Journal of Health Informatics in Africa

Prevalence and Complications Associated with Diabetes Mellitus at the Nairobi Hospital, Nairobi City County, Kenya

Amos Olwendo [1*], George Otieno[1], Kenneth Rucha[1]

Department of Health Management & Informatics, Kenyatta University, P.O. Box. 43884, 00100, Nairobi, Kenya.

Background and Purpose: Diabetes mellitus (DM) is a lifestyle disease and a global health challenge. About 14.2 million people in Africa had the disease in 2015. Kenya is presently experiencing an increase in mortality and morbidity related to diabetes.
Methods: This research employed a retrospective cross-sectional study design that sampled records of confirmed cases of diabetes mellitus collected during routine care between January 2012 and December 2016 at the Nairobi Hospital located in Nairobi city, Kenya. A stratified sample of 652 records of male and female patients were retrieved from the EHR database and analyzed in this research. The dataset was subjected to pre-processing; that involved handling cases of missing values, smoothing for the removal of noise, identification and removal of outliers, and resolving cases of inconsistencies. Data were normalized using the z-score standardization and analyzed based on dimensions of EHR data quality and through cluster analysis using Density-Based Spatial Clustering of Applications with Noise (DBSCAN).
Results: The prevalence of T2DM is at 92% and the most common complications of diabetes include; retinopathy (12%), neuropathy (11%), and cardiovascular (11%). Hypertension was present in 39% of cases of diabetes.
Conclusion: Diabetes is increasingly becoming a health problem in Kenya thus there is need for increased public awareness of the dangers of diabetes mellitus. Members of the public need to sensitized on the usefulness of physical exercise and dietary requirements to slow the development and progression of diabetes. Also, there is need for understanding the causal relationship between T1DM and T2DM and hypertension.

Keywords: Diabetes Mellitus; Complications; Computational Phenotyping; Density-based Clustering; DBSCAN;

1 Introduction

Diabetes refers to a group of diseases that affect how the body uses glucose. Glucose is vital to health because it is an important source of energy for the cells that make up muscles and tissues. DM is a global health challenge and about 14.2 million people in Africa had the disease in 2015. The number of DM cases are expected to rise in Africa especially in countries transitioning from low to middle income economies such as Kenya. A number of DM cases stay undiagnosed for long in Kenya and the prevalence of DM among Kenyans aged 27-79 was 2.2% (approximately 484,000 persons) in 2015 (Mohamed et al., 2018; Mwangi et al., 2017). The principal cause of diabetes varies by type. Nevertheless, all types of diabetes can lead to excess sugar in the blood. Diabetes mellitus is a leading cause of mortality and morbidity that is characterized by insulin deficiency. Diabetes can be classified as type 1, 2, 3, and 4. Type 1 diabetes mellitus (T1DM), characterized by insulin deficiency and develops at any point during an individual's lifetime, is currently on the rise worldwide. On the other hand, T2DM affects millions of people worldwide. T2DM makes about 90% of all diabetes cases in Kenya and is known to develop in a person at about age 40 (Mwangi et al., 2017). Other forms of diabetes include gestational, chemical-induced. Chronic diabetes conditions include T1DM and T2DM. Potentially reversible diabetes conditions include prediabetes which occurs when the blood sugar levels are higher than normal, but not high enough to be classified as diabetes. Also, gestational diabetes, which occurs during pregnancy but may resolve after the baby is delivered (American Diabetes Association, 2016; Conti et al., 2017; Szendroedi et al., 2016; WHO, 2014).

*Corresponding author address: Email: aolwend@gmail.com, Tel: +254712655407

J Health Inform Afr. 2020;7(2):47-57. DOI: 10.12856/JHIA-2020-v7-i2-290

The rising urbanization has resulted in the change in lifestyles especially with regards to nutrition hence the rising cases of diabetes is associated with the change in the lifestyle. Diabetes is the next epidemic in low income countries owing to the changing lifestyles triggered by a number of factors such as unhealthy diets that encompass consumptions of high calories, and sedentary lifestyles due to socioeconomic development. Sub-Saharan Africa is reported to be experiencing the fastest growing rates of diabetes worldwide. The development and progression of diabetes mellitus is characterized by a number of complications which include cardiovascular diseases such as coronary heart disease with chest pain, heart attack, stroke, and atherosclerosis; neuropathy; nephropathy, retinopathy; skin infections; hearing impairment; Alzheimer's disease; preeclampsia, macrosomia (Conti et al., 2017; Kharono et al., 2017; Mwangi et al., 2017; Yadav et al., 2017).

