

RESEARCH ARTICLE

Identifying patients at increased risk of hypoglycaemia in primary care: Development of a machine learning-based screening tool

Stijn Crutzen¹  | Sunil Belur Nagaraj¹  | Katja Taxis²  | Petra Denig¹ 

¹Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

²Unit of Pharmacotherapy, Epidemiology and Economics, Groningen Research Institute of Pharmacy, University of Groningen, Groningen, The Netherlands

Correspondence

Petra Denig, Clinical Pharmacy and Pharmacology, EB70, University Medical Center Groningen, PO Box 30.001, 9700 RB Groningen (NL)

Funding information

The Royal Dutch Pharmacists Association (KNMP)

Abstract

Introduction: In primary care, identifying patients with type 2 diabetes (T2D) who are at increased risk of hypoglycaemia is important for the prevention of hypoglycaemic events. We aimed to develop a screening tool based on machine learning to identify such patients using routinely available demographic and medication data.

Methods: We used a cohort study design and the Groningen Initiative to Analyse Type 2 diabetes Treatment (GIANTT) medical record database to develop models for hypoglycaemia risk. The first hypoglycaemic event in the observation period (2007–2013) was the outcome. Demographic and medication data were used as predictor variables to train machine learning models. The performance of the models was compared with a model using additional clinical data using fivefold cross validation with the area under the receiver operator characteristic curve (AUC) as a metric.

Results: We included 13,876 T2D patients. The best performing model including only demographic and medication data was logistic regression with least absolute shrinkage and selection operator, with an AUC of 0.71. Ten variables were included (odds ratio): male gender (0.997), age (0.990), total drug count (1.012), glucose-lowering drug count (1.039), sulfonylurea use (1.62), insulin use (1.769), pre-mixed insulin use (1.109), insulin count (1.827), insulin duration (1.193), and antidepressant use (1.05). The proposed model obtained a similar performance to the model using additional clinical data.

Conclusion: Using demographic and medication data, a model for identifying patients at increased risk of hypoglycaemia was developed using machine learning. This model can be used as a tool in primary care to screen for patients with T2D who may need additional attention to prevent or reduce hypoglycaemic events.

KEYWORDS

artificial intelligence, hypoglycaemia, type 2 diabetes

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. Diabetes/Metabolism Research and Reviews published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Strict glucose control in patients with type 2 diabetes (T2D) is essential for preventing long-term micro- and macro-vascular complications, such as retinopathy, neuropathy, and cerebrovascular, kidney and heart disease.¹⁻⁴ However, strict control of T2D with insulin and insulin secretagogues, especially in frail and older patients, increases the risk of hypoglycaemia.⁵⁻⁷ Commonly, three levels of hypoglycaemia are distinguished. Glucose levels between 3.9 mmol/L and 3.0 mmol/L are defined as mild hypoglycaemia and are considered an alert value.⁸ Glucose levels below 3.0 mmol/L are defined as moderate hypoglycaemia and are considered clinically important events, whereas severe hypoglycaemia is defined by a mental or physical impairment that requires external assistance. Hypoglycaemia is not only associated with reduced quality of life and higher healthcare costs but also hinders strict glucose control.⁹⁻¹² Severe cases of hypoglycaemia can lead to unconsciousness, hospitalization, and even brain death. Hypoglycaemia during hospital admissions has been associated with prolonged hospital stay and increased mortality.¹³ In addition to the mortality directly caused by hypoglycaemic events, several studies have shown an association of hypoglycaemia with all-cause and cardiovascular mortality.¹⁴⁻¹⁶

In T2D patients, the risk of hypoglycaemia strongly varies among patients. Not all T2D patients require insulin or insulin secretagogues to control their glucose level.¹⁷ Moreover, differences in co-medication can influence the risk of hypoglycaemia. For instance, non-selective beta-blockers reduce the effects of epinephrine, which in turn, increases hypoglycaemia unawareness,¹⁸ and selective serotonin reuptake inhibitors can increase sensitivity to insulin.¹⁹ In addition to differences in prescribed medications, differences in comorbidities can also influence the risk of hypoglycaemia. Dementia, for instance, reduces patients' ability to manage their glucose-lowering medications and increases the risk of medication errors,^{20,21} and depression can lead to poor self-care and self-monitoring.²² Furthermore, decreased renal and liver functions can diminish the clearance of both endogenous and exogenous insulin.^{18,23}

Several models have been developed to identify patients at high risk of hypoglycaemia.²⁴⁻³⁰ Most of these models have been developed using a mix of type 1 diabetes patients and T2D patients, focus on selecting high risk patients in inpatient settings, and often include previous hypoglycaemic events to predict future events. Most hypoglycaemic events occur outside the hospital setting and therefore need to be managed by primary care.^{31,32} In primary care, stratifying patients based on hypoglycaemia risk can provide an approach in which treatment, glycaemic goals and education are tailored to the individual, which ultimately helps to reduce hypoglycaemia.³³ Previous research has shown that pharmacy-led interventions can be effective in reducing the number of hypoglycaemic events in T2D patients.^{34,35} Using routinely available pharmacy data to screen for patients at increased risk of hypoglycaemia can improve the effectiveness and efficiency of

these interventions. Generally, pharmacy data include information about patients' medication history and all drugs dispensed during a specific period. Conventional approaches to using such information in risk models include creating simple measures of medication use at the substance or therapeutic class level.³⁶ Recent developments in machine learning provide statistical tools to include multiple characterizations of medication use in one model, for example, 'current insulin use', 'insulin count', 'insulin type', and 'insulin use duration'. In this way, more information can be utilized for risk modelling.

In this study, we aimed to develop a screening tool to identify T2D patients at increased risk of hypoglycaemic events based on machine learning using demographic and medication data available in community pharmacies. Additionally, we investigated whether the performance of the model could be improved by adding electronic clinical data that are available in general practice. We also compared the results of these models with results from traditional regression models.

