Leveraging Aggregate Datasets from Electronic Health Records (EHR) to Assess Burden of Illness in Chicago

Onyinyechi Uchechi Enyia
B.S. Lewis University, 2006
B.A. Lewis University 2006
M.A. Northwestern University, 2013

THESIS

Submitted as partial fulfillment of the requirements for the degree of Masters of Science in Public Health Sciences
In the Graduate College of the University of Illinois at Chicago, 2015

Chicago, Illinois

Defense Committee:
Edward Mensah, Chair and Advisor
Michael Cailas Environmental and Occupation Health Sciences
John Bing-Canar Health Policy Administration

This thesis is dedicated to my wonderful husband, Stephen Daniel, my daughter Anase, as well as the entire Enyia family. I could not have done it without you all.

ACKNOWLEDGEMENTS

I would like to thank my thesis committee—Dr. Edward Mensah, Michael Cailas, and Kevin Croke—for their unwavering support and assistance. Dr. Mensah in particular provided strict and unrelenting guidance, and always held me to the highest standards of academic performance. His "no excuses" policy shaped my academic training, and has certainly made me a better researcher. Without this strict yet tempered approach, I would not have been able to accomplish my goals. I would also like to acknowledge Michael Cailas, who provided very valuable feedback and guidance throughout the process.

Table of Contents

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

LIST OF ABBREVIATIONS

EHR	Electronic Health Record
T2D	Type 2 Diabetes
GIS	Geographic Information System
CDC	Centers for Disease Control and Prevention
BRFSS	Behavioral Risk Factor Surveillance Survey
MEPS	Medical Expenditure Panel Survey

SUMMARY

This paper used Electronic Health Record (EHR) data in the Chicago Health Atlas database, coupled with the Department of Health and Human Services Inpatient Prospective Payment System (IPPS) Provider Summary for the top 100 Diagnosis –Related Groups (DRG) for 2011 and visualization software to estimate the economic burden, and visualization of diabetes in Chicago.

Data for this study was obtained from the Chicago Health Atlas (CHA) database. CHA is a shared resource with IRB approval to extract data such as diagnoses, medications, and laboratory tests for patients seen at six healthcare institutions throughout Chicago. This data is extracted from the Electronic Health Record, and is de-identified prior to entry into CHA. Patients are assigned a unique cluster ID and were identified as diabetic or pre-diabetic based on the American Diabetes Federation standards of medical care in diabetes. EHR data was accessed through Structured Query Language queries, using local data extraction methods. Total covered charges were calculated, resulting in projected total covered charges, and total payments. Next, diabetic/prediabetic patients along with elements of the built environment were visualized using ESRI ArcGIS to locate diabetes hotspots along with resource density (resources are defined as grocery stores with produce sections, farmer's markets, and parks) within Chicago.

A total of 16,216 diabetic and pre-diabetic patients were identified through CHA. The volume of EHR data provided by the participating institutions offers a representative sample of the entire city of Chicago. Geographic locations with fewer built environmental resources to support a healthy lifestyle were associated with higher prevalence of diabetes/prediabetes.

I. INTRODUCTION

A. Background

Increasingly, the United States and other developed countries are adopting electronic health records (EHR) as a way to reduce medical errors, improve health outcomes, and improve quality and convenience of patient care(healthIT.gov, 2013) According to the American Recovery and Reinvestment Act (ARRA), which was signed into law by President Obama in 2009, Electronic Health Records are defined as an electronic record of health-related information on an individual that is created, gathered, managed, and consulted by authorized health care clinicians and staff (ARRA, 2009). Electronic Health Records are useful for identifying patients that may be due for screenings or immunizations, improving the quality of care, and tracking data over time. Electronic Health Records can also provide a metric by which to measure patient parameters and compare to the larger population. The use of EHRs in this study is particularly significant because it draws on EHRs from several healthcare institutions in a large metropolitan city, and allows for visualization of the burden of diabetes in Chicago. In this study, I use the Chicago Health Atlas (CHA) database to estimate the cost of illness of diabetes in Chicago based on the Centers for Medicare and Medicaid Services (CMS) data of Medicare provider charges. The CHA is a shared data resource spanning six large health institutions in Chicago, that provides data visualization and extracts diagnoses, medications, and laboratory tests seen at all participating institutions, maintaining patient and provider anonymity, and preserving the uniqueness of the patients (CHA, 2012). The CMS database provides average covered charges, and average total payments. Average covered charges are the provider's average charge for services covered by Medicare for all discharges in the diagnosis-related group (DRG). Average total payments are the average of Medicare payments to the provider for the

DRG including the DRG amount, teaching, disproportionate share, capital, and outlier payments for all cases (CMS.gov, 2013).

B.Diabetes impact and cost

Furthermore, diabetes has reached epidemic proportions in some parts of Chicago. On the South Side, for example, Type 2 diabetes is 19.3 percent, which is more than three times the overall Chicago average of 7% (Andreachhia and Shine, 2013) Studies have shown a correlation between racially segregated communities and community health outcomes that may explain disparities in health outcomes, healthcare access, and quality of care (LaVeist, 1993) Carter, Nunlee-Bland and Callender, 2011). Disparities also exist among racial/ethnic minorities that can create barriers to quality care, and diabetes monitoring and management. This study differs from other cost of illness studies because the data utilized in this study is gleaned directly from EHR data stemming from routine clinical care, as opposed to other data gathering methods, such as the Centers for Disease Control and Prevention (CDC) Behavioral Risk Factor Surveillance System (BRFFS). BRFSS is a telephone survey that collects data about health-related risk behaviors, chronic health conditions, and use of preventive services across 50 states, and interviews over 400,000 adults annually (CDC.gov, 2013). Another example, the Agency for Healthcare Research and Quality (AHRQ) Medical Expenditure Panel Survey (MEPS) “collects data on health services, frequency of use, costs of services, payment methods, and data on cost scope and breadth of health insurance”(AHRQ.gov, 2013). This survey consists of a household component and an insurance component. The household component collects data from a nationally representative sample of households, and is based on self-report (AHRQ.gov, 2013). Self-report as opposed to EHR collection is a key difference in the two information gathering methodologies. This paper is particularly important because it differs from BRFFS and MEPS on two fronts. First, the CHA data is free from recall bias that may be a confounding factor in BRFSS and MEPS. Second, this paper offers an estimate of the economic burden of diabetes at the municipal level, whereas BRFSS and MEPS provide state and national survey data. Finally, the significance of this paper also lies in the fact that cost and charge data can be compared across institutions in Chicago.

C. **Purpose of study**

Diabetes mellitus is a metabolic disorder that results from the body's inability to produce and/or use insulin. Type 1diabetes, usually classified as childhood diabetes, comprises about 5% of the entire diabetic population, and occurs when the body does not produce insulin (ADA, 2013). On the other hand, Type 2 diabetes develops when the body ignores or does not produce enough insulin, thus prohibiting the body from using insulin to produce glucose (American Diabetes Association, 2013). While there are several variations of diabetes, i.e. drug or chemical-induced, genetic defects in insulin action etc., this paper is focused on Type 2 diabetes because of its high prevalence. African Americans, Latinos, Native Americans, Asian Americans and Pacific Islanders are at an increased risk for Type 2 diabetes, which makes up the vast majority of the diabetic population within the United States (LaVeist, 2003). Diabetes can lead to several complications, including ketoacidosis, neuropathy, eye complications, foot complications, and hypertension. As such, diabetes can be quite costly to manage for both patients and healthcare providers (American Diabetes Association, 2013). Diabetes management requires consistent patient education in order to address the myriad issues, including blood glycemic control, weight management, nutrition, and medication management (LaVeist, 2003). Furthermore, the recommendations, according to the standards of medical care in diabetes- include screening, diagnostic and therapeutic actions to enhance health outcomes of patients with diabetes (ADA, 2013).

Researchers are increasingly exploring factors that contribute to diabetes and other chronic diseases, including genomics, built environmental factors, diet and lifestyle (Moudon and Daniel, 2009). Use of geospatial methods to assess chronic disease prevalence is critical to the influence of policy, best practice, and continued surveillance of chronic disease (Carter, Nunlee-Bland and Callender, 2011) (CDC, 2012). Furthermore, use of geospatial data can provide an avenue by which health researchers and policy analysts can measure the availability of built resources such as access and distance to health institutions, accessibility of healthful food options, and availability of parks and environments for increased physical activity to help prevent metabolic diseases (Onsurd, Poole, Rugg, Taupier and Wiggins, 2005).

