1-2018

The Impact of Big Data on Chronic Disease Management

Niharika Bhardwaj

Bezawit Wadajo

Anthony Spano

Alberto Coustasse

THE IMPACT OF BIG DATA ON CHRONIC DISEASE MANAGEMENT

ABSTRACT

Introduction: Population health management – and specifically chronic disease management – depend on the ability of providers to identify patients at high risk of developing costly and harmful conditions such as diabetes, heart failure, and chronic kidney disease (CKD). The advent of big data analytics could help identify high-risk patients which is really beneficial to healthcare practitioners and patients to make informed decisions in a timelier manner with much more evidence in hand. It would allow doctors to extend effective treatment but also reduces the costs of extending improved care to patients.

Purpose: The purpose of this study was to identify current applications of big data analytics in healthcare for chronic disease management and to determine its real-world effectiveness in improving patient outcomes and lessening financial burdens.

Methodology: The methodology for this study was a literature review. Six electronic databases were utilized and a total of 49 articles were referenced for this research.

Results: Improvement in diagnostic accuracy and risk prediction and reduction of hospital readmissions has resulted in significant decrease in health care cost. Big data analytic studies regarding care management and wellness programs have been largely positive. Also, Big data analytics guided better treatment leading to improved patient outcomes.

Discussion/Conclusion: Big data analytics shows initial positive impact on quality of care, patient outcomes and finances, and could be successfully implemented in chronic disease management.

Keywords: Big data, Big data analytics, Chronic diseases, Chronic disease management

INTRODUCTION

Chronic diseases in the United States (U.S.) such as heart disease, stroke, cancer, Type 2 Diabetes Mellitus (T2DM), obesity, and chronic lung diseases have been the greatest preventable drivers of morbidity and mortality (Bauer, Briss, Goodman, & Bowman, 2014; CDC, 2014a). As of 2012, roughly 50% of the U.S. population (117 million people) had at least one chronic condition with 25% suffering from more than one (Ward, 2014). In 2014, 7 of the top 10 causes of mortality in the U.S. were chronic diseases, with heart disease and cancer contributing to about 47% of deaths nationwide (NCHS, 2016). Although short-term improvements in preventable hospitalizations and cardiovascular deaths have been noted, there have been alarming increases in the rates of obesity (7.2% from 2013 to 2015) and self-reported diabetes (5.6% in the last 20 years) (CDC, 2014b; UHF, 2015).

Apart from this morbidity and mortality burden, chronic medical illnesses accounted for 86% of the U.S. healthcare expenditures in 2010. The total cost (direct medical plus lost productivity costs) to the U.S. was $315.4 billion for heart disease and stroke, $157.77 billion for cancer in 2010 and $345 billion for diabetes and prediabetes in 2012 (Mariotto, Yabroff, Shao, Feuer, & Brown, 2011; ADA, 2013; Go, et al., 2014).

Population health management – and specifically chronic disease management – have depended on the ability of providers to identify patients at high risk of developing costly and harmful conditions such as diabetes, heart failure, and chronic kidney disease. These basic risk stratification tasks have been traditionally performed through non-electronic means, such as patient questionnaires, manual chart reviews, and in-person assessments (Bresnick, 2015).

However, the advent of big data analytics has drastically changed the way providers can develop risk scores, monitor patients, and even divide cohorts into extremely narrow subgroups to ensure precision care (Bates, Saria, Ohno-Machado, Shah, & Escobar, 2014). It has been envisaged that the ability to obtain and analyze big data could glean information that could identify high-risk individuals, inform more effective treatments, and pinpoint cost reduction areas across the healthcare system (Burg, 2014).

Within healthcare, big data has been defined as high volume and high diversity biological, clinical, environmental, and lifestyle information collected from single individuals to large cohorts, in relation to their health and wellness status, at one or several time points (Auffray, et al., 2016). Big data analytics has involved various analytical techniques ideal for analyzing a large proportion of text-based health documents and other unstructured clinical data (e.g., physician's written notes, prescriptions and medical imaging) such as descriptive, diagnostic, predictive, and prescriptive analytics (Groves, Kayyali, Knott, & VanKuiken, 2013). Predictive modelling has been noted to be vital to transforming large clinical datasets, or “big clinical data,” into actionable knowledge for various healthcare applications. Such models could guide clinical decision making and personalized medicine (Bauer, et al., 2014). For instance, asthma patients at high risk for hospitalization could be enrolled in an asthma case management program by determining their risk of hospitalization within the next year (Luo, Stone, Sakaguchi, Sheng, & Murtaugh, 2015).

Furthermore, healthcare organizations must ensure the proper acquirement of tools, infrastructure, and techniques in order to best utilize big data to optimize their business and avoid risking losing millions in revenues and profits on useless data techniques (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011).

