

ELPP 2016: Big Data for Healthcare

Prashant Dhamdhere (VMWare), Jeremiah Harmsen (Google), Raaghav Hebbar (NetApp), Srinath Mandalapu (Yahoo), Ashish Mehra (NetApp), Suju Rajan (Yahoo)

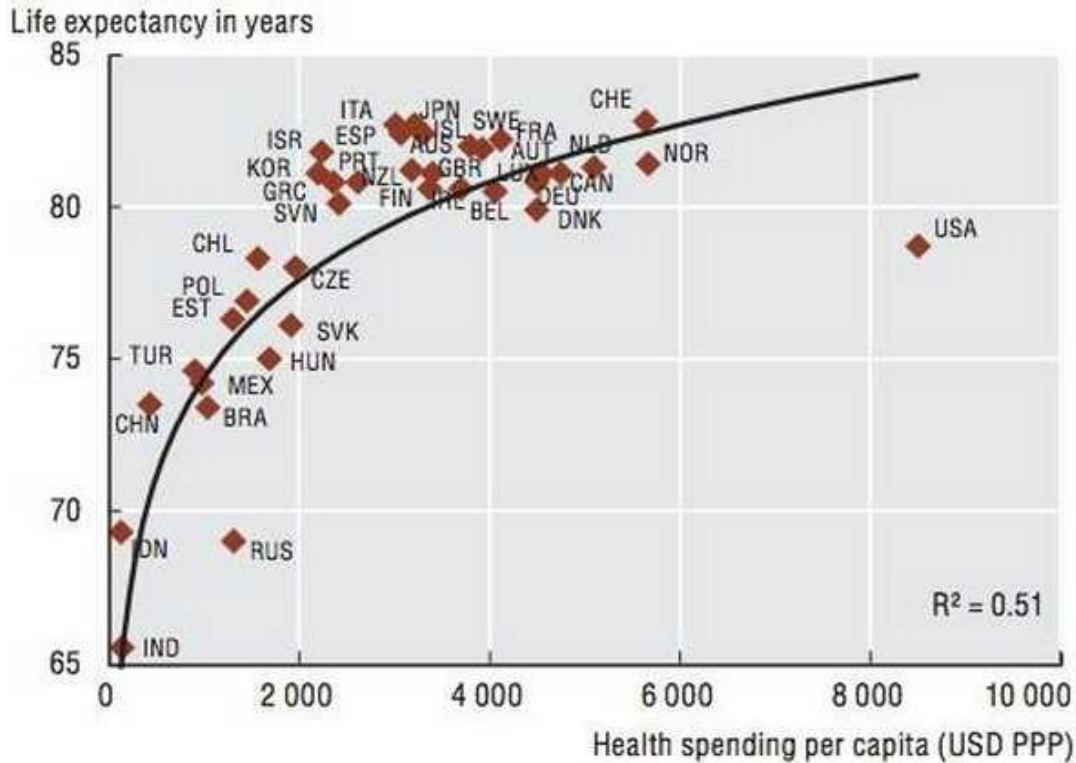


This work was created in an open classroom environment as part of a program within the Sutardja Center for Entrepreneurship & Technology and led by Prof. Ikhlaz Sidhu at UC Berkeley. There should be no proprietary information contained in this paper. No information contained in this paper is intended to affect or influence public relations with any firm affiliated with any of the authors. The views represented are those of the authors alone and do not reflect those of the University of California Berkeley.

Healthcare Industry: Inefficiencies

Good health is fundamental to human existence and productivity. Needless to say, humans are always in search of medical interventions and processes that would enable us to lead healthy and long lives. This need is reflected in the amount of money that we spend on healthcare. In 2015, the amount of money spent per person on healthcare in the USA reached ~10K, making it a \$3T industry that in 2020 is projected to hit \$12T. Despite the enormous amount of money that is being poured in, the healthcare industry, specifically in the US, is rife with inefficiencies. In this report, we will attempt to highlight some of the inefficiencies in the healthcare system in the US and attempt to address how big data can be used to address these inefficiencies. We will also identify the top societal challenges to implementing big data solutions and conclude with a prediction on who stands to win or lose in the attempt to use big data for healthcare.

Despite spending ~17% of GDP on healthcare, the average life expectancy at US is fortieth ranked in the world.



Source: OECD Health Statistics 2013

A number of factors have been identified on what causes the US healthcare system to be as inefficient. One estimate, claims that 30% of the money spent on healthcare is wasteful, which easily translates into billions of dollars in savings that could be diverted to core medical research. Some of the commonly cited reasons¹ for the increased spending are:

1. Practice of defensive medicine: Where inappropriate or unnecessary tests are done primarily to protect against lawsuits.
2. Inefficient healthcare administration: Healthcare industry is notorious in having the largest proportion of non-expert administrators. By one estimate, in every hospital ~30% of the employees are in purely administrative roles.
3. No focus on preventative medicine: There is very little incentive for practitioners to focus on prevention (low cost, high impact) as opposed to treatment (high cost, high impact).
4. Payment by third-parties: Since the cost for particular treatments come from insurance companies, there is very little incentive for patients to choose low cost alternatives.

¹ <http://www.pwc.com/us/en/healthcare/publications/the-price-of-excess.html>

While there are several detailed reports² on what makes the healthcare industry inefficient, in this report, we want to focus specifically on the problem of how to leverage "big data" to help remove or alleviate some of these inefficiencies. In the following sections, we will first discuss big data in the healthcare industry and then identify the healthcare areas that can leverage this data to become more efficient.

Healthcare Data

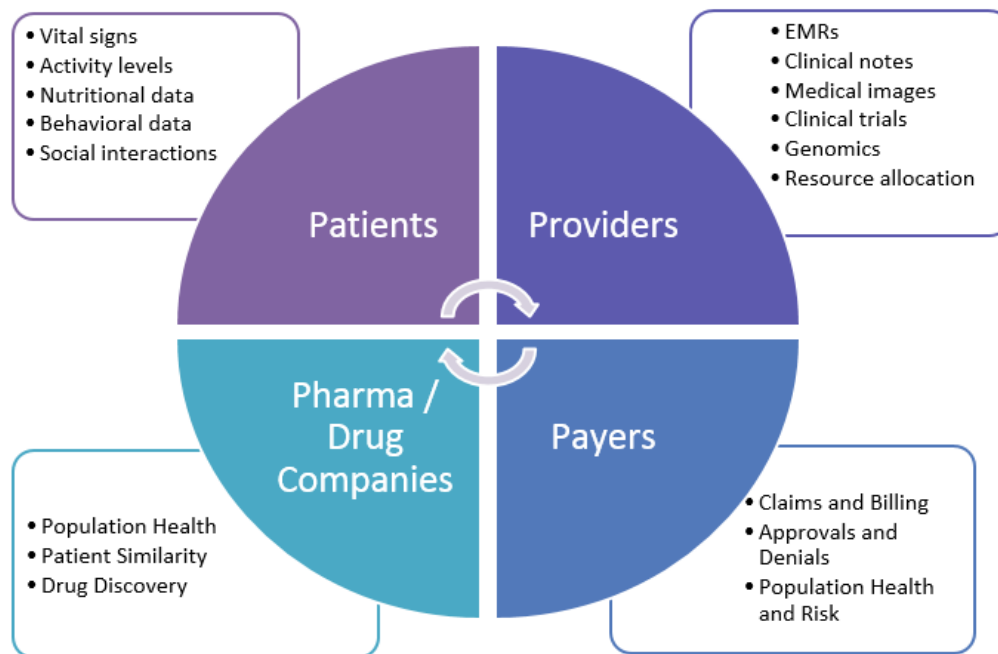
Before healthcare data can be shared, processed, mined, visualized, etc... it must be collected. Today this happens through a dizzying array of channels. A doctor may scribble a patient's answer in their clinical notes which are then digitized and added to an electronic health record. Information in these records can also be produced from lab tests, prescriptions, medical imaging, etc...

Data types range from structured to unstructured where *structured data* has a well defined representation making it amenable to data processing whereas *unstructured data* is far more difficult to leverage. Examples of each include,

- ❖ Structured data
 - International Classification of Diseases Codes
 - Current Procedural Terminology
 - Lab work
 - IoT sensed (e.g., personal heart rate monitor)
- ❖ Semi-structured
 - Medications
- ❖ Unstructured
 - Clinical notes
 - MRI, CT-scan, imaging
 - Social media data

A widely accepted metric is that *80% of all healthcare data is unstructured* meaning that this data cannot be trivially used in big data systems. Additionally this data is fragmented and siloed in numerous formats, organizations, and stakeholders. As illustrated below, healthcare data has several key stakeholders and owners: patients, providers (e.g., doctors and hospitals), payers (i.e., insurance companies), and pharma / drug companies. Each of these stakeholders primarily owns, cares about, and handles specific types of healthcare data.

