1-6-2023

Modifiable and Non-modifiable Factors Associated with DKA among Children and Adolescents with Type 1 Diabetes: A Machine Learning Exploration Using the T1D Exchange Data Set

Bridget Bassett

ABSTRACT

Modifiable and non-modifiable factors associated with DKA among children and adolescents with Type 1 diabetes: A machine learning exploration using the T1D Exchange data set

By

Bridget Bassett

Dec 8, 2022

INTRODUCTION: Diabetic ketoacidosis (DKA) is a serious life-threatening complication among pediatric patients with Type 1 diabetes. Even one instance of DKA can predispose a patient to more episodes of DKA in the future compounding the complications and risks.

AIM: The aim of this study is to use LASSO, a new variable selection method, to determine novel risk factors for DKA.

METHODS: The T1D Exchange dataset was used for a new variable selection technique for diabetic ketoacidosis (DKA) among pediatric patients in the United States. With DKA as a binary outcome, the HPGENSELECT procedure was used while LASSO or L1 regression was employed to create sparse models for variable selection.

RESULTS: The following modifiable variables were selected: number of blood glucose checks per patient per day, BMI, albumin creatinine ratio, systolic and diastolic blood pressure, BUN levels, having a hypoglycemic event in the previous three months and lipid levels (HDL/LDL/total cholesterol/triglycerides). The non-modifiable variables that were selected in the model are the following: age, diabetes duration in years, height and months from exam date. The model did produce an acceptable AUC for predictive ability.

DISCUSSION: The problem of finding modifiable risk factors for pediatric patients continues to be challenging, even if it is vitally important. The data in this study were both collected retrospectively and voluntarily, and their use for a predictive model should be used with caution. Machine learning techniques offer the potential to identify novel risk factors for DKA among pediatric patients if EHR are used and the dataset is large enough.

MODIFIABLE AND NON-MODIFIABLE FACTORS ASSOCIATED WITH DKA AMONG CHILDREN AND ADOLESCENTS WITH TYPE 1 DIABETES: A MACHINE LEARNING APPROACH

by

BRIDGET J. BASSETT

B.CE., GEORGIA INSTITUTE OF TECHNOLOGY

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

MODIFIABLE AND NON-MODIFIABLE FACTORS ASSOCIATED WITH DKA AMONG CHILDREN AND ADOLESCENTS WITH TYPE 1 DIABETES: A MACHINE LEARNING APPROACH

by

BRIDGET J. BASSETT

Approved:

Ruiyan Luo

Dr. Ruiyan Luo
Committee Chair

Dr. Ike Okosun
Committee Member

Dec 8, 2022
Date

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

___Bridget J. Bassett____
Signature of Author

ACKNOWLEDGMENTS

I would like to thank my family for supporting me getting this MPH, my thesis advisor Dr. Luo, Dr. Okosun and the many faculty members at Georgia State that have given me advice and support during my time there.

LIST OF TABLES

INTRODUCTION

Type 1 Diabetes (T1D) is an autoimmune disease characterized by progressive loss of pancreatic beta cells resulting in insulin deficiency and high blood sugar. This high blood sugar is called hyperglycemia. Prior to 1921, patients diagnosed with Type 1 diabetes were given no hope of long-term survival. The discovery of artificial insulin was one of the greatest breakthroughs in medical history (Rosenfeld, 2002). Since the invention of insulin, there have been continuous improvements in diabetes care and technology enabling better quality of life and long-term outcomes. These breakthroughs include continuous glucose monitors, subcutaneous insulin pumps and the hormone glucagon for hypoglycemic emergencies (Lauritzen et al., 1979). A healthy life is now possible with these advancements for individuals with Type 1 diabetes.

Despite these advances, few adult or pediatric patients maintain optimal levels of blood glucose. The gold standard for measuring an average blood glucose over the previous three months is the HbA1c value. Current guidelines suggest an endocrinology visit every three months for all persons with Type 1 diabetes that includes an HbA1c test. The vast majority of pediatric patients do not maintain an HbA1c value less than the value recommended by the American Diabetes Association and the International Society of Pediatric and Adolescent Diabetes of 7.5% (Chiang et al., 2014). Only 17% of patients under the age of 18 have an A1C below 7.5% (Foster et al., 2019).

There are widely known short- and long-term complications of poorly managed Type 1 diabetes in the medical literature. Long term complications are usually the result of years of mild to moderate hyperglycemia and are considered microvascular complications. These microvascular complications target organ systems that place a high demand on microvasculature such as the eyes, kidneys and nerves of the extremities (Nathan, 2014). The findings from the Diabetes Control and Complication Trial (DCCT) show that patients could slow the development of eye, kidney and nerve disease with intensive diabetes treatment. This type of intensive diabetes treatment is difficult to manage and maintain over a long period of time.

The two most widely known acute complications of Type 1 diabetes are extreme hypoglycemia and diabetic ketoacidosis (DKA). They are emergencies and are fatal if not treated quickly. Both complications are widely considered preventable. The aim of this current study uses a novel variable selection technique using machine learning and LASSO regression for DKA occurrence in pediatric patients. Using a large dataset from the T1D Exchange Registry, we hope to determine modifiable and non-modifiable risk factors for pediatric and adolescent patients to improve the body of evidence for future studies.

What is DKA?

Diabetes ketoacidosis is a complication of severe insulin deficiency. This insulin deficiency leads to hyperglycemia, dehydration and ketogenesis leading to acidosis, coma, and death if untreated. DKA can occur at diagnosis or anytime during the life of a person with Type 1 as

Type 1 is a chronic autoimmune illness. DKA occurs because of inappropriate management of blood sugar levels or accidental or intentional omission of insulin.

Current known risk factors for DKA include omission of insulin, limited access to medical services and unrecognized interruption of insulin delivery in patients using an insulin pump (Wolfsdorf et al., 2018). A population-based cohort study in Switzerland over 10 years showed that the risk was highest at around 15 years of age for DKA. The risk was most pronounced during adolescence and for girls.

Not maintaining optimal glycemic control has been shown to be a major risk factor for DKA in pediatric patients. Worsening glycemic control in puberty is common. Poor glycemic control during adolescence has been shown to be caused by poor diabetes acceptance, psychiatric disorders (depression, anxiety, eating disorders) and risk-taking behavior that is seen in adolescent development (Ebrahimi et al., 2022). Risk for DKA is also increased in children who omit insulin, children with poor metabolic control or previous episodes of DKA, gastroenteritis with persistent vomiting and inability to maintain hydration, children with psychiatric disorders, including eating disorders, children with difficult or unstable family circumstances, peripubertal and adolescent girls, binge alcohol consumption and children with limited access to medical services (Raghupathy, 2015).

