Review Article

Big Data Analytics as Applied to Diabetes Management

Lidong Wang[1, *], Cheryl Ann Alexander[2]

[1]Department of Engineering Technology, Mississippi Valley State University, Itta Bena, Mississippi, USA

[2]Technology and Healthcare Solutions, Inc., Itta Bena, Mississippi, USA

Email address:

lwang22@students.tntech.edu (Lidong Wang)

[*]Corresponding author

Abstract: Type 2 Diabetes Mellitus (DM) affects many people in the U.S. Among the most affected include women, older adults, and some ethnicities/racial groups. Data from numerous sources are used to detect DM and determine self-care activities. In the following paper we discuss Type 2 Diabetes Mellitus, the role of new technologies in diabetes care, diabetes self-management, and Big Data analytics in diabetes management. It was determined by the data in several articles that by using big data we can predict or diagnose diabetes among undiagnosed patients. A wide variety of data can be managed using big data including the Electronic Medical Record (EMR), pharmacy reports, and laboratory reports, among other data. Also there are new mHealth apps that allow the tracking and reporting of data on a secure, wireless connection, through the cloud, etc. Finally, we need to apply the use of big data in future research to determine the significance of our findings.

Keywords: Diabetes, Big Data Analytics, Electronic Medical Record (EMR), Obesity

1. Introduction

One of the most important disease processes that we can study is Diabetes Mellitus, a complex non-communicable disease that is prevalent among minorities, the elderly, and the obese. The universal surge in the number of people who are obese has stimulated a fast rise in the incidence of Type 2 DM which now affects a projected 91% of the over-all populace with the diagnosis of DM in high-income countries. A predicted 193 million adults with DM are at present without treatment; one in two adults suffering from DM [1].

DM is a syndrome of metabolic disease processes distinguished by high blood glucose (BG) which is a result of decreased insulin secretion, insulin action, or both. The chronic high BG associated with DM results in long-term damage, abnormal function, and failure of multiple organs, especially the eyes, kidneys, nerves, heart, and blood vessels, both large and small. The pathogenesis of DM is complicated and involves several processes, ranging from the autoimmune destruction of the β-cells of the pancreas with consequent insulin deficiency to abnormalities that cause insulin resistance. Abnormalities in carbohydrate, fat, and protein metabolism, common to DM, often results in deficiency in insulin action on target tissues. In most cases it is uncertain which abnormality, either alone or a combination of the two, is responsible for the primary diagnosis of DM. Certain infections and a deficiency of growth can also be associated with chronic high BG. People with diabetes may also have a higher chance of atherosclerotic cardiovascular, peripheral arterial, and cerebrovascular disease in addition to hypertension and anomalies of lipoprotein metabolism. Obesity is highly prevalent in Type 2 DM and is the cause of some degree of insulin resistance. Type 2 DM may go undiagnosed for many years frequently because the hyperglycemia is developed gradually and early on in the state of prediabetes, where BG may not be over 126 mg/dl, the classic symptoms of diabetes may not be evident. Symptoms such as polyuria, polydipsia, and polyphagia may be ignored or hidden by food cravings in the obese, or some other factors. Age, obesity, and lack of exercise contribute to the risk of developing DM and in patients with hypertension or dyslipidemia, race/ethnicity is a risk factor. A strong genetic predisposition can increase the risk of DM as well. Note that hemoglobin (A1C) is not recommended for use at present (ADA, 2012). Table 1 [26] demonstrates characteristics of a

diabetes dataset. An immense body of research has validated a host of interventions geared to improve DM. Screening, diagnosis, and therapeutic actions considered to positively affect the health and outcomes of patients with DM are discussed in this paper. Research into new data collection techniques, big data, and Big Data analytics as applied to diabetes is also considered in the paper. The purpose of the study in the paper is to give an overview of how current researchers will benefit from applying Big Data analytics to diabetes management. New horizons for diabetes management are introduced.

Table 1. Components of a Diabetes Dataset.

Components	Description or examples
Past Medical History	• Age & characteristics of onset of disease • Diabetes education history
Physical Examination	• Height, weight, BMI • Vital signs, etc.
Laboratory Tests	• HgA1C • Liver profile, thyroid, fasting lipids, etc.
Referrals	• DSME • Mental health professional as needed

2. New Challenges in Outcome Data for Diabetes

The key to DM management is early diagnosis in prediabetes and diabetes. Increasing in incidence, both diseases are common and inflict substantial public health drain (ADA, 2013). Quality standards, which fosters best practice, inserts the individual at the epicenter of his or her care choices. Patient from Hispanic, African-American, and Asian families are at higher danger of acquiring Type 2 DM. The National Institute for Health and Care Excellence advocates in patient-centered care and better education to encourage self-management of this disease [1].

Glycemic burden is a good indicator of poor outcomes and efficient interventions do exist to decrease the incidence and progression of prediabetes and DM. The successful incidence of early detection through the action of massive early testing of asymptomatic patients has yet to be absolutely proven, and rigorous trials to provide proof are fairly unlikely to occur. A strong prediction for adverse outcomes is the duration of glycemic burden yet efficient interventions do exist to decrease the chance that DM will progress from prediabetes and decrease the incidence of the complications arising from DM. Mass testing of asymptomatic patients has not been established and rigorous trials to gain the proof is unlikely to happen (ADA, 2013). An establish risk factor for Type 2 DM is obesity and unfortunately many patients are too frequently prescribed with antidiabetic medications that increase the appetite and cause weight gain. The most common drugs associated with weight gain in treating DM are sulfonylureas, thiazolidinediones (TZDs), and insulin [2]. Current research has also indicated that hypothyroidism is a risk factor for new-onset DM2, and is also connected to the use of statins, a

widely-used class of cholesterol-lowering drugs. The next generation of research for drug efficacy and therapy development endeavors will most likely benefit from applying cutting-edge knowledge discovery technologies which use Artificial Intelligence and Big Data analytics on the molecules, genetics, mechanism of action, and electronic medical record (EMR), which is confirmed by the outcomes. The art of forecasting clinical outcomes and risk factors can be practically employed in evolving targeted therapies and creating clinical trials, as well as helping produce better-differentiated results with higher therapeutic and commercial options. Therefore, pre-clinical experimental data can be directly linked with potential clinical outcomes data [3].

