

An Overview of Health Analytics

Wullianallur Raghupathi^{1*} and Viju Raghupathi²

¹Graduate School of Business, Fordham University, New York, NY, USA

²Finance and business management, Brooklyn College, City University of New York, Brooklyn, NY, USA

Abstract

Objectives: We examine the emerging health analytics field by describing the different health analytics and providing examples of various applications.

Methods: The paper discusses different definitions of health analytics, describes the four stages of health analytics, its architectural framework, development methodology, and examples in public health.

Results: The paper provides a broad overview of health analytics for researchers and practitioners.

Conclusions: Health analytics is rapidly emerging as a key and distinct application of health information technology. The key objective of health analytics is to gain insight for making informed healthcare decisions.

Keywords: Data warehousing; ETL; Descriptive analytics; Discovery analytics; Health analytics; Informed decision; Insight; Predictive analytics; Prescriptive analytics

Introduction

The past several years have seen an exponential increase both in the number of available health information technology (HIT) applications and their use by health care providers [1,2]. This trend is supported by several developments: government stimulation in the way of laws and regulations (e.g., adoption of electronic health records and meaningful use), the proliferation and availability of various applications, the rapidly declining costs of acquisition and storage of structured and unstructured health data, and the adoption of interoperability and standards. It's not surprising, therefore, that data acquisition is increasing substantially across different types of HIT, including electronic health records [3], clinical decision support systems [4,5], medical imaging [6], public health databases, and the proprietary systems of health care providers (e.g., physicians, insurance companies, HMOs, hospitals, government agencies, and others).

Data are also being accumulated in various web 2.0 and social media applications such as Twitter, Facebook, YouTube, blogs, and wikis, as well as email messages and mobile applications. These high volumes of data are collected for, among other reasons, compliance and regulation reporting. Stakeholders recognize the opportunity and potential for exploiting their large health data sets with advanced analysis to gain insight for making informed health care decisions [7-10], thereby improving quality of care and reducing costs [11-14]. Furthermore, the positive influence on health outcomes, thanks to sophisticated decision support [10], propels the practice of evidence-based medicine [15], personalized medicine, and improved public health.

While aspects of such health analytics as the use of statistical models, data mining [16,17], and clinical decision support have existed for decades, only recently have more advanced and sophisticated information technologies and tools been readily available to support decision making in an integrated fashion. In this context, health analytics is a vendor-driven term that consolidates the various dimensions of "analysis". The range of technologies and tools-databases, electronic health record systems, data warehouses, web applications, clinical decision support systems and others-are integrated for interoperability and seamless processing of health data for insight [18]. Such integrated applications help doctors, care givers, and patients understand and make sense of health data as needed for diagnosis, treatment, health

management, and preventive care. In one scenario, for example, the use of health analytics technologies can ensure that emergency room doctors are briefed and ready to treat patients prior to their arrival by ambulance.

Diagnostic and current health data can be downloaded by hospital staff from a wide variety of systems to develop a patient profile that includes past illnesses, chronic conditions, allergies, blood and tissue typing. With this information as well as a constant stream of vital sign data fed directly by paramedics en route to the hospital, receiving doctors can make better decisions about the care processes and treatment plans for each patient prior to arrival in the emergency room. And by compiling all incoming information into a dashboard, doctors can model a patient's situation by comparing it to similar cases, outlining possible complications and readying necessary resources and medical equipment.

Analytics is also key in helping operating room doctors better predict the effects of anesthetics during surgical procedures. In a unique pilot project at The Ottawa Hospital, analytics are used in a patient safety learning system [19]. If a patient experiences a reaction from an anesthetic administered during a surgical procedure, the hospital administers a counter measure used to negate a narcotic overdose. With advanced analytics and modeling technology, clinicians can pinpoint the type of patients prone to narcotic reactions and avoid the use of certain anesthetics with susceptible patients. With analytics, doctors can identify vulnerable patients early, in triage, limiting the use of narcotics counter measures and more effectively managing the use of narcotics in the operating room. The Ottawa Hospital administrators can also predict trends in hospital use, more accurately forecasting days of higher than normal patient volume and allocate resources and staff accordingly, ensuring that departments run at peak efficiency while

***Corresponding author:** Raghupathi W, Graduate School of Business, Fordham University, 113 W. 60th Street, New York, NY, 10023, USA, E-mail: raghupathi@fordham.edu

Received August 08, 2013; **Accepted** September 11, 2013; **Published** September 18, 2013

Citation: Raghupathi W, Raghupathi V (2013) An Overview of Health Analytics. J Health Med Informat 4: 132. doi:10.4172/2157-7420.1000132

Copyright: © 2013 Raghupathi W, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

enabling doctors and medical professionals to do what they do best—provide optimal patient care [19].

This article provides an overview of this emerging discipline of health analytics [20,21]. It is organized as follows. Section 2 contains a definition of health analytics with examples. Section 3 describes the four stages of health analytics. This is followed by section 4 which explains the architectural framework. Section 5 describes the methodology. Section 6 contains examples of health analytic case studies in public health. Section 7 identifies the challenges and limitations. Finally, section 8 offers conclusions and future directions.

Health Analytics

HIMSS describes healthcare analytics as the “systematic use of data and related clinical and business (C&B) insights developed through applied analytical disciplines such as statistical, contextual, quantitative, predictive, and cognitive spectrums to drive fact-based decision making for planning, management, measurement and learning [22,23]”. Health analytics applications can be considered a “collection of decision support technologies for the health care provider aimed at enabling knowledge workers such as physicians, nurses and health officials, health policy makers and pharmacists to gain insight and make better and faster health decisions”. We propose another definition: “health analytics is the use of data, information technology, statistical analysis, quantitative methods, and mathematical computer-based models to help health care providers gain improved insight about these patients and make better, fact-based decisions adapted from [24]. Yet another view is of health analytics as a “way of transforming data into actions through analysis and insights in the context of the health care decision making and problem solving”. These definitions share a common goal: to gain insight for making informed healthcare decisions [20].

Historically, healthcare delivery organizations already have been applying descriptive analytics to cases. Using query and reporting tools and technologies, health care workers have gathered information on past performance, enabling classification and categorization of typically structured data. Now, data warehouses merge disparate data to create health dashboards, clinical data repositories and individual patient views. Health entities are moving toward predictive analytics, building on the capabilities of descriptive analytics to forecast future events using various models and what-if analyses. In the long term, the same entities will utilize prescriptive analytics to forecast possible outcomes and allow providers to make proactive decisions. Too, discovery analytics will help healthcare providers, pharmaceutical companies, and researchers identify unknown diseases and medical conditions and seek new or alternative treatments and drugs. These four types of health analytics are discussed in the next section in detail.

Examples of health analytics applications abound in the literature. IBM categorizes these applications in terms of advancing patient safety, improving clinical outcomes, and promoting wellness and disease management. Patient safety issues [23] arise mainly from medical errors and result in increased patient deaths and financial and legal consequences for the providers. A key objective for providers, therefore, is to effectively comply with care options and protocols, as well as compliance & regulation reporting requirements to predict and mitigate adverse episodes and minimize readmissions [25]. How, then, to collect and analyze safety and outcome data and track key performance metrics? The answer lies in health analytics. Structured as well as unstructured data reside in patient charts, physician notes, handwritten prescriptions and elsewhere. Health analytics have the potential to facilitate early alerts on adverse episodes and minimize

infections or outbreaks; allow physicians, nurses and others identify patients at risk for readmission; and anticipate medication allergies and thus prevent adverse drug reactions. Duke University Health System, for example, uses drill-down analytics of millions of clinical records to identify near-miss cases and develop predictive analytics that flag possibly high-risk future events automatically. Analytics also benchmark the best clinical practices, suggesting areas for training of hospital personnel on patient safety [25].

