Diabetes Ther (2020) 11:681–699
https://doi.org/10.1007/s13300-020-00759-4

ORIGINAL RESEARCH

Application of Machine Learning Models to Evaluate Hypoglycemia Risk in Type 2 Diabetes

Luke Mueller · Paulos Berhanu · Jonathan Bouchard · Veronica Alas · Kenneth Elder · Ngoc Thai · Cody Hitchcock · Tiffany Hadzi · Iya Khalil · Lesley-Ann Miller-Wilson

Received: October 16, 2019 / Published online: February 3, 2020

ABSTRACT

Introduction: To identify predictors of hypoglycemia and five other clinical and economic outcomes among treated patients with type 2 diabetes (T2D) using machine learning and structured data from a large, geographically diverse administrative claims database.

Methods: A retrospective cohort study design was applied to Optum Clinformatics claims data indexed on first antidiabetic prescription date. A hypothesis-free, Bayesian machine learning analytics platform (GNS Healthcare REFS™: Reverse Engineering and Forward Simulation) was used to build ensembles of generalized linear models to predict six outcomes defined in patients' 1-year post-index claims history, including hypoglycemia, antidiabetic class persistence, glycated hemoglobin (HbA1c) target attainment, HbA1c change, T2D-related inpatient admissions, and T2D-related medical costs. A unified set of 388 variables defined in patients' 1-year pre-index claims history constituted the set of predictors for all REFS models.

Results: The derivation cohort comprised 453,487 patients with a T2D diagnosis between 2014 and 2017. Patients with comorbid conditions had the highest risk of hypoglycemia, including those with prior hypoglycemia (odds ratio [OR] = 25.61) and anemia (OR = 1.29). Other identified risk factors included insulin (OR = 2.84) and sulfonylurea use (OR = 1.80). Biguanide use (OR = 0.75), high blood glucose (> 125 mg/dL vs. < 100 mg/dL, OR = 0.47; 100–125 mg/dL vs. < 100 mg/dL, OR = 0.53), and missing blood glucose test (OR = 0.40) were associated with reduced risk of hypoglycemia. Area under the curve (AUC) of the hypoglycemia model in held-out testing data was 0.77. Patients in the top 15% of predicted hypoglycemia risk constituted 50% of observed hypoglycemic events, 26% of T2D-related inpatient admissions, and 24% of all T2D-related medical costs.

Conclusions: Machine learning models built within high-dimensional, real-world data can predict patients at risk of clinical outcomes with a high degree of accuracy, while uncovering important factors associated with outcomes that can guide clinical practice. Targeted interventions towards these patients may help reduce hypoglycemia risk and thereby favorably

Enhanced Digital Features To view enhanced digital features for this article go to https://doi.org/10.6084/m9.figshare.11527902.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13300-020-00759-4) contains supplementary material, which is available to authorized users.

L. Mueller (✉) · V. Alas · K. Elder · N. Thai · C. Hitchcock · T. Hadzi · I. Khalil
GNS Healthcare, Cambridge, MA, USA
e-mail: lmueller@gnshealthcare.com

P. Berhanu · J. Bouchard · L.-A. Miller-Wilson
Sanofi-Aventis U.S. LLC, Bridgewater, NJ, USA

impact associated economic outcomes relevant to key stakeholders.

Keywords: Healthcare costs; Hypoglycemia; Machine learning; Resource utilization; Type 2 diabetes; Value-based

Key Summary Points

Type 2 diabetes (T2D) is associated with significant healthcare resource utilization, especially among patients with sub-optimal management, treatment-related adverse events including hypoglycemia, and comorbid health conditions. Value-based initiatives offer a unique solution to this problem, but additional evidence is needed to design and support these initiatives.

A Bayesian machine learning platform, Reverse Engineering Forward Simulation (REFS^{TM}), was applied to administrative claims data to identify predictors of key clinical and economic outcomes in T2D.

Machine learning models such as REFS have the potential to guide the provision of data-driven, individualized care with these results establishing the importance of ensuring that patients with T2D are appropriately treated with evidence-based interventions to ensure more favorable outcomes as well as control of healthcare resource utilization and costs.

INTRODUCTION

Rising healthcare costs in the USA and around the world are a growing concern to global health economists. In 2016, the USA spent 17.8% of its gross domestic product on healthcare—nearly twice as much as other high-income countries—while presenting consistently lower clinical outcomes relative to these countries [1].

Diabetes is a growing health problem and primary contributor to these costs, afflicting 24.7 million Americans—90–95% of which have type 2 diabetes (T2D)—and accounting for an estimated $327 billion in direct medical costs and reduced productivity in 2017. In fact, patients with diabetes incur medical expenditures approximately 2.3 times that of patients without diabetes [2].

To address uncontrolled spending and inadequate quality of care in the industry, health systems are shifting their reimbursement models to focus on value, rewarding quality and cost reduction. In the USA, the Affordable Care Act, the most extensive healthcare reform seen in years, stimulated massive changes in health insurance coverage and care delivery, and reinforced the adaptation of novel concepts, such as patient-centered medical homes, integrated care delivery, and outcomes-based reimbursements.

Although in theory value-based contracting and reimbursement seem to be a desirable mitigation to excessive healthcare spending, in reality such solutions are difficult to implement largely because of high implementation costs, measurement challenges, and absence of a suitable data infrastructure [3].

To date, there is no gold standard for value-based arrangements [4], and the status and performance of established arrangements are relatively unknown as many exist outside the public domain [3, 5]. Indeed, global trends suggest payers are increasingly requiring pharmaceuticals demonstrate greater certainty and value of their products in the real world, as costs of new therapies continue to rise.

Continued exponential growth and integration of "big data" sources and maturity of analytic tools offer new opportunities to overcome these practical barriers, through precise estimation of both clinical and economic value, and insights driven by real-world data. In particular, legislative action to increase transparency and accessibility of data, increased adoption of electronic health records (EHR) by health systems, and new technologies capable of storing and processing these data provide researchers rich sources of real-world patient populations

without the expense and time required by traditional observational cohort studies [6, 7].

This study sought to contribute to the development of value-based contracts between pharmaceuticals and payers by applying machine learning models to a nationally representative claims database. In particular, a predictive Bayesian network model was trained on six diabetes-related clinical and economic outcomes: hypoglycemic events, antidiabetic medication persistence, T2D-related inpatient admissions, glycated hemoglobin (HbA1c) target attainment, change from baseline HbA1c, and T2D-related medical costs, using a consistent set of variables within a derived T2D population. Variables most predictive of outcomes were extracted for inference. Finally, a link between outcomes was established by evaluating the concentration of observed outcomes within the highest risk groups of the hypoglycemia model, providing an estimate of the potential value offered by a model-driven value-based strategy.

METHODS

Data Source

The data source for this study was Optum Clinformatics Data Mart (CDM) (Eden Prairie, MN, USA). Optum's CDM is a closed system of de-identified health claims data that includes over 13 million lives annually, and contains patients' medical, prescription drug, laboratory, and eligibility information beginning in year 2000. The data comes from a large, national US health insurer and is certified as de-identified. For this study, a data extract between 1 January 2014 and 30 September 2017 was received from Optum.

