ABSTRACT

Patients suffer the same form of illness, but their responses to therapy may differ greatly. This can be attributed to their genetics or other environmental factors. Personalized Medicine may improve the outcomes of treatment by understanding the individuals, causes of their disease, and the best treatment for their cure.

KEYWORDS: Personalized medicine, Clinical data mining and warehousing, Genetics, Pharmagenomics, Data security

INTRODUCTION

The Health Information Technology for Economic and Clinical Health (HITECH) Act and the Affordable Care Act (ACA) established a number of programs intended to accelerate the transformation of the United States Health care delivery system. Some of these programs include the adoption of the Electronic Health Record and its meaningful use, Participation in Accountable Care Organization and the use of the Clinical Decision Support (CDS) system to improve care at the point of service. These programs and initiatives are major drivers to clinical data quality and availability. Currently there is an immense amount of data already available to health care delivery organizations as a byproduct of the practice of medicine and possibly the adoption of the programs aforementioned. It is the hope of health care practitioners, health policy makers and care organizations that this amount of data will be of value to the improvement of the health care systems in the United States and possibly globally. The three major areas of enhancing performance with data are improving patient care, reducing costs and managing performance. According to Pogorelc (2013), “By 2015, the average hospital will produce more than 665 terabytes of data which is equivalent to 697,303,040 megabytes” (Pogorelc, 2013). This rapid growth in availability of data in health care can be an opportunity as well as a challenge to the industry.

This paper will discuss the data warehousing and data mining in health care, the opportunities and challenges inherent in the use of medical data, and specifically the application of these concepts in optimizing personalized medicine.
LITERATURE REVIEW

Data Warehousing and Mining in Healthcare
In recent times, organizations have understood the opportunities inherent in making meaningful use of data from their information systems to make better decisions, save cost, view trends and predict future outcomes—hence the need to store huge amounts of data in what is currently known as a “data warehouse” (Pedersen & Jensen, 1998). According to Pedersen, the functionality offered by data warehousing has traditionally been used in business in the areas of retail and finance, but the technology is now increasingly being used in more scientific area (Pedersen & Jensen, 1998). Prior to now, data warehousing as a concept was used more in finance and management where the need to see a corporate perspective of information was key. Initially, this was a bit more complicated, as organizations held their data in disparate systems, making it difficult to have an integrated data to extract information from (Inmon, 2007). Data warehousing thus required that many systems that are unintegrated become integrated.

“Data warehousing” as a term was first used by Barry Devlin, but became popularized by Bill Inmon who defined it as: “A data warehouse is a subject oriented, integrated, non-volatile and time-variant collection of data in support of management’s decision” (Pedersen & Jensen, 1998).

Data in the data warehouse is optimized for data analysis rather than data entry. It is therefore used to understand and manage the enterprise both at a strategic and tactical level.

Gray and Watson (1998) posited that data warehousing is fundamentally created to provide a dedicated source of data to support decision making applications. And that data warehouse provides a single version of the truth by providing users and applications access to the same data (Watson, 2007).

A major question would be: Is a clinical or healthcare data warehouse really needed? The answer to this question would be a definite “yes” because, just like Gray and Watson (1998) reasoned, it provides the same version of data for various needs, which saves different users the time and resources to write programs to answer questions on different kinds of data—hence increasing the focus on data analysis rather than data collection and cleansing.

In trying to understand healthcare issues and solutions, unbiased and timely data must be available. Healthcare data is usually stored in diverse formats both logically and physically, and it is limited in scope of reports—therefore making it ripe for the application of data warehousing technology, to integrate these data sources and thereby support more effective healthcare decision-making at both individual and community levels. Many clinical uses of data warehousing have also been reported in the last 15 years, in addition to the administrative uses. These uses include recruitment for clinical trials, gene-disease association, family health history data patterns, public health, trends in drug use/cost/interactions, infection surveillance and so on. It has also been found to be a fundamental tool in the personalization of patient care (Evans et al, 2012).

