Biochimica et Biophysica Acta 1863 (2017) 1046–1053

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbadis

Methods needed to measure predictive accuracy: A study of diabetic patients☆

Hafiz M.R. Khan [a,*], Sarah Mende [a], Aamrin Rafiq [b], Kemesha Gabbidon [c], P. Hemachandra Reddy [a,d,e,f,g,h,i]

[a] Department of Public Health, Texas Tech University Health Sciences Center, 3601 4th Street, MS 9424, Lubbock, TX 79430, United States
[b] Department of Computer Science, Texas Tech University, Lubbock, TX 79409, United States
[c] Department of Health Disparities, Cancer Prevention Research Training Program, Division of Cancer Prevention & Population Sciences, UT MD Anderson Cancer Center, TX, United States
[d] Garrison Institute on Aging, Texas Tech University Health Sciences Center, 3601 4th Street, MS 9424, Lubbock, TX 79430, United States
[e] Cell Biology & Biochemistry Department, Texas Tech University Health Sciences Center, 3601 4th Street, MS 9424, Lubbock, TX 79430, United States
[f] Neuroscience & Pharmacology Department, Texas Tech University Health Sciences Center, 3601 4th Street, MS 9424, Lubbock, TX 79430, United States
[g] Neurology Department, Texas Tech University Health Sciences Center, 3601 4th Street, MS 9424, Lubbock, TX 79430, United States
[h] Speech, Language, and Hearing Sciences Department, Texas Tech University Health Sciences Center, 3601 4th Street, MS 9424, Lubbock, TX 79430, United States
[i] Garrison Institute on Aging, South West Campus, Texas Tech University Health Sciences Center, 6630 S. Quaker Suite E, MS 7495, Lubbock, TX 79413, United States

ARTICLE INFO

Article history:
Received 7 July 2016
Received in revised form 6 January 2017
Accepted 9 January 2017
Available online 11 January 2017

Keywords:
Diabetes
Bootstrapping sampling methods
Predictive errors

ABSTRACT

Diabetes is one of the leading causes of morbidity and mortality and it can result in several complications such as kidney failure, heart failure, stroke, and blindness making it a major medical and public health concern in the United States. Statistical methods are important to detect risk factors and identify the best sampling plan to determine predictive bounds for diabetic patients' data. The main objective of this paper is to identify the best fit bootstrapping sampling method and to draw the predictive bound considering diabetes patient data. A random sample was used from the National Health and Nutritional Examination Survey (NHANES) for this study. We found that there were significant relationships between age, marital status, and race/ethnicity with diabetes status ($p < 0.001$) and no relationship was observed between gender and diabetes status. We ran the logistic regression to identify the risk factors from the data. We identified that the significant risk factors are age ($p < 0.001$), total protein ($p < 0.001$), fast food ($p < 0.0339$), and direct HDL ($p < 0.001$). This study provides evidence that the parametric bootstrapping method is the best fit method compared with other methods to estimate the predictive error bounds. These findings will be of great significance for identifying the best sampling methods, which can increase the statistical accuracy of laboratory clinical research of diabetes. This will also allow for the determination of precise risk factors that will best represent the data by detecting mild and extreme outliers from disease observations. Therefore, these results will be useful for researchers and clinicians to select the best sampling methods to study diabetes and other diseases in order to maximize the accuracy of their results. This article is part of a Special Issue entitled: Oxidative Stress and Mitochondrial Quality in Diabetes/Obesity and Critical Illness Spectrum of Diseases - edited by P. Hemachandra Reddy.

1. Introduction

Diabetes is one of the most prevalent non-communicable diseases worldwide. Three types of diabetes can occur – Type 1, Type 2, and gestational. These diseases occur due to an insufficient production of the hormone insulin or due to an inability of the body to use insulin effectively. Insulin is critical for the body's cells to uptake glucose for use as energy [1]. For this reason, diabetes can be classified as a metabolic disorder, characterized by underutilization and overproduction of blood glucose. A condition also known as hyperglycemia, for which the cause of varies among each form of diabetes [2].

