ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Decision-Analytic Modeling: Past, Present, and Future

Using Machine Learning Applied to Real-World Healthcare Data for Predictive Analytics: An Applied Example in Bariatric Surgery

Stephen S. Johnston, MA,[1,*] John M. Morton, MD,[2] Iftekhar Kalsekar, PhD,[1] Eric M. Ammann, PhD,[1] Chia-Wen Hsiao, MPH,[3] Jenna Reps, PhD[4]

[1]Epidemiology, Medical Devices, Johnson & Johnson, New Brunswick, NJ, USA; [2]Department of Surgery, Stanford University, Stanford, CA, USA; [3]Ethicon, Johnson & Johnson, Somerville, NJ, USA; [4]Epidemiology, Johnson & Johnson, Titusville, NJ, USA

ABSTRACT

Objectives: Laparoscopic metabolic surgery (MxS) can lead to remission of type 2 diabetes (T2D); however, treatment response to MxS can be heterogeneous. Here, we demonstrate an open-source predictive analytics platform that applies machine-learning techniques to a common data model; we develop and validate a predictive model of antihyperglycemic medication cessation (validated proxy for A1c control) in patients with treated T2D who underwent MxS.

Methods: We selected patients meeting the following criteria in 2 large US healthcare claims databases (Truven Health MarketScan Commercial [CCAE]; Optum Clinformatics [Optum]): underwent MxS between January 1, 2007, to October 1, 2013 (first = index); aged ≥18 years; continuous enrollment 180 days pre-index (baseline) to 730 days postindex; baseline T2D diagnosis and treatment. The outcome was no antihyperglycemic medication treatment from 365 to 730 days after MxS. A regularized logistic regression model was trained using the following candidate predictor categories measured at baseline: demographics, conditions, medications, measurements, and procedures. A 75% to 25% split of the CCAE group was used for model training and testing; the Optum group was used for external validation.

Results: 13 050 (CCAE) and 3477 (Optum) patients met the study inclusion criteria. Antihyperglycemic medication cessation rates were 72.9% (CCAE) and 70.8% (Optum). The model possessed good internal discriminative accuracy (area under the curve [AUC] = 0.778 [95% CI = 0.761-0.795] in CCAE test set N = 3527) and transportability (external AUC = 0.759 [95% CI = 0.741-0.777] in Optum N = 3477).

Conclusion: The application of machine learning techniques to real-world healthcare data can yield useful predictive models to assist patient selection. In future practice, establishment of prerequisite technological infrastructure will be needed to implement such models for real-world decision support.

Keywords: prediction, machine learning, type 2 diabetes, metabolic surgery, antihyperglycemic medication

VALUE HEALTH. 2019; 22(5):580–586

Introduction

Among individuals with type 2 diabetes (T2D) and body mass index (BMI) of ≥35 kg/m^2, laparoscopic metabolic surgery (including Roux-en-Y gastric bypass and sleeve gastrectomy, also commonly referred to as "bariatric surgery") is recognized by the American Diabetes Association as one of the most effective means to achieve substantial, sustained weight loss, improved glycemic control, and in many cases remission of T2D.[1]

Nevertheless, not all patients with T2D who undergo metabolic surgery experience remission. This uncertainty of response to surgery, combined with a wide variety of other factors, including nonuniform health insurance coverage and access, high out-of-pocket costs, self-perception, and patient fear of undergoing surgery or complications, have all contributed to a low uptake of metabolic surgery among those who are surgically eligible (~1%).[2–7]

Patient-level prediction (PLP) models provide a means by which an individual can be assigned a predicted probability of experiencing a particular outcome (beneficial or harmful) based on a set of predictors such as their demographics, comorbidities, or other factors. These models may be used in clinical and economic decision support to target treatments to individuals for whom there is the high probability of benefit, or away from individuals for whom there is a high probability of harm.

In the context of metabolic surgery in individuals with T2D, a variety of PLP tools have been developed for the outcome of T2D remission—most of which are operationalized as discrete integer

* Address correspondence to: Stephen S. Johnston, MA, Epidemiology, Medical Devices, Johnson & Johnson, 410 George Street, New Brunswick, NJ 08901, USA.
Email: sjohn147@its.jnj.com

https://doi.org/10.1016/j.jval.2019.01.011

scores, few of which have been externally validated in cohorts other than those in which they were developed, and most of which have been evaluated within relatively small samples (N ranging from 31 to 900).[8–15]

To our knowledge, however, no previous study has attempted to harness the vast amount of information contained in "big data" healthcare databases, such as administrative claims data, to predict outcomes after metabolic surgery.

