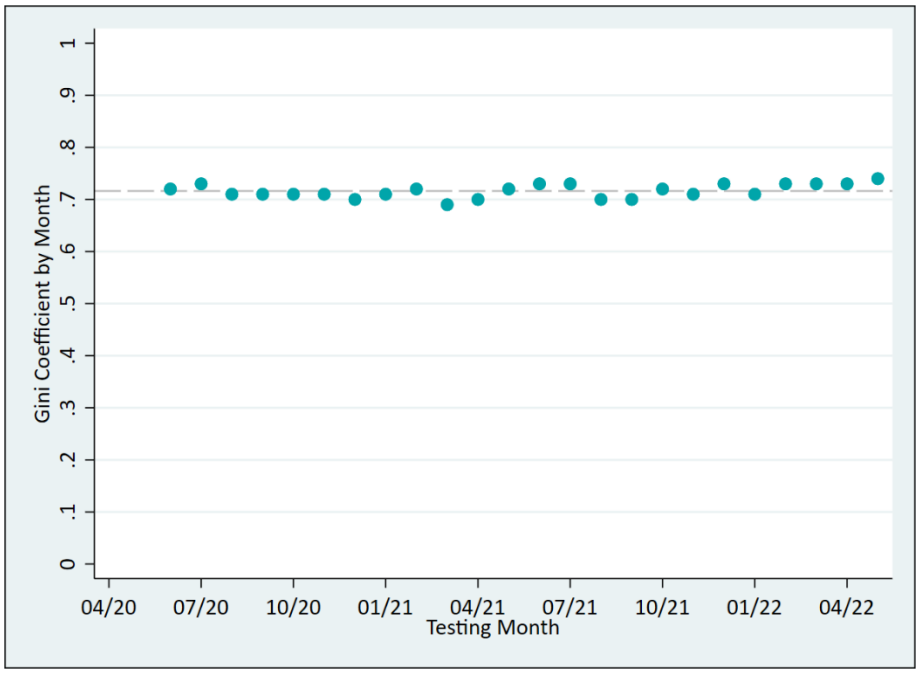


Figure 9 plots the Gini coefficients from 23 months of data based on the concentration curves estimated on the 20% holdout sample from each month. The final month of holdout data was not used due to incomplete information on severe diabetes complication events in that month. These scores indicate that model performance is steady over time, with a very slight upward trend. This is consistent with the risk factors becoming more predictive as additional claims history becomes available.

Figure 9. Pre-DC MDPCP Gini Scores by Month



Production Data Testing

The risk scores based on the Pre-DC Model were released in October 2022; as of the time of writing, data are not available to assess the predictive power in a production environment.

Return to Table of Contents

Pre-HE: Model Operations

The Hilltop Institute's Pre-HE Model™ is designed to support proactive advanced care planning discussions by estimating a patient's risk of death within the next six months. Every month, Hilltop estimates risk scores and reasons for risk for all attributed patients of MDPCP-participating practices with the goal of identifying patients who are potentially suitable for hospice care and providing care teams with information that can guide the sensitive and difficult conversations about end-of-life care with patients and their families. It was initially developed by The Hilltop Institute, in conjunction with the Maryland Department of Health, to support the care management efforts of primary care providers enrolled in MDPCP.

The Hilltop Pre-HE Model™ risk scores were first deployed for the MDPCP population in October 2022. Patient-level risk scores are provided to participating practices monthly for their attributed beneficiaries via CRISP and are known as the "Hospice Eligibility and Advanced Care Planning (Pre-HE)" scores.

Data Sources

The Hilltop Pre-HE Model™ is built on the Hilltop Pre-AH Model™ and therefore uses the same administrative claims data, supplemented with various publicly available environmental data sets used to generate the environmental risk factors.

Hilltop uses SAS 9.4 to create the production version of the model and Stata 16.1 to create a validation version as an additional quality check. The risk factors for the Hilltop Pre-HE Model™ are described below and in greater detail in Appendix 1.

Risk Factors

Because the Pre-HE Model™ is built on the Hilltop Pre-AH Model™, it uses all the risk factors from the Pre-AH Model™, in addition to 18 other risk factors created from facility-related claims, provider-related claims, and prescription drug-related claims. While some of these risk factors are eliminated in the variable selection step, this process is data-driven, and all risk factors are included in the pool of *potential* risk factors to be used in the model. A high-level description of risk factors, as well as the process for identifying them, is provided in the sections below. For a description of each risk factor, along with data source and sample statistics, see Appendix 1.

Literature Review

As part of the development process for this model, Hilltop conducted a literature review in order to identify potential risk factors for inclusion in the model. This is a crucial element of model development: including high-quality risk factors as predictors can improve model performance, transparency, and interpretability. The review was intended to survey the existing literature and locate risk factors for which there is statistical evidence of association with mortality within a short or moderate time horizon. In early 2022, the research team searched PubMed to identify

published literature that identifies risk factors for mortality.[30] This review proceeded in three phases: a title screen, an abstract screen, and a full-text review. All records were reviewed by two independent reviewers on the research team, and any disagreements were reconciled through additional reviewer discussion.

We identified 80 articles that met the search criteria and conducted title and abstract screens on this pool of results. This process yielded 22 papers for full-text review. In the risk factor extraction process, we excluded as candidate risk factors those that were similar in substance to those already in the Pre-AH risk factor library, and then grouped similar remaining risk factors. The risk factor extraction yielded 18 unduplicated risk factors that have been shown to be highly predictive of mortality within a short or moderate time horizon.

Facility-Related Risk Factors

Facility-related risk factors cover information on admissions over the past 12 months, nursing home stays over the past 12 months, and certain procedures. These claims are used to construct the diagnostic condition flags (e.g., indicating pancreatic cancer), as well as the claims-based frailty index. These condition flags rely on diagnostic information from hospital, nursing home, physician, and lab claims in conjunction with CCW coding specifications to generate beneficiary-level risk factors that represent underlying disease states that increase risk for mortality within a short or moderate time horizon. Fifteen risk factors that use these claims were added for the Pre-HE Model™.

Provider-Related Risk Factors

Provider-related risk factors cover utilization of certain services (such as vaccinations, lab tests, or J-code procedures), place of service (for example, urgent care or rural health clinic), and provider specialty (for example, endocrinology or oncology). Physician and lab-related claims are used to construct diagnostic and procedural condition flags (e.g., dialysis). All 18 risk factors added for the Pre-HE Model™ use these claims.

Prescription Drug-Related Risk Factors

Prescription drug-related risk factors index utilization of drugs identified in Hilltop's literature review as potential risk factors for mortality within six months. One risk factor that uses these claims was added for the Pre-HE Model™.

Beneficiary Demographics-Related Risk Factors

Information from the beneficiary demographics files, such as date of birth, race, and sex, are used to create potential risk factors for mortality within six months. We geocode the address listed in beneficiary demographic data to allow the social and environmental risk factors to be

[30] We used the following search strings: "(mortality/frailty) AND (predict*) AND Medicare*"; "(mortality) AND (machine learning) AND Medicare*"

linked with a person's administrative claims. These data were used to construct the mortality outcome event for the Pre-HE Model™ but were not used to create any additional risk factors.

Environmental Risk Factors

The last category of risk factors indexes social and environmental risk. These risk factors identified during the literature review are individual-level demographic and socioeconomic factors that are unavailable in administrative claims data (e.g., food insecurity). Hilltop did not identify any social/environmental risk factors that were highly predictive of mortality in previous research that were not already included in the existing pool of social/environmental features. See Appendices 1 and 2 for more detail.

Outcome: Death within Six Months

Death within six months (1/0) is defined at the person-month level as the presence of a date of death for a beneficiary in the Beneficiary Demographics file that is within six months of the last day of each person-month. This means that for each beneficiary who has died, the flag for this event will be a 1 for the six months prior to their death.

Table 19. Example Scenario for Modeling Death within 6 Months

	Jun 2020	Jul 2020	Aug 2020	Sep 2020	Oct 2020	Nov 2020	Dec 2020
Presence of a Date of Death	-	-	-	-	-	-	X
Death within 6 Months Flag	0	1	1	1	1	1	1

Pre-HE: MDPCP Model Performance

Coefficients

Table 20 presents risk factor coefficient estimates for Model 1 for the training performed in September 2023. Model 1 includes all six types of risk factors: diagnostic, pharmacy, procedural, utilization-based, geographic, and demographic. The risk factors in this table are those that were included in the final model. All other risk factors were eliminated in the variable selection step due to insufficient predictive power. Note that the risk factor coefficients are presented as odds ratios. Odds ratios can be interpreted in terms of a multiplicative impact: for example, an odds ratio of 1.05 indicates that if that risk factor were to increase by one unit, then the risk of all-cause mortality within six months would increase by 5%.

Table 20. Pre-HE MDPCP Risk Model Odds Ratios for Model 1

Risk Factor	Odds Ratio
Indicator for hospice enrollment	41.779
Severity of Frailty	34.951
Prior hospitalization discharge status - other	5.236

Risk Factor	Odds Ratio
Indicator for metastatic cancer	2.56
Indicator for Cancer of Pancreas	2.512
Indicator for Cancer of Liver and Intrahepatic Bile Duct	2.372
CCW indicator for muscular dystrophy	2.316
Indicator for Cancer of Brain and Nervous System	1.928
CCW indicator for alcohol use disorders	1.919
Indicator for having Received Chemotherapy	1.917
Indicator for Cancer of Esophagus	1.842
Beneficiary race - Native American	1.806
CCW indicator for intellectual disabilities and related conditions	1.698
Indicator for original Medicare eligibility for a non-age-related cause	1.624
Indicator for Cancer of Ovary	1.597
Indicator for Cancer of Bronchus; Lung	1.562
Indicator for psychosocial problems	1.543
Indicator for Cancer of Bone and Connective Tissue	1.533
CCW indicator for leukemias and lymphomas	1.529
CCW indicator for non-Alzheimer's dementia	1.489
CCW indicator for heart failure and non-ischemic heart disease	1.461
Indicator for frailty	1.45
Prior hospitalization admission type - emergency	1.402
CCW indicator for autism spectrum disorders	1.389
CCW indicator for Parkinson's Disease or Secondary Parkinsonism	1.386
Prior hospitalization admission type - urgent	1.367
Beneficiary race - White	1.35
CCW indicator for pressure and chronic ulcers	1.35
Chronic Renal Insufficiency/ESRD	1.345
CCW indicator for cerebral palsy	1.343
Prior hospitalization discharge status - transferred to inpatient care	1.341
CCW indicator for multiple sclerosis and transverse myelitis	1.34
Indicator for protein-calorie malnutrition	1.295
CCW indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis	1.289
Indicator for Cancer of Stomach	1.272
Indicator for oncologist visit	1.267
Recent Increase in Frailty severity	1.266
CCW indicator for tobacco use	1.26
Indicator for having Received Dialysis	1.253
CCW indicator for Alzheimer's disease	1.238
CCW indicator for liver disease, cirrhosis and other liver conditions (except viral hepatitis)	1.238
CCW indicator for anemia	1.225