2 | MATERIAL AND METHODS

2.1 | Study design

We conducted a cohort study using demographic and medication data to predict the occurrence of hypoglycaemia in T2D patients. The performance of this model was compared to that of a model that additionally included clinical data. Data from the Groningen Initiative to ANalyse Type 2 Diabetes Treatment (GIANTT) were used (www.giantt.nl). This database includes data from a longitudinal dynamic cohort of more than 60,000 patients with a general practice (GP)-confirmed diagnosis of T2D. The data were extracted from the electronic medical records of 180 GPs in the northern parts of the Netherlands.³⁷ At the end of our study period, approximately 80% of all the GPs in this region contributed to the GIANTT database. The database contains patients' demographic data, data on prescribed medications (similar to medication data in a pharmacy information system), medical history based on the International Classification of Primary Care (ICPC)³⁸ and routinely available clinical measurements, such as blood pressure and low haemoglobin A1c (HbA1c) levels.

2.2 | Outcome and predictor variables

The outcome of this study was the first hypoglycaemic event during the observation period between January 2007 and January 2014. Hypoglycaemic events were defined by the ICPC code for hypoglycaemia (T87) or by a recorded glucose measurement below 3.9 mmol/L. We therefore included all the levels of hypoglycaemia that were recorded in the medical records by GPs since mild events can also be relevant for our screening tool. Recorded hypoglycaemic events in free text parts of the medical record were used to enrich the T87 codes in the GIANTT database. The date of the first hypoglycaemic event was used as the index date. Patients without any hypoglycaemic events were assigned

a random index date in the observation period. When assigning these random dates, the distribution of the index dates of the event group was taken into account to correct for temporal influences. Based on the literature, potential predictor variables related to demographics, medication, and clinical information were selected. An overview of the selected variables and their definitions can be found in Appendix SI. Any prescription in the 4 months prior to the index date was used to determine current medication use, the latest measured value within 1 year prior to the index date was used for most clinical measurements and active ICPC codes at the time of the index date were used for comorbidities. For the medication data, potential predictor variables included not only the current use of medication but also other aspects of medication use, including duration of use, counts of medication classes, and potential drug-drug interactions.

2.3 | Study population

This study included T2D patients who were 18 years or older and were treated for diabetes in a primary care setting. We excluded patients registered with a GP with poor documentation of hypoglycaemia, as demonstrated by a lack of recordings for hypoglycaemia (T87) for any of their T2D patients. The following patient exclusion criteria were applied: (A) included in the database for less than 1 year prior to their index date; (B) no prescription of a glucose-lowering medication 1 year prior to the index date; (C) no T2D diagnosis before or within 6 months after the index date; and (D) missing all of the following measurements 1 year prior to the index date: estimated glomerular filtration rate (eGFR), creatinine, albumin to creatinine ratio, HbA1c, weight, body mass index, high density lipoprotein, low density lipoprotein, total cholesterol, systolic blood pressure, and diastolic blood pressure.

Due to the imbalanced dataset ($N = 2,523$ and $11,344$ with and without a hypoglycaemic event, respectively), which could severely bias the performance of machine learning algorithms, we used a sample-size equalization method to balance the data. In this method, we randomly divided the patients without a hypoglycaemic event into four equal subdatasets, corresponding to the number of patients in the hypoglycaemic group, as shown in Figure 1 ($11,344/2,523 = 4$ subdatasets). Model training was performed using these balanced datasets.

2.4 | Analysis

We first evaluated the prediction performance of all the individual variables and evaluated the performance of traditional logistic regression using either all the demographic and medication data or all the demographic, medication and clinical data. Next, to develop the screening tool, we trained several machine learning algorithms on either demographic and medication data or demographic, medication, and clinical data. We evaluated the performance of the following algorithms: logistic regression with backwards selection, ridge logistic

regression, elastic net logistic regression ($\alpha = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, \text{ or } 0.9$), least absolute shrinkage and selection operator (LASSO) logistic regression and random forest (RF). These algorithms were used because they are relatively simple to interpret, easy to implement in practice, and, in the cases of elastic net, LASSO and RF are suitable for dealing with multidimensional data.^{39,40} Missing clinical measurement data were imputed using the K-nearest neighbour ($k = 5$) multiple imputation method (10-fold). The list of the predictor variables and the percentage missing for each variable are given in Appendix SI. We performed fivefold cross-validation (80% training data and 20% testing data) to evaluate the performance of the models using area under the curve (AUC) as a metric. The best performing model without clinical data was compared with the best performing model including clinical data. Stata 15.0 was used for data cleaning, logistic regression, and RF. R version 3.5.1 was used with the glmnet package for LASSO, elastic net, and ridge logistic regression.⁴¹ For RF, out of the bag error was used to find the optimal number of trees and the optimal number of variables considered at each node. No maximum depth was set, and the minimum number of observations per leaf was set to one. We performed fivefold cross-validation to identify optimal values for the lambda plus one standard error for LASSO, elastic net, and ridge logistic regression. Second, we performed fivefold cross-validation to identify the optimal α for the elastic net between 0.1 and 0.9 with steps of 0.1. Third, we performed fivefold cross-validation to select the balanced subdataset and the imputed dataset that resulted in the best performance. The predictor variables were standardized for the elastic net, ridge, and LASSO logistic regression analyses. A heatmap was created to show the importance of the variables used in the best performing model.

2.5 | Ethics statement

Based on the research code of conduct in the Netherlands, research using anonymous medical record data requires no ethics committee approval.

3 | RESULTS

We extracted 40,124 patients from the GIANTT database. After exclusion, 13,876 patients remained, 19.2% of whom had at least one hypoglycaemic event during the observation period (Appendix SII). The subdatasets contained either 5201 or 5202 patients. Table 1 shows that patients with and without a hypoglycaemic event were similar in age and clinical measurements, but patients with an event used insulin more often and had a longer diabetes duration.