Thus, using administrative claims data and an innovative PLP modeling platform (*Patient Level Prediction* software, an open-source R package[16]) developed by the Observational Health Data Sciences and Informatics (OHDSI) network, we sought to examine whether we could yield a useful PLP model for predicting complete antihyperglycemic medication cessation (as a proxy for A1c control) after metabolic surgery in patients with T2D and antihyperglycemic medication treatment.

Methods

OHDSI Network and PLP Framework

OHDSI is an open science collaborative with an international network of researchers and data partners who focus on methodological research, open-source analytics development, and clinical applications to advance the generation and dissemination of reliable medical evidence from observational data.[17] We followed the OHDSI PLP framework to develop and validate our PLP model conforming to community-defined best practices.[18]

The OHDSI PLP framework has several key innovations and strengths. First, the OHDSI PLP framework and software are applied to the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). The OMOP CDM is an open community standard for organizing the content of observational datasets into a homogenous data structure.[19] Use of the OMOP CDM within the framework enables a PLP study to be developed and executed in one database and then consistently executed across different healthcare databases stored in the OMOP CDM for external validation or clinical application purposes. Second, the PLP framework also uses the OHDSI Standard Vocabulary[20]—a set of standard clinical taxonomies for diagnosis codes, medications, observations, and so on (eg, SNOMED, LOINC, RxNorm)—to automatically generate a very high-dimensional feature set of candidate predictors (often numbering in the tens of thousands of predictors) for use in a PLP model. The feature set is based on the collection of observed diagnoses, medications, observations, and so on, that are present for the cohort in which the PLP model is being trained. A priori defined predictors may also be created and used in the PLP model; however, the automatic generation of a feature set is both very efficient from a programming perspective and allows for predictors with previously unknown predictive associations to contribute to the PLP model. Finally, the PLP framework and software generate a complete set of code, which is portable from researcher to researcher, to facilitate efficient replication of the PLP model and minimization of reproducibility errors.

Overview of Predictive Modeling Approach

We undertook the following steps to develop and validate our PLP model.[18] First, we first defined our target population in whom we wish to predict the outcome (patients undergoing metabolic surgery who had a baseline diagnosis of T2D and antihyperglycemic medication treatment); second, we defined our outcome (complete cessation of antihyperglycemic medication from 365 to 730 days after metabolic surgery); third, we selected a database in which we could obtain data on a large sample of the target population (Truven MarketScan Commercial Claims and Encounters [CCAE] Database); fourth, among the target population extracted from the database, we subdivided the sample into a training set (comprising 75% of the sample) in which we initially developed the PLP model and a separate test set (comprising the remaining 25% of the sample) in which we assessed the internal performance (discrimination/calibration) of the model; fifth, we selected a separate database in which we could obtain data on another large sample of the target population (Optum Clinformatics Database [Optum]) and applied the trained PLP model to assess the external validity of the model.

Sources of Data

We trained the model in the CCAE database and externally validated the model in the Optum database. The CCAE database comprises health insurance claims and encounter records for commercially insured employees and dependents in the United States and is the largest database for privately insured patients with longitudinal follow-up in all 50 states. The Optum database comprises health insurance claims data for a combination of US private insurance and Medicare Advantage beneficiaries. Both databases contain de-identified data derived from health plan members' enrollment data and facility, physician, and pharmacy claims. The CCAE and Optum databases include approximately 23 million and 13 million covered lives annually, respectively.

Target Population

The target population comprised patients meeting all following criteria: underwent laparoscopic metabolic surgery (either Roux-en-Y gastric bypass or sleeve gastrectomy) between January 1, 2007, to October 1, 2013 (first observed surgery during this period = index); aged ≥18 years at index; continuous observation (ie, insurance enrollment) of 180 days before (baseline) to 730 days after index; ≥1 baseline condition occurrence of T2D; and ≥1 baseline prescription fill for antihyperglycemic medication. Above, these criteria are defined in terms of the OMOP CDM nomenclature, but what underlies these criteria are sets of codes lists which are native to given databases and data sources (eg, International Classification of Diseases [ICD] and Current Procedural Terminology procedure codes, ICD diagnosis codes, and National Drug Codes), which have been mapped to the taxonomies of the Standard Vocabulary (eg, SNOMED and RxNorm) and can therefore translate across databases with disparate underlying native code systems.

Outcome and Validation of Outcome

We defined the outcome as no prescription fills for any antihyperglycemic medication from 365 to 730 days after metabolic surgery. This included both oral and injectable antihyperglycemic medications. We used this outcome as a proxy for A1c control, chosen owing to a lack of datasets that contain information on A1c (to measure A1c control/remission more precisely), have records for enough patients undergoing metabolic surgery to adequately train and externally validate a predictive model, or have comprehensive longitudinal data on enrollment and prescription medication use (as in the case with many electronic health record [EHR] databases). We conducted a small, ancillary outcome validation study to assess whether cessation of antihyperglycemic medication following metabolic surgery is an acceptable proxy for A1c control without continued use of antihyperglycemic medication (see Appendix 1 for methods and results in Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.01.011).