II. CONCEPTUAL FRAMEWORK

A. Review of related literature

Diabetes research is heavy on genomic and genetic studies, as opposed to studies on the contribution of the built environment to the disease. Genome Wide Association Studies (GWAS) explore the genetic factors correlated to various diseases. An example is the Use of Diverse Electronic Medical Record Systems to Identify Genetic Risk for Type 2 Diabetes Within a Genome-Wide Association Study by Abel Kho and colleagues Kho, Hayes, Rasmussen, Tovik et al, 2012). This study was particularly significant because it utilized EMR data collected through routine clinical care to identify cases and controls and target genomic risks for type 2 diabetes. Ultimately, the researchers across the five participating institutions were able to replicate the TCF7L2 gene associated with type 2 diabetes. However, exploration of these genetic factors falls short of explaining the environmental factors that may be associated with the disease, as well as the disparities associated with the disease. In contrast, Patel and colleagues conducted a different study looking at environmental factors associated with type 2 diabetes. In this study, the "environment" explored consisted of 266 environmental "loci" measured across diabetics and controls through the use of environmental assays, utilizing the Centers for Disease Control (CDC) National Health and Nutrition Examination Survey (NHANES) (Patel, Bhattacharya and Butte, 2012). The difference between the Patel study and this paper is that this paper focuses on the external built environment, health disparities and situation awareness coupled with EHR data and the contribution to diabetes prevalence, and economic burden. Situation awareness is critical in order to understand how environmental factors impact disease prevalence. Perhaps one of the most well-known theoretical situational awareness models was developed by Endsley (Endsley, 2000). Endsley defines three levels of situational awareness:

- Perception of the elements in the environment: This level requires perception of *important* information.
- Comprehension of the current situation: This process involves how people combine, interpret, store and retain information. Beyond data and bits of information, it involves the integration of multiple pieces of information, and prioritization of that information relative to the goals or objectives. In the context of public health, practitioners must be able to glean the most relevant information, prioritize it, and assign meaning to information and data elements.
- Projection of future status of the situation: Ultimately, projection implies a high level of understanding of the situation, and timely decision making.

Endsley (2000) also defines situational awareness as "the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 2000) (Woodcock & Toy, 2011). More simply, situational awareness is knowing what is going on around you, through the interpretation of data and information relevant to a given circumstance. In the public health context, situational awareness is critical to the enhancement of public health outcomes through the practical application of actionable public health initiatives utilizing effective information gathering, interpretation, and analysis. Furthermore, visualization is directly associated with situational awareness, as it is the method by which, in the public health context, we are able to familiarize ourselves with the built environment, assigning space and location to elements of our environment, as well as disease prevalence, progression, and burden. Visualization of public health information promotes situation awareness in that it orients us to the factors surrounding public health issues, thus allowing us to target vectors that cause disease. From a policy perspective, it allows us to identify gaps in public health prevention strategies.

Through the exploration of situational awareness and the application of health data collected through EHR to provide data visualization, policymakers and decision makers can be equipped with the necessary tools to allocate resources that could potentially decrease the prevalence of disease or mitigate gaps in health service allocation or access (Dubowitz, 2011). Of particular importance is the use of geospatial technology as a methodological tool to visualize the built environment that can impact the risk for chronic diseases such as diabetes. Data visualization is being used around the world as a tool for decision support for allocation of public health resources. More specifically, health geography is a broad emerging field that uses geospatial configuration to explore morbidity and mortality, sites and systems of care such as hospitals and health care services, and constructs of wellness, identity and place experience Kearns and Collins, 2010). Exploring situation awareness in public health setting is gaining popularity within the United States, as well as other developed countries. An example is the use of roads networks to map community health centers in Slovenia. Lavrac et al. researched the use of GIS data to visualize the Slovenia national road network, and locations of community health centers in Slovenia. This access map provided critical information for decision makers to identify areas with low accessibility to community health centers (Lavrac, Bohanec and Pur, 2007). This is an example of the use of visualization to address gaps in health care provision.

B. Electronic health records

During the previous and current presidential administrations, the use of EHRs has grown exponentially. The Health Information Portability and Accountability Act (HIPAA) enacted in 2009 is a measure to protect personal health information held by covered entities. According to the U.S. Department of Health and Human Services, it specifies the various safeguards to ensure the confidentiality and integrity of protected health information (U.S. Department of Health and Human Services, 2013). In February of 2009, the Health Information Technology for Economic and Clinical Health Act (HITECH) was enacted to promote meaningful use of health information technology. It also addresses information security and privacy concerns, and imposes penalties for entities that violate HIPAA and HITECH. The US Office of the National Coordinator has outlined specific criteria to foster the meaningful use of EHR. Some of those objectives include the following:

- Improve quality, safety, efficiency, and reduce health disparities
- Engage patients and family
- Improve care coordination, and population and public health
- Maintain privacy and security of patient health information

EHR implementation is currently being conducted in stages to streamline the process and provide incentives for providers and health care entities. Stage 1, which occurred from 2011-

2012 focused mainly on wide adoption of EHR for data capture and sharing (HHS, 2012). Stage 2 of meaningful use scheduled for 2014, involves the advance of clinical process, and stage 3 slated for 2016 involves the goal of improved outcomes. The significance of EHR implementation can be seen currently in areas of public health surveillance, immunization, and in various studies of incidence and prevalence of disease (Hripcsak, Soulakis, Li, Morrison, Lai et al, 2009).

C. Quality of care and cost reduction

Ultimately, EHR and GIS can be leveraged to inform public health policy when combined with economic and financial data from healthcare institutions. Health researchers and policymakers have studied the economic burden of chronic diseases on society using health expenditure data and national surveys and reports from the U.S. government (Olin, Machlin and Rhoades, 2008) (Finkelstein, Fiebelkorn and Wang, 2003). Many of these reports aim to estimate costs that could be avoided if cases of the disease were prevented (Honeycutt, Segel, Hoerger and Finkelstein, 2009).

Furthermore, the use of EHR has cost implications for the overall national healthcare expenditures. The expanding Federal deficit has resulted in difficulty controlling Medicare and Medicaid expenditures. According to the CMS, Medicare spending has increased by 7.9% to $502 billion in 2008 and 2009. Medicaid spending increased by 4.9% to $374 billion, due to a 7.4% increase in Medicaid enrollment (CMS.gov, 2011) The unsustainability of this bleak economic backdrop has necessitated research into measures that could potentially reduce healthcare spending and costs overall. Research has shown that the use of EHR can be instrumental in healthcare cost reduction (Payne, Bates, Berner, et al, 2012). More importantly, EHR data coupled with data visualization can create an avenue to illustrate health situation awareness, therefore targeting "hot spots" of disease and illness for prevention efforts. It also has potential to reduce health disparities, particularly for diseases such as diabetes that disproportionately impact communities of color through an increase in patient care quality and access. An example of the use of EHR to facilitate cost reduction is Kaiser Permanente's EHR implementation geared towards enhanced patient satisfaction, improved workflow and process, and streamlining processes and eliminating inefficiencies (Liang, 2010). This work demonstrated the positive impact that implementation of EHR can have on entire processes to facilitate better health care delivery, and reduce unnecessary costs. On the other hand other research has suggested that there may be unintended consequences to the implementation of EHR. High cost, low physician/clinician utilization, and inadequate training were some of the obstacles to favorable implementation of EHR (Middleton, Bloomrosen, Dente, Hasmat, Overhage et al, 2013). In response to some of the unintended consequences of EHR, Middleton et al. compiled recommendations to improve the usability of EHRs. The authors focused on EHR usability-associated medical errors, safe and effective use of EHR, and EHR usability overall (Middleton, Bloomrosen, Dente, Hasmat, Overhage et al, 2013). Cebul et al conducted a study comparing the use of EHR over paper records for achieving improved quality standards for diabetes practices. Overall, the study concluded that achievement of composite standards for outcomes was significantly higher, and EHR sites were associated with higher achievement on eight of nine component standards (Cebul, Love, Jain, et al, 2011) (Buntin, Burke, Hoaglin, MC et al, 2011).

D. Cost of illness studies

There are two main approaches used most often to conduct COI studies: attributable risk based approach applied to aggregate cost data, and the regression based approach applied to individual level cost data.[15] Using an indicator variable and coefficient estimates, the regression based approach models medical spending by attempting to predict individual spending if the disease were eliminated. The attributable risk method identifies medical diseases associated with the disease in question, and attributable factors are calculated for each condition. It then represents the portions of disease prevalence that are caused by the disease in question. These reports have used various sources to conduct these studies, such as the Centers for Medicare and Medicaid Services (CMS), National Center for Health Statistics (NCHS), and the Medical Expenditure Panel Survey (MEPS). However, very little has been documented regarding economic burden of specific chronic diseases at the municipal level. An estimation of disease-attributable costs is useful for quantifying the direct and indirect costs associated with the disease. This paper puts forth an alternative approach to COI analysis at the municipal level. It provides a method of analysis using payer information directly from healthcare institutions, which allows for the comparison of costs across institutions in Chicago. This paper uses the Chicago Health Atlas database, which will be discussed below, coupled with the Department of Health and Human Services Inpatient Prospective Payment System (IPPS) Provider Summary for the top 100 Diagnosis –Related Groups (DRG) to estimate the direct cost of diabetes.

