What is BIG DATA ?

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy.

Often highcosts, lost opportunity, poor decision making, governance and customer service are the pitfalls of poor data capture and management.

But datacapture and management is not so simple and it involves several overlapping fields to get a data insights and make real difference.

Why do we need Big Data in Health ?

Paradigm Shift

Standard medical practice is moving from relatively ad-hoc and subjective decision making to evidence-based healthcare.

Economic Advantage

It is widely believed that the use of information technology can reduce the cost of healthcare while improving its quality by making care more preventive and personalized and basing it on more extensive (home-based) continuous monitoring. McKinsey estimates that a savings of 300 billion dollars every year in the US alone.

Also recently Accenture reported that cybersecurity threats in Health Data are very real. To prevent DataBreaches, Big Data is needed to plug the loop holes.

Indirect Social Benefits

It will improve social mood and less physical violence, more innovations, better health policy formulations in general

Big Data workflow

Health Data capture through WEARABLE DEVICES

SENSORS
TODAY

Clip slide

Jawbon
e

iHealth wireless
pulse oximeter

Withings smart body
analyser

iBGStar

SENSORS
TODAY

Clip sli

Jawbon
e

iHealth wireless
pulse oximeter

Withings smart body
an

iBGStar

EATING HABITS

HEART RATE

SLEEP

CALORIES

BLOOD SUGAR

BODY TEMPERATURE

SENSORS
IN FUTURE

Clip slide

DETECTION

ANALYSIS

DIAGNOSTICS

CELL CULTURE

DRUG DELIVERY

THERAPEUTICS

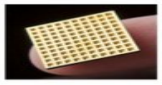
SENSORS
IN FUTURE

Clip slide

Continuous MicroCHIPS
Glucose Monitoring

Sensor-Laden
Transdermal patch

Google lens



Parathyroid
hormone
microchip injection

MIT batteryless power
source

What can BIG DATA DO IN HEALTHCARE ?

1. Preventing Organ Failure : By early diagnosis and intervention
2. Cost-Effective Treatment Of Chronic Disease : By Doing comparative analysis cross different geographies , alternative cures ..cost effective holistic health care approach can be practiced rather approaching disease at critical stages when the costs are alarming.
3. Reducing Drug Reactions :Transparency in medical records and research will mitigate drug overdosage reactions.
4. Avoiding Readmissions : Admission only when the patient is really critical will help doctors concentrate better and cut medical costs and insurance costs.

By aggregating and analyzing health data from disparate sources, such as clinical, financial and administrative data, the outcome of treatments in relation to the resource utilization can be monitored.

What sources are holding HEALTH BIG Data ?

Clinical data is a staple resource for most health and medical research. Clinical data is either collected during the course of ongoing patient care or as part of a formal clinical trial program.

Clinical data falls into six major types

- Electronic health records -- Administrative data -- Claims data
- Disease registries -- Health survey -- Clinical trials data

Electronic health records

An electronic health record (EHR) is a digital version of a patient's paper chart. EHRs are real-time, patient-centered records that make information available instantly and securely to authorized users.

- Contains a patient's medical history, diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results

- Allow access to evidence-based tools that providers can use to make decisions about a patient's care

Electronic medical records (EMRs)

They are a digital version of the paper charts in the clinician's office. An EMR contains the medical and treatment history of the patients in one practice. EMRs have advantages over paper records. For example, EMRs allow clinicians to:

- Track data over time
- Easily identify which patients are due for preventive screenings or checkups
- Check how their patients are doing on certain parameters—such as blood pressure readings or vaccinations
- Monitor and improve overall quality of care within the practice
- But the information in EMRs doesn't travel easily *out* of the practice. In fact, the patient's record might even have to be printed out and delivered by mail to specialists and other members of the care team. In that regard, EMRs are not much better than a paper record.
- With fully functional EHRs, all members of the team have ready access to the latest information allowing for more coordinated, patient-centered care. With EHRs:
- The information gathered by the primary care provider tells the emergency department clinician about the patient's life threatening allergy, so that care can be adjusted appropriately, even if the patient is unconscious.
- A patient can log on to his own record and see the trend of the lab results over the last year, which can help motivate him to take his medications and keep up with the lifestyle changes that have improved the numbers.
- The lab results run last week are already in the record to tell the specialist what she needs to know without running duplicate tests.
- The clinician's notes from the patient's hospital stay can help inform the discharge instructions and follow-up care and enable the patient to move from one care setting to another more smoothly.