Data warehousing in healthcare is very challenging because the industry is ripe with often incompatible medical standards and coding schemas. These data come from many sources and are delivered in many forms including different data formats and individual spread sheets which are highly sensitive, making privacy and security issues paramount. Another major issue is that the health care industry is widely decentralized and largely autonomous (Berndt et al., 2001).
As clinical data warehousing is being adopted rapidly by health care systems, the value of these warehouses cannot be harnessed without applying data mining technology to them. Data mining tools are search tools that originated in statistics, computer science, and other non-biomedical disciplines to find associations among variables that may be useful in making management decisions (Mullins et al., 2006).

Data mining is the process of selecting, exploring and modeling large amounts of data. Different authors and researchers have defined data mining in so many different ways. “Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner” (Yoo et al., 2012). Kaur and Wasan (2006) also defined it as the non-trivial extraction of implicit previously unknown and potentially useful information about data (Fayyad et al., 1996). Data mining enables the generation of scientific hypotheses from large experimental data sets and from biomedical literature, as it also bridges the gap between data availability and the use of knowledge that can be derived from it (Yoo et al., 2012). Often times, the concept of data mining and Knowledge Discovery in Databases (KDD) have been used interchangeably, but it is very important to distinguish both concepts despite their similarities. According to Fayyad et al., (1996) data mining is one of the steps (Selection, Pre-processing, Transformation, Data mining and Interpretation/Evaluation) of KDD, which involves fitting models to or determining patterns from observed data, while KDD refers to the overall process of discovering useful knowledge from data, such as data cleaning and incorporation of heuristics (Fayyad et al., 1996).

In health care practice, data mining has been used to reduce adverse drug effects and also to suggest cheaper and safer alternatives. Areas where data mining techniques have been applied successfully in healthcare management include, but are not limited to: executive information systems for health care; forecasting treatment costs and demand of resources; anticipating patient’s future behavior given their history; Public Health Informatics; e-governance structures in health care; health insurance; etc. (Kaur & Wasan, 2006).

An emerging field of healthcare, where data warehousing and mining is of absolute importance, is personalized medicine. Personalized medicine is a new trend of patient care, where patients are treated not only based on the symptoms, but also on the causes of the disease, its rates of progression and how their bodies will react towards the medications and diets prescribed. The data mining techniques will help to get a better understanding of what causes a disease, what its particular structure is and how to design personalized therapies (Kuchenreuther & Sackman, 2014).

A data warehouse is a central research repository for personalized medicine with respective services for collecting, sharing and further elaborating annotated anonymized clinical data and other research-relevant data from diverse heterogeneous sources, such as in particular clinical trials and electronic patient records from health information systems or electronic health records (Schera et al, 2014).

**Personalized Medicine**

Over the past six decades, much evidence has emerged indicating that a substantial portion of variability in drug response is genetically determined, with age, nutrition, health status, environmental exposure, epigenetic factors and concurrent therapy playing important contributory roles. To achieve individual drug therapy with a reasonably predictive outcome, one
must further account for different patterns of drug response among geographically and ethnically distinct populations (Vogenberg et. al 2010). The observation in variability of drug response in the early 1950s led to the evolution of pharmacogenetics, which is a confluence of genetics, biochemistry and pharmacology. According to Vogenberg et al., advances in molecular medicine spawned the newer field of pharmacogenomics, which seeks to understand all of the molecular underpinnings of drug response. Vogenberg posits that personalized medicine came about as a result of the commercialization of the aforementioned research application.

According to the United States Food and Drug Administration, the term “personalized medicine” is described as “providing the right patient with the right drug at the right doses at the right time.” They also described it as tailoring of medical treatment to the individual characteristics, needs and preferences of a patient during all stages of care, including prevention, diagnosis, treatment and follow-up. With the scientific breakthroughs and technological advancements in recent times, personalized healthcare has the capacity to detect the onset of disease at its earliest stages; pre-empt the progression of disease; and at the same time, increase the efficiency of the health care system by improving the quality, accessibility and affordability (personalized medicine coalition).