3.1 | Models with demographic and medication data

Using the demographic and medication data, the best predicting individual variable was 'insulin use duration', with an AUC of 0.64

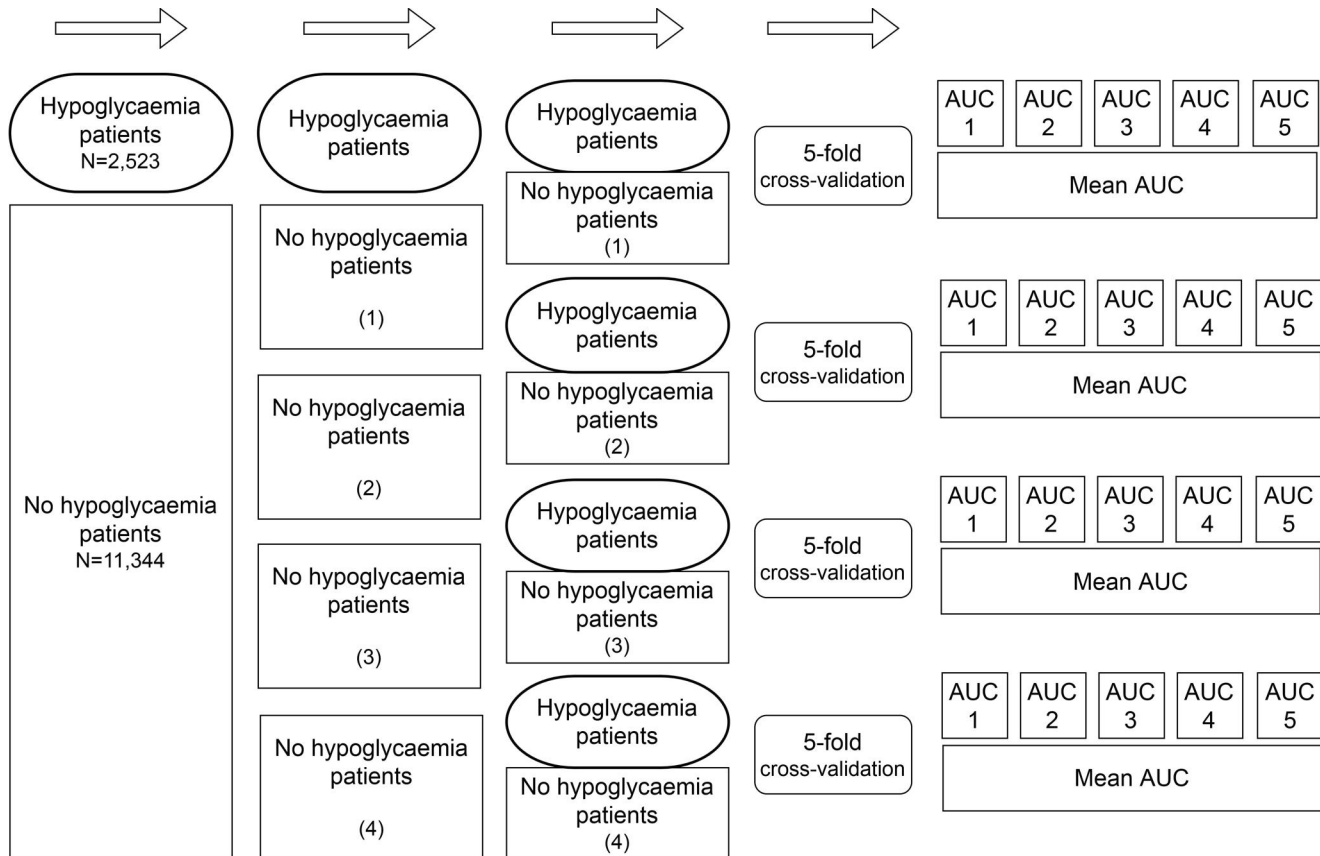


FIGURE 1 Schematic overview of the sample-size equalization method and fivefold cross-validation to evaluate and compare the performance of the different machine learning models. To balance the data, the non-hypoglycaemia patients were divided in four equal groups, which were each matched with the hypoglycaemia patients. This was followed by fivefold cross-validation in each of the four subdatasets to determine which machine learning method in which subdataset resulted in the best performing model using area under the curve (AUC) as metric

(Figure 2). Out of the 25 individual variables, 3 had an AUC above 0.6, 6 had an AUC between 0.60 and 0.55, and the remaining 16 variables had an AUC below 0.55. Using traditional logistic regression without variable selection resulted in an AUC of 0.69. The best performing machine learning algorithm was the LASSO logistic regression algorithm, with a mean AUC of 0.71 (± 0.019) using 10 variables (Table 2). The LASSO logistic regression algorithm selected age, sex, six diabetes medication-related variables, and two co-medication-related variables. The most important variable was 'insulin use', followed by 'sulfonylurea use' and 'insulin use duration' (Figure 3). Although sex and age were selected as predictors in most folds, their importance was relatively low. Logistic regression with backward selection resulted in a mean AUC of 0.69, ridge logistic regression resulted in an AUC of 0.69, the best performing elastic net logistic regression model ($\alpha = 0.6$) resulted in an AUC of 0.70, and RF resulted in an AUC of 0.68.

3.2 | Models with additional clinical data

The best predictive individual clinical variable was diabetes duration, with an AUC of 0.57 (Appendix SIII). When including additional clinical

data, traditional logistic regression without variable selection resulted in an AUC of 0.71. The best performing machine learning algorithm was the LASSO logistic regression algorithm, with an average AUC of 0.71 (± 0.024). The performance of the resulting model was similar to that of the model without clinical data. The model included nine of the same demographic and medication variables; three additional medication variables: antipsychotic, antibiotic, and oral corticosteroid use; and 10 additional clinical variables: diabetes duration, weight, eGFR, HbA1c, total cholesterol, depression, high blood pressure, non-chronic infection, hypercholesterolaemia, and albuminuria (Table 2). Logistic regression with backward selection resulted in an average AUC of 0.71, and ridge logistic regression resulted in an AUC of 0.70. For the best performing elastic net logistic regression model ($\alpha = 0.5$), the AUC was 0.71, and RF resulted in an AUC of 0.66.