Nutrition is an important factor influencing the risk of developing T2DM and to some extent T1DM. Excess availability of metabolites such as free fatty acids (sources include dark green leafy vegetables, olive oil, whole grain foods, and eggs) and branched-chain amino acids (sources include chicken, fish, eggs, beans, nuts, and soya) induce whole-body insulin resistance hence minimize the development of diabetes. DM is one of the diseases that require biomarker discovery and translation research to determine the clinical characteristics of their sub-phenotypes right from onset to the manifestation of its complications (American Diabetes Association, 2016; Jones, 2013; Kharono et al., 2017; Szendroedi et al., 2016; Tenenbaum & Avillach, 2016; Yadav et al., 2017).

The long-term complications of diabetes develop gradually and the longer a person lives with diabetes with the blood sugar level less controlled the higher the risk of complications. Diabetes complications may be disabling or even life-threatening. Some of the most common complications of diabetes include: neuropathy, nephropathy, retinopathy, the risk of developing cardiovascular diseases, skin conditions, foot damage, hearing impairment, and depression. Neuropathy is the damage that occurs to the nerve damage. Excess sugar in the blood can wound the walls of the capillaries that nourish the nerves, especially on the legs. This may lead to tingling, numbness, burning or pain that usually begins at the tips of the toes and gradually spreads upward (Daga et al., 2015). On the other hand, nephropathy occurs when diabetes damages the glomeruli, the vessels that filter waste from the blood. Severe damage can lead to kidney failure or irreversible end-stage kidney disease, which may require kidney transplant (Daga et al., 2015). In addition, diabetes can damage the blood vessels of the retina that result in diabetic retinopathy leading to blindness. Diabetes may also increase the risk of other serious vision conditions, such as cataracts and glaucoma (Daga et al., 2015; Mwangi et al., 2017). Moreover, diabetes also increases the risk of various cardiovascular problems, including coronary artery disease with chest pain, heart attack, stroke and narrowing of arteries (atherosclerosis). Persons experiencing cardiovascular complications are more likely to have heart disease or stroke. Left untreated, you could lose all sense of feeling in the affected limbs. Damage to the nerves related to digestion can cause problems with nausea, vomiting, diarrhea or constipation. For men, it may lead to erectile dysfunction. Foot damage occurs when the nerves in the feet are damaged resulting in poor blood flow to the feet. Left untreated, cuts and blisters can develop serious infections, which often heal poorly thus may result into amputation of the affected area. Diabetes may also leave the body susceptible to skin problems and hearing problems. Finally, persons with diabetes may experience depression and increase the risk of dementia, such as Alzheimer's disease (Conti et al., 2017; Richesson et al., 2014; Szendroedi et al., 2016).

1.1 Secondary Utility of EHR Data in Clinical Informatics Research

Precision Medicine (PM) is medical care designed to optimize efficiency and/or therapeutic benefit for a group of patients thus an effort for improve healthcare quality. The primary goal of PM is to uncover disease sub-phenotypes defined by distinct molecular mechanisms that underlie various disease manifestations. However, critical disease subtype distinctions may also be impacted by nonmolecular factors such as socioeconomic status (Tenenbaum & Avillach, 2016).

The field of medicine has undoubtedly grown over the years. One of the many appreciated reasons for the advances in medicine is the applications of health information technologies in various departments within hospitals for purposes such as keeping patient records, medical imaging, and health information exchange. Electronic Health Records (EHRs) are increasingly employed for the management of patient data in primary care worldwide due to the fact there are standards for the design and development of an EHR (van der Bij et al., 2017). This has led to an explosion of electronic patient records collected during routine care. Moreover, historical EHR data may be used for secondary purposes such as in conducting research.