2.1. Obesity as a Target for Treatment

Now targeted as a disease in its own right, obesity merits diagnosis, evaluation, treatment, and prevention [2]. An option that many patients are now seeking is bariatric surgery, which will help patients lower not only their weight but their BG and HgA1C as well. Indeed, losing weight induces a positive approach for the treatment of DM and has many benefits, including improved insulin sensitivity, improved mortality rates, and restoration of beta-cell function. However, some ethnicities, such as South Asian people, are still hard-to-treat in terms of clinical outcomes and self-care intervention. As for researchers who want to study DM further, the most suitable interventions for this ethnic group should be considered. Research has demonstrated that a 12-week diabetes weight loss intervention can produce substantial weight loss within a primary care setting [4].

Methods to reduce BG include medications such as metformin, a-glucosidase inhibitors, orlistat, and thiazolidinediones, which have been shown to reduce the disease process in several degrees, as well as intensive lifestyle modification programs (ADA, 2013). A primary factor in the treatment of the hyperglycemia of Type 2 DM as highlighted by the updated American Diabetic Association (ADA) position statement is tight BG control. Coupled with decreasing cardiovascular disease (CVD) risk factors such smoking cessation, control of blood pressure, cholesterol control (using therapy such as statins), and if appropriate, antiplatelet therapy, control of BG can be achieved. The control of hyperglycemia decreases the chance of microvascular and macrovascular complications and research has shown the long-term management of high BG can result in better cardiovascular outcomes. The elderly and frail, however, are at risk for hypoglycemia, which is not appropriate for those patients who cannot tolerate it. Those patients with advanced disease may also be more susceptible to too-low BG [1].

2.2. Diabetes Self-Management Education (DSME)

Diabetes self-management education (DSME) is the method of enabling the information, ability, and capacity essential for DM self-care. The primary DSME is

characteristically delivered by a health care professional, while continuing support can be offered by staff within a practice and a host of community-based resources. DSME/S curriculum is intended to refer to the patient's health care views, cultural needs, current knowledge, physical limitations, emotional concerns, family support, financial status, medical history, health literacy, numeracy, and a wide variety of other factors that impact each individual's skill in meeting the trials of self-management [5].

Cost-effective savings of inpatient admissions and readmissions, as well as a projected life-time of health care expenses associated with a decreased risk for complications are just a few of the advantages of DSME/S. Improving A1c (HgA1C) by as more than 1% in patients with Type 2 DM is linked to DSME/S. In addition to this significant decrease, DSME imposes a positive effect on numerous areas such as other clinical, psychosocial, and behavioral aspects of DM care. Managed care corporations, physician offices, and medical homes are now incorporating DSME/S into practice. Four significant stages to assess, engage, and evaluate/readjust DSME/S include but are not limited to: 1) new diagnosis of Type 2 DM, 2) yearly for health maintenance and to avoid the chance of complications, 3) when new complicating factors impact self-management, and 4) when transitions in care occur [5].

Many health care disparities, inconsistent health care resource allocation and consumption, the quality of DM care, dietary habits, levels of physical activity, real and perceived self-care-efficacy, and socioeconomic factors are associated with a divide between treatment advice and self-care. One highly recognized reason for the increasing rate of complications, including microvascular disease processes, is an inadequate control of BG. Self-management of DM consists of obedience to pharmacotherapy, diet, exercise, BG monitoring, foot care, dental care, eye care, immunization, and retrieval of diabetic resources. Researchers have very little information about what personal factors which may encourage or prevent the habit of self-care among African Americans. Additionally, African Americans with DM have a substantial inverse connection with engaging in self-care behaviors. There is a considerable inverse relationship between becoming involved in self-care actions and the objective measurement of BG control over an extended period of time, as well as lowering HgA1c for African Americans. One way to judge determination to perform DM self-care behaviors is attitude [6].

The most investigated model of DM care is the Chronic Care Model (CCM), a structured model of care for patients with chronic illnesses. Originating in the U. S., the CCM has six elements: healthcare organization, self-management support, delivery system design, decision support, clinical information systems, and community resources and policy [7]. This type of action research and the methods used, will give us further opportunity for all researchers to study and understand the DM disease process; collaboratively elaborate on the desired model; and augment the power of providers to develop and internal drive and commitment to help their patients