The National Institute of Health also defined personalized medicine as “an emerging practice of medicine that uses an individual’s genetic profile to guide decisions made in regard to the prevention, diagnosis, and treatment of disease. Knowledge of a patient's genetic profile can help doctors select the proper medication or therapy and administer it using the proper dose or regimen” (National Institute of Health).

The National Cancer Institute, however, defines personalized medicine as “a form of medicine that uses information about a person’s genes, proteins, and environment to prevent, diagnose, and treat disease.” In cancer cases it uses specific information about a person’s tumor to help diagnose, plan treatment, find out how well treatment is working or make prognosis (Josko, 2014).

The most prominent examples of modern personalized medicine are genomic tests designed to guide treatment choices, such as testing for human epidermal growth factor receptor type 2 (HER2, also referred to as HER2/neu), to select patients with breast cancer who will benefit from trastuzumab; and testing for the KRAS mutation, to determine who is likely to benefit from therapies inhibiting the epidermal growth factor receptor (Garber and Tunis 2009).

Personalized medicine works by making use of information from genomes and their derivatives (RNA, proteins and metabolites) to guide medical decision making. Such medical decisions should be done while people are still healthy or at the earliest stages of the disease, and therefore be considered as preventive healthcare (Ginsburg and Willard, 2009). Such kind of preventive healthcare is now very affordable. The cost of sequencing a human genome has dropped to $1000 compared to $400 million in the last decade (Collins, 2015). There has been demonstrated and documented evidence of success for several conditions and treatments. The question now is: Will personalized medicine be able to achieve its acclaimed widespread benefits?

Benefits of Personalized Medicine
• **Disease susceptibility:** Pharmacogenetics has made genetic linkage studies in families with hereditary breast and ovarian cancer syndromes, as well as families with hereditary colon cancer, possible. It has also led to the identification of several important loci that are used for screening, disease risk counseling and preventive treatment programs. For instance, women who carry mutations in either BRCA1 or BRCA2 (*human genes that produce tumor suppressor proteins*- these proteins help repair damaged DNA and, therefore, play a role in ensuring the stability of the cell’s genetic material. When either of these genes is mutated, or altered, such that its protein product either is not made or does not function correctly, DNA damage may not be repaired properly. As a result, cells are more likely to develop additional genetic alterations that can lead to cancer) have a high risk for breast and ovarian cancer, and it is now recommended that women in such families have the opportunity to undergo genetic testing to make decisions about surveillance or even surgical approaches to mitigating a high risk of developing breast cancer. Similarly, people in families with a strong history of colon cancer can undergo testing for genes such as MLH1 and MSH2 that may identify individuals who have a risk as high as 60% for colon cancer. Early and regular screening colonoscopy in these individuals, as opposed to the recommendation for the general population to begin screening at a particular age, may enable the early detection of colon cancer (Ginsburg & Willard, 2009).

• **Reduction in adverse drug reactions:** Personalized medicine has been shown to reduce the adverse effect of certain drugs by being able to interpret genetic information of individuals to determine the likelihood of positive response to a particular treatment option. A typical case is that of panitimumab, a drug for the treatment of colon cancer, which has shown to be effective only in the absence of the KRAS mutation in a tumor. Also, the drug maraviroc is only beneficial for HIV patients with certain strains (CCR5) of the HIV disease. These examples, among others, demonstrate how genomic information may lead to tailored treatments with fewer adverse effects and a positive impact on personal health and well-being (Najafzadeh et.al, 2012).

**Issues and Problems**

Personalized medicine promises so many opportunities in the improvement of health care, from cost savings to prevention and reduction of adverse drug reactions. In the midst of all these promises, there are some many challenges that need to be addressed in order to fully harness the tremendous benefits of personalized medicine. These obstacles include public policy issues relating to uncertain regulatory requirements, insufficient insurance reimbursements for diagnostic tests linked to pre-emptive care, incomplete legal protections to prevent genetic discrimination, the lack of a comprehensive healthcare information technology system, and a lack of integration of the this knowledge into the medical education system, thereby making it difficult for physicians to incorporate personalized medicine diagnostics or pharmacogenomics into their practices (PMC).