J Health Inform Afr. 2020;7(2):47-57. DOI: 10.12856/JHIA-2020-v7-i2-290

The use of EHR data to conduct retrospective study designs would undoubtedly reduce research costs and promote patient-centered research. However, lack of standards for EHR data and the increasing demand of EHR software worldwide has led to the introduction of software products that record data with questionable quality. Presently, EHR data are characterized by noisy data that comes as a result of erroneous inputs and coding inaccuracies. EHR data are normally inaccurate, redundant, incomplete, and/or irrelevant. Moreover, EHR data may also experience fragmentation in records due to inconsistent patient visits to various healthcare providers without data integration plans. Therefore, clinical data with questionable qualities may not be suitable for secondary uses such as understanding the natural history of disease, cohort identification, biomarker discovery and computational phenotyping just to mention a few (Farrell et al., 2017; Richesson et al., 2014; Tenenbaum & Avillach, 2016; N G Weiskopf & Weng, 2013; Yadav et al., 2017).

1.2 Dimensions for Evaluation of the Quality of EHR Data

The domains of EHR data quality are generally categorized as; conformance, completeness, timeliness, and consistency (Feder, 2017; Kahn et al., 2016; N G Weiskopf & Weng, 2013). However, considering its objectives, this research was limited to assessing chose conformance, completeness, and consistency of the EHR data.

1.3 Conformance

Conformance of data is investigated to ascertain whether its value meets syntactic or structural constraints such as the expected format and data values for each data element. Conformance of EHR data element are categorized as; value conformance, relational conformance, and computational conformance. Value conformance determines whether the data element is a true representation of the expected value. For example, age is expressed using positive integer within an acceptable value range. On the other hand, relational conformance determines whether the data value conforms to relational constraints based on external standards. Finally, computational conformance determines whether the computed data values match validation values defined by external standards (Kahn et al., 2016; Zozus et al., 2014).

1.4 Completeness

Data completeness assesses the presence or absence of data at a single moment over time. Data completeness not only checks the absence of data but also the underlying reason for which such data is missing. Data could be missing due to imputation failure or the personnel failing to document and/or enter such data or due to the patient in question having failed to provide the required data (Kahn et al., 2016; Nicole G Weiskopf et al., 2013; Zozus et al., 2014).

1.5 Consistency

The consistency in data is its constancy with regards to the stipulated data validation rules. That is, the absence of a difference when comparing two or more representations of a data element.

Consistency of data is both affected by the training offered to personnel and the relevancy of the data definitions. Furthermore, data consistency may also be affected by the guidelines and procedures that guide its collection. Data consistency is measured through evaluation of the; procedure for measurement, data measures, and the granularity of the data values. Procedure for measurement evaluates the technique employed for data collection and documentation. On the other hand, data measure evaluates consistency of a variable's unit of measurement and reference range. Finally, data granularity evaluates the degree of detail of the given data element (Kahn et al., 2016; van Engen-Verheul et al., 2016; Nicole G Weiskopf et al., 2013; Zozus et al., 2014).

1.6 Computational Phenotyping

On the other hand, computational phenotyping is the practice of learning latent relationships from raw data without human intervention. Computational phenotyping is achieved through identification of disease phenotypes and sub-phenotypes from healthcare data for adequate management of patient health. The goal of this research was to assess the quality of EHR and determine the suitability of such data to conduct computational phenotyping of diabetes mellitus. Computational phenotyping tasks include; discovering and stratifying new disease sub-phenotypes and;

discovering specific phenotypes for improving classification under existing disease boundaries and definitions. This research was limited to the development of an unsupervised learning model using the DBSCAN algorithm to explore the model's ability to discover and stratify diabetes mellitus cases to help in improving categorization of cases of diabetes under existing complications (Che & Liu, 2017; Denaxas et al., 2017; Ghosh et al., 2016; Richesson et al., 2014; Tenenbaum & Avillach, 2016; Yadav et al., 2017).

Computational phenotyping is achieved through clustering data which involves grouping of objects such that the objects in a group (cluster) are similar (or related) to one another and different from (or unrelated to) the objects in other groups. DBSCAN is a density-based clustering algorithm which determines important properties about the distribution of the dataset. Thereafter, the algorithm constructs a model and a target function that when supplied with the unencountered dataset x, it's upon the target function to determine the cluster(s) to which each record in the dataset is to be assigned. A cluster is a dense region of points which is separated by low-density regions from other regions of high density. Density-based clustering is best applied in the case that clusters are irregular or intertwined, and when noise and outliers are present in the data (Ester et al., 1996; Richesson et al., 2014; Schubert et al., 2019; Yadav et al., 2017; Zozus et al., 2014).

