International Journal of Biomolecules and Biomedicine (IJBB)
ISSN: 2221-1063 (Print), 2222-503X (Online)
http://www.innspub.net
Vol. 8, No. 1, p. 1-9, 2018

INNSPUB

REVIEW PAPER OPEN ACCESS

A review of predictive analysis techniques of diabetes prevalence

Nicholas Musau[1], Josiah Ochieng Kuja[*2]

[1]*School of Computing and Information Technology, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya*

[2]*School of Biological Sciences Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya*

Key words: IBM watson analytics, Diabetes, Prevalence.

Article Published: 03 December 2018

Abstract

Diabetes is a significant cause of mortality and morbidity in different continents of the world. Many diabetes victims are found in developing countries like Sub-Saharan Africa. However, some developed nations like United States and Europe record significant records on diabetes prevalence. Studies project a dramatic increase of the infection spread in the world. Also, it provides visible results on the effects of the infection among the victims and the society at large. Studies of type 2 diabetes prevalence indicate minimal rates in rural population and moderate results in the developed regions of the same country. Such results create an alarm to the unaffected regions. The frequent observation of modestly high prevalence of impaired glucose tolerance in areas with low prevalence of diabetes indicate risk of early stage of diabetes epidemics.

***Corresponding Author:** Josiah Ochieng Kuja ✉ kujahjos@gmail.com

Introduction

Researched data on global epidemiology of diabetes projection indicate that the greatest increase in prevalence and the numbers of affected people happens in developing nations. However, the prevalence rates are expected to shoot by 48% by the year 2020. The upgrade is expected to rise and increase the number of diabetes individuals in the developing countries. A projection done in Sub-Saharan Africa showed 18% increase of diabetes prevalence rate in the entire continent. Predictions of diabetes impact on Sub Saharan Africa ought to take into consideration the ruin brought by HIV epidemic. Currently, Africa continent accounts for about three-quarters of the world's HIV infection. Present findings of the moderate to high prevalence of reduced glucose tolerance in nations with low diabetes prevalence might signpost early stages of diabetes epidemic.

Nevertheless, the prediction techniques used in different types of research present incorrect statistics of the diabetes prevalence in various locations. Currently, many researchers use some old softwares for data analysis and prediction of the infections prediction. This paper ought to introduce new technological techniques which are useful in data analytics and infection prevalence prediction. The present world entails a lot of data growing rapidly from different sources. The invention of electronic health record systems led to establishment of electronic health databases that host huge amounts of data. The traditional statistical analysis tools such as R, SPSS, STATA and EXCELL could not analyses the huge data fast and give immediate and effective results. As a result, IBM Watson Analytics tool is essential in analyzing and predicting infections' prevalence in certain regions.

Causes of high diabetes preference and the general trend

The preference is cultivated by risk factors such as age, urbanization, and family history of infection. Other contributors include failure to participate in physical activities and adiposity. According to (Alotaibi *et al.*, 2017), type 1 diabetes prevalence data indicate low rates in Sub-Saharan Africa. Deaths resulting from diabetes cases are unacceptably high. Additionally, the mortality rates are disproportionally because of infective and preventable acute metabolic causes. Extensive duration of disease infection increase the frequency of hypertension as well as microvascular complexities. Many institutions work towards managing the diabetes issue, but their efforts are limited by the inadequate clinical infrastructure. Further, limited medical supply and educational programs inhibit proper treatment of diabetes.

Hall *et al.*, (2011) observed that the highest prevalence rate of diabetes is expected to happen in North Africa and Middle East. The predictions are based on the rapid economic development, changes in lifestyle and urbanization within the regions. This global epidemic is expected to affect the Kingdom of Saudi Arabia. A report released by the ministry of health in Saudi Arabia indicated an estimate of 0.9 million citizens diagnosed with diabetes infection around 1992. The figure increased to 2.5 million in 2010. Moreover, around 4660 diabetic patients attended medical clinics and family clinics across Saudi Arabia. Diabetes cases in Saudi Arabia are associated with increased obesity and high number of aging generation. Diabetic patients experience other chronic infections that worsen individual health condition. For instance, the end stage of renal infection is high among diabetic victims. The infection accounts for 24% and 51% of the individuals getting renal replacement therapy (Hall *et al.*, 2011).