2.3. The Challenges of Treating Older Diabetics

The health care of older people presents a great many challenges to advanced practice nurses or nurse practitioners, and nurses who have achieved the doctoral level in the science of nursing (DSNs). DM itself is even more challenging today as the elderly are living longer and developing more comorbidities alongside DM. DM and the treatment of the elderly is particularly challenging for DSNs, who conduct research on DM in general as well as self-care behaviors. DSNs are becoming more and more responsible for discharge planning and safe discharge after an inpatient stay; the care of older adults is presenting more challenges in the management and treatment of DM. Research has shown that team work improves patient safety, reduces waste, makes services more maintainable, and increases staff retention as well as increases patient and staff satisfaction. Team work among caregivers has the capacity to give better care, make care safer, more efficient, and sustainable, yet decrease duplication and gaps, especially in the older population where many agencies are implicated in individual patient care. The ever rising challenges of frailty, comorbidities among patients with DM, and the social isolation that many elders experience as a result of living alone, many having lost life-long spousal partners, present real difficulties for the elder patient with DM. However, many elderly patients are able to manage their DM effectively and live independent lives [8]. Due to the challenges, there is a great number of older adults who do not achieve or maintain proper glycemic control. Maintaining a proper BG involves many complicated actions such as eating right/proper nutrition, a regularly modulated exercise program, self-monitoring activities such as checking BG, and medication adherence. Cognitive abilities are required to complete these tasks. Yet despite these problems faced by the elderly with DM, IT-based interventions which require perceptual-motor skills and eye-hand coordination, can help many older adults conquer the difficulties and reach true independence with the right training and system modifications [9].

A significant problem for patients with DM, associated with lower glycemic control, is impaired sleep quality (PSQI). Additionally, a more casual attitude toward self-care activities which are necessary for the optimal achievement of glycemic control and management of DM, a poor attitude toward dietary management and food choices, and a decreased positive attitude toward managing the disease and self-reporting is associated with less sleep. The Functional Outcomes of Sleep Questionnaire (FOSQ), includes the following components: the capacity to realize a lifestyle that is active and productive, maintain social relationships with friends and family, sustain vigilance to required tasks, and continue healthy intimate sexual relationships, even after controlling for age, race, BMI, marital status [10]. The FOSQ identifies patients at risk for poor disease management and outcomes. Older people receive less sleep than younger or middle aged adults and are more affected by a lack of sleep.

A significant problem for the elderly who are diabetic is

dementia. It is a particular challenge for providers who manage older adults with diabetes; the prevalence recorded as 1 in 25 in the 70–79 age group growing to 1 in 6 in the over 80 age group, with numbers projected to grow exponentially by 2050 [8].

Clinical Outcome Search Space (COSS) is used for several reasons including: methodical drug repositioning and intellectual property generation, to identify adverse events—differentiating the causes of adverse events among a drug or the primary disease. One study involved Phase one which included high-throughput in silico processing of a huge group of biomedical data to recognize risk factors for the occurrence of statin-associated DM. The elderly, women, and Asians have been identified through epidemiological studies to be more susceptible to develop statin-associated DM [3].

2.4. The Mindful Attention Awareness Scale (MAAS)

The Mindful Attention Awareness Scale (MAAS) scores patients on the basis of age, sex, race/ethnicity, family history of diabetes, and childhood socioeconomic status and reflects glycemic control and self-care. Mindfulness is typically defined by the patient's ability to attend to everyday normal tasks in a nonjudgmental way while attending to their physical and mental processes. An increased versus lower MAAS means a patient is considerably more likely to have good glycemic control with normal BG and not substantially at risk for Type 2 DM. Better BG control is connected to dispositional mindfulness, mostly due to the decreased incidence of obesity and perceived sense of better control in patients with greater levels of mindfulness. "Dispositional mindfulness" denotes an intrinsic, yet variable, trait, where most everyone has multiple capabilities to focus on and be aware of the "moment" or events that are happening presently. In Type 2 DM, there is little research about the relationship of mindfulness and its relationship to Type 2 DM. Interventions and actions which indicated better glycemic control educated diabetics in mindfulness, while pointing out actions that improve glycemic control such as the following: heir attention The interventions that showed significant improvements in glucose regulation trained diabetic patients in mindfulness, while also directing their attention to behaviors that improve glucose regulation such as the following: 1) diet; 2) physical activity, glucose monitoring; and 3) diabetes medication adherence [11].

3. Big Data Analytics in Diabetes Management

Big Data encompasses the aggregation and merging of large and heterogeneous datasets while education analytics includes an examination for patterns in educational practice or performance in single or aggregate datasets. Health care providers now need to be educated on the handling of complex and powerful dynamics of analytics and Big Data in order to ensure professional educational programs are accountable for all programs run and developed. Although using digital technology has commonly created some trace of the events in which they were employed and even how they were applied within those actions. Big Data exploits multiple data resources and also suggests that most of the big data which could be utilized in health care provider education research most likely was not discovered for education purposes. For example, the data that could be gathered from multiple institutions and merged with other data, including the following: population demographics or health outcomes data, to identify broader patterns of behavior and its impact than possible using educational data alone [12]. Table 2 shows the difference between traditional research methods and the big data method.

Table 2. A Comparison of traditional research methods and the big data method [12].

Aspects	Traditional Methods	Big Data Method
Data collection instruments	Purpose-built data gathering instruments	Opportunistic mining of pre-existing and dynamically accumulating data
Data cleaning	Scrupulous confirmation of data quality	Allowing uncertainty in data quality
Data collection intrusiveness	High	Low
Data acquisition cost per subject	High	Low
Methodology	Governs what data is essential	Reacts to what data is available
Analysis	Prevailing desktop approaches and tools (SPSS, NVivo, Atlas.ti	Potentially new tools required to parse and report on very large datasets
Sample properties	Tightly defined	Loosely defined (depending on markers in the data to specify population features)
Contexts	One to a few	Many (hinges on how a context is defined)
Replicability	Potentially problematic to sustain or recreate study context, resources	Comparatively easy to access and reanalyze data; experiments may be open-ended
Sample size	Variable, tending to small	Variable, tending to very large
Temporality	Static final reports	Dynamic dashboards