² http://www.nehi.net/writable/publication_files/file/waste_clinical_care_report_final.pdf

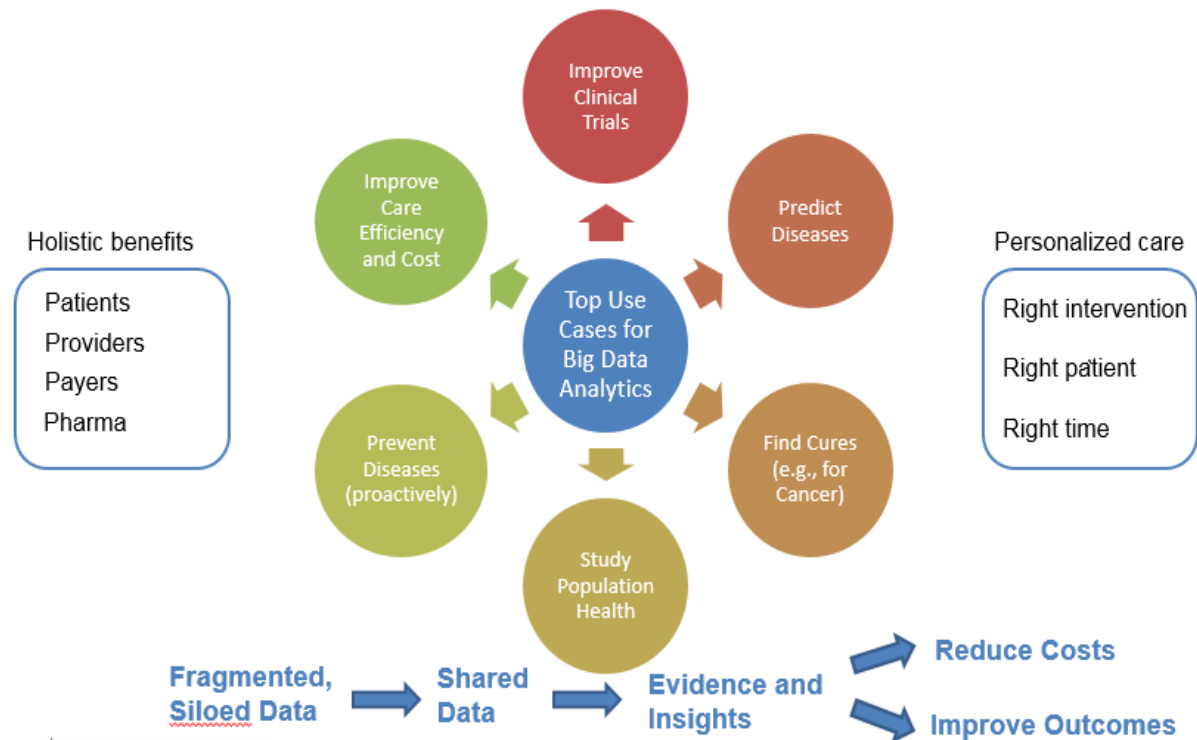


As an example, patients typically generate and are concerned with vital signs, physical activity levels, nutritional information, and online social interactions. Providers, however, work extensively with electronic medical records (EMRs), clinical notes, and medical images while also concerned with improving resource allocation and usage. Similarly, payers typically deal with claims processing (approvals and denials), billing, and population health and risk assessment. The total amount of data in healthcare is growing rapidly as well, “in 2012, worldwide digital healthcare data was estimated to be equal to 500 petabytes and is expected to reach 25,000 petabytes in 2020”³.

This mix and explosion of data requires the full focus of emerging big data tools such as text mining and natural language processing to give structure to unstructured data, visualization for insights, machine learning for predictions, etc.

³ Health-care hit or miss? Nature. 2011;470(7334):327–329. [[PubMed](#)]

Key Use Cases - Areas of Opportunities



Based on our research and analysis⁴, we've identified 6 top use cases that can benefit significantly from big data analytics of healthcare data (i.e., cognitive computing⁵) to process fragmented and siloed data, derive meaningful, actionable insights, and drive cost reductions and improved outcomes. These are:

1. Improve care efficiency and reduce cost
2. Improve the efficiency, cost and effectiveness of clinical trials
3. Find cures for diseases such as Cancer
4. Predict the risk and likelihood of diseases
5. Data ownership and sharing to study population health
6. Prevent diseases through proactive lifestyle changes

Taken together, these use cases promise to achieve holistic benefits for each of the primary stakeholders in healthcare, and at the same time deliver personalized care with the right intervention to the right patient at the right time. We focus on each of these top use cases in the sections below.

⁴ <http://www.healthcarediver.com/news/5-ways-artificial-intelligence-is-changing-the-face-of-healthcare/415424/>

⁵ <https://www.explorys.com/news/is-cognitive-computing-in-health-care-ready-for-prime-time>

Improve Care Efficiency and Cost

Health care that is efficient allots the optimal level of care, while minimizing expenditures. Expenditures in the case of health care efficiency are not limited to monetary costs, however. While monetary cost can be considered as an input, research from both the Agency for Healthcare Research and Quality (AHRQ)⁶ and the American Medical Association (AMA) caution against considering a cost-effectiveness analysis or ratio to be analogous to efficiency of care delivery. Efficiency is also one of the domains of quality developed by the Institute of Medicine. Another way to conceptualize efficiency, however, is to understand it as the reduction of waste, such as reducing prescriptions of antibiotics for children in cases where they are not necessary⁷.

- ❖ About 85 percent of children will never need specialty care, so we've developed a network that is about 80 - 85% primary care and 15-20% specialty – and of course, the hospital is pretty much dedicated to specialty situations.
- ❖ In the high-volume area of physician billing, we have ~750,000 claims a year, so getting a great first-pass rate is huge for us, because then we're cleaning up fewer problems.
- ❖ If you get care in the hospital, it's inherently a lot more expensive than other settings. First of all, the charges are a lot higher and you have a lot more overhead. But also, it's the way the insurance processes the co-insurance and other parts of the patient responsibility⁸. That can be very difficult for a lot of patients to manage.

Examples of care efficiency :

- ❖ Optimal utilization of resources
- ❖ Reduce prescription of antibiotics for children
- ❖ Charges based on claim pattern
- ❖ Reduce care cost by Genomic analytics
- ❖ Decreased claim denials
- ❖ Revenue cycle management⁹

A Case Study 1 - Resource Optimization:

⁶<http://www.ahrq.gov>

⁷ <http://www.hrsa.gov/healthit/toolbox/childrenstoolbox/improvingquality/efficiencyofcare.html>

⁸ <http://healthitanalytics.com/news/hca-obamacare-brings-accountable-care-patient-responsibility>

⁹ <http://revcycleintelligence.com/>








For the case study we chose the startup LeanTaas¹⁰ that specializes in providing resource optimization to improve the **Care Efficiency** using their **Machine Learning** platform carried out in Stanford Infusion center and Colorado Cancer Center. Objectives of the case study were to overcome:

- ❖ Imbalance between Infusion appointment, supply and demand
- ❖ Frequent mid-day peaks
- ❖ Unpredictable long wait
- ❖ Sub-optimal utilization of resources
- ❖ Not able match usage profile lead to over-staffing

Results of using the LeanTaas platform:









- ❖ **Stanford Infusion¹¹:**
 - 25% higher patients, at 15% lower costs
 - 30% wait time reduction at mid-day.