Another key objective for healthcare providers is improving the clinical outcomes of patients via treatments and protocols, an area of increasing importance given the advent of Accountable Care Organizations (ACOs). Health analytics has the potential to identify new and effective treatments and care protocols and assess compliance against benchmarks. The intention here is to maintain top rankings for reimbursement as well as gain understanding regarding patient outcomes [25]. Analytics objectives include the evaluation of the transition from a fee-for-service to a performance-based reimbursement model; development of panel data views; provision of insight to physicians and others to make informed decisions; and the monitoring of patient compliance with treatment protocols. The cardiac research program at California Pacific Medical Center in San Francisco has initiated numerous projects to promote treatments and survival rates of patients with cardio vascular disease.

Collectively, these projects generate and analyze huge amounts of medical data to treat patients, develop sophisticated cardiac risk models to improve outcomes with the overall goal to reduce patient stays and reduce costs. In another example, Southeast Texas Medical Associates (SETMA) is reportedly utilizing health analytics to reduce hospital readmission rates. This practice group analyzes clinical data to help it identify the causes of patient readmission and develop more comprehensive treatment and intervention care plans thus lowering the number of patients who return to the hospital. In just six months, SETMA reduced the number of hospital readmissions by 22% [25].

A third major objective is the promotion of wellness and disease management. Studies show patients with chronic conditions account for more than three quarters of health expenditures. More than 125 million Americans suffer from at least one chronic condition, and 75 million have two or more chronic conditions [25]. Disease management and wellness programs can improve the overall health of those with chronic diseases, resulting in lowered healthcare delivery costs. Health analytics can assist in the identification of populations of patients in different disease categories, thereby facilitating the management of patient wellness. For example, populations with a specific chronic disease, such as diabetes or heart disease, can be identified and monitored to prevent the development of other medical conditions associated with the disease.

Analytics can help identify the best candidates for wellness and disease management programs; determine the patient and clinical information needed to better promote wellness or manage diseases; and aggregate the information required to demonstrate better outcomes and qualify for bundled payments. An example reported in the literature is that of Seton Healthcare in central Texas [25]. The center identified which patients with congestive heart failure were most likely to be readmitted to the hospital. These patients were provided with proactive disease management to reduce cost and mortality rates resulting in improved quality of life. Seton utilizes analytics from unstructured data—physician notes and discharge summaries, for example—to extract key information. Sophisticated analytics techniques are then applied to gain clinical insight. Trends and patterns in patient care and outcomes

are identified by detecting correlations or disparities previously hidden (in plain sight) in accessible free-text files. This information allows Seton to improve disease management processes and prevent avoidable readmissions.

Another example of the practical use of health analytics is in the neo-natal intensive care unit at Toronto's Hospital for Sick Children. There, nurses chart a baby's heart rate every hour. A baby's heart beats almost 120 times a minute, and it is the pattern of those beats over time that provides signals as to whether something is wrong. With the help of IBM's Watson system, live streams of heartbeats are analyzed. Watson identified patterns revealing signs of infection 24 hours before the babies showed any visible symptoms. In premature babies, advancing treatment by as little as an hour can be life saving [26].

As the healthcare delivery industry transforms itself under the influence of regulatory changes, marketplace shifts, as well as the renewed focus on quality outcomes and need to reduce costs, healthcare providers will have to focus on the collection, storage, integration, and analysis of different types of health data. This focus in turn will result in greater insight and more effective patient outcome decisions while reducing the costs of health care delivery. The analytics phenomenon is expected to be pervasive across the health care delivery system including in the functions of performance & risk management, clinical decision support, acute and long-term care, chronic disease management, preventive medicine, and population health.

Health analytics can assist administrators, case managers, clinicians, and even patients in better understanding healthcare processes and outcomes across multiple facilities, providers, stakeholders, and regulatory agencies. Look, for example, at the care of diabetes patients. The de-identification of individual patient data and the analysis of alternative treatment protocols and outcomes across a large number of diabetes patients, leads the health care delivery organization to analytics that identify locations with the most successful patient outcomes. Instead of limiting a patient's health providers to one individual's data, health analytics takes in the landscape of insulin protocols, compliance, drug combinations and interactions, care models, and other useful data.

In addition to enhancing the quality of health care delivery, reducing medical errors and improving patient safety, by using health analytics, health care providers can contain costs, comply with mandatory reporting requirements, meet benchmarks and standards, monitor patient satisfaction, evaluate physician performance and offer incentives and promote health talent management. Furthermore, health analytics capabilities can facilitate dynamic fraud detection (e.g., insurance claims) and assist in behavior modification to improve healthier lifestyle choices.

Nearly every stakeholder and participant in the healthcare system (patient, physician, HMO, hospital, etc.) makes decisions, whether with regard to medical diagnosis and treatment or with respect to financial, insurance and/or case management. And these decisions likely impact both the patient and the overall health care delivery system, whether the decision involves cost, outcome, and/or quality of care. Considering the disparate sources of health data and the uncertain and imperfect nature of the data, decisions are often complex and difficult to make.

The onus is on stakeholders to have good information and good insight for making informed decisions that will impact patient care positively. Complicating health decisions today is the overwhelming amount of data available - that collected by providers and data found on the internet and through social media. The unwieldy amount of information is one reason why the emerging discipline of health

analytics matters greatly in today's health care delivery environment supported by such tools as advanced spreadsheet models, statistical software packages and more complex business intelligence & analytics software (such as Cognos, Hyperion, Business Objects, MicroStrategy, Teradata etc.).

Health Analytics Stages

Here we describe a four-stage model of health analytics. Health analytics begins with the collection, organization and manipulation of health and medical data [24,27] and is supported by the four major stages described below.

Descriptive analytics

Descriptive analytics is the most commonly used and most well understood type of analytics [24,27]. It was the earliest to be introduced and the simplest by far, being easy to implement and understand since it describes data "as is" without complex calculations. Descriptive analytics is more data-driven than the other models. Most health care entities start with descriptive analytics, using data to understand past and current health care decisions and make informed decisions. The models in descriptive analytics categorize, characterize, aggregate and classify data, converting it into useful information for understanding and analyzing health care decisions, outcomes and quality. Such data summaries might be in the form of meaningful charts and reports that, for example, illustrate patient hospitalizations, physician performance, utilization management, and the like. Descriptive analytics uses a lot of visualization. One could, for example, obtain standard and customized reports and drill down into the data, running queries to better understand the usage and effect of a new drug. Descriptive analytics helps answers such questions as, How many patients did we treat in each facility? What was our revenue and cost last quarter? How many and what types of medical conditions did we treat? Which facility has highest cost-quality ratios? What do my patients look like? Which patients should be targeted for drug (treatment) promotion (trial)? and How can one identify high-risk patients?

Predictive analytics

Predictive analytics is a slightly more advanced type of analytics and emphasizes the use of information (v. data) [24,27,28]. It looks at past performance in an effort to predict the future by examining historical or summarized health data, detecting patterns of relationships in these data, and then extrapolating these relationships to forecast. A physician might wish to predict the response of different patient groups to different drug (dosages) of reactions (clinical trials), for instance. Predictive analytics can anticipate risk and find relationships in health data not apparent with descriptive analytics alone. Advanced techniques such as data mining allow predictive analytics to detect hidden patterns in large quantities of data, necessary for segmenting and grouping data into coherent sets to predict behavior and detect trends. A health professional could ask, Which drugs should I use for the trial (treatment). Which drugs should I replenish in anticipation of an epidemic? Which of my patients are most likely to get well (based on a protocol)? When one drug fails, which others are most likely to fail too? Who is most likely to have another heart attack? What will happen if drug dosage is reduced, or a cocktail of drugs is given (for cancer treatment)? How do I predict health outcomes for my patients to improve my service? How can I predict need and allocate resources to ensure I am delivering health services effectively? These and other questions are typically answered through predictive analytics.