The significance of this study lies in the visualization of diabetes diagnoses within the Chicago, and the ability to compare cost of treatment and economic burden of diabetes across health care institutions. The volume of data provided by the participating institutions offers a representative sample of the entire city of Chicago, and differs from national studies that focus on larger geographical areas, rather than municipalities. In addition, this study allows for the visualization of patient location narrowed down to zip code and elements of the built environment. This is useful not only for policy and decision makers, but for academic research as well. Increasing numbers of healthcare institutions are adopting EHRs in order to be in compliance with the Federal mandate to decrease the costs of healthcare through increased efficiency of health care delivery. Researchers will have greater access and ability to use data mining techniques and geographical information systems (GIS) visualization strategies to carry out studies that can measure the burden of illness at a more granular level, and empower individuals as well as healthcare providers to develop health situation awareness and thus modify behaviors to increase health care outcomes.

E. Justification of patient inclusion from Electronic Health Record

Diabetes diagnosis has been historically based on fasting plasma glucose, or the 2-h value in the 75-g oral glucose tolerance test (American Diabetes Association, 2013). The American Diabetes Association, International Diabetes Federation, and the European Association for the Study of Diabetes recommended the use of the A1C test to diagnose diabetes, with a threshold of >=6% in 2009 (Diabetes Care, 2012) As mentioned previously, there are increased risk factors that contribute to diabetes diagnosis, such as certain ethnic groups, obesity, and other co-morbidities associated with diabetes such as hypertension and elevated cholesterol. HA1C has also been used to diagnose pre-diabetes. Prospective studies have used HA1C to predict progression to diabetes, demonstrating a strong association between elevated HA1C and progression to diabetes. Results from research involving 44,203 individuals from 16 cohort studies with a follow-up of an average of 5.6 years, demonstrated that “with an A1C between 5.5 and 6.0% had a substantially increased risk for diabetes with 5-year incidences ranging from 9-25%. HA1C range of 6.0 to 6.5% had a 5 –year risk of developing diabetes between 25 to 50% and relative risk 20 times higher compared with an HA1C of 5.0%” (American Diabetes Association, 2013). As I will demonstrate in this paper, pre-diabetes based on HA1C results included as part of the criteria for measuring diabetic risk and burden.

Given the prevalence of diabetes worldwide, we can conclude that the costs associated with diabetes will continue to grow exponentially. One of the goals of this paper is to estimate the Cost of Illness as a measure of the economic burden of diabetes in Chicago. This Cost of Illness estimate could potentially serve as a useful tool for policy makers to target interventions specifically tailored to diabetes prevention efforts. In addition, the use of geospatial technology to illustrate average distance that these diabetic and pre diabetic patients travel to receive care helps provide perspective in terms of access to quality healthcare, given the continuous care necessary to manage chronic diseases such as diabetes. I use the Chicago Health Atlas Database to estimate the direct costs of diabetes for a representative sample of Chicago residents utilizing 5 different healthcare institutions in the city of Chicago. The Chicago Health Atlas (CHA) is a shared data resource with IRB approval to extract data such as diagnoses, medications, and laboratory tests for patients seen at six large healthcare institutions throughout Chicago. Data is extracted from the Electronic Health Record (EHR), and is de-identified prior to entry into CHA. In addition, patients are assigned a unique clusterID. CHA is sponsored by the Otho S. Sprague Memorial Institute. Diabetes inpatient cost data was obtained from the Centers for Medicare and

Medicaid System (CMS) Inpatient Prospective Payment System (IPPS) Provider Summary for the Top 100 Diagnosis-Related Groups (DRG).

III. DATA AND METHODS

All institutions contributing data to CHA obtained approval from their Institutional Review Boards to contribute data. The CHA contains data from 2006 to 2012, and continues to amass data at regular intervals. Data extracted contains patient data for all patients above age 18, excluding pregnant women. Institutions are randomly coded to preserve integrity of cost information. Diabetes diagnosis was classified according to the International Classification of Diseases, 9th revision, clinical modification (ICD-9-CM) diagnostic codes. Across sites that provided diagnosis codes, patients with ICD-9-CM codes of 250.xx were used. Furthermore, in alignment with the appropriate national guidelines defining pre-diabetic patients, these patients were identified as those with hemoglobin A1c levels > 5.7, or abnormal blood glucose levels (>110 mg/dl). As previously mentioned, patients with HA1c levels greater than 5.7 and or blood glucose levels greater than 100 are a significantly higher risk of progression to diabetes (ADA, 2013).[6] In this study, patients with ophthalmic disease, cardiovascular disease, or renal disease, which could have potentially made up a large percentage of patients with diabetes, were not included because charge data from CMS was not yet complete for renal or ophthalmic diagnoses. Due to variability in the data collected from each institution, each site was extracted individually, and analyzed separately according to extracted data and the charge data for that particular institution. EHR data was accessed through Structured Query Language (SQL) queries, using local data extraction and analytical tools described below. The first step in the analysis was to prepare the CHA datasets for analysis and to explore associations in the data. First, I extracted the tables with the appropriate data elements. Data elements extracted are summarized in the following table:

TABLE I

Demographic	**Laboratory**	**Diagnosis**	**Payer**	**Site**
Race	HA1C	Diabetes (ICD-9: 250)	Insurance Provider Information	Inpatient
Ethnicity	Glucose Hemoglobin			Office
Gender				Outpatient
Cluster ID				Emergency

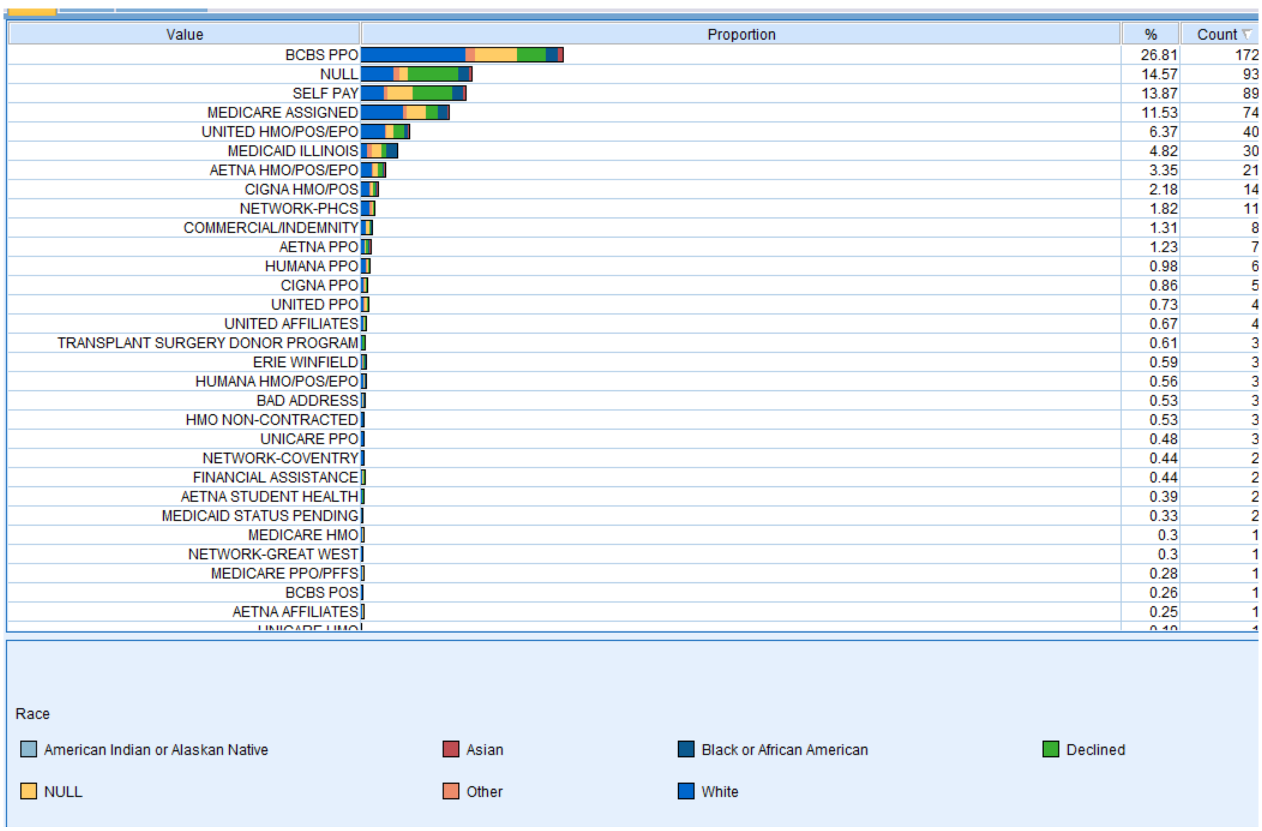
In addition, I used the Department of Health and Human Services Inpatient Prospective Payment System (IPPS) Provider Summary for the top 100 Diagnosis –Related Groups (DRG) to estimate the direct cost of diabetes. The Centers for Medicare and Medicaid Services (CMS) have released data of Medicare provider charges. These are hospital-specific charges from more than three thousand hospitals receiving payments through the Medicare Inpatient Prospective Payment System from the top one hundred most frequently billed discharges. The IPPS contains the DRG definition, Provider ID, Provider Name, Provider Street Address, City, State, Zip Code, Total Discharges, Average Covered Cost, and Average Total Payments (Medicare.gov, 2013) (CMS.gov, 2013). Total discharges are the number of discharges billed by the provider for inpatient hospital services. "Average covered charges are defined as the provider's average charge for services covered by Medicare for all discharges in the DRG, and vary from hospital to hospital due to differences in charge structure" (CMS.gov, 2013). "Average total payments are the average of Medicare payments to the provider for the DRG including the DRG amount, teaching, disproportionate share, capital, and outlier payments for all cases. Also included are co-payment and deductible amounts for which the patient is responsible" (Medicare.gov, 2013).