4 | DISCUSSION

4.1 | Main findings

A model including 10 demographic and medication variables based on machine learning showed an acceptable performance to screen for an

TABLE 1 Characteristics of patients with and without a hypoglycaemic event

Variables	Hypoglycaemia patients	No hypoglycaemia patients
Total number of patients, n (%)	2,523 (19.1)	10,713 (80.9)
Age, years	66.4 (12.5)	67.9 (12.1)
Female, %	45.0	50.2
Diabetes duration, years	8.3 (6.5)	6.8 (5.6)
Insulin use, ^a %	34.0	12.1
Sulfonylurea use, ^a %	44.6	37.1
Metformin use, %	75.5	79.0
Number of medicines	6.5 (3.5)	5.8 (3.3)
HbA1c, %	7.1 (1.0)	7.0 (1.0)
BMI, kg/m ²	30.2 (5.6)	30.0 (5.5)
SBP, mmHg	138.6 (17.0)	139.6 (17.7)
DBP, mmHg	77.3 (9.6)	78.1 (10.1)
LDL, mmol/L	2.5 (0.88)	2.5 (0.91)
HDL, mmol/L	1.2 (0.34)	1.2 (0.35)
Total cholesterol, mmol/L	4.4 (1.0)	4.4 (1.2)
eGFR, ml/min/173 m ²	75.1 (22.8)	76.7 (23.2)

Note: Values are reported as the mean with the standard deviation (sd) unless reported otherwise. Abbreviations: HbA1c, haemoglobin A1c; BMI, body mass index; DBP, diastolic blood pressure; eGFR = estimated glomerular filtration rate; HDL, high density lipoprotein; LDL, low density lipoprotein; SBP, systolic blood pressure.

^aOf the hypoglycaemia patients, 71.4% used insulin and/or sulfonylurea, and 46.9% of the patients without hypoglycaemia used insulin and/or sulfonylurea.

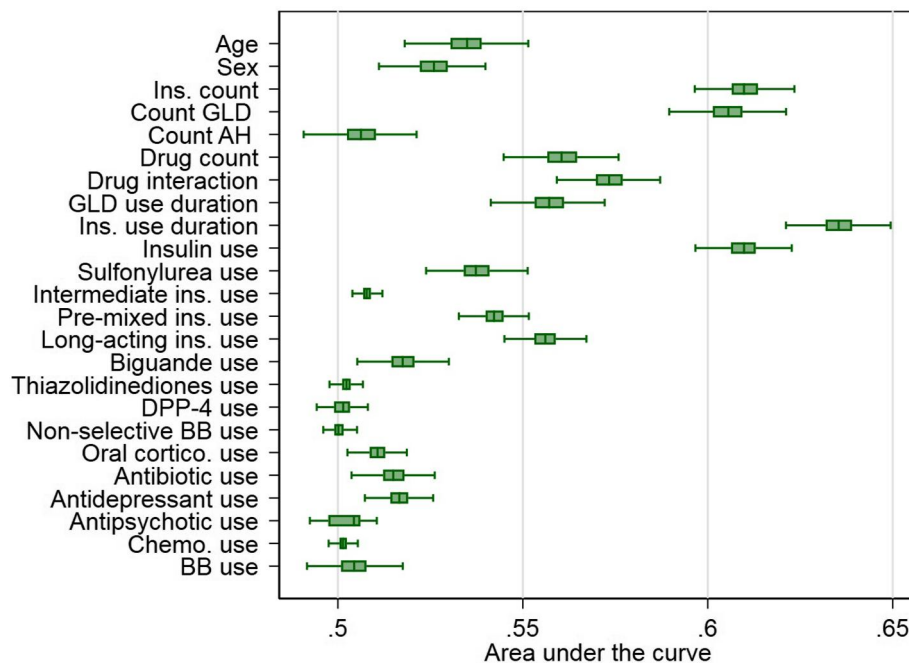


FIGURE 2 Boxplot of the area under the curve of the individual predictors, based on 1000 bootstraps. AH, antihypertensive; BB, beta-blocker; cortico., corticosteroid; chemo., antineoplastic or immunomodulating agent; DPP-4, dipeptidyl peptidase 4 inhibitor; GLD, glucose-lowering drugs; drug interaction, interaction between insulin and/or sulfonylurea with co-medication; ins., insulin

TABLE 2 Odds ratios of the predictor variables for the best performing models (LASSO) using only demographic and medication data or additional clinical data

Predictors	Odds ratio
Demographic and medication data	
Sulfonylurea use	1.620
Insulin use	1.769
Pre-mixed insulin use	1.109
Insulin use duration capped at 5 years (years)	1.193
Count of different types of insulin	1.827
Count of glucose-lowering drugs	1.039
Count of all drugs	1.012
Antidepressant use	1.050
Age (years)	0.990
Sex (0 = F/1 = M)	0.997
Intercept	0.965
Additional clinical data	
Sulfonylurea use	2.001
Insulin use	1.660
Pre-mixed insulin use	1.481
Insulin use duration capped at 5 years (years)	1.244
Count of different types of insulin	1.120
Count of glucose-lowering drugs	1.016
Count of all drugs	1.008
Antidepressant use	1.184
Age (years)	0.986
Antibiotic use	1.047
Oral corticosteroid use	1.016
Antipsychotic use	0.922
Diabetes duration capped at 5 years (years)	0.969
Weight (kilogramme)	0.994
eGFR (ml/min/1,73 m ²)	0.999
HbA1c (%)	0.948
Total cholesterol (mmol/L)	0.973
Depression	0.956
High blood pressure	1.107
Non-chronic infection	1.536
Hypercholesterolaemia	0.909
Albuminuria	1.012
Intercept	3.196

Abbreviations: eGFR, estimated glomerular filtration rate; F, female; HbA1c, haemoglobin A1c; LASSO, least absolute shrinkage and selection operator; M, male.

increased risk of hypoglycaemia in T2D patients in primary care. The inclusion of additional clinical data did not improve the performance of the model.