Diabetes can lead to many complications such as heart disease, renal failure, blindness, and stroke [3–5]. It has become a major health concern due to the increase in new diabetic patients and premature deaths due to diabetes. In 2013, half of diabetes related deaths were in patients under age 60 [6].

Diabetes mellitus, commonly referred to as diabetes, is a major public health issue; it ranks among the top 10 killers in the United States. Nationally, diabetes is responsible for 234,051 recorded deaths annually [7]. In the past three decades there has been an exorbitant increase in the prevalence of early onset of diabetes [8]. Approximately 27% of

Abbreviations: NHANES, National Health and Nutrition Examination Survey; OLS, ordinary least square; GLM, generalized linear model.

☆ This article is part of a Special Issue entitled: Oxidative Stress and Mitochondrial Quality in Diabetes/Obesity and Critical Illness Spectrum of Diseases - edited by P. Hemachandra Reddy.

* Corresponding author.
E-mail address: hafiz.khan@ttuhsc.edu (H.M.R. Khan).

http://dx.doi.org/10.1016/j.bbadis.2017.01.007
0925-4439/

non-diabetics may have undiagnosed diabetes, which could have significant implications in diabetes research [9]. Incidence rates of diabetes are on the rise throughout developed and developing countries. According to the World Health Organization (WHO), 346 million people worldwide have diabetes and it is expected to increase by 51% by the year 2030 [10,11].

Diabetes is a group of metabolic diseases characterized by a state of hyperglycemia resulting from a defect in insulin secretion, insulin action or both. Diabetes induced hyperglycemia is associated with a myriad of adverse health effects such as: neuropathy, retinopathy, nephropathy, heart and blood vessel damage, and if left untreated organ failure is imminent [12]. The etiology of diabetes mellitus is multifactorial, and is classified into subgroups including gestational, Type 1, and Type 2 diabetes. Gestational diabetes is another subclass, which takes place during pregnancy, and begins when the body is unable to create and use all the insulin that is needed for pregnancy. According to a 2014 analysis by the Centers of Disease Control and Prevention (CDC), the prevalence of gestational diabetes is as high as 9.2% for pregnant women [13]. Type 1 occurs when there is a complete lack of insulin (β-cell destruction), characterized by the pancreas' inability to produce a sufficient insulin response. Type 2 refers to insulin resistance, where the body is unable to respond to the insulin that is created and loses the ability to normalize glucose levels in the blood [14].

This study will focus on application of statistical methods that can be used in research of Type 2 diabetes. The insulin resistance that characterizes Type 2 diabetes has been associated with pancreatic β-cell function, particularly β-cell dedifferentiation [15]. The dedifferentiation of β-cells can explain many aspects particular to this form of diabetes, especially its apparent reversibility that are often possible with lifestyle changes. Risk factors associated with diabetes that are related to lifestyles choices can often be modified, such as, obesity, high sedentary behaviors, and lack of physical activity. There are also other risk factors that are not modifiable including, family history, ethnicity, and age [13]. Diabetes also increases the risk for comorbidities, such as cardiovascular disease, hypertension, and kidney failure [16].

The following are some possible risk factors related to diabetes which are described in details.

1.1. Age and marital status

Age is an important factor in the development of diabetes and though there has been an increase in the prevalence of diabetes among adolescents, research over the last 20 years indicates that some risk factors increase with age. A pooled analysis that studied the effects of age on various risk factors related to the development of diabetes found that the effect of body mass index (BMI) on diabetes increased with age, especially among western cultures [17]. However, the effects of age and the development of diabetes and other comorbidities can be slowed by other factors, such as marital status. A separate study found that for married patients already diagnosed with diabetes their quality of life declined at a much slower rate than unmarried diabetes patients [18]. Though no causal relationship has been established with these risk factors and diabetes, their effect on diabetes and other associated risk factors could provide information on how to prevent or slow the progression of the disease.