IBM SPSS Modeler 10.0 was used to analyze the data. Once the tables were extracted, duplicate values were removed and data was cleaned and validated to ensure accurate analysis. There was some variability in data collected from each site with regard to laboratory and/or diagnosis information used to ascertain diabetes/pre-diabetes. Institution A contained glucose labs. Institution C contained both ICD-9-CM codes and HA1c labs. Institution B, Institution D and Institution E contain ICD-9-CM codes. I sought to identify patients that were diagnosed with diabetes based on the ICD-9-CM code for diabetes (250.xx), and extracted this information for Institution A and Institution C. For institutions collecting HA1c data, I created an algorithm that selected out patients with HA1c results greater than 5.7. I then re-ran the data audit to ensure that the data was clean.

Once the data was cleaned, and validated it was useful to perform a distribution analysis of the data to explore overall associations and generate descriptive statistics. This information was useful in assessing patient payer information, race, age and other demographic factors that could contribute to healthcare disparities with regard to diabetes and diabetes care. Distribution analysis illustrates payer overlaid by race. Details of the associations will be described further in the discussion section. Next, in order to identify pre-diabetic patients for further analysis, I selected the unfiltered laboratory data that had been derived to produce HA1c levels greater than 5.7 and collecting sites. I used the performed an algorithm to ensure that only one value per unique cluster ID was included in the analysis to prevent duplicate values within the same year.

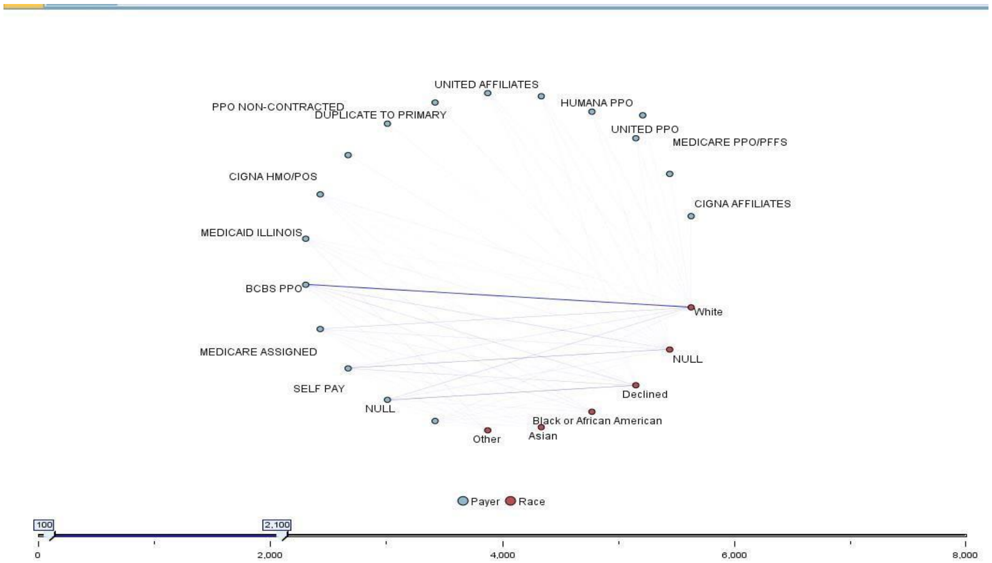
For Institution A, patients were extracted according to glucose level in 2011.The unique cluster ID, lab value, zip code, payer information, ethnicity, race, and gender was collected for each patient. I removed all duplicate values to ensure that only unique laboratory results were included in the analysis. There were a total of 6416 unique patients at Institution A. At Institution E, patients were identified by unique cluster ID, ICD-9-CM specifying diabetes, race, ethnicity, zip code, and payer information. There were a total of 448 unique patients. Descriptive statistics include payer distribution grouped by race, and distribution based on race and gender. Institution B patients were identified for the year 2009. The deviation in years stems from the fact that 2009 contained the most complete diagnosis data for each patient. Patients were identified based on whether or not they had specific ICD-9-CM diagnosis for diabetes. Variables extracted include cluster ID, gender, race, ethnicity, zip code, site of visit and payer information, and there was a total of 5815 unique patients. Institution C patients were extracted as well for the year 2011, and each patient had an ICD-9-CM code diagnosis for diabetes. Variables included cluster ID, payer information, ethnicity, race, zip code, gender, as well as Glucose lab values for that particular visit. There were 841 unique patients. Institution D patients were extracted for the year 2011 as well, based on ICD-9-CM diagnosis code for diabetes, and there was a total of 2696 diabetic patients at Institution D. Variables extracted included unique cluster ID, payer information, site of encounter, and diagnosis. The figures below show the descriptive statistics of diabetic/pre-diabetic at each site. Figure 1 shows the descriptive statistics by race/ethnicity at Institution A.

