Spring 4-12-2021

# Identification of Variables for Hemoglobin A1c Prediction

Erica A. Richards
erichards@unmmg.org

# Identification of Variables for Hemoglobin A1c Prediction

Erica A. Richards

A scholarly project submitted to the College of Nursing

In partial fulfillment of the requirements

For the degree

Doctor of Nursing Practice

University of New Mexico

College of Nursing

Albuquerque, New Mexico

| | |
|---|---|
| Scholarly project chair: | Dr. Christine Delucas DNP, MPH, RN, NEA-BC |
| Scholarly project committee member: | Dr. Melissa Cole DNP, MSW, RN-BC, NEA-BC, FACHE |
| Date of submission: | April 12, 2021 |

COLLEGE
OF NURSING

# "Identification of Variables for Hemoglobin A1c Prediction"

**Erica Richards, MSN, MBA**

A. Christine Delucas. DNP, MPH, RN, NEA-BC (Chair)



Melissa Cole, DNP, MSW, RN-BC, NEA-BC, FACHE (Member)

**Abstract**

An evolution in healthcare is taking place throughout our country. Healthcare providers are required to achieve an ever-growing number of regulatory and metric-driven standards. This transformation aims to improve both the quality and cost of healthcare in the United States. Diabetes mellitus (DM) is one of the most costly chronic diseases in the United States. Since DM has high mortality, high morbidity, and low quality of life rates, it is no wonder that DM is one of the focuses of the National Committee for Quality Assurance performance improvement tools known as the Healthcare Effectiveness Data and Information Set (HEDIS). In addition, the HEDIS set of metrics is increasingly being used to judge the quality of healthcare providers and health systems, as well as link this performance to reimbursement. For healthcare providers to remain financially viable, they need to score well on these metrics. The aim of this study was to identify which common quantitative laboratory data elements in the Cerner Health Facts database correlate to poorly controlled glycosylated hemoglobin (HbA1c) levels in adult patients with DM that could be used in the future to create a mathematical model to predict which patients may have poorly controlled HbA1c levels.

## Dedication

There are many individuals who have influenced me over the years in my personal life, education, and professional career—far too many to mention. I am forever changed for having met and learned from each of them. I consider myself fortunate to have had the wide range of experiences and opportunities that I have, as they have shaped me and brought me to where I am today.

I am honored to have a loving and supportive family that have journeyed through this life with me. I am grateful for my parents, Nancy and Marcus, for instilling in me the value of education and hard work. They have always believed in me and supported me even when it appeared that my goal was daunting. They taught me that quitting was not an option and that success was mostly achieved through perspiration, both of which have served me well over the years.

I owe my inspiration to complete my doctorate to my children. They always stood by me and encouraged me to get my work done even when there were other things we would rather be doing. Thinking of them, all six of them, has kept me moving forward as I really have done this for them—to show them that you are never too old for school. I love each and every one of you.

In addition to my children, my husband has stood by me and helped me whenever and wherever I have asked. Everything from driving the kids, making dinner, and trying to help me understand statistics. Thank you for putting up with me though these two years. I am sure it was not easy for you.

Deborah Walker, you, very simply, have expanded how I view the profession of nursing. Your devotion to the profession has inspired me and brought me back to the basics of what I went into nursing for—helping others. You have shown me that next level of what helping others looks like and given me a potential direction for the next steps in my career.

I would also like to thank Dr. Delucas and Dr. Cole for providing guidance and support even when I was not hearing what you were trying to tell me. You still stood by me while I was figuring things out the hard way.

Last, but by no means least, a special thank you to my cohort. You all are truly why I made it to this point in the program. Each one of you touched me in your own way. Being able to problem-solve with you and to see that all of you were not going to give up even when life made things more challenging was what encouraged me to stay in the program when life was challenging for me. I will miss our time together most of all after we graduate.

I consider each of these groups my families. And with that, I dedicate the following work to my families. I know that I could not possibly have touched all of your lives as much as you have touched mine, but I do hope that I have made you proud.

**Acknowledgements**

This project was supported in part by the National Center for Research Resources and the National Center for Advancing Translational Science of the National Institutes of Health through grant number UL1TR00449, the University of New Mexico Clinical and Translational Science Center.

## Table of Contents

List of Tables ..... 9

List of Acronyms ..... 10

CHAPTER 1: INTRODUCTION AND BACKGROUND ..... 12

Problem Statement ..... 13

Study Purpose and PICOT Question ..... 13

Objectives and Goals ..... 14

CHAPTER 2: REVIEW OF THE LITERATURE ..... 15

Search Strategy ..... 15

Literature Review ..... 15

Demographics ..... 16

Provider Characteristics ..... 17

Socioeconomic Factors ..... 18

Comorbidities ..... 18

Duration of Disease ..... 19

Medication ..... 20

Intensification ..... 20

HbA1c/Glucose Levels ..... 21

Healthcare Utilization ..... 21

EMR Systems/Predictive Data ..... 22

Literature Synthesis ..... 23

CHAPTER 3: THEORETICAL MODEL AND METHODOLOGY ..... 25

Theoretical Model ..... 25
Methodology ..... 25
Study Population ..... 25
Sampling Procedure ..... 26
Research Design ..... 27
Ethical Issues ..... 27
Data Protection Plan ..... 27
Statistical Analysis ..... 27
Budget ..... 28
CHAPTER 4: RESULTS AND DISCUSSION ..... 29
Interpretation of Findings ..... 31
Discussion ..... 32
Implications for Practice ..... 33
Limitations and Strengths of the Study ..... 34
Limitations ..... 34
Strengths ..... 34
Suggestions for Further Research ..... 35
Concluding Remarks ..... 35
References ..... 37
Appendix ..... 42

## List of Tables

Table 1. 2018 HEDIS HbA1c Results 13

Table 2. Level of HbA1c Control by Mean Age 16

Table 3. Demographic Characteristics of Participants 28

Table 4. Association of HbA1c Levels to Other Common Laboratory Values 29

## List of Acronyms

| | |
|---|---|
| ALC | absolute lymphocyte count |
| ALP | alkaline phosphatase |
| ALT | alanine aminotransferase |
| AMC | absolute monocyte count |
| ANC | absolute neutrophil count |
| AST | aspartate aminotransferase |
| AUC | area under the receiver operating characteristic curve |
| BMI | body mass index |
| BUN | blood urea nitrogen |
| CART | classification and regression trees |
| DM | diabetes mellitus |
| EMR | electronic medical record |
| eGFR | estimated glomerular filtration rate |
| HbA1c | Hemoglobin A1c |
| HDL | high-density lipoprotein cholesterol |
| HEDIS | Healthcare Effectiveness Data and Information Set |
| HFDB | Cerner Health Facts database |
| HIPAA | Health Insurance Portability and Accountability Act |
| HMO | health maintenance organization |
| IQR | interquartile range |
| LDL | low-density lipoprotein cholesterol values |
| LR | logistic regression |
| MCV | mean corpuscular volume |
| MPV | mean platelet volume |

| | |
|---|---|
| NCQA | National Committee for Quality Assurance |
| OR | odds ratio |
| PPO | preferred provider organization |
| TPC | theory of planned change |
| VLDL | very-low-density lipoprotein |

## CHAPTER 1: INTRODUCTION AND BACKGROUND

An evolution in healthcare is taking place throughout our country. Healthcare providers are required to achieve an ever-growing number of regulatory and metric-driven standards. This transformation aims to improve both the quality and cost of healthcare in the United States. Healthcare spending in the United States accounted for 18% of the gross domestic product from 2009 through 2019, which is approximately twice as much as other high-income countries, despite similar utilization rates (Centers for Medicare and Medicaid Services, 2020; Papanicolas et al., 2018). Americans might be willing to pay more for a higher quality of care, but, despite the high cost, our overall quality ranks last when compared to industrialized countries (Paterson, 2013; The Commonwealth Fund, 2020).