DKA and especially recurrent DKA can cause permanent cognitive deficits in all patients but particularly in pediatric patients. The desire for autonomy during adolescence, refusing parental

support for diabetes management and peer acceptance can also play a behavioral role in the care of diabetes management for teenagers. The risk for DKA in established Type 1 diabetes patients is 1 to 10% per patient per year (Wolfsdorf et al., 2018).

Recurrent DKA is associated with older age (in relation to pediatric patients or the early teenage years), higher HbA1c levels and higher insulin doses. DKA was found to be most frequent in adolescents and associated with higher HbA1c values, nonwhite race, lack of private health insurance and lower household income (Jefferies et al., 2015).

Children and teens who have Type 1 diabetes have an increased mortality compared to the general population and DKA is the leading cause of death for pediatric patients with Type 1 particularly if it is complicated by cerebral edema (Edge et al., 1999). Even one occurrence of moderate to severe DKA in young children has been found to be associated with lower cognitive scores and altered brain growth. Cerebral edema because of DKA in the study by Edge, et al. is one of the common causes of death among children under 20. The incidence of ketoacidosis was 8 per 100 person-years and increased with age in girls, with some stratification among age for factors associated with it in a study among a cohort of 1243 children in Denver over 19 years (Rewers et al., 2007).

Since the basic treatment of DKA involves insulin administration and the correction of fluid and electrolyte imbalances, most patients are managed in a hospital intensive care unit to be monitored closely. For this reason, DKA treatment in the United States has an average annual

cost of more than $5 billion dollars (*National Diabetes Statistics Report 2020. Estimates of Diabetes and Its Burden in the United States.*, 2020). Resource use for treating DKA among pediatric patients is yet another reason why more information is needed to determine new risk factors for DKA. Eighty percent of DKA episodes occurred among 20% of those children with recurrent events. In 2014, the DKA hospitalization rate among persons with diabetes aged less than 45 was approximately 27 times the rate among persons older than 65 years old (Benoit et al., 2018).

Modifiable versus Nonmodifiable Risk Factors

Most of the research on the incidence of DKA for pediatric patients and the risk factors that have been determined are not modifiable by pediatric or adolescent patients as they are dependent on factors outside of their control. Family income, race/ethnicity and insurance type are some examples of non-modifiable risk factors. BMI and HbA1c are modifiable with proper continued education or family support. Other biomarkers such as blood pressure, lipid levels and measures of kidney health such as BUN level are also modifiable. Our study aims to use machine learning to help distinguish between these modifiable and non-modifiable risk factors.

There are many documented options to provide education and support to patients and family members including a care coordination approach. There are also several different diabetes education methods to help address HbA1c levels, BMI, correct insulin dosage and sick day management. There is some evidence on cognitive -behavioral interventions and family systems approaches as options to address modifiable risk factors (Wagner et al., 2015).

The extended health belief model can help explain the socio-psychological factors at play for pediatric and adolescents with Type 1 diabetes. The extended health belief model describes both coping mechanisms and self-management behaviors that are critically related to overall health (Harvey & Lawson, 2009). The Gillibrand, et al (2006) study found that those young adult patients with the highest family support had the highest scores in self-care of diabetes. This indicates the need to support the whole family in diabetes education (Gillibrand & Stevenson, 2006).

Not only are a high percentage of pediatric patients with Type 1 living with higher than optimal HbA1c levels, but more young people are also being diagnosed with the disease every year. The global incidence of Type 1 is increasing 3% per year in children and adolescents and 5% per year in preschoolers (Patterson et al., 2014).

REVIEW OF THE LITERATURE

Machine learning is an inherently different method for data analysis. Risk prediction models prior to using machine learning were based on a literature review and clinical knowledge of the disease or diagnosis to inform the model. This method builds a model with variables that are based on current knowledge. Machine learning allows researchers to look for new explanatory variables that might help explain some new or unknown effects or variables. The effort in this

study has been to use a new method for variable selection for a desired outcome. It is an exploration of machine learning.

Variable selection is an important step in any analysis. Variable selection methods include filter, wrapper and embedded methods. Filter methods allow the user to filter out the data for feature selection based on general characteristics of the data. Filtering methods require the user to determine a cut-off value or threshold for variable inclusion. These can include F-ratios, selectivity ratios and VIP scores. Wrapper methods evaluate subsets through iterations of the algorithm and return the best performing subset. This includes forward and backward selection. Embedded methods include forms of regularization, such as L1 or LASSO as we have done in this analysis (Kavakiotis et al., 2017; Sorochan Armstrong et al., 2022).

There are many examples of employing the wrapper method for DKA detection. Abakar, et al. used a tree-based classifier as a method for reducing the number of features in their DKA prediction model. They divided their features into categories and used a random forest algorithm to reduce the number of features from 28 to 5 (Abaker & Saeed, 2020).

Fralick, et al. used two different machine learning algorithms for classification and for feature selection. Gradient boosted trees defined predictors for DKA as having had a previous DKA, baseline HbA1c, baseline creatinine level, use of medications for dementia and baseline bicarbonate level. They used LASSO regression and found that prior DKA, digoxin use, use of medications for dementia and recent hypoglycemia were predictors for DKA (Fralick et al.,

2021). This analysis used a combination of the predictors they found and then assessed them in a logistic regression.

Fan, et al. used LASSO regression to identify candidate predictors for predicting acute kidney failure among patients in intensive care units for DKA. The Medical Information Mart for Intensive Care III database was used for this study. A multivariate logistic regression was done using the predictors discovered from the LASSO regression. A nomogram or a graphical representation as a function of several variables was constructed for the outcome of acute kidney disease (Fan et al., 2021).

In a study on cerebral oedema as a complication diabetic ketoacidosis in children, researchers used stepwise variable selection for their logistic regression model. Their study used biochemical variables at hospital admission and treatment variables while hospitalized. This study proved that treatment decisions while in the hospital played a role in whether patients developed cerebral edema while in DKA. Although the outcome is slightly different for this study, the aims and methods used highlight how important the variable selection method is when doing any type of analysis (Edge et al., 2006).

A similar study evaluating predictors of dehydration in children with DKA used forward stepwise logistic regression to identify clinical and biometric predictors for severe dehydration(Trainor et al., 2021) While both Trainor, et al. and Edge, et al. did not use a machine learning algorithm for variable selection they both illustrate the use of a filter variable selection.

Outside of variable selection, machine learning algorithms are currently used in academic research to improve the lives for those living with diabetes. Published research has included the use of machine learning to forecast future blood glucose levels in adults with Type 1 (De Paoli et al., 2021). In a viewpoint article in the JAMA in 2018, Beam et al, compared several methods of machine learning, deep learning, risk calculators, results from randomized clinical trials among others comparing the various methods and noting the benefits of machine learning (Beam & Kohane, 2018). This provided a valuable resource to objectively compare the methods and results of these most common ways to analyze data. Another work by Vehi, et al. in 2020 developed multiple methods of machine learning to prevent and predict hypoglycemic events for adults with Type 1. This study used four different machine learning methods that might enhance tight glycemic control without severe hypoglycemic events (Vehí et al., 2020).