Table 1. Baseline characteristics of target populations

	CCAE (N = 13 050)		Optum (N = 3477)	
Age, mean and standard deviation	49.3	8.7	51.5	10.0
Female, n and %	9035	69.2%	2375	68.3%
5 most common baseline antihyperglycemic medications, n and %				
Metformin/biguanides	10 163	77.9%	2.635	75.8%
Sulfonylureas	4592	35.2%	1234	35.7%
Insulins and analogues	3977	30.5%	904	26.0%
Thiazolidinediones	3551	27.2%	874	25.1%
Dipeptidyl peptidase 4 (DPP-4) inhibitors	1745	13.4%	404	11.6%
5 most common baseline comorbid conditions, n and %				
Hypertensive disorder	9472	72.6%	3143	90.3%
Inflammation of specific body systems	7763	59.5%	2396	68.9%
Pain	7218	55.3%	2426	69.8%
Finding related to sleep	6676	51.2%	2244	64.5%
Hyperlipidemia	6673	51.1%	2792	80.3%
Roux-en-Y gastric bypass at index, n and %	10 012	76.7%	2473	71.1%
Adjustable gastric banding revision during baseline, n and %	261	2.0%	80	2.3%

CCAE, Truven MarketScan Commercial Claims and Encounters database.

Predictors

All candidate predictors used the patients' history recorded in the 180 days before index (baseline), unless otherwise noted. The candidate predictors included age group at index, sex, index month (to account for seasonality), the Diabetes Complications Severity Index,[21] SNOMED condition occurrences, SNOMED measurements and procedures, drug exposures mapped to ingredient level, and metabolic surgery-specific factors (Roux-en-Y gastric bypass vs sleeve gastrectomy at index; prior primary adjustable gastric banding, prior adjustable gastric banding revision).

Statistical Analysis Methods

We trained a lasso logistic regression model, a type of machine learning, within the CCAE database using one repetition of 10-fold cross-validation. Cross-validation is a resampling procedure used to select the optimal hyperparameter value while minimizing the optimistic area under the receiver operating characteristic curve (AUC) that would occur with overfitting. In 10-fold cross-validation, each lasso logistic regression with a specific variance (hyperparameter) has its performance estimated by splitting a training dataset into 10 sets, holding out 1 of these sets while training the model on the other 9 sets, and then evaluating the model on the holdout set. This process is iteratively repeated in each set to give an estimated cross-validation performance of the lasso logistic regression with the specific variance. The training dataset comprised a randomly selected 75% subset of the total dataset, with the remaining 25% held out to enable an internal validation of the model. We assessed model discrimination using the AUC and model calibration by inspecting a calibration plot. We externally validated the trained model by applying it in the Optum database and evaluating its model discrimination therein. The PatientLevelPrediction R package version 2.0.0 was used for model training, internal validation, and external validation.[16]

Sensitivity Analyses

As benchmarks for the internal validation, we also fit predictive models using Gradient Boosting Machine, Random Forest, and AdaBoost techniques, performing a grid search to select the optimal hyperparameters for each. We compared the resultant test AUCs for these models to that of the primary lasso logistic regression approach.

Results

Target Population

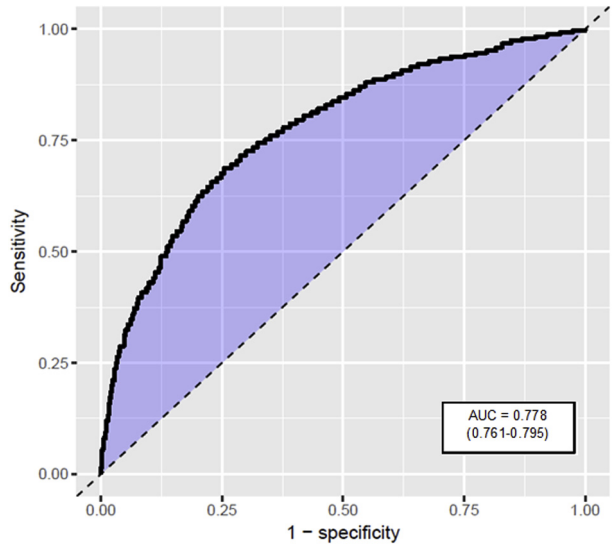
The study included 13 050 and 3477 patients from the CCAE and Optum databases, respectively. Among patients from the CCAE database, 9523 (72.9%) attained complete antihyperglycemic medication cessation between 1 and 2 years after metabolic surgery; in Optum, these numbers were 2461 (70.8%). Table 1 presents the baseline characteristics of the target population in the CCAE and Optum datasets.