Sample Selection

This extract was used to construct a population of patients with T2D following a retrospective cohort study design. Patients were eligible for inclusion if they met the following criteria: at least one medical claim with a diagnosis of T2D (International Classification of Diseases, 9th revision (ICD-9) codes 250.x0, 250.x2; International Classification of Diseases, 10th revision (ICD-10) code E11) anytime during patients' available claims history, at least one prescription claim for a qualifying antidiabetic medication (Generic Product Identifier (GPI) 27) anytime during patients' available claims history, at least 18 years old as of the index date, and continuously enrolled with a medical and pharmacy benefit during the 12-month pre-index and 12-month post-index periods. The index date was the date of first prescription claim for an antidiabetic medication.

Outcomes

Six unique outcomes were chosen for modeling, all captured in a 12-month post-index period; these included hypoglycemic events, antidiabetic medication class persistence, HbA1c target attainment, HbA1c change from baseline, T2D-related inpatient admissions, and T2D-related medical costs.

Hypoglycemia was defined by a combination of relevant ICD-9 and ICD-10 codes for hypoglycemia (Table A1 in the supplementary material) and available blood glucose measurements as follows: if a patient had at least one medical claim with any of the ICD-9 or ICD-10 codes associated with hypoglycemia as listed in Table A1 during the post-index period or if a patient had a least one blood glucose measurement of 70 mg/dL or below, we inferred that the patient experienced a hypoglycemic event. The outcome was coded as a binary variable (i.e., 1 if at least one hypoglycemic event in the post-index period; 0 otherwise). Definitions for remaining study outcomes are detailed in the supplementary material.

Statistical Analyses

A unified set of covariates was built for all outcomes using patients' available pre-index medical, pharmacy, and laboratory claims information. Covariates were not necessarily diabetes-related; rather, patients' complete claims history was extracted leaving variable

selection to the models. Furthermore, in order to extract a meaningful signal from each variable, and to ensure variables were defined consistently across varying time periods, claims codes were aggregated together using a variety of widely used code sets. Covariate structures are briefly described below:

- *Demographics* age, gender, race, insurance type, product type, region, and low-income subsidy status.
- *Diagnoses* ICD-9 and ICD-10 diagnosis codes from medical claims aggregated via Clinical Classification Software (CCS) codes [8, 9].
- *Pharmaceutical utilization* National Drug Code (NDC) product codes from pharmacy claims aggregated via Generic Product Identifier (GPI) codes [10].
- *Procedures* ICD-9, ICD-10, CPT, and Healthcare Common Procedure Coding System (HCPCS) codes from procedures in medical claims aggregated via Berenson-Eggers Type of Service (BETOS) codes [11].
- *Laboratory* Logical Observation Identifiers Names and Codes (LOINC) codes from laboratory data aggregated via the LOINC hierarchies.
- *Healthcare resource utilization* Acute inpatient admissions, inpatient length of stay, outpatient visits, office visits, visits with an endocrinologist, and emergency department visits, total medical costs, total pharmacy costs (including copay of the index prescription and total out-of-pocket costs), outpatient costs, and emergency department (ED) costs. Visit counts were categorized into 0, 1, 2, 3 or more visits, while costs were discretized into quartiles to account for heavily skewed distributions.

A machine learning analytic platform, Reverse Engineering and Forward Simulation (REFS^{TM}) (GNS Healthcare, Cambridge, MA, USA), was used for all modeling. Briefly, each REFS model was an ensemble consisting of 128 generalized linear models, constructed using Markov chain Monte Carlo sampling of the full Bayesian posterior distribution of models, given the available data—i.e., P(model|data) [12–16]. In high-dimensional data sets, attempts to identify a single best model will inevitably lead to overfitting; thus, our approach identifies an ensemble of models, each scored by both its goodness-of-fit to the observed data and its complexity. This approach allows incorporation of previous knowledge regarding the different types of data (e.g., medical claims, pharmacy claims, laboratory data, etc.), including the expected relative contribution of each type, into the final models. Therefore, examination of large numbers of variables with low signal (i.e., medications and diagnoses) is possible without overwhelming the signal of the fewer—although potentially more directly informative—variables (i.e., demographics, indexing antidiabetic class, etc.).

The 128 constituent models within the ensemble looked at combinations of the available parameters, including linear additive and quadratic terms, and up to second-order interactions to accommodate non-linear effects against outcomes. To prevent overfitting, the complexity (i.e., the number of terms) of each model in the ensemble was penalized by specifying a maximum entropy prior over the number of unique predictors selected per given variable class.

From the ensemble for each study outcome, predictors can be ranked and evaluated by their relative selection frequency (proportion of models in the ensemble in which the variable was selected) and distributions of effect estimates. A high selection frequency for a given predictor represents an increased probability of a true predictive association with the outcome [14].

REFS models were trained on 80% of the final study cohorts sampled at random, while the remaining 20% were reserved for model validation and performance assessment. Out-of-sample validation provides a more realistic estimate of true model performance and supports generalized use of models in new patient samples.

Model prediction performance was measured via the area under the receiver operating characteristic curve (AUC) metric for all outcomes except T2D-related total medical costs. The AUC metric illustrates the performance of a binary classifier system and is used to examine the overall specificity and sensitivity of the

classification approach, i.e., the trade-off between false positive and negative predictions of a model. AUC values range between 0 and 1 where greater AUC values indicate better model performance (an AUC of 0.5 is associated with a random predictor). For the T2D-related total medical costs outcome, which is continuous, Pearson R^2 (predicted vs. observed costs) was used to assess performance.

In addition, credible intervals (CIs) for performance metrics were calculated for each REFS ensemble by generating a distribution of performance metrics across all 128 models in each ensemble. This is a unique characteristic of ensemble modeling—granting insight into model consensus across the ensemble—as well as suggesting best- and worst-case scenarios for potential deployment of these models in new patient samples.

Finally, the utility of the hypoglycemia ensemble was estimated by evaluating the concentration of costs and visits, as well as other study outcomes within 20 equally distributed bins of predicted hypoglycemic event risk (henceforth referred to as "ventiles"), as applied in similar contexts [17]. More specifically, each ventile contained 5% of patients in the study cohort (N = 22,672) and patients were assigned to a ventile according to his or her predicted probability of hypoglycemia. Patients with the lowest probability of hypoglycemia—in the bottom 5th percentile—were assigned to the first ventile, whereas patients with the highest probability of hypoglycemia were assigned to the 20th ventile. The proportions of visits, costs, or other study outcomes within each ventile out of the total number of visits, costs, or other study outcomes were calculated as a measure of concentration of risk.

Compliance with Ethics Guidelines

The Optum CDM data used in this study was compliant with the Health Insurance Portability and Accountability Act. Fully anonymized retrospective data were obtained from Optum via license agreement, and no other data were used in this study. As such, this study was deemed exempt from ethical approval.

RESULTS

Population Characteristics

Population characteristics are presented in Table 1. The full sample included 453,487 patients with T2D after applying inclusion and exclusion criteria listed above; population counts for each layer of the inclusion and exclusion criteria are presented in Fig. 1. Within the full sample, 221,473 (48.8%) patients had at least one post-index HbA1c laboratory value for the HbA1c target attainment outcome, and 36,263 (8.0%) additionally had uncontrolled baseline HbA1c for the assessment of the change from baseline HbA1c outcome. Demographic characteristics were comparable across the samples for most variables, though regional distributions shifted after additionally requiring HbA1c testing in the pre- and post-index periods. For example, 42.3% of patients were from the South in the full sample, whereas 49.9% of patients were from the South in the sample requiring pre- and post-index HbA1c tests.