Big data is data "whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it [13]." The characterization supported by the working group

on "Big Data analytics and data mining" of the International Informatics Association includes an aggregation of multiple datasets. Succeeding said characterization, which includes that the primary components of big data are on the one hand, data which is multifactorial in character, and on the other hand suggests technological transformation. More to the point, it is universally agreed that "big data" is More precisely, it is widely agreed that the term "big data" is associated with the simultaneous incidence of the V properties, the most significant of which are volume, variety and velocity, but likewise together with veracity and variability exist in the dataset. Data are "large" in size, and volume is evident, but variety depends on different formats by which the data are gathered, from structured material to text, or even images or signals. The necessity to process data at a pace that is rapid enough to withstand decision-making is indicated by velocity. The uncertain or suspect nature of connected data, often huge amounts, very often, or data collected with partial quality control or without processing describes veracity. The time variance of data, most notably in the situation where they are gathered from real-time processing or in the situation of continuous data stream is defined by variability. Therefore, the concomitant presence of two or more factors as noted above defines Big Data analytics, and as a direct outcome, is to alter the technological architecture expended to handle the data and the algorithms formed to analyze them [13].

A goal is to continue to explore analytics to gain what researchers may describe as "intelligent case management," which in today's terms includes targeting, recruiting, retention and compliance. Through the control of the skill to use big data skillfully enough to drive smarter predictive representations, researchers can now begin to examine the efficiency of multiple arrangements of clinical, claims and public data to generate better, more influential, predictive models [14].

Health care researchers discovered, and were among the first researchers to pioneer the concept of randomized clinical studies, unfortunately we have dropped off behind other more industrious researchers and organizations that have been utilizing randomization and analytics in their businesses for many years. The most promising fact about data-driven health care is the capacity to combine the right individuals with the right resources, yet maintaining the ability to deliver better outcomes [15].

Clinical trial modeling has numerous benefits including a reduction in cost and time savings where existing data can precisely simulate a trial and its probable outcomes without the necessity for a real trial. Notably a model of clinical trial outcomes only provides indirect confirmation of the possible outcomes that need to be established. Having said that, a prevailing alternative to real-time clinical trials is modelling and simulation, which is substantially excellent for early decision-making. It also may be a good tool for developing personalized medicines, in which situation it would be hard to enlist a statically significant number of people from distinctive subpopulations into a clinical trial [16].

3.1. Information as Data

All kinds of information can be analyzed using big data. This information includes such data as vital signs, diagnoses, medications prescribed, and numerous sorts of reported symptoms and diseases can be analyzed using big data. Researchers have made the discovery that a diagnosis of a sexual or gender disorder amplified the risk of Type 2 DM by 130%; almost the same rate as suffering from high blood pressure, a well-established risk factor for diabetes [17].

Health data includes a variety of information, both digital and paper. Data can be any of the following: Electronic Health Reports (EHR) of patient's data, clinical reports, emergency care data, doctor's prescription, diagnostic reports, medical images, pharmacy information, health insurance related data, any other clinical data from the Computerized Patient Order Entry (CPOE) system, data from social medias (including Twitter feeds (so-called tweets), logs, status updates on Facebook and other platforms, and web pages); and less patient-specific information such as news feeds, and articles in medical or nursing journals. Data which is raw and can be analyzed further includes: glucose measurements, blood pressure readings, and various other measurements or readings [18, 19]. For example, provider actions and decisions can be validated through an analytical description on patient characteristics which will save costs and be more clinically accurate for provider decision-making; and tools and analysis can be offered to health care providers to provide personalized medicine; this process is applying advanced analytics to the client's report (e.g., segmentation and predictive modeling) [19].

In addition, these multiple datasets can be handled using Big Data analytics thereby now producing results across institutions, media, and areas of clinical data. although veracity adopts the simultaneous scaling up in performance of the architectures and platforms, algorithms. and tools to balance the requirement of big data [19]. Data mining has been used by multiple researchers to create predictive models for diabetes. Huge unstructured input data has been mined from the following datasets: EHR / Patient Health Record (PHR), clinical systems and external sources (government sources, laboratories, pharmacies, insurance companies etc.), in countless forms (flat files, .csv, tables, ASCII/text, etc.), and existing at numerous locations [18, 19].

Current day scientists have discovered that a patient's record can be analyzed using state-of-the-art data analytics to predict the future risk of metabolic syndrome as well as diabetes in patients who are not currently diagnosed with the disease. Using data such as medical claims records, demographics, pharmacy claims, lab tests and biometric screening results over a specified time period can lead researchers to be able to predict the patient's risk of developing these diseases which impact society and health care through increase medical costs for those who suffer from Type 2 DM, metabolic syndrome, and the existing comorbidities [20]. Researchers can also use both claims data or biometric data to determine whether the patient will most

certainly get worse, improve, or remain the same.

In addition, the researchers at the University of California at Los Angeles (UCLA) developed a tool that used the EHR to predict the future of DM and was overall 2.5% better in predicting data. team using the full EHR was 2.5 better at predicting whether a person had diabetes. UCLA researchers collected EHR data from 9948 patients to develop a pre-screening tool that predicts current diagnosis of Type 2 DM. This tool was developed to use multivariate logistic regression and a random-forests probabilistic model for out-of-sample validation. Instead of using basic covariates alone for the detection and diagnosis of patients with Type 2 DM, the team at UCLA decided a full comparison of EHR models would be superior and the following information was included: commonly prescribed medications, diagnoses (as ICD9 categories), and conventional predictors. A more restrictive EHR model for diagnosis and prediction excluded medications, and another more conventional model contained basic predictors and their interactions (BMI, age, sex, smoking status); among these models using the full EHR was far better. Findings from the UCLA study indicated that migraines, depot medroxyprogesterone acetate, and cardiac dysrhythmias were not linked with Type 2 DM, yet sexual and gender identity disorder diagnoses, viral and chlamydial infections, and herpes zoster infections were connected definitively. For random forest machine learning out-of-sample prediction, precision improved during the use of EHR phenotypes. Additionally, using EHR phenotypes was linked to a significantly improved detection rate for diagnosing Type 2 DM, even if there was missing or unsystematically documented data, based on the ROC curves. Furthermore, patients who required more laboratory screening could be identified more effectively by using EHR phenotypes. By integrating EHR phenotype screening to the number of undiagnosed people currently living in the U.S, it is probable that an estimated 400,000 patients with active, untreated diabetes could be identified in contrast to the conventional pre-screening models [17].