Final Model Specification and Internal Validation Performance

From among 22 099 candidate predictors in the CCAE dataset, 125 predictors were selected through Lasso regression. The model's run time, hardware and processers used, and coefficients for each predictor are available in Appendix 2 (see Supplemental Materials found at https://doi.org/10.1016/j.jval.2019.01.011); however, the coefficients must be interpreted with caution owing to potential for multicollinearity, which is inherent with machine-learning based predictive models. The 3 predictors with the strongest negative association with antihyperglycemic medication cessation were baseline use of insulins and analogues, prior adjustable gastric banding revision, and increasing Diabetes Comorbidity Severity Index. The 3 predictors with the strongest positive association with antihyperglycemic medication cessation were baseline use of noninsulin glucose-lowering drugs, having undergone Roux-en-Y gastric bypass (vs sleeve gastrectomy), and younger age.

The internal validation of the model showed that the model possessed very good internal discriminative accuracy, with an AUC of 0.778 (0.761-0.795). Figures 1 and 2 present the receiver operating characteristic (ROC) plot and calibration plot for the external validation of the model, respectively. Figure 3 presents a scatterplot of the model's sensitivity, positive predictive value

Figure 1. Receiver operating characteristic plot for the internal validation of the model in the 25% CCAE validation set (N = 3527).



(PPV), specificity, and negative predictive value (NPV) by predicted probability thresholds. Under the circumstance of using a ≥50% predicted probability threshold to select patients for metabolic surgery, the model possesses 91.5% sensitivity, 79.4% PPV, 61.0% specificity, and 36.1% NPV.

External Validation Performance

The external validation of the model in Optum showed that the model possessed good transportability, with an external AUC of 0.759 (0.741-0.777). The ROC plot and calibration plot for the external validation of the model are presented in Figures 4 and 5, respectively.

Figure 2. Calibration plot* for the internal validation of the model in the 25% CCAE validation set (N = 3527). *The sample was split into 10 deciles, and the mean predicted probability of the outcome was plotted against the observed probability of the outcome for each decile. The dotted line represents perfect model calibration, with the expected risk neither under- nor overestimated across risk deciles.

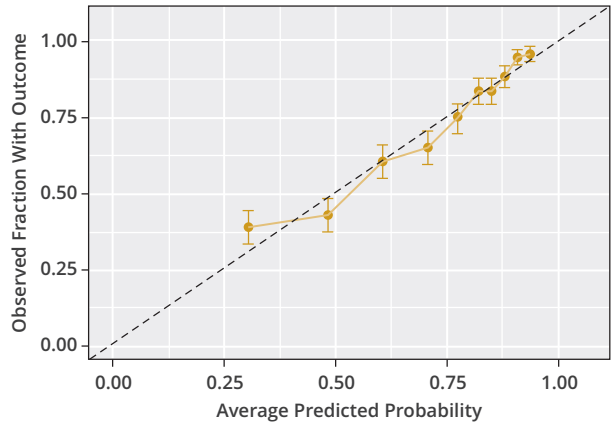
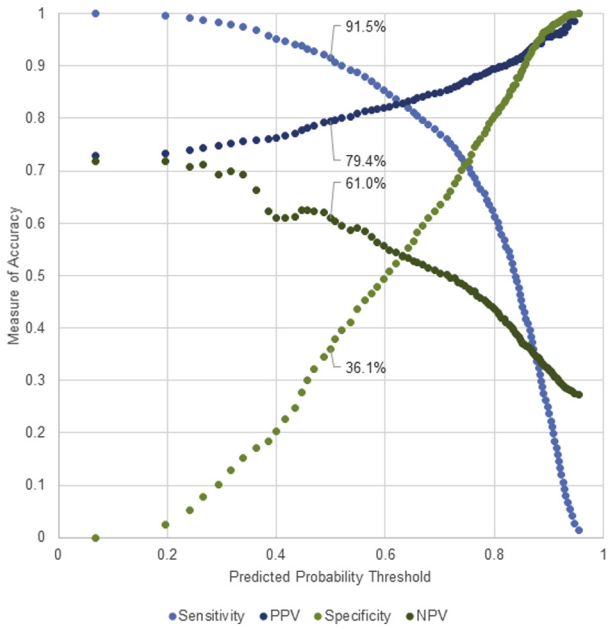


Figure 3. Sensitivity, positive predictive value, specificity, and negative predictive value of the trained model in the 25% CCAE validation set (N = 3527).