Outcome Distributions

In the full sample, 3.6% of patients had at least one hypoglycemic event in the post-index period, 18.2% of patients were persistent with their respective indexing antidiabetic medication class throughout the post-index period, and 8.4% of patients had at least one T2D-related inpatient admission in the post-index period. Mean T2D-related total medical costs per patient in the post-index period were $4274 (standard deviation (SD) = $12,196). In the sample requiring at least one post-index HbA1c value, 72.8% of patients met their HbA1c target in the post-index period (29.5% of which were uncontrolled or missing HbA1c laboratory data in the pre-index period). In the sample additionally requiring uncontrolled baseline HbA1c, 28.4% of patients had a change in HbA1c between the pre- and post-index period greater than the required threshold. Nearly all characteristics presented in Table 1 were significantly associated with all outcomes at α level 0.05—

Table 1 Selected characteristics by study outcome

Variable	Overall	Hypoglycemic event	Persistent to antidiabetic class	T2D-related inpatient admission	High T2D-related medical cost[b]	HbA1c target attainment	Change from baseline HbA1c
Population size	453,487	453,487	453,487	453,487	453,487	221,473	36,263
With outcome	–	16,227 (3.6%)	82,689 (18.2%)	37,884 (8.4%)	113,366 (25.0%)	161,230 (72.8%)	10,281 (28.4%)
Age group							
18–34	4891 (1.1%)	139 (0.9%)	1693 (2.0%)	259 (0.7%)	903 (0.8%)	1163 (0.7%)	162 (1.6%)
35–44	20,892 (4.6%)	438 (2.7%)	6407 (7.7%)	873 (2.3%)	3684 (3.2%)	4459 (2.8%)	793 (7.7%)
45–54	58,369 (12.9%)	1390 (8.6%)	15,055 (18.2%)	3128 (8.3%)	11,774 (10.4%)	13,336 (8.3%)	1979 (19.2%)
55–64	96,878 (21.4%)	2800 (17.3%)	19,840 (24.0%)	7065 (18.6%)	22,506 (19.9%)	25,166 (15.6%)	2492 (24.2%)
65–74	155,837 (34.4%)	5794 (35.7%)	21,011 (25.4%)	13,094 (34.6%)	39,772 (35.1%)	65,467 (40.6%)	3028 (29.5%)
≥ 75	116,620 (25.7%)	5666 (34.9%)	18,683 (22.6%)	13,465 (35.5%)	34,727 (30.6%)	51,639 (32.0%)	1827 (17.8%)
Age, mean (SD)	66.1 (12.4)	69.2 (11.9)	63.1 (13.7)	69.4 (11.7)	67.9 (12)	69.0 (11.4)	62.4 (12.9)
Gender							
Female	224,498 (49.5%)	8829 (54.4%)	41,563 (50.3%)	19,563 (51.6%)	59,967 (52.9%)	83,969 (52.1%)	4479 (43.6%)
Male	228,989 (50.5%)	7398 (45.6%)	41,126 (49.7%)	18,321 (48.4%)	53,399 (47.1%)	77,261 (47.9%)	5802 (56.4%)
Race							
Asian	21,728 (4.8%)	685 (4.2%)	4198 (5.1%)	1015 (2.7%)	3749 (3.3%)	9860 (6.1%)	513 (5.0%)

Table 1 continued

Variable	Overall	Hypoglycemic event	Persistent to antidiabetic class	T2D-related inpatient admission	High T2D-related medical cost[b]	HbA1c target attainment	Change from baseline HbA1c
Black	48,148 (10.6%)	2152 (13.3%)	10,842 (13.1%)	4693 (12.4%)	14,108 (12.4%)	15,358 (9.5%)	1075 (10.5%)
Hispanic	76,709 (16.9%)	3106 (19.1%)	15,967 (19.3%)	4466 (11.8%)	18,057 (15.9%)	31,245 (19.4%)	2402 (23.4%)
Unknown	39,697 (8.8%)	1413 (8.7%)	4720 (5.7%)	3214 (8.5%)	9487 (8.4%)	16,432 (10.2%)	859 (8.4%)
White	267,205 (58.9%)	8871 (54.7%)	46,962 (56.8%)	24,496 (64.7%)	67,965 (60.0%)	88,335 (54.8%)	5432 (52.8%)
Region							
Midwest	97,363 (21.5%)	2436 (15.0%)	15,985 (19.3%)	10,189 (26.9%)	26,710 (23.6%)	22,874 (14.2%)	1217 (11.8%)
Northeast	50,876 (11.2%)	1684 (10.4%)	8411 (10.2%)	5031 (13.3%)	13,891 (12.3%)	17,818 (11.1%)	888 (8.6%)
South	192,002 (42.3%)	8048 (49.6%)	37,931 (45.9%)	16,361 (43.2%)	49,914 (44.0%)	69,635 (43.2%)	5116 (49.8%)
Unknown	2374 (0.5%)	107 (0.7%)	487 (0.6%)	137 (0.4%)	481 (0.4%)	1046 (0.6%)	61 (0.6%)
West	110,872 (24.4%)	3952 (24.4%)	19,875 (24.0%)	6166 (16.3%)	22,370 (19.7%)	49,857 (30.9%)	2999 (29.2%)
Insurance type							
Commercial	172,317 (38.0%)	3837 (23.6%)	42,619 (51.5%)	9953 (26.3%)	33,567 (29.6%)	40,403 (25.1%)	4770 (46.4%)
Medicare	282,054 (62.2%)	12,423 (76.6%)	40,250 (48.7%)	27,978 (73.9%)	79,992 (70.6%)	121,198 (75.2%)	5524 (53.7%)

Table 1 continued

Variable	Overall	Hypoglycemic event	Persistent to antidiabetic class	T2D-related inpatient admission	High T2D-related medical cost[b]	HbA1c target attainment	Change from baseline HbA1c
Product type							
EPO	19,291 (4.3%)	453 (2.8%)	5289 (6.4%)	1064 (2.8%)	3856 (3.4%)	5376 (3.3%)	636 (6.2%) [a]
HMO	131,125 (28.9%)	6030 (37.2%)	22,381 (27.1%)	8173 (21.6%)	30,937 (27.3%)	63,956 (39.7%)	3592 (34.9%)
IND	6345 (1.4%)	160 (1.0%)	974 (1.2%)	835 (2.2%)	1745 (1.5%)	703 (0.4%)	16 (0.2%) [a]
OTH	150,765 (33.2%)	5817 (35.8%)	19,891 (24.1%)	18,239 (48.1%)	46,418 (40.9%)	55,390 (34.4%)	2366 (23%)
POS	127,665 (28.2%)	2702 (16.7%)	31,638 (38.3%)	7108 (18.8%)	25,069 (22.1%)	28,583 (17.7%)	3428 (33.3%)
PPO	28,794 (6.3%)	1410 (8.7%)	4636 (5.6%)	3362 (8.9%)	8001 (7.1%)	10,480 (6.5%)	462 (4.5%) [a]
Low income subsidy							
Yes	74,392 (16.4%)	4150 (25.6%)	11,086 (13.4%)	9102 (24%)	25,826 (22.8%)	27,714 (17.2%)	1563 (15.2%)
CCI, mean (SD)	2.7 (2.2)	3.9 (2.6)	2.5 (2.2)	3.7 (2.6)	3.4 (2.4)	3.0 (2.3)	2.8 (2.2)
Indexing antidiabetic class							
Amylin analogue	10 (0.0%)	1 (0.0%)	1 (0.0%)	2 (0.0%)	4 (0.0%)	2 (0.0%)	1 (0.0%)
Alpha-glucosidase inhibitor	563 (0.1%)	57 (0.4%)	144 (0.2%)	36 (0.1%)	153 (0.1%)	252 (0.2%)	8 (0.1%)
Biguanide (metformin)	241,678 (53.3%)	4942 (30.5%)	48,744 (58.9%)	17,047 (45.0%)	52,536 (46.3%)	95,501 (59.2%)	4140 (40.3%)