People with chronic health conditions account for 90% of national annual healthcare expenditures. Of the chronic diseases, diabetes mellitus (DM) is one of the most costly, at $237 billion every year. With such a high cost, DM is more expensive than cancer, obesity, Alzheimer's, heart disease, or stroke (Centers for Disease Control and Prevention, 2019). Additionally, DM has high mortality, high morbidity, and low quality of life rates (Trikkalinou et al., 2017; Wilkinson et al., 2014). It is no wonder that DM is one of the focuses of the National Committee for Quality Assurance (NCQA) performance improvement tools known as the Healthcare Effectiveness Data and Information Set (HEDIS; Centers for Disease Control and Prevention, 2017).

HEDIS is used by many organizations, including the Centers for Medicare and Medicaid Services, to benchmark quality of care in more than 90 measures across 6 domains (NCQA, 2019). It has become commonplace for healthcare providers and health systems to be judged on these measures, and reimbursement for services is increasingly linked to performance in these areas. For healthcare providers to remain financially viable, they need to score well on these metrics.

As shown in Table 1, NCQA reported for the 2018 HEDIS (which includes both type 1 and type 2 DM patients) that the percentage of patients who had a glycosylated hemoglobin (HbA1c) result that was considered controlled at less than 8% ranged from a low of 49% (Medicaid health maintenance organization [HMO] patients) to a high of 68% (Medicare preferred provider organization [PPO] patients). In addition, the percentage of patients who had an HbA1c that was considered poorly controlled at greater than 9% ranged from a low of 20% (Medicare PPO patients) to a high of 41% (Medicaid HMO patients) (NCQA, 2020). Therefore, approximately one in three patients with DM in the report had poorly controlled HbA1c.

**Table 1**

*2018 HEDIS HbA1c Results*

| Level of Control | Medicaid HMO (%) | Medicare PPO (%) |
|---|---|---|
| Controlled (HbA1c < 8%) | 48.7 | 68.4 |
| Poorly Controlled (HbA1c > 9%) | 41.2 | 19.9 |

## Problem Statement

Due to the perceived difficulty in controlling this disease and the high costs involved both monetarily and healthwise, it is important to be able to identify the variables available in electronic medical record (EMR) administrative data, which enables clinicians to predict and eventually improve the HbA1c control of adult DM patients. The definition of administrative data for the purposes of this research was any data that are generated and collected from any encounter in a healthcare system (Cadarette & Wong, 2015).

## Study Purpose and PICOT Question

The aim for this project was to identify which common quantitative laboratory data elements in the Cerner Health Facts database (HFDB) correlate to poorly controlled HbA1c levels of greater than 8% in adult patients with DM that could be used in the future to create a mathematical model to predict which patients may have a poorly controlled HbA1c. The

question posed was: Which common quantitative laboratory variables in the HFDB correlate with poorly controlled HbA1c (defined as HbA1c greater than or equal to 8%) in adult diabetic patients' data between July 1, 2016 and July 1, 2017? For the purposes of this study, common laboratory variables are the components of a complete blood count, a comprehensive metabolic panel, and a lipid profile.

## Objectives and Goals

The goal of this quantitative study was to identify quantitative laboratory variables that correlate with uncontrolled HbA1c in adult diabetic patients found in the HFDB. An objective of the study was to add to the body of literature related to laboratory values correlated with uncontrolled HbA1c. A literature review revealed that little research has been conducted on the quantitative laboratory values found in many EMRs and their correlation to uncontrolled HbA1c.

# CHAPTER 2: REVIEW OF THE LITERATURE

A literature review was conducted to determine what current research exists on HbA1c values and the variables that influence it. The initial review focused on literature about the relationship of blood glucose to other variables. Since literature directly related to this topic is limited, additional studies were included that reviewed the predictive ability of EMR data.

## Search Strategy

The first step in this research was describing the administrative data variables that are identified in the literature. The literature search was conducted on the PubMed and the Cumulative Index of Nursing and Allied Health Literature databases. This search focused on the question: What administrative data variables influence HbA1c control in adult diabetic patients? Search filters included English language only and human studies limited to adults 19 years of age and older. Key words included *A1C* and *administrative data*. Due to the low number of articles returned, the search time period was set to the past 10 years. This search resulted in a total of 70 articles. After title review, four articles were identified for this paper, as they appeared to most closely address the research question.

Due to the low number of applicable studies, an additional key word search was performed with the same filters for the past 5 years changing the search words to *EMR data* and *predict* to determine if EMR data has been used to predict outcomes. This search resulted in a total of 130 articles. After removal of duplicates from the previous search and title review, an additional four articles were identified for this paper, as they appeared to include data related to using EMR data to predict future outcomes.

## Literature Review

A wide range of variables were found in the reviewed literature. Main themes among the variables were demographics, provider characteristics, socioeconomic factors, comorbidities, duration of disease, medication, intensification, HbA1c/glucose levels, healthcare utilization,

and EMR systems/data. Considering the diversity of variables found, it is helpful to consider the findings for each type of variable in turn.

**Demographics**

Demographics are common data points found in administrative data and frequently used in the investigation of health outcomes correlations. Demographic data points customarily found in health system EMRs are age, gender, race, ethnicity, and geographic area. Less frequently found are educational attainment and income level. In this literature review, age was the most frequently assessed demographic in relation to HbA1c control (five studies). Three out of the eight studies reviewed (Juarez et al., 2014; Kamal et al., 2014; LeBlanc et al., 2015) found that those with controlled HbA1c tended to be older. Even though patient ages varied slightly among the studies, the findings were relatively similar, as shown in Table 2.

**Table 2**

*Level of HbA1c Control by Mean Age*

| Level of Control | Juarez et al., 2014 | Kamal et al., 2014 |
|---|---|---|
| Optimal | 63.1 | 66.5 |
| Poor | 59.8 | 58.4 |

In addition, LeBlanc et al. (2015) found that each additional year of age increased the odds of HbA1c being in optimal glycemic control by 2%. In contrast, Lee et al. (2018) reported that of the patients with poorly controlled HbA1c, 57% were aged 45 to 64, and only 27% were aged greater than or equal to 65. Related to this, Li et al. (2019) found that patients aged 85 or older were negatively associated with hypoglycemia (odds ratio [OR] 0.6). Considering that age is not a laboratory value, its relationship to HbA1c values was not reviewed in this study.

Another commonly found demographic in EMRs is gender. Three of the eight studies reviewed gender in relationship to HbA1c control. All three found that males were more likely to have poor HbA1c control (Juarez et al., 2014; Kamal et al., 2014; Lee et al., 2018). The fourth

study (LeBlanc et al., 2015) reported that males who had higher baseline HbA1cs, used insulin, and had a lengthier duration of DM were more likely to have poor control. In regard to duration of disease, Juarez et al. (2014) found that those with DM for 10 years or more were more likely to have HbA1c above the American Diabetes Association guidelines' goal of less than 7%. Considering that gender is not a quantitative variable, its relationship to HbA1c values was not reviewed in this study.

