An Overview of Health Analytics

Wullianallur Raghupathi, Ph.D, Fordham University, New York, NY, USA

Viju Raghupathi, Ph.D, Brooklyn College, City University of New York, New York, NY, USA

Keywords: data warehousing, ETL, descriptive analytics, discovery analytics, health analytics, informed decision, insight, predictive analytics, prescriptive analytics

Summary

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.

Correspondence to:
W. Raghupathi
Graduate School of Business
Fordham University
113 W. 60th Street.
New York, NY, 10023
USA
E-mail: raghupathi@fordham.edu
1. 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 [4, 23]. 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 [6], clinical decision support systems [8, 22], medical imaging [1], public health databases, and the proprietary systems of health care providers (e.g., physicians, insurance companies, HMOs, hospitals, government agencies, and others). Data is 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 [3, 11, 12, 17], thereby improving quality of care and reducing costs [2, 5, 9, 14]. Furthermore, the
positive influence on health outcomes, thanks to sophisticated decision support
[17], propels the practice of evidence-based medicine [28], personalized medicine,
and improved public health. While aspects of such health analytics as the use of
statistical models, data mining [18, 24], 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 [19]. 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 [21]. 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 enabling doctors and medical professionals to do what they do best - provide optimal patient care [21].
This article provides an overview of this emerging discipline of health analytics [16, 25]. First, we define and discuss health analytics with examples. Then we describe the four stages of health analytics applications. Third, we explain the architectural framework of health analytics. We follow this section with a description of the health analytics application development process. Fifth, we provide examples of health analytics in public health. Sixth, we identify the challenges and limitations, and finally, we offer conclusions and future directions.

2. 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 [13, 27].” 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.
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 [16].

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 [27] 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 [15]. 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 [15].

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 [15]. 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 cardiovascular 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% [15].

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. 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 [15]. 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 examples - 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 [7].

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.).

3. 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 [10, 20] and is supported by the four major stages described below.

1. Descriptive analytics is the most commonly used and most well understood type of analytics [10, 20]. 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?

2. Predictive analytics is a slightly more advanced type of analytics and emphasizes the use of information (v. data) [10, 20, 26]. 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.

3. **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 [10, 20]. 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.

4. Discovery (wisdom) 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 (e.g., correlation between a drug and 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, 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.
4. 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:

![Health Analytics Architectural Framework Diagram](image)

**Fig. 1 Health Analytics Architectural Framework**

1. *Health data sources*

Health data is 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 this data needs to be organized, structured and otherwise standardized in a way that readies it for analytics.

2. Health data transformation

For analytics purposes, data has to be pooled. The data is in a raw state and requires processing or transformation [6]. 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 is available in real-time. Health analytics tasks usually need to be executed incrementally as new data arrives, 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 is cleansed and readied [6].

3. Platform & tools

Once the data is 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.

4. 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.

5. 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.

Table 1. Outline of Health Analytics Methodology

<table>
<thead>
<tr>
<th>Stage One</th>
<th>Concept Design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Establish need for health analytics project</td>
</tr>
<tr>
<td></td>
<td>• Define problem Statement</td>
</tr>
<tr>
<td></td>
<td>• Why is project important and significant?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage Two</th>
<th>Proposal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Abstract</td>
</tr>
<tr>
<td></td>
<td>- Summarize proposal</td>
</tr>
<tr>
<td></td>
<td>• Introduction</td>
</tr>
<tr>
<td></td>
<td>- What is problem being addressed?</td>
</tr>
<tr>
<td></td>
<td>- Why is it important and interesting?</td>
</tr>
<tr>
<td></td>
<td>- Why health analytics approach?</td>
</tr>
<tr>
<td></td>
<td>• Background material</td>
</tr>
<tr>
<td></td>
<td>- Problem domain discussion</td>
</tr>
<tr>
<td></td>
<td>- Prior projects and research</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage Three</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Hypothesis development</td>
</tr>
<tr>
<td></td>
<td>• Data sources &amp; collection</td>
</tr>
<tr>
<td></td>
<td>• Variable selection (independent and dependent variables)</td>
</tr>
</tbody>
</table>
- ETL and data transformation
- Platform/Tool Selection
- Analytic techniques
- Expected results & conclusions
- Policy implications
- Scope & limitations
- Future research
- Implementation
  - Develop conceptual architecture
    - Show and describe component (e.g., Figure 1)
    - Show and describe analytic platform/tools
  - Execute steps in methodology
  - Import data
  - Perform various analytics using various techniques (queries, reports, analysis, data mining, etc.)
  - Gain insight from outputs
  - Draw conclusion
  - Derive policy implications
  - Make informed decisions

Stage Four
- Presentation and walkthrough
- Evaluation

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 is 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.

1. Health Analytics Case Studies

Over the past three years, the Center for Digital Transformation 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. 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 had low infant mortality rates, they did not necessarily have low adult mortality rates (Figure 2).

![Fig. 2 Population, government healthcare spending and adult mortality](image)

Also, countries with high GDP did not necessarily have low infant mortality. As seen in Figure 3, the United States has a GDP that is three times that of Japan, but its infant mortality is also higher. The results show that high economic indicators do not necessarily correlate positively with public health related issues.
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. 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 of diseases (Figure 4).

![Table showing health data by state](image)

Fig. 4 Overall healthcare by state

Also, on a national level, the majority of states have not reduced the percentage of the population that lacks access to healthcare (Figure 5). 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.