Statistically, diabetes victims have a two of four chances higher to develop cardiovascular disease (Murray & Lopez, 1997). The patients are also 2 to 5 times more likely to pass on from the infection (Mozaffarian *et al.*, 2015). Besides, it affects individual's lives. Diabetes infection places an important burden to the society and the healthcare service. Globally, diabetes infection account for around 11% out of the total healthcare expenditure (WHO, 2016). It is imperative to understand the epidemiology of diabetes to identify public health priorities. Further, understanding of the diabetes epidemiology can help in generating policy initiatives

and evaluation of the effects of the services in reducing individual and societal burden of diabetes. Despite the fact that prevalence estimates are offered by International Diabetes Federation, there exist substantial differences in periodical trends because the estimates are grounded on imputation.

Diabetes cases in various continents

According to data released by Center for Disease Control and Prevention (CDC) in the U.S, more than 29 million adults in the State are diabetes victims (Centers for Disease Control and Prevention, 2014). Around 25% of the diabetic victims lack the awareness of personal infection (Naicker & Ashuntantang, 2017). United States comprise of 86 million people living with prediabetes which is a serious health issue that increases an individual's risk of chronic infections such as type 2 diabetes (Naicker & Ashuntantang, 2017). CDC is researching towards reversing the U.S diabetes epidemic by tracking the infection trends. The organization is focusing on prevention, identification of treatment mechanisms and improvement of medical care in the State. Diabetes victims fail to produce sufficient insulin and also fail to use insulin appropriately. Insulin facilitates entry of blood sugar into the cells to be used as energy (Kalwat *et al*, 2016). When the body lacks sufficient insulin, it builds up within the blood. Such increased blood levels lead to stroke, heart disease, blindness, amputation of toes and kidney failure.

A study done by Merriam *et al.* (2009), in the U.S showed that Latinos represent the minority group in the U.S and their presence represents 13.7% of the entire population. Nevertheless, CDC analyzed data from Behavioral Risk Factor Surveillance System (BRFSS) and discovered that Hispanic people have high records of prevalence of diabetes. Among the 13.7%, 7.4% Hispanics victims understand they have diabetes (Merriam *et al*, 2009). The increasing prevalence of obesity and sedentary lifestyle predicts an increase in the number of people with type 2 diabetes mellitus (Merriam *et al*, 2009). The results show a heavy burden to the Latino community because of the adverse effects of diabetes disease. It is evident that Latinos stand a high chance of diabetes attack in the United States when compared to average Americans. As a result, prevention mechanisms of type 2 diabetes in Latino communities is a significant priority.

A randomized clinical trial approach called Diabetes Prevention Program (DPP) demonstrated a modest weight loss as an essential approach of reducing diabetes infections (Merriam *et al.*, 2009). The program also indicated increased physical activities as better mechanisms to manage incidences of diabetes among the prediabetes victims. Nevertheless, the recruitment and intervention methodologies were expensive. Other diabetes management programs like Lawrence Latino Diabetes Prevention Project (LLDPP) targets to designing and testing of a less intensive intervention that targets weight loss via physical activity and dietary change. Such approaches will reduce the risk of type 2 diabetes among the Latino community. LLDPP research approach was designed to reduce the extra cost of screening and recruitment activities experienced in the DPP methodology. LLDPP was an accurate and less costly screening procedure grounded on predictive equation (Merriam *et al.*, 2009).

Sinnott *et al.*, (2017) observed that diabetes mellitus (DM) is a global epidemic in the current society. World Health Organization (2016) survey discovered that the apparent epidemic of diabetes is highly contributed by economic changes and lifestyle. Many people are at risk of developing complications. Diabetes type 2 is a serious disease signified by chronic hyperglycemia (Sinnott *et al.*, 2017). Most of hyperglycemia cases are associated with assortment of environmental, population explosion, genetic factors, and increased geriatric population. Professionals lack predictive approaches to use to produce quality intervention methodologies for this colossal health issue. However, nurses use Evidence-Based Practices to encourage sensible healthcare approaches among the victims to manage the progression risk of the infection.