4.2 | Comparison with previous research

Several models have been developed for predicting hypoglycaemia.^{24–30} However, these models mainly focus on the inpatient setting or do not make a distinction between type 1 and T2D. Our model is intended for primary care using demographic and medication data that are widely available to screen for T2D patients with an increased risk of hypoglycaemia. By including only T2D patients in the development of the model, variables that are specific for T2D patients, such as sulfonylurea use, can contribute to the model.

The performance of our model, which had an AUC of 0.71, is considered acceptable⁴² and comparable with several previously developed models,^{24–26} although some models have shown higher AUCs than ours.^{28–30,43} The higher performance of these models may be due to the availability of richer data in clinical trials and inpatient settings in comparison to data that are routinely available in outpatient settings. For example, daily glucose measurements may be available for diabetes patients who are admitted to a hospital but not for T2D patients in primary care. More importantly, all of the better performing models included prior hypoglycaemic events as a variable, which was one of the, if not the most, important predictor in these models. Since we aimed to screen for patients at increased risk of a first hypoglycaemic event, we did not use prior events as a predictor. Predicting the first hypoglycaemic event is more difficult, but it is essential to identify at-risk patients who are not already known to healthcare professionals.

Many of the variables in our models are known risk factors for hypoglycaemia, such as using different types of insulin or using pre-mixed insulin. Previous research has shown that a longer diabetes duration is predictive of hypoglycaemia.³⁰ In our models, as in previous research, a longer duration of insulin use was associated with a higher risk of hypoglycaemia.⁴⁴ When clinical data were added to that model, diabetes duration—in addition to a longer insulin use duration—was associated with a lower risk. Surprisingly, lower age contributed to an increased risk of hypoglycaemia in our models, whereas in other models, higher age was predictive of hypoglycaemic events.^{26,28–30} This difference might be due to differences in the severity of the hypoglycaemia events used as outcomes. Previous studies mainly used severe hypoglycaemic events as the outcome, while we included any events regardless of severity. It has been found that younger patients may report more hypoglycaemic events, in general, but not more severe hypoglycaemic events.⁴⁵ This finding might be due to the increased hypo-unawareness in older patients.⁴⁶ The inclusion of antidepressant use as a predictor could be explained

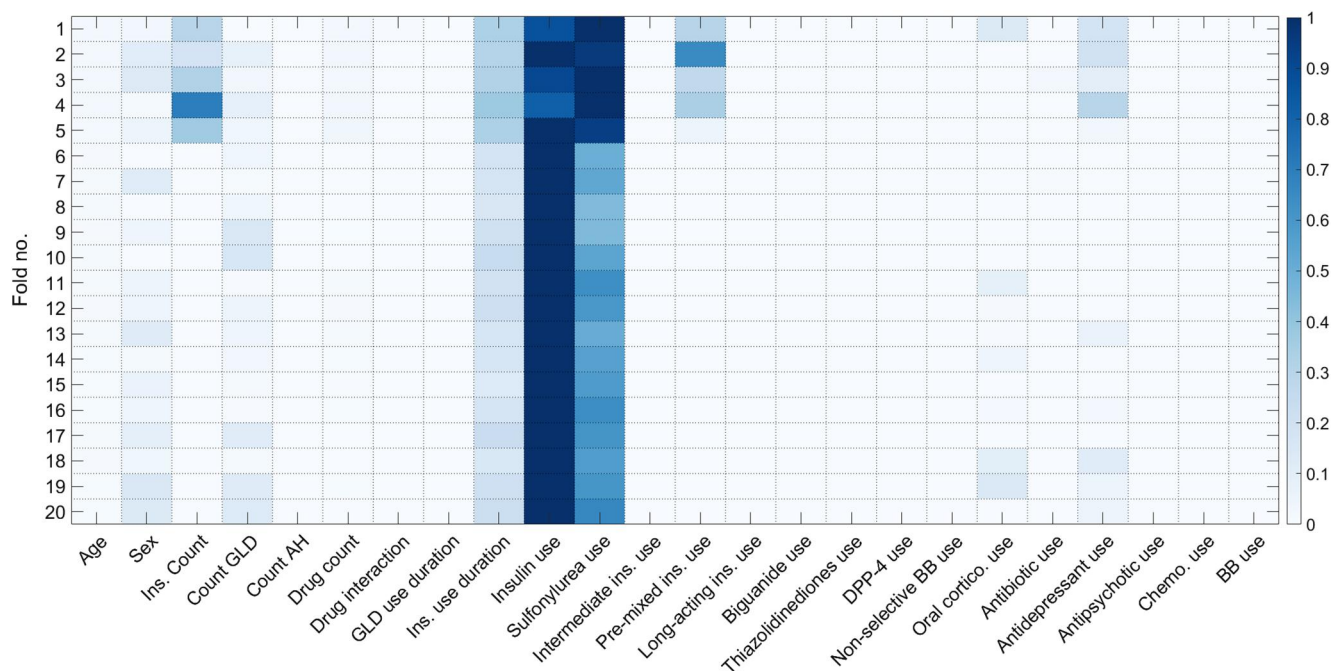


FIGURE 3 Heatmap showing the set of variables selected in the different folds of the fivefold cross-validation in the four subdatasets (5x4 folds) based on least absolute shrinkage and selection operator (LASSO) logistic regression algorithm. A darker blue colour represents a higher weight assigned by LASSO. This is indicative of a higher importance of a variable for predicting hypoglycaemia. AH, antihypertensive; BB, beta-blocker; cortico., corticosteroid; chemo., antineoplastic or immunomodulating agent; DPP-4, dipeptidyl peptidase 4 inhibitor; drug interaction, Interaction between insulin and/or sulfonylurea with comedication; GLD, glucose-lowering drugs; ins., insulin

by the increase in insulin sensitivity caused by selective serotonin reuptake inhibitors.¹⁹ Another explanation might be the association of severe depression with severe hypoglycaemia in T2D patients.²²

Of note is our finding that clinical data, such as the HbA1c level, did not improve the performance of the model. Additionally, the HbA1c level was similar for those who did and those who did not have a hypoglycaemic event. This finding suggests that the level of glucose control is not a good measure to inform healthcare professionals about the risk of hypoglycaemia, which is in line with previous research that showed that HbA1c is not related to hypoglycaemia risk in older T2D patients who use insulin.⁴⁷ More generally, it is important to realize that by developing a model based on a large amount of data, some patient characteristics that are considered clinically relevant may not contribute to the performance of the model as a whole.