Risk Factor	Odds Ratio
CCW indicator for drug use disorders	1.221
Indicator for fluid and electrolyte imbalance	1.195
Indicator for immunosuppressive drug use	1.195
CCW indicator for chronic kidney disease	1.194
CCW indicator for atrial fibrillation and flutter	1.186
Indicator for Oxygen Usage in DME	1.173
Indicator for no vaccination (flu or pneumonia)	1.165
CCW indicator for peripheral vascular disease	1.161
Indicator for pulmonary circulatory disorder	1.158
Indicator for problems with care provider dependency	1.158
Indicator for Paraplegia or Hemiplegia	1.157
Indicator for solid tumor without metastasis	1.154
CCW indicator for diabetes	1.131
CCW indicator for acute myocardial infarction	1.128
Beneficiary race - Black	1.127
CCW indicator for sensory (blindness and visual) impairment	1.126
Indicator for oral corticosteroid use	1.113
CCW indicator for hypertension	1.103
Indicator for rheumatoid arthritis/collagen vascular disease	1.084
Indicator for dual eligibility with Medicaid	1.074
Age	1.074
Number of avoidable hospitalizations	1.069
Indicator for albuminuria	1.067
Indicator for warfarin use	1.067
Number of emergency department visits within the past 6 months	1.057
Indicator for beta blocker use	1.055
Indicator for urinary tract infection	1.053
Located in whole county mental health care shortage area	1.052
Indicator for diabetes with complications	1.051
CCW indicator for pneumonia, all-cause	1.045
Indicator for arrhythmia	1.038
Indicator for presence of a for-profit hospital	1.036
Number of home health visits	1.03
Prior admission length of stay	1.014
Number of medications	1.012
National ranking of deprivation	1.001
Continuity of primary care - Duration	1.001
Total health spending	1.000
Percent single mothers	1.000
Median household income	1.000
Physician diversity	1.000

Risk Factor	Odds Ratio
Taxable interest per capita	1.000
Number of outpatient visits	.998
Percent aged 65 and over	.998
Number of specialist visits	.996
Social workers per 1000 residents	.994
Number of lab tests	.973
Indicator for cerebrovascular disease	.971
Indicator for provider administered drug	.966
Indicator for sepsis	.958
CCW indicator for colorectal cancer	.936
CCW indicator for sensory (deafness and hearing) impairment	.928
CCW indicator for depression, bipolar, and other depressive mood disorders	.921
Indicator for mental health use	.916
Number of HbA1c tests	.915
Indicator for sleep apnea	.912
CCW indicator for benign prostatic hyperplasia	.91
CCW indicator for fibromyalgia, chronic pain and fatigue	.9
Number of urgent care visits	.898
CCW indicator for breast cancer	.896
Indicator for anti-diabetes medication use	.895
Indicator for prior surgery	.893
Indicator for problems with social environment	.89
CCW indicator for osteoporosis with or without pathological fracture	.887
CCW indicator for glaucoma	.885
Indicator for respiratory infection	.881
CCW indicator for asthma	.877
CCW indicator for hip/pelvic fracture	.865
Indicator for losartan use	.859
Indicator for gastroesophageal reflux disease	.85
CCW indicator for personality disorders	.85
CCW indicator for rheumatoid arthritis/osteoarthritis	.843
CCW indicator for migraine and chronic headache	.84
CCW indicator for obesity	.84
Indicator for statin use	.828
Prior hospitalization discharge status - home	.804
Indicator for other problems with primary support group	.803
CCW indicator for cataracts	.794
CCW indicator for hyperlipidemia	.785
Indicator for prior nursing home stay	.782
Indicator for Hospital Bed Usage in DME	.75
CCW indicator for prostate cancer	.712

Risk Factor	Odds Ratio
CCW indicator for other developmental delays	.693
Beneficiary gender - female	.647
Discontinuity of primary care - Index	.621
Indicator for occupational exposure to risk factors	.545
Indicator for insulin use	.0003

Predicted Probabilities

The outcome of the Pre-HE Model is a set of probabilities that estimate the patient-specific risk of all-cause mortality within the next six months. In general, the predicted probabilities are relatively low and tend to persist across time.

Table 21. Summary Statistics for Pre-HE Scores in the MDPCP Population

Score Date	Model Date	Cohort Size	Events per Month	Average Score	N > 1% Risk	Monthly Correlation
Oct 2022	July 2022	361,192	9,102	.01784	115,988	NA

The key metric that should be used to allocate care resources is the relative risk: no matter the absolute risk of the patient panel, the efficient allocation of care resources requires the identification of the riskiest patients. At the time of writing, we are unable to calculate the month-over-month stability of risk scores because only one month of Pre-HE Model™ risk scores has been released; future versions of this document will update this field.

Predictive Power

Validation Data Testing

Figure 10 shows the concentration curve for these scores on all-cause mortality events as of December 2021. We find that the top 10% riskiest patients account for approximately 60% of the all-cause mortality events in the following month.

Figure 10. Pre-HE MDPCP Concentration Curves as of December 2021

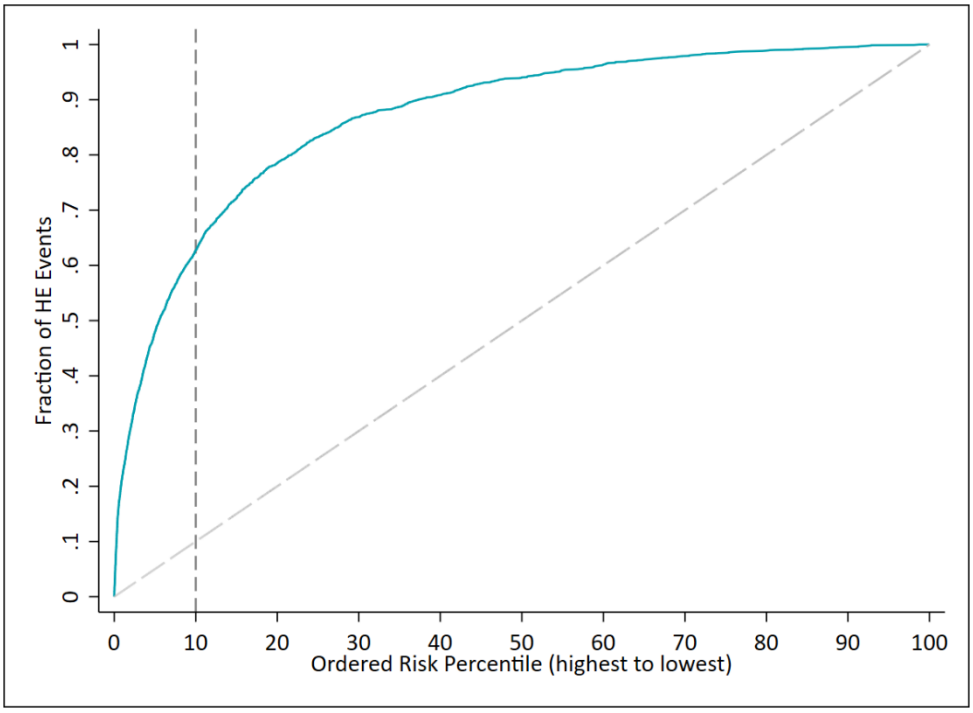
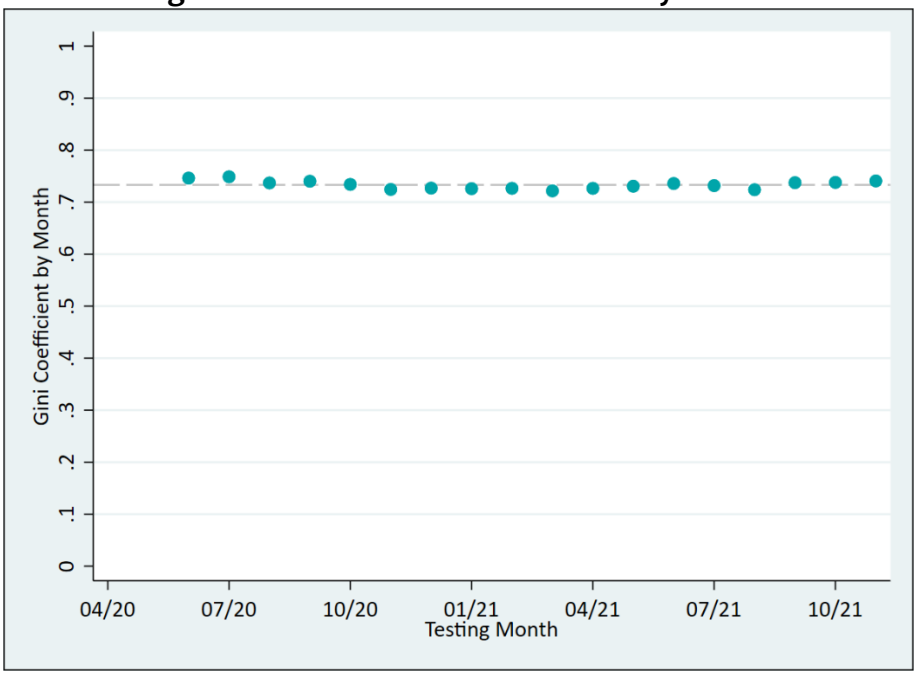


Figure 11 plots the Gini coefficients from 19 months of data based on the concentration curves estimated on the 20% holdout sample from each month. The three months of holdout data were not used due to incomplete information on mortality events due to data lag. These scores indicate that model performance is steady over time, with a very slight upward trend. This is consistent with the risk factors becoming more predictive as additional claims history becomes available.

Figure 11. Pre-HE MDPCP Gini Scores by Month



Production Data Testing

The risk scores based on the Pre-HE Model™ were released in October 2022; at the time of writing, data are not available to assess the predictive power in a production environment.

Return to Table of Contents

Section 4. Limitations

There are three main limitations of the Hilltop Pre- Models that are important to consider when implementing the models for guiding care coordination services: the timing lag, the lack of clinical risk factors, and the granularity of the environmental risk factors. These are discussed in detail below.

Timing Lag

Hilltop receives the Medicare CCLF claims and the Medicaid MMIS2 claims with a lag of over one month. Claims that arrive in late October 2019, for example, cover utilization through mid-September 2019. Hilltop uses these data to calculate risk factors based on utilization in September 2019 and then applies the risk model coefficients to estimate the risk of incurring an avoidable hospitalization in October 2019. These scores are then deployed in mid-November 2019 for use by providers and care managers. This raises two distinct, but related, issues:

- By providing the one-month predictions (in this example, predicting October 2019 events) to care managers over a two-month time horizon (here, in November 2019), the risk predictions may be "outdated" by the time they are used by care managers and providers.
- The risk predictions do not incorporate the most recent patient experience, which may degrade the quality of the risk scores.

Hilltop does not believe that this issue—the possibility that the risk score quality is impaired due to either being "outdated" or missing relevant recent information—substantively impacts the utility of our risk scores, for the following four reasons:

- The time-variant risk factors in the predictive models are all estimated with a look-back period of at least one year. This has the consequence that risk factors tend to change slowly, meaning that the risk scores also change slowly. As a result, there is high consistency of risk scores across months: patients that have high risk scores in October 2019 will also have high risk scores in November 2019.
- Internal testing has verified that applying one-month predictions on a two-month time horizon is substantively equivalent to directly estimating two-month predictions.
- Hilltop has determined that our predictive models have good discrimination and calibration in the production environment: applying the one-month predictions to a two-month horizon results in the identification of almost 45% of all avoidable hospital events in the riskiest 10% of patients.
- To the extent that structural factors determine the risk of incurring an avoidable hospital or COVID-19 hospitalization event, it is likely that high-risk *behavior* persists across time; that is, most individuals will not suddenly "become" high-risk in the interval between the most recent claims data and receipt of the risk scores by care managers and providers.

Finally, it is possible that improvements in revenue management systems will reduce the claims delivery lag so that, for example, risk scores estimated using September 2019 data can be provided to participating providers in early October 2019. In this case, the prediction time horizon would be one month (as in the current configuration of the models). Reconfiguring the model now to generate risk scores at a two-month horizon would entail development costs; then, should the time lag be reduced, reconfiguring the model again to generate risk scores at a one-month horizon would entail additional costs. Therefore, Hilltop will continue to estimate the next month's outcome events. However, we will continue to monitor this issue and update the models as needed.

Clinical Data

Administrative claims data do not include information on vital statistics—for example, blood pressure or lab results—meaning that Hilltop is unable to incorporate those clinical risk factors into our predictive models. It is likely that development of clinical risk factors would improve the predictive power of the models, although researchers have documented only relatively modest improvements in model performance for claims-and-clinical models relative to claims-only prediction models for heart failure patients (Hammill et al., 2011). Hilltop hopes to include this information in future versions of the model once the level of information exchange between electronic health records allows.