Additionally, diabetes patients are requested to reduce the impact of diabetes by use of Impaired Fasting Glucose.

The management approach provides extensive preventive benefits even when the victim halts the counselling sessions. Studies record that communities with deprived socioeconomic positions have high chances of increased diabetic prevalence. Women have a greater incidence rates of mortality resulting from type 2 diabetes. Systematic review and meta-analysis indicated that augmented risk of type 2 diabetes mellitus was linked to low education level, income and occupation (Bharati *et al.*, 2011).

A South Indian research by Bharati *et al.* (2011) showed that Oral Glucose Tolerance Test is the best test for diagnosing type 2 diabetes mellitus. However, the approach is inconvenient, expensive and with weak reproducibility. Population-level management of type 2 diabetes mellitus should lie in the risk groups to thwart macroangiopathic changes. A Canadian study by Bharati *et al.* (2011) revealed that sex-adjusted diabetes and age-adjusted prevalence increased by 69% from 5.2% in 1994 to 8.8% in 2005. Moreover, the prevalence increased from 6.9% to 8.8% between 2000 and 2005 indicating an increase rate of 27% in a period of five years. Diabetes infection showed high prevalence among people aged 50 and above.

Methodology in various continents

A study in the U.S by Merriam *et al.* (2009) entailed a research which involved a primary recruitment outreach approach. The method drew information from GLFHC database. Patient panes through identification of eligible patients who received an email inviting them for study examination from community coordinators. The experts generated a mailing list every 6-8 weeks by sending a screening query of the present GLFHC database. The methodology targeted identifying potentially eligible criteria. Another query was executed to eliminate patients who had participated in the study. This time around an approximate of 250 patients were selected randomly for individual mailing. Experts designed personalized patient screening invitation letters and signed by patients' primary care physician (PCP) then mailed to individuals.

Experts eliminated patients by their PCP if they appeared ineligible. Primary Care Physicians engaged in regular updates through communications from the community-PI and provider meetings. The selected GLFHC patient records were downloaded into a Lotus Notes tracking database and distributed to community coordinators for frequent telephone outreach. Other outreach approach entailed public service announcement on public access on television. Advertisement and guest spots also enabled the research experts to reach many individuals during the survey. Bilingual newspapers and Senior Center newsletter were also used to create awareness of the study among the public.

Methodology in the U.S

United States records significant number of diabetes cases in the world. As Menke *et al.* (2015) notes, a series of stratified and multistage probability surveys called NHANES were designed to represent the U.S noninstitutionalized population and civilians. Researchers used NHANES III to collect data over 11 years between 1988 and 1999. Professionals collected data through in-home interviews and physical visits to mobile examination centers. Data from 2011- 2012 was used to estimate the recent preferences of diabetes trends. In this case, the response rate ranged from 70% to 80% for examination and 73% to 86% for interview processes. Participants were selected randomly to participate in examination processes. Pregnant women were excluded in the survey because pregnancy influences glucose measurements. All participants offered written consent besides research ethics according to the National Center for Health Statistics (NCHS) which approved all the protocols.

NHANES used a standardized questionnaire to collect individual information like age, race, gender, income level and education level. Family income level was classified into regions. During the assessment, height and weight were measured and body mass index (BMI) calculated as weight (Kgs) divided by height (meters squared). Participants were asked if they had diagnosed with diabetes before. Such questions helped to understand family and individual history. A phlebotomist was used to obtain blood sample from individual participants with regard to standardized protocols.

Hemoglobin A1c was measured and the interassay coefficient of variation was 0.7% - 3/1%. Moreover, fasting glucose was assessed and the variation ranged from 0.8% to 3.7% (Hennink *et al*., (2017).