The reliance of U.S. healthcare system on big data has been anticipated to continue its climb to enable a more complete view with information on care coordination, outcomes-based reimbursement models, health management, and patient engagement (IHTT, 2013; Archenaa &

Anita, 2015). Massive datasets have been created in the medical field via the myriad data sources and electronic health records due to innovations. These innovations have included laparoscopic and robotic surgery, smart homes for patient self-monitoring, smart applications (apps) and softwares for body signal analysis and other mHealth technologies for biological, behavioral, and environmental data collection (Barbash & Glied, 2010; Peters & Buntrock, 2014; Theoharidou, Tsalis, & Gritzalis, 2014).

Based upon early successes of big data, McKinsey and Company has estimated healthcare cost savings of 12% to 17% - $348 billion to $493 billion when extrapolated to healthcare expenses that reached $2.9 trillion in 2013 (Groves et al., 2013). Most of these savings were expected to come from clinical operations and Research and Development with $165 billion and $108 billion in waste, respectively (Manyika, et al., 2011).

Overall, the purpose of this study was to identify current applications of big data analytics in healthcare for chronic disease management and to determine its real-world effectiveness in improving patient outcomes and lessening financial burdens.

METHODOLOGY

The conceptual framework for this research conformed to the steps and research framework followed by Yao, Chu and Li (2010). The framework elucidates the course of Big Data analytics adoption for chronic disease management. For an investigation into whether big data analytics can lead to improved patient outcomes in a cost-effective manner for people suffering from chronic diseases, their preliminary benefits and correspondent costs must be pinpointed at first. The process of technology adoption begins when problems in the existing system necessitate assessment of needs subsequently resulting in the creation and institution of a solution. It involves an evaluation of the benefits of and barriers to utilization of big data analytics once it is adopted, and is repeated to permit appraisal of benefits and addressal of each of the barriers (see Figure 1). This conceptual framework is suitable for the present study as it centers on means of application of new technology in healthcare settings. Moreover, there is support for the internal validity of this approach evident from its successful replication in past studies (Coustasse, Tomblin, & Slack, 2013; Deslich & Coustasse, 2014; Porterfield, Engelbert, & Coustasse, 2014).

This study hypothesized that the introduction of Big Data Analytics for chronic condition management should demonstrate benefits in terms of improved diagnostic accuracy and risk prediction, more effective and targeted therapy along with reduced readmissions and expenses.

This research utilized literature review as its methodology. The electronic databases of PubMed, Academic Search Premier, CINAHL Complete, ProQuest, EBSCOHost and Google Scholar were searched for the terms 'Big data analytics' or 'Predictive analytics' and 'Chronic disease' or 'Chronic medical illness'. Articles were also retrieved from the American Diabetes Association and Centers for Disease Prevention and Control websites. Literature was selected to include the advantages of big data analytics for chronic disease management in regards to both outcomes and cost. To stay current in the research study, the search was limited to sources attainable as full texts published between 2007 and 2016 in the U.S. in English. Original articles, reviews and research studies including primary and secondary data were included. After a review of the abstracts, relevant articles were chosen. From a total of 90 references found, 49 resources were selected for this research. The literature search was conducted by A.S, B.W., N.B. and S.N. and validated by AC who acted as a second reader and double-checked that the references met the research study inclusion criteria.

The results were categorized into the following subheadings: Big Data Analytics: Reduction of Readmissions, Big Data Analytics: Improvement in diagnostic accuracy and risk prediction and Big Data Analytics: Treatment guidance, patient outcome improvement and financial implications.

RESULTS

As per a study conducted in 2015, 15% of providers have been using predictive modelling with 92% of them utilizing the outputs to predict patient risk or illness. The illnesses and conditions most often targeted included: readmissions, patient deterioration, sepsis, and general patient health (CHIME, 2015). Per a study conducted by McKinsey and Company in 2013, more than 200 new businesses have developed innovative healthcare big data analysis apps since 2010 - 40% aimed at direct health interventions or predictive capabilities instead of apps formerly centered on data management and retrospective data analysis (Kayyali, Knott, & VanKuiken, 2013). (See Figure 2). Table 1 lists the study design and purpose and summarizes the findings of the studies included in this review.