Environmental Risk Factors

To control for environmental factors in the Hilltop Pre- Models, we have developed a rich set of ZIP code-level and census tract-level covariates derived from publicly available sources. These data have two main limitations:

- 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. Hilltop plans, in the future, to use the address-level information available in the CCLF claims to disaggregate (and refresh) the area-level risk factors as much, and as frequently, as possible. Additionally, if available in the future, individual-level social welfare screening data will be added to provide a more robust individual-level risk prediction.
- ZCTA-level risk factors 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, ZIP code-level risk factors offer little predictive power. The census tract-level of the covariates are more granular; however, they are currently only available for MDPCP beneficiaries. Hilltop intends to extend the geocoding procedure necessary to link individual claims and census tract-level environmental risk factors to the HealthChoice population soon.

Return to Table of Contents

References

Alba, A. C., Agoritsas, T., Walsh, M., Hanna, S., Iorio, A., Devereaux, P. J., McGinn, T., & Guyatt, G. (2017). Discrimination and calibration of clinical prediction models: Users' guides to the Medical literature. *JAMA*, *318*(14), 1377–1384. https://doi.org/10.1001/jama.2017.12126

Andersen, R. M. (1995). Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior*, *36*(1), 1–10. https://doi.org/10.2307/2137284

Anderson, G. F., Ballreich, J., Bleich, S., Boyd, C., DuGoff, E., Leff, B., Salzburg, C., & Wolff, J. (2015). Attributes common to programs that successfully treat high-need, high-cost individuals. *The American Journal of Managed Care*, *21*(11), 4.

Bagherzadeh-Khiabani, F., Ramezankhani, A., Azizi, F., Hadaegh, F., Steyerberg, E. W., & Khalili, D. (2016). A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *Journal of Clinical Epidemiology*, *71*, 76–85. https://doi.org/10.1016/j.jclinepi.2015.10.002

Baker, J. M., Grant, R. W., & Gopalan, A. (2018). A systematic review of care management interventions targeting multimorbidity and high care utilization. *BMC Health Services Research*, *18*, 65. https://doi.org/10.1186/s12913-018-2881-8

Berkowitz, S. A., Parashuram, S., Rowan, K., Andon, L., Bass, E. B., Bellantoni, M., Brotman, D. J., Deutschendorf, A., Dunbar, L., Durso, S. C., Everett, A., Giuriceo, K. D., Hebert, L., Hickman, D., Hough, D. E., Howell, E. E., Huang, X., Lepley, D., Leung, C., … Brown, P. M. C. (2018). Association of a care coordination model with health care costs and utilization. *JAMA Network Open*, *1*(7), e184273. https://doi.org/10.1001/jamanetworkopen.2018.4273

Billings, J., Zeitel, L., Lukomnik, J., Carey, T. S., Blank, A. E., & Newman, L. (1993). Impact of socioeconomic status on hospital use in New York City. *Health Affairs*, *12*(1), 162–173. https://doi.org/10.1377/hlthaff.12.1.162

Brophy, J. M., Joseph, L., & Rouleau, J. L. (2001). β-blockers in congestive heart failure: A Bayesian meta-analysis. *Annals of Internal Medicine*, *134*(7), 550–560. https://doi.org/10.7326/0003-4819-134-7-200104030-00008

Center for Medicare & Medicaid Innovation. (2019). *MDPCP payment methodologies: Beneficiary attribution, care management fee, performance-based incentive payment, and comprehensive care payment. Version 1.0p*.

Center for Medicare and Medicaid Innovation. (2017). *CPC+ payment methodologies: Beneficiary attribution, care management fee, performance-based incentive payment, and payment under the Medicare physician fee schedule. Version 1*. https://innovation.cms.gov/files/x/cpcplus-methodology.pdf

Centers for Disease Control and Prevention. (2020). *National diabetes statistics report 2020.* (p. 32). U.S. Dept of Health and Human Services. https://www.cdc.gov/diabetes/data/statistics-report/index.html

Centers for Medicare and Medicaid Services. (2018). *Risk adjustment in Medicare Advantage* [Report to Congress]. Author. https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Downloads/RTC-Dec2018.pdf

Chang, H.-Y., Weiner, J. P., Richards, T. M., Bleich, S. N., & Segal, J. B. (2012). Validating the Adapted Diabetes Complications Severity Index in claims data. *The American Journal of Managed Care*, *18*(11), 10.

Din, A., & Wilson, R. (2020). Crosswalking ZIP codes to census geographies: Geoprocessing the U.S. Department of Housing & Urban Development's ZIP code crosswalk files. *Cityscape*, *22*(1). https://www.huduser.gov/portal/periodicals/cityscpe/vol22num1/ch12.pdf

Edwards, S. T., Peterson, K., Chan, B., Anderson, J., & Helfand, M. (2017). Effectiveness of intensive primary care interventions: A systematic review. *Journal of General Internal Medicine*, *32*(12), 1377–1386. https://doi.org/10.1007/s11606-017-4174-z

Furumoto, A., Ohkusa, Y., Chen, M., Kawakami, K., Masaki, H., Sueyasu, Y., Iwanaga, T., Aizawa, H., Nagatake, T., & Oishi, K. (2008). Additive effect of pneumococcal vaccine and influenza vaccine on acute exacerbation in patients with chronic lung disease. *Vaccine*, *26*(33), 4284–4289. https://doi.org/10.1016/j.vaccine.2008.05.037

Glasheen, W. P., Renda, A., & Dong, Y. (2017). Diabetes Complications Severity Index (DCSI)—Update and ICD-10 translation. *Journal of Diabetes and Its Complications*, *31*(6), 1007–1013. https://doi.org/10.1016/j.jdiacomp.2017.02.018

Hammill, B. G., Curtis, L. H., Fonarow, G. C., Heidenreich, P. A., Yancy, C. W., Peterson, E. D., & Hernandez, A. F. (2011). Incremental value of clinical data beyond claims data in predicting 30-day outcomes after heart failure hospitalization. *Circulation: Cardiovascular Quality and Outcomes*, *4*(1), 60–67. https://doi.org/10.1161/CIRCOUTCOMES.110.954693

Hedlund, J., Christenson, B., Lundbergh, P., & Örtqvist, Å. (2003). Effects of a large-scale intervention with influenza and 23-valent pneumococcal vaccines in elderly people: A 1-year follow-up. *Vaccine*, *21*(25–26), 3906–3911. https://doi.org/10.1016/S0264-410X(03)00296-2

The Hilltop Institute. (2020). *Evaluation of the Maryland Medicaid HealthChoice Program: CY 2014 to CY 2018*. Baltimore, MD: UMBC. https://hilltopinstitute.org/wp-content/uploads/publications/EvaluationOfTheHealthChoiceProgram-July2020.pdf

Hong, C. S., Siegel, A. L., & Ferris, T. G. (2014). Caring for high-need, high-cost patients: What makes for a successful care management program? *Issue Brief (Commonwealth Fund)*, *19*, 1–19.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366.

Kaiser Family Foundation. (2021). *Medicaid managed care penetration rates by eligibility group*. State Health Facts. https://www.kff.org/medicaid/state-indicator/managed-care-penetration-rates-by-eligibility-group/

Liaw, W., Moore, M., Iko, C., & Bazemore, A. (2015). Lessons for primary care from the first ten years of Medicare coordinated care demonstration projects. *The Journal of the American Board of Family Medicine*, *28*(5), 556–564. https://doi.org/10.3122/jabfm.2015.05.140322

Llorca, J., & Delgado-Rodríguez, M. (2002). Visualising exposure-disease association: The Lorenz curve and the Gini index. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, *8*(10), MT193-197.

Mauguen, A., & Begg, C. B. (2016). Using the Lorenz curve to characterize risk predictiveness and etiologic heterogeneity. *Epidemiology (Cambridge, Mass.)*, *27*(4), 531–537. https://doi.org/10.1097/EDE.0000000000000499

McCarthy, D., Ryan, J., & Klein, S. (2015). Models of Care for high-need, high-cost patients: An evidence synthesis. *Issue Brief (Commonwealth Fund)*, *31*, 1–19.

Moss, J. L., Johnson, N. J., Yu, M., Altekruse, S. F., & Cronin, K. A. (2021). Comparisons of individual- and area-level socioeconomic status as proxies for individual-level measures: Evidence from the Mortality Disparities in American Communities study. *Population Health Metrics*, *19*, 1. https://doi.org/10.1186/s12963-020-00244-x

Nichol, K. L., Nordin, J., Mullooly, J., Lask, R., Fillbrandt, K., & Iwane, M. (2003). Influenza vaccination and reduction in hospitalizations for cardiac disease and stroke among the elderly. *New England Journal of Medicine*, *348*(14), 1322–1332. https://doi.org/10.1056/NEJMoa025028

Office of Policy Development and Research. (2021). *HUD USPS ZIP code crosswalk files*. HUD User. https://www.huduser.gov/portal/datasets/usps_crosswalk.html#codebook

Pelser, C., Henderson, M., & Stockwell, I. (2019). *Risk factors for potentially avoidable hospital admissions and emergency department visits: A literature review*. The Hilltop Institute at UMBC. https://www.hilltopinstitute.org/wp-content/uploads/publications/MDPCPAvoidableHospitalizationLiteratureReview-April2019.pdf

Schrader, D., Haft, H., Perman, C., Sowinski-Rice, A., Bowden, S., Gruber, E., Anand, R., Hope, K., & Roa, J. (2021). *Maryland Primary Care Program (MDPCP) 2020 annual report*. Maryland Department of Health. https://health.maryland.gov/mdpcp/Documents/MDPCP_2020_Annual_Report.pdf

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology (Cambridge, Mass.)*, *21*(1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, *49*(11), 1225–1231. https://doi.org/10.1016/S0895-4356(96)00002-9

Walter, S., & Tiemeier, H. (2009). Variable selection: Current practice in epidemiological studies. *European Journal of Epidemiology*, *24*(12), 733–736. https://doi.org/10.1007/s10654-009-9411-2

Young, B. A., Lin, E., Von Korff, M., Simon, G., Ciechanowski, P., Ludman, E. J., Everson-Stewart, S., Kinder, L., Oliver, M., Boyko, E. J., & Katon, W. J. (2008). Diabetes Complications Severity Index and risk of mortality, hospitalization, and healthcare utilization. *The American Journal of Managed Care*, *14*(1), 15–23.

Appendix 1. Risk Factor Codebook

Age: For each person-month, this variable records person age as of the end of the month.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Age	74.048	21	109	24.175	0	65

Air pollution level: For each person, this variable records the average daily fine particulate matter (PM 2.5) concentration from the EPA's Downscaler Model for 2011-2015 in the person's Census Tract or ZCTA of residence.