PPV, positive predictive value; NPV, negative predictive value.

Sensitivity Analyses

Sensitivity analyses of the internal validation step using alternative predictive model approaches yielded the following test AUC values: gradient boosting machine test AUC = 0.744; random forest test AUC = 0.788; and AdaBoost test AUC = 0.758. We retained lasso logistic regression as the primary analysis approach given that it yielded an internal test AUC of 0.778 and has the benefit of providing a comparatively more parsimonious and interpretable result versus the alternative approaches.

Discussion

We used a state-of-the-art R-based platform, applied to administrative claims data, to develop a PLP model for cessation of antihyperglycemic medication after laparoscopic metabolic surgery among patients with T2D. Our model performed well, with internal AUC of 0.778. It also generalized well to another dataset and population, with external AUC of 0.759.

Previous models predicting diabetes remission after metabolic surgery (for most of which formal measures of discrimination are not reported) have reported internal AUCs varying from 0.69 to 0.95 in samples with 46 to 103 patients and external AUCs varying from 0.71 to 0.86 in samples with 502 to 900 patients.[8,10,11] Our model's internal and external AUCs fall within the range of the aforementioned prior models; however, direct comparisons between our model and the prior models are limited by (1) differences in the definitions of diabetes remission, and (2) the inability to directly implement the prior models in administrative claims data, which lack the required information on BMI, C-peptide, duration of diabetes, and A1c.

Several considerations arise when contemplating how to practically implement PLP models for real-world decision support in general. Surely, many different stakeholders would benefit from

Figure 4. Receiver operating curve plot for the external validation of the model in Optum (N = 3477).

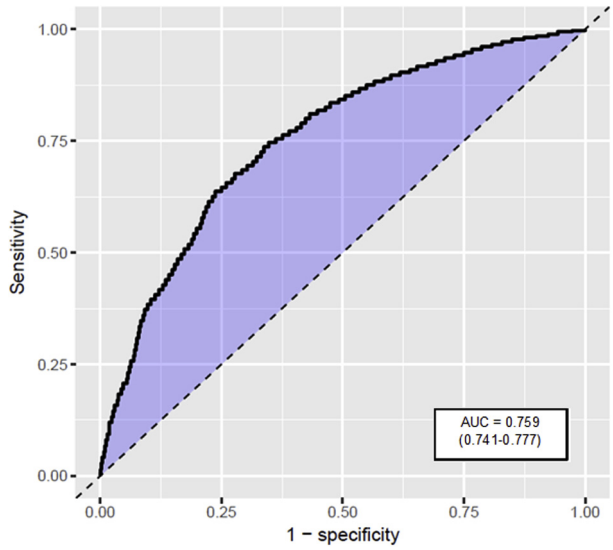
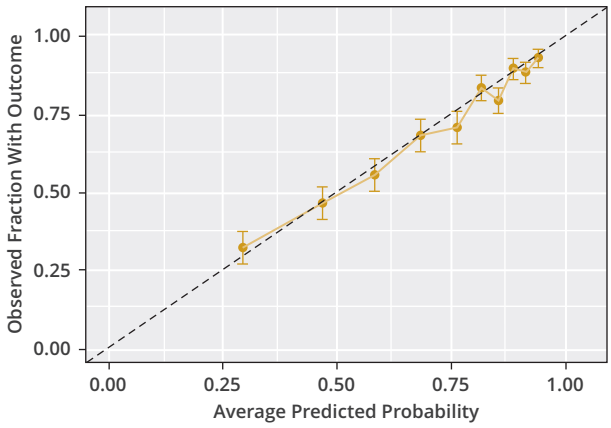


Figure 5. Calibration* plot for the external validation of the model in Optum (N = 3477). *The sample was split into 10 deciles, and the mean predicted probability of the outcome was plotted against the observed probability of the outcome for each decile. The dotted line represents perfect model calibration, with the expected risk neither under- nor overestimated across risk deciles.



the information a PLP model provides: patients wishing to know whether they can expect a desirable outcome if they undergo what they might perceive as an invasive surgical procedure; a physician who wishes to appropriately set expectations for their patients or who wishes to target metabolic surgery to only those in whom they feel confident it will meet treatment goals; a payer that may not provide universal coverage of elective metabolic surgery but wishes to maximize the expected benefits of the surgery when it is approved.

A first logical question is: *Who* would implement this model from a technical standpoint, and what are the prerequisite circumstances to do so? With the present form of our model, a prerequisite data requirement is access to a 6-month history of administrative claims data for the individual for whom the prediction will be made. Today, such data are most likely to lie in the hands of a payer. Healthcare providers who are members of integrated delivery networks may also be able to gain access to the needed history of their patients' administrative claims data; however, it is far more likely in the current and near term that healthcare providers will have access primarily to a patient's EHR. In the limitations section below, we discuss the subject of applying our PLP model in EHR data.