Prescriptive analytics

Prescriptive analytics come into play when health/medical problems involve too many choices or alternatives for a provider to effectively consider descriptive or predictive analytics [24,27]. Prescriptive analytics uses health and medical knowledge in addition to data or information. Prescriptive analytics is also normative, addressing the question of what should be. Prescriptive analytics is used in many areas of health care, including drug prescriptions and treatment alternatives. For example, one may determine the maximum dosage of the drug that is effective to maximize treatment outcome, or alternative surgical options can be considered weighing the pros and cons of each. Personalized medicine and evidence-based medicine are both supported by prescriptive analytics.

Discovery (Wisdom) analytics

Discovery analytics utilizes knowledge about knowledge, or wisdom, to discover new drugs (drug discovery), previously unknown diseases, alternative treatments, etc. In the case of discovering previously unknown facts such as correlation between a drug and its side effect, the application “learns” associations and flags the pharmaceutical company. Alternatively, the analysis may lead to the discovery of previously unknown diseases and medical conditions. In new drug discovery, the use of computer simulations with what-if analysis is emerging rapidly as an example of discovery analytics. Likewise is the use of computer simulations to augment clinical trials and speed up the study of efficacy of new drugs. Predicting the future in the context of such uncertain public health situations as epidemics is another example. Although the models and tools used in descriptive, predictive, prescriptive, and discovery analytics are different, many applications involve all four.

Architectural Framework

Figure 1 provides an overview of the health analytics architectural framework, highlighting the different components and how they work together to enable analytics. There are four key elements in the health analytics architectural framework as shown in the figure.

Health data sources

Health data are extracted from both internal and external data sources. Internal data sources include patient data, hospital data, diagnoses and treatment data. External data sources include benchmarks, publicly available data sets, as well as data from government agencies, the WHO and others that augment the internal data sources. The data over which analytics tasks are performed often come from diverse sources, including multiple clinical laboratories, radiologists, insurance companies and HMOs, patient databases, public health systems (e.g. CDC, HHS, WHO); there may be multiple providers and hospitals internally and externally.

The diverse sources contain data of varying quality and use inconsistent representations, codes and names, and these have to be reconciled. The problem of integrating, cleansing and standardizing data in preparation for analytics tasks presents a challenge. Additionally, data can come from multiple, disparate systems (e.g., electronic health record systems, clinical decision support systems, etc.) found in various locations within the wider health system. And health data may be either structured (quantitative) or unstructured (text). All these data need to be organized, structured and otherwise standardized in a way that readies it for analytics.

Health data transformation

For analytics purposes, data has to be pooled. The data are in a raw

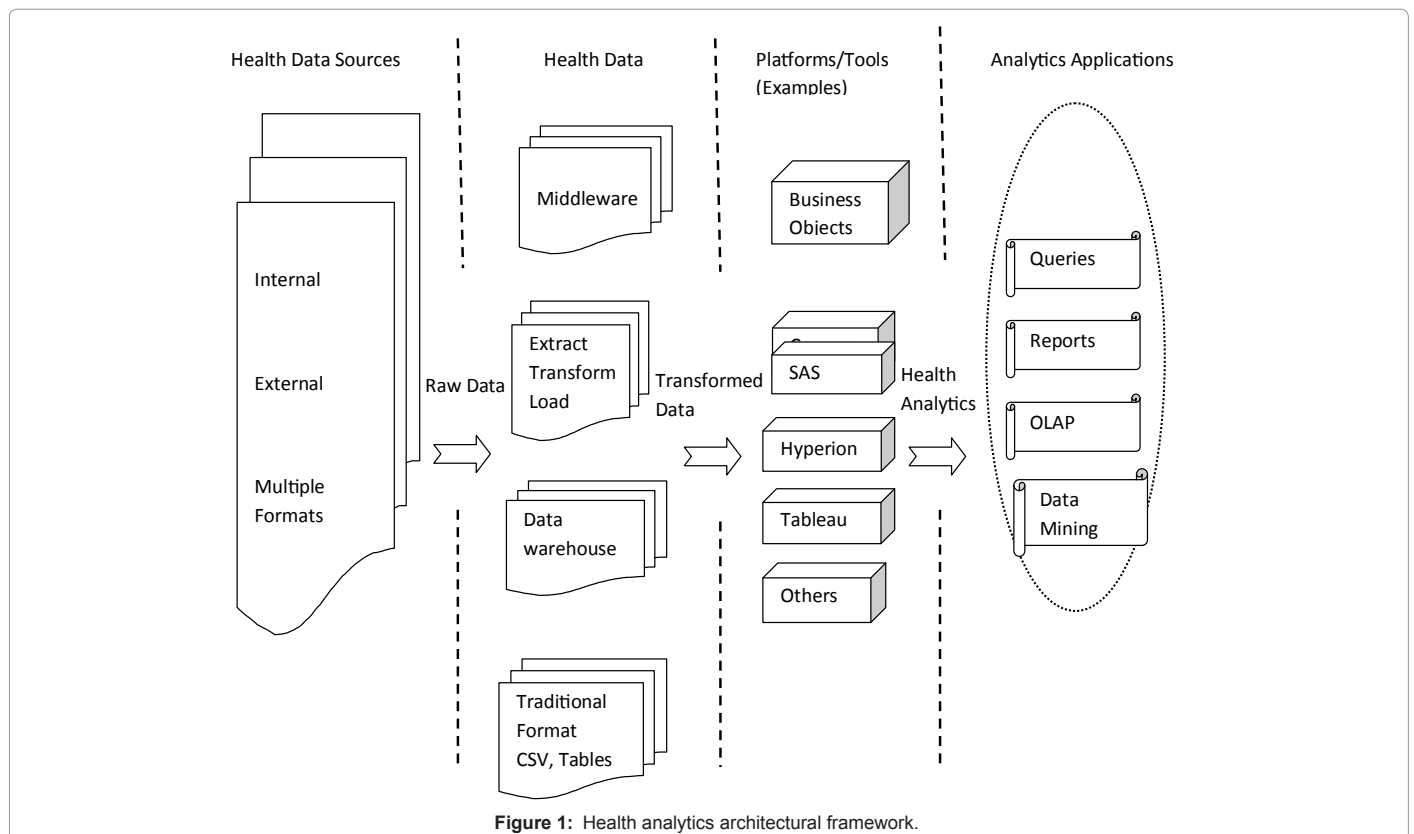


Figure 1: Health analytics architectural framework.

state and requires processing or transformation [3]. Several options are available. A service-oriented architectural approach combined with web services is one route. The data stays raw and services that perform the routine and standardized processes across the organization are used to call, retrieve and process the data. Another approach is data warehousing (middleware), whereby data from various sources is aggregated and made ready for processing. However, not all the data are available in real-time. Health analytics tasks usually need to be executed incrementally as new data arrive, as, for example, when data for discharged patients and closed cases are made available. Efficient and scalable data loading and refresh capabilities are, therefore, imperative for provider (health system) analytics, via the steps of extract, transform, and load (ETL), data from diverse sources are cleansed and readied [3].

Platform & tools

Once the data are cleansed and made ready for analytics, a suite of tools is utilized to perform the four types of analytics. These may range from statistical tools (e.g., SPSS, SAS, R, etc.) to more advanced business intelligence and data mining tools (Business Objects, Cognos, Hyperion, Tableau, etc.). Typically, after a needs and comparative analysis, the user chooses from several available vendors and tools. Usually, a combination of different types of tools is needed to perform the four types of analytics.

