PREDICTIVE MODELING OF HYPOGLYCEMIA RISK AMONG PATIENTS WITH TYPE 2 DIABETES (T2D) USING ENSEMBLE-BASED, HYPOTHESIS-FREE PREDICTIVE MODELING

NGOC H. THAI[1], LJ WEI[2], VERONICA ALAS[1], IYA KHALIL[1], PAULOS BERHANU[3], MEHUL R. DALAL[4], JENNIFER C. SUNG[3]

[1]GNS HEALTHCARE, INC., CAMBRIDGE, MA; [2]HARVARD T.H. CHAN SCHOOL OF PUBLIC HEALTH, BOSTON, MA; [3]SANOFI US, INC., BRIDGEWATER, NJ; [4]MILLENNIUM PHARMACEUTICALS, INC., A WHOLLY OWNED SUBSIDIARY OF TAKEDA PHARMACEUTICAL COMPANY LIMITED, CAMBRIDGE, MA; USA

INTRODUCTION

- Use of antidiabetes therapies is associated with a risk of hypoglycemia, which can affect quality of life and result in increased rates of mortality and morbidity in patients with type 2 diabetes (T2D).[1-3]
- Hypoglycemia – and fear of hypoglycemia – can limit the use of antidiabetes therapies prescribed to achieve and maintain the glycemic control needed to prevent diabetes complications.[2,4,5]
- Not much has been published around identifying key predictors of hypoglycemia. Accurate estimation of the incidence of hypoglycemia in T2D patients is difficult due to differences in definitions, methods of data collection, patient populations, and drug type and regimen used to treat patients.[2,6]
  - hypoglycemia occurs most frequently with insulin and sulfonylureas, but treatment with other agents (including those that are typically associated with lower rates of hypoglycemia, which are taken by a large number of patients) can also lead to severe hypoglycemia[4,7]
  - selecting antidiabetes agents with a relatively low risk of hypoglycemia has the potential to reduce hypoglycemia events[4]
- Identifying patients at risk of hypoglycemia, followed by actions to reduce the risk of hypoglycemia, has the potential to improve outcomes in patients with T2D, lower associated health care costs, and increase options for early therapy intensification.

OBJECTIVE

- To establish and validate a prediction model for hypoglycemia events over a 1-year follow-up period from patients' baseline information collected 6 months prior to starting medication using REFS™ (Reverse Engineering and Forward Simulation).

METHODS

Data Source and Study Population

- Data for this retrospective cohort study were collected between January 1, 2008, and December 31, 2013, consisting of administrative claims and laboratory records from the Truven Health MarketScan® Commercial, Medicare Supplemental, and Laboratory databases.
- Included patients were adults with an eligible T2D diagnosis code (ICD-9-CM: 250.xx[8]) who had a pharmacy claim for an antidiabetes medication.
  - the index date is the date of the first observed antidiabetes medication claim
- Patients were required to have continuous enrolment for 18 months (6 months pre- and 12 months post-index).