3.2. Significant Development in Big Data Datasets

An alternative topic that is increasing in popularity for using Big Data to process large amount of datasets is genomics data repositories. Significant data resources such as the database for genotypes and phenotypes (dbGaP) are highly noteworthy because they hold human phenotype and genotype information. These general datasets, such as the dbGaP, are important due to the broad range of information contained within. The Electronic Medical Records and Genomics (eMERGE) network includes data from researchers with a wide variety of skill in genomics, informatics, statistics, ethics, and clinical medicine and is designed to merge genomics biorepositories with the EHR data to sustain biomedical research. One unique path for investigators is the large patient data repositories available to promote research. However, privacy and data safety may interfere with practical execution of a very large data assimilation strategy. Because of the heterogeneous characteristics of data that may be combined and gathered,

overconfident conclusions can be reached and disseminated [13].

There is an increased interest in big data management from researchers and scientists in different countries and from multiple agencies, such as the National Institute of Health (NIH) and can be attributed to only three combined realities:

(1) Health care is gathering increased quantities of data from EHRs, home monitoring, and biomedical research which includes genomics;

(2) Advances in technology and procedures for data supervision and the newer analytics that are being developed in the private division allows us to store, retrieve, analyze, and remove knowledge from big data groups;

(3) We might attain even greater datasets by assuring our understanding of diseases at a considerably higher pace than we are currently at [13].

A holistic understanding of the diabetic's condition and comorbidities can be achieved by using Big Data analytics and using wearable monitoring devices, wireless telecommunication, and devices that are processed by big data analytics. The function of environment, a role best understood by applying geo-referencing of the datasets, can provide a view of the evolution of diabetes as a disease process. The increasing necessity for advanced IT architecture that relies on agent-based technology and is able to disseminate data storage and computing has already been established by artificial pancreas (AP) projects. The combined accessibility of wearable monitoring devices, wireless telecommunication, and tools capable of managing big data unlocks the opportunity for maintaining a holistic view of the individual with DM [13]. Today's researchers are adopting three primary Big Data oriented resolutions and from an IT point of view. There are three main categories. Cloud computing is an affordable solution for obtaining high computational functioning. Parallel programming, fast becoming easier and more effective, provides the second of the three. And "MapReduce," which is a programming and computing model that implements algorithms on distributed environments, is another widely and successfully used paradigm [13].

mHealth systems include mobile phones, patient monitoring devices, tablets, personal digital assistants, and other wireless devices. BG data can now be automatically gathered, transmitted via phone or Internet, combined with other physiologic datasets, investigated, and represented as actionable data. Mobile decision supporting software applications (apps) are used to help or guide providers in decision making or patients can actually make independent decisions using the apps without waiting on a provider to respond. Five steps are necessary in order for providers to advise remotely based on data from these apps; data must be: (1) collected, (2) transmitted, (3) analyzed, (4) stored, and (5) presented. Medication event information, timing of food intake, amount of carbohydrate intake, amount of exercise, or hypoglycemic symptoms is data that can be gathered via a mobile device system to facilitate automatic patient advice, but an embedded decision support system software package must be available for use, and then the information must

endure a sixth step in the cycle. Aggregation occurs after transmission and before analysis [21]. Now there is a new group of glucose monitors that allow patients to transmit their BG data to the cloud on an individualized, secure website where they are able to review it at any time. The ability for understanding the data from continuous glucose monitoring data, considering the situation in which the data was attained (pre-post prandial, fasting, etc.), can give us more insight into the disease of DM itself and self-management/self-care behaviors. Actually, a patient can adjust their insulin pump by wearing a device fitted with a Geographical Position System (GPS) which localizes their position during exercise and accounts for heart rate, speed, pace, and elevation integrated into an open source database. A complete view of this phenomena occurring to diabetics is Mosaic (Models and Simulation Techniques for Discovering Diabetes Influence Factors, which is a project funded in Europe designed to facilitate the use of big data in diabetes management [13, 21]. Wireless or cable transmission to their smartphone followed by optional automatic transmission to the cloud or by way of automatic transmission to the cloud through a global system for mobile communication radio chip that will be built into the glucose monitor. The information can then be used to help individualize medications and schedules, food choices, or exercise. The apps could solve the following three diabetes management issues: (1) tracking self-monitored BG checks; (2) guide insulin or other medication doses; and/or (3) help determine prandial bolus insulin dosages [21].

The five pathways to value generation; right living, right care, right provider, right value and right innovation are empowered by the use of Big Data. Other important examples do not link precisely use big data but by using support technologies such as cellphones or wearable sensors to track, record, and analyze physical movement, phone activity (calls, texts, etc.), and monitor behavioral therapy for chronic illness such as diabetes, link into Big Data analytics. In addition, another big data application gives asthmatics the opportunity to utilize a GPS-enabled tracker to supervise inhaler usage [22].