Table 1 continued

Variable	Overall	Hypoglycemic event	Persistent to antidiabetic class	T2D-related inpatient admission	High T2D-related medical cost[b]	HbA1c target attainment	Change from baseline HbA1c
Antidiabetic combination	20,876 (4.6%)	571 (3.5%)	4588 (5.5%)	1331 (3.5%)	4573 (4.0%)	6851 (4.2%)	716 (7.0%)
DPP4 inhibitor	19,171 (4.2%)	668 (4.1%)	3938 (4.8%)	1951 (5.1%)	5565 (4.9%)	6859 (4.3%)	438 (4.3%)
Dopamine receptor agonist	46 (0.0%)	4 (0.0%)	16 (0.0%)	3 (0.0%)	13 (0.0%)	15 (0.0%)	1 (0.0%)
GLP-1 receptor agonist	10,399 (2.3%)	332 (2.0%)	939 (1.1%)	707 (1.9%)	2710 (2.4%)	2945 (1.8%)	346 (3.4%)
Insulin-sensitizing agent	9513 (2.1%)	397 (2.4%)	1568 (1.9%)	742 (2.0%)	2281 (2.0%)	3924 (2.4%)	318 (3.1%)
Insulin	47,722 (10.5%)	4243 (26.1%)	4179 (5.1%)	6563 (17.3%)	18,091 (16.0%)	10,287 (6.4%)	1493 (14.5%)
Meglitinide analogue	1495 (0.3%)	81 (0.5%)	60 (0.1%)	171 (0.5%)	490 (0.4%)	579 (0.4%)	23 (0.2%)
SGLT2 inhibitor	7376 (1.6%)	197 (1.2%)	1894 (2.3%)	397 (1.0%)	1579 (1.4%)	1878 (1.2%)	300 (2.9%)
Sulfonylurea	94,591 (20.9%)	4719 (29.1%)	16,618 (20.1%)	8928 (23.6%)	25,353 (22.4%)	32,129 (19.9%)	2496 (24.3%)

T2D type 2 diabetes, *SD* standard deviation, *EPO* exclusive provider organization, *HMO* health maintenance organization, *IND* indemnity, *OTH* other product type, *POS* point-of-service, *PPO* preferred provider organization, *CCI* Charlson comorbidity index score, *DPP* dipeptidyl peptidase 4, *GLP-1* glucagon-like peptide 1, *SGLT2* sodium/glucose cotransporter 2

[a] Not significant at α level 0.05, comparing outcome to non-outcome within respective variable group

[b] High T2D-related medical cost refers to patients whose post-index T2D-related medical claims costs were in the top 25th percentile of costs within this sample

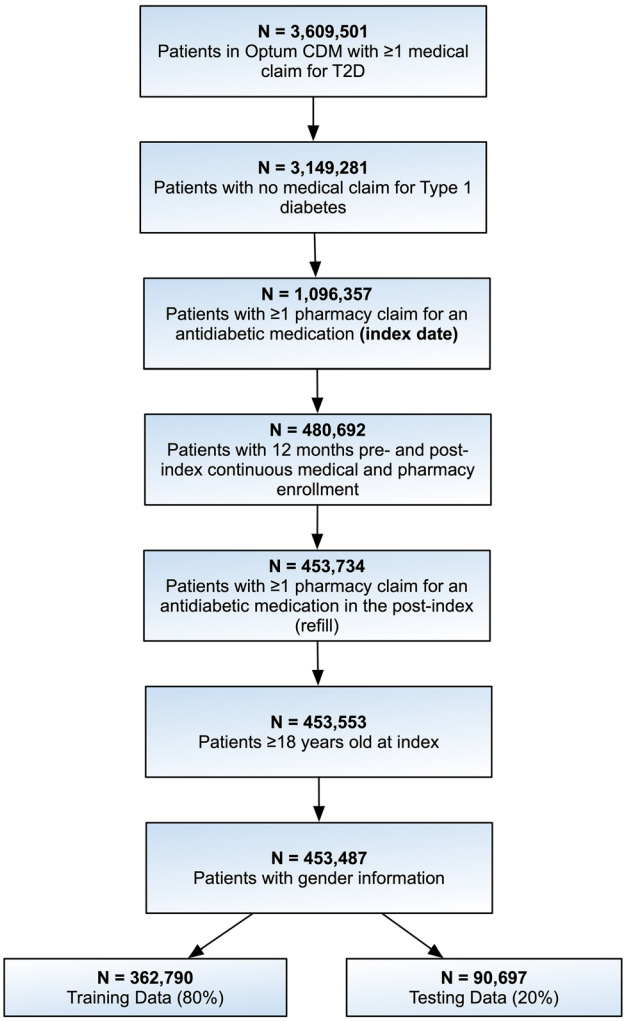

Fig. 1 Sample selection

though some associations may not be clinically relevant because of the high sample size.

Covariate Set

In total, 6907 potential predictors were extracted from the study cohort. After preprocessing (i.e., removal of variables with high missingness, multicollinearity, etc.), 388 predictors remained, including 13 demographic variables, 89 diagnosis variables, 180 pharmacy variables, 68 procedure variables, 30 laboratory variables (4 categorical values, 26 indicators for presence of test), and 8 utilization variables. In addition, REFS models explored the space of all pairwise interactions between these 388 predictors.

Model Performance

Performance was moderately strong to very strong across outcomes in out-of-sample testing data. In order of accuracy, out-of-sample performance tested as follows: HbA1c target attainment (AUC 0.867, 95% CI 0.867–0.867), hypoglycemia (AUC 0.773, 95% CI 0.772–0.774), T2D-related inpatient admission (AUC 0.735, 95% CI 0.734–0.736), change from baseline HbA1c (AUC 0.709, 95% CI 0.706–0.711), and antidiabetic class persistence (AUC 0.675, 95% CI 0.674–0.675). Performance of the T2D-related medical costs model was also strong in out-of-sample tests (R^2 0.235, 95% CI 0.234–0.235).