Davis et al. offers a slightly different perspective to the challenges of personalized medicine through an investigation carried out in 2009 on the challenges and promises of personalized medicine. Their findings highlighted the three following major obstacles that have held back the advancement of personalized medicine:

• Scientific challenges (for example, poor understanding of molecular mechanisms or a lack of molecular markers associated with some diseases)
• Economic challenges (that is, incentives that are poorly aligned between stakeholders)
• Operational issues - electronic tracking of diagnostic information, privacy concerns, reimbursement/coding issues and provider/patient education).

Although scientific challenges remain, it now seems that the economic challenges and operational issues present the most significant obstacles to the further development of personalized medicine. In many cases, operational issues can largely be resolved within a particular stakeholder group (Davis et al, 2009).

**Ethical Issues:** Unlike many other diagnostic tests, genomic sequencing for personalized medicine allows for the identification of individual and related family members. Thus, concerns regarding data security, confidentiality and privacy are raised and linked to genomic patient data. Genomic patient data must be restricted to use in clinical practice (De Lecea & Rossbach, 2012).

For patients or individuals to be willing to take advantage of personalized medicine, they need to make sure that their privacy is protected to a great extent. Certain questions arise on the ethical issues or the extent to which genetic information of an individual should be used for personalized medicine services. Collecting personal information and sharing it with other family members who share the same genetics increases the likelihood of learning a lot of personal information of other relatives. Furthermore, this kind of data, even after you remove the identifiers, can provide enough information to identify someone. Therefore, the privacy can be at risk.

Another major issue that may arise from personalized medicine is that insurance companies may require genetic testing to pay for targeted treatment, even though it might be against a patient’s choice. Secondly, insurance companies might not also agree to pay for a particular treatment or drug because it is not a known treatment for all for a particular disease, but can be used for a unique individual (Josko, 2014).

**Practitioners’ Perceptions**
According to Raghavan and Vassy (Raghavan & Vassy, 2014), there are many issues or factors affecting physicians’ rates of the adoption of personalized medicine. Most important is the highly variable knowledge and comfort with genetic concepts broadly and their applications to clinical medicine specifically. For example, in a survey of U.S. physicians, including generalists and specialists, Selkirk et al. (Selkirk et al., 2013) found that 79% and 69% of primary-care and non-primary-care physicians, respectively, report that “lack of knowledge about genomic medicine” is a barrier to its incorporation in practice. In another study, Haga et al. found that, while only 43% of respondents to a survey of U.S. primary care physicians reported inadequate knowledge of genomic testing, feeling well-informed about genomic testing was a very strong predictor of genomic test utilization (an odds ratio of 4.6 of ordering a genomic profile for a patient). Also, in a recent focus group conducted by Najafzadeh et al. (Najafzadeh et al., 2013) in Canada, out of 28 physicians in the group sample, about 36% of physicians self-reported that they were not familiar with the concept of personalized medicine prior to participation in the focus groups. The physicians expressed different opinions on the actual meaning of personalized medicine: some physicians indicated that the term referred to tailoring treatment based on physician’s knowledge about patient-specific information (e.g., health record, co-morbidities, hereditary, familial risks, etc.), while others emphasized on the role of genetic information in personalized
A few physicians speculated that personalized medicine referred to improved treatments for socio-economically advantaged patients.

**Patient's Perception**
In a study by De Marco in 2010 (DeMarco et al., 2010), patients’ attitudes towards personalized medicine, genetic testing and race-based prescribing, and whether they differed between white and African-American participants, were examined and their results suggested that personalized medicine and genetic testing, though not well understood by lay persons, were considered positive advances in medicine. However, participants also voiced a range of concerns. For example, participants felt that, while promising, the use of genetic testing to personalized medicine might be too expensive to be accessible to the general public. The African-American group discussed more about medical mistrust by marginalized populations, which may affect the acceptability of personalized medicine when it becomes widely available. A welcome development in this technology was that African-Americans viewed personalized medicine and genetic testing as positive developments that could reduce the trial-and-error nature of prescribing and the possibility of side effects. In particular, patients need to understand the purpose of genetic testing and how it can be used together with other clinical information to determine the best treatment. Health professionals will need to draw from many disciplines, such as psychology, anthropology and advertising, to create social marketing campaigns that address the numerous barriers to acceptance and utilization. Even in the face of overall positive views of personalized medicine, some participants were wary of participating in any sort of medical testing, believing that they might be exploited or the privacy of test results be compromised. One way to address this concern is to provide patients with information about the Genetic Information Nondiscrimination Act (GINA), legislation signed into law in 2008 to protect patients from discrimination by their health insurers and employers, based on their genetic information. This approach is similar to the dissemination of information during health care visits on the Health Insurance Portability (HIPPA).