A second-order matter related to the prerequisite data is the *timeliness* of data. In our analysis, we used administrative claims data with service dates leading up until the day before the surgery was conducted. In practice, presurgical predictions to inform the decision of whether to undergo a surgery would likely be made weeks or even months ahead of the decision to plan a surgery. Furthermore, administrative claims data, unlike EHR data, typically have a lag associated with the adjudication of the claims, sometimes extending several months for inpatient claims. Thus, any prediction made based on administrative claims data would likely reflect the patient's status as of the recent past as opposed to the day on which the prediction was made. We investigated how this phenomenon might affect our predictive model in a post-hoc sensitivity analysis. In this post-hoc analysis, we trained the model using predictors measured over 365 days until 180 days before metabolic surgery. Use of less recent data did result in slight numeric decrement in the model AUC, from 0.778 to 0.755 for the internal validation analysis, and from 0.759 to 0.745 in the external validation analysis. The small decreases in the C-statistic indicate that it would be feasible to build a predictive model for antihyperglycemic medication cessation based on older claims data that would be available before the planned surgery date.

Regardless of *who* implements the prediction on what data, another important prerequisite to making the prediction is the technological capability. To easily translate the PLP model from one dataset to another, the datasets need to share a common data model (in our case, the OMOP CDM) and must have an interface into which they can "plug" a patient's data to compute the prediction. With respect to the former requirement, several CDMs exist, including the OMOP CDM, the PCORnet CDM, and the Sentinel CMD, among others; however, portability of PLP models across CDMs may not be possible.[22,23] Without cross-CDM translation, the implementation of PLP models may be fragmented by variation in CDM adoption. The latter requirement is relatively trivial given the simple computation required to generate a prediction once the underlying data are coupled with the PLP model. Thus, it is incumbent on us as a community of researchers to continue to develop and facilitate such tools which would enable the practical application of PLP models.

Envisioning the future when someone *will* implement this model in practice, what will they do with the prediction? If a patient has a 75% predicted probability of antihyperglycemic medication cessation, what does this mean to the patient, her healthcare provider, or her payer? What is the predicted probability threshold below which the patient would not undergo, be recommended for, or be approved for the metabolic surgery versus above which they would? From the patient perspective, preference studies and benefit-risk analysis may be useful to answer this question. From the economic perspective, a concept referred to as "target efficiency," which combines PLP models with economic cost-benefit analysis, may be used to compute the specific prediction threshold at which targeting the intervention only to individuals who meet the threshold will yield the maximum expected economic net benefit (ENB).[24,25] Such an approach would likely be most useful to a payer that wishes to

target scarce resources in a way that maximizes economic efficiency.

Finally, PLP models may serve to generate hypotheses regarding ways to adapt current clinical practice to optimize outcomes. For example, given that insulin use was strongly associated with not achieving antihyperglycemic medication cessation, this suggests that earlier intervention with metabolic surgery in the disease process for diabetes (before insulin use) may result in significantly higher diabetes remission rates.

Limitations

This study was subject to additional limitations that deserve consideration. First, although there are methods that can predict multiple outcomes simultaneously, PLP models tend to focus on a single aspect (risk/probability of a binary outcome). The same is true for the present PLP model, which focuses on antihyperglycemic medication cessation but does not consider surgical risks nor does it consider prognosis in the absence of surgery. As described above, the expected net benefit of an intervention stems not only from the potential benefits of the intervention but also from the costs to implement it and its potential iatrogenic impact. Thus, this model alone could not function as an adequate decision support tool to decide whether to undergo metabolic surgery. Ideally, information from multiple PLP models—on both benefits and risks—could be used to inform patient care in the future.

Second, we based our example on 2 very large real-world administrative claims databases, chosen primarily on the basis that they would provide large samples in which the PLP model could trained, internally validated, and externally validated. Administrative claims databases lack complete information on a variety of predictors that could potentially be very important to antihyperglycemic medication cessation or diabetes remission, including baseline BMI and A1c. A critical aim for future research would be to implement this model using EHR data, in which a more objective measure of diabetes remission may also be possible. Future studies comparing the relative predictive power of administrative claims data (which may contain a larger breadth of information across the continuum of care) versus that of EHRs (which may sometimes provide narrower insights due to their reflection of a single practice but may also provide richer clinical detail such as baseline A1c) would be very informative for future decisions regarding implementation. Furthermore, it may be possible in the future to train a model using an integrated dataset that includes both administrative claims data and EHR data linked together for the same patient, or which links in other information such as patient reported lifestyle data. Indeed, continual incremental improvement of PLP models will be possible in the future and will be key to advancing a learning healthcare system.