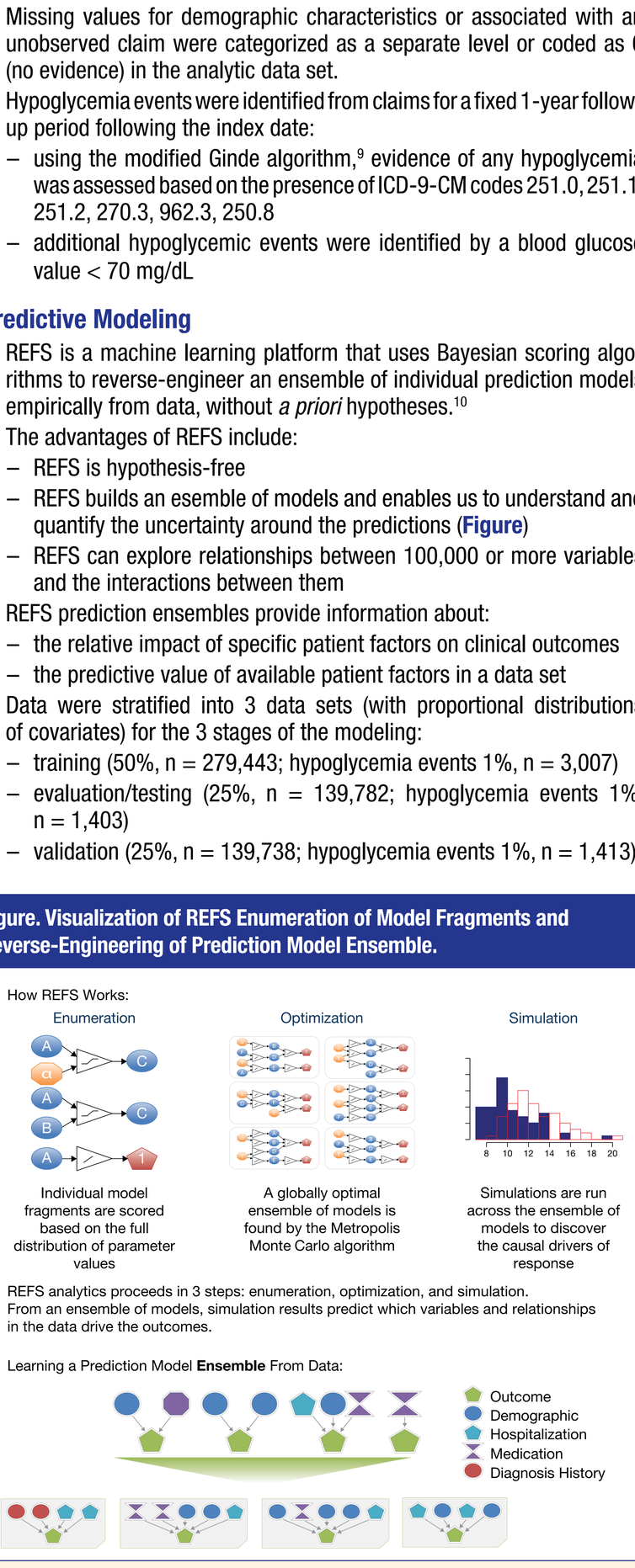
Study Measures

- Baseline data were collected during the 6-month pre-index period comprising:
  - demographic characteristics
  - clinical characteristics: comorbidities (Charlson Comorbidity Index [CCI], all available diagnoses via ICD-9-CM codes), history of hypoglycemia (prior to index date)
  - medications: Medi-Span drug classes identified via National Drug Codes
  - health care resource utilization: inpatient, outpatient, and emergency room [ER] services, specialist visits, health care costs
- Missing values for demographic characteristics or associated with an unobserved claim were categorized as a separate level or coded as 0 (no evidence) in the analytic data set.
- Hypoglycemia events were identified from claims for a fixed 1-year follow-up period following the index date:
  - using the modified Ginde algorithm,[9] evidence of any hypoglycemia was assessed based on the presence of ICD-9-CM codes 251.0, 251.1, 251.2, 270.3, 962.3, 250.8
  - additional hypoglycemic events were identified by a blood glucose value < 70 mg/dL

Predictive Modeling

- REFS is a machine learning platform that uses Bayesian scoring algorithms to reverse-engineer an ensemble of individual prediction models empirically from data, without *a priori* hypotheses.[10]
- The advantages of REFS include:
  - REFS is hypothesis-free
  - REFS builds an esemble of models and enables us to understand and quantify the uncertainty around the predictions (**Figure**)
  - REFS can explore relationships between 100,000 or more variables and the interactions between them
- REFS prediction ensembles provide information about:
  - the relative impact of specific patient factors on clinical outcomes
  - the predictive value of available patient factors in a data set
- Data were stratified into 3 data sets (with proportional distributions of covariates) for the 3 stages of the modeling:
  - training (50%, n = 279,443; hypoglycemia events 1%, n = 3,007)
  - evaluation/testing (25%, n = 139,782; hypoglycemia events 1%, n = 1,403)
  - validation (25%, n = 139,738; hypoglycemia events 1%, n = 1,413)

Figure. Visualization of REFS Enumeration of Model Fragments and Reverse-Engineering of Prediction Model Ensemble.