Source: Environmental Protection Agency (2011-2015)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Air pollution level	11.244	0	192	80.46	0	267

Beneficiary gender - female: Beneficiary gender is female.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary gender - female	.592	0	1	.535	0	1

Beneficiary gender - male: Beneficiary gender is male.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary gender - male	.408	0	1	.465	0	1

Beneficiary race - Asian: Beneficiary's Research Triangle Institute (RTI) race code is Asian.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - Asian	.033	0	1	.058	0	1

Beneficiary race - Black: Beneficiary's Research Triangle Institute (RTI) race code is Black.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - Black	.198	0	1	.363	0	1

Beneficiary race - Hispanic: Beneficiary's Research Triangle Institute (RTI) race code is Hispanic.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - Hispanic	.023	0	1	.07	0	1

Beneficiary race - Native American: Beneficiary's Research Triangle Institute (RTI) race code is Native American.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - Native American	0	0	1	.005	0	1

Beneficiary race - Other: Beneficiary's Research Triangle Institute (RTI) race code is Other.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - Other	.008	0	1	.002	0	1

Beneficiary race - Unknown: Beneficiary's Research Triangle Institute (RTI) race code is Unknown.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - Unknown	.029	0	1	.258	0	1

Beneficiary race - White: Beneficiary's Research Triangle Institute (RTI) race code is White.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Beneficiary race - White	.708	0	1	.244	0	1

CCW indicator for ADHD, conduct disorders, and hyperkinetic syndrome: For each person-month, this variable records whether the person meets the CCW clinical criteria for ADHD, conduct disorders, and hyperkinetic syndrome. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for ADHD, conduct disorders, and hyperkinetic syndrome	.01	0	1	.06	0	1

CCW indicator for Alzheimer's disease: For each person-month, this variable records whether the person meets the CCW clinical criteria for Alzheimer's Disease. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for Alzheimer's disease	.015	0	1	0	0	1

CCW indicator for HIV/AIDS: For each person-month, this variable records whether the person meets the CCW clinical criteria for HIV/AIDS. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for HIV/AIDS	.005	0	1	.005	0	1

CCW indicator for Parkinson's Disease or secondary Parkinsonism: For each person-month, this variable records whether the person meets the CCW clinical criteria for Parkinson's Disease or secondary Parkinsonism. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for Parkinson's Disease or Secondary Parkinsonism	.014	0	1	0	0	1

CCW indicator for acute myocardial infarction: For each person-month, this variable records whether the person meets the CCW clinical criteria for acute myocardial infarction. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for acute myocardial infarction	.007	0	1	.001	0	1

CCW indicator for alcohol use disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for alcohol use disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for alcohol use disorders	.003	0	1	.025	0	1

CCW indicator for anemia: For each person-month, this variable records whether the person meets the CCW clinical criteria for anemia. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for anemia	.208	0	1	.048	0	1

CCW indicator for anxiety disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for anxiety disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for anxiety disorders	.176	0	1	.123	0	1

CCW indicator for asthma: For each person-month, this variable records whether the person meets the CCW clinical criteria for asthma. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for asthma	.081	0	1	.075	0	1

CCW indicator for atrial fibrillation and flutter: For each person-month, this variable records whether the person meets the CCW clinical criteria for atrial fibrillation and flutter. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for atrial fibrillation and flutter	.116	0	1	.003	0	1

CCW indicator for autism spectrum disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for autism spectrum disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for autism spectrum disorders	.004	0	1	.012	0	1

CCW indicator for benign prostatic hyperplasia: For each person-month, this variable records whether the person meets the CCW clinical criteria for benign prostatic hyperplasia. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for benign prostatic hyperplasia	.136	0	1	.003	0	1

CCW indicator for breast cancer: For each person-month, this variable records whether the person meets the CCW clinical criteria for breast cancer. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for breast cancer	.054	0	1	.002	0	1

CCW indicator for cataracts: For each person-month, this variable records whether the person meets the CCW clinical criteria for cataracts. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for cataracts	.315	0	1	.003	0	1

CCW indicator for cerebral palsy: For each person-month, this variable records whether the person meets the CCW clinical criteria for cerebral palsy. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for cerebral palsy	.003	0	1	.001	0	1

CCW indicator for chronic kidney disease: For each person-month, this variable records whether the person meets the CCW clinical criteria for chronic kidney disease. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for chronic kidney disease	.17	0	1	.01	0	1

CCW indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis: For each person-month, this variable records whether the person meets the CCW clinical criteria for chronic obstructive pulmonary disease (COPD) and bronchiectasis. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for chronic obstructive pulmonary disease (COPD) and bronchiectasis	.11	0	1	.014	0	1

CCW indicator for colorectal cancer: For each person-month, this variable records whether the person meets the CCW clinical criteria for colorectal cancer. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for colorectal cancer	.014	0	1	.001	0	1

CCW indicator for cystic fibrosis and other metabolic developmental disorders: For each person-month, this variable records whether the person meets the CCW

clinical criteria for cystic fibrosis and other metabolic developmental disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for cystic fibrosis and other metabolic developmental disorders	.007	0	1	.002	0	1

CCW indicator for depression, bipolar, and other depressive mood disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for depression, bipolar, or other depressive mood disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for depression, bipolar, and other depressive mood disorders	.183	0	1	.127	0	1

CCW indicator for diabetes: For each person-month, this variable records whether the person meets the CCW clinical criteria for diabetes. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for diabetes	.293	0	1	.042	0	1

CCW indicator for drug use disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for drug use disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for drug use disorders	.003	0	1	.064	0	1

CCW indicator for endometrial cancer: For each person-month, this variable records whether the person meets the CCW clinical criteria for endometrial cancer. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for endometrial cancer	.007	0	1	0	0	1

CCW indicator for epilepsy: For each person-month, this variable records whether the person meets the CCW clinical criteria for epilepsy. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for epilepsy	.023	0	1	.01	0	1

CCW indicator for fibromyalgia, chronic pain and fatigue: For each person-month, this variable records whether the person meets the CCW clinical criteria for fibromyalgia, chronic pain and fatigue. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for fibromyalgia, chronic pain and fatigue	.225	0	1	.048	0	1

CCW indicator for glaucoma: For each person-month, this variable records whether the person meets the CCW clinical criteria for glaucoma. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for glaucoma	.21	0	1	.009	0	1

CCW indicator for heart failure and non-ischemic heart disease: For each person-month, this variable records whether the person meets the CCW clinical criteria for heart failure and non-ischemic heart disease. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for heart failure and non-ischemic heart disease	.084	0	1	.007	0	1

CCW indicator for hip/pelvic fracture: For each person-month, this variable records whether the person meets the CCW clinical criteria for hip/pelvic fracture. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for hip/pelvic fracture	.008	0	1	.001	0	1

CCW indicator for hyperlipidemia: For each person-month, this variable records whether the person meets the CCW clinical criteria for hyperlipidemia. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for hyperlipidemia	.768	0	1	.073	0	1

CCW indicator for hypertension: For each person-month, this variable records whether the person meets the CCW clinical criteria for hypertension. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for hypertension	.743	0	1	.092	0	1

CCW indicator for hypothyroidism: For each person-month, this variable records whether the person meets the CCW clinical criteria for hypothyroidism. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for hypothyroidism	.209	0	1	.016	0	1

CCW indicator for intellectual disabilities and related conditions: For each person-month, this variable records whether the person meets the CCW clinical criteria for intellectual disabilities and related conditions. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for intellectual disabilities and related conditions	.009	0	1	.004	0	1

CCW indicator for ischemic heart disease: For each person-month, this variable records whether the person meets the CCW clinical criteria for ischemic heart disease. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for ischemic heart disease	.204	0	1	.009	0	1

CCW indicator for learning disabilities: For each person-month, this variable records whether the person meets the CCW clinical criteria for learning disabilities. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for learning disabilities	.002	0	1	.021	0	1

CCW indicator for leukemias and lymphomas: For each person-month, this variable records whether the person meets the CCW clinical criteria for leukemias and lymphomas. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for leukemias and lymphomas	.017	0	1	.001	0	1

CCW indicator for liver disease, cirrhosis and other liver conditions (except viral hepatitis): For each person-month, this variable records whether the person meets the CCW clinical criteria for liver disease, cirrhosis and other liver conditions (except viral hepatitis). If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for liver disease, cirrhosis and other liver conditions (except viral hepatitis)	.053	0	1	.015	0	1

CCW indicator for lung cancer: For each person-month, this variable records whether the person meets the CCW clinical criteria for lung cancer. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for lung cancer	.012	0	1	0	0	1

CCW indicator for migraine and chronic headache: For each person-month, this variable records whether the person meets the CCW clinical criteria for migraine and chronic headache. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for migraine and chronic headache	.037	0	1	.022	0	1

CCW indicator for mobility impairments: For each person-month, this variable records whether the person meets the CCW clinical criteria for mobility impairments. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for mobility impairments	.022	0	1	.003	0	1

CCW indicator for multiple sclerosis and transverse myelitis: For each person-month, this variable records whether the person meets the CCW clinical criteria for multiple sclerosis and transverse myelitis. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for multiple sclerosis and transverse myelitis	.006	0	1	.001	0	1

CCW indicator for muscular dystrophy: For each person-month, this variable records whether the person meets the CCW clinical criteria for muscular dystrophy. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for muscular dystrophy	0	0	1	0	0	1

CCW indicator for non-Alzheimer's dementia: For each person-month, this variable records whether the person meets the CCW clinical criteria for non-Alzheimer's dementia. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for non-Alzheimer's dementia	.046	0	1	.001	0	1

CCW indicator for obesity: For each person-month, this variable records whether the person meets the CCW clinical criteria for obesity. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for obesity	.247	0	1	.088	0	1

CCW indicator for osteoporosis with or without pathological fracture: For each person-month, this variable records whether the person meets the CCW clinical criteria for osteoporosis with or without pathological fracture. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for osteoporosis with or without pathological fracture	.125	0	1	.002	0	1

CCW indicator for other developmental delays: For each person-month, this variable records whether the person meets the CCW clinical criteria for other developmental delays. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for other developmental delays	.001	0	1	.01	0	1

CCW indicator for peripheral vascular disease: For each person-month, this variable records whether the person meets the CCW clinical criteria for peripheral vascular disease. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for peripheral vascular disease	.131	0	1	.004	0	1

CCW indicator for personality disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for personality disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for personality disorders	.01	0	1	.006	0	1

CCW indicator for pneumonia, all-cause: For each person-month, this variable records whether the person meets the CCW clinical criteria for pneumonia, all-cause. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for pneumonia, all-cause	.	.	.	.008	0	1

CCW indicator for post-traumatic stress disorder: For each person-month, this variable records whether the person meets the CCW clinical criteria for post-traumatic stress disorder. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for post-traumatic stress disorder	.01	0	1	.026	0	1

CCW indicator for pressure and chronic ulcers: For each person-month, this variable records whether the person meets the CCW clinical criteria for pressure and chronic ulcers. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for pressure and chronic ulcers	.035	0	1	.003	0	1

CCW indicator for prostate cancer: For each person-month, this variable records whether the person meets the CCW clinical criteria for prostate cancer. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for prostate cancer	.052	0	1	.001	0	1

CCW indicator for rheumatoid arthritis/osteoarthritis: For each person-month, this variable records whether the person meets the CCW clinical criteria for rheumatoid arthritis/osteoarthritis. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for rheumatoid arthritis/osteoarthritis	.391	0	1	.035	0	1

CCW indicator for schizophrenia and other psychotic disorders: For each person-month, this variable records whether the person meets the CCW clinical criteria for schizophrenia and other psychotic disorders. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for schizophrenia and other psychotic disorders	.015	0	1	.017	0	1

CCW indicator for sensory (blindness and visual) impairment: For each person-month, this variable records whether the person meets the CCW clinical criteria for sensory (blindness and visual) impairment. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for sensory (blindness and visual) impairment	.004	0	1	0	0	1

CCW indicator for sensory (deafness and hearing) impairment: For each person-month, this variable records whether the person meets the CCW clinical criteria for sensory (deafness and hearing) impairment. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for sensory (deafness and hearing) impairment	.089	0	1	.006	0	1

CCW indicator for spina bifida and other congenital anomalies of the nervous system: For each person-month, this variable records whether the person meets the CCW clinical criteria for spina bifida and other congenital anomalies of the nervous system. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for spina bifida and other congenital anomalies of the nervous system	.002	0	1	.002	0	1

CCW indicator for spinal cord injury: For each person-month, this variable records whether the person meets the CCW clinical criteria for spinal cord injury. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for spinal cord injury	.005	0	1	.001	0	1

CCW indicator for stroke/ischemic transient attack: For each person-month, this variable records whether the person meets the CCW clinical criteria for stroke/ischemic transient attack. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for stroke/ischemic transient attack	.064	0	1	.005	0	1