Big Data Analytics: Reduction of Readmissions

Hospital readmissions are an important quality indicator for health systems, being a significant contributor to health care costs (Parikh, Kakad, & Bates, 2016). All-cause 30-day readmissions cost the US hospitals more than $41.3 billion in 2011 (Hines, Barrett, Jiang, & Steiner, 2011). While studying the 30-day risk of readmission for congestive heart failure patients using predictive analytics on more than 3 million patient records, Zolfaghar, et al., (2013), reported an accuracy of 77% and recall of 61%. Parkland Health and Hospital System, Dallas, TX, utilized Electronic Medical Record data to predict readmission risk for heart failure patients and noticed 26% relative reduction in risk-adjusted odds of readmission among patients enrolled post intervention versus those enrolled pre-intervention (Amarasingham, et al., 2013). In another study, a model predicting risk of rehospitalization within 30 days utilizing 15 years of medical data was seen to have a positive predictive value of 54% (Hoch & Karpati, 2013). (See Table 1).

Big Data Analytics: Improvement in Diagnostic Accuracy and Risk Prediction

Ross, Shah, Dalman, Nead, & Leeper (2016) showed that use of machine learning based predictive analytics provided better results than the gold standard risk prediction scores for identifying undiagnosed Peripheral Arterial Disease (PAD) (Machine learning - 84%; PAD risk score - 71%) and predicting future risk of major adverse cardiac events (Machine learning - 70%, Framingham score - 56%). Another study by Rusin, et al., (2016) used continuous high resolution recordings to measure the effectiveness of automated, intelligent analysis of standard physiologic data in real-time to detect signs of clinical deterioration too subtle to observe for clinicians in children with parallel systemic and pulmonary circulation (such as hypoplastic left heart syndrome) reported that the algorithm was 91% accurate for detecting impending events and could possibly serve as an early warning indicator for such patients. (See Table 1).

Razavian, et al., (2015) mentioned that applying a population-level risk prediction model for T2DM on readily available administrative data improved positive predictive value by at least 50% in predicting diabetes when compared to classical diabetes risk prediction algorithms on very large populations with inadequate data. It also identified novel risk factors for T2DM, such as chronic liver disease (Odds Ratio [OR] 3.71), high alanine aminotransferase (OR 2.26), esophageal reflux (OR 1.85), and history of acute bronchitis (OR 1.45). Kupersmith, et al., (2007) analyzed the U.S. Veterans Health Administration’s EHR data to identify a high rate of mental illness comorbidity (24.5%) among patients with diabetes by exploring the influence of surrogate attributes. (See Table 1).

A retrospective analysis of big data stored within Optum Labs data warehouse by McCoy, et al., (2016) indicated that intensive treatment almost doubled severe hypoglycemia risk among complicated patient cases. Data Mining Algorithms were applied to a large clinical dataset and an increase in T2DM classification accuracy was noticed from 78.71% to 86.64% for Hemoglobin A1C (HbA1c) at 6.5% if an oxidative stress marker was included in the algorithm and to 85.63% when interleukin-6 was included but with lower optimal HbA1c range between 5.73 and 6.22% (Jelinek, Stranieri, Yatsko, & Venkatraman, 2016). (See Table 1).

Using Explorys, a novel big data storage system - it was observed that patients with nicotine dependence, obesity, depressive disorders, and alcohol abuse had a relative risk of 4.489, 6.007, 5.511, and 3.326 for low back pain, respectively, compared to patients without it. (Shemory, Pfefferle, & Gradisar, 2016). Van Fossen, Wilhelm, Eaton, & McHenry, (2013) also utilized Explorys and reported an increase in prevalence of subsequent breast and renal cell cancer in thyroid cancer patients (female: 0.67- and two-fold; M - 29- and 4.5-fold respectively) and of thyroid cancer in breast and renal cell cancer patients (Female: twofold and 1.5-fold; Male - 19- and threefold respectively). (See Table 1).

Dinov, et al., (2016) explored risk of Parkinson's disease utilizing the Parkinson's Progression Markers Initiatives' (PPMI) unique archive containing complex imaging, genetics, clinical and demographic data. It was found that model-free Big Data machine learning-based classification methods were significantly powerful in predicting Parkinson's disease in the PPMI subjects with an accuracy, sensitivity, and specificity consistently exceeding 96%. The Cleveland Clinic developed dashboards for their care managers that utilized patient data from EHR and excel sheets to pinpoint about 1,000 out of 54000 local, at high risk patients not currently covered by care coordination by using filters based on geography, condition and many more. These patients could then be engaged, have appointments fixed and their medication adherence monitored (Health Data Management, 2016). (See Table 1).