Administrative data

These are primarily hospital discharge data reported to a government agency like AHRQ.

Claims data

- Claims data describe the billable interactions (insurance claims) between insured patients and the healthcare delivery system. Claims data falls into four general categories: inpatient, outpatient, pharmacy, and enrollment. The sources of claims data can be obtained from the government (e.g., Medicare) and/or commercial health firms (e.g., United HealthCare).
- Basic Stand Alone (BSA) Medicare Claims Public Use Files (PUFs)
- Medicaid Statistical Information System
- PharMetrics Integrated Database (IMS Rx Benchmark)

Disease registries

They are clinical information systems that track a narrow range of key data for certain chronic conditions. The most common conditions captured include cancer, diabetes, heart disease, and asthma. Registries often provide critical information for managing patient conditions.

- Global Alzheimer's Association Interactive Network (GAAIN)
- National Cardiovascular Data Registry (NCDR)
- National Program of Cancer Registries
- National Trauma Data Bank

Health surveys

- National Health & Nutrition Examination Survey (NHANES)
- National Long Term Care Survey National Medical Expenditure Survey
- National Center for Health
- CMS Data
- National Health and Aging Trends Study (NHATS)

CLINICAL Trial Data

- Current Controlled Trials
- ClinicalTrials.gov
- European Union Clinical Trials Database
- IFPMA Clinical Trials Portal (Pharmaceutical Manufacturers & Associations)

- WHO International Clinical Trials Registry Platform
- Immune Tolerance Network TrialShare
- Drug Delivery Clinical Trials Database

WITHOUT BIG DATA WE ARE DOOMED SOONER OR LATER

Medicare Penalties: Medicare penalizes hospitals that have high rates of readmissions among patients with Heart failure, Heart attack, Pneumonia.

Hospitalizations account for more than 30% of the 2 trillion annual cost of healthcare in the United States. Around 20% of all hospital admissions occur within 30 days of a previous discharge. – not only expensive but are also potentially harmful, and most importantly, they are often preventable.

BIG DATA IS REAL AND ENDORSED BY GOVT AND INDUSTRY

BRAIN Initiative: Find new ways to treat, cure, and even prevent brain disorders, such as Alzheimer's disease, epilepsy, and traumatic brain injury. A new bold $100 million research initiative designed to revolutionize our understanding of the human brain.

- The US President unveiled a new bold $100 million research initiative designed to revolutionize our understanding of the human brain. BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative.
- Find new ways to treat, cure, and even prevent brain disorders, such as Alzheimer's disease, epilepsy, and traumatic brain injury.

- *"Every dollar we invested to map the human genome returned $140 to our economy… Today, our scientists are mapping the human brain to unlock the answers to Alzheimer's."*

-- President Barack Obama, 2013 State of the Union.

Industry Initiatives

Heritage Health Prize: Develop algorithms to predict the number of days a patient will spend in a hospital in the next year. http://www.heritagehealthprize.com

GE Head Health Challenge: Methods for Diagnosis and Prognosis of Mild Traumatic Brain Injuries. Develop Algorithms and Analytical Tools, and Biomarkers and other technologies. A total of $60M in awards.

Big Data Opportunities

- Analysis of data sets can find new correlations, to "spot business trends, prevent diseases, combat crime and so on simulations,[and biological and environmental research.

Big Data Challenges

- Data sets grow in size in part because they are increasingly being gathered by cheap and numerous information-sensing mobile devices aerial remote sensing software logs, cameras, microphones, radio-frequency identification(RFID) readers, and wireless sensor networks

- Digital curation : It is the selection, preservation, maintenance, collection and archiving of digital assets. Digital curation establishes, maintains and adds value to repositories of digital data for present and future use. This is often accomplished by archivists, librarians, scientists, historians, and scholars.

- Overall IT usage : Big data success relies on healthcare IT usage statistics in health care. If IT dependency is less then it is even more challenging to implement Big Data. A study showed that on average just 55% of healthcare providers in Germany use healthcare IT within primary care settings and 60% in secondary care settings .This would be considerably lower for underdeveloped countries.