- In the training stage, the models were selected for inclusion in the REFS ensemble.
  - 128 logistic regression models (each consisting of ≤ 25 variables) were selected for the ensemble
- In the evaluation stage, the most predictive model was selected.
  - cross-validated area under the curve (AUC) was used to evaluate all of the 128 candidate models
  - the best logistic regression model from the REFS ensemble was the one that achieved the highest AUC
- In the validation stage, hypoglycemia risk score was fixed to a 5% threshold (≤ 5%, > 5%).
  - point estimates for 1-year hypoglycemia risk and 95% confidence interval (CI) were calculated for each risk interval

RESULTS

Demographics, Comorbidity, and Medication Use

- A total of 558,963 patients with T2D were included in the analysis (**Table 1**):
  - mean age 54.6 years, 48.4% female, mean (SD) CCI score 1.90 (1.16), 5.2% using insulin
  - a total of 1.0% of patients (n = 5,823) had ≥ 1 hypoglycemia event over a 1-year follow-up period

Table 1. Demographics, Comorbidity, and Medication Use.

Variable	No Hypoglycemia n = 553,140 (99.0%)	Any Hypoglycemia n = 5,823 (1.0%)	*P* Value
Demographic characteristics			
Age, mean (SD), years	54.6 (12.6)	54.0 (14.8)	< 0.001
Gender			
Male	285,911 (51.7)	2,491 (42.8)	< 0.001
Female	267,229 (48.3)	3,332 (57.2)	
Comorbidities			
Hypertension	268,415 (48.5)	2,917 (50.1)	0.018
Dyslipidemia	231,730 (41.9)	2,535 (43.5)	0.012
Chronic pulmonary disease	45,571 (8.2)	793 (13.6)	< 0.001
Obesity	43,288 (7.8)	696 (12.0)	< 0.001
Depression	32,192 (5.8)	591 (10.2)	< 0.001
Retinopathy	28,638 (5.2)	403 (6.9)	< 0.001
Cancer	21,962 (4.0)	288 (5.0)	< 0.001
Nephropathy	20,652 (3.7)	435 (7.5)	< 0.001
CCI score			
Mean (SD)	1.9 (1.2)	1.8 (1.3)	< 0.001
Medications			
Biguanides	372,549 (67.4)	3,307 (56.8)	< 0.001
Sulfonylureas	78,156 (14.1)	1,185 (20.4)	< 0.001
Insulin	28,570 (5.2)	573 (9.8)	< 0.001
DPP-4 inhibitors	20,633 (3.7)	231 (4.0)	0.361
Sulfonylurea–biguanide combinations	11,275 (2.0)	173 (3.0)	< 0.001
GLP-1 receptor agonists	7,573 (1.4)	93 (1.6)	0.152
Meglitinide analogs	2,388 (0.4)	46 (0.8)	< 0.001
α-glucosidase inhibitors	660 (0.1)	132 (2.3)	< 0.001

Values are n (%) unless stated otherwise.
DPP-4, dipeptidyl peptidase-4; GLP-1, glucagon-like peptide-1; SD, standard deviation.

Model Training

- Model candidates comprised 128 logistic regression models.
- Key predictors identified by REFS (selection frequency ≥ 90%) were (**Table 2**)
  - young age, thyroid disorders, baseline inpatient costs, baseline hypoglycemia, sulfonylurea use, sulfonylurea–biguanide combination use, baseline inpatient hospital visits (T2D-related), and upper gastrointestinal disorders

Table 2. Key Predictors Identified in the REFS Ensemble.