Big Data Analytics: Treatment Guidance, Patient Outcome Improvement and Financial Implications

Big data analytic studies of care management and wellness programs have been largely positive. For example, per Berg (2015), number of members that would have had an asthma controller medication prescription decreased by 7.3% and those with a prescription for statin medication decreased by 16.6% without the care management program (Berg, 2015). In another study, Chronic Obstructive Pulmonary Disease (COPD) patients at risk for exacerbation were identified using big data from the HealthCore Integrated Research Environment and followed for 12 months. COPD exacerbation risk was reduced by 22% with lower COPD-related healthcare resource utilization and costs of $4,084 vs $5,656 per patient-year in association with the initiation of budesonide/formoterol combination versus tiotropium, respectively, was noticed (Trudo, et al., 2015). Propeller Health, a leading digital platform for respiratory health management, has used sensors for asthma inhalers, along with mobile apps and advanced analytics on big data, to help providers identify at-risk asthma patients before an attack occurs. During a 12-month study that measured real-world effectiveness of this platform to reduce use of Short Acting Beta Agonists (SABA) and improve asthma control, the study arm monitoring SABA use with the Propeller Health system significantly decreased SABA used (daily mean number of SABA users 0.41 vs 0.31), increased SABA-free days (21% vs 17%), and improved asthma control test scores (63% vs 49%) (Merchant, Inamdar, & Quade, 2016). (See Table 1).

The EMR data from the Nihon University School of Medicine clinical big data warehouse has been used for several retrospective analysis studies. One study found an increase in the risk of gastrointestinal bleeding associated with a combination of clopidogrel and aspirin versus aspirin alone (RR - 2.06) for stroke risk reduction (Takahashi, Nishida, Nakayama, & Asai, 2013). Nishida, et al., (2013) found losartan had a beneficial effect on serum uric acid levels (decrease of 0.14 mg/dl) compared to an increase caused by other angiotensin receptor blockers in hypertensive T2DM patients. (See Table 1).

DISCUSSION

This literature review aimed to look for signs that big data analytics could prove beneficial for chronic disease management and ultimately lead to improved outcomes, reduced disease burden and decrease treatment costs. Improving outcomes for patients suffering from chronic diseases should begin with gaining an understanding of the disease epidemiology from the massive pool of data through appropriate application of big data analytics. Determination of treatable patient risk factors and risk for readmission will help providers identify and monitor at-risk patients to prevent and manage chronic disease more effectively.

The publications included in our review applied analytics to gain insights such as risk predictions for diseases like diabetes, cancer, heart disease, low back pain and prediction of re-hospitalization/readmission risk and to provide information to enable readmission reduction and treatment guidance for diseases like hypertension, diabetes, stroke. Also, the studies that predicted risk of disease or diagnosed diseases all had an accuracy rate exceeding 70%. This could be applied to detect undiagnosed cases of chronic disease before they become too severe.

In a paper, Raghupathi & Raghupathi (2014) concluded that the field of big data analytics has shown potential for extracting insight from enormous data sets and improving outcomes while minimizing costs. The results from this literature review concur with their conclusions. There is consistent evidence that big data analytics can improve patient outcome, enhance diagnostic accuracy and reduce costs, thus, demonstrating potential benefits for its use for chronic disease management.

This literature review had a number of limitations. To begin with, there is a scarcity of quality big data sources even after EHR implementation for data analysis. Also, the initial cost of conducting big data analytics currently limits its use to large academic hospitals. Thus, there were a small number of articles demonstrating real-world application & benefits of big data analytics. Further, some of the studies applied analytics only to gain deeper understanding of the chronic disease without reporting a positive effect on patient outcomes. Additionally, as articles were evaluated to establish relevancy, publication and researcher bias cannot be ruled out.

All stakeholders in healthcare including healthcare organizations, payers, physicians, and patients have much to gain from implementation of big data analytics for chronic disease management. It can help monitor clinical indicators for decision support and explore which treatment option is most likely to result in clinical improvements for patients, thereby, reducing readmissions and improving quality of care/outcomes. Ultimately, it may lead to substantial savings for healthcare facilities. In order to realize these rewards, an increase in the support and investment into such a system must be considered. Many hospitals and health systems may already have an existing data source such as DW. Yet, these have not been utilized to their full capacity. With employment of big data analytics, full benefits realization can be achieved.

CONCLUSION

Big data analytics seem to have a positive impact on chronic disease care. Despite concerns of implementation costs and efficacy of data analysis, hospitals and other chronic care settings have seen promising initial results in terms of improvement in clinical outcomes as well as financial performance by adapting and integrating data analysis with healthcare information technology.

REFERENCES

Amarasingham, R., Patel, P. C., Toto, K., Nelson, L. L., Swanson, T. S., Moore, B. J., & Drazner, M. H. (2013). Allocating scarce resources in real-time to reduce heart failure readmissions: a prospective, controlled study. *British Journal of Medicine quality & safety*, *22*(12), 998-1005.

American Diabetes Association (ADA). (2013). Economic costs of diabetes in the US in 2012. *Diabetes care*, *36*(4), 1033.

Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. *Procedia Computer Science*, *50*, 408-413. *doi: 10.1016/j.procs.2015.04.021*

Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., & Del Signore, S. (2016). Making sense of big data in health research: Towards an EU action plan. *Genome medicine*, *8*(1), 71.

Barbash, G. I., & Glied, S. A. (2010). New technology and healthcare costs—the case of robot-assisted surgery. *New England Journal of Medicine*, *363*(8), 701-704.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in healthcare: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, *33*(7), 1123-1131.

Bauer, U. E., Briss, P. A., Goodman, R. A., & Bowman, B. A. (2014). Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA. *The Lancet*, *384*(9937), 45-52.

Berg, B. G. (2015, July 17). *3 ways big data is improving healthcare analytics.* Retrieved September 26, 2016, from http://www.healthcareitnews.com/blog/3-ways-big-data-improving-healthcare-analytics

Bresnick, J. (2015, December 24). *Big Data Analytics Improves Chronic Disease Risk Stratification*. *HealthITAnalytics*. Retrieved September 24, 2016, from http://healthitanalytics.com/news/big-data-analytics-improves-chronic-disease-risk-stratification

Burg, N. (2014, November 10). CastlightVoice: How Big Data Will Help Save Healthcare. Retrieved September 25, 2016, from http://www.forbes.com/sites/castlight/2014/11/10/how-big-data-will-help-save-healthcare/#391b75a71187

Centers for Disease Control and Prevention (CDC). (2014b). National diabetes statistics report: estimates of diabetes and its burden in the United States. *US Department of Health and Human Services.* https://www.cdc.gov/diabetes/pubs/statsreport14/national-diabetes-report-web.pdf (accessed September 23, 2016)

Centers for Disease Prevention and Control (CDC). (2014a). Chronic diseases and health promotion. *US Department of Health and Human Services*. Available at: http://www.cdc.gov/chronicdisease/overview/ (accessed September 23, 2016)

College of Health Informatics Management Executives (CHIME). (2015). Jvion Releases Findings from Latest Predictive Analytics in Healthcare Survey. *Healthcare IT*. Retrieved September 26, 2016, from https://chimecentral.org/jvion-releases-findings-latest-predictive-analytics-healthcare-survey/

Coustasse, A., S. Tomblin, & Slack, C. (2013). Impact of radio-frequency identification (RFID) technologies on the hospital supply chain: a literature review. *Perspectives in Health Information Management*, 1.

Deslich, S., & Coustasse, A. (2014). Expanding technology in the ICU: the case for the utilization of telemedicine. *Telemedicine and e-Health*, *20*(5), 485-492.

Dinov, I. D., Heavner, B., Tang, M., Glusman, G., Chard, K., Darcy, M., & Foster, I. (2016). Predictive Big Data Analytics: A Study of Parkinson's Disease Using Large, Complex, Heterogeneous, Incongruent, Multi-Source and Incomplete Observations. *PloS one*, *11*(8), e0157077.

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M. J., & Fullerton, H. J. (2014). Heart disease and stroke statistics-2014 update. *Circulation, 129*(3), e29. *doi: 10.1161/01.cir.0000441139.02102.80*

Groves, P., Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The 'big data revolution in healthcare. *McKinsey Quarterly*, *2*.

Health Data Management. (2016). *Big Data and Analytics.* New York: SourceMedia Marketing Solutions Group. Retrieved September 27,2016, from http://editiondigital.net/publication/?i=317633#{"issue_id":317633,"page":10}

Hines, A. L., Barrett, M. L., Jiang, H. J., & Steiner, C. (2011). Conditions with the largest number of adult hospital readmissions by payer. *HCUP Statistical Brief, # 172.* Retrieved September 27, 2016, from https://www.ncbi.nlm.nih.gov/books/NBK206781/

Hoch, I., & Karpati, T. (2013). PS2-37: Development and Use of a Predictive Analytics Tool in a Large Healthcare Organization. *Clinical Medicine & Research*, *11*(3), 154–155. http://doi.org/10.3121/cmr.2013.1176.ps2-37

Institute for Health Technology Transformation (IHTT) (2013). Transforming Healthcare through Big Data Strategies for leveraging big data in the healthcare industry. http://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.r93.cf2.rackcdn.com/iHT2_BigData_2013.pdf Retrieved on September 25, 2016

Jelinek, H. F., Stranieri, A., Yatsko, A., & Venkatraman, S. (2016). Data analytics identify glycated haemoglobin co-markers for type 2 diabetes mellitus diagnosis. *Computers in biology and medicine*, *75*(1), 90-97. Retrieved on November 4, 2016, from http://dx.doi.org/10.1016/j.compbiomed.2016.05.005

Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The big-data revolution in US healthcare: Accelerating value and innovation. *McKinsey & Company*, 1-13. Retrieved September 26, 2016, from https://digitalstrategy.nl/wp-content/uploads/E2-2013.04-The-big-data-revolution-in-US-health-care-Accelerating-value-and-innovation.pdf

Kupersmith, J., Francis, J., Kerr, E., Krein, S., Pogach, L., Kolodner, R. M., & Perlin, J. B. (2007). Advancing evidence-based care for diabetes: lessons from the Veterans Health Administration. *Health Affairs*, *26*(2), w156-w168.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, *52*(2), 21.