4.3 | Strengths and limitations

By using machine learning to develop our models, we were able to include several related variables, which allowed us to make better use of the information available when only using medication data. Particularly, considering more information about insulin than just current use improved the performance of the model. By using LASSO to select variables, a simpler model could be obtained.

LASSO logistic regression is able to select predictors without having to rely on *p*-values, which are highly dependent on power. Another strength is the use of cross-validation, which provides a less biased estimation of the performance of models. The use of a large database, which consists of routinely available primary care data, has two advantages. First, using routinely available data will mimic daily practice more closely, resulting in a more accurate estimation of the performance of the prediction models. Second, by using these data, only predictors that are commonly available and well documented in primary care are included in the prediction model.

Several limitations of the study should be noted. First, it is likely that a portion of the control patients were misclassified. By using routinely available data, we missed hypoglycaemic events that were not documented by GPs. Although we excluded GPs who never documented hypoglycaemic events, the included GPs may not have consistently documented such events. This misclassification most likely resulted in a lower performance of the models. Bias could have been introduced when the missed events were not random. For instance, GPs might be more primed to recognize hypoglycaemia in patients who use insulin, causing an underestimation of hypoglycaemia risk in patients who do not use insulin. Second, some patients may have been falsely classified as not using insulin or sulfonylurea because the 4-month look-back period was too short for prescriptions that can last for a longer period. When this period was increased to twelve months, insulin

use increased by approximately five percent points in the hypoglycaemia group. A third limitation is that we only performed an internal validation. Finally, four predictors that were selected by LASSO in the best performing model with clinical data were imputed, namely, weight, eGFR, HbA1c, and total cholesterol. This limits the applicability of the model using clinical data. In addition, potential predictors not documented, for example, related to the level of education or self-management, were not included, and the use of newer diabetic medication was not included in our primary care dataset in the study period.

4.4 | Implication for practice

Both the models including demographic and medication data and the model with additional clinical data could be used to identify patients with type 2 diabetes in primary care who are at increased risk of hypoglycaemia. For implementation in clinical practice, we recommend a screening tool based on the model without clinical data because it requires less information and shows a similar performance to the model with clinical data. Our model can be easily implemented in the electronic information systems of community pharmacies as well as GPs, converting the ORs from the LASSO model into a risk score. The model is not intended to replace or mimic the clinical judgement of a healthcare professional. Instead, this model is intended to be used as a screening tool to assist healthcare professionals identify patients who may need more support and monitoring. By identifying patients at increased risk for hypoglycaemia, additional support and monitoring to prevent or reduce the severity of hypoglycaemic events can be targeted to patients who are most in need of such support. By opting for a relatively low cut-off point, a high sensitivity can be achieved, ensuring that few patients at increased risk of hypoglycaemic events are missed. It should be noted that such a tool is not appropriate for detecting patients at risk of minor hypoglycaemic events that are not reported to GPs.

5 | CONCLUSION

We developed a model for identifying patients at increased risk of hypoglycaemia in primary care using demographic and medication data based on LASSO, which can be used as a screening tool. This model had an acceptable performance, outperformed individual predictor variables and performed similarly to a model that used additional clinical data.

ACKNOWLEDGEMENTS

An unconditional grant was provided by the Royal Dutch Pharmacists Association (KNMP), they had no role in the execution of this study or in the drafting of the article.

CONFLICTS OF INTEREST

The authors have no conflicts of interest.

DATA AVAILABILITY STATEMENT

Requests to receive data from the GIANTT database can be directed at the GIANTT database steering commission <https://www.giantt.nl/>.

AUTHOR CONTRIBUTIONS

Stijn Crutzen, Katja Taxis, and Petra Denig: research idea and study design. Stijn Crutzen, Petra Denig: Data acquisition. Stijn Crutzen, Sunil Belur Nagaraj, Katja Taxis, and Petra Denig: analysis and interpretation. Sunil Belur Nagaraj, Petra Denig, and Katja Taxis: supervision or mentorship. All authors contributed substantially to the intellectual content during manuscript drafting and revision. All authors approved the manuscript and this submission.