 LeanTaas  iQueue used advanced data science

5 types of infusion appointments...					
	1 hour	2 hour	3-5 hour	6-8 hour	9+ hour
...each with a different daily volume of patients...	20	20	15	10	5
...resulting in a total of 256 possible appointment slots...	 64 slots 7:00, 7:10, etc. until 5:30pm	X			Max of 4 patients can start at a given time

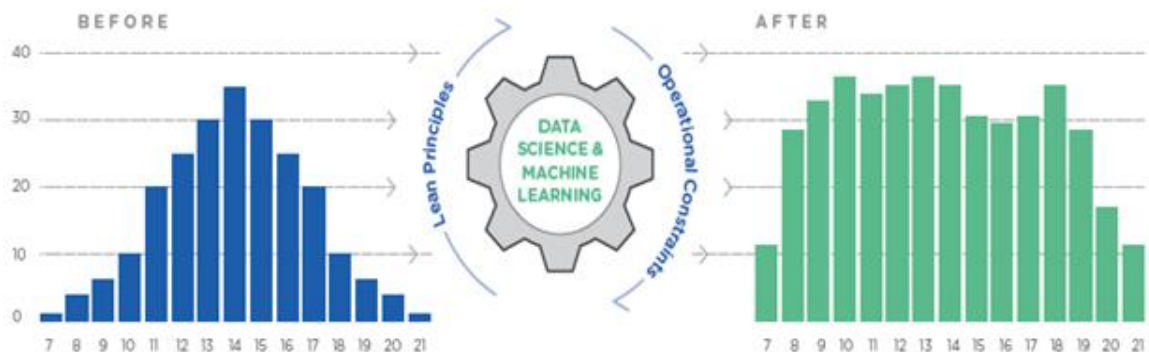
¹⁰ <http://www.leantaas.com/>

¹¹ http://www.leantaas.com/uploads/iQueue_CaseStudy_Stanford.pdf

5 types of infusion appointments...					
	1 hour	2 hour	3-5 hour	6-8 hour	9+ hour
...each with a different daily volume of patients...	20	20	15	10	5
...resulting in a total of 256 possible appointment slots...	 64 slots 7:00-7:10, etc. until 5:30pm   Max of 4 patients can start at a given time				

❖ **Colorado Cancer Center¹²:**

- 16% higher patients
- 28% lower overtime
- Significant staff satisfaction



Cost:

Purchasing health insurance is not easy. Collectively, American health insurance companies offer thousands of different policies through hundreds of intermediaries. While many health insurance consumers' choices are circumscribed by the whims of

¹² http://www.leantaas.com/uploads/iQueue_CaseStudy_Colorado.pdf

their employers' group health insurance¹³ providers, individuals who must purchase health insurance on the open market¹⁴ are liable to face a dizzying array of choices¹⁵.

If you're looking for affordable health insurance, you'll need to take an honest look at your demographic profile and medical history. Young, healthy people need to pay far less for health insurance than their older counterparts. However, it should be noted that many health insurance companies offer less-than-ideal coverage for their younger customers.

The increased cost of health insurance is a central fact in any discussion of health policy and health delivery. Annual premiums edged beyond \$17,500 for an average family. For those Americans who are fully-covered, these cost realities affect employers, both large and small, plus the "pocket-book impact" on ordinary families. Yet for 2016 among the roughly 85 percent of HealthCare.gov consumers with premium tax credits, the average monthly net premium increased just \$4, or 4 percent, from 2015 to 2016, according to an HHS report¹⁶.

2016 costs at a glance ¹⁷	
Part A premium	Most people don't pay a monthly premium for Part A (sometimes called "premium-free Part A" ¹⁸). If you buy Part A, you'll pay up to \$411 each month. Calculate my premium ¹⁹ .
Part A hospital inpatient deductible and coinsurance	You pay:

¹³ <http://thelawdictionary.org/group-health-insurance/>

¹⁴ <http://thelawdictionary.org/open-market/>

¹⁵ <http://thelawdictionary.org/article/how-much-does-health-insurance-cost-for-the-average-adult/>

¹⁶ http://www.ncsl.org/research/health/health-insurance-premiums.aspx#Exchange_premiums

¹⁷ <http://www.ncsl.org/research/health/health-insurance-premiums.aspx>

¹⁸ <https://www.medicare.gov/your-medicare-costs/part-a-costs/part-a-costs.html>

¹⁹ <http://medicare.gov/eligibilitypremiumcalc/>

	<ul style="list-style-type: none"> • \$1,288 deductible for each benefit period • Days 1-60: \$0 coinsurance for each benefit period • Days 61-90: \$322 coinsurance per day of each benefit period • Days 91 and beyond: \$644 coinsurance per each "lifetime reserve day" after day 90 for each benefit period (up to 60 days over your lifetime) • Beyond lifetime reserve days: all costs
Part B premium	Most people pay \$104.90 each month.
Part B deductible and coinsurance	\$166 per year. After your deductible is met, you typically pay 20% of the Medicare-approved amount for most doctor services (including most doctor services while you're a hospital inpatient), outpatient therapy, and durable medical equipment.
Part C premium	The Part C monthly premium varies by plan. Compare costs for specific Part C plans²⁰.
Part D premium	The Part D monthly premium varies by plan (higher-income consumers may pay

²⁰ <https://www.medicare.gov/find-a-plan/questions/home.aspx>

more). Compare costs for specific Part D plans²¹.

Providers :

Because of the significant out-of-pocket payments required by traditional Medicare, a booming market of private-sector insurance products has grown up around the government programs. These Medicare-related insurance products are one of the fastest-growing segments of the U.S. health insurance industry, and they are the part of the market on which a smart consumer should focus his or her attention. Medicare Providers is here to help seniors, and other Medicare eligible individuals, understand these products and provide tools to assist in the decision making process.

Most popular providers are Aetna, UnitedHealthCare, Humana, BCBS, Cigna, Anthem, AARP, and one can find the specific provider in their area at medicare-provider.net²².

Improve Clinical Trial Efficiency

Clinical trials refer to biomedical or behavioral research study of human subjects to answer specific questions about interventions (vaccines, drugs, treatments, devices, or new ways of using known drugs, treatments, or devices) to improve state of health. The trials are conducted by drug manufacturers under the purview of FDA. Each clinical trial goes through four phases as outlined in the table below²³.

Phase (% of drugs passing Phase)	Details	Duration	Size (Number of participants)

²¹ <https://www.medicare.gov/find-a-plan/questions/home.aspx>

²² <http://www.medicare-providers.net/>

²³ <http://www.fda.gov/ForPatients/Approvals/Drugs/ucm405622.htm>

Phase 1 (70%)	Evaluate safety & observe bodily reactions. Identify safe dosage range and possible side effects.	Several months	20-80
Phase 2 (33%)	Determine efficacy and further evaluate safety with randomizations in test subjects	Months - few years	100's
Phase 3 (25-30%)	Compare against other standard interventions, monitor adverse effects and establish safe use. Request FDA approval if successful.	Several years	1000's
Phase 4	Monitor effectiveness in the general population and collect information about adverse effects	Ongoing	General Population

Inefficiencies in Clinical Trials:

Cost of running a trial: Clinical trials are typically funded by the federal government (ex. National Institutes of Health) and by privately held pharmaceutical and biotechnology companies. There are wildly varying estimates on the cost of developing a drug, varying from \$800M to \$2.6B per drug. A majority of the costs comes from drug discovery and development (48% of the cost), foregone returns in investment while the drug is in development (41% of the cost) and additional research & development needed after release (1% of the cost).

Success rates: As can be seen from the table above, the percentage of drugs that can successfully pass all four phases of clinical trials is less than 10%. Besides the low success rate, the time it takes to complete a trial can also be 10-15 years.

There are a number of factors that can contribute to the high cost and low success rate:

1. **Enrolling test subjects:** Identifying humans to participate in clinical trials is a laborious process. Around 10-15% of the trials do not manage to enroll even a single patient and about two-thirds fail to meet their original goals for the study. The most common factors are fear among test subjects of long

term damage to health and unequal payments (test subjects have no idea about how “risky” a trial is). Besides drugs that intend to target a specific condition needs to identify patients who are willing to get on an experimental drug sometimes at the expense of pausing or interfering with their regular treatment.

2. **Collecting data from test subjects:** During the administration of the clinical trial, the test subjects are required to show up at the clinics frequently for several weeks in order to monitor their overall health, gather data on the effect the drug has on metabolic functions (for example, by monitoring concentration in bodily fluids) and transmit the data along to the principal investigator/researcher who is conducting the clinical trial.
3. **Improper trial set-up:** The ideal group of test subjects should be a uniform sample drawn from several different dimensions (for example, genomic, lifestyle and metabolic levels). Not accounting for any of the confounding factors could end up in test results that are difficult to interpret and extrapolate from. However, it is only very recently that data along all of these different dimensions is even being gathered at an individual level.

Big Data to address inefficiencies:

With the availability of real-time and comprehensive data and the ability to quickly transmit the data there are a number of possible ways by which big data can be leveraged to improve the inefficiencies in clinical trials. While all of these proposals are fine in theory, the actual implementation still runs into blockers that are ubiquitous to healthcare data overall.