Predictor	Selection Frequency, %
Age group	100
Thyroid disorders	100
Baseline inpatient costs	100
Baseline hypoglycemia	100
Sulfonlyureas	100
Sulfonlyurea–biguanide combinations	99
Baseline inpatient hospitalization (T2D-related)	95
Upper gastrointestinal disorders	93
Anxiety disorders	72
Headache, including migraine	58
Human insulin	57
HMG-CoA reductase inhibitors (statins)	52
Biguanides	47
Baseline inpatient hospital visits	43
Baseline pharmacy costs (T2D-related)	37
Diseases of the heart	24
Baseline ER costs	18
Baseline inpatient costs (T2D-related)	17
Mood disorders	16
Disorders of lipid metabolism	12
Baseline ER visits	10
ACE inhibitors and thiazide/thiazide-like therapy	9
Other lower respiratory disease	7
Other upper respiratory disease	5

Only (main-effect) predictors with selection frequency ≥ 5% are shown.
ACE, angiotensin-converting enzyme; HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A.

Model Evaluation

- The best logistic regression model chosen had a cross-validated AUC of 0.73.
- Predictors from the best-performing model were (**Table 3**):
  - diseases of the female genital organs, oldest vs youngest age (≥ 75 vs < 35 years), thyroid disorders, mental disorders, other lower respiratory disease, upper gastrointestinal disorders, baseline inpatient hospitalizations (T2D-related), baseline hypoglycemia, insulin use, sulfonylurea use, sulfonylurea–biguanide combination use, baseline ER costs, and baseline inpatient costs

Model Validation

- Patients were classified into 2 categories based on the predicted risk of hypoglycemia (≤ 5%, > 5%), and corresponding mean risk estimates were calculated (**Table 4**).
- The mean risk estimate of patients with predicted risk ≤ 5% is 1.3% (95% CI 1.2-1.4), while the mean risk estimate of patients with predicted risk > 5% is 14.0% (95% CI 12.5-15.6).

Table 3. Predictors Identified in the Best Model Selected From Model Evaluation.

Predictor	Odds Ratio	*P* Value
Age 18-34 years	1.54	< 0.001
Age 45-54 years	0.91	0.105
Age 55-64 years	0.82	0.001
Age 65-74 years	0.81	0.011
Age ≥ 75 years	1.15	0.071
Diseases of the female genital organs	1.26	< 0.001
Thyroid disorders	1.52	< 0.001
Mental disorders	1.90	< 0.001
Upper gastrointestinal disorders	1.41	< 0.001
Other lower respiratory disease	1.33	< 0.001
Other lower respiratory disease + baseline hypoglycemia	0.54	< 0.001
Baseline inpatient hospitalizations (T2D-related)	0.79	< 0.001
Baseline hypoglycemia	24.68	< 0.001
Human insulin	1.63	< 0.001
Sulfonylureas	1.49	< 0.001
Biguanides	0.91	0.163
Sulfonylurea–biguanide combinations	2.00	< 0.001
Sulfonylureas + biguanides	1.77	< 0.001
Baseline ER costs	1.04	< 0.001
Baseline inpatient costs	1.10	< 0.001
Baseline inpatient hospitalizations (T2D-related) + biguanides	0.67	< 0.001
Baseline hypoglycemia + human insulins	0.40	< 0.001
Baseline hypoglycemia + sulfonylureas	0.46	< 0.001

Table 4. Risk Estimated From the Validation Data Set (n = 139,738).

Hypoglycemia Risk	Number of Patients	Number of Patients With Hypoglycemia	Mean Risk Estimate, %	Lower 95% CI, %	Upper 95% CI, %
≤ 5%	137,659	1,788	1.3	1.2	1.4
> 5%	2,079	291	14.0	12.5	15.6

Risk prediction was calculated for each patient in the validation set; mean risk estimated across all patients is reported. 95% CI for the rate of hypoglycemia was calculated based on the hypoglycemia event status of the patients in the follow-up period.