Big Data analytics in healthcare are granular data accumulation, temporal abstraction, multimodality, unstructured data, and integration of multi-source data. They are distributed across the 'enable, produce, consume' activities noted above and includes close to a hundred features such as demographics, socio-economic variables, education background, clinical variables such as blood pressure, body-mass-index (BMI), kidney function, sensory-motor function as well as blood glucose levels, cholesterol profile, inflammatory markers, oxidative stress markers, and use of medication [22].

Table 3. Features of Big Data [22].

Aspects	Description	Examples from healthcare
Volume	Size of data	Cohorts of patients, multiple conditions and treatment plans
Variety	Diverse formats and types (numbers, images, text)	Medical, clinical and omics data and images from patients with assorted conditions
Velocity	The rate at which data arrives and changes (streams, batches, infrequent intervals)	Wearable sensors and diagnostics communicating patient behavior
Veracity	Unpredictability of innately inaccurate data types	Patient feedback and clinician notes on patient's condition
Variability	Different interpretations of the same data	Clinical data on the same condition affecting a various group of patients
Value	Intrinsic value addition to the organization against the costs to acquire/accumulate.	Degree of value addition to clinical decision-support and translational research
Sparseness	Low density of useful content (missing or null values)	Variability of patient feedback on symptoms and progress
Complexity	Hierarchies, linkages between entities and recurrent data structures	Multi-pharmacy and multi-morbidity

Applying Big Data analytics can result in substantial cost savings and the same length of stay (LOS) for diabetics and on readmission rates. For diabetics who are admitted to a hospital, there is a significant necessity for diabetic inpatient teaching prior to discharge. It takes an interdisciplinary health care team to provide services for a large population of individuals with a chronic disease such as DM. The ADA encourages learning a set of "survival skills" for diabetics to encourage sufficient instruction and skill in home care after admission to the hospital. The most preferred place to teach diabetes care is on an outpatient basis however, it does require some lifestyle modification. There is a set content for the survival skills that should be focused on prior to discharge and those are: 1) medications, 2) glucose monitoring, 3) hypoglycemia recognition and treatment, and 4) post-discharge contact information [23].

In an examination of the literature for articles published on self-care and diabetes management, several interesting facts surfaced. An exploration of several databases (MEDLINE, CINAHL, Scopus, PsycINFO and PubMed) for articles published from 2002 to 2012 was conducted using the following terms: diabetes, self-management, self-care, barriers and intervention. It was revealed in the literature that the most influential factors on diabetes self-care behaviors were interventions using concepts of self-efficacy, self-determination and proactive coping, and interventions incorporating information technology, and resulted in better health care outcomes [9].

It is quite effective to use multiple datasets when applying

big data analytics to diabetes self-care management. By utilizing administrative claims, pharmacy records, health care use, and laboratory results as variables, the complete health care status and history of every patient in a cohort can be learned and described. Machine learning can then be applied to boost predictive variable set and fit models that may can help researchers find diabetics among records of the undiagnosed. Additionally, notes of hospitalizations, outpatient visits, laboratory orders, and medication fulfillment, for all beneficiaries, and laboratory test results can be combined with the any of the following criteria to increase the accuracy of prediction: International Classification of Diseases, Clinical Modification (ICD-9-CM) code of 250.xx, recorded as a hospital discharge diagnosis or physician clinical encounter; use of a diabetes medication and medication records such as time stamp and duration, drug purchases (ATC codes, defined daily doses, amount, time stamp),; HbA1C value; outpatient encounters (time stamp, regional codes); and GPS data ([13, 24]. Other variables such as age (continuous variable); gender (binary indicator); weight status (overweight—binary indicator; underweight—binary indicator); comorbidity of obesity, hypercholesterolemia, dyslipidemia, uncontrolled hypertension or other cardiovascular disease, sleep apnea, acute bronchitis, hypothyroidism, and anemia (binary indicator), hypercholesterolemia history (binary indicator); high blood alcohol; prediabetic fasting BG levels, increased levels of C-reactive protein and a protective HDL-C; and increased levels of serum alanine aminotransferase will also help scientists who apply big data analytics or machine learning to medical records find undiagnosed diabetes [24].

Figure 1. Big Data Analytics [22].

According to recent research, patients who experience poor quality of sleep and sleep disorders, including insomnia, obstructive sleep apnea (OSA), and restless legs syndrome (RLS), are more likely to receive a diagnosis of DM in their lifetime. Indeed, an observational study of older adults reports that many of these patients suffer from self-reported medication nonadherence as discovered by the data from a 4-item Morisky Medication Adherence Scale and in patients with sleep disturbances, this number was raised by almost 50%. In OSA, most patients are obese, a strong risk factor for Type 2 DM. Other risk factors for OSA included older age and being male. Worse self-care was associated more strongly with increased subjective reports of daytime sleepiness (Epworth Sleepiness Scale) and presented many mental health issues, including depression most [10]. Unfortunately, diabetes seems to be a risk factor for mental health patients as diabetics were almost twice as likely to have depression or other mental health issues than nondiabetic patients [25].

4. New Horizons for Diabetes Management

Clearly we can see significant research on the horizon for diabetes and mental health and diabetes and sleep as the magnitude of this problem seems to be increasing yearly [25]. When analyzing very large long-term monitoring data, novel methods need to be designed for data analysis. The methods include both exploratory and prediction tools. This area promises strong development in the upcoming years [13].