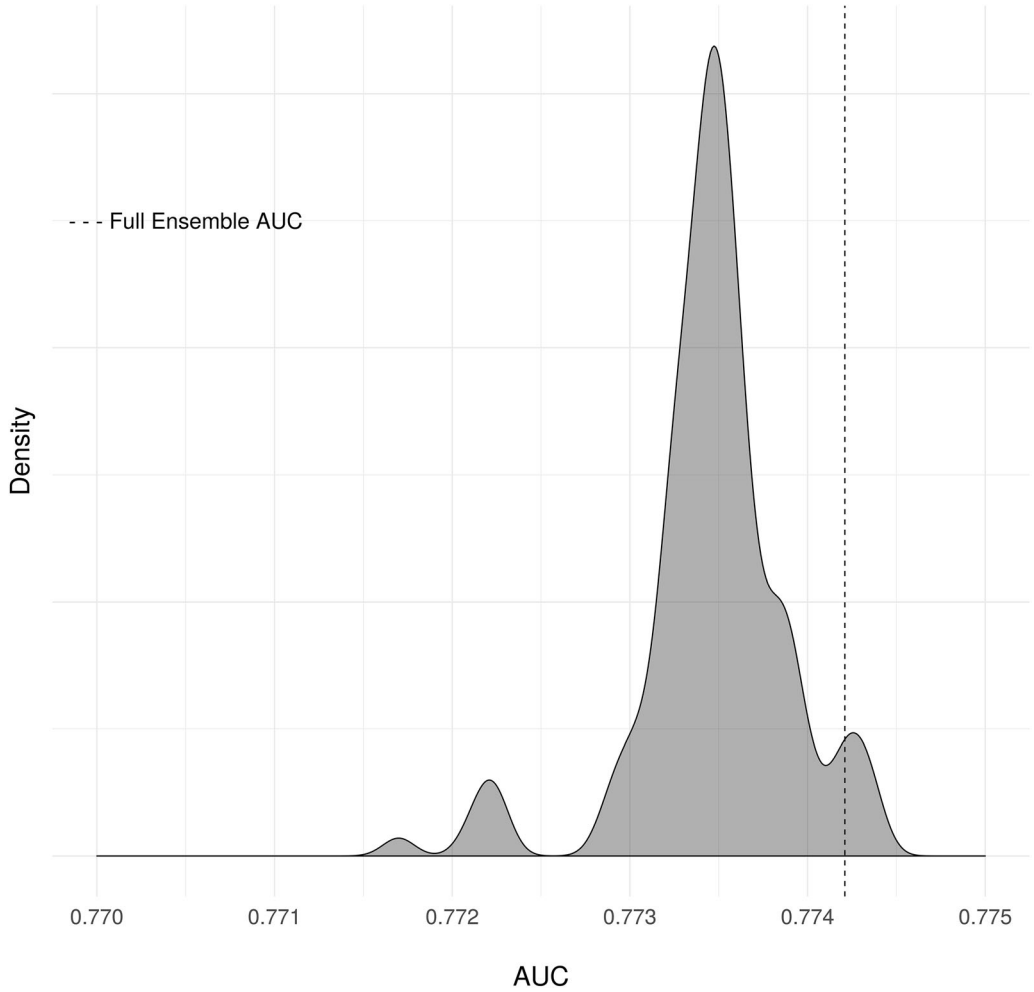

Fig. 2 Hypoglycemia model performance across REFS ensemble. Hypoglycemia model ensemble performance: AUC was calculated within each of the 128 models comprising the full ensemble, and separately across the full ensemble (indicated by the dotted line). The full ensemble AUC generally performs better than most individual models in the ensemble as it combines information across diverse models. Several single models performed slightly better but are more prone to overfitting

Figure 2 summarizes the distribution of out-of-sample AUC values across networks in the hypoglycemia model ensemble, with the AUC averaged over the full ensemble indicated by the dotted line.

Top Predictors

Figure 3 summarizes the ensemble for the hypoglycemia model through its top predictors (i.e., those with at least 90% selection frequency across the ensemble). Outputs for other outcomes are included in the supplementary material. The figure is split into two parts to account for large effect sizes on prior hypoglycemia and its interactions. As each model in the ensemble potentially consists of different sets of variables, the effect of a particular variable may vary across the ensemble. For example, "CCI score (1)" had an odds ratio for hypoglycemia of 1.28 in one model, but 0.67 in another model after controlling for different sets of covariates. The median odds ratio is reported alongside its interquartile range (IQR) across the ensemble to account for this variation.

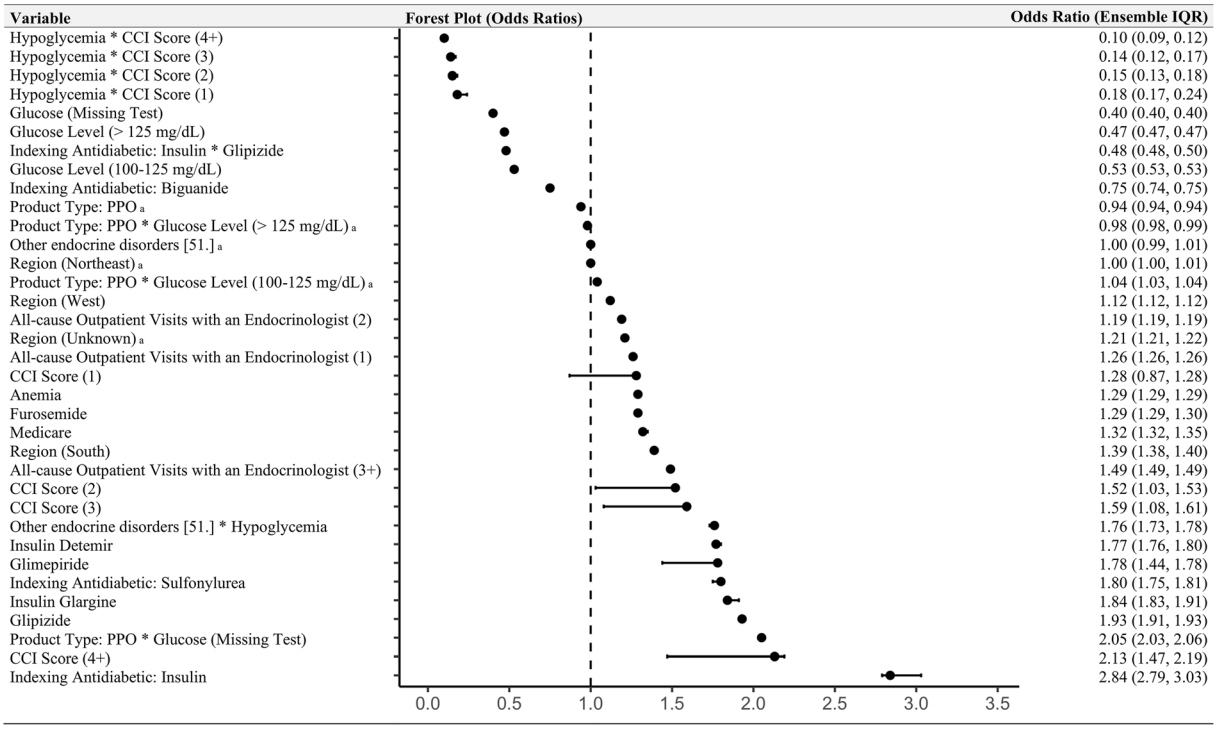

Fig. 3 Hypoglycemia ensemble summary of predictors. CCI Charlson comorbidity index score, PPO preferred provider organization, IQR interquartile range. Categorical variable reference levels: CCI score, 0; region, Midwest; indexing antidiabetic, any other antidiabetic; outpatient visits with an endocrinologist, 0 visits; glucose level, < 100 mg/dL. Interaction terms are indicated by asterisks. Not shown: prior hypoglycemia (odds ratio = 25.61, IQR = 23.55–25.61). [a]Median p value for variable > 0.05

Interaction Effects

For all model ensembles, REFS additionally explored all pairwise interactions between variables as potential covariates. Several such interactions were present in the hypoglycemia model. For example, an interaction was discovered between "indexing antidiabetic: insulin" and "glipizide" (OR 0.48, IQR 0.48–0.50). In isolation, the main effects of both "indexing antidiabetic: insulin" (2.84, 2.79–3.03) and "glipizide" (1.93, 1.91–1.93) were positively associated with hypoglycemia. However, the negative direction of the interaction coefficient implies an antagonistic rather than synergistic interaction, i.e., the joint effect was less than the sum of their individual effects. This suggests that while patients taking insulins and/or glipizide (a sulfonylurea) had higher odds of hypoglycemia compared to patients taking neither of these prescriptions, their concomitant use was not purely additive.