Third, because the outcome of interest was antihyperglycemic medication cessation between 1 and 2 years after surgery, only patients who underwent metabolic surgery and for whom at least 2 years of follow-up data were available could be used in the present model. The 13 050 (CCAE) and 3477 (Optum) patients meeting the study criteria came from samples of 23 480 and 6028 metabolic surgery patients with any duration of follow-up data available; thus, the attrition rate over the 2-year period was approximately 42 % to 44%. It is possible that the patients used for training the model may differ from the overall group of patients considering metabolic surgery, which limits generalizability.

Finally, the use of machine learning techniques to develop predictive models typically involves a tradeoff: greater predictive power is gained at the cost of interpretability and parsimony. Our PLP model selected from among over 22 000 candidate predictors, many of which represent overlapping concepts and many of which may be highly multicollinear with other predictors that may or may not have made it into the ultimate PLP model specification. Thus, the individual interpretation of a single underlying predictor may sometimes appear unintuitive and should not be taken as being representative of an *effect estimate* for that predictor. Although many of the predictors in our final model do make clinical sense (eg, indicators for insulin use and increasing values of the Diabetes Comorbidity Severity Index were associated with lower probability of antihyperglycemic medication cessation) some carried associations that may not be easily explained (eg, inflammatory disorder of head was associated with higher probability of antihyperglycemic medication cessation). This aspect of the underlying PLP model may be unpalatable to those seeking a parsimonious score by which to stratify patients, such as the DiaRem or ABCD scores. Nevertheless, in the context of predictor selection from many thousands of candidate predictors, if those that are chosen by a data-driven means (and their corresponding coefficients) yield an accurate prediction in model training, internal validation, and external validation, their predictive utility may be more important than their alignment with clinical intuition.

Our research adds several unique contributions to the literature. First, to our knowledge, no prior study has reported the development and validation of an administrative claims data-based predictive model in the context of metabolic surgery. These data are a mainstay of outcomes research, and although they lack certain clinical elements such as BMI, duration of diabetes, and A1c, which would be ideal for the prediction problem at hand, we have demonstrated their potential utility in this application. Moreover, we accomplished this application using machine learning techniques applied to a very large feature set of over 22 000 potential predictors, as opposed to a traditional approach of using a priori model specification. This also represents one of the first published reports of a clinical prediction application of the OHDSI network's PLP framework. As noted above, the framework's underlying use of the OMOP CDM, the OHDSI Standard Vocabulary, and a network data infrastructure enabled us to efficiently conduct an external validation of the model within a separate database (Optum) from the one in which the model was trained (CCAE). Indeed, the present analysis' code is available to any member of the OHDSI network who wishes to replicate or expand upon our research (note: anyone may become a member of the OHDSI network [https://www.ohdsi.org/join-the-journey/]).

Conclusions

We developed a well-performing PLP model to predict antihyperglycemic medication cessation after metabolic surgery. PLP models based on readily available real-world healthcare data hold promise for healthcare decision support and may serve to generate hypotheses regarding ways to adapt current clinical practice to optimize outcomes. Barriers currently exist to the practical implementation of PLP models as clinical decision support tools. Nevertheless, as a community, patients, providers, payers, and researchers can work together to overcome these barriers and benefit from the full potential of PLP models based on large real-world data sources.

Acknowledgment

This study was funded by Johnson & Johnson.

Disclosure

Stephen Johnston, Iftekhar Kalsekar, Eric Ammann, and Jenna Reps are employees of Johnson & Johnson; Chia-Wen Hsiao is an employee of Ethicon, a part of the Johnson & Johnson family of companies; John M. Morton has no financial interests to declare.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2019.01.011.