- Data heterogeneity (reports, lab reports, images, sensor data, etc.) and Unstructured formats

- Manual data without semantic annotations is unstructured and cannot be processed automatically.

- Data Sharing : As of today, a lot of health data is stored in data silos. A seamless exchange and aggregation of the data often relies on individualized solutions due to the lack of standards and flexible interfaces as well as the heterogeneous nature of the data.

- Although it is feasible from a technical point of view to exchange health data by e.g. using HL76 CDA (Health Level 7 – Clinical Document Architecture), as of today health data is hardly shared across organizations due to non-technical reasons.

- Data Privacy and Security : Big Data security and privacy challenges, which include secure computations in distributed programming frameworks, secure data storage and transaction logs, real-time security and compliance monitoring, scalable and composable privacy preserving data managing and analysis approaches and granular access control and audits.

Algorithms used in Health care

- Genetic algorithm, Support vector machines, Decision tree, Neural network and Cluster Analysis, to disclose the hidden patterns inside the large data set.

- Genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a metaheuristic is routinely used to generate useful solutions to optimization and search problems.[Genetic algorithms belong to the larger class of evolutionary algorithms (EA), which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover.

- For Genomic studies.. By wiki definition, genome is the genetic material of an organism

  Gene is hereditary unit transferred from parent to offspring.

- Genetic Algorithms are explored in medical applications to characterize patterns and results. For example, optimizing image analysis such as, assessing classes of cells in blood cell microscope images or for facilitating magnetic resonance tomography (MRT) treatment planning and 3D visualization of image data. Genetic algorithms can be used for optimizing pharmaceutical products. Recently, it was shown that Genetic Algorithms were able to identify additional anti-bacterial peptides with a high activity during a study.

- Finally, it was shown that Genetic Algorithms enhance the precision of artificial neural networks (ANNs) such as for hip-bone fracture prediction or for optimizing efficient search strategies of ANNs to predict and discriminate pneumonia within a training group.

- Genetic Algorithms are very useful for preparation of Antibiotics

Predictive analytics extracts information from preexisting data sets to determine future patterns and trends. Provided specific patient data is located in a single data database, algorithms can create personalized health experiences. Following the evolution of new data and technology, healthcare analytics encourages patients to pursue active lifestyles, select highly rated providers, and follow data driven results rather than clinical judgment.

HARDWARE/TECHNOLOGY FOR big data

- Enterprise-grade servers that are well suited for modern big data analytics workloads have:
- Higher compute intensity (high ratio of operations to I/O)
- Increased parallel processing capabilities
- Advanced virtualization capabilities
- Modular systems design
- Elastic scaling capacity
- Enhancements for security and compliance and hardware-assisted encryption
- Increased memory and processor utilization

Important Tier 1 vendors such as Cisco, Dell, HP, and IBM.

Big Data storing and Processing Technology

- Apache Hadoop is an open-source software framework written in Java for distributed storage and distributed processing of very large data sets on computer clusters built from commodity hardware.
- HDFS is like the bucket of the Hadoop system: You dump in your data and it sits there all nice and cozy until you want to do something with it. Hadoop is more of a data warehousing system - so it needs a system like MapReduce to actually process the data and it also doesn't use SQL like queries.
- On a Hadoop cluster, the data within HDFS and the MapReduce system are housed on every machine in the cluster. This has two benefits: it adds redundancy to the system in case one machine in the cluster goes down, and it brings the data processing software into the same machines where data is stored, which speeds information retrieval.

Conclusion

Each year, lots of money is wasted is wasted in duplicate records. On an average, each hospital has been spending around1.5 million USD per year, Fortunately duplicate records can now be effectively removed through innovative electronic medical record (EMR) systems. Similarly it has been shown that around 20% of radiology tests are duplicates and waste approximately $20 billion per year. By using appropriate technology, this waste can be prevented.

Big data eliminate costs of duplicate records, predictive analytics can more effectively identify early treatments in a patient's history, prevent future medical episodes and avoid future readmission. Currently preventable diseases such as smoking and obesity cost $2 trillion worldwide. By integrating big data, providers can incorporate family history and current conditions for more effective preemptive care. As a result, it is reported this could reduce total healthcare costs by over $38 billion.