Concentration of Risk and Outcome Interactions

Figure 4 and Table 2 demonstrate a link between outcomes by assessing the relative proportion of observed events across risk ventiles within the hypoglycemia ensemble. Notably, patients in the top 15th percentile of hypoglycemia risk (i.e., top three ventiles) constituted 50% of all observed hypoglycemic events, 26% of T2D-related inpatient admissions, and 24% of T2D-related medical costs in the full sample of patients with T2D. Hypoglycemia-related visits were even more densely concentrated within the high-risk ventiles, with patients in the top 5th percentile of hypoglycemia risk (i.e., 20th ventile) constituting

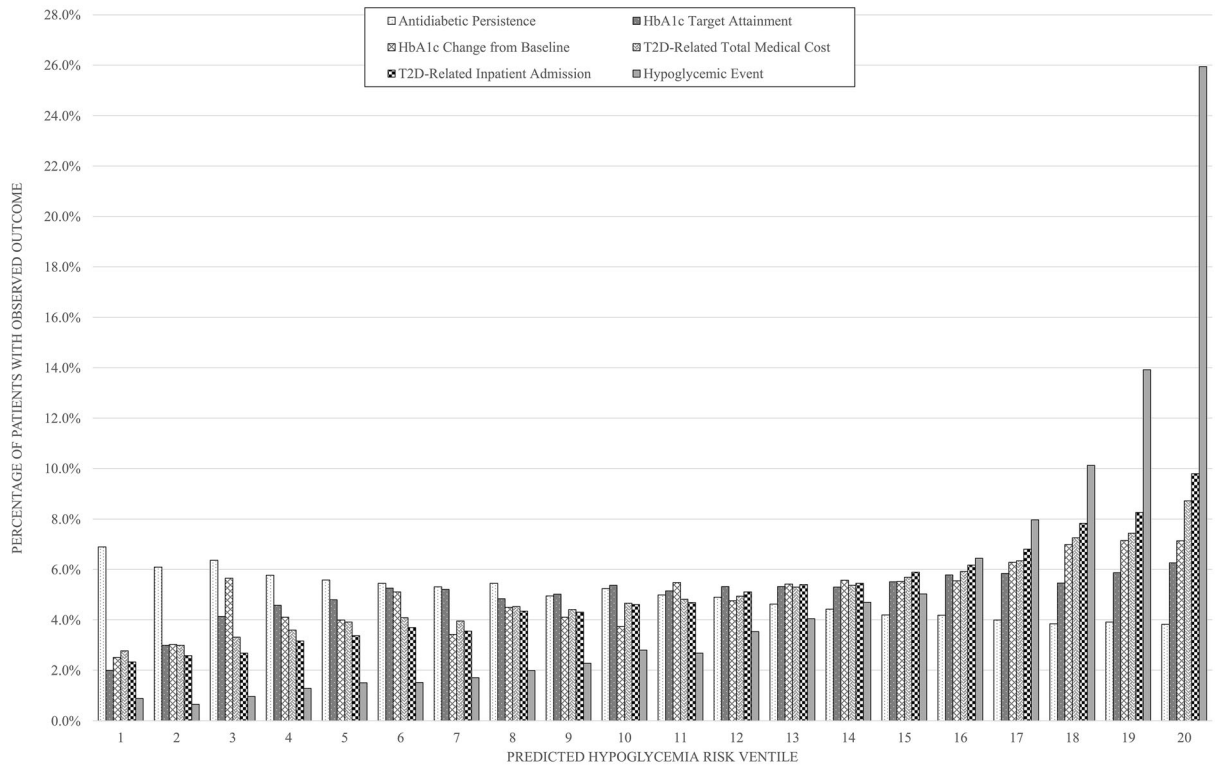

Fig. 4 Relative proportions of study outcomes by hypoglycemia risk ventile. The probability of hypoglycemia, as estimated by the model, was split into 20 groups (every 5th percentile). Then, for each outcome, the number of observed events within each risk group was summed and divided by the total number of observed events across the study population. For example, 2965 (9.8%) T2D-related inpatient admissions occurred in the top 5th percentile of predicted hypoglycemia risk out of 30,265 T2D-related inpatient admissions in the study population. For T2D-related medical costs, the sum of costs within ventiles was divided by the total sum of costs across the study population

26.7% of hypoglycemia-related inpatient admissions, 24.7% of hypoglycemia-related ER visits, and 27.4% of hypoglycemia-related outpatient visits. Average healthcare costs (medical + pharmacy) per patient per year were $36,567 in this group, compared to $12,038 for patients in the bottom 5th percentile of hypoglycemia risk.

Conversely, patients in the bottom 15th percentile of hypoglycemia risk constituted 19% of patients persistent to their respective antidiabetic classes, 9% of HbA1c target attainers, and 11% patients who had the met the appropriate threshold for change from baseline HbA1c; these results suggested a positive correlation between HbA1c reduction and hypoglycemia risk.

DISCUSSION

In this study, six predictive model ensembles, including hypoglycemic events, antidiabetic medication class persistence, HbA1c target attainment, HbA1c change from baseline, T2D-related inpatient admissions, and T2D-related medical costs, were trained on a T2D population derived from administrative claims data. A retrospective cohort study design and hypothesis-free Bayesian network models were used to explore associations between thousands of unique variables, their interactions, and outcomes to identify those variables most predictive of each outcome. Model performance was moderately strong to very strong across model ensembles in held-out testing data. Notably, the

Table 2 Relative proportions of post-index study outcomes, visits, and costs by hypoglycemia risk ventile

Post-index variable[a]	Highest risk stratum (20th ventile)		19th ventile		11–18th ventiles		Lowest risk stratum (1–10th ventiles)		Overall
	N[b]	%[c]	*N*	%	*N*	%	*N*	%	*N*
Number of patients	22,672	5.0	22,672	5.0	181,376	40.0	226,719	50.0	453,439
Study outcomes									
Hypoglycemic event	3348	25.9	1797	13.9	5745	44.5	2014	15.6	12,904
Antidiabetic persistence	2532	3.8	2590	3.9	23,250	35.1	37,790	57.1	66,162
HbA1c target attainment	8082	6.3	7573	5.9	56,345	43.7	56,982	44.2	128,982
HbA1c change from baseline	1175	7.1	1177	7.2	7498	45.6	6602	40.1	16,452
T2D-related inpatient admission	2965	9.8	2501	8.3	14,328	47.3	10,469	34.6	30,263
T2D-related total medical cost	$4114	8.7	$3509	7.4	$2691	45.6	$1801	38.2	$2358
Visits: all-cause									
Inpatient admissions	5608	10.2	4590	8.4	26,082	47.5	18,644	33.9	54,924
ER visits	7243	7.4	6451	6.6	43,955	44.9	40,313	41.2	97,962
Outpatient visits	22,420	5.1	22,376	5.1	177,857	40.3	218,783	49.6	441,436
Visits: hypoglycemia-related									
Inpatient admissions	501	26.7	254	13.5	886	47.2	236	12.6	1877
ER visits	570	24.7	323	14.0	1144	49.6	271	11.7	2308
Outpatient visits	2164	27.4	1192	15.1	3381	42.9	1153	14.6	7890
Costs									
Healthcare	$36,567	10.7	$30,090	8.8	$19,406	45.4	$12,038	35.2	$17,114
Medical	$28,652	11.3	$23,081	9.1	$14,173	44.8	$8825	34.8	$12,669
Pharmacy	$7915	8.9	$7009	7.9	$5233	47.1	$3213	36.1	$4446