**Future of Personalized Medicine**
Genomic information can also predict illnesses in healthy individuals. In the near future, physicians can look forward to a medical landscape in which the pairing of affordable, efficient DNA sequencing and electronic health records could be used to inform a lifetime of health care strategies. Combined with the use of mobile health technology to assist in real-time monitoring of factors, such as diet, exercise, blood pressure, heart rate and blood chemistries, this approach could lead to more precise ways of preventing and managing chronic diseases (Collins, 2015).

Personalized medicine is a chance to revolutionize health care, but this will require a team effort by innovators, entrepreneurs, regulators, payers and policymakers. A future platform of Personalized Medicine includes many stakeholders, who are not limited to the healthcare industry only. This platform does not include only hospitals and the treatment they provide, but also drug development effort (in the early stages); integration of electronic health care system (which includes travel information, emergency rooms etc.); responsiveness towards therapies, in specifics disease states; cost – effective treatment data (Jakka & Rossbach, 2013); changing policies related to the ownership of medical information and the patient’s role in making future medical decisions; redefining the reimbursements and pricing policies; creating organizational structures and programs that support such platform.
The Platform of Personalized Medicine

<table>
<thead>
<tr>
<th>Medical Properties</th>
<th>Legal Properties</th>
<th>Economic Properties</th>
<th>Ethical Properties</th>
<th>Social Properties</th>
<th>Organizational Properties</th>
</tr>
</thead>
</table>

Table 1: A Suggested Platform of Personalized Medicine

The future research in this topic will be towards providing directions and suggestions on what each area properties should include and expending this suggested platform with other necessary areas which help on creating and sustaining the concept of Personalized Medicine.

Conclusions

The future of healthcare in the United States and beyond looks very promising with the advent of personalized medicine. Personalized medicine has the potential to reduce the cost and pain associated with trial and error/one-size fits all approach of the current health care systems, introducing new treatment protocols of using molecular elements that signal the risk of disease at the genetical level. Through early detection, prevention, and accurate risk assessments and efficiencies in care delivery, it promises to address current inefficiencies in health care delivery thereby having a significant impact on the economies of nations.

For personalized medicine to deliver on its promises, it will be heavily dependent on efficient data mining tools and techniques because of large collection of molecular data from patients and use of enormous amount of data from EHRs. According to Chris Edwards on using data for personalized cancer treatment, “one of the problems of personalized medicine is obtaining enough data to work out how different treatments fare under different conditions, getting a database of 10,000 to 20,000 patients with 50 to 100 common tumor types, amounting to at least one million patients.”

The future of personalized medicine will be based on a combination of personal data which include genomic information as well as longitudinal documentation of all possible molecular components. It is therefore pertinent that efficient data collection, storage and mining tools are developed and applied in order to speed up mining of realms of clinical data for meaningful use in personalized medicine and also not forgetting data security which is a major issue in dealing with patients’ genomic information in personalized medicine.

In order to optimally reap the benefits of personalized medicine, there should be supportive public policy environment that would address each of the issues raised, and provide incentives to reinforce emerging business models that accelerate the co-development of drugs and diagnostic tests. There should also be strategic alliance amongst the different disciplines involved in promoting personalized medicine- patients, scientists and drug companies. Last but very important, there should be sufficient non competing resources and a strong, sustained commitment of time, energy, and ingenuity from the scientific, medical, and patient communities for the full potential of personalized medicine to be realized. (Collins, 2015).

References:
References available upon request