2 Materials and Methods

A retrospective cross-sectional study design that utilized 652 records of confirmed cases of diabetes mellitus that were collected and stored in the EHR database during routine care at The Nairobi Hospital between January 2012 and December 2016. The Nairobi Hospital is located in Nairobi city and it is one of the leading hospitals in Kenya. It has a client-base comprising of persons from middle and upper social classes. Nairobi hospital was chosen because it is one of the few hospitals that have utilized EHR for a period of no less than five years and also embraces research. The data sample size was determined through a stratified sampling of considering both genders. The dataset was subjected to pre-processing focusing on data entry and typing errors. Moreover, missing values were replaced with the arithmetic mean and inconsistencies in data values were also addressed appropriately. Furthermore, outliers were identified and replaced by the arithmetic mean as well. Finally, data were smoothened for the removal of noise in the data and normalized through z-score standardization method. The quality of the dataset was evaluated based on the dimensions of EHR data quality; conformance, completeness, and consistency. Descriptive statistics were measured using SPSS version 21. On the other hand, cluster analysis was conducted using the DBSCAN algorithm and cluster results verified using the International Classification of Diseases version ten (ICD 10) codes assigned to each data record during diagnosis.

3 Results

3.1 Description of the EHR Dataset

A total of 652 records of confirmed cases of diabetes mellitus were extracted from the EHR database. The attributes of the dataset comprised of; Age, Gender, Body Mass Index (BMI), BSA, Pulse, Systolic, Diastolic, Random Blood Sugar (RBS), SPO2 (Oxygen saturation), Temperature, and Respiration as summarized in Table 1.

Table 1. Descriptive statistics of the attributes of the EHR data set.

Attribute	N	Minimum	Maximum	Mean	Std. Deviation
Age	652	21	82	53.4	11.1
Gender	652	0	1	0.5	0.5
BMI	652	16.5	311.5	30.0	17.3
BSA	652	.70	5.6	1.9	0.2
Pulse	652	55	118	82.7	11.2
Systolic	652	52	203	132.9	18.3

J Health Inform Afr. 2020;7(2):47-57. DOI: 10.12856/JHIA-2020-v7-i2-290

Diastolic	652	45	111	80.9	10.7
RBS	652	2.1	22.0	8.2	3.8
SPO2	652	10	100	97.3	3.9
Temperature	652	20.0	37.2	35.9	0.9
Respiration	652	12	181	18.4	6.6

3.2 Evaluation for quality based on domains of EHR data quality

Both Conformance and Consistency of the EHR data elements was evaluated to beyond average except for values for Temperature and BMI which were out of range. Errors with BMI data were as a result in wrong recordings of body Weight and Height or both. However, data incompleteness (as a result of an unexpected value or no value present at all) was the main challenge to the quality of EHR data. Completeness of EHR data was evaluated at 75% and data attributes with cases of Incompleteness were mainly due to data entry errors. The rest of the details are as summarized in Table 2.

Table 2. A summary of the evaluation for quality based on domains of EHR data quality.

Dimension	Parameter	Frequency N= 652	Percentage
Conformance	Value conformance	646	99%
	Relational conformance	652	100%
	Computational conformance	646	99%
Consistency	Procedure for measurement	652	100%
	Data measure	646	99%
	Data Granularity	642	98%
Completeness	Completeness	489	75%

3.3 Distributions of the types of diabetes in the dataset

The distributions of cases of diabetes mellitus in the dataset based on assigned ICD 10 codes included; gestational (1%), Prediabetes (0.4%), T1DM (30%), and T2DM (92%). on the other hand, co-morbidities identified in the dataset were variations of hypertension (67%). Amongst T2DM cases, 60% had hypertension and only 18% of T2DM cases had T1DM. More of the details are as summarized in Table 3.

Table 3. Distribution of the types of diabetes mellitus in the EHR dataset.

Category	Number of Cases	Percentage (%)
Prediabetes	3	0.4
Gestational diabetes	7	1
Hypertension only	25	4
T1DM only	19	3
T1DM with Hypertension	0	0

T2DM only	171	26
T2DM with Hypertension	255	39
T1DM and T2DM without Hypertension	18	3
T1DM and T2DM with Hypertension	155	24

3.4 Complications of diabetes identified from the dataset based on ICD 10 Codes

Retinopathy (12%) was the leading complication associated with diabetes followed by neuropathy (11%) and cardiovascular complications (11%). Other complications in the data set include kidney damage (nephropathy), foot damage, skin conditions, and depression. However, it is interesting to note that 40% of cases of diabetes were not associated with any complications yet only 0.4% of the EHR dataset were prediabetic as summarized in Table 3 and Table 4.