1. **Better randomization of test subjects:** As highlighted earlier, one of the challenges in conducting a successful clinical trial is to select subjects uniformly across different dimensions. As an example, understanding the cell genotype and the metabolic response to drug molecules will enable researchers to identify candidates who are likely to respond well to a particular treatment²⁴. In addition, monitoring biomarkers during the course of a trial can allow patients to be moved in and out of treatment and control groups. An added advantage is developing a more comprehensive profile of the control subjects which can be shared amongst different clinical studies so that they can be enrolled quickly²⁵.

²⁴<http://radar.oreilly.com/2013/08/cancer-and-clinical-trials-the-role-of-big-data-in-personalizing-the-health-experience.html>

²⁵<http://social.eyeforpharma.com/research-development/leveraging-big-data-simplify-clinical-trials-and-fix-rd>

2. **Leveraging wearables:** The quick rise in the adoption of wearables, smart applications and at-home monitors that can transmit to a central location have all made it possible to gather real-time data, even in between clinical visits on the patient's response to treatments. This data can help in two ways a) with more data the time taken for a trial to converge can be shortened b) the investigators/researchers can respond more quickly to issues that occur during the trial. Another cost savings is where the test subjects do not have to rely on human clinicians to gather data, instead relying on the wearable/at-home monitors to constantly transmit the data in an automated fashion. The savings from removing humans from the data gathering loop more than offsets the cost of transmitting and storing the additional data.

Some players in the Big Data for Clinical Trials space:

- **GNS Healthcare²⁶:** Uses machine learning and statistics to create software models that let users predict the outcome of “what if” scenarios. This helps determine which treatment or line of action will be best for individuals and for the health system as a whole.
- **Ayasdi²⁷:** Visualizes data in a multidimensional graphic that readily shows outliers as well as high and low-response groups in the data, even without pre-specifying the characteristics of those clusters. The outliers can represent unknown biomarkers, or subgroups of patients that would be well (or poorly) suited to a clinical trial of a particular drug.
- **Medidata²⁸:** Uses a SaaS platform to create a “Clinical Cloud” (8B clinical records from ~2M patients, across 9K studies. Adding 500K records daily). They help streamline the clinical trial process (data capture, data transfer, visualization, train researchers, patient monitoring etc.)
- **NuMedii²⁹:** Uses data mining to correlate disease information and drug data to predict drug efficacy. The company's database includes billions of points of disease, pharmacological, and clinical data, and it mines this using network-based algorithms. This should de-risk the development process, increasing the chance of drugs making it through to the market.
- **MedChemica³⁰:** Uses data mining of pre-competitive shared data while

²⁶ <http://www.gnshealthcare.com/solutions/foundations/>

²⁷ <http://drbonnie360.com/post/53413359353/ayasdi-using-the-power-of-math-to-solve-medical>

²⁸ <https://www.mdsol.com/en/what-we-do>

²⁹ <http://numedii.com/technology/>

³⁰ <http://www.medchemica.com/>

maintaining the security of each individual partner's intellectual property.

Find Cures

We are at inflection point. Genomic research combined with Government's initiative to make illegible handwriting into electronic health record database has opened up several avenues of research & innovation. With recent advances in knowledge, technology, computational power, we are at tipping point and there is a real possibility to compliment biological research with technology in identifying cures for diseases like Cancer, Alzheimer, and Diabetes.

Many companies are squeezing every last algorithmic drop out of big data to help them make more informed decisions, accessing and analyzing voluminous flows of structured and unstructured market data to guide the optimal allocation of resources. This same promise now beckons for medical research to provide new insights into finding cures for the most deadly of diseases.

Let's look at impact of Big Data on 2 most devastating diseases- Cancer and Alzheimer that many people throughout the world have no choice but to confront.

A. Cancer:

Summary: Every year, 8 million people around the world die from cancer, and 14 million people find out they have the disease. Cancer is genetic disease and won't respond to 20th century one-size-fits-all treatment.

Initiative: DNA sequencing is helping doctors and researchers take major leaps forward in understanding the complexity of the disease. Decoding genomes is helping identify the specific mutations that drive cancer growth. The American Society of Clinical Oncology (ASCO), a non-profit, professional organization of oncologists, is on the brink of translating huge amounts of raw cancer data into knowledge that can be applied to best help future patients. Latest report from American Cancer Society show that cancer mortality is down more than 20% over the past 20 years. Many patients are living longer thanks to better treatment and earlier detection. Science is tipping the odds of survival in favor of patients.

Case Study:

CancerLinQ is a cutting-edge health information technology platform that will provide real-time quality feedback to providers, feed personalized insights to doctors, and uncover patterns that can improve care. Similarly, American Association for Cancer Research's (AACR) Project GENIE is an international data-sharing project. By aggregating clinical-grade sequencing data, GENIE will improve patient treatment decisions and catalyze clinical and translational research.

B. Alzheimers:

Summary: Alzheimer's is a global crisis. Nearly 44 million people worldwide suffer from dementia, and this number will spike to 115 million by 2050. The crisis is not particular to the rich world. Nearly 60 percent of the burden of dementia is in low- and middle-income nations. And this percentage will rise.

Initiative: The Global CEO Initiative on Alzheimer's Disease (CEOi) has taken up Big Data challenge to identify best bouquet of predictors of cognitive decline and of neuroprotective genotypes. Technology is leveraged for semantic Analysis and Machine Learning for Early Detection of Alzheimer.

Case Study: Mobile applications or software that can track cognitive function over time could greatly reduce the amount of time doctors spend diagnosing patients. One of many such programs, Lumosity³¹, provides a brain game platform, and data from user scores could be valuable to correlate with diagnosis and conversion to Alzheimer's disease. This may lead to efficient early detection, which is essential if we are to intervene in neuronal loss before it becomes irreversible.

Clear Hope:

1. Discover bio-markers & Improve efficacy of treatment

Clinical trials are not effective to prove efficacy of targeted therapy drugs. Trials of N=1 mostly end up in failure. ASCO's CancerlinQ³² initiative connects Doctors with Data helping them to make informed decisions in real time. By capturing data on a multitude of cancer patients and analyzing their tumors in depth through sequencing and other techniques, this information can help researchers characterize the various tumors. This, in turn, allows them to create more targeted therapies per characterized tumor &

³¹ <http://www.lumosity.com/>

³² <http://cancerlinq.org/>

individual. Through predictive analytics, we can better understand diseases and pre-emptively treat them, ultimately on a more personalized basis.

2. Individualized Treatment & Prognosis

Diseases are heterogeneous in their causes, rates of progression, and their response to drugs. Simply put, each person's disease is potentially unique, and therefore, that person needs to be treated as an individual. Knowing a patient's prognosis allows medical teams to decide how aggressively to treat each cancer and what steps to take after a tumor is gone. Having access to up-to-date information on how thousands of other similar patients have reacted to proposed treatment plan will enable Doctors to tailor treatment to individual patients and provide best chance of positive outcome. Health care will become very personal. Rather than a particular drug for everyone to take for a particular disease, the drug itself will be particular — to you.

3. Drug Discovery

The ability to sequence the DNA of large numbers of tumors³³ has allowed researchers to understand the genetic changes underlying certain cancers. Scientists can use this information to help test potential new medicines that might target some of these genetic changes or drivers of cancer growth. Big data sets from preclinical studies are being used to help predict which medicines or combinations of medicines might be good candidates to move forward into clinical trials in humans.

4. Dr. Gini (AI- Driven Diagnosis)

IBM's Watson Analytics engine to match cancer patients with treatments that most likely to help them. As well as recommending the relevant cancer drug most likely to treat a particular patient's cancer, Watson can even recommend drugs that have not been used to treat cancer before. Since it is programmed with specific details of how thousands of medicines interact with the human body, Watson can suggest anything which it thinks might interact beneficially with the cell affected by mutation which is causing the cancer. Of course, a doctor will probably have to take many other issues into consideration before prescribing whatever the AI-driven Watson suggests, but it surely will speed up the process. Idea is to answer questions that you

³³ <http://cancergenome.nih.gov/abouttcga/overview>

didn't know to ask.

5. Public

Health

Epidemiologists study the causes and patterns of human diseases including cancer. In the era before big data, they discovered that smoking causes the vast majority of lung cancer cases. Now big data is allowing them to answer even bigger questions in cancer research. This new era³⁴ of epidemiology takes advantage of the availability of large collections of hospital records and genomic data, leading to new insights into diverse cancers in diverse populations.