Health analytics applications

Using various tools, the four types of analytics can be performed in the context of queries, reports, online analytics processing (OLAP), and data mining. The example queries provided in the previous section can be answered using these tools. To answer such a query as “identify patients who have incurred medical expenses during the coverage year whose overall amount exceeds the average medical expense amount by at least 50%”, various tools can be used. Likewise, numerous reports required for health compliance and regulation and for monitoring key performance and outcome indicators can be generated and viewed as dashboards and scorecards. Patient diagnosis and treatment, real-time clinical decision support, can be provided with OLAP. Lastly, such data mining techniques as association, classification, and clustering provide advanced decision support.

Methodology

While several different methodologies are being developed in this rapidly emerging discipline, here we outline a practical hands-on methodology. Table 1 shows the main stages of this methodology. In Stage One, the interdisciplinary health analytics team develops a concept design. This is a first cut at establishing the need for such a project. A problem statement is followed by a description of the project’s significance. Developers will note that trade-offs include cheaper options, risk, problem-solution alignment, and so on. Once the concept design is approved in principle, the team can proceed to Stage Two, the proposal development stage. Here, more details are filled in. Taking the concept design as input, an abstract highlighting the overall methodology and implementation process is outlined. This is followed by an introduction to the health analytics domain: What is the problem being addressed? Why is it important and interesting to the organization? What is the case for an analytics approach? Because the complexity and cost of analytics techniques are significantly higher compared to traditional approaches, it is important to justify their use. The project team also should provide background information on the problem domain and identify prior projects and research conducted in this domain.

Stage One	Concept Design <ul style="list-style-type: none"> • Establish need for health analytics project • Define problem Statement • Why is project important and significant?
Stage Two	Proposal <ul style="list-style-type: none"> • Abstract <ul style="list-style-type: none"> - Summarize proposal • Introduction <ul style="list-style-type: none"> o What is problem being addressed? o Why is it important and interesting? o Why health analytics approach? • Background material <ul style="list-style-type: none"> o Problem domain discussion o Prior projects and research
Stage Three	Methodology <ul style="list-style-type: none"> • Hypothesis development • Data sources & collection • Variable selection (independent and dependent variables) • ETL and data transformation • Platform/Tool Selection • Analytic techniques • Expected results & conclusions • Policy implications • Scope & limitations • Future research • Implementation <ul style="list-style-type: none"> o Develop conceptual architecture <ul style="list-style-type: none"> -Show and describe component (e.g., Figure 1) -Show and describe analytic platform/tools o Execute steps in methodology o Import data o Perform various analytics using various techniques (queries, reports, analysis, data mining, etc.) o Gain insight from outputs o Draw conclusion o Derive policy implications o Make informed decisions
Stage Four	<ul style="list-style-type: none"> • Presentation and walkthrough • Evaluation

Table 1: Outline of health analytics methodology.

Both the concept design and the proposal are evaluated in terms of the four Cs:

- **Completeness**–Is the concept design complete?
- **Correctness**–Is the design technically sound? Is correct terminology used?
- **Consistency**–Is the proposal cohesive, or does it appear choppy? Is there flow and continuity?
- **Communicability**–Is the proposal formatted professionally? Does the report communicate design in easily understood language?

Next, in Stage Three, the steps in the methodology are fleshed out and implemented. The problem statement is broken down into a series of hypotheses. Note that the hypotheses are not quantitatively defined as in the case of statistical approaches. Rather, the hypotheses are developed to help guide the health analytics process. The independent and dependent variables are identified during this stage as well. In terms of the analytics, there is no great need to classify the variables, but it helps to identify causal relationships or correlations. The data sources as outlined in Figure 1 are identified; the data are collected (longitudinal data, if necessary), described, and transformed to make it ready for analytics. An important step at this point is platform/tool evaluation and selection. There are several options, including Business Objects, Cognos, Hyperion, and Tableau. A major consideration is whether the platform is available on a desktop or on the cloud. The next step is to

apply the various health analytics techniques to the data. These are not so different from the routine analytics; they're only scaled up to large data sets. Through a series of iterations and what-if analyses, insight is gained from the analytics, and from the insight, informed decisions can be made and policy shaped. In the final steps, conclusions are offered, scope and limitations are identified, and the policy implications discussed. In *Stage Four*, the project and its findings are presented to stakeholders for action.

The health analytics project is validated using the following criteria:

- Robustness of analyses, queries, reports, and visualization
- Variety of insight
- Substantiveness of research question
- Demonstration of health analytics application
- Some degree of integration among components
- Sophistication and complexity of analysis

The implementation is a staged approach with feedback loops built in at each stage to minimize risk of failure, and users should be involved. Implementation is also an iterative process, especially in the analytics step, wherein the analyst performs what-if analysis. The next section describes several health analytics projects undertaken in the public health arena at the Center for Digital Transformation at Fordham University.

Health Analytics Case Studies

Over the past three years, the Center for Digital Transformation

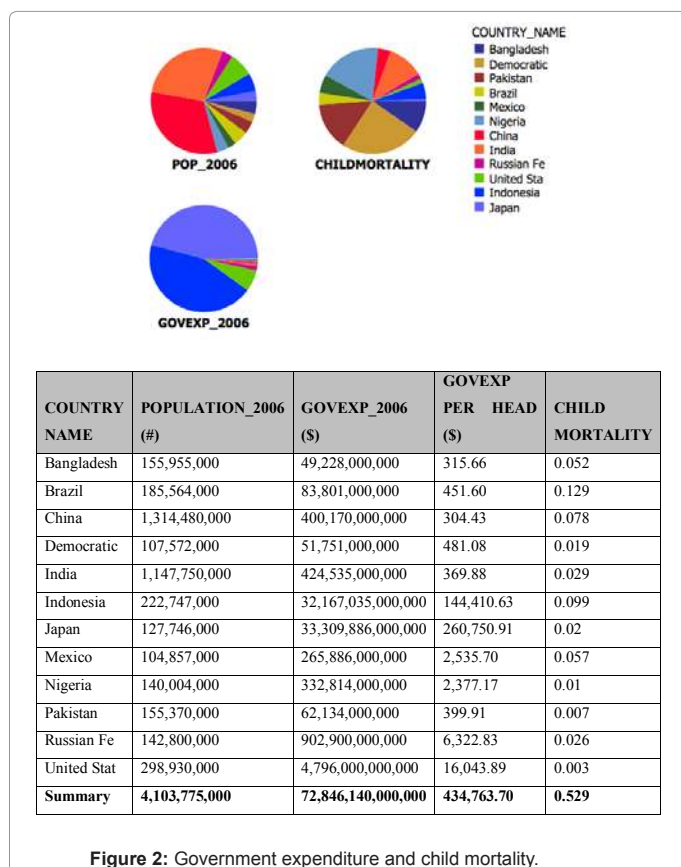


Figure 2: Government expenditure and child mortality.

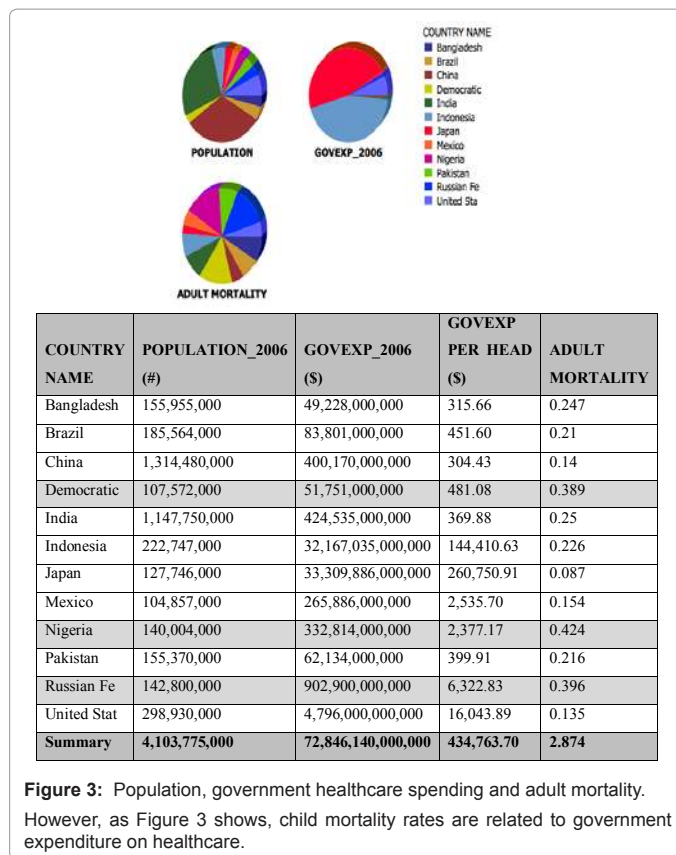


Figure 3: Population, government healthcare spending and adult mortality.