A seminal work in the use of machine learning for DKA prediction was done by Lin, et al. Their study used six different types of flexible machine learning (XGBoost, distributed Random Forest and feedforward network) along with conventional machine learning (logistic regression and LASSO). They gathered 3400 DKA cases from adults in the Optum de-identified electronic health record system. The models they created using machine learning demonstrated similar performance and identified "overlapping, but different" top 10 predictors for adults with Type 1 diabetes developing DKA (Li et al., 2021). Unfortunately, their study did not include pediatric patients for DKA prediction but did show a unique set of different machine learning options that all produced overlapping, but different predictors for DKA.

Conversely, Schwartz, et al. created a simple risk prediction model for DKA from variables derived from EHR data for pediatric patients. They developed an automatic risk index using the variables of previous DKA, most recent HbA1c and type of insurance coverage. Machine learning was not used to develop this Risk index (Schwartz et al., 2022). They used electronic health record data that included ICD-9 and ICD-10 codes to classify Type 1. The team used a 70/30 split for training and validation as was done for the analysis described here.

Most of the studies in literature for machine learning and Type 1 diabetes focused on glycemic management, carbohydrate counting using a mobile application, infection detection, using momentary assessment for increasing self-management among pediatric patients, pediatric diabetic retinopathy, real time hypoglycemia prediction and detecting intentional insulin omission for weight loss among girls with type 1. There were no articles using a LASSO method for variable selection on DKA for a pediatric population.

An important role of variable selection and prediction models is to inform patients and families of the potential for an event that can be prevented (Steyerberg, 2009). It is generally considered that DKA is a preventable outcome. The advent of a risk score or marker to alert the care team of a change in mode of education methods, family support, diabetes supplies support or otherwise could prevent some of these instances of DKA. There are many prediction models being used in present day clinical practice. These include the Framingham Risk Score,

Ottawa Ankle Rules, EuroScore, Nottingham Prognostic Index and the Simplified Acute Physiology Score (Steyerberg, 2009).

METHODS AND PROCEDURES

Study Population

The data used in this study comes from the T1D Exchange(*T1D Exchange – T1D Exchange*, n.d.). The T1D Exchange began as a non-profit that focused on improving care and outcomes for those living with Type 1 diabetes. They maintain a longitudinal dataset of patient information from July 2007 to April 2018 called the T1D Registry. Participation in the dataset is voluntary across 83 academic sites across the United States. The database includes 34,013 participants. All patients that attend any of the Registry clinic sites are eligible to be entered in the Registry. For patients under 18 years of age, a parent or guardian provided consent for medical information release at regular office visits by both medical record extraction and by completing optional questionnaires. No tests or procedures were required to be part of the T1D Registry. The data is deidentified and publicly available. Permission to use the T1D Registry has been given through the Georgia State University Internal Review Board and is attached in the appendices.

The T1D Exchange Registry dataset contains values for five different possible study visits. These visits include Annual, Enrollment 1, Enrollment 2, Year 1 and Year 5. This analysis looked at Year

1 data. As DKA is common at diagnosis and the data collected is retrospective, we chose to use Year 1 data for this study. The response variable was the variable Pt_DKAFg where the value is 1 for the presence of a DKA diagnosis in the previous 3 months before a visit. For this dataset, we only know if a patient was given a diagnosis of DKA within the 3 months preceding the study visit for the T1D Exchange.

Outcome and Predictor Variables

The original dataset contains values from 140,461 observations. As it is common to be diagnosed with DKA at the time of diagnosis, we limited the observations that were labeled "Year 1" and those with age less than 26. Literature states that HbA1c levels remain stable through age 26. This ensured that we could maximize the number of observations in the analysis. This reduced the number of observations in the dataset to 13,644. The original dataset contained 97 variables or features that were candidates for the model selection. Some variables were removed logically as this study aims to predict DKA for patients under the age of 26. Removing these variables reduced some of the missingness seen in the original dataset, even though it is generally discouraged to remove variables when doing model selection using machine learning.

There were multiple options for insurance in the dataset as well. Since it is possible to have more than one type of insurance, we created a new variable which aggregated the insurance

type. Doing this allowed us to look at the presence of insurance as an effect for DKA rather than the specific type of insurance.

Summary statistics (mean or median) for continuous variables and percentages and frequency for categorical variables were computed for the baseline characteristics. The primary outcome variable was having a diagnosis of DKA in the three months prior to the appointment. The non-modifiable risk factors chosen from the variable selection are age, diabetes duration in years, months from consent date to date of HbA1c measurement, height and months from exam date. These were assessed from clinic study questionnaires and clinic data.

The modifiable risk factors chosen from the variable selection are the number of blood sugar checks per day as reported by the patient, BMI, the albumin creatinine ratio, HbA1c, lipid values (HDL, LDL, total cholesterol, triglycerides), weight in kilograms, systolic and diastolic blood pressure, BUN levels and having at least one severe hypoglycemic event in the past three months. These were also assessed from the clinic data and study questionnaires.

Method

Analysis was performed using SAS Studio. The HPGENSELECT procedure was used with Selection=LASSO. The HPGENSELECT procedure fits and builds generalized linear models and is designed for predictive modeling. It is considered a high-performance procedure. The LASSO

method can produce sparser and more interpretable models than some of the other model selection methods such as forward, backward and stepwise (Rodriguez, n.d.).

The HPGENSELECT procedure does not have graphics options, so PROC LOGISTIC was used with the variables chosen in the PROC HPGENSELECT procedure to assess the goodness of fit and graphics. Additionally, a /missing statement was added to allow the procedure to run and use the variables that were available. AUC and ROC values were used for assessing the predictive quality and goodness of fit for the resulting logistic regression.

Group lasso procedures allow for variables within a group (such as categorical predictors) to be removed. It is important to note that we did not allow for split variables. All the levels of the variable were chosen or none of them were chosen (Schreiber-Gregory et al., n.d.).

Each learning algorithm contains a loss function, an optimization criterion based on the loss function and an optimization routine that leverages training data to find a solution to the optimization criteria. Underfitting the model does not allow the model to predict well, or it has a high bias. Overfitting predicts well from the training but poorly for the test set. This can also be called high variance. In some cases, the “event” being modeled is relatively uncommon. This can sometimes lead to overfitting of a model. An overfit model underestimates the probability of an event in low-risk patients and overestimates it in high-risk patients (Pavlou et al., 2015).