Luo, G., Stone, B. L., Sakaguchi, F., Sheng, X., & Murtaugh, M. A. (2015). Using computational approaches to improve risk-stratified patient management: rationale and methods. *Journal of Medical Internet Research research protocols*, *4*(4), e128.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute.* Retrieved September 26, 2016, from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Mariotto, A. B., Yabroff, K. R., Shao, Y., Feuer, E. J., & Brown, M. L. (2011). Projections of the cost of cancer care in the United States: 2010–2020. *Journal of the National Cancer Institute.*103 (2):117-28. doi: 10.1093/jnci/djq495.

McCoy, R. G., Lipska, K. J., Yao, X., Ross, J. S., Montori, V. M., & Shah, N. D. (2016). Intensive Treatment and Severe Hypoglycemia among Adults with Type 2 Diabetes. *Journal of American Medical Association internal medicine.176* (7):969-978. doi:10.1001/jamainternmed.2016.2275

Merchant, R. K., Inamdar, R., & Quade, R. C. (2016). Effectiveness of Population Health Management Using the Propeller Health Asthma Platform: A Randomized Clinical Trial. *The Journal of Allergy and Clinical Immunology: In Practice*,*4*(3), 455-463.

National Center for Health Statistics (NCHS). (2016). Health, United States, 2015: with special feature on racial and ethnic health disparities. Retrieved September 25, 2016 from http://www.cdc.gov/nchs/data/hus/hus15.pdf

Nishida, Y., Takahashi, Y., Susa, N., Kanou, N., Nakayama, T., & Asai, S. (2013). Comparative effect of angiotensin II type I receptor blockers on serum uric acid in hypertensive patients with type 2 diabetes mellitus: a retrospective observational study. *Cardiovascular diabetology*, *12*(1), 1.

Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating predictive analytics into high-value care: the dawn of precision delivery. *JAMA*, 315(7), 651-652.

Peters, S. G., & Buntrock, J. D. (2014). Big data and the electronic health record. *The Journal of ambulatory care management*, *37*(3), 206-210.

Porterfield, A., Engelbert, K., & Coustasse, A. (2014). Electronic prescribing: Improving the efficiency and accuracy of prescribing in the ambulatory care setting. *Perspectives in Health Information Management / AHIMA, American Health Information Management Association*, *11*, 1.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, *2*(1), 1.

Razavian, N., Blecker, S., Schmidt, A. M., Smith-McLallen, A., Nigam, S., & Sontag, D. (2015). Population-Level Prediction of Type 2 Diabetes from Claims Data and Analysis of Risk Factors. *Big Data*, *3*(4), 277-287. doi:10.1089/big.2015.0020.

Ross, E. G., Shah, N., Dalman, R. L., Nead, K., & Leeper, N. J. (2016). VESS18. Use of Predictive Analytics for the Identification of Latent Vascular Disease and Future Adverse Cardiac Events. *Journal of Vascular Surgery*, *63*(6), 28S-29S.

Rusin, C. G., Acosta, S. I., Shekerdemian, L. S., Vu, E. L., Bavare, A. C., Myers, R. B., & Penny, D. J. (2016). Prediction of imminent, severe deterioration of children with parallel circulations using real-time processing of physiologic data. *The Journal of thoracic and cardiovascular surgery*. *152*(1), 171-177. doi:10.1016/j.jtcvs.2016.03.083

Shemory, S. T., Pfefferle, K. J., & Gradisar, I. M. (2016). Modifiable Risk Factors in Patients With Low Back Pain. *Orthopedics*. 39(3), e413-e416. DOI:10.3928/01477447-20160404-02

Takahashi, Y., Nishida, Y., Nakayama, T., & Asai, S. (2013). Comparative effect of clopidogrel and aspirin versus aspirin alone on laboratory parameters: a retrospective, observational, cohort study. *Cardiovascular diabetology*, *12*(1), 1.