Table 4. A summary of complications of diabetes mellitus based on ICD 10 codes.

ICD 10 Code and Description	Complication	Percentage
F31-Bipolar affective disorder,	Depression	4%
B35.3 - athletes foot B35.3 - Athlete's foot, also known as tinea pedis M21.6 - deformities of foot	Foot damage	4%
H52.4 - Presbyopia-long-sightedness H52.1 - Nearsightedness (myopia) H52.0 - a condition of the eye H35.0 - retinopathy and retinal…	Retinopathy	12%
N08.3-Glomerular disorders in diabetes mellitus I13 - hypertensive heart and kidney diseases	Nephropathy	3%
L30 - unspecified dermatitis – skin L20 - Atopic dermatitis R23 - other skin changes	Skin conditions	2%
I10-neuropathy G63.2 - Diabetic polyneuropathy	Neuropathy	11%
H45 - Transient cerebral ischaemic attacks I64 – Stroke I65 - Occlusion and stenosis of precerebral arteries, … I50 - Heart failure	Cardiovascular	11%

E78.0 - Pure hypercholesterolaemia	High cholesterol	5%
E03.9 – Hypothyroidism E03 - other Hyperthyroidism	Thyroid	1%
E78 - Disorders of lipoprotein metabolism and other lipidaemias	Lipids	7%
Others – no other ICD 10 code except for the diabetic classes	Unclassified	40%

3.5 Clusters identified from the dataset using DBSCAN algorithm

The DBSCAN algorithm categorized 88% of the EHR dataset as noise and the remaining 12% of the dataset were categorized into 23 clusters. The other details are as summarized in Table 5.

Table 5. Clusters identified from cases of diabetes mellitus using DBSCAN algorithm.

Cluster	ICD 10 Codes	Complication
1	E11 – T2DM E10 – T1DM I10 – Essential Hypertension I75 - Atheroembolism H52.4 - Presbyopia-long-sightedness	Eye damage
2	E11 – T2DM	None
3	E11 – T2DM K30 – Functional dyspepsia H40 – Glaucoma – eye condition	Eye damage
4	E11 – T2DM M75.4 – impingement syndrome of shoulder M21.6 – deformities of foot	Neuropathy Foot damage
5	E11 – T2DM E10 – T1DM I10 – Essential hypertension I79 - Disorders of arteries, arterioles and capillaries…	Cardiovascular disease
6	E11 – T2DM E10 – T1DM I10 – Essential hypertension I86 - Varicose veins of other sites E78 - Disorders of lipoprotein metabolism and other lipidaemias	Cardiovascular disease Lipids

J Health Inform Afr. 2020;7(2):47-57. DOI: 10.12856/JHIA-2020-v7-i2-290

7	E11 – T2DM E10 – T1DM I10 – Essential hypertension	None
8	E10 – T1DM	None
9	E11 – T2DM E10 – T1DM I10 – Essential hypertension I74 – I78 - Diseases of capillaries	Cardiovascular disease
10	E11 – T2DM M54.5 - Low back pain	Neuropathy
11	E11 – T2MD F31 - Bipolar affective disorder B35.3 - Athlete's foot, also known as tinea pedis	Depression Foot damage
12	E11 – T2DM I10 – Essential hypertension E78 - Disorders of lipoprotein metabolism and other lipidaemias	Lipids
13	E11 – T2DM I10 – Essential hypertension E78 - Disorders of lipoprotein metabolism and other lipidaemias E66 – Obesity	Obesity Lipids
14	E11 – T2DM E10 – T1DM I10 – Essential hypertension F31- Bipolar affective disorder M13.9 – arthritis – unspecified	Depression Neuropathy
15	E11 – T2DM I10 – Essential hypertension M47 – Spondylosis- Spondylosis-arthritis to the spine	Neuropathy
16	E11 – T2DM	None
17	E11 – T2DM	None
18	E11- T2DM E10 – T1DM H52.4 – Presbyopia	Eye damage
19	E11 – T2MD E10 – T1DM I10 – Essential hypertension	None