ORCID

Stijn Crutzen  <https://orcid.org/0000-0002-0537-0858>

Sunil Belur Nagaraj  <https://orcid.org/0000-0002-6409-4101>

Katja Taxis  <https://orcid.org/0000-0001-8539-2004>

Petra Denig  <https://orcid.org/0000-0002-7929-4739>

REFERENCES

- Selvin E, Marinopoulos S, Berkenblit G, et al. Meta-analysis: glycosylated hemoglobin and cardiovascular disease in diabetes mellitus. *Ann Intern Med.* 2004;141(6):423-432.
- Stratton IM, Adler AI, Neil HAW, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. *Br Med J.* 2000;321(7258):405-412.
- Kirkman MS, Mahmud H, Korytkowski MT. Intensive blood glucose control and vascular outcomes in patients with type 2 diabetes mellitus. *Endocrinol Metab Clin North Am.* 2018;47(1):81-96. <https://doi.org/10.1016/j.ecl.2017.10.002>.
- Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *N Engl J Med.* 2008;358(24):2545-2559. <https://doi.org/10.1056/NEJMoa0802743>.
- Thorpe CT, Gellad WF, Good CB, et al. Tight glycemic control and use of hypoglycemic medications in older veterans with type 2 diabetes and comorbid dementia. *Diabetes Care.* 2015;38(4):588-595. <https://doi.org/10.2337/dc14-0599>.
- Finfer S, Chittock D, Li Y, et al. Intensive versus conventional glucose control in critically ill patients with traumatic brain injury: long-term follow-up of a subgroup of patients from the NICE-SUGAR study. *Intensive Care Med.* 2015;41(6):1037-1047. <https://doi.org/10.1007/s00134-015-3757-6>.
- UKPDS. Intensive Blood-glucose control with sulfonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes. *Endocrinologist.* 1999;9(2):149. <https://doi.org/10.1097/00019616-199903000-00016>.
- International hypoglycaemia study group. Glucose concentrations of less than 3.0 mmol/L (54 mg/dL) should be reported in clinical trials: a joint position statement of the American diabetes association and the European association for the study of diabetes. *Diabetes Care.* 2017;40(1):155-157. <https://doi.org/10.2337/dc16-2215>.
- Williams SA, Pollack MF, DiBonaventura M. Effects of hypoglycemia on health-related quality of life, treatment satisfaction and health-care resource utilization in patients with type 2 diabetes mellitus. *Diabetes Res Clin Pract.* 2011;91(3):363-370. <https://doi.org/10.1016/j.diabres.2010.12.027>.
- Wild D, von Maltzahn R, Brohan E, Christensen T, Clauson P, Gonder-Frederick L. A critical review of the literature on fear of