FIGURE 1



*Institution A distribution analysis

Figure 2



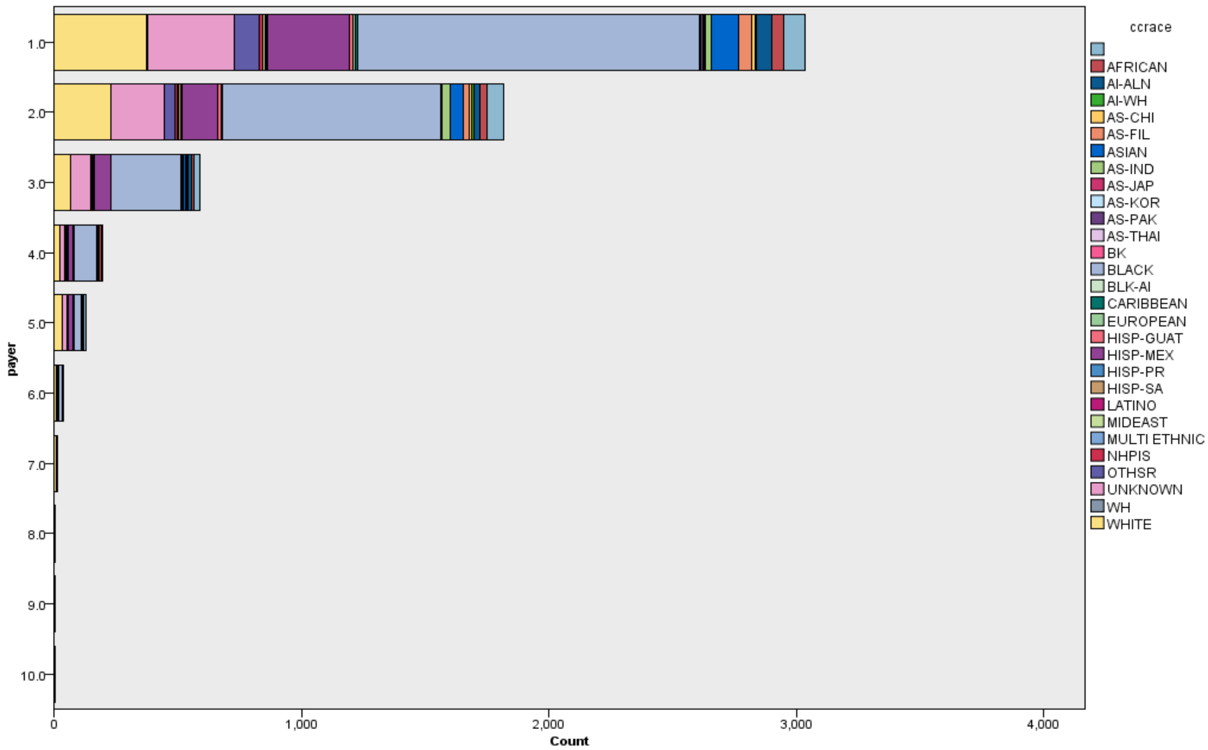
*Institution E distribution analysis

Figure 3.1

1	Medicare
2	Medicaid
3	Private insurance
4	Self-pay
5	No charge
6	Other

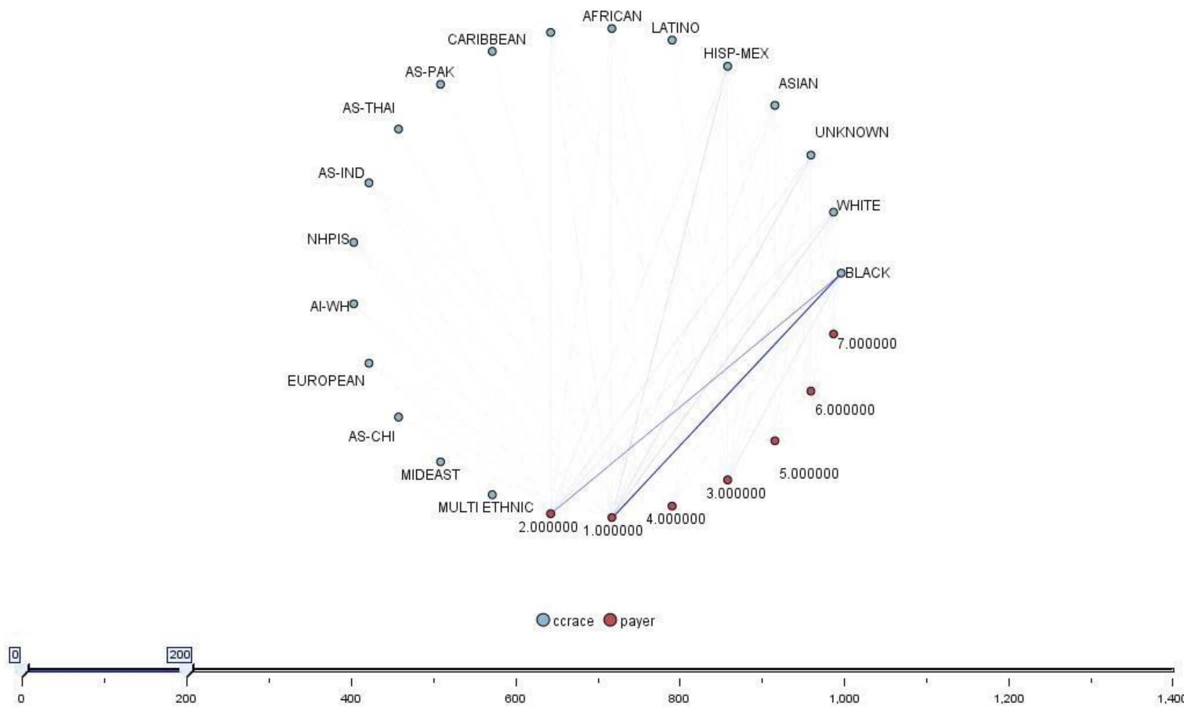
*Institution B payer key

FIGURE 3.2



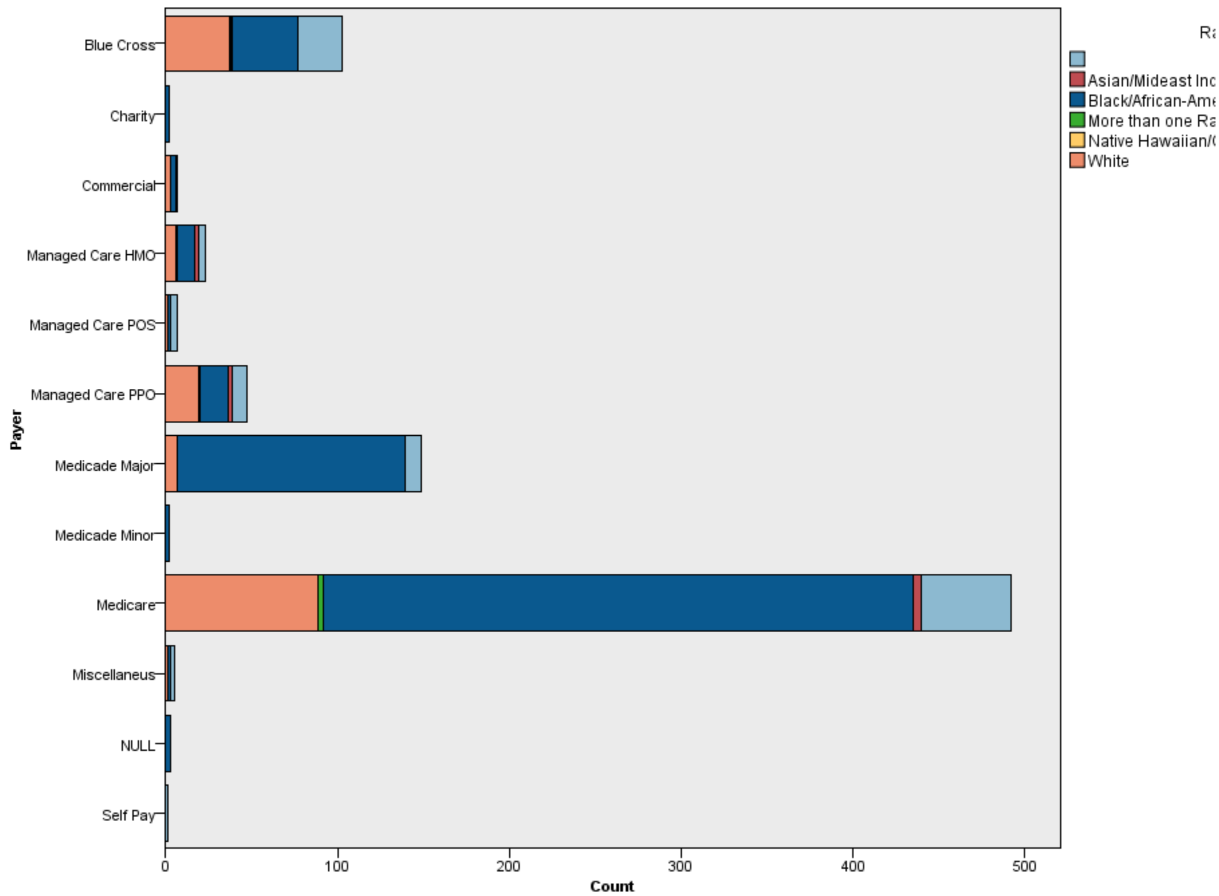
*Institution B patient summary distribution analysis of payer overlaid by race

Figure 3.3



*Institution B payer distribution analysis

Figure 4



*Institution C distribution analysis

Next, the CMS charge data was prepared for analysis. Using IBM SPSS modeler, I used the CMS national charge data and extracted all charge data for the city of Chicago. I then selected charge data for each of the 5 institutions that are part of CHA. I then combined the data, and selected charge data for diabetes, coded as '638 – Diabetes W CC' in the IPPS-DRG. I then aggregated the results, obtaining average covered charges, number of discharges, and average total payments for each institution.

In order to assess burden of illness at each institution, the average covered charges and average total payments at each institution was projected based on the number of diabetes and pre-diabetic patients discharged at each institution. However, I made two assumptions prior to the calculation of COI. First, I assumed that the treatment cost did not change during the year. I also assumed that the number of diabetic patients would remain stable. Charge and payments were projected over ten years. The rationale for ten years is based on studies that have demonstrated that most persons diagnosed with pre-diabetes can be expected to be diagnosed with diabetes within 10 years of the pre-diabetic assessment (ADA, 2013)[6] For the projected treatment cost, the average rate of inflation for healthcare for 2011 to 2021 is 2.3 (CBO.gov, 2011) In addition, the Congressional Budget Office has determined that healthcare costs between 2011 and 2021 will increase at an average annual rate of 5.7% (Keehan, Cuckler, Sisko, Madison et al, 2012). Each institutions average charge data and payment data was multiplied by the number of diabetic (and pre-diabetic where applicable) patients extracted from the EHR for the year 2011 (2009 for Insitution B).