To help alleviate these concerns, we can regularize or penalize the model to build a simpler or less complex model. This type of regularization for variable selection is considered an embedded method as described in the literature review above.

To regularize a model, a penalty term is added to the objective function whose value is higher when the model is more complex. One type of regularization is called L1 or Least Absolute Shrinkage and Selection Operator (LASSO.) LASSO regularizes the L1 norm of the vector of regression coefficients, i.e., the sum of the absolute values of coefficients. Restricting the L1 norm will shrink some of the coefficients to zero and hence remove the corresponding predictors from the model, which will enhance prediction accuracy (Pavlou et al., 2015).

Specifically, for the logistic regression on K predictors with LASSO regularization, the parameters are estimated by minimizing the following objective function

$$-2\log L + \lambda \sum_{[k=1]}^{K} |\beta_k| \qquad (1)$$

where L denotes the likelihood function, β_k, k=1…, K, are the slope coefficients to be estimated, and $\lambda > 0$ is the regularization parameter. We call (1) the penalized likelihood. The penalty term $\lambda \sum_{[k=1]}^{k} |\beta_k|$ controls the sparsity of the model. With this penalty term, some coefficients or β values are shrunk to zero. When λ is equal to 0, the penalty is 0 so minimizing (1) is equivalent to maximizing the likelihood to estimate parameters. The optimal regularization parameter is unknown and can be selected by cross-validation procedure or other criteria such as AIC, BIC. To do this, a dataset is divided into three parts; a training set, a

validation set and a testing set. In cases where there are relatively few events that are being used, it is most optimal to use 70% for the training/validation and 30% for the test set rather than using a 50/50% approach. We use the former ratio for this analysis. Then a model is employed using the optimal regularization parameter. In the HPGENSELECT procedure, a partition statement is utilized to do cross-validation.

RESULTS

The following variables resulted from the LASSO variable selection method used in this analysis. The response variable is the presence of DKA in the previous three months. The maximum regularization parameter in the model was given as 0.90329 and the chosen regularization parameter for the model was 0.010404. At each step of variable selection, the lambda is shown in selection details starting with a lambda of 1 at Step 0 and ending with a lambda of 0.0115 at Step 20 as seen in Table 2.

Variables included in the final model include age, albumin creatinine ratio (a measure of total amount of protein in blood), the average number of blood sugar tests per day, diastolic blood pressure, systolic blood pressure, body mass index (BMI), a blood urea nitrogen test (BUN) score that is a measure of urea nitrogen in the blood as a measure of kidney waste product, duration of diabetes in years, HbA1c, high density lipoprotein (HDL), height in centimeters, low density lipoprotein (LDL), having one severe hypoglycemic event in the three months prior to

the appointment, weight in kilograms, total cholesterol and triglyceride levels. Effects can be found in Table 3 at the end of the document.

Most of the effects chosen in the model have known associations to DKA including HbA1c values, BMI and age. This analysis is unique in that it offers both demographic as well as biological biomarkers for every patient. The novel biomarkers included in the model are the albumin/creatinine ratio, diastolic and systolic blood pressure, BUN, HbA1c, HDL, LDL, total cholesterol, and triglyceride values. Unfortunately, these are not markers that are collected at every normal 3-month endocrinology checkup, so their inclusion in a model would only be helpful when they were available. HbA1c, systolic and diastolic blood pressure are collected at every 3-month endocrinology visit. The American Diabetes Association recommends screening for dyslipidemia at diagnosis and every 3 years if found abnormal (American Diabetes Association Professional Practice Committee, 2021). More research is needed on whether these biomarkers for DKA are seen using prospective data rather than retrospective data for DKA models.

The given ROC value or Area under the curve (Figure 1) for the fitted model found is 0.7105. The greater the number for the ROC, the better the prediction. See Figure 1. Additionally, the model shows a high value for the Hosmer Goodness of Fit test (p-value is 0.7092). This test A low p-value for this test would indicate that there is no evidence that the model lacks fit with the data. As the model p-value is not low, we can conclude that the model fits the data accordingly.

Many of the variables found from the LASSO regression and included in the model have known relationships both to each other and to DKA. Several of the variables had odds ratios over 1, indicating that the odds of DKA, holding all the other variables constant, was greater than not having DKA. There were three variables that had odds ratios that had p-values less than 0.05 and are considered statistically significant. The value for diastolic blood pressure had an odds ratio of 1.045 (p-value=0.0302). The odds ratio for HbA1c is 1.368 (p-value=0.0007). Finally, the odds ratio for the variable indicating a severe hypoglycemic event is 0.239 (p-value=0.0003). As this odds ratio is less than one, rather than a risk factor for DKA, this indicates that it is a protective factor.

DISCUSSION

Interpretation of Findings

HbA1c and diastolic blood pressure were found to be significant in this analysis, in particular the finding of HbA1c. Gosmanov, et al. found that adults in the United States with an HbA1c above 9.0% had a 12-fold higher incidence of DKA (Gosmanov et al., 2021). Additionally, Pettus, et al. found that for pediatric patients, age and HbA1c were found to have similar results and that DKA instances increased with increased levels of HbA1c (Pettus et al., 2019). The odds ratio found with this analysis was 1.368 (p-value=0.007) for HbA1c. The Schwartz, et al. study also found HbA1c to be significant in their variable selection method. Fralick, et al.

found that baseline HbA1c was a strong predictor for DKA. Here LASSO regression was used for variable selection as well even though the population looked at Type 2 diabetes among adults (Fralick et al., 2021). Finally, it has been said that HbA1c is a very common predictor for DKA because it is one of the few biomarkers that is reliably available compared to many other markers, both demographic and biometric (Kavakiotis et al., 2017). As HbA1c is a requirement to receive continued insulin prescription refills, it is logical that it is reliably available. Pairing the availability with its definition it makes sense that HbA1c is often found to be a predictor for DKA.

The Abaker, et al. study found five variables in their variable selection method. None of these were the same as those found in our model. The variables found include age, infection years, BMI, sugar, symptom days with a response variable of DKA occurrence. The dataset used in this analysis consisted of only 937 cases with 27 possible selection variables (Abaker & Saeed, 2020). A small dataset could be a reason for the difference in variables chosen than those found in our analysis.

The Lin, et al. study examined several different machine learning methods for variable selection. This study found systolic blood pressure as a predictor in one of the five methods that were examined. Interestingly, three of the five methods picked HbA1c as the top predictor for DKA. The Fan, et al. study included both HbA1c in their list of variables chosen using LASSO regression for variable selection. Unfortunately, the outcome variable for this analysis was acute kidney failure for those with DKA.