CCW indicator for tobacco use: For each person-month, this variable records whether the person meets the CCW clinical criteria for tobacco use. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for tobacco use	.069	0	1	.056	0	1

CCW indicator for traumatic brain injury and nonpsychotic mental disorders due to brain damage: For each person-month, this variable records whether the person meets the CCW clinical criteria for traumatic brain injury and nonpsychotic mental disorders due to brain damage. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for traumatic brain injury and nonpsychotic mental disorders due to brain damage	.003	0	1	.001	0	1

CCW indicator for urologic cancer: For each person-month, this variable records whether the person meets the CCW clinical criteria for urologic cancer. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for urologic cancer	.008	0	1	0	0	1

CCW indicator for viral hepatitis: For each person-month, this variable records whether the person meets the CCW clinical criteria for viral hepatitis. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
CCW indicator for viral hepatitis	.008	0	1	.007	0	1

Cardiovascular disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for cardiovascular disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Cardiovascular disease	NA	NA	NA	.02	0	1

Chronic Renal Insufficiency/ESRD: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for chronic renal insufficiency or ESRD within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Chronic Renal Insufficiency/ESRD	.003	0	1	.001	0	1

Coagulation Defect: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for coagulation defect within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Coagulation Defect	NA	NA	NA	.004	0	1

Continuity of primary care - Duration: For each person-month, this variable calculates the average time interval between primary care visits over the past 12 months. Visits that occur within 14 days are aggregated. Individuals with no primary care visits over the past 12 months are assigned a value of 365. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Continuity of primary care - Duration	88.518	14	365	221.01	16	365

Coronary Artery Disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for coronary artery disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Coronary Artery Disease	NA	NA	NA	.005	0	1

Cumulative Number of Days for Inpatient Stays: For each person-month, this variable records the number of days for inpatient hospital stays in the previous 12 months.

Source: Part A claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Cumulative Number of Days for Inpatient Stays	.832	0	220	NA	NA	NA

DCSI Score - Metabolic: For each person-month, this variable records the person's metabolic DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication; 2 = had at least 1 severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score - Metabolic	.011	0	2	NA	NA	NA

DCSI Score - Peripheral Vascular Disease: For each person-month, this variable records the person's PVD DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication; 2 = had at least 1 severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score - Peripheral Vascular Disease	.155	0	2	NA	NA	NA

DCSI Score – Cardiovascular: For each person-month, this variable records the person's cardiovascular DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication; 2 = had at least 1 severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score – Cardiovascular	.62	0	2	NA	NA	NA

DCSI Score – Cerebrovascular: For each person-month, this variable records the person's cerebrovascular DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication; 2 = had at least 1 severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score – Cerebrovascular	.226	0	2	NA	NA	NA

DCSI Score – Nephropathy: For each person-month, this variable records the person's nephropathy DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication; 2 = had at least 1 severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score – Nephropathy	.125	0	2	NA	NA	NA

DCSI Score – Neuropathy: For each person-month, this variable records the person's neuropathy DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score – Neuropathy	.131	0	1	NA	NA	NA

DCSI Score – Retinopathy: For each person-month, this variable records the person's retinopathy DCSI score over the past 12 months: 0 = had no complications; 1 = had at least 1 non-severe complication; 2 = had at least 1 severe complication.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
DCSI Score – Retinopathy	.129	0	2	NA	NA	NA

Diabetes Duration: For each person-month, this variable records the time since the person's first recorded diagnosis of diabetes.

Source: Beneficiary Demographics

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Diabetes Duration	20.513	0	92	NA	NA	NA

Discontinuity of primary care - Index: For each person-month, this variable calculates (1 - the continuity of care index), from Boxerman and Bice, 1977. This score ranges from 0 to 1 and is intended to measure dispersion in person-provider contact. If the person sees the same provider for all visits, indicating highly continuous care, the index score is 0; if the person sees a different physician for every visit, indicating fragmented care, the index score is 1. If a person has no primary care visits within the past year, they are assigned a value of 0. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Discontinuity of primary care - Index	.73	0	1	.282	0	1

Discontinuity of primary care - Proportion: For each person-month, this variable estimates (1 - the fraction of primary care visits within the past 12 months provided by the same provider). For example, if a person had 10 primary care visits over the past 12 months, and four visits were with the same provider, then this measure would take a value of (1 - .4) = .6. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Discontinuity of primary care - Proportion	.561	0	1	.545	0	1

General internists per 1000 residents: For each person, this variable records the number of general internists per 1000 residents in the Census Tract or ZCTA of residence.

Source: National Provider Identifier Database, American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
General internists per 1000 residents	.828	0	194	.805	0	33

Immunosuppressive disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for immunosuppressive disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Immunosuppressive disease	NA	NA	NA	.032	0	1

Indicator for Cancer of Bone and Connective Tissue: For each person-month, this variable records whether the person has had a diagnosis for cancer of the bone or connective tissue in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Bone and Connective Tissue	0	0	1	NA	NA	NA

Indicator for Cancer of Brain and Nervous System: For each person-month, this variable records whether the person has had a diagnosis for cancer of the brain or nervous system in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Brain and Nervous System	.001	0	1	NA	NA	NA

Indicator for Cancer of Bronchus; Lung: For each person-month, this variable records whether the person has had a diagnosis for cancer of the bronchus/lung in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Bronchus; Lung	.009	0	1	NA	NA	NA

Indicator for Cancer of Esophagus: For each person-month, this variable records whether the person has had a diagnosis for cancer of the esophagus in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Esophagus	.001	0	1	NA	NA	NA

Indicator for Cancer of Liver and Intrahepatic Bile Duct: For each person-month, this variable records whether the person has had a diagnosis for cancer of the liver or intrahepatic bile duct in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Liver and Intrahepatic Bile Duct	.001	0	1	NA	NA	NA

Indicator for Cancer of Ovary: For each person-month, this variable records whether the person has had a diagnosis for cancer of the ovary in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Ovary	.002	0	1	NA	NA	NA

Indicator for Cancer of Pancreas: For each person-month, this variable records whether the person has had a diagnosis for cancer of the pancreas in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Pancreas	.001	0	1	NA	NA	NA

Indicator for Cancer of Stomach: For each person-month, this variable records whether the person has had a diagnosis for cancer of the stomach in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Cancer of Stomach	.001	0	1	NA	NA	NA

Indicator for Hospital Bed Usage in DME: For each person-month, this variable records whether the person has a DME claim for a home hospital bed in the previous 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Hospital Bed Usage in DME	.008	0	1	NA	NA	NA

Indicator for Morphine Use: For each person-month, this variable records whether the person has received or been prescribed morphine in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A, B, and D claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Morphine Use	.038	0	1	NA	NA	NA

Indicator for Oxygen Usage in DME: For each person-month, this variable records whether the person has a DME claim for home oxygen therapy in the previous 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Oxygen Usage in DME	.083	0	1	NA	NA	NA

Indicator for Paraplegia or Hemiplegia: For each person-month, this variable records whether the person has had a diagnosis for paraplegia or hemiplegia in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Paraplegia or Hemiplegia	.012	0	1	NA	NA	NA

Indicator for Use of Fibrates: For each person-month, this variable records whether the person has been prescribed a fibrate in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Use of Fibrates	.016	0	1	NA	NA	NA

Indicator for Use of Insulin AND Another Glucose-Lowering Medication: For each person-month, this variable records whether the person has been prescribed both insulin AND another glucose-lowering drug within the same month in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for Use of Insulin AND Another Glucose-Lowering Medication	.021	0	1	NA	NA	NA

Indicator for albuminuria: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for albuminuria within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for albuminuria	.025	0	1	.003	0	1

Indicator for arrhythmia: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for arrhythmia within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for arrhythmia	.231	0	1	.033	0	1

Indicator for cerebrovascular disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for cerebrovascular disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for cerebrovascular disease	.131	0	1	.008	0	1

Indicator for cilostazol use: For each person-month, this variable takes the value of 1 if a person incurred a claim for cilostazol within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for cilostazol use	.002	0	1	0	0	1

Indicator for diabetes with complications: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with

any diagnosis for diabetes with complications within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for diabetes with complications	.198	0	1	.025	0	1

Indicator for diabetic foot procedure: For each person-month, this variable takes the value of 1 if a person incurred an inpatient diabetic foot procedure over the last 12 months and 0 otherwise.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for diabetic foot procedure	.001	0	1	0	0	1

Indicator for diabetic ulcer: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for diabetic ulcer within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for diabetic ulcer	.044	0	1	.002	0	1

Indicator for difficulty with life management: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for difficulty with life management within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for difficulty with life management	.003	0	1	.001	0	1

Indicator for dual eligibility with Medicaid: For each person-month, this variable takes the value of 1 if a person was dually eligible for both Medicaid and Medicare within the past 12 months, and 0 otherwise.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for dual eligibility with Medicaid	.131	0	1	0	0	1

Indicator for durable medical equipment (DME) use: For each person-month, this variable takes the value of 1 if a person used any durable medical equipment in the previous 12 months, and 0 otherwise.

Source: Part B DME claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for durable medical equipment (DME) use	.283	0	1	.02	0	1

Indicator for endocrinologist visit: For each person-month, this variable takes the value of 1 if a person visited an endocrinologist within the past 12 months, and 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for endocrinologist visit	.085	0	1	.013	0	1

Indicator for fluid and electrolyte imbalance: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for fluid and electrolyte imbalance within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for fluid and electrolyte imbalance	.14	0	1	.036	0	1

Indicator for frailty: For each person-month, this variable takes the value of 1 if a person meets the definition for frailty within the past twelve months, and 0 otherwise. The clinical definition for frailty is derived from Kim and Schneeweiss 2014.

Source: Part A, B, and Part B DME claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for frailty	.35	0	1	.072	0	1

Indicator for gastroesophageal reflux disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for gastroesophageal reflux disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for gastroesophageal reflux disease	.232	0	1	.039	0	1

Indicator for gastroparesis: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for gastroparesis within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for gastroparesis	.004	0	1	.001	0	1

Indicator for having received chemotherapy: For each person-month, this variable records whether the person has received chemotherapy in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for having Received Chemotherapy	.013	0	1	NA	NA	NA

Indicator for having received dialysis: For each person-month, this variable records whether the person has been on dialysis in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for having Received Dialysis	.006	0	1	NA	NA	NA

Indicator for having diabetes and being over 40 years old.: For each person-month, this variable takes the value of 1 if a person meets the CCW clinical criteria for diabetes within the past 12 months and is over the age of 40 years, and 0 otherwise.

Source: Part A and B claims and Beneficiary Demographics

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for having diabetes and being over 40 years old.	NA	NA	NA	.01	0	1

Indicator for hospice enrollment: For each person-month, this variable takes the value of 1 if a person enrolled in hospice within the past 12 months, and 0 otherwise.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for hospice enrollment	.002	0	1	NA	NA	NA

Indicator for immunosuppressive drug use: For each person-month, this variable takes the value of 1 if a person incurred a claim for an immunosuppressive drug within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for immunosuppressive drug use	NA	NA	NA	.16	0	1

Indicator for insulin use: For each person-month, this variable takes the value of 1 if a person incurred a claim for insulin within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for insulin use	.039	0	1	.013	0	1

Indicator for leukotriene receptor modifier use: For each person-month, this variable takes the value of 1 if a person incurred a claim for leukotriene receptor modifiers within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for leukotriene receptor modifier use	.037	0	1	.023	0	1

Indicator for lifestyle problems: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for lifestyle problems within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for lifestyle problems	.026	0	1	.015	0	1

Indicator for metastatic cancer: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for metastatic cancer within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for metastatic cancer	.018	0	1	.001	0	1

Indicator for neuropathy: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for neuropathy within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for neuropathy	.063	0	1	.005	0	1

Indicator for no VA clinic or VA medical center: For each person, this variable records whether the person's Census Tract or ZCTA of residence does not contain at least one VA clinic or medical center.