VI.RESULTS

Calculations for each institution based on number of diabetic and pre-diabetic patients are shown in table 1 below:

TABLE II

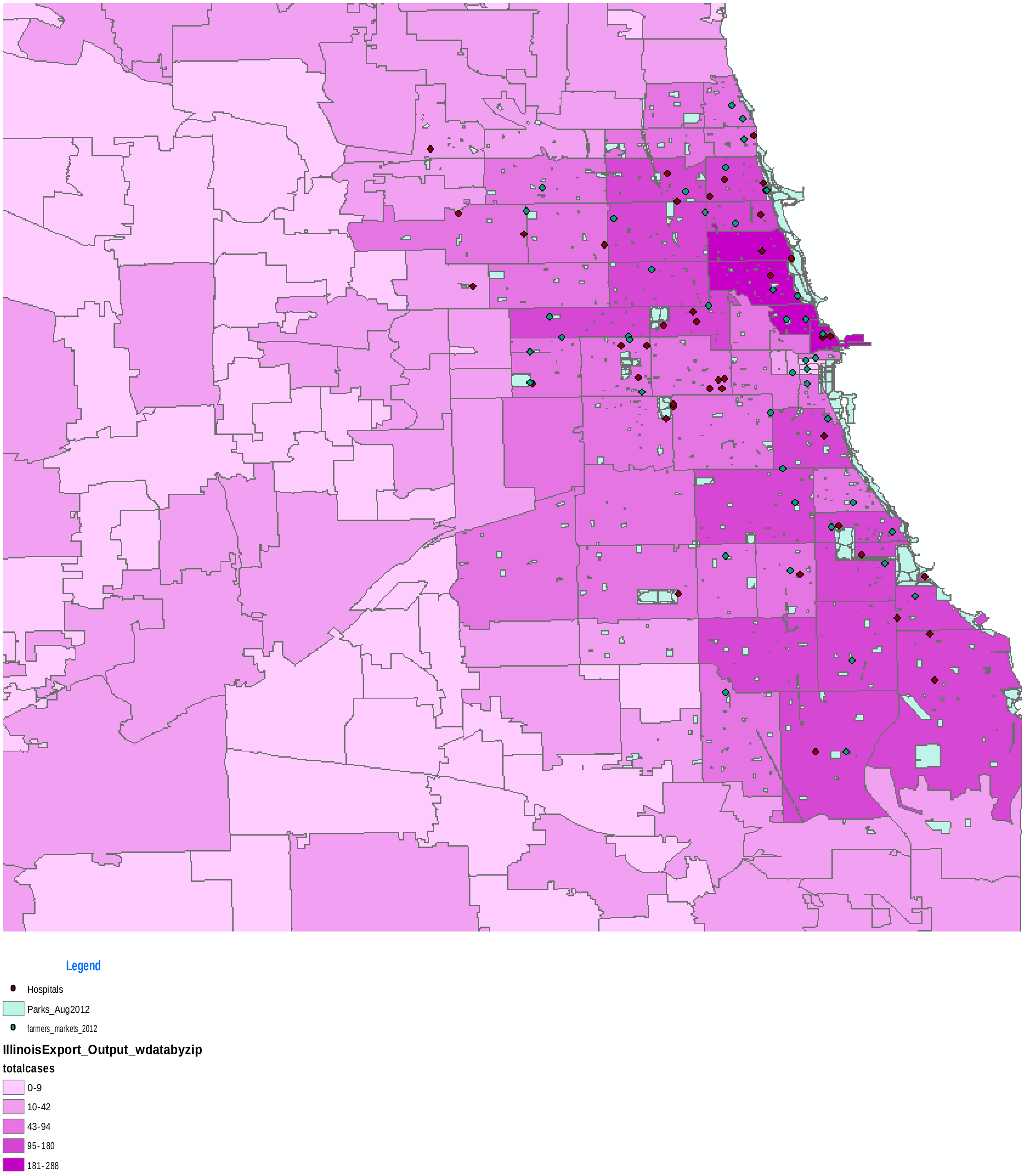
Institution	Total Discharges	Average Covered Charges	Diabetic/Prediabetic Institution based on Electronic Health Record	Total Covered Charges Based on Electronic Health Record	Average Total Payments	Total Payments Based on Electronic Health Record
Institution C	25	$39,164.92	841	$32,937,697.72	$8,826.96	$7,423,473.36
Institution D	46	$18,282.93	2696	$49,290,779.28	$7,773.98	$20,958,650.08
Institution B	12	$14,308.67	1387	$19,846,125.29	$9,402.42	$13,041,156.54
Institution E	33	$21,643.55	448	$9,696,310.40	$10,320.30	$4,623,494.40
Institution A	62	$22,345.21	6416	$143,366,867.40	$7,293.19	$46,793,107.04

TABLE III

Institution	Projected institution Diabetic/Prediabetic Charges CPI 2021	Charge with Projected Health Spending Growth average of 5.7%	Institution Diabetic/Prediabetic Predicted Payment(2011)	Projected Institution Diabetic/Prediabetic Payment CPI 2021	Charge with Projected Health Spending Growth Average of 5.7%
Institution C	$41,998,499.09	$44,392,413.54	$7,423,473.36	$9,465,589.91	$10,005,128.53
Institution D	$628,501,135.01	$66,432,592.71	$20,958,650.08	$26,724,146.11	$28,221,772.44
Institution B	$25,254,683.71	$26,694,200.68	$13,041,156.54	$16,595,193.21	$17,541,119.22
Institution E	$12,363,659.63	$13,068,388.23	$4,623,494.40	$5,895,367.28	$6,231,403.22
Institution A	$182,805,528.82	$193,225,444.00	$46,793,107.04	$59,665,380.38	$63,066,307.06

Next, using ESRI ArcGIS, I created a visual map of the data elements. Patients were identified based on zip code location. I created a patient layer in order to visualize where majority of patients came from to access medical care. This provided a view of the density of diabetic and pre-diabetic patients in various zip code boundaries throughout Chicago. Next, I overlaid the 5 institutions onto the map to visualize distance between institutions. In order to visualize the built environment, I accessed the City of Chicago Data Portal to obtain shapefiles for farmers markets within Chicago. The rationale for visualizing farmer's markets is the ability to purchase fresh fruits and vegetables, and access to produce can mitigate some of the issues associated with access to healthful food options. Gallagher conducted a study on food deserts that highlighted increased obesity, diabetes and other co-morbidities associated with a lack of healthful produce and food options. I also included a layer of Chicago hospitals. The rationale for including hospitals was to observe full services healthcare centers where diabetic patients can go for care.

Figure 5: Chicago EHR-Identified Diabetic Population and Built Environment



V. DISCUSSION

According to the American Diabetes Association, total estimated costs of diagnosed diabetes have increased to $245 billion in 2012 from $174 billion in 2007, representing a 41% increase over five years. Additionally, government insurance, including Medicare, Medicaid and the Military account for 62.4% of cost for diabetes care in the US, with the remainder covered by private insurance. Diabetes affects all segments of the U.S. population, and needs to be addressed across all ethnic groups. However, the burden of diabetes and its complications are disparately carried across racial-ethnic groups. This study and others have demonstrated that there are increasing economic costs associated with diabetes, for both patients and providers. (Payne, Bates et.al, 2012) (Carter, Nunlee-Bland, Callender, 2011). I have also shown that there are not just biological, but psychosocial and environmental considerations for explaining the increased burden of diabetes in various communities in Chicago. This study also confirms the ability to observe patterns of racial-ethnic disparities in access to diabetes care and also illustrates the costs associated with the disease, in addition to ten year projection of costs using EHR data. Observation of figure 5 illustrates that in areas of the South Side with high diabetes prevalence, the number of full service health care facilities are very few. As a result, diabetic patients must travel farther for care. The key aim of this study was to use EHR data to estimate the economic burden of diabetes in Chicago. It was also to illustrate the built environmental factors that may contribute to increased prevalence of the disease, and highlight increased burden of disease in communities of color using EHR data and geospatial technology.

This study is significant because, while cost of illness studies have been conducted at the national level, a cost of illness study has not been conducted for the city of Chicago. More significant is the fact that the study is based on data taken directly from EHR from routine clinical care. Another significant factor is the use of CMS cost and charge data based on diabetes diagnosis, linking it to the EHR data to provide an estimate and projection of the cost of diabetes in Chicago. This study goes even further to use geospatial technology linked to the EHR data to visualize diabetes "hotspots" within Chicago, as well as a few key environmental factors that studies have shown may be associated with the disease (Wen, Kowalseki-Jones, 2012). Based on the visualization of built elements, I observed that certain zip codes within Chicago with higher prevalence of Diabetes also have limited access to full service hospitals, and fewer farmer's markets within a 5-mile radius. In the near future, integration of EHR information and access to that information for providers and patients will have major implications on the way we deliver care, as well as cost-reduction methods. This study can help highlight areas where policy and decision makers can focus resources and additional research to improve public health outcomes through efficient and meaningful technologies.