J Health Inform Afr. 2020;7(2):47-57. DOI: 10.12856/JHIA-2020-v7-i2-290

20	E11 – T2DM	None
21	E11 – T2DM	None
22	E11 – T2DM E10 – T1DM I10 – Essential hypertension	None
23	E11 – T2DM E10 – T1DM I10 – Essential hypertension	None

4 Discussion

The prevalence of prediabetes (0.4%) in the EHR dataset was quite unbelievable (Mohamed et al., 2018). As a matter of fact, 40% of the cases of diabetes were reported unclassified (based on the assigned ICD 10 codes) meaning that they had not developed into any of the complications of diabetes. Moreover, cases of Prediabetes are believed that usually develop into T2DM thus prevalence of T2DM at 92% was not matching with the prevalence of Prediabetes at 0.4% (Chung et al., 2020). Therefore, the task of diagnosis and assignment of the associated ICD 10 codes must have been compromised. However, effective management of diabetes requires early diagnosis of the disease. Therefore, the few cases of prediabetes show that the actual statistics of persons living with prediabetes maybe incorrect hence the need for urgent measures for the identification of all possible cases of prediabetes cases among the general population.

Moreover, it was interesting to observe that there were no identified cases with both hypertension and T1DM. On the other hand, hypertension was observed in 60% of cases of T2DM. Also, 70% of the cases of T2DM had T1DM and hypertension as co-morbidities. The presence of T1DM, which has genetic predispositions and usually develops at any time right from birth, is increasing among patients that had never until the time of the diagnosis of T2DM. This calls for the need for analysis of the causal relationship between T1DM and T2DM and the causal relationship T2DM and hypertension.

Moreover, the complications arising from cases of diabetes mellitus such as retinopathy; which causes eye damage that results into blindness and complications like neuropathy, and cardiovascular challenges need to be identified early in time since the aftermaths of such complications would not only burden the healthcare system but also lead to a rise in mortality and also affect the productivity of the general population. As a matter of urgency, the government of Kenya should come up with the measures for targeted case finding for not only undiagnosed cases of diabetes but also cases that would develop into retinopathy and cardiovascular complications.

The clustering algorithm determined that 88% of the records had noise hence only 12% of the dataset were utilized in the construction of the target function for the model. This could have been as a result of incompleteness that was evaluated at 25% and the inconsistencies and lack of conformance recorded for data attributes such as Temperature and BMI. Height, Weight, and Temperature are measured values and the challenges experienced with them must have come from the human resource doing the measurements. However, BMI calculations must have been questionable since it is calculated from both Weight and Height whose measurements had already been determined to be questionable.

Moreover, the clustering algorithm seems to agree with the results in Table 3 and takes it further to show the connectedness among the various cases of diabetes. Results in Table 4 show the traditional view of the categories of diabetes mellitus. However, the algorithm showed that not only are the types of diabetes related but also that the complications of diabetes mellitus are intertwined. Despite the wanting quality of the EHR data, clusters identified from 12% of the dataset disclose that computational phenotyping would be realizable from EHR data.

In Table 5, Clusters 7, 19, 22, and 23 show similarities in the disease and associated complications based on the ICD 10 codes. However, the algorithm determined that these clusters are significantly distinct hence form distinct sub-phenotyping groups. Similarly, clusters 2, 16, 17, and 20, based on the characteristics of the dataset look similar. However, the algorithm identified each of them as distinct. This means there is an underlying difference in the data that is not obvious thus only known to the learning algorithm. clusters; 1, 3, and 18 are distinct sub-groups of retinopathy. On the hand, clusters; 5, 6, and 9 are distinct sub-groups of cardiovascular complications. Moreover, clusters; 12 and 13 are distinct sub-groups of lipids. This shows that the task of computational phenotyping from routine healthcare data is achievable since as the algorithm has not only identified the phenotypes but also the sub-types of the given phenotypes. As a result, clinical decision support systems could easily be developed from such backgrounds to assist physicians in the delivery of healthcare services. Moreover, minimal effort to reduce the amount of "noise" from EHR data would probably guarantee much better results.