In nutshell, it's a fight between health science & technology and far from won but significant advances are being made. Just this year a study concluded that thanks to the advances in spotting and treating cancer, by 2050 no one under 80 will be dying from the disease. Big Data-powered research and treatment programs will undoubtedly play a part in that victory, just as they continue to give us answers in every field of science.

Predict Diseases

Case Study 2 - a collaborative engine for practical disease prediction

At the University of Notre Dame, Nitesh V. Chawla, an associate professor of computer science, and his doctoral student Darcy A. Davis, recently developed the Collaborative Assessment and Recommendation Engine (CARE)³⁵. The primary purpose of the system is to sift through mountains of health population data and capture patient similarities to generate a personalized disease risk profile for individual patient.

The monumental cost of health care, especially for chronic disease treatment, is quickly becoming unmanageable. This crisis has motivated the drive towards preventative medicine, where the primary concern is recognizing disease risk and taking action at the earliest signs. Collaborative Assessment and Recommendation Engine (CARE) which relies only on patient's medical history using ICD-9-CM codes

³⁴ http://epimonitor.net/Samet_Delivers_NIH_Gordon_Lecture.htm

³⁵ Predicting individual disease risk based on medical history - DA Davis, NV Chawla, N Blumm, N Christakis, Albert-Laszlo Barabasi

in order to predict future disease risks. CARE uses collaborative filtering methods to predict each patient's greatest disease risks based on their own medical history and that of similar patients. Each user is a patient whose profile is a vector of diagnosed diseases. Using collaborative filtering, it can generate predictions on other diseases based on a set of other similar patients.

Database comprises the Medicare records of 13M elderly patients in the United States with a total of 32M hospital visits. The testing patient is the individual for whom CARE is making predictions based on the histories of the training patients. Thus, the individual medical history is the testing patient while other patients' medical histories is the set of training patients. Collaborative filtering is performed on the resulting group, generating predictions for the future visits of the testing patient. The output after CARE is a ranked list of diseases for the subsequent visits of the testing patient, ranked in order from the highest risk score to the lowest. If a sampling of future diagnoses can be provided to a practitioner, appropriate medical tests can be ordered sooner and lifestyle adjustments can be adopted by the patient proactively. This will not only result in improving the quality life for the patient, but also in reducing the health care costs.

Case Study 3 - Novel Predictive Models for Metabolic Syndrome: Aetna

A new study by Aetna's Innovation Labs and GNS Healthcare uses "big data" analytics to predict patients at risk for metabolic syndrome³⁶. Metabolic syndrome is the name for a group of five risk factors that raise your risk for heart disease and other health problems, such as diabetes and stroke:

- Large waist size
- High blood pressure
- High triglycerides
- Low high-density lipoprotein
- High blood sugar

Patients who exhibit three of these five factors are classified as having metabolic syndrome. People who have metabolic syndrome are 2x as likely to have a heart attack or stroke and are 5x as likely to develop diabetes as those who don't. All together, these conditions account for ~20% of healthcare costs in the United States. As a managed healthcare company, Aetna is very interested in personalized

³⁶ Steinberg et al. Novel Predictive Models for Metabolic Syndrome Risk: A "Big Data" Analytic Approach. Am J Manag Care. 2014 Jun 26

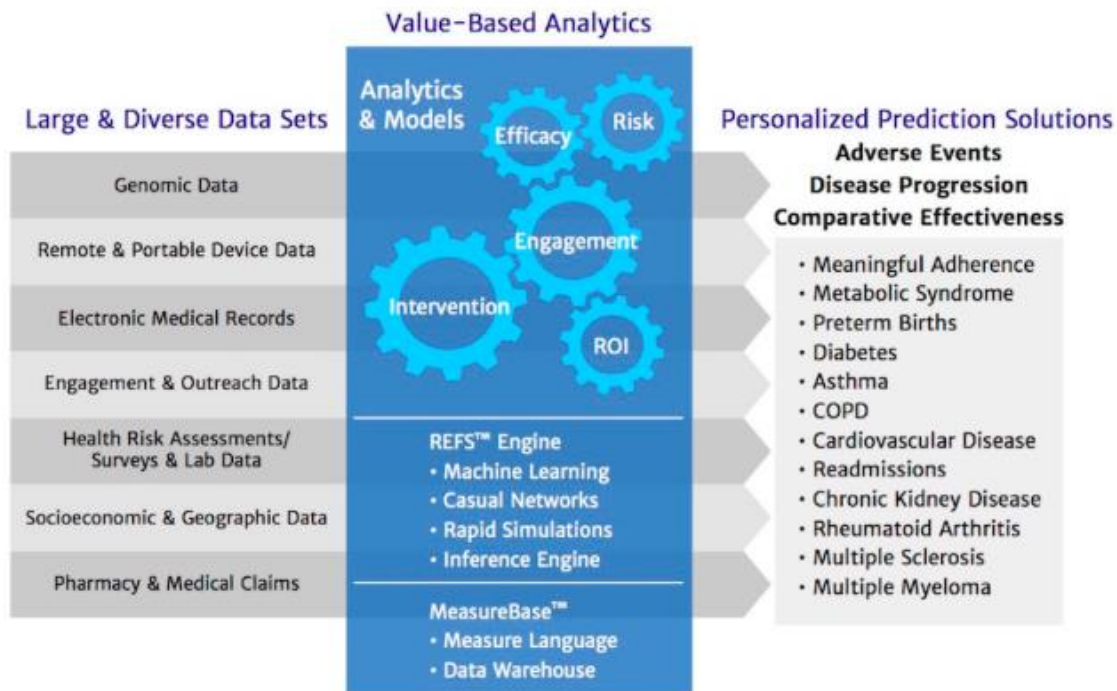
interventions that could reduce risk and decrease costs associated with metabolic syndrome.

Aetna's Innovation Labs collaborated with GNS Healthcare to build computer models and evaluate information from 37,000 members of one of Aetna's employer customers. The information included medical and pharmacy claims, demographics, lab tests and biometric screening results (blood pressure and cholesterol) over a two-year period. The models were used to predict future risk of metabolic syndrome on both a population and an individual level. Detailed risk profiles were built for each subject that included which combination of the five metabolic syndrome factors that person exhibited and are at risk for developing. The models were also be used to create personalized exercise, weight management, and care management programs.

GNS Healthcare³⁷ MAX platform drives precision-targeted personalized interventions that maximize financial ROI while optimizing clinical outcomes. MAX can uniquely integrate very large and diverse data sets, including medical and pharmacy claims, lab results, EHR, intervention engagement and outreach, survey, socio-economic and genomic data to construct highly accurate predictions of intervention outcomes.

³⁷ <http://www.gnshealthcare.com/technology-overview/technology/>

GNS Healthcare MAX™ Architecture



Data Ownership and Sharing

Many data collection methods are relatively “high-touch” and involve a large commitment from the patient. For example, to answer a doctor’s question the patient is typically physically present in the doctor's office.

Ubiquitous access to the internet and always connected devices has created a new channel for healthcare data creation. Devices such as smartwatches or healthcare apps are “low-touch” in that they allow for the collection of healthcare data with very little or no burden on the patient.

- **Case: Apple HealthKit, ResearchKit and CareKit.** HealthKit is a SDK which addresses the collection and display of sensor data from apps. ResearchKit allows medical researchers to create apps with an aim of capturing and sharing meaningful research data. CareKit is a framework for creating apps

that allow patients to manage medical conditions and share them with their care team.

- **Case: Conversa Health.** Similar to ResearchKit plus CareKit, Conversa provides system that allows patients to continue the conversation with their providers in a very lightweight manner. The app allows patients to do things such as track progress and answer simple questions. These “digital checkups” can save patients and providers vast amounts of time. The providers are able to view individual progress to identify needs for intervention and also track aggregate population trends.

Trend: Greater ownership of data by patients

We see a trend and opportunity in patients taking more ownership of their data. This means an expectation to be able to view, analyze and share their own health data. Two major drivers of the trend are presented.

Trend Driver #1: Patient owned data collection devices

Healthcare data is increasingly being captured and shared by devices owned by the patients themselves. For example a patient may purchase a smartwatch and enable continuous heart rate monitoring. Since the patient is responsible for purchasing and using the device they will have a strong expectation that they should be able to access the data produced by that device.