However, as Figure 3 shows, child mortality rates are related to government expenditure on healthcare.

at Fordham University has conducted several pilot health analytics studies in the public health arena. These studies were supported by IBM's Smarter Planet grants to the first author, but to be clear these are academic studies and IBM was in no way involved with the studies themselves. The Cognos 8.4 platform was used to conduct the health analytics, and the data for each study was downloaded from various public health sources into DB2 databases and analyzed using the query, report and analysis studios of Cognos. Below are excerpts of descriptions and sample outputs of each study conducted.

Global healthcare

In this study, national level economic data on government expenditure on healthcare, private expenditure on healthcare, gross domestic product (GDP), and the total population, was analyzed for a correlation with national health issues of infant and adult mortality and immunization availability Figure 2. The data relating to 100 countries in 2006 was extracted from the World Health Organization (WHO) and International Monetary Fund (IMF). Analytics showed that while countries with large population and high government spending per head (such as Nigeria and Russian Federation) had low infant mortality rates they did not necessarily have low adult mortality rates (Figure 3).

Behavioral habits and health risks

This study emphasizes the importance of preventive healthcare by identifying behavioral factors that may be linked to developing certain diseases. By recognizing these behavioral factors at an early stage and changing them, healthcare organizations can prevent the onset of potential diseases. Using analytics, the study investigates the relationship between the health variables of exercise, healthcare access/coverage, alcohol consumption and fruit/vegetable consumption with

developing diseases such as cholesterol, diabetes, and overweight/obesity (measured using Body Mass Index). The data was downloaded from the Center for Disease Control's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) Maps, a database that has tracked health conditions and behavioral risk factors for all states in the U.S. since 1984. The analyses show that, as expected, high exercise, high access to healthcare, high consumption of fruits/vegetables, and low consumption of alcohol, are all negatively correlated to the diseases of diabetes, cholesterol and obesity, as shown by the aggregate numbers in Figure 4. Additionally, the data reveals the disparity in healthcare between the Northern and Southern states in the U.S. and the fact that little has been done to reverse this trend over the seven-year analysis period. In terms of overall health, the southern states show the highest aggregate number of diseases (Figure 4).

Also, on a national level, the percentage of population without access to healthcare has not really reduced a great deal, as shown in Figure 5. On an average, Minnesota has the lowest percentage of residents (7.25%) without access to healthcare. Virgin Islands have the

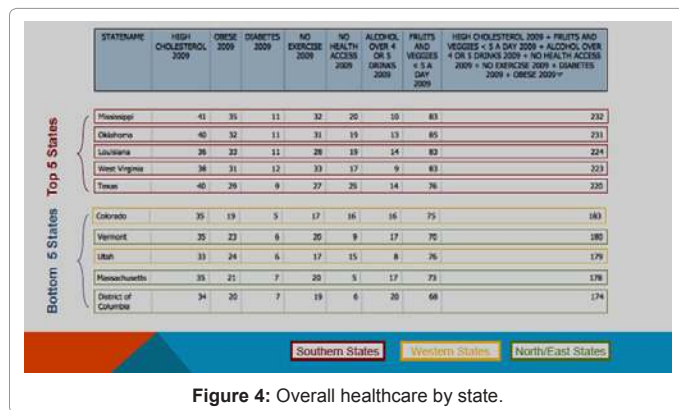


Figure 4: Overall healthcare by state.

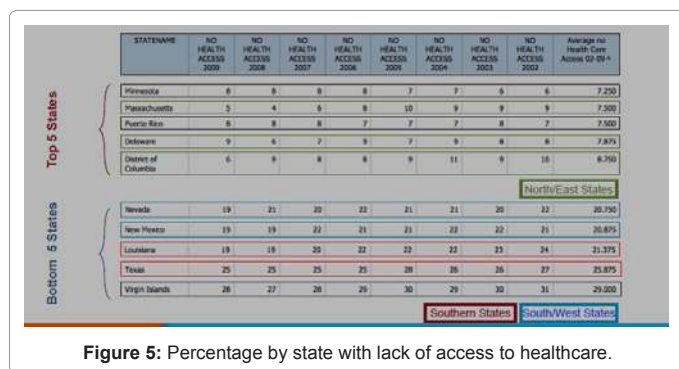


Figure 5: Percentage by state with lack of access to healthcare.

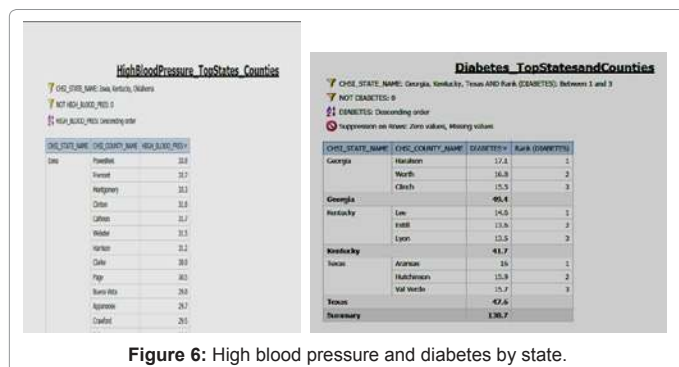


Figure 6: High blood pressure and diabetes by state.

Age Group	Brain & other Nervous System (%)	Breast (%)	Leukemia (%)	Melanoma of the skin (%)	Non-Hodgkin Lymphoma (%)	Prostate (%)
<1 year	37.33	1.22	42.98	6.90	9.74	1.83
1-4	30.46	0.11	57.21	2.58	8.40	1.24
5-9	38.41	0.20	41.37	4.92	14.50	0.61
10-14	33.78	1.21	37.39	8.45	18.95	0.21
15-19	24.88	3.39	30.73	18.88	21.11	1.00
20-24	19.22	7.01	21.04	31.91	20.26	0.56
25-29	16.02	19.89	13.75	32.55	17.13	0.65
30-34	10.82	36.62	9.44	27.57	14.81	0.74
35-39	7.15	50.39	6.95	21.52	12.78	1.21
40-44	5.03	57.59	5.28	16.85	10.61	4.65
45-49	4.16	55.72	4.69	13.36	9.74	12.33
50-54	3.49	46.16	4.48	10.46	9.07	26.35
55-59	4.13	14.71	5.88	11.42	11.22	52.65
60-64	2.68	33.17	4.56	7.30	8.51	43.77
65-69	2.44	28.27	4.79	6.61	8.33	49.56

Figure 7: Cancer incidence by age.

highest percentage (29%) of population without access to healthcare between the years 2002 to 2009. For future studies, one can consider benchmarking health care best practices for use in certain Southern states. Also, we can add more risk factors and diseases to the database as progress allows for more sophisticated analytics.