Quantitative methodologies, such as HbA1C measurements, other lab data, BG, and qualitative data that explore appropriate interventions for self-care and the management of diabetes, including the psychosocial aspect need to be developed for future research area. Numerous psychosocial factors can influence DM management including motivation, socioeconomic status, literacy, knowledge, social and the support of health care workers, especially for older adults are crucial factors in self-efficacy, motor skill, and literacy [9]. There are only several mHealth apps for diabetes that have been tested for use and have been shown potentially beneficial, but higher level research into the use of mHealth needs to be conducted. Factors that need to be investigated and explored are (1) privacy to please supervisors; (2) clinical advantage to assure clinicians; and (3) economic advantage to convince payers. Any devices can be used; the most common devices are mobile phones, patient monitoring devices, tablets, personal digital assistants, or other wireless devices [21]. Evidently, a lack of proof exists for treating sleep disorders to improve patient-centered outcomes, or whether treatment of sleep disorders could be a potential benefit for managing the disease. Further research into this area is certainly necessary to advance the use of higher level technologies in diabetes management. There is a lack of evidence regarding the potential effects of treating sleep disorders on patient-centered outcomes, suggesting that further research is necessary to evaluate whether sleep disorder treatment could be an
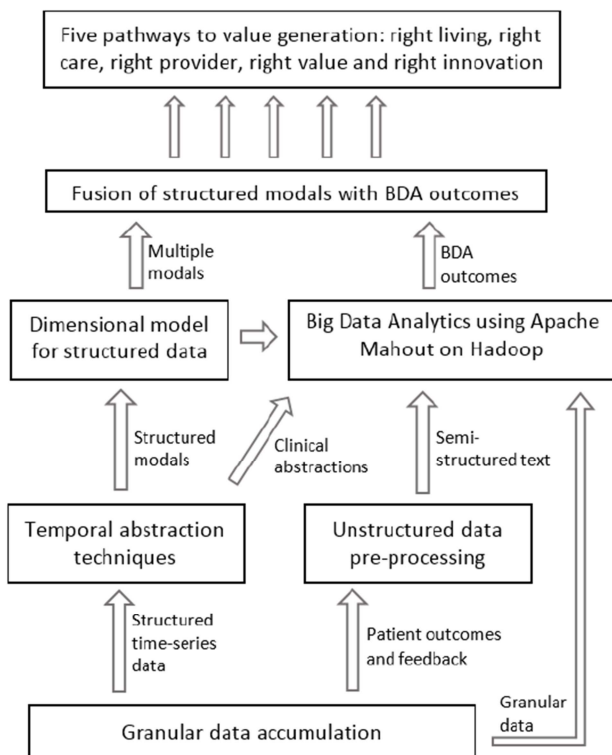
effective strategy for addressing this potential barrier to effective diabetes management [10].

Higher level studies need to replicate current findings to establish causality and determine potential benefit for mindfulness-based interventions to reduce Type 2 DM. Behavioral interventions to stop or reverse population-wide increases in Type 2 DM have stress that behavioral interventions are only mildly, if at all, effective. Novel targets for interventions are necessary to increase effectiveness and finally impact the management of DM on a population-wide level [11].

Naturally, using big data in the management of diabetes is a novel approach and holds the promise of new data to better understand the disease or to uncover new knowledge. A large number of variables can be gathered, examined, and integrated that may identify and pinpoint certain factors which highlight the multifaceted aspects of diabetes management, which is, of course, of increased interest to the health care community. Big data allows the collection of observational, "nonexperimental" information which takes into account the potential biases and unclear facts hidden in the data [13]. Predictive models have been developed for management of diabetes and its complications. Multiple logistic or a similar linear regression is often used for prediction model development [27[. Although extensive effort has been put in to building these predictive models, there is much work to do for impact studies, especially in the big data area.

5. Conclusion

In this paper, we have examined diabetes management from a host of viewpoints and the role of using new technologies in diabetes management. We have also discussed the use of big data analytics in managing diabetes. From our research, we can determine that Type 2 Diabetes Mellitus affects a significant number of adults and results in a wide variety of comorbidities and self-care issues. Older adults, some racial/ethnic groups, and women are particularly vulnerable to the effects of DM. Through the application of big data, we can now see trends and data that was never before available as large amounts of information is processed from a variety of sources, not only the EHR. Not only can we discover data about current diabetics, we are now successful at giving health care providers the ability to predict or diagnose undiagnosed diabetics from raw data fed into a big data program. Further research needs to be conducted to elucidate the role of big data analytics more clearly for future health care providers.

References

[1] Phillips, A. (2016). Optimizing the person-centered management of type 2 diabetes. *British Journal Of Nursing*, *25*(10), 535-538 4p.

[2] Lahiri, S. W. (2016). Management of Type 2 Diabetes in the Setting of Morbid Obesity: How Can Weight Gain Be Prevented or Reversed? *Clinical Diabetes*, *34*(2), 115-120 6p. doi:10.2337/diaclin.34.2.115

[3] PR, N. (2015, June 23). New Biovista Inc. Publication: EMR Data Confirm Big Data Analytics Prediction That Hypothyroidism is a Risk Factor for New-onset Diabetes Mellitus. *PR Newswire US*.

[4] Huntriss, R., & White, H. (2016). Evaluation of a 12-week weight management group for people with type 2 diabetes and pre-diabetes in a multi-ethnic population. *Journal Of Diabetes Nursing*, *20*(2), 65-71 7p.

[5] Powers M, Bardsley J, Vivian E, et al. Diabetes Self-Management Education and Support in Type 2 Diabetes: A Joint Position Statement of the American Diabetes Association, the American Association of Diabetes Educators, and the Academy of Nutrition and Dietetics...Reprinted with permission from Diabetes Care 2015; 38: 1372–1382. *Clinical Diabetes* [serial online]. Spring2016 2016; 34(2): 70-80 11p. Available from: CINAHL Complete, Ipswich, MA. Accessed June 20, 2016.