Trend Driver #2: Improved care

Greater patient involvement and ownership of data can lead to improved care. The OpenNotes initiative is a research project aimed at understanding the benefits of sharing doctors clinical visit notes. Details from the project show major improvements³⁸ in care such as,

- “More than 60% of patients reported doing better with taking medications as prescribed because of open notes”
- “More than 77% of patients reported that open notes helped them feel more in control of their care”

³⁸ <http://www.opennotes.org/what-is-opennotes-2/why-open-visit-notes/>

An excellent example of a company capitalizing on this trend is **PicnicHealth**³⁹. For a fee PicnicHealth will do the “legwork” of collecting, standardizing and presenting a customer’s health data. Having data like this in the hands of patients, in standard formats will drive the need for tools to analyze and share this data.

Secondary Trend: Sharing of data by patients

The trend of greater patient ownership of data will drive a secondary trend of patients *sharing* that data. As patients own increasing amounts of data they will have more opportunities and incentives to donate or be compensated for this data. For example a patient may choose to make their heart rate monitor data available to a research team for free or for a fee. This will create markets for health data and the need for platforms to match patients selling data with researchers buying data.

An excellent early example of such a platform is **HealthBank**⁴⁰. HealthBank is a citizen-owned cooperative based in Switzerland which allows for users to store their health data and share with providers, family members and researchers. This “health data transaction platform” gives patients the ability to share, monetize, track and revoke data at a highly granular level.

We expect to see **strong, positive cross-side network effects** with health data sharing where patients seeking to be compensated will produce and share more data, leading to more researchers buying this data and increasing the value of the data which will lead further to increase the patient desire to produce it.

Prevent Diseases

The earlier use cases pertain to diagnosing, treating and helping patients and the population at large when suffering from or at risk from diseases or other medical conditions. This use case is about preventing diseases altogether (or at least delaying their onset significantly) through proactive intervention and lifestyle changes via a combination of actionable insights, recommendations, engagement and formation of good habits. A core aspect of addressing this use case is to collect and aggregate user-generated data, infer insights from this data and correlate meaningfully with population health, and deliver the insights with action plans to consumers in a format that they’d find easy-to-consume and highly engaging. The

³⁹ <https://picnichealth.com/>

⁴⁰ <https://www.healthbank.coop/>

focus is not just on losing weight or eating less / right or exercising more, but on achieving more holistic improvements in lifestyles and daily choices.

Note here that we refer to consumer as “user” and not just “patient” - this is because the use case applies also to those consumers who don’t currently suffer from specific diseases that require medical intervention.

Relevant user data

This includes a variety of user information from individuals, data generated by multiple tracking devices, sensors and apps per individual, and derived data based on this information. For example:

- Profile: gender, height, weight, age, BMI, and ethnicity
- Vital signs: heart rates, blood pressure, cholesterol and glucose levels
- Activity: calories consumed, calories burned, exercise frequency and intensity, diet, nutritional composition, meal frequency, hydration, sleep pattern and depth/quality, respiratory rate
- Other physiological and behavioral indicators such as mood, stress levels, depression, etc.

User data collection and aggregation

With the advent of quantified self movement, increasing adoption of apps, wearables and sensors⁴¹, there is a lot of continuously generated (not just episodic) user data to capture, aggregate, and analyze in real-time to generate insights and make informed predictions. As mentioned elsewhere, there are emerging mobile kits, platforms and services that enable centralized data collection and aggregation from a wide set of devices and apps, and sharing anonymized (de-identified) data with others so that this can be aggregated and analyzed at the population level.

User and population data analytics and analysis

Digital biomarkers⁴² are defined as consumer-generated physiological and behavioral measures collected through connected digital tools. From vital signs, activity levels and digital biomarkers, the goal is to extract clinically relevant info from data recorded using wearables and generate “phenotypic signatures” to establish a user baseline that can be a reference point for what is considered normal or expected. Using such a baseline and data analytics, one can then look for what the user should be doing to improve lifestyle, derive intra and inter-attribute correlations within the user’s data, and also compare with prior trends and observations. This helps detect if something has changed from the normal pattern, identify possible avenues to explore further, and provide suggestions to the user for better awareness and non-medical guidance to encourage a healthier lifestyle.

⁴¹ <https://rockhealth.com/reports/digital-health-consumer-adoption-2015/>

⁴² <https://rockhealth.com/reports/the-emerging-influence-of-digital-biomarkers-on-healthcare/>

Another compelling dimension of data analytics to prevent diseases is to correlate and compare the individual user's data with that observed with a population similar in profile to this user. This can help provide insights into how a user's lifestyle compares with that of the population (e.g., do any risks apply and what is specifically relevant for me?), identify generalized recommendations that seem effective at the population level, and offer customized recommendations to the user with proper rationale to encourage and motivate. When presenting to consumers, it is important to validate findings and recommendations with medical experts, and confirming clear linkage to clinical research and studies when possible.

User engagement, incentives and compliance

To achieve effective outcomes for individuals, just having digital health tools and data from wearables, sensors and apps won't suffice⁴³. Behavior change is key for sustained long-term user engagement through habit formation, social motivation, and goal reinforcement⁴⁴. Also important is providing a great user experience and compelling incentives that drive initial and continued engagement.

In addition to many leading apps focused on fitness, diet / nutrition, meditation and wellness, there are several emerging players and platforms that use data-driven insights to achieve proactive health improvements through preventative lifestyle changes. For example, Ginger.io offers mobile application that integrates information from mobile sensors with other sources of data to assist with behavioral therapy. Other notable examples are Kaiser HealthConnect (for patient engagement), Apple HealthKit and CareKit (for data aggregation and sharing), Omada Health (focused on Diabetes management and prevention), and Welltok (for consumer engagement and lifestyle changes). We briefly provide a case study on Welltok's offerings.

Case study

Welltok⁴⁵ is a startup that has raised more than \$145M and offers the Caféwell Concierge and Insights apps for personalized health solutions for everyday consumers. To offer an effective solution, Welltok has partnered with Watson Health from IBM to provide the "cognitive computing" backend in the cloud. Taken together Welltok has created an engagement and analytics focused Caféwell Health Optimization platform for population health managers and consumers. Through this

⁴³ <http://endeavourpartners.net/assets/Endeavour-Partners-Wearables-and-the-Science-of-Human-Behavior-Change-Part-1-January-20141.pdf>

⁴⁴ <http://endeavourpartners.net/the-future-of-activity-trackers-part-3-the-secret-to-long-term-engagement/>

⁴⁵ <http://welltok.com/>

platform, they offer each consumer with a personalized plan that engages consumers in healthy behaviors through rewards, streamlined communication, and actionable insights. This action plan includes an Intelligent Health itinerary - an action plan with recommendations that get better with use and over time. Welltok's offering is powered in the backend by Watson Health which provides natural language processing to assess user's activity goals and provide prompts for improved user engagement. Ongoing pilots across the State of Colorado co-sponsored by United Healthcare and Kaiser, have delivered promising initial results:

- 50% participation in the program
- 650% increase in health assessment completion
- \$100K awarded for healthy behaviors

It is relatively early in this program to measure overall effectiveness and more results are expected by summer 2016. Successful roll-outs are also underway with Centura Health, Community Health Plan (Washington), Coventry HealthAmerica.

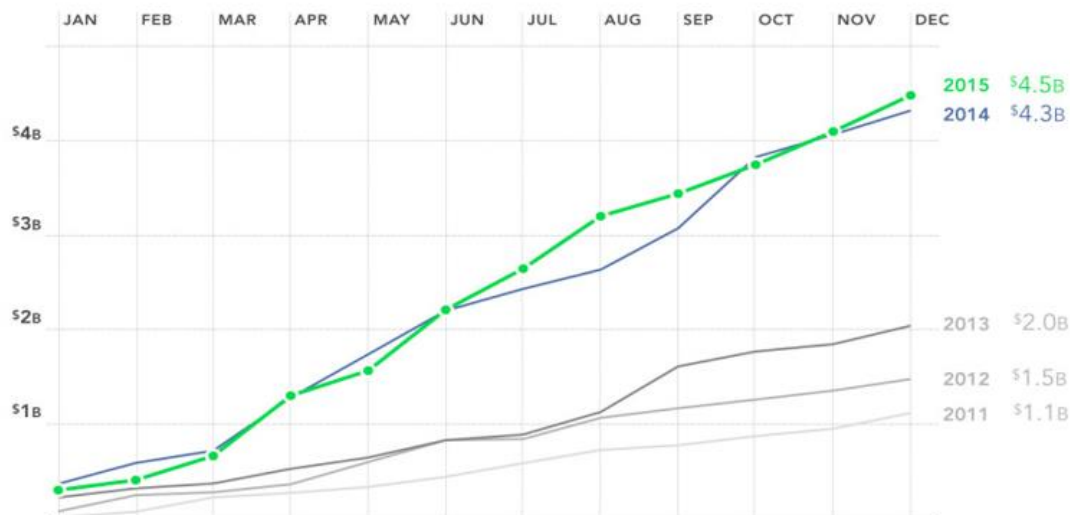
Digital Health Funding

To capitalize on the opportunity for disruption and positive impact in Healthcare in the US, the push for Digital Health has been fueled by huge investments in emerging startups.