In another related study on behavioral factors and health risks, data from the U.S. Department of Health and Human Services' Community Health Status Indicators Project report was analyzed to investigate relationships between different behavioral factors (such as exercise, smoking, drug use and unemployment) and the health risks of developing diseases (such as diabetes, obesity, blood pressure, and depression). The data was collected for 2008. While not a 100% accurate predictor of disease given these behavioral factors, there are distinct similarities between the groupings of states by diseases and behaviors. Kentucky was in the top three states for three out of the four diseases of obesity, diabetes, and high blood pressure (Figure 6) and was rated highest for occurrences of smoking and lack of exercise. California, Texas, and New York were the top three states, respectively, for major depression and for occurrences of unemployment and recent drug use. In the drill down by behavior, the top states for unemployment were California, Texas, and New York; for smoking-Kentucky, Georgia and Indiana; for lack of exercise-Kentucky, Georgia and Missouri; and for recent drug use-California, Texas and New York.

Cancer incidences

In this study, data for the years 2002 to 2006 from the Center for Disease Control and Prevention's online database CDC Wonder was analyzed for relationships between six leading cancer types and variables such as age (Figure 7), sex, ethnicity, gender and incidence count.

Figure 7 shows that certain types of cancer such as brain and other nervous system cancer, leukemia, and non-Hodgkin lymphoma, are more prevalent in the younger age group below 20 years, while melanoma, breast cancer and prostate are more prevalent in the age group 25-59.

The analyses for gender shows that cancer is more among females (about 51%) as shown in Figure 8.

INCIDENCE_COUNT % of total (INCIDENCE_COUNT)			
SEX	CANCER_TYPE	INCIDENCE_COUNT	% of total (INCIDENCE_COUNT)
Female	Brain and Other Nervous System	40,118	2.01%
	Breast	754,092	37.79%
	Leukemias	50,422	2.53%
	Melanoma of the Skin	97,577	4.89%
	Non-Hodgkin Lymphoma	81,554	4.09%
Female		1,023,763	51.31%
Male	Brain and Other Nervous System	50,964	2.55%
	Breast	8,228	0.41%
	Leukemias	71,280	3.57%
	Melanoma of the Skin	112,960	5.66%
	Non-Hodgkin Lymphoma	106,538	5.34%
	Prostate	621,566	31.15%
	Male		971,536
Summary		1,995,299	100.00%

Figure 8: Cancer incidence by gender.

STATE	Brain & Other Nervous System %	Breast %	Leukemia %	Melanoma of the Skin %	Non-Hodgkin Lymphoma %	Prostate %
Alabama	4.26	37.97	5.49	9.58	8.93	33.78
Alaska	8.16	32.90	11.45	6.92	13.55	27.01
Arizona	7.98	54.30	9.23	16.39	12.11	0.00
Arkansas	4.44	32.94	5.81	7.13	9.07	40.60
California	5.63	53.86	7.92	15.64	12.65	4.30
Colorado	1.80	13.23	2.27	4.59	3.08	75.02
Connecticut	4.69	35.72	5.72	11.66	9.24	32.97
Delaware	5.42	33.29	6.59	11.80	9.99	32.91
District of Columbia	7.86	34.35	10.58	7.92	10.46	28.84
Florida	4.93	45.62	6.81	11.79	10.89	19.98
Georgia	4.03	37.07	5.54	11.93	8.24	33.20
Hawaii	5.40	41.47	7.18	13.16	9.62	23.17
Idaho	5.15	33.22	6.54	13.02	9.09	32.97
Illinois	4.02	39.51	5.91	8.34	9.00	33.22
Iowa	5.50	41.28	6.62	11.15	10.50	24.96
Indiana	5.13	37.26	6.88	10.41	9.79	30.51
Iowa	4.97	37.04	7.22	11.26	10.43	29.09
Kansas	4.76	36.93	6.44	12.00	9.64	30.24
Louisiana	3.92	35.76	6.00	7.68	10.05	36.60
Maine	4.33	37.00	6.48	11.84	9.32	31.03
Maryland	3.95	38.98	4.61	11.84	8.12	32.51
Massachusetts	4.55	33.99	5.33	11.76	8.85	35.52
Michigan	4.17	35.72	5.82	9.62	9.33	35.35
Minnesota	4.70	34.61	6.91	10.57	9.40	33.81
Mississippi	4.98	35.42	6.10	8.57	8.83	36.09
Missouri	4.85	39.42	5.95	9.23	9.38	31.17
Montana	5.29	32.69	7.15	10.00	9.41	35.45
Nebraska	4.88	37.34	7.52	10.09	10.65	29.52
Nevada	5.51	34.79	7.26	10.52	10.33	31.59
New	3.96	36.21	6.28	14.52	9.57	29.47

Figure 9: Cancer incidence by state.

Additionally, drill down features of Cognos reveals that the incidence of certain types of cancer such as leukemia, melanoma of the skin and non-Hodgkin lymphoma aggregate to about 12% in women and about 15% in men (Figure 8). Analytics also shows the distribution of different cancer types across states (location) (Figure 9). The most prevalent cancers in most states were breast cancer and prostate cancer.

Incidence Count		Female (#)	Male (#)	Summary	Female (%)	Male (%)
Brain & Other Nervous System	American Indian or Alaska Native	1,000	1,064	2,064	48.45	51.55
	Asian or Pacific Islander	2,384	2,645	5,029	47.41	52.59
	Black or African American	4,282	4,575	8,857	48.35	51.65
	Other Races and Unknown Combined	1,936	2,184	4,120	46.99	53.01
	White	30,516	40,496	71,012	42.97	57.03
Brain & Other Nervous System		40,118	50,964	91,082	44.05	55.95
Breast	American Indian or Alaska Native	4,519	96	4,615	97.92	2.08
	Asian or Pacific Islander	25,698	584	26,282	97.78	2.22
	Black or African American	83,049	1,556	84,605	98.16	1.84
	Other Races and Unknown Combined	10,171	480	10,651	95.49	4.51
	White	630,655	5,512	636,167	99.13	0.87
Breast		754,092	8,228	762,320	98.92	1.08
Leukemia	American Indian or Alaska Native	1,296	1,472	2,768	46.82	53.18
	Asian or Pacific Islander	2,945	3,713	6,658	44.23	55.77
	Black or African American	5,847	7,027	12,874	45.42	54.58
	Other Races and Unknown Combined	2,451	3,162	5,613	43.67	56.33
	White	37,883	55,906	93,789	40.39	59.61
Leukemia		50,422	71,280	121,702	41.43	58.57
Melanoma of the Skin	American Indian or Alaska Native	912	960	1,872	48.72	51.28
	Asian or Pacific Islander	1,287	1,038	2,325	55.35	44.65
	Black or African American	1,736	1,568	3,304	52.54	47.46
	Other Races and Unknown Combined	6,398	6,459	12,857	49.76	50.24
	White	87,244	102,935	190,179	45.87	54.13
Melanoma of the Skin		97,577	112,960	210,537	46.35	53.65
Non-Hodgkin Lymphoma	American Indian or Alaska Native	1,213	1,498	2,711	44.74	55.26
	Asian or Pacific Islander	3,669	4,249	7,918	46.34	53.66
	Black or African American	8,825	11,213	20,038	44.04	55.96
	Other Races and Unknown Combined	2,797	3,315	6,112	45.76	54.24
	White	65,050	86,263	151,313	42.99	57.01
Non-Hodgkin Lymphoma		81,554	106,538	188,092	43.36	56.64
Prostate	American Indian or Alaska Native		2,988	2,988		100.00
	Asian or Pacific Islander		9,778	9,778		100.00
	Black or African American		99,354	99,354		100.00
	Other Races and Unknown Combined		19,517	19,517		100.00
	White		489,929	489,929		100.00
Prostate		621,566	621,566	621,566	100.00	0
Summary		1,023,763	971,536	1,995,299		

Figure 10: Cancer incidence by ethnicity.