A. Limitations

There were several limitations that, while important, can be addressed and possibly resolved through further research. The first limitation of the study was in the completeness of the data, as mentioned previously. Institution B contained most complete data from 2009, as opposed to 2011. As a result, analysis was done for the year 2009, which was then projected to the year

2011. Next, accurateness of projected costs may be skewed because the study assumes that all of the patients are still alive at the end of ten years. Thus, costs may be inflated. In addition, the study does not include costs associated with co-morbidities linked to diabetes, such as ophthalmic issues, cardiovascular disease, amputations, etc. Inclusion of these co-morbidities in future research may result in more accurate analysis of cost, as well as a clearer visualization of the burden and/or disparities of these comorbidities throughout Chicago. In addition, the study does not take into account costs associated with diabetes for small community clinics. This information could provide greater insight into the costs for diabetic patients that may not receive care at one of the 5 institutions. Another limitation is the fact that dialysis centers were not included in this study.

VI. CONCLUSION AND RECOMMENDATIONS

Overall, this study, and similar studies have significant policy implications. The use of EHRs will be mandatory in the near future, not just for enhanced quality of care delivery, but for cost reduction. Moreover, access to EHR data can be used in academic research and policy studies to gather population data on diseases and health status in order to inform resource allocation and policies. It has major potential to target disease hotspots, and eliminate many health disparities associated with location and socioeconomic factors. With regard to policy, my recommendation is an in-depth look into the design of EHRs from vendors to ensure that data can be seamlessly extracted and utilized securely for secondary use. This would require standards and interoperability across the numerous EHR vendors. It would also require collaboration with public health agencies. Currently, public health agencies receive surveillance and immunization data from EHRs. However, in the future, EHRs can provide additional vital information, such as asthma or other chronic disease diagnoses and other factors that could provide insight into illness in specific communities.

Furthermore, I would recommend a closer collaborative relationship between healthcare providers at the bedside, and public health agencies using EHR and GIS to provide greater insight into population health. This research is limited in scope, as it does not look at other built environmental factors, such as restaurant and grocery store quality and ranking, prevalence of crime and how it may make physical activity prohibitive, and other factors. Future research will analyze diabetes prevalence, looking specifically at restaurant risk ratings, as well as grocery store ratings and prevalence of grocery stores with minimal produce sections in those communities. Ultimately, this additional research will serve to provide a vivid outline of the environmental, social and physical factors that may influence prevalence of diabetes, and thus economic burden of the disease in Chicago. Finally, access to environmental and cost data in real time is ideal for tackling chronic diseases, whose burden may shift over time. Development of an internet-based application that can provide cost details as well as patient population statistics for chronic diseases such as diabetes, can help policy-makers shift and allocate resources to those communities with the most need. The shift from static data to dynamic, useful EHR information will then be crucial in the allocation of resources, and reduction of healthcare costs.

Appendix

Sample of data, excluding zip code information

Cluster_ID	gender	ccrace	ccethnic	year	site	payer	icd9
22410	0	BLACK	Non-Hispa	2009	O	3	250
52462	0			2009	O	1	250
61323	0	ASIAN		2009	O	1	250
62686	0	BLACK	Non-Hispa	2009	O	1	250
63511	0	UNKNOW	Hispanic	2009	O	1	250
69413	0	UNKNOW	Hispanic	2009	O	2	250
104690	0	HISP-PR		2006	O	2	250
106070	0	HISP-MEX		2009	O	1	250
108399	1	BLACK		2009	O	2	250
121632	0	HISP-MEX		2009	O	3	250
122650	1	OTHSR	Non-Hispa	2009	O	1	250
123284	0	WHITE	Non-Hispa	2009	O	1	250
126325	0	AS-FIL		2009	O	1	250
126869	0	HISP-MEX		2009	O	1	250
128439	1	HISP-GUA		2009	O	2	250
141288	1	BLACK		2009	O	1	250
141699	0	BLACK	Non-Hispa	2009	O	4	250
145381	0	WHITE	Hispanic	2009	O	2	250
153488	1	WHITE		2009	O	2	250
160444	1	WHITE	Non-Hispa	2009	O	1	250
170481	0	WHITE	Hispanic	2009	O	1	250
177278	0	HISP-MEX	Hispanic	2009	O	1	250
191031	0	UNKNOW	Hispanic	2009	O	1	250
280070	1	WHITE		2009	O	2	250
304454	0	BLACK	Non-Hispa	2009	O	2	250
304703	1	BLACK	Non-Hispa	2009	O	3	250
309211	0			2009	O	2	250
310075	1	WHITE	Hispanic	2009	O	1	250
310639	0	BLACK	Non-Hispa	2009	O	2	250

Works Cited

Andreachhia, S., Shine, K. (2013).. Combat ready: Fighting diabetes on Chicago's South Side Jan 30, 2013

Buntin, M. B., M. F. Burke, and M. C. Hoaglin. "The Benefits of Health Information Technology: A Review of the Recent Literature Shows Predominantly Positive Results." *Health Affairs* 30 (2011): 64-71. Print.

Carter, E., Nunlee-Bland, G., & Callender, C. (2011). A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. *Perspectives in Health Information Management.*

Cebul, R. D., & Jain, A. K. (2011). Electronic health records and quality of diabetees care. *New England Journal of Medicine*, *365*, 825-833.

Centers for Disease Control and Prevention. (2013, March 19). Retrieved from http://www.cdc.gov/brfss/

Centers for Disease Control and Prevention (2012, April 20) Retrieved from http://www.cdc/org/mmwr.

Centers for Medicare Medicaid Services. (n.d.). *Centers for Medicare Medicaid Services*. Retrieved April 15, 2013, from http://www.slideserve.com/netis/centers-for-medicare-medicaid-services

Chicago Health Atlas. (n.d.). *CHITREC Chicago Health Atlas Comments*. Retrieved from http://chitrec.org/research/chicago-health-atlas/

Medical Expenditure Panel Survey Background. (2013, June 20). Retrieved from Medical Expenditure Panel Survey: http://meps.ahrq.gov/mepsweb/about_meps/surve?back.jsp

Dubowitz, T. W. (2011). Using geographic information systems to match local health needs with public health services and programs. *American Journal of Public Health*, 1664-1665.

Finkelstein E.A, F. I. (2003). National Medical Expenditures Attributable to Overweight and Obesity: How Much and Who's Paying? *Health Affairs*, 219-296.

Honeycutt, A. S. (2009). Comparing cost-of illness eestimates from alternative approaches: an application to diabetes. *Health Services Research*, 303-320.

Hripcsak, G. S. (2009). Syndromic Surveillance using ambulatory electronic health records. *Journal of the American Medical Informatics Association*, 354-361.

Huang, E. B. (2009). Projecting the future diabetes population size and related costs for the U.S. *Diabetes Care*, 2225-2229.

Keehan, S. C. (2012). National Health Expenditure Projections: modest annual growth until coverage expands and economic growth accelerates. *Health Affairs*, 1600-1612.

Kho, A. H. (2012). Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *Journal of the American Medical Informatics Association*, 212-218.

LaVeist, T. (2003). Racial segregation and longevity among African Americans: an individual-level analysis. *Health Services Research*, 1719-1734.

Lavrac, N. B. (2007). Data mining and visualization for decision support and modeling of public health-care resources. *Journal of Biomedical Informatics*, 438-447.

Liang, L. (2010). *Connected for Health: Using electronic health records to transform care delivery.* San Francisco: Jossey-Bass.

Matthews, S. M. (2009). Work Group II: using geographic information systems for enhancing research relevant to policy on diet, physical activity, and weight. *American Journal of Preventive Medicine*, 171-176.

Middleton, B. B. (2013). Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. *Journal of the American Medical Informatics Association*, 2-8.

n.d. (2012). Standards of medical care in diabetes-2012. *Diabetes Care*.

n.d. (2011) Congressional Budget Office Economic Projections for Calendar Years 2011-2021.Retrieved from http://www.cbo.gov/sites/default/files/cbofiles/attachments/EconomicTables.pdf June 27, 2013.

n.d. (2013, June 20). *American Recovery and Reinvestment Act of 2009, Title XII- Health Information Technology Subitle B- Incentives for the Use of Health Infomration Technology, Section 3031, State Grants to Promote Health Information Technology. State Health Information Exch.* Retrieved from grants.gov: http://www.grants.gov/search/search.do?oppID=58990&mode=VIEW

Olin, G. M. (2008, May). *Estimating the cost of illiness: the case of diabetes.* Retrieved from Agency for Healthcare Research and Quality: http://gold.ahrq.gov Accessed May 29, 2013

Onsurd, H. P. (2005). The future of the spatial information infrastructure. In R. U. McMaster, *A research agenda for geographic information science.* Boca Raton: CRC Press.

Patel, C. B. (n.d.). An environment-wide association study (EWAS) on type 2 diabetes mellitus. *Plos One*.