Source: Veterans Affairs Facility Listing

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for no VA clinic or VA medical center	.014	0	1	.92	0	1

Indicator for anti-diabetes medication use: For each person-month, this variable takes the value of 1 if a person incurred a claim for anti-diabetes medication within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for anti-diabetes medication use	.135	0	1	.967	0	1

Indicator for beta blocker use: For each person-month, this variable takes the value of 1 if a person incurred a claim for beta blockers within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for beta blocker use	.257	0	1	.965	0	1

Indicator for no federally qualified health center: For each person, this variable records whether the person's Census Tract or ZCTA of residence does not contain at least one active federally qualified health center.

Source: CMS Provider of Service Files (December 2020)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for no federally qualified health center	.072	0	1	.598	0	1

Indicator for losartan use: For each person-month, this variable takes the value of 1 if a person incurred a claim for losartan within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for losartan use	.131	0	1	.979	0	1

Indicator for no mental health center: For each person, this variable records whether the person's Census Tract or ZCTA of residence does not contain at least one active community mental health center.

Source: CMS Provider of Service Files (December 2020)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for no mental health center	.021	0	1	.897	0	1

Indicator for mental health use: For each person-month, this variable takes the value of 1 if a person incurred a visit with a mental health professional over the past 12 months, and 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for mental health use	.05	0	1	.899	0	1

Indicator for no rural health clinic: For each person, this variable records whether the person's Census Tract or ZCTA of residence does not contain at least one active rural health clinic.

Source: CMS Provider of Service Files (December 2020)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for no rural health clinic	.002	0	1	.996	0	1

Indicator for statin use: For each person-month, this variable takes the value of 1 if a person incurred a claim for statins within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for statin use	.445	0	1	.93	0	1

Indicator for no vaccination (flu or pneumonia): For each person-month, this variable takes the value of 1 if a person did not receive a vaccination (flu or pneumonia) within the past 12 months, 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for no vaccination (flu or pneumonia)	.669	0	1	.879	0	1

Indicator for occupational exposure to risk factors: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for occupational exposure to risk factors within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for occupational exposure to risk factors	0	0	1	0	0	1

Indicator for oncologist visit: For each person-month, this variable takes the value of 1 if a person visited an oncologist within the past 12 months, and 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for oncologist visit	.123	0	1	.005	0	1

Indicator for oral antibiotic use: For each person-month, this variable takes the value of 1 if a person incurred a claim for oral antibiotics within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for oral antibiotic use	.348	0	1	.321	0	1

Indicator for oral corticosteroid use: For each person-month, this variable takes the value of 1 if a person incurred a claim for oral corticosteroids within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for oral corticosteroid use	.149	0	1	.073	0	1

Indicator for original Medicare eligibility for a non-age-related cause: Beneficiary was originally eligible for Medicare for a reason other than age.

Source: Beneficiary Demographics

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for original Medicare eligibility for a non-age-related cause	.152	0	1	NA	NA	NA

Indicator for other problems with primary support group: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for other problems with primary support group within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for other problems with primary support group	.006	0	1	.005	0	1

Indicator for pancreatitis: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for pancreatitis within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for pancreatitis	.017	0	1	.003	0	1

Indicator for peptic ulcer disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for peptic ulcer disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for peptic ulcer disease	.01	0	1	.002	0	1

Indicator for peripheral and visceral atherosclerosis: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for peripheral and visceral atherosclerosis within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for peripheral and visceral atherosclerosis	.117	0	1	.004	0	1

Indicator for pneumonia: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for pneumonia within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for pneumonia	.032	0	1	.013	0	1

Indicator for presence of a for-profit hospital: For each person, this variable records whether the person's Census Tract or ZCTA of residence contains at least one active (short term or critical access or transplant) for-profit hospital.

Source: CMS Provider of Service Files (December 2020)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for presence of a for-profit hospital	.101	0	1	.427	0	1

Indicator for previous conservative diabetic wound procedure: For each person-month, this variable takes the value of 1 if a person underwent at least one conservative diabetic procedure within the past 12 months, and 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for previous conservative diabetic wound procedure	.01	0	1	.001	0	1

Indicator for prior nursing home stay: For each person-month, this variable takes the value of 1 if a person incurred a nursing home stay within the last 12 months, and 0 otherwise.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for prior nursing home stay	.026	0	1	.001	0	1

Indicator for prior readmission: For each person-month, this variable takes the value of 1 if a person incurred an all-cause 30-day hospital readmission within the last 12 months, and 0 otherwise. We define readmission as two inpatient stays occurring fewer than 30 days apart.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for prior readmission	.019	0	1	.01	0	1

Indicator for prior surgery: For each person-month, this variable takes the value of 1 if a person underwent a surgery within the past 12 months, and 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for prior surgery	.644	0	1	NA	NA	NA

Indicator for problems with care provider dependency: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for problems with care provider dependency within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for problems with care provider dependency	.082	0	1	.002	0	1

Indicator for problems with education and literacy: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for problems with education and literacy within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for problems with education and literacy	0	0	1	.002	0	1

Indicator for problems with employment and unemployment: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for problems with employment and unemployment within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for problems with employment and unemployment	.001	0	1	.003	0	1

Indicator for problems with housing and economic conditions: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for problems with housing and economic conditions within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for problems with housing and economic conditions	.003	0	1	.013	0	1

Indicator for problems with social environment: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for problems with social environment within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for problems with social environment	.005	0	1	.001	0	1

Indicator for problems with upbringing: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for problems with upbringing within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for problems with upbringing	0	0	1	.003	0	1

Indicator for protein-calorie malnutrition: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for protein-calorie malnutrition within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for protein-calorie malnutrition	.015	0	1	.003	0	1

Indicator for provider administered drug: For each person-month, this variable takes the value of 1 if a person received a provider-administered drug as defined by a 'J code' in the past 12 months, and 0 otherwise.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for provider administered drug	.265	0	1	.049	0	1

Indicator for psychosocial problems: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for psychosocial problems within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for psychosocial problems	0	0	1	.002	0	1

Indicator for pulmonary circulatory disorder: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for pulmonary circulatory disorder within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for pulmonary circulatory disorder	.042	0	1	.004	0	1

Indicator for respiratory infection: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for respiratory infection within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for respiratory infection	.138	0	1	.203	0	1

Indicator for retinopathy: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for retinopathy within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for retinopathy	.002	0	1	0	0	1

Indicator for rheumatoid arthritis/collagen vascular disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for rheumatoid arthritis/collagen vascular disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for rheumatoid arthritis/collagen vascular disease	.068	0	1	.008	0	1

Indicator for rivaroxaban use: For each person-month, this variable takes the value of 1 if a person incurred a claim for rivaroxaban within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for rivaroxaban use	.021	0	1	.001	0	1

Indicator for sepsis: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for sepsis within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for sepsis	.03	0	1	.008	0	1

Indicator for sickle cell anemia: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for sickle cell anemia within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for sickle cell anemia	0	0	1	.001	0	1

Indicator for sleep apnea: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for sleep apnea within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for sleep apnea	.17	0	1	.036	0	1

Indicator for solid tumor without metastasis: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for solid tumor without metastasis within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for solid tumor without metastasis	.132	0	1	.006	0	1

Indicator for urinary tract infection: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for urinary tract infection within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for urinary tract infection	.112	0	1	.03	0	1

Indicator for use of Anti-Hypertensive Treatment: For each person-month, this variable records whether the person has been prescribed an anti-hypertensive in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for use of Anti-Hypertensive Treatment	.473	0	1	NA	NA	NA

Indicator for use of Anticoagulants: For each person-month, this variable records whether the person has been prescribed an anticoagulant in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for use of Anticoagulants	.085	0	1	NA	NA	NA

Indicator for use of Sulfonylureas: For each person-month, this variable records whether the person has been prescribed a sulfonylurea in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for use of Sulfonylureas	.041	0	1	NA	NA	NA

Indicator for use of Thiazolidinediones: For each person-month, this variable records whether the person has been prescribed a thiazolidinedione or glitazone in the past 12 months. If so, this variable takes the value 1; if not, then 0.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for use of Thiazolidinediones	.008	0	1	NA	NA	NA

Indicator for warfarin use: For each person-month, this variable takes the value of 1 if a person incurred a claim for warfarin within the past 12 months, and 0 otherwise.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Indicator for warfarin use	.015	0	1	.001	0	1

Interaction of ADRD and Frailty Index: For each person-month, this variable records the interaction between whether a person has a dementia diagnosis AND their frailty index score.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Interaction of ADRD and Frailty Index	.004	0	1	NA	NA	NA

Interstitial lung disease: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for interstitial lung disease within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Interstitial lung disease	NA	NA	NA	0	0	1

Located in partial county mental health care shortage area: For each person, this variable takes the value of 1 if the person's Census Tract or ZCTA of residence is located in a county that is designated by HRSA in 2018 to be a partial-county mental health care shortage area. The variable takes the value of 0, otherwise. If the census tract lies in two counties, the value is estimated as a weighted average of the county-level attributes.

Source: Area Health Resources File

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Located in partial county mental health care shortage area	.68	0	1	.816	0	1

Located in partial county primary care shortage area: For each person, this variable takes the value of 1 if the person's Census Tract or ZCTA of residence is located in a county that is designated by HRSA in 2018 to be a partial-county primary care shortage area. The variable takes the value of 0, otherwise. If the census tract lies in two or more counties, the value is estimated as a weighted average of the county-level attributes.

Source: Area Health Resources File

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Located in partial county primary care shortage area	.892	0	1	.958	0	1

Located in whole county mental health care shortage area: For each person, this variable takes the value of 1 if the person's Census Tract or ZCTA of residence is located in a county that is designated by HRSA in 2018 to be a whole-county mental health care shortage area. The variable takes the value of 0, otherwise. If the census tract lies in two or more counties, the value is estimated as a weighted average of the county-level attributes.

Source: Area Health Resources File

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Located in whole county mental health care shortage area	.174	0	1	.145	0	1

Located in whole county primary care shortage area: For each person, this variable takes the value of 1 if the person's Census Tract or ZCTA of residence is located in a county that is designated by HRSA in 2018 to be a whole-county primary care shortage area. The variable takes the value of 0, otherwise. If the census tract lies in two or more counties, the value is estimated as a weighted average of the county-level attributes.

Source: Area Health Resources File

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Located in whole county primary care shortage area	.001	0	1	0	0	1

Median household income: For each person, this variable records the median household income in the person's Census Tract or ZCTA of residence (pooled from 2015-2019).

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Median household income	95270.664	0	250000	78049.164	0	250000

National ranking of deprivation: For each person, this variable records the national ranking of deprivation for the person's Census Tract or ZCTA of residence. This index 'includes factors for the theoretical domains of income, education, employment, and housing quality.' See https://www.neighborhoodatlas.medicine.wisc.edu/ for additional detail. Higher values indicate a greater degree of deprivation.

Source: Neighborhood Atlas

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
National ranking of deprivation	31.387	1	100	39.886	1	97

Number of HbA1c tests: For each person-month, this variable counts the number of visits within the past 12 months in which a person received a Hemoglobin A1C (HbA1c) test. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of HbA1c tests	.761	0	13	.034	0	9

Number of Previous Severe Type 2 Diabetes Complications: For each person-month, this variable records the number of severe type-2 diabetes complications in the previous 12 months.

Source: Part A and B claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of Previous Severe Type 2 Diabetes Complications	.143	0	53	NA	NA	NA

Number of avoidable hospitalizations: For each person-month, this variable counts the number of avoidable hospitalizations incurred within the prior 12 months (not including the month in which the avoidable hospitalization occurred).