While it remains true that the incidence of cancer may be proportional to the population of each state, the analysis offers a ball park to healthcare experts with regard to the magnitude of the prevalence. Also, this is only an exploratory study and future studies can perform a comprehensive analysis on the distribution by location.

In Figure 10, for melanoma of the skin, prevalence is in whites (54.13%) followed by American Indian or Alaskan native (51.28%). The prevalence of leukemia is fairly close among the different ethnicities with whites having the highest (59.6%) followed by other races and unknown combined (56.33%), Asian or Pacific islander (55.77%), Black or African American (54.58%), and American Indian or Alaskan native (53.18%).

Demographic characteristics and childhood immunization rates

This study investigates the relationship in the U.S. between the socio-demographic characteristics (such as race, income, public/private healthcare, healthcare costs, and state/region of residence) and childhood immunization rates. The analysis also considered the rate of parents opting out of recommended immunizations for their children (opt-out rate). Data for the years 1999 to 2010 was downloaded from the CDC. Analysis addressed several questions, including whether a relationship exists between certain demographic characteristics and the choice to opt-out of these recommended childhood immunizations and whether a relationship exists between immunization levels and various demographic characteristics, and healthcare costs. The results show that areas with high opt-out rates have higher median incomes (Figure 11).

Most races followed the same trend over the 12-year period, except for Native Americans (Figure 12). All races except for Whites and

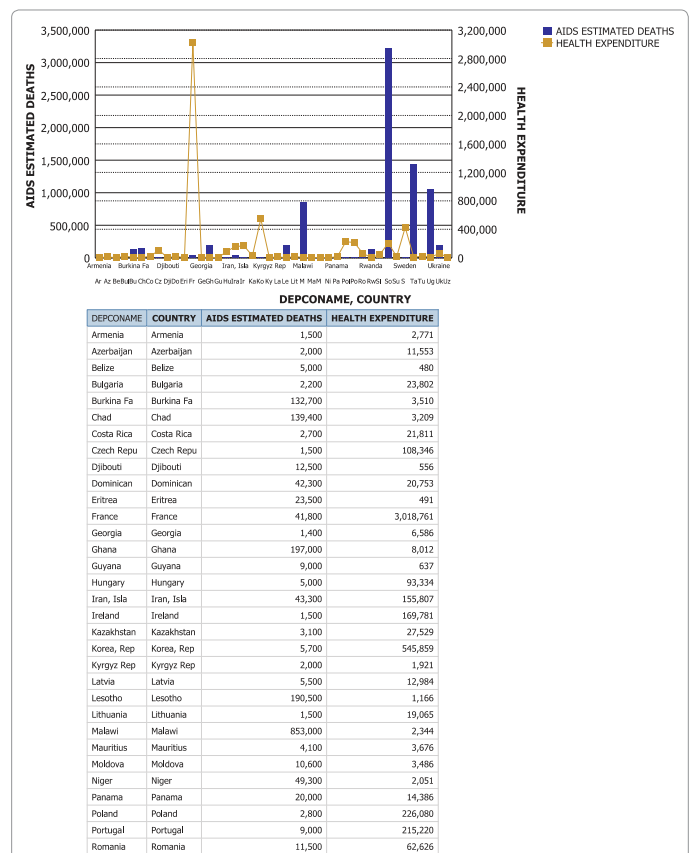


Figure 13: Health expenditures and Estimated AIDS related deaths.

Asians showed a net positive MMR vaccination rate. In the analysis for income and immunization rates, it showed that there was an increase in the percentage of children below the poverty line receiving the MMR vaccine. The analysis for healthcare costs by state shows the highest rates for Alaska, Montana and Kentucky.

Disease and quality of life

This study investigates the correlation between the quality of life in various countries and the levels of disease in those countries. Countries were grouped as developed, developing, and undeveloped. Data for the years 1996 through 2009 was collected from the World Bank website (<http://www.worldbank.org>). Data on health, nutrition, and population in relation to quality of life indices was collected for each of the categories. The data was then analyzed using Cognos 8.4, to identify correlations between wealth and education levels of a country and the impact on the health of its citizens. It was generally expected that high health expenditures, high secondary school enrollment, and high labor force percentages would positively correlate with lower levels of disease, and lower levels of disease-related deaths. Figure 13 shows the analysis for health expenditures and Aids related deaths. The aggregate of health expenditures (as a percentage of GDP) and the aggregate of Aids estimated deaths (per 100,000 people) for the time period are shown. Results show that countries that have high expenditure in healthcare have the lowest Aids estimated deaths (e.g., France, Sweden, Portugal and Bulgaria), and countries that have low health expenditures have much higher Aids estimated deaths (e.g., Burkina, Chad, Lesotho and Malawi). Although it is possible for an extraneous factor such as the country's population to skew the numbers and interfere with these

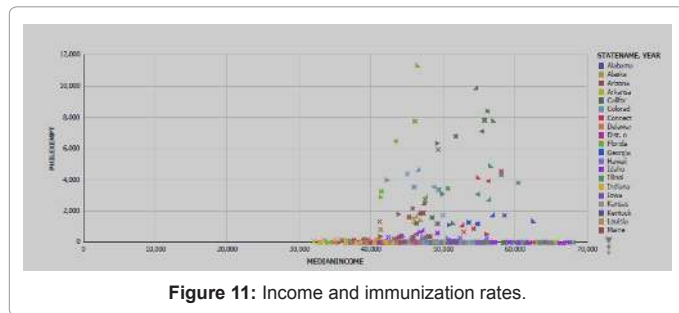


Figure 11: Income and immunization rates.

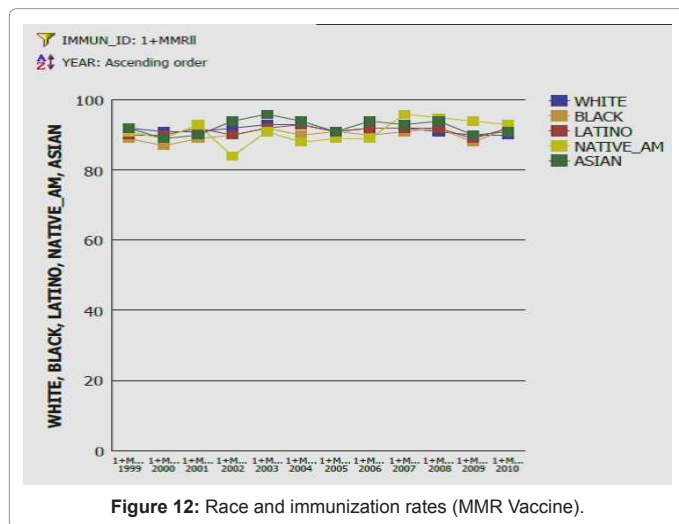


Figure 12: Race and immunization rates (MMR Vaccine).

	TUBDEATHRATE	SECONDARYSCHOOLEN
Armenia	148	1,160
Azerbaijan	205	1,230
Belize	96	857
Bulgaria	50	1,359
Burkina Fa	567	141
Chad	916	200
Costa Rica	21	1,038
Czech Repu	6	1,384
Djibouti	935	260
Dominican	172	762
Eritrea	105	345
France	4	1,658
Georgia	80	1,138
Ghana	514	496
Guyana	204	1,047
Hungary	36	1,433
Iran, Isla	55	1,008
Ireland	0	1,661
Kazakhstan	364	1,209
Korea, Rep	120	1,467
Kyrgyz Rep	268	1,102
Latvia	154	1,385
Lesotho	191	513
Lithuania	173	1,444
Malawi	348	401
Mauritius	10	1,119
Moldova	273	1,180
Niger	569	114
Panama	82	897
Poland	27	1,291
Portugal	25	1,446
Romania	126	1,242
Rwanda	765	164
Slovak Rep	9	1,342
South Afri	641	1,062
Sudan	219	330
Sweden	0	1,907
Tanzania	209	150
Tunisia	18	1,156
Uganda	415	229
Ukraine	261	1,054
Uzbekistan	211	1,145
Average(INDCONAME)	228.38095238	965.38095238

Figure 14: Tuberculosis and secondary school enrollment.

correlations, it can be seen that the more the investment in healthcare the lower the number of disease related deaths (Figure 13).