Table 2 continued

Post-index variable[a]	Highest risk stratum (20th ventile)		19th ventile		11–18th ventiles		Lowest risk stratum (1–10th ventiles)		Overall
	N[b]	%[c]	*N*	%	*N*	%	*N*	%	*N*
Hypoglycemia-related medical	$395	30.1	$167	12.7	$75	45.6	$15	11.6	$66

[a] All variables were collected in the 1 year following the index date

[b] Represents the count of patients with the relevant study outcome event or visit within ventiles. Patients with multiple visits were only counted once. For cost variables, represents the average costs per patient per year ($/PPY) within ventiles

[c] Represents the percentage of patients with the relevant study outcome event or visit within ventiles out of the total number of relevant study outcome events or visits. For costs, represents the percentage of costs within ventiles out total costs

hypoglycemia model performed comparably to other machine learning models applied in a similar, but more homogeneous cohort [18]. In addition, we determined that significant healthcare resource utilization was densely concentrated among patients with the highest risk of hypoglycemia, providing an estimate for the utility of our models if deployed in real-world patient populations.

Previous Value-Based Analyses

Historically, clinical trials and cost-effectiveness research have been the primary sources of evidence supporting value-based contracts. However, the former often fails to capture real-world scenarios involving patients outside strict control of the trial, as well as information on healthcare resource utilization and cost; while the latter (in the form of quality-adjusted life years (QALYs), incremental cost-effectiveness ratios (ICERs), etc.) often fails to capture important details of the individual patient experience [19]. This study offers an alternative solution, by leveraging predictive models applied to real-world data to (1) target interventions or risk-sharing strategies towards patients on the basis of identified adverse and/or protective factors of outcomes; or (2) target interventions or risk-sharing strategies towards patients on the basis of model predictions directly.

Utility of "Top Predictors" Selected by Model

Among variables most frequently selected by models across the ensemble, several are well established in previous literature. Notably, patients who indexed on insulin prescriptions (2.84, 2.79–3.03) and sulfonylurea prescriptions (1.80, 1.75–1.81) had increased risk of hypoglycemia, whereas patients who indexed on biguanide prescriptions (0.75, 0.74–0.75) had reduced risk. This is consistent with use of insulin being the prime cause of hypoglycemia, and sulfonylureas being a known cause, most frequently in combination with other medications. Likewise, pre-index hypoglycemia was an

expected risk factor for hypoglycemia (25.61, 23.55–25.61), given that patients in this study were prevalent diabetics taking antidiabetic medications where recurrent hypoglycemia is common.

Apart from antidiabetic medications, furosemide prescriptions (1.29, 1.29–1.30) were highly predictive and positively associated with hypoglycemic events and may merit further study. In particular, the association of furosemide with other conditions which are indicative of more serious diabetes could suggest an epiphenomenon, or the potential impact of kidney disease on drug metabolism.

Other risk factors frequently selected by the hypoglycemia models were more specific to healthcare administration data, including Medicare plan type (vs. Commercial 1.32, 1.32–1.35), and Southern (vs. Midwest 1.39, 1.38–1.40) and Western (vs. Midwest 1.12, 1.12–1.12) regions. In addition to having the highest prevalence of diabetes in the USA, health in the South is generally worse and associated with critical risk factors including obesity and sedentary lifestyle [20]. In these models, region may proxy for these risk factors, which are not captured in claims data. Alternatively, these variables may also reflect differences in healthcare administration, or access to care across the USA and may be particularly important for rollout of a value-based contract on a national scale.

Model Accuracy and Ensembling

An important aspect of predictive models is their ability to perform accurately in diverse patient populations. Model ensembling is an attractive solution to the problem of overfitting on a single data set, and in some cases can be invaluable for practitioners seeking to deploy models across multiple healthcare entities—a likely case in the US marketplace. For example, an interaction between "low-income subsidy status" and "product type: health maintenance organization (HMO)" was in 7 of 128 models in the full T2D-related inpatient admissions ensemble, implying this relationship was captured by a different set of variables in the remaining 121 models. Therefore, if the model was applied to data sets where this variable did not exist, its performance would not be significantly altered.

Diverse Outcomes for Value-Based Contracting

A diverse set of outcomes were explored in this study in order to capture both the clinical and economic impact of T2D. Each model, representing a distinct outcome measure, contains a set of variables that REFS determined were most predictive of the associated outcome. Importantly, many of these predictors co-occur across different outcome models. Patients with high baseline HbA1c levels, for example, were less likely to hit their HbA1c target thresholds, but more likely to experience significant changes in their HbA1c and be persistent to their antidiabetic prescriptions.

Similar conclusions can be drawn from the concentration of observed outcomes within predicted risk categories of different models. In particular, patients in the top percentiles of hypoglycemia risk also constitute large proportions of other adverse clinical and economic outcomes. This may be a particularly useful result to health policymakers requiring simpler risk-adjustment schema; that is, targeting patients for preventive interventions in T2D can be done effectively through a single outcome measure (e.g., hypoglycemia) rather than attempting to solve a complex optimization problem across multiple outcomes. Furthermore, as value-based contracts must ultimately be adopted by patients and physicians to be successful, they should necessarily present clear clinical advantages and objective outcome measures, respectively [5].

Limitations

One limitation of this study related to sample selection. The HbA1c target attainment outcome required at least one post-index laboratory measurement for HbA1c in addition to inclusion/exclusion criteria listed above. The change from baseline HbA1c outcome similarly

required both pre- and post-index laboratory measurements for HbA1c, and additionally that patients' pre-index HbA1c level was at least 8%. As such, models built for these outcomes have slightly different study populations and should be interpreted accordingly.

A second limitation of this study relates a fundamental aspect of predictive modeling; specifically, it cannot answer questions of a causal nature. This is especially true for administrative claims data, where medications, diagnoses, procedures, laboratory tests, etc. are heavily confounded by a variety of factors, notably:

- Access to healthcare—proximity to medical facilities, cultural obstacles related to treatment, educational and socioeconomic status
- Bias on the part of the physician—past experiences with a particular treatment, history with a particular patient
- Patients' underlying disease severity (i.e., confounding by indication)
- Non-random treatment assignment for novel therapies (i.e., channeling bias) [21]
- Healthy adherer bias [22].

In the context of this study, antidiabetic prescriptions (such as insulin) can be stronger predictors of patients' underlying T2D severity than T2D diagnosis codes extracted from medical claims, as they represent a decision made on behalf of the provider to intensify treatment of the patient's T2D. Likewise, missing laboratory values can often be proxies for healthy patients, as they potentially represent a decision by the provider *not* to order the laboratory test given the patient's current disease status. For example, "glucose (missing test)" was negatively associated with hypoglycemia (0.40, 0.40–0.40). Indeed, in some cases these proxies may be effective controls for unmeasured confounding [23]. Nevertheless, these variables serve to build highly predictive models whose output can be valuable despite confounding.