Nelson (2002) reported that "disparities in healthcare exist even when insurance status, income, age and severity of conditions are comparable" (p. 666). These disparities have a greater impact on death rates from DM in racial and ethnic minorities than in whites (Nelson, 2002). Race is commonly captured in EMRs and was evaluated in three of the reviewed studies. Juarez et al. (2014) found that in a sample of patients enrolled in a large health plan in Hawaii, Native Hawaiians were more likely to be above the American Diabetes Association's goal for HbA1c than other groups, such as whites, Japanese, and other Pacific Islanders. Lee et al. (2018) found that the proportion of minority residents in a census tract was a better predictor (with 86% accuracy) of locations with poor HbA1c control, with the portion of non-Hispanic black residents having the highest specificity (91%). Additionally, Li et al. (2019) found that African-Americans were more likely to have hypoglycemia (OR 1.8), while being Hispanic was negatively associated with hypoglycemia (OR 0.7). Considering that race is not a quantitative variable, its relationship to HbA1c values was not reviewed in this study.

**Provider Characteristics**

Provider characteristics are not usually found in EMRs. However, one of the studies reviewed, LeBlanc et al. (2015), examined provider characteristics for relationships with HbA1c control. Le Blanc et al. looked at provider demographics, duration of employment with the practice group, type of primary care training (internal medicine versus family practice), degree (medical doctor versus nurse practioner/physician assistant), number of patients on the provider's panel, and percentage of patients with DM on their panels. The researchers found

that none of the provider characteristics were associated with poor HbA1c control. They went so far as to suggest that focusing on individual providers' characteristics may not be a good use of resources. Considering that these provider characteristics are not a quantitative variable, their relationship to HbA1c values was not studied in this study.

**Socioeconomic Factors**

Socioeconomic factors such as income, education, employment, community safety, social supports, and, indirectly, insurance coverage are generally considered contributing factors in health outcomes (County Health Rankings, 2019). However, many of these factors are not captured in most EMRs. Only two of the studies reviewed utilized socioeconomic factors in their research. Lee et al. (2018) used census-tract-level socioeconomic factors and demographics of minority race/ethnicity, poverty, low education, and low median income as predictor variables in mapping "hot spots" of poor glycemic control (an HbA1c of greater than 9%). They found that these characteristics were less accurate than the proportion of minority residents in predicting hot spots of poor HbA1c control. The researchers reported that the accuracy of these socioeconomic factors ranged from 70% to 78%, which might be the reason some of the demographic factors, which had an accuracy of 86%, were better predictors. Additionally, Li et al. (2019) found that having Medicaid coverage was positively associated with hypoglycemia (OR 1.5). Since these factors are not quantitative variables or are not available in the HFDB, their relationship to HbA1c values was not reviewed in this study.

**Comorbidities**

Comorbidities such as heart disease and cardiovascular risks are known to be associated with worse health outcomes (Valderas et al., 2009) and are available in the EMR administrative data. Three of the reviewed studies assessed the association of certain comorbidities and blood sugar. Juarez et al. (2014) looked at this association using the Johns Hopkins Adjusted Clinical Group methodology for the *International Classification of Diseases, Ninth Revision, Clinical Modification* to determine which patients were considered to have high morbidity. In patients

with comorbidities, the researchers found that high morbidity was significantly associated with well-controlled HbA1c levels of less than or equal to 7%. Kamal et al. (2014) reviewed the association of obesity, low-density lipoprotein cholesterol values (LDL), high-density lipoprotein cholesterol values (HDL), smoking status, and depression with HbA1c control. They found that patients with poorly controlled HbA1c had a higher prevalence of obesity (body mass index [BMI] greater than 30), LDL levels greater than or equal to 100 mg/dl (36%) and HDL less than or equal to 40 mg/dl (51%). Smoking status did not have an association with HbA1c control, but higher prevalence of depression was present among those with poorly controlled HbA1c. Additionally, Li et al. (2019) found that infection within 30 days (OR 2.5), diabetic neuropathy (OR 1.6), alcohol consumption (OR 1.6), chronic heart failure (OR 1.3), dementia (OR 1.5), and hypoglycemia in the previous 12 months (OR 2.4) were all positively associated with hypoglycemia. These variables that are quantitative common laboratory values were reviewed in this study.

**Duration of Disease**

The duration of time that a patient has had DM is not typically found in EMR administrative data, as it is not typically a discrete field. Two of the reviewed studies assessed the relationship of the duration of a patient's having DM with the control of HbA1c. Juarez et al. (2014) found that longer durations of DM were significantly associated with poor HbA1c control. For those with poor control, 48% had DM for 10 or more years, as compared to 35% with HbA1c of less than or equal to 7%. LeBlanc et al. (2015) used the length of time a patient was in a diabetic registry as a substitute for the length of time the patient had DM. Their research found that for those in the diabetes registry with poorly controlled HbA1c, the odds of poor control increased by 3% for each year they were in the registry. Since the duration of DM is not directly available in the HFDB, it was not a variable reviewed in this study.

**Medication**

The type of medication a patient is using for glucose control is a data point that can be found in most EMRs, assuming the patient has been prescribed a medication while in the care of that health system. Two reviewed studies assessed the correlation between medication type and HbA1c control. Le Blanc et al. (2015) showed that "users of insulin were more than twice as likely to have poor control than nonusers" (p. 603). Conversely, Juarez et al. (2014) found that insulin users made up more than half of those with HbA1c in control. Li et al. (2019) found that use of non-long-acting insulin and no antibiotic use within 30 days were positively associated with hypoglycemia (OR 2.2) in patients, whereas use of long-acting insulin within 90 days (OR 0.7) was negatively associated with hypoglycemia. Further, Juarez et al.'s (2014) study indicated that of those who had HbA1c levels that were considered out of control, 44% were on a single oral medication, 7% were on multiple oral medications, and 23% were on oral medication and insulin together. In contrast, respective figures for those considered controlled were 35%, 3%, and 7%.

Differences among these study populations and different dosing of medications may account for the differences in these studies. In addition, any administrative data for a health system's EMR only indicates what medication(s) were prescribed in an outpatient environment. No objective data exist on whether or not a patient administered the medication or administered it correctly. As type of medication prescribed is not a quantitative variable, it was not reviewed in this study.

**Intensification**

LeBlanc et al. (2015) looked at intensification of medication therapy. They defined intensification as "the addition of a new antihyper-glycemic medication or an increase in dosage of a current oral medication" (p. 598) within 90 days of an elevated HbA1c. The researchers did not establish any correlation between provider characteristics and intensification once a patient's HbA1c was above goal. They did, however, find that older patients with a greater

change in their HbA1c values were more likely to have their medication therapy intensified (OR 1.54). In addition, the researchers found that a greater change in HbA1c in patients with an initial controlled HbA1c was most strongly associated with the probability of intensification (OR 1.54). Since intensification is not a quantitative variable, it was not reviewed in this study.

**HbA1c/Glucose Levels**

There are generally a variety of laboratory values available in an EMR. HbA1c is a laboratory value that is frequently found for DM patients. LeBlanc et al. (2015) reviewed baseline HbA1c levels in relation to poor control. They found that patients with a baseline HbA1c of greater than or equal to 7% had twelve times higher odds of poor control (OR 12.4). In addition, Li et al. (2019) found a negative association between hypoglycemia and serum calcium (OR 0.5). Due to the fact that baseline HbA1c is not readily identifiable in the HFDB, this variable was not considered in this study.