Methodologies in Europe

A research done by Place & Lane (2000), indicated that many people in Europe, especially in North Liverpool amid year population of around 176, 682 were diabetes victims. The research targeted individuals with diabetes during a period of one year. Diabetes was diagnosed clinically using the WHO approach. Researchers obtained diabetes patient data from general practices around the area. Researchers used patients' data from Diabetes Centers and hospitals for statistical predictions. Moreover, discharge data offered essential information. Patients' records attending hospitals and Retinal Clinics was used to evaluate diabetes prevalence in Europe. Besides the mentioned methodologies, researchers researched about the list of stroke in-patients diagnosed with diabetes. During the study, the list of children with acute diabetes cases attending children hospital provided a good source of information. Most of the patients' health records entailed patient name, postcode, gender and date of birth.

Methodologies in South Asia

Hennink *et al*. (2017) carried out a literature review on population-based and cross-sectional study and observed that there existed differing results in urban and rural regions of South India. A medical field study in Mahatma Gandhi Medical College and Research Institute in Puducherry showed significant results of predictive diabetes data research. Professionals used simple random sampling techniques to select individuals aged around 20 years and above. The objective in the study was to discover the prevalence of diabetes and correlations of diabetes among patients. Further predesigned and pretested questionnaires were also valued in eliciting family information and individual sociodemographic variables. Research professionals identified height, waist, weight and hip circumference as valuable details in examining diabetes prevalence among patients. The details were used to measure the level of fasting glucose and blood cholesterol.

Prevalence rates of diabetes in urban and rural areas of South India were calculated by experts at WHO to use on prediction activities in national level. The prevalence revealed that diabetes mellitus had 5% error and 20% absolute error occurred within a calculated range of 1370 sample size. The sampling was extracted by simple random sampling approach from a population in the field practice area. Data collection was also done through interview schedule developed at the research institute with the help of faculty members and other professionals.

The predesigned questionnaires contained questions related to family information characteristics. Critical details of the questionnaire included family history of diabetes mellitus, area of residence, family type, and family history of chronic infections, personal details like age, gender, and education. Other significant details were occupation, type of food and dietary habits. Health workers were informed about the diabetes prevalence survey and motivated the family members to participate in the survey. All the participants were explained on the significance of the survey and were ensured of strict confidentiality. Researchers obtained an informed consent from the families and also gave the participants a chance to fail to respond to the questionnaires. Interestingly, all families responded to the questionnaire.

Methodology in sub-Saharan Africa

Literature review of previous studies provided systematic report of the diabetes prevalence in the Sub-Saharan Africa. Sub-Saharan Africa entailed all the mainland of island nations of Madagascar. Hall *et al*. (2011) reports that articles providing diabetes data around the region were used to collect information regarding diabetes prevalence in the region. The researched papers recorded information about diabetes prevalence, infection, access to diabetes diagnosis care and economic burden resulting from diabetes infection. Experts run a combined keyword search on the PubMed database and identified 1102 papers. A review of other articles were extracted from the website of World Health Organization, World Bank and International Diabetes Federation.

Discussion

The prevalence of diabetes around the world accounts for a greater percentage. Diabetes disease is associated with other infections such as gestational diabetes and typical ketosis diabetes. Malnutrition also results from diabetes infections. The results obtained in various countries can be used to show low medium prevalence. Huge amounts of data collected is entered into a computer and analyzed using statistical software. Such datasets indicates log-likelihood above 3.84. The total diabetic population can be calculated by summing the numbers of the missing cases to the aggregated cases. Statistical software used in this case, divide the number of cases by the total population within the group.

Many people lack knowledge about natural history and clinical course of diabetes in Africa. Mortality rates record high in developing countries. There is need for control and management of diabetes prevalence mortality figures ranging from 7.6 % to 11.8%. Records on known diabetes patients continue to increase to around 41.5% and the new cases rise to 15.7%. High proportions of deaths resulting from renal causes in various countries.