Study Strengths and Limitations

When using machine learning, a large amount of data is needed since it will be divided into a test set, a training set and a validation set. Our response variable was rare, so the need was even greater to have a large number to start in the overall dataset. The dataset contained over 13,000 patients for use in this analysis. The T1D Exchange dataset also allowed us to have biomarkers as well as demographic data. Additionally, the care given at academic diabetes centers might ensure fidelity of the study materials.

Most of the publicly available government datasets such as National Health and Nutrition Examination Survey (NHANES), Kids' Inpatient Database (KID) and National Inpatient Sample (NIS) simply list "any diagnosis of diabetes" as a variable. A diagnosis of diabetes for these datasets usually means "Has a doctor ever told you have diabetes?" or having a biomarker such as HbA1c level that is higher than 6.5%. While persons with Type 2 diabetes can and do develop DKA over the course of their disease, the research question we were looking for included Type 1 diabetes only. The T1D Exchange Registry offered a sample of a population with Type 1 that was easily identifiable and available.

There are some limitations of the use of this dataset. First, as the study sites are large research medical sites with academic centers associated with them, the patient populations that utilize

these sites may have some underlying differences from patients who get their care from somewhere different than an annexed diabetes center. This may play a role in whether this is generalizable to the pediatric population with Type 1 diabetes. Finally, the dataset comes from 83 study sites across the country, but there was not a location or clinic site variable in the dataset available to the public. If academic center data was available, it would have been possible to do a nested or multi-level model and use that data.

Additionally, while there are multiple methods of imputation used in analysis, it is important to note the variables that have missingness and assess why they are so. Unfortunately, in the initial stages of analysis, every record in the dataset contained some missing data. This is likely due to the nature of the Registry. It is voluntary and there are only 5 maximum entries per participant. At each of five possible visits, participants provided information to the researcher. The visit options were Annual, Enrollment 1, Enrollment 2, Year 1 and Year 5. This analysis looked at Year 1 data. There were 3 variables that did not have any entries for the population we were interested in. These included a value for TSH, GFR and a value for the total number of units of insulin used. These values were likely not collected for pediatric patients. After their removal, we added a /missing statement in SAS that allowed the model to run with the variables that were in the dataset. We did not impute any missing data in this analysis. Imputing data was not logical as we were looking for novel unknown variables to be placed in the model.

Some of the biomarkers included in the study can be understood to provide insight into both general health, but also cardiac and endocrine system health such as systolic and diastolic blood pressure, the albumin/creatinine ratio and urea level in the urine (BUN). Since the values were recorded after a possible DKA event (both questions were asked or recorded in the previous 3 months), it is possible that these biomarkers changed because of the DKA diagnosis rather than them providing predictive ability. These biomarkers can be altered by an occurrence of DKA.

Lastly, it is possible that there was more than one instance of DKA in the previous 3 months for the variable chosen. The question asked of the patients was "did you have one or more DKA events in the 3 months prior to the appointment?" There are only a few effects in our new model that are collected at every 3-month endocrinology clinic visit or are already known. These include weight, height, BMI, HbA1c, insurance status, age at diabetes diagnosis, diabetes duration and patient education level.

Future Study

Considering these limitations, future research could benefit from the use of electronic health records to have more robust and complete data for variable selection and modeling. As most large health systems have electronic healthcare databases of patient clinical and demographic data, doing an analysis using one of those would be an appropriate next step.

Using a large Electronic Health Record (EHR) dataset, rather than a voluntary retrospective dataset might provide some insights into the variables that can predict future incidences of DKA.

Using a multilevel model using the T1D exchange dataset would enable analysis for within patient DKA and to detect biometric changes over time. Replicating this data on the Year 5 data as well as the Enrollment data might provide some important data for the diabetes community. Finally, if location or a time element could be added to this T1D exchange dataset, it would be possible to do different and more complex analysis relating to DKA among pediatric patients.

Conclusions

Using the T1D Exchange dataset and a LASSO regression for variable selection, we selected a model with 16 variables. Six of these variables were deemed to be non-modifiable. Ten of the variables can be considered modifiable. Two of the variables had statistically significant odds ratios for risk factors for DKA. These included HbA1c and diastolic blood pressure. The model chosen showed that it offered appropriate goodness of fit for assessing and targeting interventions to reduce the modifiable risk factors found in this analysis.

The problem of finding modifiable risk factors for pediatric patients continues to be challenging, even if it is vitally important. The data in this study were both collected retrospectively and voluntarily, and their use for a predictive model should be used with caution. Machine learning

techniques offer the potential to identify novel risk factors for DKA among pediatric patients if EHR are used and the dataset is large enough.

Finding modifiable risk factors for DKA could potentially lower the overall personal and societal cost of DKA to pediatric and young adult patients and their families. Reducing the incidence of DKA for pediatric type 1 patients is in the best interest of patients, their caregivers, and the health system. This analysis additionally contributes to the body of work regarding DKA prevention among pediatric patients with Type 1 diabetes.

Figures and Tables

Table 1: Baseline Patient Characteristics

Baseline Characteristics	Mean (or %), SD
Age (years)	13.9 (5.0)
Albumin Creatinine Ratio	20.1 (76)
Average number of Blood sugar checks per day	5.7 (2.5)
Systolic Blood Pressure	67.0 (8.6)
Diastolic Blood Pressure	113.0 (12.5)
BMI	21.9 (5.1)
BUN	13.7 (5.6)
Duration of Diabetes (years)	6.2 (4.6)
HbA1c (%)	8.5 (1.6)
HbA1c Imputed date (months)	14.5 (4.7)
HDL	58.2 (14.8)
Height (Cm)	155.0 (20.0)
LDL	91.7 (28.0)
Weight (Kg)	55.5 (21.4)
Total Cholesterol	166.0 (32.0)
Triglycerides	99.0 (75.0)
Hypoglycemic event in past 3 months (%)	7.2

Table 2: Progression of Variable Selection as Lambda Changes

Selection Details											
Step	Description	Effects In Model	Lambda	AIC	AICC	BIC	ASE	Validation AIC	Validation AICC	Validation BIC	Validation ASE
0	Initial Model	1	1	141.034	141.047	144.815	0.052	87.145	87.172	90.155	0.074
1	Triglyc entered	2	0.8	142.803	142.840	150.364	0.052	89.163	89.245	95.184	0.074
2	AlbCreatRat_mggNew entered	3	0.64	144.240	144.315	155.583	0.052	90.990	91.154	100.021	0.074
3		3	0.512	143.716	143.791	155.058	0.052	90.924	91.089	99.956	0.074
4	WeightKg entered	4	0.4096	145.385	145.510	160.508	0.052	92.976	93.252	105.019	0.074
5		4	0.3277	144.830	144.956	159.953	0.052	93.494	93.769	105.536	0.074
6	TotChol entered	5	0.2621	146.162	146.351	165.066	0.052	96.043	96.460	111.097	0.075
7		5	0.2097	145.613	145.801	164.516	0.052	96.601	97.018	111.654	0.075
8	BldPrDia entered	6	0.1678	146.893	147.158	169.577	0.052	98.749	99.337	116.813	0.075
9	HeightCm entered	7	0.1342	147.656	148.011	174.122	0.052	100.327	101.115	121.401	0.075
10	ExamDtMnC entered	9	0.1074	149.713	150.286	183.740	0.051	104.284	105.570	131.380	0.075
	LDL entered	9	0.1074	149.713	150.286	183.740	0.051	104.284	105.570	131.380	0.075
11		9	0.0859	147.979	148.552	182.005	0.051	104.448	105.734	131.544	0.076
12	bmi entered	10	0.0687	147.988	148.691	185.795	0.050	106.522	108.105	136.628	0.076
	HbA1c entered	10	0.0687	147.988	148.691	185.795	0.050	106.522	108.105	136.628	0.076
	TotChol removed	10	0.0687	147.988	148.691	185.795	0.050	106.522	108.105	136.628	0.076
13	age entered	14	0.055	152.429	153.788	205.360	0.050	115.162	118.273	157.311	0.077
	BldPrSys entered	14	0.055	152.429	153.788	205.360	0.050	115.162	118.273	157.311	0.077
	BUN entered	14	0.055	152.429	153.788	205.360	0.050	115.162	118.273	157.311	0.077
	HDL entered	14	0.055	152.429	153.788	205.360	0.050	115.162	118.273	157.311	0.077
	LDL removed	14	0.055	152.429	153.788	205.360	0.050	115.162	118.273	157.311	0.077
	TotChol entered	14	0.055	152.429	153.788	205.360	0.050	115.162	118.273	157.311	0.077
14	WeightKg removed	13	0.044	145.772	146.946	194.921	0.049	114.484	117.161	153.622	0.078
15	diabDur entered	15	0.0352	147.608	149.166	204.319	0.048	119.500	123.082	164.659	0.079
	WeightKg entered	15	0.0352	147.608	149.166	204.319	0.048	119.500	123.082	164.659	0.079
16	BGTestAvgNumPtRep entered	16	0.0281	146.815	148.587	207.307	0.048	122.654	126.744	170.824	0.079
17	LDL entered	17	0.0225	145.986	147.986	210.259	0.047	126.341	130.977	177.522	0.080
18		17	0.018	143.588	145.588	207.861	0.046	128.225	132.862	179.406	0.082
19		17	0.0144	141.666	143.666	205.938	0.046	130.335	134.971	181.515	0.083
20	Pt_SHFlg entered	18	0.0115	144.129	146.629	215.963	0.046	136.468	142.314	193.670	0.084

Maximum Regularization Parameter	0.90329
Chosen Regularization Parameter	0.010414

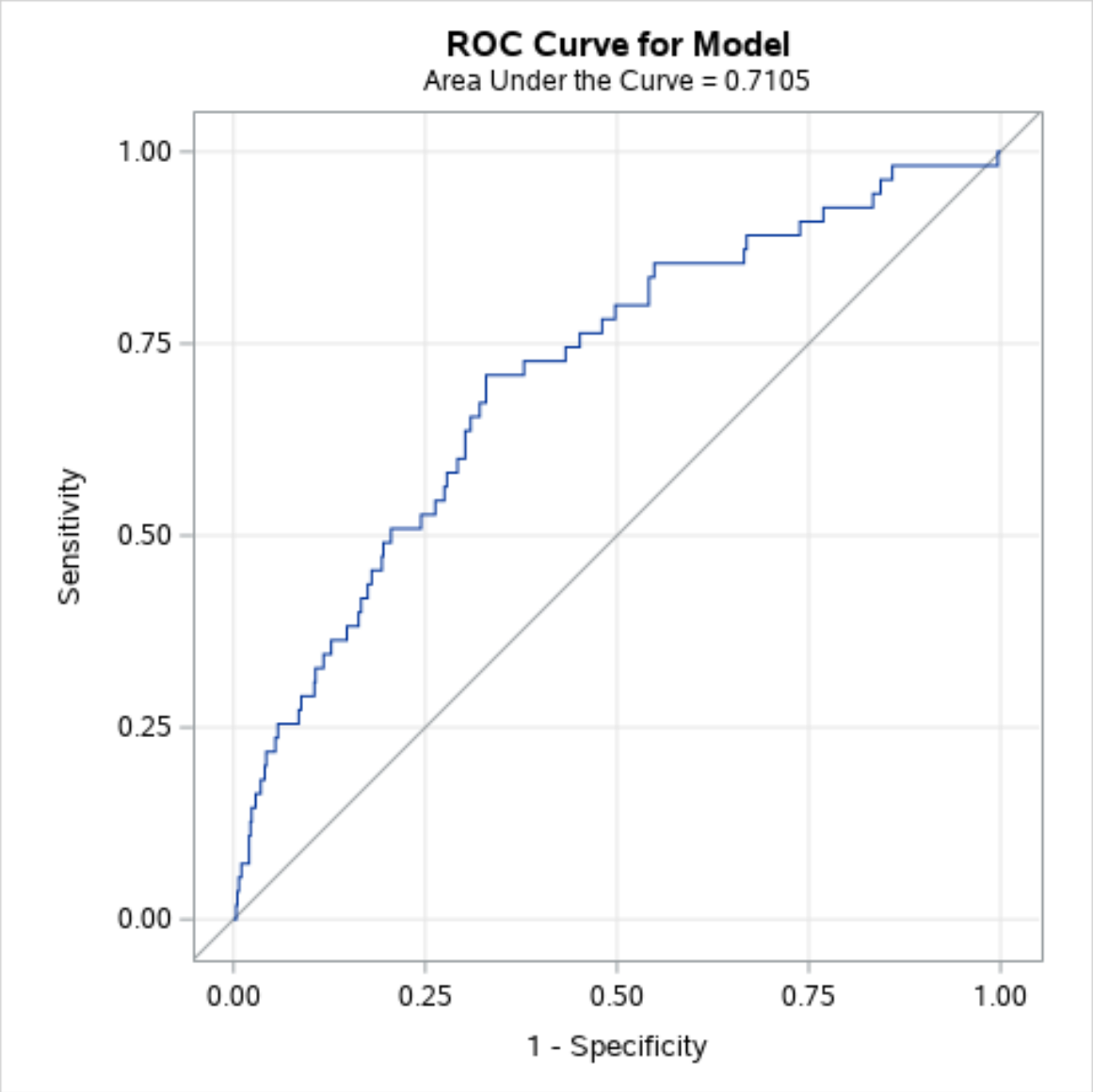
Table 3: Odds Ratio estimates of the effects chosen in the Lasso regression