**Healthcare Utilization**

If a health system has its scheduling platform integrated with its EMR, healthcare utilization data are available for study. These data can be difficult to interpret depending on how the scheduling system is designed, thus making it potentially difficult to study utilization. Three of the studies reviewed here were able to access utilization data in order to assess their relationship to HbA1c control. Kamal et al. (2014) found that the average number of office visits per patient were higher in those with poor HbA1c control (5.15) than among all DM patients in the sample (4.85) and all patients in the primary care group (2.34). LeBlanc et al. (2015) found that patients with at least one visit to an endocrinologist were twice as likely to have poor control (OR 2.31), but those with at least one visit to a specialist other than an endocrinologist were more likely to have HbA1c that was controlled (OR 1.13).

Lee et al. (2018) sought to identify areas of poor glycemic control in New York City for unique patients in a citywide A1c registry using indirect measures instead of a population-based A1c registry. The researchers found that healthcare utilization measures more accurately

identified geographic areas of poor glycemic control than demographic and socioeconomic factors. Lee et al.'s research indicates that rates of inpatient hospitalizations and emergency department visits should be included as quantitative variables that may correlate to glycemic control. Because utilization data are not a quantitative laboratory value, they were not reviewed in this study.

**EMR Systems/Predictive Data**

This last theme group focused on EMR systems being used to support decision-making and predictive modeling. Three of the reviewed studies addressing this theme found that predictive models can be successful using EMR data (Li et al., 2019; Sahni et al., 2018; Zhao et al., 2019). Li et al. (2019) found that multiple logistic regression (LR), classification and regression trees (CART), and random forest models were successful in and had similar performance in predicting hypoglycemia when using EMR laboratory values of glucose and serum calcium (random forest mean AUC 90%, LR mean AUC 89%, CART mean AUC 88%).

Sahni et al. (2018) studied the ability of EMR data to be used to predict 1-year mortality in patients emergently admitted to the hospital. The researchers reported that a random forest model using EMR data points of age, blood urea nitrogen (BUN), platelet count hemoglobin, and creatinine was successful (AUC 0.86) at predicting 1-year mortality of this group. Zhao et al. (2019) reported that a random forest regression model including the EMR data of recent estimated glomerular filtration rate (eGFR), age, BMI, gender, obesity, hypertension, and diabetes successfully predicted future eGFR (mean coefficient of determination 0.95). Even though these studies do not directly address the study question of this project, they do support the concept that EMR data can be used to predict future laboratory values and patient outcomes, which is of value to this study.

In the fourth reviewed study addressing this theme, Adaji et al. (2008) performed a systematic review of 25 articles related to DM and medical records systems. These researchers found that EMR systems have been successfully used to support clinicians with timely access to

data, as well as with reminders for evidence-based care, both of which improve the process measures of DM care. However, Adaji et al. found that systems designed to remind clinicians of evidence-based care did not improve patients' laboratory values of HbA1c.

These studies do not show any associations between quantitative laboratory data and DM. However, that is not the value these studies bring to the current project. Their value here lies in the fact that they indicate EMR administrative data can be and has been used to successfully predict patient laboratory values using a variety of statistical models.

## Literature Synthesis

As evidenced by numerous studies on variables that correlate to blood glucose and HbA1c, a limited set of quantitative and qualitative variables that impact blood glucose values have been frequently studied. However, very few studies have narrowly focused on the correlation of quantitative administrative data found in EMRs to HbA1c values, and even fewer on the predictive ability of EMR administrative data. As healthcare evolves, health systems and providers are in need of this research, as their reimbursements are becoming increasingly tied to the attainment of outcome measures, such as those of HEDIS. Additionally, utilizing this research will help health systems attain their goal of providing better healthcare. In the reviewed studies that addressed this question, the study populations tended to be so specific that the studies' generalizability to the general population or other populations is questionable. Therefore, more research on the subject is needed. The current project was intended to help fill that gap.

Thus far, existing research has indicated that potentially multiple variables have a statistically significant impact on blood glucose or HbA1c levels. Those potential variables are patient age, duration of disease, medication type, life style factors, BMI, comorbidities, race/ethnicity, LDL, HDL, HbA1c levels, serum calcium, healthcare utilization, the diagnosis of depression, and infection within the past 30 days, not all of which are quantitative laboratory values. However, these variables may very well be specific to the populations that were studied,

and some of the impact results are contradictory. Therefore, clinicians need clearer guidance on which data points found in their EMR's administrative data are potential predictors of which patients are more likely to have HbA1c levels that are out of control.

# CHAPTER 3: THEORETICAL MODEL AND METHODOLOGY

## Theoretical Model

DM is a disease in which the actions of the person with the disease have a great impact on the outcomes and trajectory of the disease's effects. Since the course of this disease is highly contingent on the patient, healthcare providers have an obligation to help patients achieve the best outcome possible. Therefore, the theory of planned change (TPC) was chosen as the theoretical framework for this study.

Kurt Lewin, a social psychologist of the early 20th century who is considered the father of social psychology, developed the TPC (Lewin, 1997; Shirey, 2013). The TPC consists of three stages or phases: unfreezing, moving or transition, and refreezing (Shirey, 2013). Within these three phases, there are driving forces and restraining forces (Shirey, 2013). Driving forces help move patients toward the desired change or goal, while restraining forces hinder patients from moving toward the change or goal (Shirey, 2013).

In the phase of unfreezing, healthcare providers can assist patients in moving toward change (moving phase) by helping them recognize the potential negative outcome from the trajectory of their disease. Identifying the quantitative variables that indicate when patients are on a negative trajectory with their disease process gives healthcare providers additional driving force to share with patients: a quantifiable forecast of their HbA1c.

## Methodology

### Study Population

This retrospective qualitative study used de-identified patient data from the HFDB. The HFDB consists of over 158,300,000 patients from across the nation. Approximately 435,476 unique patients with DM with at least one HbA1c lab result existed in the HFDB between July 1, 2016 and July 1, 2017. Of those, 21,108 had an HbA1c result greater than 8% after the date of diagnosis of DM. For this study, only the first HbA1c value after the diagnosis of DM and the

independent variables closest in time to that HbA1c value (regardless of directionality) were used.

Power analysis performed using G*Power (Heinrich Heine Universität Düsseldorf, 2020) indicated that this sample of 21,108 should have sufficient (80%) statistical power to detect very small effect sizes of Cohen's $f^2 = .0004$ using a threshold of $p = .05$. The racial and gender make-up of this population was varied due to the large sample size and geographic spread of the facilities that report to the database; however, it does not match the racial make-up of the population of the United States. This decreases the generic applicability of this study. Honing the study to specific populations would be a step for further research.

**Sampling Procedure**

The HFDB is a database comprised of patient records from over 600 participating hospitals and clinics that have a Cerner EMR. The longitudinal, relational data from over 106 million unique patients are de-identified and available at the patient level. Data in HFDB are extracted directly from the EMR from hospitals with which Cerner has a data use agreement. Encounters may include pharmacy, clinical and microbiology laboratory, admission, and billing information from affiliated patient care locations. All admissions, medication orders and dispensing, laboratory orders, and specimens are date and time stamped, providing a temporal relationship between treatment patterns and clinical information. Cerner Corporation has established operating policies compliant with the Health Insurance Portability and Accountability Act (HIPAA) to establish de-identification for HealthFacts (University of New Mexico Health Sciences Center, n.d.-b).