The current world era seats in a world of explosive data generation. The data is foreseen to continue growing and involve many industries. The data exploration need use of newer data collection tools and efficient data analytic tools and procedures. Initially, data analysis required traditional data analytics knowledge and substantial background in essential subjects like computer science and statistics. In 2015 IBM introduced the IBM Watson Analytics (IBMWA) software. The software came with advanced capabilities of delivering advanced statistical procedures for Big Data. The application uses Statistical Packages for the Social Science (SPSS). The new release of Watson Analytics provides users with advanced functions absent in the existing programs like SPSS, R, STATA, and EXCEL.

IBM Watson Analytics analyzes datasets automatically and examines the data quality. Additionally, it determines an optimal statistical approach.

The availability of IBM Watson Analytics allows a user to request exploratory, visual analytics and predictive results from a certain dataset. Users can use natural language to ask questions related to their objectives because the tool uses Natural Language Processing technology (NLP). Most of the information generated by diabetes researchers is unstructured. As a result data analysts spend time coding the data and cleaning it before analyzing. Such preparations lead to time consumption. Alternatively, IBM Watson Analytics is designed to handle structured and unstructured data. It can load mass of data (Big Data) in seconds of importing from a local or clouds source. IBM Watson has the capability to give visualizations and predictions immediately the data is loaded. The IBM tool gives an opportunity to discover and visualize diabetes patterns derived from the data.

IBM Watson Analytics tool uses natural language. As a result, researchers with limited statistical knowledge can use it for prediction activities. Moreover, users need not be experts in data science to start using the tool. The powerful feature in this tool is that users can ask questions using natural language. Such capabilities enable many diabetic researchers to generate more specific results from data. Diabetes can now be addressed in a deeper manner with the use of IBMWA than while using previous tools.

Razavian *et al.* (2015) noted that, the tool presents a new diabetes prediction methodology in the world. In this approach, data-driven predictive models are used for the results of type 1 and 2 diabetes. The approach facilitates risk assessment from electronic data of large populations with the absence of additional screening costs. The proposed approach reveals early and late-stage risk factors in diabetes control. IBM Watson is able to methodically improve predictive variables sets and fit other models predicting the onset of diabetes type 2. The developers used machine learning to methodically enhance variable set and fit models for predicting the onset of type 2 diabetes. The enhanced methodology can increase model of the AUC curve to 0.80 with around 900 variables selected for prediction. Research officers can observe improvements for the new model in predicting diabetes onset on 1-3 and 2-4 years after the baseline.

The improved model enhances positive predictive value by 50% and identifies novel surrogate risk factors for type 2 diabetes. IBM Watson Analytics identifies others infections such as chronic liver disease, high alanine, esophageal reflux and a history of acute bronchitis. The model could help in predicting many other infections besides diabetes. Liver risk factors developed later in the process of diabetes development compared with obesity-related factors like high hemoglobin A1c and hypertension. The population-level risk prediction for types 1 and 2 diabetes appeared readily available while using the administrative data. The new model showed better prediction performance than the traditional prediction models used in research centers. The new model allows intervention allocation at the national scale easily and accurately. Additionally, it recovers potentially novel risk factors at various stages before disease onset.

The recent growing need for electronic health records has enable clinical officers to collect large amounts of datasets in healthcare databases. Such an opportunity offers an unprecedented occasion to apply predictive analytics to enhance medical practice and manage potentially novel risk factors. The model has deployed large-scale risk assessment models like hospital readmission models. Moreover, it provides a prediction of healthcare utilization and costs. Traditional well-known models for prediction of diabetes onset according to Hennink *et al.* (2017) include ARIC, San-Antonio, FINDRISC, and AUSDRISK. Nevertheless, the new model provides potential solutions for more accurate risk assessment of clinical infections. The model is associated with a time-consuming and high management cost. The improvement of health condition using new IBM Watson Analytic too helps to improve diabetes cases prediction mechanisms. Such methodologies will create an opportunity for the development of diabetes interventions.