Effect	Point Estimate for Odds Ratio	95% Wald Confidence Limits	
Age	0.858	0.746	0.985
Albumin Creatinine Ratio	1.002	0.998	1.006
Number of Blood Glucose Tests per day reported by the patient	0.995	0.847	1.169
Diastolic Blood Pressure	1.045	1.004	1.008
Systolic Blood Pressure	1.014	0.980	1.050
BUN (Blood Urea Nitrogen Test)	1.032	0.960	1.109
Diabetes Duration in Years	0.997	0.918	1.083
Body mass index (BMI)	1.411	0.946	2.103
Time in Months since Exam date	1.030	0.959	1.106
HbA1c	1.368	1.141	1.641
HDL	1.016	0.976	1.058
Height (in cm)	1.111	0.992	1.249
LDL	1.022	0.988	1.057
Severe Hypoglycemic Event in the previous three months	0.239	0.111	0.517
Weight in Kg	0.877	0.753	1.012
Total Cholesterol	0.978	0.946	1.102
Triglyceride level	1.001	0.993	1.008

Table 4 : Modifiable and Non-Modifiable Variables

Non-Modifiable	Modifiable
Age	Number of Blood Glucose checks reported by the patient
Diabetes duration in years	BMI
Months from consent date to date of HbA1c measurement	Albumin Creatinine ratio
Height	HbA1c
Months from exam date	HDL/LDL/Total Cholesterol/Triglycerides
	Weight
	Systolic Blood Pressure
	Diastolic Blood Pressure
	BUN
	At least one hypoglycemic event in the past 3 months

Figure 1: ROC Curve for model chosen.



REFERENCES

Abaker, A. A., & Saeed, F. A. (2020). Towards transparent machine learning models using feature sensitivity algorithm. *Jurnal Informatika*, *14*(1), 15. https://doi.org/10.26555/jifo.v14i1.a16983

American Diabetes Association Professional Practice Committee. (2021). 14. Children and Adolescents: Standards of Medical Care in Diabetes—2022. *Diabetes Care*, *45*(Supplement_1), S208–S231. https://doi.org/10.2337/dc22-S014

Beam, A. L., & Kohane, I. S. (2018). Big Data and Machine Learning in Health Care. *JAMA*, *319*(13), 1317–1318. https://doi.org/10.1001/jama.2017.18391

Benoit, S. R., Zhang, Y., Geiss, L. S., Gregg, E. W., & Albright, A. (2018). Trends in Diabetic Ketoacidosis Hospitalizations and In-Hospital Mortality—United States, 2000–2014. *Morbidity and Mortality Weekly Report*, *67*(12), 362–365. https://doi.org/10.15585/mmwr.mm6712a3

Chiang, J. L., Kirkman, M. S., Laffel, L. M. B., & Peters, A. L. (2014). Type 1 Diabetes Through the Life Span: A Position Statement of the American Diabetes Association. *Diabetes Care*, *37*(7), 2034–2054. https://doi.org/10.2337/dc14-1140

De Paoli, B., D'Antoni, F., Merone, M., Pieralice, S., Piemonte, V., & Pozzilli, P. (2021). Blood Glucose Level Forecasting on Type-1-Diabetes Subjects during Physical Activity: A

Comparative Analysis of Different Learning Techniques. *Bioengineering (Basel, Switzerland)*, *8*(6), 72. https://doi.org/10.3390/bioengineering8060072

Ebrahimi, F., Kutz, A., Christ, E. R., & Szinnai, G. (2022). Lifetime risk and health-care burden of diabetic ketoacidosis: A population-based study. *Frontiers in Endocrinology*, *13*. https://www.frontiersin.org/articles/10.3389/fendo.2022.940990

Edge, J. A., Ford-Adams, M. E., & Dunger, D. B. (1999). Causes of death in children with insulin dependent diabetes 1990–96. *Archives of Disease in Childhood*, *81*(4), 318–323. https://doi.org/10.1136/adc.81.4.318

Edge, J. A., Jakes, R. W., Roy, Y., Hawkins, M., Winter, D., Ford-Adams, M. E., Murphy, N. P., Bergomi, A., Widmer, B., & Dunger, D. B. (2006). The UK case–control study of cerebral oedema complicating diabetic ketoacidosis in children. *Diabetologia*, *49*(9), 2002–2009. https://doi.org/10.1007/s00125-006-0363-8

Fan, T., Wang, H., Wang, J., Wang, W., Guan, H., & Zhang, C. (2021). Nomogram to predict the risk of acute kidney injury in patients with diabetic ketoacidosis: An analysis of the MIMIC-III database. *BMC Endocrine Disorders*, *21*, 37. https://doi.org/10.1186/s12902-021-00696-8

Foster, N. C., Beck, R. W., Miller, K. M., Clements, M. A., Rickels, M. R., DiMeglio, L. A., Maahs, D. M., Tamborlane, W. V., Bergenstal, R., Smith, E., Olson, B. A., & Garg, S. K. (2019). State of Type 1 Diabetes Management and Outcomes from the T1D Exchange in 2016–2018. *Diabetes Technology & Therapeutics*, *21*(2), 66–72. https://doi.org/10.1089/dia.2018.0384

Fralick, M., Redelmeier, D. A., Patorno, E., Franklin, J. M., Razak, F., Gomes, T., & Schneeweiss, S. (2021). Identifying Risk Factors for Diabetic Ketoacidosis Associated with SGLT2 Inhibitors: A Nationwide Cohort Study in the USA. *Journal of General Internal Medicine*, *36*(9), 2601–2607. https://doi.org/10.1007/s11606-020-06561-z

Gillibrand, R., & Stevenson, J. (2006). The extended health belief model applied to the experience of diabetes in young people. *British Journal of Health Psychology*, *11*(1), 155–169. https://doi.org/10.1348/135910705X39485

Gosmanov, A. R., Gosmanova, E. O., & Kitabchi, A. E. (2021). Hyperglycemic Crises: Diabetic Ketoacidosis and Hyperglycemic Hyperosmolar State. In *Endotext [Internet]*. MDText.com, Inc. https://www.ncbi.nlm.nih.gov/books/NBK279052/

Harvey, J. N., & Lawson, V. L. (2009). The importance of health belief models in determining self-care behaviour in diabetes. *Diabetic Medicine*, *26*(1), 5–13. https://doi.org/10.1111/j.1464-5491.2008.02628.x

Jefferies, C. A., Nakhla, M., Derraik, J. G. B., Gunn, A. J., Daneman, D., & Cutfield, W. S. (2015). Preventing Diabetic Ketoacidosis. *Pediatric Clinics of North America*, *62*(4), 857–871. https://doi.org/10.1016/j.pcl.2015.04.002

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. *Computational and Structural Biotechnology Journal*, *15*, 104–116. https://doi.org/10.1016/j.csbj.2016.12.005

Lauritzen, T., Faber, O. K., & Binder, C. (1979). Variation in 125I-insulin absorption and blood glucose concentration. *Diabetologia*, *17*(5), 291–295. https://doi.org/10.1007/BF01235885

Li, L., Lee, C.-C., Zhou, F. L., Molony, C., Doder, Z., Zalmover, E., Sharma, K., Juhaeri, J., & Wu, C. (2021). Performance assessment of different machine learning approaches in predicting diabetic ketoacidosis in adults with type 1 diabetes using electronic health records data. *Pharmacoepidemiology and Drug Safety*, *30*(5), 610–618. https://doi.org/10.1002/pds.5199

Nathan, D. M. (2014). The Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Study at 30 Years: Overview. *Diabetes Care*, *37*(1), 9–16. https://doi.org/10.2337/dc13-2112

National Diabetes Statistics Report 2020. Estimates of diabetes and its burden in the United States. (2020). 32.