Prior to the requesting of the data for this study, Institutional Review Board permission was obtained from the University of New Mexico Health Sciences Center (see Appendix). The sample for this study included all unique adult patients within the database identified with the diagnosis of DM. The time period for this study was July 1, 2016 and July 1, 2017. With a final

sample size of 21,108 patients, the power of this study is that it is able to detect very small effect size correlations in the data.

**Study Design**

The study design was a retrospective quantitative analysis with multiple linear regression to assess the presence of a relationship between the predictor variables (independent variables) and the outcome variables of HbA1c greater than 8%. The regression analysis identified the variables with statistically significant correlations. In addition to the correlations, the effect size of the correlation was also identified for each of the variables using Cohen's *f*-squared. Regression beta coefficients were also identified in the analysis.

**Ethical Issues**

This study did not present any specific ethical concerns. Through the use of the HFDB and an honest broker, participant confidentiality was guaranteed since the database contained only de-identified patient data. Data points contained a patient identifier that was not traceable back to the patient for the purpose of identifying which data came from which patient.

**Data Protection Plan**

The data protection plan for this study was governed by the University of New Mexico Health Science Center's HFDB Data Use Agreement. All HFDB data was only used by the recipient and approved study team. Data were not used to re-identify any person, nor were they redistributed by any method. Data were kept on a secured drive behind the firewall at the University of New Mexico Health Sciences Center and accessed only by the approved study team.

**Statistical Analysis**

All data requested from the HFDB met inclusion criteria. A significant number of patient records did not include an HbA1c value greater than 8% after the date of diagnoses with DM. These patient records were not included in the final analysis. Due to the large data size, statistical analysis and computer power had to be obtained from the University of New Mexico's

Clinical and Translational Science Center.

Descriptive statistics were used to summarize basic demographic information within the sample from the categorical variables. Multiple linear regression analysis was used on the quantitative laboratory values to determine correlations with the HbA1c values. The adjusted *R*-squared was converted to a Cohen's *f*-squared using the formula described by Cohen (1988) ($f^2 = R^2 / [1 - R^2]$) since G*Power, which was used for the power analysis, relies on Cohen's f-squared (Faul et al., 2009).

The *R*-squared was adjusted for age, race/ethnicity, marital status, and urban/rural status. Therefore, any correlations found are for the entire model, including the adjustment variables, rather than the independent variable alone.

**Budget**

The budget for this analysis was $3,000 for 40 hours of biostatistician time to run the statistical analysis at the University of New Mexico's Clinical and Translational Science Center. This cost was paid for by the student-investigator. No other costs were incurred.

## CHAPTER 4: RESULTS AND DISCUSSION

Of the 21,108 patients in the final sample, 10,736 (51%) were male and 10,371 (49%) were female. One (<1%) patient record missing gender was kept in the data. Median age of the patients represented in the data was 60 years. The majority were white (*n* = 14,481, 69%), male (*n* = 10,736, 51%), single (*n* = 9,962, 47%), and living in an urban environment (*n* = 15,375, 73%). Table 3 presents the demographic characteristics of the participants.

**Table 3**

*Demographic Characteristics of Participants (n = 21,108)*

| Demographic Characteristic | Median (IQR) or *n* (%) |
|---|---|
| Age | 60 (51, 69) |
| Race/Ethnicity | |
| White | 14,481 (68.6%) |
| Black | 3,965 (18.8%) |
| Asian | 2,145 (10.2%) |
| Native | 393 (1.9%) |
| Hispanic | 124 (0.6%) |
| Gender | |
| Male | 10,736 (50.9%) |
| Female | 10,371 (49.1%) |
| Marital Status | |
| Single | 9,962 (47.2%) |
| Widowed | 5,796 (27.5%) |
| Divorced/separated | 2,059 (9.8%) |
| Other | 2,592 (12.3%) |
| Urban | 15,374 (72.8%) |
| Rural | 5,734 (27.2%) |

The multiple linear regression analysis is shown in Table 4. The independent variables absolute lymphocyte count (ALC), absolute monocyte count (AMC), and absolute neutrophil count (ANC) are represented on Table 4 but did not have multiple linear regression run due to the small sample size.

**Table 4**

*Association of HbA1c Levels to Other Common Laboratory Values*

| EMR Laboratory Value | *n* | Correlation Coefficient (95% C.I.) | *p*-Value | Adjusted $R^{2\,a}$ | Adjusted Cohen's $f^{2\,a}$ |
|---|---|---|---|---|---|
| Albumin | 15,937 | -0.03 (-0.03 to -0.02) | <0.0001 | 0.04 | 0.04 |
| ALC | 8 | NA | NA | NA | NA |
| ALP | 17,929 | 3.53 (3.26 to 3.80) | <0.0001 | 0.05 | 0.05 |
| ALT | 16,880 | -0.39 (-0.51 to -0.27) | <0.0001 | 0.05 | 0.05 |
| AMC | 456 | NA | NA | NA | NA |
| ANC | 55 | NA | NA | NA | NA |
| Anion Gap | 15,092 | 0.14 (0.10 to 0.18) | <0.0001 | 0.04 | 0.04 |
| AST | 16,319 | -0.28 (-0.38 to -0.18) | <0.0001 | 0.01 | 0.01 |
| Bilirubin | 15,520 | 0.01 (0.00 to 0.01) | <0.0001 | 0.01 | 0.01 |
| BUN | 18,994 | 0.14 (0.07 to 0.20) | <0.0001 | 0.12 | 0.14 |
| Calcium | 14,990 | 0.00 (-0.01 to 0.00) | 0.2176 | 0 | 0 |
| Chloride | 18,255 | -0.49 (-0.53 to -0.45) | <0.0001 | 0.05 | 0.05 |
| Carbon Dioxide | 14,989 | -0.22 (-0.25 to -0.19) | <0.0001 | 0.04 | 0.04 |
| Creatinine | 16,140 | 0.00 (0.00 to 0.00) | 0.2769 | 0.11 | 0.12 |
| Glucose | 16,460 | 21.66 (20.83 to 22.49) | <0.0001 | 0.17 | 0.2 |
| Hematocrit | 16,407 | 0.05 (0.00 to 0.09) | 0.0311 | 0.05 | 0.05 |
| HDL | 15,657 | -0.12 (-0.25 to 0.00) | 0.049 | 0.02 | 0.02 |
| Hemoglobin | 17,642 | 0.06 (0.05 to 0.08) | <0.0001 | 0.09 | 0.1 |
| LDL | 6,613 | 2.37 (1.83 to 2.91) | <0.0001 | 0.05 | 0.05 |
| MCV | 17,241 | -0.20 (-0.25 to -0.15) | <0.0001 | 0.07 | 0.08 |
| MPV | 13,514 | 0.02 (0.01 to 0.03) | 0.0029 | 0.01 | 0.01 |
| Platelets | 17,281 | 0.60 (-0.05 to 1.24) | 0.0687 | 0.03 | 0.03 |
| Potassium | 18,266 | 0.00 (0.00 to 0.00) | 0.8895 | 0.02 | 0.02 |
| Red Blood Cells | 14,238 | 0.02 (0.01 to 0.02) | <0.0001 | 0.08 | 0.09 |
| Total Protein | 16,954 | 0.00 (0.00 to 0.01) | 0.4124 | 0.04 | 0.04 |
| Triglycerides | 13,641 | 4.34 (3.42 to 5.26) | <0.0001 | 0.04 | 0.04 |
| VLDL | 4,425 | 0.72 (0.40 to 1.04) | <0.0001 | 0.03 | 0.03 |
| White Blood Cell | 14,274 | 0.06 (0.03 to 0.09) | <0.0001 | 0.02 | 0.02 |

*Note*. Acronyms can be found in the List of Acronyms on page 11.