The advancement of technology in research and data analysis help to improve the clinical research and deliver quality health results. Chong (2018), defines the Dexcom tool for monitoring the rate of diabetes infections and the number of sugars in the body. The results are used to monitor and evaluate the level of blood sugars in a diabetic patient. The assessment uses the upper and the lower limits in analyzing sugar levels. A diabetic victim must respond to the alarm by consuming additional carbohydrates of administering insulin. The Dexcom tool for data collection offers rate-of-change limits that provide rudimentary proactive capabilities. The information gathered by the tool can be used to proactively administer insulin before a patient hits the upper limit. The IBM Watson Analytics is able to weed out the various items of concern such as the ability to collect data. Other factors of considerations include data collection, analytical techniques, and the type of data analyzed. The IBM's tool provides promising results that could be used for the continued research and improvement of health condition among patients.

Conclusions and Recommendations

Data scientists find much application of the predictive analytics in the healthcare sector. The emergence of data collection and analytics tools like Dexcom and IBM Watson Analytics respectively can help in managing patients' condition by providing quality predictive results to researchers. The new analytic model shows the importance of Big Data and predictive tools used in the evaluation of data. Data science knowledge and predictive analytics can be utilized for more optimized datasets or even prediction result to proper solution establishment on diabetes prevalence and spread. They can be used to impact patients' lives on a daily basis. It is essential to apply the new analytic tools to assess the prevalence of diabetes in a patient. The utilization of the IBM Watson Analytics can help in predicting the future occurrence of diabetes. Such tools can help improve health condition and statistics among diabetes patients.

According to (LIEBERT, 2018), Big Data is transforming the healthcare industry especially with electronic recording and evaluation of diabetes prevalence. The new focus on Big Data has led to the development of innovative methodologies for predicting the prevalence of diabetes in different places. Many researchers have managed to monitor, record, analyze and integrate information related to human health into clinical evaluation processes.

The current population of patients generates a lot of data that is stored in hospital databases. Masses of data require powerful analytical tools to make proper decisions and predictions. The current data-driven approach is used to obtain health population health. The new technology uses machine learning to develop risk factors and predictive models for the occurrence of type 2 diabetes. IBM Watson Analytics uses machine learning technology and is able to interpret electronic health data.

The model is capable of analyzing data collected from a pharmacy, healthcare records and laboratory results of more than 4.1 million patients. The new Big Data Analytics model identifies risk factors for both types 1 and 2 diabetes. The tool is better for predicting disease onset in various regions. The availability of electronic health records offers an opportunity to use predictive analytics to improve the practice of medicine and to infer potential risk factors. The model has succeeded in various areas of health care such as hospital readmission and the prediction of healthcare costs. Victims of type 2 diabetes continue to increase and are estimated to reach 366 million by 2030 (Harvey, *et al.*, 2002). Around 15 current statistics indicate that many diabetic patients suffer from type 2 diabetes. Diabetes Prevention Program (DPP) revealed that intensive lifestyle intervention involving exercise and weight loss was effective in reducing type 2 diabetes.

Interventions showed that patients could experience cost-effective benefits when the targeted population records are analyzed well. During diabetes assessment, the DPP program selected participants on basis of obesity and elevated glucose levels. The study discovered that only 11% of the surveyed participants without lifestyle developed diabetes within three years. Such results emphasized the need for better models with enhanced risk assessment for diabetes prevalence. Traditionally, researchers used various models to assess the prevalence of type 2 diabetes. Some traditional models used include ARIC, AUSDRISK, San-Antonio, and FINDSISC (DeFronzo, *et al.*, 2013). Such models offered good solutions for accurate risk assessment. However, they require much time and were costly in the screening phase. Machine learning helps to discover a surrogate for variables that could be missing (Razavian, *et al.*, 2015).

The use of IBM Watson Analytics tool will be a perfect approach for analysis and prediction of the prevalence of diabetes in different parts of the world. IBM Watson does not need a deep understanding of data science or even analytics skills because it uses natural language program (NLP). It is possible to ask a question from a big dataset attached in IBM Watson tool. The tool embeds NLP technology into the analytic platform. It is a cloud-based platform that could be used anywhere in the world. What is required is the data to be analyzed and a simple excel sheet to start. Since no other systematic reviews that have supported the use of IBM Watson Analytics, it is essential that all research institutions employ the technology to enhance diabetes prediction techniques. Further, the technology will provide better techniques for risk assessment and management of diabetes occurrence from massive sets of datasets.