Theoharidou, M., Tsalis, N., & Gritzalis, D. (2014). Smart home solutions for healthcare: privacy in ubiquitous computing infrastructures. *Handbook of smart homes, health care and well-being*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.723.4834&rep=rep1&type=pdf accessed on September 23, 2016

Trudo, F., Kern, D. M., Davis, J. R., Tunceli, O., Zhou, S., Graham, E. L., Williams, S. A. (2015). Comparative effectiveness of budesonide/formoterol combination and tiotropium bromide among COPD patients new to these controller treatments. *International Journal of Chronic Obstructive Pulmonary Disease*, *10*(1), 2055–2066. http://doi.org/10.2147/COPD.S90658

United Health Foundation (UHF) & American Public Health Association (APHA). America's Health Rankings. A call to action for individuals and their communities. (2015).

http://cdnfiles.americashealthrankings.org/SiteFiles/Reports/2015AHR_Annual-v1.pdf (accessed 23 September 2016)

Van Fossen, V. L., Wilhelm, S. M., Eaton, J. L., & McHenry, C. R. (2013). Association of thyroid, breast and renal cell cancer: A population-based study of the prevalence of second malignancies. *Annals of Surgical Oncology, 20*(4), 1341-1347. doi:10.1245/s10434-012-2718-3

Ward, B. W. (2014). Multiple chronic conditions among US adults: a 2012 update. *Preventing chronic disease,volume 11(1), e62.* DOI: http://dx.doi.org/10.5888/pcd11.130389

Yao, W., Chu, C. H., & Li, Z. (2010). The use of RFID in healthcare: Benefits and barriers. In *RFID-Technology and Applications* (RFID-TA), 2010 *IEEE International Conference* on (pp. 128-134). DOI 10.1109/RFID-TA.2010.5529874

Zolfaghar, K., Meadem, N., Teredesai, A., Roy, S. B., Chin, S. C., & Muckian, B. (2013, October). Big data solutions for predicting risk-of-readmission for congestive heart failure patients. In *Big Data, 2013 IEEE International Conference on* (pp. 64-71).

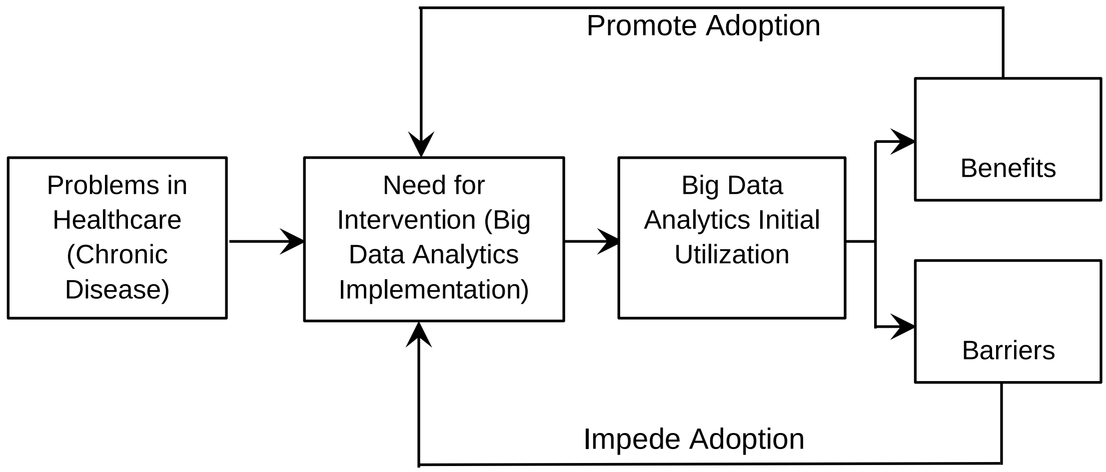
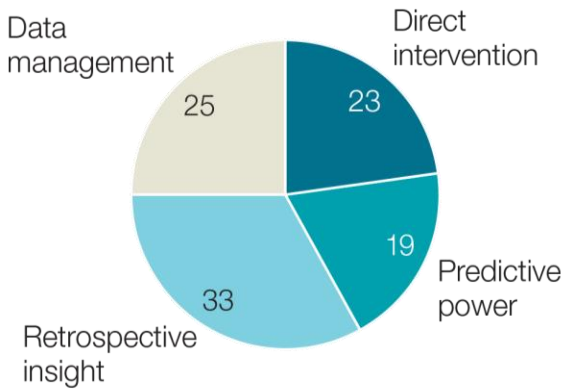


Figure 1: Conceptual Framework adapted from Yao, Chao-Hsien & Li, 2010



Key: 100% = 132

Source: 2010-11 submissions to Health Data Initiative Forum; Rock Health; Standard & Poor's Capital IQ; McKinsey analysis.