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of avoidable hospitalizations	.05	0	17	.025	0	32

Number of emergency department visits within the past 6 months: For each person-month, this variable counts the number of emergency department visits incurred within the prior 6 months.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of emergency department visits within the past 6 months	.163	0	68	.171	0	81

Number of heart-related procedures: For each person-month, this variable counts the number of heart-related procedures incurred over the past year.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of heart-related procedures	.021	0	11	.004	0	18

Number of home health visits: For each person-month, this variable counts the number of home health visits incurred within the past 12 months. We apply a logarithmic transformation to non-zero values. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of home health visits	.063	0	6	.035	0	6

Number of hospital beds per 1000 residents: For each person, this variable records the number of active (short term or critical access or transplant) hospital beds per 1000 residents in the person's Census Tract or ZCTA of residence.

Source: CMS Provider of Service Files (December 2020) American Community Survey (2017, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of hospital beds per 1000 residents	3.462	0	1045	3.972	0	2349

Number of hospitals: For each person, this variable records the number of active (short term or critical access or transplant) hospitals in the person's Census Tract or ZCTA of residence.

Source: CMS Provider of Service Files (December 2020)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of hospitals	.093	0	7	.65	0	10

Number of hospitals per 1000 residents: For each person, this variable records the number of active (short term or critical access or transplant) hospitals per 1000 residents in the person's Census Tract or ZCTA of residence.

Source: CMS Provider of Service Files (December 2020) American Community Survey (2017, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of hospitals per 1000 residents	.018	0	5	.02	0	5

Number of lab tests: For each person-month, this variable counts the number of visits within the past 12 months in which a person received any laboratory test. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of lab tests	.173	0	26	.017	0	14

Number of medications: For each person-month, this variable counts the number of distinct medications (as measured by NDC codes) for which there are part D claims within the past 12 months.

Source: Part D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of medications	8.049	0	104	4.497	0	148

Number of outpatient visits: For each person-month, this variable counts the number of visits in an outpatient setting incurred within the past 12 months. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of outpatient visits	17.483	0	251	7.146	0	676

Number of previous COVID hospitalizations: For each person-month, this variable counts the number of covid-related hospitalizations incurred within the prior 12 months (not including the month in which the covid hospitalization occurred).

Source: Part A claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of previous COVID hospitalizations	NA	NA	NA	0	0	1

Number of primary care physicians per 1000 residents: For each person, this variable records the number of primary care physicians per 1000 residents in the person's Census Tract or ZCTA of residence.

Source: National Provider Identifier Database, American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of primary care physicians per 1000 residents	1.761	0	375	1.674	0	42

Number of primary care visits: For each person-month, this variable counts the number of primary care visits within the past 12 months. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of primary care visits	10.768	0	220	4.249	0	334

Number of prior admissions: For each person-month, this variable counts the number of all inpatient hospital admissions incurred within the past twelve months.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of prior admissions	.158	0	24	.08	0	24

Number of rural clinic visits: For each person-month, this variable counts the number of rural clinic visits incurred within the past 12 months. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of rural clinic visits	0	0	2	.001	0	14

Number of specialist visits: For each person-month, this variable counts the number of specialist visits incurred within the past 12 months. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of specialist visits	5.415	0	337	.469	0	168

Number of specialty care physicians per 1000 residents: For each person, this variable records the number of specialty care physicians per 1000 residents in the person's Census Tract or ZCTA of residence.

Source: National Provider Identifier Database, American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of specialty care physicians per 1000 residents	1.878	0	637	1.777	0	50

Number of urgent care visits: For each person-month, this variable counts the number of urgent care visits incurred within the past 12 months. We define visits as unique combinations of person-provider-day.

Source: Part B claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Number of urgent care visits	.226	0	31	.485	0	86

Part D OOP spending: For each person-month, this variable records the total amount of out-of-pocket spending for prescriptions in the previous 12 months.

Source: Part D claims

Models: Pre-DC

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Part D OOP spending	322.03	0	41517	NA	NA	NA

Percent Hispanic, ages 65+: For each person, this variable records the percent of the population aged 65 and above in the person's Census Tract or ZCTA of residence that is Hispanic.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent Hispanic, ages 65+	2.581	0	68	3.69	0	79

Percent Native American: For each person, this variable records the percent of the population in the person's Census Tract or ZCTA of residence that is Native American.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent Native American	.953	0	29	1.069	0	29

Percent aged 0-4: For each person, this variable records the percentage of individuals in the person's Census Tract or ZCTA of residence aged 0-4 (pooled from 2013-2017).

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent aged 0-4	5.562	0	35	6.347	0	35

Percent aged 65 and over: For each person, this variable records the percentage of individuals in the person's Census Tract or ZCTA of residence aged 65 and over (pooled from 2013-2017).

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent aged 65 and over	17.854	0	100	14.567	0	100

Percent foreign born: For each person, this variable records the percent of individuals who are foreign-born in the person's Census Tract or ZCTA of residence.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent foreign born	11.621	0	66	15.335	0	57

Percent in poverty: For each person, this variable records the percentage of families whose income in the past 12 months is below the poverty level in the person's Census Tract or ZCTA of residence (pooled from 2015-2019).

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent in poverty	5.637	0	56	8.316	0	62

Percent in poverty age 65+: For each person, this variable records the percentage of people age 65+ whose income in the past 12 months is below the poverty level in the person's Census Tract or ZCTA of residence (pooled from 2015-2019).

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent in poverty age 65+	3.968	0	100	5.257	0	100

Percent live alone, ages 65+: For each person, this variable records the percent of the population aged 65 and above in the person's Census Tract or ZCTA of residence that lives alone.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent live alone, ages 65+	11.737	0	66	10.713	0	46

Percent married: For each person, this variable records the percent of the population aged 15+ in the person's Census Tract or ZCTA of residence that is currently married (pooled from 2013-2017).

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent married	50.629	0	100	43.44	0	100

Percent non-English speakers: For each person, this variable records the percent of individuals who speak Spanish or other languages and who speak English less than 'very well' in the person's Census Tract or ZCTA of residence.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent non-English speakers	31.396	0	100	36.311	0	100

Percent non-white, ages 65+: For each person, this variable records the percent of the population aged 65 and above in the person's Census Tract or ZCTA of residence that is non-white.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent non-white, ages 65+	26.198	0	100	41.158	0	100

Percent of population in college group quarters - ZCTA: For each person, this variable records the percent of the 2010 population living in university group quarters in the person's ZIP code tabulation area of residence.

Source: 2010 Census

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent of population in college group quarters - ZCTA	NA	NA	NA	7.372	0	390300

Percent of population in nursing home group quarters - ZCTA: For each person, this variable records the percent of the 2010 population living in nursing facility group quarters in the person's ZIP code tabulation area of residence.

Source: 2010 Census

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent of population in nursing home group quarters - ZCTA	NA	NA	NA	.493	0	280

Percent of population that is Black - ZCTA: For each person, this variable records the percent of the 2018 population that is Black in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent of population that is Black - ZCTA	NA	NA	NA	36.982	0	97

Percent of population that is male - ZCTA: For each person, this variable records the percent of individuals in the 2018 population that is male in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent of population that is male - ZCTA	NA	NA	NA	48.149	0	100

Percent of units with 0 or 1 bedrooms - ZCTA: For each person, this variable records the percent of occupied housing units with 0 bedrooms or 1 bedroom in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent of units with 0 or 1 bedrooms - ZCTA	NA	NA	NA	12.334	0	81

Percent single mothers: For each person, this variable records the percent of women aged 15-50 giving birth within the past 12 months who are not married in the person's Census Tract or ZCTA of residence.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent single mothers	25.631	0	100	37.918	0	100

Percent speak Spanish, aged 65+: For each person, this variable records the percent of the population aged 65 and above in the person's Census Tract or ZCTA of residence that speaks Spanish.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent speak Spanish, aged 65+	2.3	0	57	3.386	0	76

Percent travel by car - ZCTA: For each person, this variable records the percent of total workers age 16+ who travel to work in a car alone or in a carpool in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent travel by car - ZCTA	NA	NA	NA	81.949	0	100

Percent travel by public transit - ZCTA: For each person, this variable records the percent of total workers age 16+ who travel to work by public transportation in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent travel by public transit - ZCTA	NA	NA	NA	10.087	0	100

Percent with a commute of 60 mins or more - ZCTA: For each person, this variable records the percent of workers age 16+ who did not work at home who have travel time to work 60 minute or more in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent with a commute of 60 mins or more - ZCTA	NA	NA	NA	15.188	0	100

Percent with less than high school education: For each person, this variable records the percent of individuals aged 18 and older with less than a high school diploma in the person's Census Tract or ZCTA of residence.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent with less than high school education	19.399	0	167	11.998	0	62

Percent with less than high school education, ages 65+: For each person, this variable records the percent of the population aged 65 and above in the person's Census Tract or ZCTA of residence that has less than a high school diploma.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Percent with less than high school education, ages 65+	13.298	0	100	17.11	0	100

Physician diversity: For each person, this variable records the percentage of medical doctors who are minorities (African Americans, Hispanics, and others, but excluding Asian-Americans). If the ZIP code tabulation area lies in two or more counties, the value is estimated as a weighted average of the county-level attributes, with weights being the fraction of the ZCTA population residing within each county.

Source: American Community Survey (2019, individual)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Physician diversity	21.904	0	1401	175.56	0	1707

Population: For each person, this variable records the population of the person's Census Tract or ZCTA of residence.

Source: American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Population	5689.727	0	70749	34191.477	0	107673

Population density: For each person, this variable records the population per square mile in the person's Census Tract or ZCTA of residence.

Source: American Community Survey (2019, 5-year estimates), Census

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Population density	3292.46	0	55433	3786.696	0	122674

Population growth: For each person, this variable records the percent population growth recorded in the person's Census Tract or ZCTA of residence from 2013 - 2019.

Source: American Community Survey (2011 and 2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Population growth	3.777	-100	233	4.952	-100	685

Population per household - ZCTA: For each person, this variable records the total population divided by total number of housing units in the person's ZIP code tabulation area of residence.

Source: American Community Survey (2018, 5-year estimates)

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Population per household - ZCTA	NA	NA	NA	2.488	0	87

Prior admission length of stay: For each person-month, this variable calculates the length of the most recently incurred hospital inpatient stay over the past 12 months. For individuals without a previous inpatient stay, this value is set to zero.

Source: Part A claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior admission length of stay	.505	0	132	.227	0	1200

Prior hospitalization admission source - none: For each person-month, this variable indicates the individual did not incur an inpatient hospital stay within the past 12 month.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission source - none	.895	0	1	.94	0	1

Prior hospitalization admission source - other: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission source was: other.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission source - other	.002	0	1	0	0	0

Prior hospitalization admission source - physician referral: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission source was: physician referral.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission source - physician referral	.09	0	1	.023	0	1

Prior hospitalization admission source - transferred from facility: For each person-month, this variable indicates that for the individual's most recently incurred inpatient

hospital stay within the past 12 months, the individual's admission source was: transferred from facility.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission source - transferred from facility	.013	0	1	.036	0	1

Prior hospitalization admission type - elective: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission type was: elective.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission type - elective	.025	0	1	.01	0	1

Prior hospitalization admission type - emergency: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission type was: emergency.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission type - emergency	.074	0	1	.019	0	1

Prior hospitalization admission type - none: For each person-month, this variable indicates the individual did not incur an inpatient hospital stay within the past 12 month.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission type - none	.895	0	1	.94	0	1

Prior hospitalization admission type - other: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission type was: other.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission type - other	0	0	1	.026	0	1

Prior hospitalization admission type - trauma center: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission type was: trauma center.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission type - trauma center	.001	0	1	0	0	1

Prior hospitalization admission type - urgent: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's admission type was: urgent.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization admission type - urgent	.004	0	1	.005	0	1

Prior hospitalization discharge status - home: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's discharge status was: home.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization discharge status - home	.084	0	1	.057	0	1

Prior hospitalization discharge status - none: For each person-month, this variable indicates the individual did not incur an inpatient hospital stay within the past 12 month.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization discharge status - none	.895	0	1	.94	0	1

Prior hospitalization discharge status - other: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's discharge status was: other.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization discharge status - other	.001	0	1	.001	0	1

Prior hospitalization discharge status - transferred to inpatient care: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's discharge status was: transferred to inpatient care.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization discharge status - transferred to inpatient care	0	0	0	0	0	1

Prior hospitalization discharge status - transferred to post-acute care: For each person-month, this variable indicates that for the individual's most recently incurred inpatient hospital stay within the past 12 months, the individual's discharge status was: transferred to post-acute care.