Smart Money Is Fueling Digital Health

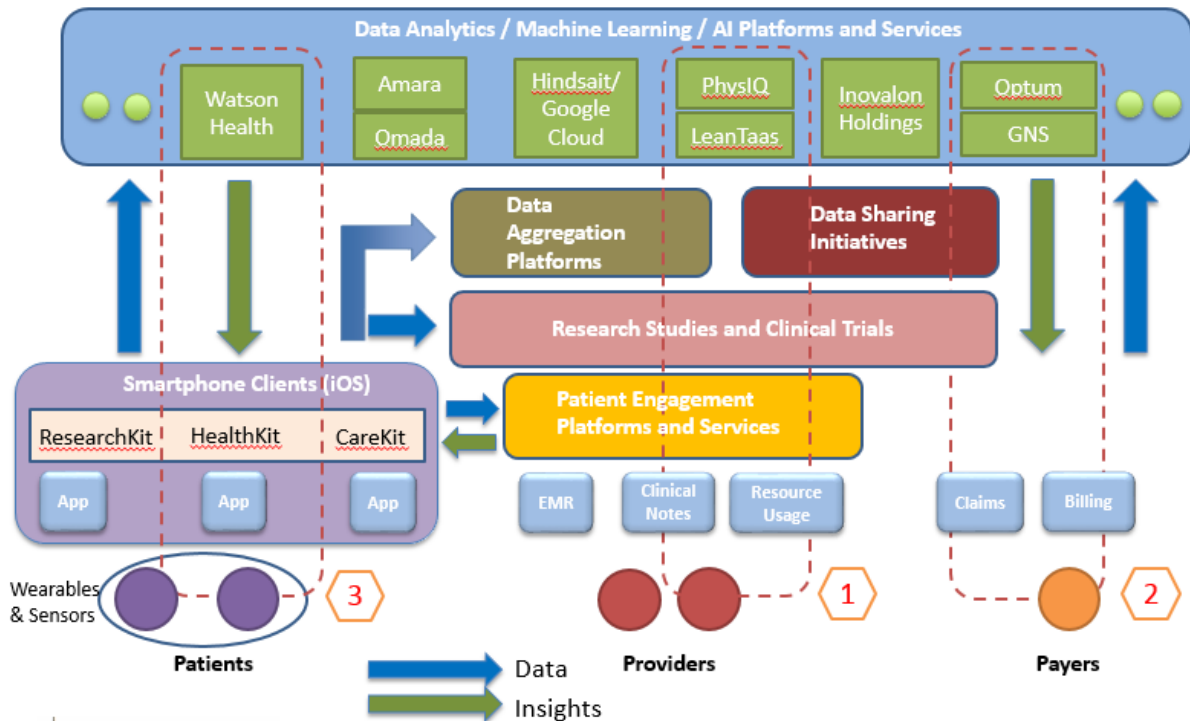
DIGITAL HEALTH FUNDING
2011-2015

ROCK
HEALTH



This is illustrated well in the figure above which shows the amount of funds invested in Digital Health startups in the past 4-5 years. More than \$4.5B were invested in Digital Health by the end of 2015 and the pace of investments continues to be higher so far in 2016.

Emerging Landscape



From our research and analysis of the key use cases and case studies presented, we have assembled the above picture to depict the emerging landscape of technology building blocks, platforms and services in digital health and the use of Big Data. To a large extent the landscape represents the “composable infrastructure” and “end to end solutions architecture” for addressing the 6 use cases we’ve presented earlier. The key constituents of this landscape are summarized next.

Patient-Interfacing Clients and Kits: Patients are increasingly using and relying on a variety of wearables and sensors to track their fitness, activity, sleep, and vital signs. These devices are often used along with many smartphone apps for health and fitness - there are over 40,000 health-and-fitness related apps in the Apple App

Store. A key problem faced by users is centrally tracking data from multiple devices and apps. Apple's HealthKit offers the ability for apps to share relevant health and fitness data - any app can publish data allowed by the user into the Health app, and retrieve data from the Health app, which makes the Health app a central hub for tracking a user's health and fitness data.

Apple also released the ResearchKit⁴⁶, an open source framework designed for medical and health research that allows app developers to launch apps to gather accurate data from volunteers that opt-in to use the app and share data. This allows researchers to rapidly recruit millions of willing volunteers for large-scale studies about asthma, breast cancer, cardiovascular diseases, diabetes, Parkinson's, and more. And very recently (April 2016), Apple released the CareKit which is also an open-source framework designed to enable developers to create "doctor-patient" apps that help people actively manage their own medical conditions and share information with their doctors and medical providers if they choose to do so. Example apps initiated created using CareKit include a Parkinson's disease app to help track the effectiveness of the medicines prescribed, and a post-surgery app to improve recovery after discharge.

Data Analytics / Machine Learning / AI Platforms and Services: This layer constitutes a set of sophisticated, specialized platforms and services in the cloud that provide capabilities in data analytics, machine learning and AI techniques for the vast amounts and different types of healthcare data. Notable examples include Watson Health, PhysIQ, Hindsait (on Google Cloud), Inovalon Holdings, Amara Health, Wanda Health, IMS Health, and many others. For example, PhysIQ is a data analytics platform that extracts intelligence from real-time patient data from wearables and sensors, and uses machine learning technology to analyze continuous vital signs such as heart rate, respiration rate, body motion, etc. As another example, Watson Health uses cognitive computing to sift through medical literature and knowledge base to find patterns that are relevant for different patients a doctor is seeing in order to help make stronger diagnoses and craft better treatment plans.

Data Aggregation Platforms: These emerging platforms aggregate data from many users, devices, sensors, and apps into a common repository that can be queried and tapped into to obtain and process data feeds. A leading example is Validic Health -

⁴⁶ <http://researchkit.org/>

Validic is a leading cloud-based digital health platform that provides convenient and quick access to patient data from in-home clinical devices, wearables, and patient healthcare applications. Validic creates value by aggregating, normalizing, and combining data from a broad array of wearable devices and providing it to the healthcare system in a HIPAA compliant manner. It has a partnership with Philips Healthcare and many other healthcare systems, and integrates data from more than 215 devices.

Data Sharing Initiatives and Alliances: These enable and simplify the sharing and interoperability of healthcare data between different stakeholders. Some of these were mentioned earlier during the use case discussion, such as Open Notes. Additional examples include HealthData APIs, CommonWell Health Alliance, NODE Health (Mt Sinai Hospital), Cal Index (by Orion Health) which is a data exchange built by California insurers.

Research Studies and Clinical Trials: User-generated data gathered from apps enabled by ResearchKit can be readily accessed, processed and analyzed for research studies and clinical trials. A recent example of this is the SleepHealth app created by IBM in partnership with the American Sleep Apnea Association (ASAA) on a study⁴⁷ to determine the connections between sleep habits and general health. Data collected from the app will drive the SleepHealth Mobile Study, which will also draw from insights gleaned by IBM's Watson cognitive computing system as new data is ingested.

Patient Engagement Platforms and Services: These play a key role in connecting the providers directly with the patients to enhance patient engagement at different levels and ensure a better transition from the hospital to a patient's home by integrating remote medical monitoring technology, enabling coordinated care, and offering data-driven insights and recommendations to patients and general consumers. The patient engagement value provided is governed by various factors such as providing educational information, reminding or alerting users, recording and tracking health information, displaying and summarizing health information, providing guidance based on user activity, enabling secure direct communications with providers and family members, providing support through social networks, and supporting behavior change through rewards (such as a points system). In

⁴⁷ <http://www.eweek.com/database/ibm-launches-sleep-study-app-on-watson-health-cloud.html>

various studies and pilots, remote monitoring been shown to have benefits such as lower risks of hospitalization, shorter hospital stays, and lower rates of subsequent medical conditions. As mentioned earlier, such engagement platforms and services can also play a key and effective role in encouraging proactive lifestyle changes to prevent diseases and improve health.