[a] Adjusted for age, race/ethnicity, marital status, and urban/rural status.

A statistically significant association existed between HbA1c and 21 of the 29 independent variables, with *p*-values ranging from 0.049 to <0.0001. The laboratory values of calcium, creatinine, platelets, potassium, and total protein did not have a significant association to HbA1c. Of the remaining statistically significant associations, all were found to have small effect sizes (Cohen's $f^2$ between 0.02 and 0.14), except for glucose, which had a medium effect size (Cohen's $f^2 = 0.2$). The independent variables were found to have mostly small correlation coefficients. Glucose, triglycerides, alkaline phosphatase (ALP), and LDL were found to have the largest correlation coefficients, respectively 21.66, 4.34, 3.53, and 2.37. The other statistically significant independent variables had smaller correlation coefficients ranging from -0.51 to 0.72.

**Interpretation of Findings**

Due to large size of the sample, statistical analysis had the ability to find some of the smallest correlations that existed in the data set between the common laboratory values and HbA1c values, as revealed in the large number of variables found to have a statistically significant association to HbA1c with a small effect size. While statistically significant, these laboratory values only explain a small portion of the variability in the HbA1c. Due to the statistical power of this analysis, the effect size and correlation coefficient were considered in addition to statistical significance. Based on the analysis of the statistical data, four variables are worthy of note.

First, a medium effect size (Cohen's $f^2 = 0.2$) and what appears to be a large correlation coefficient (21.66; 95% CI [20.83, 22.49]) existed in addition to the statistical significance between glucose values and HbA1c ($p$ <0.0001). It is logical that glucose values would be associated with HbA1c values, as an HbA1c value is a marker of average glucose values over the previous three months (Nitin, 2010). The correlation coefficient of 21.66 indicates that for every one unit increase in HbA1c, glucose rises by 21.66 units. This sounds like a significant number; however, on its own, a 21.66 unit rise in glucose would not necessarily be a large enough fluctuation to alert a clinician.

Three other variables of note are ALP, LDL, and triglycerides. Each showed a statically significant association with HbA1c values ($p < 0.0001$). The correlation coefficient for these variables is larger than those of the other statistically significant variables. The correlation coefficient for ALP indicates that for every single unit of increase in HbA1c, an associated increase in ALP of 3.53 units (95% CI [3.26, 3.80]) exists. As with the glucose variable, this number on its own may appear significant, but it is unlikely that a 3.53 unit rise in ALP would be enough for a clinician to be alerted. In addition, the effect size for this variable was small, indicating that this variable alone in the model accounts for only a very small portion of the variability in HbA1c.

The relationship of LDL and triglycerides to HbA1c is similar to that of ALP in that both have a small effect size. This indicates that each variable accounts for only a very small portion of the variability in HbA1c and the correlation coefficients may appear to be large numbers (2.37 and 4.24 respectively), as compared to the other variables in the analysis. However, on their own, a 2.37 rise in LDL and 4.24 rise in triglycerides are most likely not enough to alert a clinician.

An issue that was not considered in this study but which may impact the effect size of these statistically significant associations is chronologic directionality of the independent variables in relation to the dependent variables. The sequencing of laboratory values may have influenced the association of the variables. If the independent variable occurs prior to or after the time of the dependent variable, the statistical significance of the association, as well as the effect size, may be impacted.

## Discussion

This study used qualitative methods and sought to understand the association between uncontrolled HbA1c and commonly available EMR administrative data. To date, it is the only study to consider common EMR laboratory values (administrative data) in the HFDB in relation to HbA1c greater than 8%, and as such, contributes to the limited body of knowledge currently

available in the literature. The results of this study do not support findings from prior studies, as the study question and analysis were significantly different than those in the studies reviewed. The aim of this study was to identify which quantitative data elements in the HFDB correlate to poorly controlled HbA1c levels in adult patients with DM that could be used in the future to create a mathematical model to predict which patients may have poorly controlled HbA1c levels. This study was the first step in a multistep process to identify the predictor variables and create such a mathematical model.

The key finding from this study was that the majority of the common quantitative laboratory values (as adjusted) in the HFDB had a significant statistical association with HbA1c of greater than 8%; however, the effect sizes of those associations were small and, therefore, account for very little of the HbA1c variability. The only exception was the glucose variable. The glucose association to HbA1c was statistically significant and had a medium effect size. The association between glucose and HbA1c was expected, as HbA1c is a measure of average glucose over the preceding three months, and glucose has been previously found to have a linear relationship to HbA1c (Rohlfing et al., 2002). Therefore, this study further contributed to the body of knowledge the fact that no single quantitative laboratory value in the HFDB was strongly associated with poorly controlled HbA1c during the study period.

**Implications for Practice**

This study contributes to the body of knowledge the fact that there is not a single quantitative laboratory value in the HFDB that is strongly associated with poorly controlled HbA1c for the study time period. Glucose, with a medium effect size, has been the long-standing canary in the coal mine for patients with DM. No new single variable can be reviewed to clue in the clinician to a potential future HbA1c value of greater than 8%. Therefore, there are no immediate implications for practice change that can be deduced from this study, and clinicians need to continue to follow existing evidence-based practices for the care of the adult diabetic patients.

## Limitations and Strengths of the Study

### *Limitations*

In this study, the large sample size is not only a strength but also a limitation. The ability of this analysis to detect very small effect size, statistically significant correlations in the data is a potential limitation, in that a very small effect size means that the independent variable may or, most likely, may not contribute much to variation in the dependent variable. Also due to the large sample size, additional computing power beyond the average home or office computer system was needed for the statistical analysis of such a large dataset. Even with the large data set, this study is not generalizable to all populations due to the demographic makeup of the sample.

In addition, chronologic directionality was not considered in this study. When the independent variable occurs prior to the dependent variable, it is more likely that the independent variable may have an association with the dependent variable that may be used for predictive models. However, if the independent variable occurs after the dependent variable, the independent variable would not be able to be used for predictive models. Setting the directionality of the variables sequencing would be important for refining this study.

Additionally, other factors that were not studied or adjusted for might have impacted the associations between the independent variables. Factors such as the life span of red blood cells have been shown to impact the percentage of HbA1c (Cohen et al., 2008). Therefore, identifying the hematological factors that impact HbA1c could be a potentially important step in refining a model to predict HbA1c control.

### *Strengths*

The study had several strengths. First, with the sample coming from the HFDB through an honest broker, the risk of privacy or HIPAA compliance issues was minimized since all data had been de-identified prior to receipt. Second, with the large sample size, the statistical power detected very small effect size correlations in the data. In addition, investigator bias was limited

since there was no qualitative data interpretation and no interactions between the investigator and participants. Also, notably, the study would be relatively easy to replicate due to the ease and speed of accessing the HFDB.