Payne, T. B. (2013). Healthcare information technology and economics. *Journal of the American Medical Informatics Association*, 212-217.

Robins, K. C. (2010). *Health Geography: A Companion to health and medical geography.* Blackwell Publishing.

Wen, M. K.-J. (2012). The built environment and risk of obesity in the United States: racial-ethnic disparities. *Health and Place*, 1314-1322.

Onyinyechi Daniel
Address: 739 S. Claremont Avenue #3, Chicago, Illinois 60612 Phone: 815.342.2399 Email: oenyia@gmail.com

Highly analytical professional with diverse experience and strong educational background in clinical and public health research and health economics. Excellent leadership and management skills working in a collaborative team setting. Demonstrates strong communication skills in fostering relationships between key departments and department personnel, while presenting reports and research. Proficient in Microsoft Office Suites (Access, Excel, PowerPoint, Publisher, Visio), SQL, Toad Data Modeling Software, Extensive statistical and analytical software proficiency in SAS, Stata and SPSS. Proficient skills in visualization using Arc GIS software. My training is complemented with solid background in data analysis, database modeling and design as well as website development. Articulate in English and Nigerian Igbo, with working knowledge of Spanish language. Areas of expertise include Health Policy, Health Economics, Public Health Informatics, Data Analytics, Global Health, Clinical Research, Research Development and Data Visualization.

EDUCATION

PhD Health Policy and Administration Dual Concentrations in Public Health Informatics and Health Economics — Currently Enrolled
University of Illinois at Chicago School of Public Health

December 2013
Master of Science in Health Policy and Administration
Concentration: Public Health Informatics
University of Illinois at Chicago School of Public Health

December 2012
Master of Arts in Medical Humanities and Bioethics
Northwestern University, Chicago, IL
Full Scholarship, Northwestern Memorial Hospital Ethics Committee

May 2007
Bachelor of Science in Biology and Bachelor of Arts in Chemistry
Lewis University, Romeoville, IL

Certified Clinical Research Professional (CCRP)
Society of Clinical Research Associates (SOCRA)

March 2012

PROFESSIONAL EXPERIENCE

Senior Clinical Analyst — October 2014-Present
Presence Health Elmwood Park, IL

- Collaborates with ministry, region, and system leadership to define and standardize clinical performance metrics across the system
- Researches and ensures appropriate internal and external benchmarks are used for performance comparisons
- Analyzes data using SAS, control charts, graphs, tables, and descriptive statistics to create analyses, dashboards, scorecards, and reports that define opportunities for improvement

and make appropriate recommendations

- Participate in and provide subject matter/data expertise for ministry-specific, region-specific, and system-wide clinical initiatives and projects
- Maintains knowledge of current trends in healthcare quality, healthcare informatics, data collection requirements, public reporting, and regulatory compliance
- Facilitates education of customers at all levels of the system regarding performance metrics, indicator definitions, benchmarks, and other data-related information
- Researches, educates, and reports on all government and reporting agency metrics that impact our current and future business

---

Research Data Manager January 2013–Present
Northwestern University (Division of Infectious Disease), Chicago, IL

Responsible for writing and executing computer programs for data collection, analysis, management, and QC, using SAS. Designed, updated and maintained Access databases and associated table structures. Assisted in the creation and editing of SAS programs that perform data error and logic checks to verify accuracy and consistency of data; performed data cleaning and modifications; validate and secure databases; established quality control and coordinated data flow procedures; determined the most effective methods for data collection and analyses; developed and implemented study protocols; prepared and customized reports for Principal Investigator, Project Director, and laboratories; attended meetings and presented data as needed; maintained local repository database; communicated with coordinating center regarding the reporting, editing and transfer of project data; fielded internal and external queries and communicated with laboratory and hospital staff to resolve data collection issues; responsible for investigating, verifying, and monitoring all quality assurance questions and queries and tracking of clinical trials.

Major Accomplishments:

- ✓ Successfully executed a new system that allowed for translation of HL7 laboratory files to XML documents, for conversion to SAS data sets used for analysis of data.
- ✓ Designed and maintained Access database for newly screened and enrolled MACS participants. Successfully executed a new system that increased the consent volume of patients into critical HIV sub-studies used for research analysis.

Research Consultant Jun 2007–Dec 2012
Northwestern University (Division of Infectious Disease), Chicago, IL

Assume responsibility in coordinating the process for acquiring informed consent from HIV Outpatient Study (HOPS) patients; managed large Access database, fielding queries and resolving data issues; ensured integrity and accuracy of data and performed other data analysis; compiled and sent timely data reports for Project Manager and Principal Investigator; synchronizing and conducting infectious disease research sub-studies, including HIV viral load point of care and behavioral studies among HIV patients and NA-Accord study using cancer registry data; fielded internal and external queries and communicated with laboratory and hospital staff to resolve data collection issues; examining patient antiretroviral trends and contributed to publication of findings in several academic journals; investigating, verifying, and monitoring all quality assurance questions and queries and tracking of clinical trials; and communicating effectively with project director, IRB, CDC program staff and associates.

Major Accomplishments:

- ✓ Successfully executed a new system that increased the consent volume of patients into critical HIV sub-studies used for research analysis.
- ✓ Boosted patient enrollment by 96% in four years by amending consenting practices; making use of the time necessary to onboard patients; and cooperating with physicians, investigator, and clinic nurse staff, and streamlining Access database.
- ✓ Successfully managed large patient database; conducted queries for physicians and other clinical trials and studies.

Graduate Teaching Assistant Aug 2011-Present
University of Illinois at Chicago School of Public Health

Teaching assistant for Health Information Decision Support Systems course as part of the Public Health Informatics curriculum at the University of Illinois at Chicago. My responsibilities include maintenance of grade records, proposal of grades for class participation, proposal of grading criteria for assignments and exams, and preliminary evaluation of assignments and exams based upon instructor criteria. I also served as liaison between students and the instructor.

PRESENTATIONS

Centers for Disease Control Public Health Informatics Conference April 2014
Analysis of Electronic Health Records for Population Health Management

American Medical Informatics Association Annual Symposium Nov 2013
Leveraging aggregate datasets from Electronic Health Records to provide data visualization and estimate economic burden of Diabetes in Chicago

Bioethics in the Developing World: Concepts and Practice April 2013/2014
Guest Lecturer, Global Bioethics Undergraduate Course, Northwestern University

Ethical Concepts of Autonomy and Consent in Nigerian Healthcare Feb 2011
Presenter, Medical Ethics Rounds, Northwestern University

West African Bioethics Training Program-Ethical Oversight and Research Integrity Module Aug 2010
University of Ibadan, Ibadan, Nigeria

Awards

Northwestern University Medical Humanities and Bioethics Travel Award April 2010
Northwestern Memorial Hospital Ethics Committee Full Scholarship Sept 2008
Rodney Musselman Award University of Illinois at Chicago School of Public Health October 2013

Research Contributions

Initiation of HAART at Higher CD4 Cell Counts is Associated With A Lower Frequency of Antiretroviral Drug Resistance Mutations at Virologic Failure. Journal of Acquired Immune Deficiency Syndromes (JAIDS). August 2009. 51(4).450-453 - Contributor

Rates of AIDS-Defining Opportunistic Illnesses (OIs) and CD4 Cell Counts at OI Diagnosis in a Cohort of US Patients (1994-2006) CDC presentation September 2007 – Contributor

Hospitalizations and Associated Diagnosis of HIV Patients in the United States: Conclusion. AIDS. 2008 22(11):1345-1354.- Contributor

Original Research

Leveraging aggregate datasets from Electronic Health Records to provide data visualization and estimate economic burden of Diabetes in Chicago

Thesis: Cardiovascular Disease: The Socioeconomic Impact of Healthcare Disparities Among Racial/Ethnic Minorities — May 2006

Healthy Schools Campaign: Strategies to Combat Obesity and Overweight in Chicago's Schools — 2006

Evolutionary Genetics: The Broken Gene and its Link to Disproportionately High Rates of Hypertension in the African American Populous for Student National Medical Association — Oct 2005

PROFESSIONAL DEVELOPMENT

Certified Clinical Research Professional, Society of Clinical Research Associates

Human Subjects Research Training

Good Clinical Practices Training

Health Insurance Portability Accountability Act (HIPAA) Training

Collaborative Institutional Training Initiative (CITI)

PROFESSIONAL AFFILIATIONS

American Medical Informatics Association Member of Public Health Informatics and Global Health Informatics Working Groups

American Society of Health Economists

Society of Clinical Research Associates

Association of Clinical Research Professionals

Northwestern Memorial Hospital Ethics Committee

Chicago Transplant Ethics Consortium

Northwestern University Infectious Disease Clinical Trials Community Advisory Board (CAB)