- hypoglycemia in diabetes: implications for diabetes management and patient education. *Patient Educ Couns*. 2007;68(1):10-15. <https://doi.org/10.1016/j.pec.2007.05.003>.
11. Davis S, Alonso MD. Hypoglycemia as a barrier to glycemic control. *J Diabetes Complications*. 2004;18(1):60-68. [https://doi.org/10.1016/S1056-8727\(03\)00058-8](https://doi.org/10.1016/S1056-8727(03)00058-8).
 12. De Groot S, Enters-Weijnen CF, Geelhoed-Duijvestijn PH, Kanters TA. A cost of illness study of hypoglycaemic events in insulin-treated diabetes in The Netherlands. *BMJ Open*. 2018;8(3):6-10. <https://doi.org/10.1136/bmjopen-2017-019864>.
 13. Leibovitz E, Khanimov I, Wainstein J, Boaz M. Documented hypoglycemia is associated with poor short and long term prognosis among patients admitted to general internal medicine departments. *Diabetes Metab Syndr Clin Res Rev*. 2019;13(1):222-226. <https://doi.org/10.1016/j.dsx.2018.07.004>.
 14. Davis SN, Duckworth W, Emanuele N, et al. Effects of severe hypoglycemia on cardiovascular outcomes and death in the Veterans Affairs Diabetes Trial. *Diabetes Care*. 2019;42(1):157-163. <https://doi.org/10.2337/dc18-1144>.
 15. Zinman B, Marso SP, Christiansen E, Calanna S, Rasmussen S, Buse JB. Hypoglycemia, cardiovascular outcomes, and death: the LEADER experience. *Diabetes Care*. 2018;41(8):1783-1791. <https://doi.org/10.2337/dc17-2677>.
 16. Action to Control Cardiovascular Risk in Diabetes Study Group, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. *New Engl J Med*. 2008;358(24):2545-2559.
 17. Rauh Simone P, Rutters F, Thorsted BL, et al. Self-reported hypoglycaemia in patients with type 2 diabetes treated with insulin in the Hoorn Diabetes Care System Cohort, the Netherlands: a prospective cohort study. *BMJ Open*. 2016;6(9):e012793. <https://doi.org/10.1136/bmjopen-2016-012793>.
 18. Quilliam BJ, Simeone JC, Ozbay AB. Risk factors for hypoglycemia-related hospitalization in patients with type 2 diabetes: a nested case-control study. *Clin Ther*. 2011;33(11):1781-1791. <https://doi.org/10.1016/j.clinthera.2011.09.020>.
 19. Derijks HJ, Meyboom RHB, Heerdink ER, et al. The association between antidepressant use and disturbances in glucose homeostasis: evidence from spontaneous reports. *Eur J Clin Pharmacol*. 2008;64(5):531-538. <https://doi.org/10.1007/s00228-007-0441-y>.
 20. Feil DG, Rajan M, Soroka O, Tseng CL, Miller DR, Pogach LM. Risk of hypoglycemia in older veterans with dementia and cognitive impairment: implications for practice and policy. *J Am Geriatr Soc*. 2011;59(12):2263-2272. <https://doi.org/10.1111/j.1532-5415.2011.03726.x>.
 21. Mattishent K, Loke YK. Bi-directional interaction between hypoglycaemia and cognitive impairment in elderly patients treated with glucose-lowering agents: a systematic review and meta-analysis. *Diabetes, Obesity and Metabolism*. 2016;18(2):135-141. <https://doi.org/10.1111/dom.12587>.
 22. Katon WJ, Young BA, Russo J, et al. Association of depression with increased risk of severe hypoglycemic episodes in patients with diabetes. *Ann Fam Med*. 2013;11(3):245-250. <https://doi.org/10.1370/afm.1501>.
 23. Hsu PF, Sung SH, Cheng HM, et al. Association of clinical symptomatic hypoglycemia with cardiovascular events and total mortality in type 2 diabetes: a nationwide population-based study. *Diabetes Care*. 2013;36(4):894-900. <https://doi.org/10.2337/dc12-0916>.
 24. Ena J, Gaviria AZ, Romero-Sánchez M, et al. Derivation and validation model for hospital hypoglycemia. *Eur J Intern Med*. 2018;47:43-48. <https://doi.org/10.1016/j.ejim.2017.08.024>.
 25. Elliott MB, Schafers SJ, McGill JB, Tobin GS. Prediction and prevention of treatment-related inpatient hypoglycemia. *J Diabetes Sci Technol*. 2012;6(2):302-309. <https://doi.org/10.1177/193229681200600213>.
 26. Stuart K, Adderley NJ, Marshall T, et al. Predicting inpatient hypoglycaemia in hospitalized patients with diabetes: a retrospective analysis of 9584 admissions with diabetes. *Diabet Med*. 2017;34(10):1385-1391. <https://doi.org/10.1111/dme.13409>.
 27. Park SY, Jang EJ, Shin JY, Lee MY, Kim D, Lee EK. Prevalence and predictors of hypoglycemia in South Korea. *Am J Manag Care*. 2018;24(6):278-286.
 28. Karter AJ, Warton EM, Lipska KJ, et al. Development and validation of a tool to identify patients with type 2 diabetes at high risk of hypoglycemia-related emergency department or hospital use. 2019;94612(10):1461-1470. <https://doi.org/10.1001/jamainternmed.2017.3844>.
 29. Winterstein AG, Jeon N, Staley B, Xu D, Henriksen C, Lipori GP. Development and validation of an automated algorithm for identifying patients at high risk for drug-induced hypoglycemia. *Am J Heal Pharm*. 2018;75(21):1714-1728. <https://doi.org/10.2146/ajhp180071>.
 30. Chow LS, Zmora R, Ma S, Seaquist ER, Schreiner PJ. Development of a model to predict 5-year risk of severe hypoglycemia in patients with type 2 diabetes. *BMJ Open Diabetes Res Care*. 2018;6(1):e000527. <https://doi.org/10.1136/bmjdr-2018-000527>.
 31. Lipska KJ, Warton EM, Huang ES, et al. HbA1c and risk of severe hypoglycemia in type 2 diabetes: the diabetes and aging study. *Diabetes Care*. 2013;36(11):3535-3542. <https://doi.org/10.2337/dc13-0610>.
 32. Karter AJ, Moffet HH, Liu JY, Lipska KJ. Surveillance of hypoglycemia-limitations of emergency department and hospital utilization data. *JAMA Intern Med*. 2018;178(7):987-988. <https://doi.org/10.1001/jamainternmed.2018.1014>.
 33. Ibrahim M, Baker J, Cahn A, et al. Hypoglycaemia and its management in primary care setting. *Diabetes Metab Res Rev*. 2020(April):1-21. <https://doi.org/10.1002/dmrr.3332>.
 34. Venkatesan R, Manjula Devi A, Parasuraman S, Sriram S. Role of community pharmacists in improving knowledge and glycemic control of type 2 diabetes. *Perspect Clin Res*. 2012;3(1):26. <https://doi.org/10.4103/2229-3485.92304>.
 35. Hendrie D, Miller TR, Woodman RJ, Hoti K, Hughes J. Cost-effectiveness of reducing glycaemic episodes through community pharmacy management of patients with type 2 diabetes mellitus. *J Prim Prev*. 2014;35(6):439-449. <https://doi.org/10.1007/s10935-014-0368-x>.
 36. Vitry A, Wong SA, Roughead EE, Ramsay E, Barratt J. Validity of medication-based co-morbidity indices in the Australian elderly population. *Aust N Z J Public Health*. 2009;33(2):126-130. <https://doi.org/10.1111/j.1753-6405.2009.00357.x>.
 37. Voorham J, Denig P. Computerized extraction of information on the quality of diabetes care from free text in electronic patient records of general practitioners. *J Am Med Informatics Assoc*. 2007;14(3):349-354. <https://doi.org/10.1197/jamia.M2128>.
 38. Bentsen BG. International classification of primary care. *Scand J Prim Health Care*. 1986;4(1):43-50. <https://doi.org/10.3109/02813438609013970>.
 39. Zou H, Hastie T. Erratum: regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol*. 2005;67(5):301-320. <https://doi.org/10.1111/j.1467-9868.2005.00527.x>.
 40. Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics*. 2010;26(10):1340-1347. <https://doi.org/10.1093/bioinformatics/btq134>.
 41. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via Coordinate Descent. *J Stat Softw*. 2010;33(1):1-22. <http://www.jstatsoft.org/v33/i01/>.
 42. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed.. New York, NY: John Wiley and Sons; 2000.
 43. Schroeder EB, Xu S, Goodrich GK, Nichols GA, Connor PJO, Steiner JF. Complications predicting the 6-month risk of severe hypoglycemia among adults with diabetes: development and external validation of a prediction model. *Journal of Diabetes and Its*. 2017;31:1158-1163. <https://doi.org/10.1016/j.jdiacomp.2017.04.004>.

44. Heller SR, Choudhary P, Davies C, et al. Risk of hypoglycaemia in types 1 and 2 diabetes: effects of treatment modalities and their duration. *Diabetologia*. 2007;50(6):1140-1147. <https://doi.org/10.1007/s00125-007-0599-y>.
45. Khunti K, Alsifri S, Aronson R, et al. Global HAT study. *Diabetes, Obes Metab*. 2016;18(9):907-915.
46. Bremer JP, Jauch-Chara K, Hallschmid M, Schmid S, Schultes B. Hypoglycemia unawareness in older compared with middle-aged patients with type 2 diabetes. *Diabetes Care*. 2009;32(8):1513-1517. <https://doi.org/10.2337/dc09-0114>.
47. Munshi MN, Slyne C, Segal AR, Saul N, Lyons C, Weinger K. Liberating A1C goals in older adults may not protect against the risk of hypoglycemia. *J Diabetes Complications*. 2017;31(7):1197-1199. <https://doi.org/10.1016/j.jdiacomp.2017.02.014>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Crutzen S, Belur Nagaraj S, Taxis K, Denig P. Identifying patients at increased risk of hypoglycaemia in primary care: development of a machine learning-based screening tool. *Diabetes Metab Res Rev*. 2021;37(7):e3426. doi:10.1002/dmrr.3426