**Suggestions for Further Research**

Given that HbA1c control is tied so closely with patient health and health system reimbursement, it is important for further studies building on this study to determine if a combination of independent variables exists that might lead to a predictive model of which adult DM patients are at risk for poorly controlled HbA1c. The independent variables identified in this study as statistically significant with a small effect size may very well lead to a larger effect size when combined with other independent variables in a multiple regression model. In addition, chronologic directionality of the variables should also be considered in future research.

Other independent variables commonly found in EMRs such as age, gender, and race could also be included in future studies, as conflict currently exists in the literature as to which variables contribute to HbA1c levels. In addition, honing research to specific racial/ethnic populations would improve the generalizability to those populations and may improve the predictive ability and fidelity of any model created, as differences in laboratory values exist for different racial, ethnic, and gender groups (Bergenstal et al., 2018; Beutler & West, 2005; Lim et al., 2015; Sellami et al, 2017).

Additionally, other methods of statistical analysis should be considered in further research. Previous researchers reported success in using multiple logistic regression, CART, and random forest models. Using one of these other models for analysis could improve the identification of variable associations.

**Concluding Remarks**

The potential to create mathematical models that predict which patients are at risk for complications or diseases is no longer a figment of science fiction. However, we are still in the infancy of being able to create such mathematical models with precision and fidelity. As DM is

one of the most costly chronic diseases with some of the lowest quality of life rates in the United States, being able to reasonably predict which patients will have uncontrolled HbA1c is a worthy research endeavor and should be continued beyond this study.

This study has advanced the understanding of the association between poorly controlled HbA1c and some of the common quantitative laboratory values found in EMRs around the nation. However, this is only a starting point in a multistep process required to achieve the goal of creating a predictive model for patients at risk for uncontrolled HbA1c. Further studies are needed to determine the combination of variables that will be successful in creating a predictive model. With such a model, clinicians will be able to intervene earlier with patients at risk for poorly controlled HbA1c, thus increasing the ability of DM patients to adjust their actions and change the outcomes and trajectory of their disease.

## References

Adaji, A., Schattner, P., & Jones, K. (2008). The use of information technology to enhance diabetes management in primary care: A literature review. *Informatics in Primary Care*, *16*(3), 229–237. https://doi.org/10.14236/jhi.v16i3.698

Bergenstal, R. M., Gal, R. L., & Beck, R. W. (2018). Racial differences in the relationship of glucose concentrations and hemoglobin A1c levels. *Annals of Internal Medicine*, *168*(3), 232–233. https://doi.org/10.7326/L17-0589

Beutler, E., & West, C. (2005). Hematologic differences between African-Americans and Whites: The roles of iron deficiency and alpha-thalassemia on hemoglobin levels and mean corpuscular volume. *Blood*, *106*(2), 740–745. https://doi.org/10.1182/blood-2005-02-0713

Cadarette, S. M., & Wong, L. (2015). An introduction to health care administrative data. *The Canadian Journal of Hospital Pharmacy*, *68*(3). doi:10.4212/cjhp.v68i3.1457

Centers for Disease Control and Prevention. (2017). *Healthcare effectiveness data and information set (HEDIS)*. https://www.cms.gov/Medicare/Health-Plans/SpecialNeedsPlans/SNP-HEDIS.html.

Centers for Disease Control and Prevention. (2019). *Health and economic costs of chronic disease*. https://www.cdc.gov/chronicdisease/about/costs/index.htm.

Centers for Medicare and Medicaid Services. (2020) *Historical*. CMS.gov. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences*, 2nd edition. Lawrence Erlbaum Associates.

Cohen, R. M., Franco, R. S., Khera, P. K., Smith, E. P., Lindsell, C. J., Ciraolo, P. J., Palascak, M. B., & Joiner, C. H. (2008). Red cell life span heterogeneity in hematologically normal people is sufficient to alter HbA1c. *Blood*, *112*(10), 4284–4291. https://doi.org/10.1182/blood-2008-04-154112

County Health Rankings. (2019) *Social and economic factors*. University of Wisconsin Population Health Institute. https://www.countyhealthrankings.org/explore-health-rankings/measures-data-sources/county-health-rankings-model/health-factors/social-and-economic-factors.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Heinrich Heine Universität Düsseldorf (2020). *G*Power* (Version 3.1.9.7) [Computer software]. https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html

Juarez, D. T., Ma, C., Kumasaka, A., Shimada, R., & Davis, J. (2014). Failure to reach target glycated A1c levels among patients with diabetes who are adherent to their antidiabetic medication. *Population Health Management*, *17*(4), 218–223. doi:10.1089/pop.2013.0099

Kamal, K. M., Chopra, I., Elliott, J. P., & Mattei, T. J. (2014). Use of electronic medical records for clinical research in the management of type 2 diabetes. *Research in Social and Administrative Pharmacy*, *10*(6), 877–884. doi:10.1016/j.sapharm.2014.01.001

LeBlanc, E. S., Rosales, A. G., Kachroo, S., Mukherjee, J., Funk, K. L., & Nichols, G. A. (2015). Do patient or provider characteristics impact management of diabetes? *American Journal of Managed Care*, *21*(9), 597–606.

Lee, D. C., Jiang, Q., Tabaei, B. P., Elbel, B., Koziatek, C. A., Konty, K. J., & Wu, W. Y. (2018). Using indirect measures to identify geographic hot spots of poor glycemic control: Cross-sectional comparisons with an A1C registry. *Diabetes Care*, *41*(7), 1438–1447. doi: 10.2337/dc18-0181

Lewin, G. W. (Ed.). (1997). *Resolving social conflicts field theory in social science*. American Psychological Association.

Li, X., Yu, S., Zhang, Z., Radican, L., Cummins, J., Engel, S. S., Iglay, K., Duke, J., Baker, J., Brodovicz, K. G., Naik, R. G., Leventhal, J., Chatterjee, A. K., Rajpathak, S., & Weiner, M. (2019). Predictive modeling of hypoglycemia for clinical decision support in evaluating outpatients with diabetes mellitus. *Current Medical Research and Opinion*, *35*(11), 1885–1891. https://doi.org/10.1080/03007995.2019.1636016

Lim, E., Miyamura, J., & Chen, J. J. (2015). Racial/ethnic-specific reference intervals for common laboratory tests: A comparison among Asians, Blacks, Hispanics, and White. *Hawai'i Journal of Medicine and Public Health*, *74*(9), 302–310.

National Committee for Quality Assurance. (2019). *HEDIS and performance measurement*. https://www.ncqa.org/hedis/

National Committee for Quality Assurance. (2020). *Comprehensive diabetes care*. https://www.ncqa.org/hedis/measures/comprehensive-diabetes-care/

Nelson, A. (2002). Unequal treatment: Confronting racial and ethnic disparities in health care. *Journal of the National Medical Association*, *94*(8), 666–668.

Nitin, S. (2010). HbA1c and factors other than diabetes mellitus affecting it. *Singapore Medical Journal*, *51*(8), 616–622.