References

Alotaibi A, Perry L, Gholizadeh L, Al-Ganmi A. 2017. Incidence and prevalence rates of diabetes mellitus in Saudi Arabia: An overview. Journal of epidemiology and global health.

Bharati DR, Pal R, Kar S, Rekha R, Yamuna TV, Basu M. 2011. Prevalence and determinants of diabetes mellitus in Puducherry, South India. Journal of pharmacy & bioallied sciences **34,** 513.

Centers for Disease Control and Prevention. 2014. National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014. Atlanta, GA: US Department of Health and Human Services, 2014.

Chong R. 2018. Diabetes and Data Science: Applying Predictive Analytics to Health. Retrieved from https://clockwork-solutions.com/diabetes-data-science

De Fronzo RA, Tripathy D, Schwenke DC, Banerji M, Bray GA, Buchanan TA, Ratner RE. 2013. Prediction of diabetes based on baseline metabolic characteristics in individuals at high risk. Diabetes Care **36-11,** 3607-3612.

Hall V, Thomsen RW, Henriksen O, Lohse N. 2011. Diabetes in Sub Saharan Africa 1999-2011: epidemiology and public health implications. A systematic review. BMC public health **11-1,** 564.

Harvey JN, Craney L, Kelly D. 2002. Estimation of the prevalence of diagnosed diabetes from primary care and secondary care source data: comparison of record linkage with capture-recapture analysis. Journal of Epidemiology & Community Health **56-1,** 18-23.

Hennink MM, Kaiser BN, Sekar S, Griswold EP, Ali MK. 2017. How are qualitative methods used in diabetes research? A 30-year systematic review. Global public health **12-2,** 200-219.

Kalwat MA, Wichaidit C, Nava Garcia AY, McCoy MK, McGlynn K, Hwang IH, Cobb MH. 2016. Insulin promoter-driven Gaussia luciferase-based insulin secretion biosensor assay for discovery of β-cell glucose-sensing pathways. ACS sensors **1-10,** 1208-1212.

Liebert, M. 2018. Big data is transforming healthcare - from diabetes to the ER to research. www.eurekalert.org/pub_releases/2016-02/mali-bdi020216.php

Menke A, Casagrande S, Geiss L, Cowie CC. 2015. Prevalence of and trends in diabetes among adults in the United States, 1988-2012. Jama **314-10,** 1021-1029.

Merriam PA, Tellez TL, Rosal MC, Olendzki BC, Ma Y, Pagoto SL, Ockene IS. 2009. Methodology of a diabetes prevention translational research project utilizing a community-academic partnership for implementation in an underserved Latino community. BMC medical research methodology **9-1,** 20.

Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Howard VJ. 2015. Heart disease and stroke statistics—2016 update: a report from the American Heart Association. Circulation, CIR-0000000000000350.

Murray CJ, Lopez AD. 1997. Mortality by cause for eight regions of the world: Global Burden of Disease Study. The lancet **349-9061,** 1269-1276.

Naicker S, Ashuntantang G. 2017. End stage renal disease in Sub-Saharan Africa. In Chronic Kidney Disease in Disadvantaged Populations pp. 125-137.

Rao CR, Kamath VG, Shetty A, Kamath A. 2010. A study on the prevalence of type 2 diabetes in coastal Karnataka. International journal of diabetes in developing countries **30(2),** 80.

Razavian N, Blecker S, Schmidt AM, Smith-McLallen A, Nigam S, Sontag D. 2015. Population-level prediction of type 2 diabetes from claims data and analysis of risk factors. Big Data **3-4,** 277-287.

Sinnott SJ, McHugh S, Whelton H, Layte R, Barron S, Kearney PM. 2017. Estimating the prevalence and incidence of type 2 diabetes using population level pharmacy claims data: a cross-sectional study. BMJ Open Diabetes Research and Care **5-1,** e000288.

World Health Organization. 2016. Global report on diabetes. World Health Organization.