LIMITATIONS

- Limitations of this analysis relate mainly to the use of claims data:
  - limited visibility into history of hypoglycemia
  - difficult to ensure identification of hypoglycemia events
  - index date does not represent comparable disease stage for all patients
  - some desired variables related to T2D are not (consistently) available in claims data, such as body mass index and disease duration.

DISCUSSION AND CONCLUSIONS

- Traditional regressions are hypothesis-driven (i.e. only "relevant" data will be explored) and a limited number of variables (10-15) are selected from the literature. Insights are limited to the population- and individual level. In predictive analytics there is unlimited variable selection, no prior hypothesis, and insights are relevant to multiple populations.
- The REFS machine-learning platform is not limited by prior knowledge regarding the importance of variables, but uses all the data and can scale to explore relationships between hundreds of thousands of variables. This allows for the creation of the best set of unified hypotheses that are not based on previous research.
- For every type of scientific question, an ensemble of models was built, allowing for quantification of the uncertainty around predictions.
- Patients who tended to have a relatively high risk (> 5%) of developing hypoglycemia within a year of follow-up were those who: had experienced hypoglycemia previously; were older (≥ 75 years); used insulin; used sulfonylureas; had mental disorders; or had higher health care resource utilization and costs.
- Unexpectedly, patients who had: diseases of the female genital organs; thyroid disorders; other lower respiratory disease; upper gastrointestinal disorders; or had more use of other forms of health care utilization, were also at risk of developing hypoglycemia.
- The increasing risk across the risk strata demonstrates an incremental risk for hypoglycemia events via independent validation.
- We show that the prediction model for the hypoglycemia incidence rates performs reasonably well via validation.
- This analysis presents a systematic procedure to quantify the risk for hypoglycemia and demonstrates an incremental risk of events following the first observable antidiabetes therapies in T2D patients.
- Prevention and early identification of hypoglycemia is necessary to reduce the clinical and economic burden in patients with T2D.
- The unexpected predictors identified in this study add new information that clinicians could consider for risk–benefit of antidiabetes therapies.
- Further studies should be conducted to determine whether these findings are by-products or nuances of the modeling and statistical analysis, and if they are of clinical relevance.

REFERENCES

1. Holman RR, et al. N Engl J Med. 2008;359:1577-89.
2. Amiel SA, et al. Diabet Med. 2008;25:245-54.
3. Barnett AH, et al. Int J Clin Pract. 2010;64:1121-9.
4. Ahrén B. Vasc Health Risk Manag. 2013;9:155-63.
5. Peyrot M, et al. Diabetes Care. 2005;28:2673-9.
6. Zammitt NN, Frier BM. Diabetes Care. 2005;28:2948-61.
7. Inzucchi SE, et al. Diabetes Care. 2015;38:140-9.
8. CDC. www.cdc.gov/nchs/icd/icd9cm.htm. Accessed September 12, 2016.
9. Ginde AA, et al. BMC Endocr Disord. 2008;8:4.
10. Anderson JP, et al. J Diabetes Sci Technol. 2015;10:6-18.

ACKNOWLEDGMENTS AND DISCLOSURES

This study was funded by Sanofi US, Inc. The authors received writing/editorial support in the preparation of this poster provided by Tessa Hartog, PhD, of Excerpta Medica, funded by Sanofi US, Inc., and gratefully acknowledge the support they received from Steve Zhou, who was employee at Sanofi US, Inc., when this poster was initiated.Thai and Alas are employees of GNS Healthcare. Wei is a consultant for Sanofi, Merck, Pfizer, J&J, and Novartis. Anderson is a former employee of GNS Healthcare. Zhou and Berhanu are employees of Sanofi US, Inc. Sung is an employee of Sanofi US, Inc., and owns stock of Sanofi US, Inc. Dalal is a former employee of Sanofi US, Inc.

Email address for presenting author: nthai@gnshealthcare.com