REFERENCES

1. Rubino F, Nathan DM, Eckel RH, et al. Delegates of the 2nd Diabetes Surgery Summit. Metabolic surgery in the treatment algorithm for type 2 diabetes: a joint statement by International Diabetes Organizations. *Diabetes Care*. 2016;39(6):861–877.
2. Martin M, Beekley A, Kjorstad R, Sebesta J. Socioeconomic disparities in eligibility and access to bariatric surgery: a national population-based analysis. *Surg Obes Relat Dis*. 2010;6(1):8–15.
3. Afonso BB, Rosenthal R, Li KM, Zapatier J, Szomstein S. Perceived barriers to bariatric surgery among morbidly obese patients. *Surg Obes Relat Dis*. 2010;6(1):16–21.
4. Westerveld D, Yang D. Through thick and thin: identifying barriers to bariatric surgery, weight loss maintenance, and tailoring obesity treatment for the future. *Surg Res Pract*. 2016;2016:8616581.
5. Stanford FC, Kyle TK, Claridy MD, Nadglowski JF, Apovian CM. Interest, views and perceived barriers to bariatric surgery in patients with morbid obesity. *Obesity (Silver Spring)*. 2015;23(2):277–281.
6. Hales CM, Fryar CD, Carroll MD, Freedman DS, Ogden CL. Trends in obesity and severe obesity prevalence in US youth and adults by sex and age, 2007-2008 to 2015-2016. *JAMA*. 2018;319(16):1723–1725.
7. American Society for Metabolic and Bariatric Surgery. Estimate of Bariatric Surgery Numbers, 2011-2016. Available at: https://asmbs.org/resources/estimate-of-bariatric-surgery-numbers. Accessed June 1, 2018.
8. Zhang R, Borisenko O, Telegina I, et al. Systematic review of risk prediction models for diabetes after bariatric surgery. *Br J Surg*. 2016;103(11):1420–1427.
9. Tharakan G, Scott R, Szepietowski O, et al. Limitations of the DiaRem score in predicting remission of diabetes following Roux-En-Y gastric bypass (RYGB) in an ethnically diverse population from a single institution in the UK. *Obes Surg*. 2017;27(3):782–786.
10. Craig Wood G, Horwitz D, Still CD, et al. Performance of the DiaRem score for predicting diabetes remission in two health systems following bariatric surgery procedures in Hispanic and non-Hispanic white patients. *Obes Surg*. 2018;28(1):61–68.
11. Honarmand K, Chetty K, Vanniyasingam T, Anvari M, Chetty VT. Type 2 diabetes remission rates 1-year post-Roux-en-Y gastric bypass and validation of the DiaRem score: the Ontario Bariatric Network experience. *Clin Obes*. 2017;7(3):176–182.
12. Mehaffey JH, Mullen MG, Mehaffey RL, et al. Type 2 diabetes remission following gastric bypass: does diarem stand the test of time? *Surg Endosc*. 2017;31(2):538–542.
13. Lee WJ, Chong K, Chen SC, et al. Preoperative prediction of type 2 diabetes remission after gastric bypass surgery: a comparison of DiaRem scores and ABCD scores. *Obes Surg*. 2016;26(10):2418–2424.
14. Sampaio-Neto J, Nassif LS, Branco-Filho AJ, et al. External validation of the DiaRem score as remission predictor of diabetes mellitus type 2 in obese patients undergoing Roux-en-Y gastric bypass. *Arq Bras Cir Dig*. 2015;28(Suppl 1):19–22.
15. Aron-Wisnewsky J, Sokolovska N, Liu Y, et al. The advanced-DiaRem score improves prediction of diabetes remission 1 year post-Roux-en-Y gastric bypass. *Diabetologia*. 2017;60(10):1892–1902.
16. The Observational Health Data Sciences and Informatics. Patient-Level Prediction. Available at: https://www.ohdsi.org/methods-library/. Accessed July 1, 2018.
17. Hripcsak G, Duke JD, Shah NH, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. *Stud Health Technol Inform*. 2015;216:574–578.
18. Reps JM, Schuemie MJ, Suchard MA, Ryan PB, Rijnbeek PR. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. *J Am Med Inform Assoc*. 2018;25(8):969–975.
19. Overhage JM, Ryan PB, Reich CG, et al. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc*. 2012;19(1):54–60.
20. The Observational Health Data Sciences and Informatics. ATHENA standardized vocabularies. Available at: https://www.ohdsi.org/analytic-tools/athena-standardized-vocabularies/. Accessed December 4, 2018.
21. Young BA, Lin E, Von Korff M, et al. Diabetes complications severity index and risk of mortality, hospitalization, and healthcare utilization. *Am J Managed Care*. 2008;14(1):15–23.
22. PCORnet Common Data Model (CDM). Available at: https://pcornet.org/pcornet-common-data-model/. Accessed July 1, 2018.
23. Sentinel Common Data Model. Available at: https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model/sentinel-common-data-model. Accessed July 1, 2018.
24. Salkever DS, Johnston S, Karakus MC, et al. Enhancing the net benefits of disseminating efficacious prevention programs: a note on target efficiency with illustrative examples. *Adm Policy Ment Health*. 2008;35(4):261–269.
25. Johnston SS, Salkever DS, Ialongo NS, Slade EP, Stuart EA. Estimating the economic value of information for screening in disseminating and targeting effective school-based preventive interventions: an illustrative example. *Adm Policy Ment Health*. 2017;44(6):932–942.