[6] Kleier, J. A., & Welch Dittman, P. (2014). Attitude and Empowerment as Predictors Of Self-Reported Self-Care and A1C Values among African Americans With Diabetes Mellitus. *Nephrology Nursing Journal*, *41*(5), 487-494 8p.

[7] Praneetsin, C., Pikul, N., Wipada, K., Sirirat, P., Natapong, K., & Turale, S. (2016). Action Research: Development of a Diabetes Care Model in a Community Hospital. *Pacific Rim International Journal Of Nursing Research*, *20*(2), 119-131 13p.

[8] Williams, J. (2016). Effective team working to improve diabetes care in older people. *Journal Of Diabetes Nursing*, *20*(4), 137-141 5p.

[9] Tan, C. L., Cheng, K. F., & Wang, W. (2015). Self-care management programme for older adults with diabetes: An integrative literature review. *International Journal Of Nursing Practice*, *21*115-124 10p. doi:10.1111/ijn.12388

[10] Chasens, E. R., & Luyster, F. S. (2016). Effect of Sleep Disturbances on Quality of Life, Diabetes Self-Care Behavior, and Patient-Reported Outcomes. *Diabetes Spectrum*, *29*(1), 20-23 4p. doi:10.2337/diaspect.29.1.20

[11] Loucks, E. B., Gilman, S. E., Britton, W. B., Gutman, R., Eaton, C. B., & Buka, S. L. (2016). Associations of Mindfulness with Glucose Regulation and Diabetes. *American Journal of Health Behavior*, *40*(2), 258-267 10p. doi:10.5993/AJHB.40.2.11

[12] Ellaway, R. H., Pusic, M. V., Galbraith, R. M., & Cameron, T. (2014). Developing the role of big data and analytics in health professional education. *Medical Teacher*, *36*(3), 216-222. doi:10.3109/0142159X.2014.874553

[13] Bellazzi, R. Dagliati, A. Sacchi, L., & Segagni, D. (2015). New Opportunities for Diabetes Management Big Data Technologies J Diabetes Sci Technol. 2015 Sep; 9(5): 1119–1125. Published online 2015 Apr 24. doi: 10.1177/1932296815583505

[14] Fox, B. (2011). Using big data for big impact. Leveraging data and analytics provides the foundation for rethinking how to impact patient behavior. *Health Management Technology*, *32*(11), 16.

[15] May, E. L. (2014). The power of analytics: harnessing big data to improve the quality of care. *Healthcare Executive*, *29*(2), 18.

[16] Harrison, C. (2012). Deal watch: 'Big data' deal for diabetes clinical trial modelling. *Nature Reviews Drug Discovery*, *11*(11), 822. doi:10.1038/nrd3891

[17] Anderson AE, Kerr WT, Thames A, Li T, Xiao J, Cohen MS. Electronic health record phenotyping improves detection and screening of type 2 diabetes in the general United States population: a cross-sectional, unselected retrospective study. J Biomed Inform. 2016; 54: 162-168. DOI: http://dx.doi.org/10.1016/j.jbi.2015.12.006.

[18] Saravana kumar N M swari T, S ampath P & Lavanya S. (2015). Predictive Methodology for Diabetic Data Analysis in Big Data., *Procedia Computer Science* 50 (2015), 203–208.

[19] Wullianallur R. & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health Information Science and Systems 2014. 2-3. DOI: 10.1186/2047-2501-2-3

[20] GNS, H. (2016). The American Journal of Managed Care Publishes Results Showing Big Data Analytics Can Predict Individualized Risk of Metabolic Syndrome in Patients. *Business Wire (English)*.

[21] Klonoff, D. C. (2013). The Current Status of mHealth for Diabetes: Will It Be the Next Big Thing? J Diabetes Sci Technol. 2013 May; 7(3): 749–758. Published online 2013 May 1.

[22] De Silva, D., Burstein, F., Jelinek, H. F., & Stranieri, A. (2015). Addressing the Complexities of Big Data Analytics in Healthcare: The Diabetes Screening Case Australasian Journal of Information Systems 19 · September 2015 DOI: 10.3127/ajis.v19i0.1183

[23] Hardee, S. G., Osborne, K. C., Njuguna, N., Allis, D., Brewington, D., Patil, S. P., &... Tanenberg, R. J. (2015). Interdisciplinary Diabetes Care: A New Model for Inpatient Diabetes Education. *Diabetes Spectrum*, *28*(4), 276-282 7p. doi:10.2337/diaspect.28.4.276。

[24] Razavian Narges, Blecker Saul, Schmidt Ann Marie, Smith-McLallen Aaron, Nigam Somesh, & Sontag David. (2016). Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors., Big Data. January 2016, 3(4): 277-287. doi:10.1089/big.2015.0020.

[25] Marrero, D. G. (2016). Diabetes Care and Research: What Should Be the Next Frontier?...2015 Health Care and Education Presidential Addres. *Diabetes Spectrum*, *29*(1), 54-57 4p. doi:10.2337/diaspect.29.1.54

[26] Standards of medical care in diabetes--2013. (2013). *Diabetes Care*, 36 Suppl *1*S11-S66. doi:10.2337/dc13-S011

[27] Cichosz, S.L., Johansen, M. D, Hejlesen, O. (2015). Toward Big Data Analytics: Review of Predictive Models in Management of Diabetes and Its Complications. *J Diabetes Sci Technol.* Oct 14; 10(1): 27-34. doi: 10.1177/1932296815611680.