Notable examples include Welltok's CafeWell Concierge health app described earlier in one of the case studies, the online engagement platform designed by Welkin Health (through which providers can communicate with patients and other care team members in real-time using phone, text or email), and the MyCarolinas Tracker app created by Carolinas Healthcare. CareKit from Apple (mentioned earlier) will make it easier for the creation of patient engagement apps powered by patient engagement platforms and services.

Flow of Data and Insights: Referring to the emerging landscape picture, the relevant patient-generated data flows from the end user clients (such as smartphones, wearables, and sensors) to the other building blocks as illustrated by the green arrow. The data may flow to one or more of these: data analytics platforms and services in the cloud, research studies and clinical trials, data aggregation and sharing platforms, and/or patient engagement platforms. Any insights derived - either in the cloud or by doctors and other medical providers - flow back to the patients, providers or payers as illustrated by the green arrow.

Partnerships and Case Study Examples: Also illustrated in the landscape picture are three of the case studies presented earlier:

1. LeanTaas for improving care efficiency and cost
2. Aetna and GNS Healthcare to predict patients at risk for metabolic syndrome
3. Welltok and Watson Health to guide consumer lifestyle changes

These also illustrate the partnership ecosystems that are rapidly developing and evolving to create and deliver end-to-end solutions that leverage and integrate best-of-breed platforms and services in the "cloud backend" while providing unique value to the respective stakeholders in the "client frontend".

Societal Challenges



While sharing of healthcare data brings many benefits, it is imperative that care and caution is exercised when doing so given the serious implications involved⁴⁸. These concerns include⁴⁹:

- Privacy and security - keep data secure, without unauthorized identification
- Ownership - who owns patient data, why, and for how long?
- Relevance - how much data is enough to share and what is relevant?
- Accuracy and Risk - how accurate is the data collected by wearables, sensors, etc.? are there any health risks involved due to wearables?⁵⁰
- Misuse - can patient-generated data be misused by payers and providers?

48

http://www.slate.com/articles/technology/future_tense/2015/02/how_data_from_fitness_trackers_medical_devices_could_affect_health_insurance.html

49 <https://www.pwc.com/us/en/health-industries/top-health-industry-issues/assets/pwc-hri-wearable-devices.pdf>

50 <http://readwrite.com/2016/04/30/are-wearables-really-a-health-risk-hw1/>

As noted in the picture above, the healthcare sector has seen the biggest increase in data theft (of medical records) in the past several years. We briefly cover some highlights on this aspect below.

Electronic health record (EHR) is increasingly being implemented in many developing countries. It is the need of the hour because it improves the quality of health care and is also cost-effective. Technologies can introduce some hazards hence safety of information in the system is a real challenge. Recent news of security breaches have raised concerns that despite increased usefulness and adoption of EHRs, more attention needs to be paid to the ethical issues that arise. Securing EHR with an encrypted password is a probable option. The different ethical issues arising in the use of the EHRs and their possible solutions have been outlined by others⁵¹.

Privacy And Confidentiality

Information of a patient should be released to others only with the patient's permission or allowed by law. When a patient is unable to do so because of age, mental incapacity the decisions about information sharing should be made by the legal representative or legal guardian of the patient. Information shared as a result of clinical interaction is considered confidential and must be protected⁵². Information from which the identity of the patient cannot be ascertained for example, the number of patients with breast carcinoma in a government hospital, is not in this category⁵³.

Security Breaches

Security breaches threaten patient privacy when confidential health information is made available to others without the individual's consent or authorization. Two recent incidents at Howard University Hospital, Washington showed that inadequate data security can affect a large number of people. Security measures such as firewalls, antivirus software, and intrusion detection software must be included to protect data integrity. Specific policies and procedures serve to maintain patient privacy and confidentiality.

⁵¹ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394583/>

⁵² <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394583/#ref11>

⁵³ <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4394583/#ref12>

Winners & Losers

Big data and artificial intelligence revolution is transforming what is appropriate or right for a patient and right for the healthcare ecosystem⁵⁴. Stakeholders in traditional medical management system must adopt new practices to keep up with advances in technology.

We make the following summary observations:

Patients will win but might pay more

- Better, more efficient, and timely care with more choices
- Might pay higher premiums based on fitness and lifestyle data (for poor choices)

Providers – hospitals will win but doctors could see tradeoffs

- Higher efficiencies and lower costs
- Some specialties will benefit, but not all
- Private practices may struggle / lose
- “Intelligent machines” may complement doctors, but also squeeze them more

Payers – insurance companies could win but it depends

- Depends on mix of healthy vs unhealthy population

In the following sections, we offer a broader perspective on the benefits and risks for the key constituents in healthcare - Patients, Providers, Payers, and Manufacturers.

Patients

We believe patients & individuals will be the biggest beneficiary of new value pathways emerging from Big Data. For e.g., individuals will see improvements in;

- **Right Living:** Individuals will be able to make informed lifestyle choices that promote well being and active engagement in their own care.
- **Right care:** Evidence based care that is proven to deliver needed outcomes for each patient while ensuring safety.
- **Right Value:** Sustainable approaches that continuously enhance healthcare value by reducing the cost at the same time improving quality.
- **Right Innovation:** Innovations in R&D, discovery and development will lead to precision diagnosis and treatment.

On the downside, data transparency introduces potential risk for individuals. Some providers can take advantage of data by pursuing objectives that create value for

⁵⁴ <http://www.hhnmag.com/articles/6561-ways-artificial-intelligence-will-transform-health-care>

themselves. For e.g., Insurance providers can increase rates for individuals based on profile and culpability to particular disease. Providers can convincingly market their services, regardless of clinical needs. We see these risks as real possibly unavoidable.

Providers

Providers are at center stage of this sea change. They are not only primary point of care but also one of the primary point of data origination and capture. Advances on big data will put pressure on Healthcare providers to be data-driven, and focus on quality and outcome-based protocols to improve patient care. Providers will have to transform themselves at least in following areas:

- Data collection & sharing: Providers have to improve technical capabilities to ensure consistent and comprehensive data capture. Establish security policies, governance strategies to securely share data within and outside organization.
- Value based care: Providers will compete to move from traditional medical-management techniques to value based care. Invest in analytic capabilities that are predictive than retrospective.
- Payment structure: Traditional fee-for-service payment structure replaced with new system based on outcome-based reimbursement from insights.

Manufactures

Manufactures will be under continued pressure to define value of their product. Key priorities for manufacturers;

- Relationship: Existing relationship structure between manufacturer and Provider has to change to focus attention on actual payer and customer value.
- Design and development: Incorporate big data to to establish clear view of efficacy. Partner to make breakthrough scientific discoveries.

New HealthTech Players

Challenges in health care and advances in Big Data is inspiring new Tech companies to develop healthcare applications or similar innovations. We've mentioned and noted several new VC funded as well as non-profit health tech innovators.

In summary, we believe that the net benefit of Big Data approaches in Healthcare in the longer term is clearly positive and will provide great societal benefit. Some stakeholders in the current ecosystem benefit from the inefficiencies that lack of data provides and they may lose business as more information is rapidly analyzed,

correlated, and made publicly available. Healthcare Providers & Manufacturers will see a major impact on their business practices.

Conclusion

Healthcare is in the midst of massive disruption through waves of digital health innovation, advanced data analytics and machine learning techniques, widespread adoption of cloud computing (cheap storage and computation), and a growing degree of health consciousness amongst consumers. Given that Healthcare is the largest contributor to the US GDP, the impact of the coming changes in US Healthcare will be realized and felt for decades to come, in very profound structural and substantial ways.

In this report we have explored how Big Data and the associated technologies, platforms, services and stakeholders play a very key role in addressing a broad set of important use cases in Healthcare. Taken together, these use cases represent the bulk of the challenges and opportunities in improving healthcare to reduce costs substantially and deliver significantly better health outcomes. With the help of the latest advances in big data processing, machine learning / AI techniques, emerging platforms and services, the coming years promise to fundamentally re-architect and re-structure US healthcare and make the US compare much more favorably with other countries.

We've tried to provide a comprehensive picture and perspective in this vast and complex subject, but we also acknowledge that there is a lot more happening in Digital Health than what this report is able to cover. We end on a lighter note by showing the cartoon below which aptly depicts the (healthy) future that beckons us.