1.2. Fast food consumption and obesity

An inverse relationship exists between BMI and the age of diabetes onset, showing severe weight gain associated with a higher risk of diabetes, regardless of age [19,20]. The high rates of fast food consumption in the United States are a major contributing factor to diabetes and are associated with many risk factors of diabetes such as high BMI, obesity, and large waist circumference. This is due to the high sugar and fat content found in these foods. A diet high in sugar and fat could lead to more cases of early-onset diabetes. The vast majority of those with early onset diabetes are obese (80–92%), and it is a primary determinant of Type 2 diabetes [21–23]. Obesity is often associated with increased secretion of proinflammatory adipokines and adipose tissue inflammation. The proinflammatory adipocyte fatty acid-binding protein has been shown to play a role in Type 2 diabetes. This adipokine acts as a cytosolic lipid chaperone that can reversibly bind with high affinity to hydrophobic ligands and contributes to many of the lipid-mediated processes associated with Type 2 diabetes [23]. Waist circumference is also associated with fast-food intake and obesity, and is included with low HDL levels as components of the metabolic syndrome. The metabolic syndrome is a cluster of conditions that increase risk of life-threatening diseases such as heart attack, stroke, and diabetes. Understanding causal relationships between these conditions and diabetes is important to study, especially in the pursuit of effective prevention techniques [21]. Waist circumference is an effective way to measure the amount of visceral adipose tissue, and has been shown to have a causal relationship with diabetes because of its effect on systemic inflammation [24]. An increase in visceral fat causes an imbalance of adipokines which results in a pro-inflammatory state. Local and systemic insulin resistance can be caused by the activation of inflammatory pathways due to drainage into portal circulation [24]. This insulin resistance can then lead to diabetes and other comorbid conditions.

1.3. Race/ethnicity

In the United States the burden of diabetes falls disproportionately on racial/ethnic minorities; many ethnic minority groups are more likely than whites to be diagnosed with diabetes. Related morbidity and mortality rates vary by complication and by race [25]. American Indians/Alaska Natives are more than two times likely to be diagnosed with diabetes than non- Hispanic whites, while non-Hispanic blacks and Hispanics have rates 70% to 80% higher than non-Hispanics whites [26]. Asian Americans are at a higher risk of developing type 2 diabetes; despite lower average body mass index compared with non-Hispanic whites [27]. Another important difference across ethnicities is related to diabetes identification and prevention: the blood level of glycolated hemoglobin (HbA1c). On the β-chain of a hemoglobin molecule, carbohydrates can be added to the NH-2 terminal through a slow enzymatic process. The most common addition is that of glucose, creating HbA1c [28]. These hemoglobin levels have been used to measure blood glycemia levels over extended periods of time [29]. A cohort study revealed that African Americans, Hispanics, Native Americans, and Asians had higher HbA1c levels than whites [28]. These findings could not only affect diagnosis of disease, but can influence specific population-targeted prevention strategies as well.

1.4. HDL and protein levels

Conditions consistent with diabetics include low HDL (high density lipoprotein) blood levels, and high triglyceride levels. This collection of symptoms is called diabetic dyslipidemia, which contributes to atherosclerosis [30]. Therefore, these symptoms are not only important for the study of diabetes, but can be useful in the evaluation of other associated diseases. Recent studies have focused HDL and the increased incidence rates in various ethnic populations and age groups [31]. Several predictors that play a role in the pathophysiology of Type 2 diabetes have been analyzed, and have shown an inverse relationship between HDL and Type 2 diabetes. Results of such studies have postulated that HDL may also be directly involved in the pathogenesis of type 2 diabetes mellitus by virtue of its ability to enhance pancreatic β-cell function and glucose uptake in skeletal muscle [32]. Other studies show a possibility of reverse causation, attributing low HDL levels as a consequence of insulin resistance [30]. Diabetic dyslipidemia is caused because under conditions of insulin resistance, there is reduced hormone sensitive lipase (HSL) activity, which further reduces the catabolism of triglyceride-rich very low density lipoprotein (VLDL-TG). The raised concentration