Patterson, C., Guariguata, L., Dahlquist, G., Soltész, G., Ogle, G., & Silink, M. (2014). Diabetes in the young—A global view and worldwide estimates of numbers of children with type 1 diabetes. *Diabetes Research and Clinical Practice*, *103*(2), 161–175. https://doi.org/10.1016/j.diabres.2013.11.005

Pavlou, M., Ambler, G., Seaman, S. R., Guttmann, O., Elliott, P., King, M., & Omar, R. Z. (2015). How to develop a more accurate risk prediction model when there are few events. *BMJ*, *351*, h3868. https://doi.org/10.1136/bmj.h3868

Pettus, J. H., Zhou, F. L., Shepherd, L., Preblick, R., Hunt, P. R., Paranjape, S., Miller, K. M., & Edelman, S. V. (2019). Incidences of Severe Hypoglycemia and Diabetic Ketoacidosis and

Prevalence of Microvascular Complications Stratified by Age and Glycemic Control in U.S. Adult Patients With Type 1 Diabetes: A Real-World Study. *Diabetes Care*, *42*(12), 2220–2227. https://doi.org/10.2337/dc19-0830

Raghupathy, P. (2015). Diabetic ketoacidosis in children and adolescents. *Indian Journal of Endocrinology and Metabolism*, *19*(Suppl 1), S55–S57. https://doi.org/10.4103/2230-8210.155403

Rewers, M., Pihoker, C., Donaghue, K., Hanas, R., Swift, P., & Klingensmith, G. J. (2007). Assessment and monitoring of glycemic control in children and adolescents with diabetes. *Pediatric Diabetes*, *8*(6), 408–418. https://doi.org/10.1111/j.1399-5448.2007.00352.x

Rodriguez, R. N. (n.d.). *Statistical Model Building for Large, Complex Data: Five New Directions in SAS/STAT® Software*. 23.

Rosenfeld, L. (2002). Insulin: Discovery and controversy. *Clinical Chemistry*, *48*(12), 2270–2288.

Schreiber-Gregory, D., Foundation, H. M. J., Bader, K., & Foundation, H. M. J. (n.d.). *Regulation Techniques for Multicollinearity: Lasso, Ridge, and Elastic Nets*. 23.

Schwartz, D. D., Banuelos, R., Uysal, S., Vakharia, M., Hendrix, K. R., Fegan-Bohm, K., Lyons, S. K., Sonabend, R., Gunn, S. K., & Dei-Tutu, S. (2022). An Automated Risk Index for Diabetic Ketoacidosis in Pediatric Patients With Type 1 Diabetes: The RI-DKA. *Clinical Diabetes*, *40*(2), 204–210. https://doi.org/10.2337/cd21-0070

Sorochan Armstrong, M. D., de la Mata, A. P., & Harynuk, J. J. (2022). Review of Variable Selection Methods for Discriminant-Type Problems in Chemometrics. *Frontiers in Analytical Science*, *2*. https://www.frontiersin.org/articles/10.3389/frans.2022.867938

Steyerberg, E. (2009). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* (Vol. 19). https://doi.org/10.1007/978-0-387-77244-8

T1D Exchange – T1D Exchange. (n.d.). Retrieved November 1, 2022, from https://t1dexchange.org/

Trainor, J. L., Glaser, N. S., Tzimenatos, L., Stoner, M. J., Brown, K. M., McManemy, J. K., Schunk, J. E., Quayle, K. S., Nigrovic, L. E., Garro, A., Rewers, A., Myers, S. R., Bennett, J., Kwok, M. Y., Olsen, C., Casper, T. C., Ghetti, S., & Kuppermann, N. (2021). Clinical and Laboratory Predictors of Dehydration Severity in Children with Diabetic Ketoacidosis. *Pediatrics*, *147*(3_MeetingAbstract), 501–503. https://doi.org/10.1542/peds.147.3MA5.501

Vehí, J., Contreras, I., Oviedo, S., Biagi, L., & Bertachi, A. (2020). Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. *Health Informatics Journal*, *26*(1), 703–718. https://doi.org/10.1177/1460458219850682

Wagner, D. V., Stoeckel, M., E. Tudor, M., & Harris, M. A. (2015). Treating the Most Vulnerable and Costly in Diabetes. *Current Diabetes Reports*, *15*(6), 32. https://doi.org/10.1007/s11892-015-0606-5

Wolfsdorf, J. I., Glaser, N., Agus, M., Fritsch, M., Hanas, R., Rewers, A., Sperling, M. A., & Codner, E. (2018). ISPAD Clinical Practice Consensus Guidelines 2018: Diabetic ketoacidosis and the hyperglycemic hyperosmolar state. *Pediatric Diabetes*, *19*(S27), 155–177. https://doi.org/10.1111/pedi.12701

APPENDICES

INSTITUTIONAL REVIEW BOARD

Mail: P.O. Box 3999 In Person: 3rd Floor
Atlanta, Georgia 30302-3999 58 Edgewood
Phone: 404/413-3500 FWA: 00000129

September 02, 2022

Principal Investigator: Ruiyan Luo

Key Personnel: Bassett, Bridget J; Luo, Ruiyan

Study Department: Georgia State University, School of Public Health

Study Title: Modifiable and Non-modifiable factors associated with DKA among children and

adolescents with Type 1 Diabetes: A machine learning exploration using the T1D exchange data set

Submission Type: Application for Designation of Not Human Subjects Research

IRB Number: H23108

Reference Number: 371531

Thank you for your Application for Designation of Not Human Subjects Research. Based on the information provided, this submission has been determined to be not human subjects research. This correspondence should be maintained with your records.

Please do not hesitate to contact the Office of Research Integrity at 404-413-3500 if you have any questions or concerns.

Sincerely,

Susan Vogtner

Susan Vogtner, IRB Co-Vice Chair