5 Conclusion

Diabetes mellitus is increasingly infiltrating the health and wellbeing of the Kenyan population hence the need for urgent action by the government to come up with measures for the control of development of the disease. This could be achieved through the combined efforts from both the government and the general population. The Ministry of Health (MoH) should not only increase awareness for diabetes but also conduct targeted case finding exercises for early diagnosis and management of diabetes. According to results in Table 3, cases of prediabetes appear negligible. However, this may not be the true picture given that 40% cases of diabetes were unclassified meaning the diagnosis of associated cases of diabetes were not conclusive.

On the other hand, the MoH should instill measures to ensure all food items undergo clearance by the Kenya Bureau of Standards before they are made available for human consumption as an effort to control their cholesterol contents. Moreover, the MoH should as well increase their efforts of educating the general population on the benefits of eating balanced diets and better nutrition at large.

Funding

No funding was acquired to support this research.

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Approval

The proposal was submitted for ethical approval from Kenyatta University Ethical and Research Committee. Thereafter, the study sought a permit to conduct the research from National Commission for Science, Technology and Innovation (NACOSTI). Moreover, the study sought relevant permissions and approvals from The Nairobi Hospital administration. Finally, consent of participants was sought before their involvement in this research.

References

American Diabetes Association, G. of C. N. R. C. (2016). Standards of Medical Care in Diabetes-2016 Abridged for Primary Care Providers. *Clinical Diabetes: A Publication of the American Diabetes Association*, *34*(1), 3–21. https://doi.org/10.2337/diaclin.34.1.3

Che, Z., & Liu, Y. (2017). Deep Learning Solutions to Computational Phenotyping in Health Care. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 1100–1109. https://doi.org/10.1109/ICDMW.2017.156

Chung, W., Erion, K., Florez, J., Hattersley, A., Hivert, M.-F., Lee, C., McCarthy, M., Nolan, J., Norris, J., Pearson, E., Philipson, L., McElvaine, A., Cefalu, W., Rich, S., & Franks, P. (2020). Precision Medicine in Diabetes: A Consensus Report From the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*, *43*, 1617–1635. https://doi.org/10.2337/dci20-0022

Conti, C., Mennitto, C., Di Francesco, G., Fraticelli, F., Vitacolonna, E., & Fulcheri, M. (2017). Clinical Characteristics of Diabetes Mellitus and Suicide Risk. *Article Mini Review*, *8*(40). https://doi.org/10.3389/fpsyt.2017.00040

Daga, R. A., Naik, S. A., Maqbool, M., Laway, B. A., Shakir, M., & Rafiq, W. (2015). Demographic and Clinical Characteristics of Diabetes Mellitus among Youth Kashmir, India. *Int J Pediatr*, *3*(19), 4–1. http://