Papanicolas, I., Woskie, L. R., & Jha, A. K. (2018). Health care spending in the United States and other high-income countries. *The Journal of the American Medical Association*, 319(10), 1024–1039. doi:10.1001/jama.2018.1150

Paterson, M. (2013). Mirror mirror on the wall: From reflective to transformative practice. *Health and Social Care Chaplaincy*, 1(1). doi:10.1558/hscc.v1i1.67

Rohlfing, C. L., Wiedmeyer, H. M., Little, R. R., England, J. D., Tennill, A., & Goldstein, D. E. (2002). Defining the relationship between plasma glucose and HbA(1c): Analysis of glucose profiles and HbA(1c) in the Diabetes Control and Complications Trial. *Diabetes Care*, *25*(2), 275–278. https://doi.org/10.2337/diacare.25.2.275

Sahni, N., Simon, G., & Arora, R. (2018). Development and validation of machine learning models for prediction of 1-year mortality utilizing electronic medical record data available at the end of hospitalization in multicondition patients: A proof-of-concept study. *Journal of General Internal Medicine*, *33*(6), 921–928. https://doi.org/10.1007/s11606-018-4316-y

Sellami, M., Chamari, K., Zagatto, A. M., Kebsi, W., Chaouachi, A., & Zouhal, H. (2017). Racial differences in hemoglobin and plasma volume variation: Implications for muscle performance and recovery. *Ethnicity & Health*, *24*(2), 182–193. https://doi.org/10.1080/13557858.2017.1315375

Shirey, M. R. (2013). Lewin's theory of planned change as a strategic resource. *The Journal of Nursing Administration*, 43(2), 69–72. https://doi.org/10.1097/nna.0b013e31827f20a9

The Commonwealth Fund. (2020). *U.S. Health Care from a Global Perspective, 2019: Higher Spending, Worse Outcomes?* https://www.commonwealthfund.org/publications/issue-briefs/2020/jan/us-health-care-global-perspective-2019#1

Trikkalinou, A., Papazafiropoulou, A. K., & Melidonis, A. (2017). Type 2 diabetes and quality of life. *World Journal of Diabetes*, *8*(4), 120–129. https://doi.org/10.4239/wjd.v8.i4.120

University of New Mexico Health Sciences Center. (n.d.-b). *Data use agreement for Health Facts*. Unpublished document. Health Sciences Center.

Valderas, J. M., Starfield, B., Sibbald, B., Salisbury, C., & Roland, M. (2009). Defining comorbidity: implications for understanding health and health services. *Annals of family medicine*, *7*(4), 357–363. https://doi.org/10.1370/afm.983

Wilkinson, A., Whitehead, L., & Ritchie, L. (2014). Factors influencing the ability to self-manage diabetes for adults living with type 1 or 2 diabetes. *International Journal of Nursing Studies*, 51(1), 111–122. doi:10.1016/j.ijnurstu.2013.01.006

Zhao, J., Gu, S., & McDermaid, A. (2019). Predicting outcomes of chronic kidney disease from EMR data based on Random Forest Regression. *Mathematical Biosciences*, *310*, 24–30. https://doi.org/10.1016/j.mbs.2019.02.001

# Appendix

## UNM Institutional Review Board Approval

HEALTH SCIENCES

Human Research Protections Program

May 13, 2020
Christine Delucas
ADelucas@salud.unm.edu

Dear Christine Delucas:

On 5/13/2020, the HRRC reviewed the following submission:

| | |
|---|---|
| Type of Review: | Initial Study |
| Title of Study: | Identification of variables that impact HbA1c |
| Investigator: | Christine Delucas |
| Study ID: | 20-254 |
| Submission ID: | 20-254 |
| IND, IDE, or HDE: | None |
| Submission Summary: | Initial Study |
| Documents Approved/Acknowledged: | • ACDelucas DATA USE AGREEMENT FOR HEALTH FACTS.pdf<br>• E RICHARDS DATA USE AGREEMENT FOR HEALTH FACTS.pdf<br>• HRP-582 - Richards Research 5 12 2020.pdf |
| Review Category: | Exempt: Category (4) Secondary research on data or specimens (no consent required) |
| Determinations/Waivers: | Informed Consent Not Applicable.<br>HIPAA Authorization Addendum Not Applicable. |
| Submission Approval Date: | 5/13/2020 |
| Approval End Date: | None |
| Effective Date: | 5/13/2020 |

The HRRC approved the study from 5/13/2020 to  inclusive. If modifications were required to secure approval, the effective date will be later than the approval date. The "Effective Date" 5/13/2020 is the date the HRRC approved your modifications and, in all cases, represents the date study activities may begin.

Because it has been granted exemption, this research is not subject to continuing review.

## UNM Institutional Review Board Approval Modification #1


Human Research Protections Program

August 2, 2020
Christine Delucas
ADelucas@salud.unm.edu

Dear Christine Delucas:

On 8/2/2020, the HRRC reviewed the following submission:

| | |
|---|---|
| Type of Review: | Modification / Update |
| Title of Study: | Identification of variables that impact HbA1c |
| Investigator: | Christine Delucas |
| Study ID: | 20-254 |
| Submission ID: | MOD00012166 |
| IND, IDE, or HDE: | None |
| Submission Summary: | Modification / Update #1 for Study 20-254 to revise Protocol. |
| Documents Approved: | • A1c Variables Protocol Modification w Track Changes 7 20 2020 |
| Review Category: | EXEMPTION: Categories (4) Secondary research on data or specimens (no consent required) |
| Determinations/Waivers: | Informed Consent Not Applicable.<br>HIPAA Authorization Addendum Not Applicable.<br>Re-consent is not required. |
| Submission Approval Date: | 8/2/2020 |
| Approval End Date: | None |
| Effective Date: | 8/2/2020 |

The HRRC approved the study from 8/2/2020 to inclusive. If modifications were required to secure approval, the effective date will be later than the approval date. The "Effective Date" 8/2/2020 is the date the HRRC approved your modifications and, in all cases, represents the date study activities may begin.

Because it has been granted exemption, this research is not subject to continuing review.

**If the study meets the definition of an NIH Clinical Trial, the study must be registered in the ClinicalTrials.gov database. Additionally, the approved consent document(s) must be uploaded to the ClinicalTrials.gov database.**

## UNM Institutional Review Board Approval Modification #2


Human Research Protections Program

December 16, 2020
Christine Delucas
ADelucas@salud.unm.edu

Dear Christine Delucas:

On 12/15/2020, the HRRC reviewed the following submission:

| | |
|---|---|
| Type of Review: | Modification / Update |
| Title of Study: | Identification of variables that impact HbA1c |
| Investigator: | Christine Delucas |
| Study ID: | 20-254 |
| Submission ID: | MOD00012924 |
| IND, IDE, or HDE: | None |
| Submission Summary: | Modification / Update #2 for Study 20-254 to add investigator Eunice Cho. |
| Review Category: | EXEMPTION: Categories (4) Secondary research on data or specimens (no consent required) |
| Determinations/Waivers: | Informed Consent Not Applicable.<br>HIPAA Authorization Addendum Not Applicable. |
| Submission Approval Date: | 12/15/2020 |
| Approval End Date: | None |
| Effective Date: | 12/15/2020 |

The HRRC approved the study from 12/15/2020 to  inclusive. If modifications were required to secure approval, the effective date will be later than the approval date. The "Effective Date" 12/15/2020 is the date the HRRC approved your modifications and, in all cases, represents the date study activities may begin.

Because it has been granted exemption, this research is not subject to continuing review.

**If the study meets the definition of an NIH Clinical Trial, the study must be registered in the ClinicalTrials.gov database. Additionally, the approved consent document(s) must be uploaded to the ClinicalTrials.gov database.**

This determination applies only to the activities described in this submission and does not apply should you make any changes to these documents. If changes are being considered these must be submitted for review in a study modification to the HRRC for