CONCLUSIONS

Machine learning models built with real-world data can be informative tools for the development of value-based contracts. The models developed in this study accurately identified patients at risk of critical clinical outcomes, including hypoglycemia, while uncovering the primary risk factors of those outcomes in a real-world setting. The ability of Bayesian network models to extract and perform inference in high-dimensional data is essential for guiding clinical practice, where "black box" technology is untrustworthy and holistic approaches are merited. This study also demonstrated the potential economic impact of deploying machine learning models in practice by quantifying resource use and costs in patients with the highest predicted risk of hypoglycemia.

Indeed, targeted interventions towards high-risk patients may additionally benefit economic outcomes relevant to key stakeholders in competitive therapeutic areas like diabetes. As the healthcare industry moves toward value-based care reimbursement models, it becomes critical for pharmaceutical companies to work effectively with health plans by providing key insights into optimal treatment intervention strategies for members at risk of poor clinical and economic outcomes.

ACKNOWLEDGEMENTS

Funding. This work was supported by sanofi-aventis U.S. LLC (Bridgewater Township, NJ). Publishing and the Rapid Service Fee were funded by GNS Healthcare (Cambridge, MA).

Medical Writing and/or Editorial Assistance. The authors would like to thank Rabee Zyoud of GNS Healthcare for his assistance in data collection and processing, and Kenneth Bennett of Sanofi for his informed review of our methods and results.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship for this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Disclosures. Paulos Berhanu, Jonathan Bouchard, and Lesley-Ann Miller-Wilson have direct financial relationships with Sanofi Pharmaceuticals. Luke Mueller, Veronica Alas, Kenneth Elder, Ngoc Thai, Cody Hitchcock, Tiffany Hadzi, and Iya Khalil have indirect financial relationships with Sanofi Pharmaceuticals through their employer, GNS Healthcare, who was consulted for this work.

Compliance with Ethics Guidelines. The Optum Clinformatics Data Mart data used in this study was compliant with the Health Insurance Portability and Accountability Act. Fully anonymized retrospective data were obtained from Optum via license agreement, and no other data were used in this study. As such, this study was deemed exempt from ethical approval.

REFERENCES

1. Papanicolas I, Woskie LR, Jha AK. Health care spending in the United States and other high-income countries. JAMA. 2018;319(10):1024–39. https://doi.org/10.1001/jama.2018.1150.

2. American Diabetes Association. Economic costs of diabetes in the U.S. in 2017. Diabetes Care. 2018;41(5):917–28. https://doi.org/10.2337/dci18-0007.

3. Neumann PJ, Chambers JD, Simon F, Meckley LM. Risk-sharing arrangements that link payment for drugs to health outcomes are proving hard to implement. Health Aff (Millwood). 2011;30(12):2329–37. https://doi.org/10.1377/hlthaff.2010.1147.

4. Pugatch M, Healy P, Chu R. Sharing the burden: could risk-sharing change the way we pay for healthcare? London, Great Britain: The Stockholm Network; 2010. https://www.slideshare.net/agabutti/sharing-the-burden. Accessed 12 Nov 2018.

5. Garrison LP, Towse A, Briggs A, et al. Performance-based risk-sharing arrangements-good practices for design, implementation, and evaluation: report of the ISPOR good practices for performance-based risk-sharing arrangements task force. Value Health. 2013;16(5):703–19. https://doi.org/10.1016/j.jval.2013.04.011.

6. Groves P, Kayyali B, Knott D, Van Kuiken S. The "big data" revolution in healthcare: accelerating value and innovation. McKinsey & Company; 2013. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care. Accessed 12 Nov 2018.

7. Rose S. A machine learning framework for plan payment risk adjustment. Health Serv Res. 2016;51(6):2358–74. https://doi.org/10.1111/1475-6773.12464.

8. HCUP Clinical Classifications Software (CCS) for ICD-9-CM. Rockville, MD: Agency for Healthcare Research and Quality; 2017. www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Accessed 10 Dec 2016.

9. HCUP Clinical Classifications Software (CCS) for ICD-10-CM/PCS. Rockville, MD: Agency for Healthcare Research and Quality; 2018. www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp. Accessed 10 Dec 2016.

10. Medi-Span Generic Product Identifier (GPI). Hudson Ohio: Wolters Kluwer Clinical Drug Information, Inc.; 2018. https://www.wolterskluwercdi.com/drug-data/gpi/. Accessed 10 Dec 2016.

11. Berenson-Eggers Type of Service (BETOS). Baltimore, MD: U.S. Centers for Medicare & Medicaid Services; 2016. https://www.cms.gov/Medicare/Coding/HCPCSReleaseCodeSets/BETOS.html. Accessed 10 Dec 2016.

12. Friedman N, Koller D. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. Mach Learn. 2003;50:95–125.

13. Xing H, McDonagh PD, Bienkowska J, et al. Causal modeling using network ensemble simulations of genetic and gene expression data predicts genes involved in rheumatoid arthritis. PLoS Comput Biol. 2011;7(3):e1001105. https://doi.org/10.1371/journal.pcbi.1001105.

14. Anderson JP, Parikh JR, Shenfeld DK, et al. Reverse engineering and evaluation of prediction models for progression to type 2 diabetes: an application of machine learning using electronic health records. J Diabetes Sci Technol. 2015;10(1):6–18. https://doi.org/10.1177/1932296815620200.

15. Steinberg GB, Church BW, McCall CJ, Scott AB, Kalis BP. Novel predictive models for metabolic syndrome risk: a "big data" analytic approach. Am J Manag Care. 2014;20(6):e221–8.

16. Latourelle JC, Beste MT, Hadzi TC, et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. Lancet Neurol. 2017;16(11):908–16. https://doi.org/10.1016/S1474-4422(17)30328-9.

17. Kessler RC, Warner CH, Ivany C, et al. Predicting suicides after psychiatric hospitalization in US Army soldiers: the Army Study To Assess Risk and rEsilience in Servicemembers (Army STARRS). JAMA Psychiatry. 2015;72(1):49–57. https://doi.org/10.1001/jamapsychiatry.2014.1754.

18. Pettus J, Roussel R, Liz Zhou F, et al. Rates of hypoglycemia predicted in patients with type 2 diabetes on insulin glargine 300 U/ml versus first- and second-generation basal insulin analogs: the real-world LIGHTNING study. Diabetes Ther. 2019;10(2):617–33. https://doi.org/10.1007/s13300-019-0568-8.

19. Tsevat J, Moriates C. Value-based health care meets cost-effectiveness analysis. Ann Intern Med. 2018;169(5):329–32. https://doi.org/10.7326/M18-0342.

20. Barker LE, Kirtland KA, Gregg EW, Geiss LS, Thompson TJ. Geographic distribution of diagnosed diabetes in the U.S.: a diabetes belt. Am J Prev Med. 2011;40(4):434–9. https://doi.org/10.1016/j.amepre.2010.12.019.

21. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. Med Care. 2010;48(6 Suppl):S114–20. https://doi.org/10.1097/MLR.0b013e3181dbebe3.

22. Lobo FS, Wagner S, Gross CR, Schommer JC. Addressing the issue of channeling bias in observational studies with propensity scores analysis. Res Social Adm Pharm. 2006;2(1):143–51. https://doi.org/10.1016/j.sapharm.2005.12.001.

23. Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. Med Care. 2007;45(10 Supl 2):S143–8. https://doi.org/10.1097/mlr.0b013e318074ce79.