Source: Part A Claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Prior hospitalization discharge status - transferred to post-acute care	0	0	0	.001	0	1

Pure Hypercholesterolemia: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for pure hypercholesterolemia within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Pure Hypercholesterolemia	NA	NA	NA	.009	0	1

Recent Increase in Frailty severity: For each person-month, this variable records whether the person's claims-based frailty index score has increased compared to the previous month. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Recent Increase in Frailty severity	.267	0	1	NA	NA	NA

Rurality index: For each person, this variable records the rural/urban index for the person's Census Tract or ZCTA of residence. This data is comprised of 10 codes which "delineate metropolitan, micropolitan, small town, and rural commuting areas based on the size and direction of the primary (largest) commuting flows." Higher values indicate a greater degree of rurality.

Source: USDA Rural-Urban Commuting Area Codes

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Rurality index	1.497	1	10	1.295	1	10

Severity of Frailty: For each person-month, this variable records each patient's claims-based frailty index (CFI) score using claims from the previous 12 months. CFI calculated using methods detailed in Gautam et al., 2020, Journals of Gerontology: Medical Sciences.

Source: Part A and B claims

Models: Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Severity of Frailty	.16	0	1	NA	NA	NA

Social workers per 1000 residents: For each person, this variable records the number of social workers per 1000 residents in the Census Tract or ZCTA of residence.

Source: National Provider Identifier Database, American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Social workers per 1000 residents	1.754	0	107	1.728	0	144

Solid Organ Cancer: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for solid organ cancer within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Solid Organ Cancer	NA	NA	NA	0	0	1

Taxable interest per capita: For each person, this variable records taxable interest (tax year 2018) per person in the person's Census Tract or ZCTA of residence.

Source: IRS Statistics of Income and American Community Survey (2019, 5-year estimates)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Taxable interest per capita	1629.439	0	122491	5892.085	0	127119

Thrombocytopenia: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for thrombocytopenia within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Thrombocytopenia	NA	NA	NA	.002	0	1

Total health spending: For each person-month, this variable measures the total health spending incurred within the past 12 months. We define this as the sum of claim total charge amount (Part A), claim payment amount (Part B claim lines, aggregated to the claim level), and claim line beneficiary payment amount (part D).

Source: Part A, B, and D claims

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Total health spending	16055.767	0	3250186	NA	NA	NA

Vitamin D deficiency: For each person-month, this variable records whether the person has incurred at least one inpatient or two non-inpatient claims with any diagnosis for vitamin D deficiency within the past two years. If so, this variable takes the value 1; if not, then 0.

Source: Part A and B claims

Models: Pre-CH

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Vitamin D deficiency	NA	NA	NA	.033	0	1

Walkability index: For each person, this variable records the value of the National Walkability Index for the person's Census Tract or ZCTA of residence.

Source: Environmental Protection Agency (2020)

Models: Pre-AH, Pre-CH, Pre-DC, Pre-HE

Risk Factor	MDPCP			HealthChoice		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Walkability index	11.18	0	288	94.022	0	366

Return to Table of Contents

Appendix 2. Social Determinants of Health Data Set

Geocoding Procedure

Hilltop enhanced the granularity of the SDOH risk factors from ZCTA to census tract as part of regular improvements to the production model in October 2021. We increased the granularity of the SDOH covariates because research shows there can be substantial variability of SDOH within ZCTAs. Because census tract measures represent smaller areas, they may provide a more accurate representation of an individual's proximal environment (Moss et al., 2021), but it requires the additional (and potentially non-trivial) development step of geocoding patient addresses.

We used an automated two-step geocoding procedure to identify an individual's unique census tract. First, we used Microsoft® Azure Maps' "Get Search Address" feature to transform their home address from the CCLF data into geographical coordinates (i.e., latitude, longitude). Then, we mapped the coordinates to a census tract using the GeoPandas (v0.8.1) python package. Once a unique census tract was identified for an individual, we linked the environmental risk factors from both their census tract and their 5-digit ZCTA of residence to their individual utilization risk factors in SAS (v.9.4).

Description of Variables

Table 22. Environmental Risk Factor Sources

Risk Factor	Source	Year	Original Granularity for Census Tract	Original Granularity for ZCTA
Population	ACS; Table B01003	2019	Census Tract	ZCTA
Population Growth[1]	ACS; Table B01003	2019	Census Tract	ZCTA
Population Density[2]	ACS; Table B01003	2019	Census Tract	ZCTA
Percent Age 0-4	ACS; Table S0101	2019	Census Tract	ZCTA
Percent Married	ACS; Table S1201	2019	Census Tract	ZCTA
Percent Single Mothers	ACS; Table S1301	2019	Census Tract	ZCTA
Median Household Income	ACS; Table S1901	2019	Census Tract	ZCTA
Percent in Poverty	ACS; Table S1702	2019	Census Tract	ZCTA
Percent Less than High School Diploma	ACS; Table S1501	2019	Census Tract	ZCTA
Percent Native American	ACS; Table DP05	2019	Census Tract	ZCTA
Percent Non-English Speakers	ACS; Table S1601	2019	Census Tract	ZCTA
Percent Foreign Born	ACS; Table DP02	2019	Census Tract	ZCTA
Percent Age 65+	ACS; Table S0101	2019	Census Tract	ZCTA
Percent Age 65+ Live Alone	ACS; Table S1101	2019	Census Tract	ZCTA
Percent Age 65+ Non-White	ACS; Table B01001A	2019	Census Tract	ZCTA
Percent Age 65+ Latinx	ACS; Table B01001L	2019	Census Tract	ZCTA
Percent Age 65+ in Poverty	ACS; Table S1702	2019	Census Tract	ZCTA
Percent Age 65+ Less than High School Diploma	ACS; Table S1501	2019	Census Tract	ZCTA

Risk Factor	Source	Year	Original Granularity for Census Tract	Original Granularity for ZCTA
Rural Urban Index	USDA	2010	Census Tract	ZCTA
Area Deprivation Index	WISC	2019	Census Block[3]	ZCTA
Taxable Interest	IRS	2018	ZCTA[3]	ZCTA
Has a Mental Health Center	CMS	2021	Census Tract	ZCTA
Has a Federally Qualified Health Center	CMS	2021	Census Tract	ZCTA
Has a Rural Health Clinic	CMS	2021	Census Tract	ZCTA
Has a For Profit Hospital	CMS	2021	Census Tract	ZCTA
Number of Hospitals	CMS	2021	Census Tract	ZCTA
Hospitals/1000 Residents[4]	CMS	2021	Census Tract	ZCTA
Hospital Beds/1000 Residents[4]	CMS	2021	Census Tract	ZCTA
Has a VA Clinic or Center	VA	2021	Census Tract	ZCTA
Primary Care Providers/1000 Residents[4]	NPI	2021	Census Tract	ZCTA
Internists/1000 Residents[4]	NPI	2021	Census Tract	ZCTA
Specialists/1000 Residents[4]	NPI	2021	Census Tract	ZCTA
Social Workers/1000 Residents[4]	NPI	2021	Census Tract	ZCTA
Partial Primary Care Shortage Area	AHRF	2018	County[5]	County[3]
Whole Primary Care Shortage Area	AHRF	2018	County[5]	County[3]
Partial Mental Health Care Shortage Area	AHRF	2018	County[5]	County[3]
Whole Mental Health Shortage Area	AHRF	2018	County[5]	County[3]
Percent Physician Diversity (racial or ethnic minority, excluding Asian Americans)	ACS Individual-Level Data	2019	County[5]	County[3]
Air Pollution (average daily PM2.5 concentration)	EPA	2011-2015	Census Tract	Census Tract[3]
Walkability	EPA	2020	Census Block[6]	Census Block[3]
Percent 60+ minute commute	Census	2010	NA	ZCTA
Percent Workers who travel to work by car	Census	2010	NA	ZCTA
Percent Workers who travel to work by public transit	Census	2010	NA	ZCTA
Percent Black	Census	2010	NA	ZCTA
Percent Male	Census	2010	NA	ZCTA
Percent Population living in college group quarters	Census	2010	NA	ZCTA
Percent Population living in nursing home group quarters	Census	2010	NA	ZCTA
Percent of units with 0 or 1 bedrooms	Census	2010	NA	ZCTA
Population per household	Census	2010	NA	ZCTA

Risk Factor	Source	Year	Original Granularity for Census Tract	Original Granularity for ZCTA
ACS = American Community Survey, 5-year estimates, data table number in ( ), AHRF = Area Health Resources File, CMS = Centers for Medicare and Medicaid Services, EPA = Environmental Protection Agency, IRS = Internal Revenue Service, NPI = National Provider Identified Database, USDA = United States Department of Agriculture, VA = Veteran's Affairs, WISC = Wisconsin School of Medicine and Public Health [1]Due to data availability, population growth for census tracts is from 2013-2019 and from 2011-2019 for ZCTAs. [2]Density calculated using land area (square miles) according to the 2019 Census Gazetteer records. [3]Transformed to final geographic unit using HUDuser.gov ratios[31] and the methods from Din & Wilson (2020). [4]Calcuated using the 2019 population estimates from ACS. [5]FIPS county code was matched with the county code for each FIPS census tract. [6]Tract estimate calculated from the average value across all blocks within a tract.				

Transformation Details

For risk factors that were only available at the ZCTA-level (*N*=1) or at the census tract level (or other census polygon, including county - *N*=7), we used the Department of Housing and Urban Development (HUD) USPS ZIP Code Crosswalk files to transform the variables to the appropriate geographic unit (Din & Wilson, 2020; Office of Policy Development and Research, 2021).

Imputation of Missing Values

To facilitate training the Pre- Models, a version of the data set was also created where all missing variables were imputed using the overall mean of the variable.

Census Tract-Level

Physician Diversity: Imputation of missing variables was done for the county_pct_physician_diversity variable because, in the ACS public-use microdata (from IPUMS), counties were not identified from 1950 onwards. Therefore, IPUMS assigns county based on other low-level geographic identifiers which is not possible for all counties. To avoid large amounts of missing data, county_pct_physician_diversity was imputed from a weighted average of physician diversity from the counties in that state.

ZCTA-Level

Physician Diversity: In the ACS public-use microdata (from IPUMS), counties were not identified from 1950 onwards. Therefore, IPUMS assigns county based on other low-level geographic identifiers which is not possible for all counties. To avoid large amounts of missing data,

[31] https://www.huduser.gov/portal/datasets/usps_crosswalk.html

county_pct_physician_diversity was imputed from a weighted average of physician diversity from the counties in that state.

Taxable Interest Per Capita: Data for missing ZCTAs were imputed when possible based on a weighted average of taxable interest per capita from the other ZCTAs within the same ZIP code sorting area (first three digits of ZCTA).

Area Deprivation Index: Data for missing ZCTAs were imputed when possible based on a weighted average of the area deprivation index from the other ZCTAs within the same ZIP code sorting area (first three digits of ZCTA).

The Hilltop Institute

UMBC

Sondheim Hall, 3rd Floor
1000 Hilltop Circle
Baltimore, MD 21250
410-455-6854
www.hilltopinstitute.org