Figure 2: United States Health-care Data Apps from Top Innovators, by Type of Data/Analytic Capability, 2010-12

Table 1: Big Data Analytics – By Study Design, Purpose and Study Findings

Author	Study Design	Purpose	Findings
Kupersmith, et al., 2007	Retrospective (cohort)	Risk prediction for diabetic patients	Rate of mental illness comorbidities associated with diabetes - 24.5%
Amarasingham, et al., 2013	Prospective observational study (cohort)	Reduction of Readmissions	Relative reduction in risk-adjusted odds of readmission in the IG vs CG- 26%
Van Fossen, et al., 2013	Cohort study	Risk prediction for breast, renal, and thyroid cancer patients	Increased prevalence: Subsequent breast and renal cell cancer in thyroid cancer patients: Female: 0.67- and two-fold, Male: 29- and 4.5-fold Subsequent thyroid cancer in breast and renal cell cancer patients: Female: twofold and 1.5-fold, Male: 19- and threefold
Hoch & Karpati, 2013	Retrospective study	Risk prediction for rehospitalization within 30 days	Positive predictive value: 54%
Nishida, et al., 2013	Retrospective observational study	Treatment guidance for hypertensive T2DM patients	Change in serum uric acid level (mg/dl): Losartan: -0.124; Valsartan: 0.186; Telmisartan: 0.150; Candesartan: 0.153; Olmesartan: 0.166

Takahashi, et al., 2013	Retrospective observational cohort study	Treatment guidance for patients at high-risk for stroke	Relative risk of gastrointestinal bleeding with use of clopidogrel + aspirin vs aspirin alone - Multivariate model: 2.06; propensity adjustment:2.61
Zolfaghar, et al., 2013	Retrospective study	Prediction of readmission risk for CHF patients	Accuracy: 77%; Recall: 61%
Berg, 2015	N/A	Improvement in patient outcomes for asthmatic and high cholesterol patients	Increase in number of members that received prescription for - asthma controller: 7.3%; statin: 16.6%
Razavian, et al., 2015	Retrospective study	Improvement in diagnostic accuracy of diabetes	Improvement in positive predictive value ≥ 50% over classical algorithms
Trudo, et al., 2015	Retrospective study	Improvement in patient outcomes and reduction in costs for COPD patients	Decrease in risk of COPD exacerbation - 22% COPD-related healthcare resource utilization and costs ($/patient-year) - BFC: $4,084; Tiotropium: $5,656
Dinov, et al., 2016	Prospective cohort	Diagnosis of Parkinson's disease	Accuracy, sensitivity, and specificity > 96%

Jelinek, et al., 2016	Prospective study	Improvement in risk prediction for diabetic patients	Increase in T2DM classification accuracy: Oxidative stress marker inclusion in algorithm - 78.71% to 86.64% for HbA1C at 6.5%; Interleukin-6 - 78.71% to 85.63% with HbA1c range between 5.73% and 6.22%
McCoy, et al., 2016	Retrospective data analysis	Risk prediction	Increase in probability of severe hypoglycemia (intensive treatment vs standard): Low clinical complexity: 0.28%, High clinical complexity: 1.30%; Unnecessary intensive treatment delivery: 20%.
Merchant, et al., 2016	Pragmatic controlled study	Improvement in patient outcomes in asthmatic patients	Decrease in SABA use - IG: 0.41; CG: 0.31 Increase in SABA-free days - IG: 21%; CG: 17% Improvement in ACT scores - IG: 63%; CG: 49%
Ross, et al., 2016	Prospective observational study(cohort)	Improvement in diagnostic accuracy of undiagnosed PAD and risk prediction for future major adverse cardiac events	Accuracy of PAD prediction: MLPA: 84%, PAD Risk Score: 71%; Future major adverse cardiac events prediction: MLPA: 70%,

			Framingham score: 56%
Rusin, et al., 2016	Cross-sectional study	Risk prediction for children with parallel systemic and pulmonary circulation	Accuracy: 91%
Shemory, et al., 2016	Prospective cohort study	Risk prediction for patients with low back pain	Relative risk: Nicotine dependence: 4.489, Obesity: 6.007, Depressive disorders: 5.511, Alcohol abuse: 3.326

Key: IG=Intervention Group, CG=Control Group, BFC=Budesonide/Formoterol Combination, SABA=Short Acting Beta Agonists, CHF=Congestive Heart Failure, COPD=Chronic Obstructive Pulmonary Disease, ACT=Asthma Control Test, PAD=Peripheral Arterial Disease, AUC=Area Under Curve, MLPA=Machine Learning based Predictive Analytics, N/A=Not Applicable