In the analysis for secondary school enrollment and tuberculosis related deaths (Figure 14), the results proved inconclusive. The overall figures for the years for the countries, is displayed. Looking at the trend, while countries like Sweden, France, Ireland and the Czech Republic showed high levels of school enrollment and a low level of tuberculosis-related deaths, others such as Kazakhstan and Kyrgyz had high levels of school enrollment but also high levels of tuberculosis-related deaths, as shown in the figure.

In analyzing the correlation between labor force and Aids related deaths (per 100,000), Figure 15 shows that countries that have below average numbers of AIDS related deaths also have below average numbers of labor force. Also, it is interesting to note the outliers with an above average number of AIDS related deaths (such as South Africa and Tanzania). More than 75% of all the countries fall below the average. The top three countries that experience the highest number of AIDS

related deaths (South Africa, Tanzania, and Uganda) also experience above average numbers of labor force. There is, therefore, a positive correlation between AIDS related deaths and the labor force.

Challenges

The selection of a suitable implementation platform is a major challenge [3]. It must support, at a minimum, the key functions necessary for processing the data. Criteria for platform evaluation may include availability, continuity, ease of use, scalability, ability to manipulate at different levels of granularity, privacy and security enablement, and quality assurance. In order to take off, health analytics applications need to be shrink-wrapped, user-friendly and transparent. And the lag between data collection and processing should be addressed; health data needs to be analyzed in real-time. Meanwhile, the dynamic availability of numerous analytics algorithms, models and methods in a pull-down type of menu is also necessary for large-scale adoption. And the various options of local processing (e.g. a network, desktop/laptop), cloud computing, software as a service (SaaS), and SOA-web services delivery mechanisms require further exploration. Ownership, governance and standards are key managerial issues to be addressed [5], as do the issues of continuous data acquisition and data cleansing. In the future, ontology and other design issues must be

	AIDSESTDATHS	LABORFORCEPER
Armenia	1,500	19,792,280
Azerbaijan	2,000	56,236,398
Belize	5,000	1,269,888
Bulgaria	2,200	53,625,719
Burkina Fa	132,700	69,290,721
Chad	139,400	51,561,465
Costa Rica	2,700	26,321,548
Czech Repu	1,500	77,266,608
Djibouti	12,500	4,449,617
Dominican	42,300	47,058,379
Eritrea	23,500	22,536,897
France	41,800	418,357,136
Georgia	1,400	31,814,779
Ghana	197,000	104,265,913
Guyana	9,000	3,295,995
Hungary	5,000	62,813,210
Iran, Isla	43,300	319,096,262
Ireland	1,500	27,952,246
Kazakhstan	3,100	102,331,575
Korea, Rep	5,700	347,176,466
Kyrgyz Rep	2,000	28,541,031
Latvia	5,500	17,315,478
Lesotho	190,500	13,136,638
Lithuania	1,500	25,014,029
Malawi	853,000	70,751,052
Mauritius	4,100	7,492,858
Moldova	10,600	21,949,673
Niger	49,300	56,681,044
Panama	20,000	17,888,309
Poland	2,800	225,907,210
Portugal	9,000	74,821,104
Romania	11,500	162,904,771
Rwanda	124,900	48,152,315
Slovak Rep	1,500	39,071,700
South Afri	3,220,000	200,111,226
Sudan	68,800	126,655,187
Sweden	2,500	70,534,881
Tanzania	1,430,000	250,347,846
Tunisia	1,500	50,280,214
Uganda	1,053,000	151,999,991
Ukraine	198,700	254,795,472
Uzbekistan	2,600	123,946,849
Average(INDCONAME)	188,961.9047619	92,495,528.0952381

Figure 15: Labor force and AIDS related deaths.

discussed. Furthermore, an appliance-driven approach (e.g. access via mobile computing and wireless devices) has to be promoted.

Conclusion

In this article we have provided an overview of health analytics. While the concepts and algorithms behind much of the analytics have been around for many decades, it is only recently that the advanced and sophisticated information technology required for complex processing has been made available. Health analytics offer a panoramic view of the health data from which multidimensional analysis and insight can be obtained, leading to informed decisions regarding healthcare. In addition, the methods are assumption-free and can serve as inputs to more rigorous statistical tools and techniques. There are other health analytics technologies, such as web analytics, that enable understanding of how visitors to a health entity web site (e.g., an HMO's site) interact with the pages and features. Which "landing pages," for example, are likely to encourage the visitor (patient) to seek medical service (care). Another nascent but important area is mobile (health) analytics that presents opportunities for facilitating novel and rich analytics applications for health knowledge to work on mobile devices. Although in its early stages, health analytics has great potential to improve outcomes and quality of healthcare and reduce costs of healthcare delivery.

References

1. Caban JJ, Gotz D (2011) Workshop on visual analytics in healthcare: understanding the physician perspective. *SIGHIT Record* 2: 29-31.
2. Raghupathi W, Tan J (2008) Information Systems and Healthcare XXX: Charting a Strategic Path for Health Information Technology. *Communications of the AIS* 23: 501-522.
3. (2012) The clinical data pipeline-fueling analytics for researchers and physicians from a single source. *Deloitte*.
4. (2013) Impacting the Cost and Quality of Healthcare. *Elsevier Clinical Decision Support*. Elsevier.
5. Raghupathi W (2007) Designing clinical decision support systems in health care: a systemic view. *International Journal of Healthcare Information Systems and Informatics* 2: 44-53.
6. (2011) Medical imaging. *Accenture*.
7. Kearney AT (2013) A healthy dose of data analytics. *Ideas and Insights*.
8. (2012) The promise of healthcare analytics. *HealthLeaders Media Breakthroughs*.
9. Hersh W (2012) From implementation to analytics: the future work of informatics.
10. (2013) [iHT² Report] Analytics: The Nervous System of IT-Enabled Healthcare. *IHTT*.
11. (2011) Accenture UK Health: health analytics. *Accenture*.
12. (2012) Transforming healthcare through better use of data. *CSC*.
13. (2012) Accelerating collaborative healthcare with EMC solutions. *EMC2*.
14. Horner P, Basu A (2012) Analytics & the future of healthcare. *Analytics & the future of healthcare*.
15. Taylor J (2010) Transforming healthcare delivery with analytics. *Decision Management Solutions*.
16. McAullay D, Williams G, Chen J, Jin H, He H, et al. (2005) A delivery framework for health data mining and analytics.
17. Stephan PK (2010) Healthcare Informatics: Improving Efficiency and Productivity 279.
18. Oracle enterprise healthcare analytics. *Oracle Health Sciences*.
19. Potter D (2010) Healthcare analytics: The crystal ball to improving quality of care. *Canadian Healthcare Technology*.
20. (2012) The value of analytics in healthcare. *IBM Global Business Services*.
21. (2012) The changing role of analytics for health care providers. *Sage Growth Partners*.
22. (2013) Clinical & business intelligence: an analytics executive review. *HIMSS*
23. Sulkers P (2011) Healthcare analytics: a game-changer for North York General Hospital. *Canadian Healthcare Technology*.
24. Evans JR (2013) Business Analytics – methods, models, and decisions.
25. Using analytics and collaboration to improve healthcare quality and outcomes. *IBM*.
26. Dembosky A (2012) Data prescription for better healthcare. *Financial Times*.
27. Perer A (2012) Healthcare analytics for clinical and non-clinical settings. *CHI'12*.
28. SAIC (2011) Predictive analytics for healthcare.