Denaxas, S., Direk, K., Gonzalez-Izquierdo, A., Pikoula, M., Cakiroglu, A., Moore, J., Hemingway, H., & Smeeth, L. (2017). Methods for enhancing the reproducibility of biomedical research findings using electronic health records. *BioData Mining*, *10*(1), 31. https://doi.org/10.1186/s13040-017-0151-7
Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. 226--231. http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.121.9220
Farrell, L. J., Shimeng, D., Steege, L. M., Cartmill, R. S., Wiegmann, D. A., & Wetterneck, T. B. (2017). Understanding Cognitive Requirements for EHR Design for Primary Care Teams. *Proceedings of the 2017 International Symposium on Human Factors and Ergonomics in Health Care The*. https://doi.org/10.1177/2327857917061005
Feder, S. L. (2017). Data Quality in Electronic Health Records Research : Quality Domains and Assessment Methods. *Western Journal of Nursing Research*, 1–14. https://doi.org/10.1177/0193945916689084
Ghosh, S., Cheng, Y., & Sun, Z. (2016). Deep State Space Models for Computational Phenotyping. *2016 IEEE International Conference on Healthcare Informatics (ICHI)*, 399–402. https://doi.org/10.1109/ICHI.2016.71
Jones, T. L. E. (2013). Diabetes Mellitus: the increasing burden of disease in Kenya. *South Sudan Medical Journal*, *6*(3). http://www.southsudanmedicaljournal.com/assets/files/Journals/vol_6_iss_3_aug_13/Diabetes_in_Kenya.pdf
Kahn, M. G., Callahan, T. J., Barnard, J., Bauck, A. E., Brown, J., Davidson, B. N., Estiri, H., Goerg, C., Holve, E., Johnson, S. G., Liaw, S.-T., Hamilton-Lopez, M., Meeker, D., Ong, T. C., Ryan, P., Shang, N., Weiskopf, N. G., Weng, C., Zozus, M. N., & Schilling, L. (2016). A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Washington, DC)*, *4*(1), 1244. https://doi.org/10.13063/2327-9214.1244
Kharono, B., Nabisere, R., Kiddu Persis, N., Nakakeeto, J., Openy, A., & Bakeera Kitaka, S. (2017). Knowledge, Attitudes, and Perceived Risks Related to Diabetes Mellitus Among University Students in Uganda: A Cross-Sectional Study. *The East African Health Research Journal*, *1*(2), 105–112. http://www.ncbi.nlm.nih.gov/pubmed/29250612
Mohamed, S. F., Mwangi, M., Mutua, M. K., Kibachio, J., Hussein, A., Ndegwa, Z., Owondo, S., Asiki, G., & Kyobutungi, C. (2018). Prevalence and factors associated with pre-diabetes and diabetes mellitus in Kenya: Results from a national survey. *BMC Public Health*, *18*(S3), 1215. https://doi.org/10.1186/s12889-018-6053-x
Mwangi, N., Macleod, D., Gichuhi, S., Muthami, L., Moorman, C., Bascaran, C., & Foster, A. (2017). Predictors of uptake of eye examination in people living with diabetes mellitus in three counties of Kenya. *Tropical Medicine and Health*. https://doi.org/10.1186/s41182-017-0080-7
Richesson, R. L., Horvath, M. M., & Rusincovitch, S. A. (2014). Clinical Research Informatics and Electronic Health Record Data. *IMIA and Schattauer GmbH*, 215–223.
Schubert, E., Hess, S., & Morik, K. (2019). *The Relationship of DBSCAN to Matrix Factorization and Spectral Clustering*. http://ceur-ws.org/Vol-2191/paper38.pdf
Szendroedi, J., Saxena, A., Weber, K. S., Strassburger, K., Herder, C., Burkart, V., Nowotny, B., Icks, A., Kuss, O., Ziegler, D., Al-Hasani, H., Müssig, K., Roden, M., & GDS Group. (2016). Cohort profile: the German Diabetes Study (GDS). *Cardiovascular Diabetology*, *15*(1), 59. https://doi.org/10.1186/s12933-016-0374-9
Tenenbaum, J., & Avillach, P. (2016). An informatics research agenda to support precision medicine: seven key areas. *Journal of The*. http://jamia.oxfordjournals.org/content/23/4/791.abstract
van der Bij, S., Khan, N., ten Veen, P., de Bakker, D. H., & Verheij, R. A. (2017). Improving the quality of EHR recording in primary care: a data quality feedback tool. *Journal of the American Medical Informatics Association*, *24*(1), 81–87. https://doi.org/10.1093/jamia/ocw054
van Engen-Verheul, M. M., Peute, L. W. P., de Keizer, N. F., Peek, N., & Jaspers, M. W. M. (2016). Optimizing the user interface of a data entry module for an electronic patient record for cardiac rehabilitation: A mixed method usability approach. *International Journal of Medical Informatics*, *87*, 15–26. https://doi.org/10.1016/J.IJMEDINF.2015.12.007
Weiskopf, N G, & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment : enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, 144–151. https://doi.org/10.1136/amiajnl-2011-000681
Weiskopf, Nicole G, Hripcsak, G., Swaminathan, S., & Weng, C. (2013). Defining and measuring completeness of electronic health records for secondary use. *Journal of Biomedical Informatics*, *46*(5), 830–836. https://doi.org/10.1016/j.jbi.2013.06.010
WHO. (2014). WHO | Kenya faces rising burden of diabetes. *WHO*. http://www.who.int/features/2014/kenya-rising-diabetes/en/
Yadav, P., Steinbach, M., Kumar, V., & Simon, G. (2017). *Mining Electronic Health Records: A Survey*. http://arxiv.org/abs/1702.03222
Zozus, M. N., Hammond, W. E., Green, B. B., Kahn, M. G., Richesson, R. L., Rusincovitch, S. A., Simon, G. E., & Smerek, M. M. (2014). *Assessing Data Quality*. https://www.nihcollaboratory.org/Products/Assessing-data-quality_V1 0.pdf