![Fig. 5 Lack of access to healthcare by state](image)

**Behavioral Factors and Health Risks**

In a related study, 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, sex, ethnicity, gender and incidence count. The results show that cancer is more prevalent in the age group 50 to 69 (Figure 7) and among females (Figure 8).
Additionally, drill down features of Cognos reveal that breast cancer is more prevalent in females and prostate cancer is more prevalent in males. Also, analytics shows the distribution of different cancer types across states (location) (Figure 9).
The most prevalent cancer types in most states were breast cancer and prostate cancer. Figure 10 shows the cancer incidence by ethnicity in that cancer is most prevalent in Whites.
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 1999 to 2010 period 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 Asians showed a net positive MMR vaccination rate over the period studied. 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 that those with the highest rates are 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 analyzed and grouped as developed, developing, or undeveloped, and data for the 15-year period from 1995 to 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 of developed, developing and undeveloped countries. 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 higher life expectancy, lower levels of disease, and lower levels of disease-related deaths. Results show positive correlations between secondary school enrollment and life expectancy (Figure 13), and a negative correlation between healthcare expenditure and Aids-estimated deaths (Figure 14).

Fig. 13 School enrollment and life expectancy
Interestingly, further analyses, using the drill down functionality of Cognos, reveal that in terms of specific disease-related deaths, such as from tuberculosis, the results proved inconclusive. Some countries, such as Kazakhstan and Kyrgyz, had high levels of school enrollment but still showed high levels of tuberculosis-related deaths (Figure 15), while others, including Sweden, France, Ireland and the Czech Republic, showed a high level of school enrollment and a low level of tuberculosis-related deaths.
Fig. 15 Tuberculosis and school enrollment
Contrary to expectations, the analysis for labor force shows that countries with high levels of labor force also show high levels of AIDS-related deaths (Figure 16).

<table>
<thead>
<tr>
<th>AIDSESTDEATHS</th>
<th>LABORFORCEPER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Armenia</td>
<td>1,500</td>
</tr>
<tr>
<td>Azerbaijan</td>
<td>2,000</td>
</tr>
<tr>
<td>Belize</td>
<td>5,000</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>2,200</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>132,700</td>
</tr>
<tr>
<td>Chad</td>
<td>139,400</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>2,700</td>
</tr>
<tr>
<td>Czech Rep</td>
<td>1,500</td>
</tr>
<tr>
<td>Djibouti</td>
<td>12,500</td>
</tr>
<tr>
<td>Dominican</td>
<td>42,300</td>
</tr>
<tr>
<td>Eritrea</td>
<td>23,500</td>
</tr>
<tr>
<td>France</td>
<td>41,800</td>
</tr>
<tr>
<td>Georgia</td>
<td>1,400</td>
</tr>
<tr>
<td>Ghana</td>
<td>197,000</td>
</tr>
<tr>
<td>Guyana</td>
<td>9,000</td>
</tr>
<tr>
<td>Hungary</td>
<td>5,000</td>
</tr>
<tr>
<td>Iren, Isla</td>
<td>43,300</td>
</tr>
<tr>
<td>Ireland</td>
<td>1,500</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>3,100</td>
</tr>
<tr>
<td>Korea, Rep</td>
<td>5,700</td>
</tr>
<tr>
<td>Kyrgyz Rep</td>
<td>2,000</td>
</tr>
<tr>
<td>Latvia</td>
<td>5,500</td>
</tr>
<tr>
<td>Lesotho</td>
<td>190,500</td>
</tr>
<tr>
<td>Lithuania</td>
<td>1,500</td>
</tr>
<tr>
<td>Malawi</td>
<td>853,000</td>
</tr>
<tr>
<td>Mauritius</td>
<td>4,100</td>
</tr>
<tr>
<td>Moldova</td>
<td>10,600</td>
</tr>
<tr>
<td>Niger</td>
<td>49,300</td>
</tr>
<tr>
<td>Panama</td>
<td>20,000</td>
</tr>
<tr>
<td>Poland</td>
<td>2,800</td>
</tr>
<tr>
<td>Portugal</td>
<td>9,000</td>
</tr>
<tr>
<td>Romania</td>
<td>11,500</td>
</tr>
<tr>
<td>Rwanda</td>
<td>124,900</td>
</tr>
<tr>
<td>Slovak Rep</td>
<td>1,500</td>
</tr>
<tr>
<td>South Afri</td>
<td>3,220,000</td>
</tr>
<tr>
<td>Sudan</td>
<td>68,800</td>
</tr>
<tr>
<td>Sweden</td>
<td>2,500</td>
</tr>
<tr>
<td>Tanzania</td>
<td>1,430,000</td>
</tr>
<tr>
<td>Tunisia</td>
<td>1,500</td>
</tr>
<tr>
<td>Uganda</td>
<td>1,083,000</td>
</tr>
<tr>
<td>Ukraine</td>
<td>198,700</td>
</tr>
<tr>
<td>Uzbekistan</td>
<td>2,600</td>
</tr>
</tbody>
</table>

Average(INDCONNAME): 188,961,9047619 92,495,528,0952381

Fig. 16 Labor force and AIDS related deaths
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 discussed. Furthermore, an appliance-driven approach (e.g. access via mobile computing and wireless devices) has to be promoted.

7. 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 enabling novel and rich analytics applications for health knowledge works 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. Accenture. Medical imaging; 2011a (updated 2011; cited 2013-06-17)
http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture_UK_Health_Analytics.pdf

http://www.atkearney.com/documents/10192/1044787/A+Healthy+Dose+of+Data+Analytics.pdf/e3830106-cfb7-4d4b-886a-60972770de8f


5. CSC. Transforming healthcare through better use of data; 2012 (updated 2012; cited 2013-06-26) Available from: 


http://www.ft.com/intl/cms/s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz2W9cuwajK


9. EMC. Accelerating collaborative healthcare with EMC solutions; 2012 (updated 2012; cited 2013-06-17) Available from:


