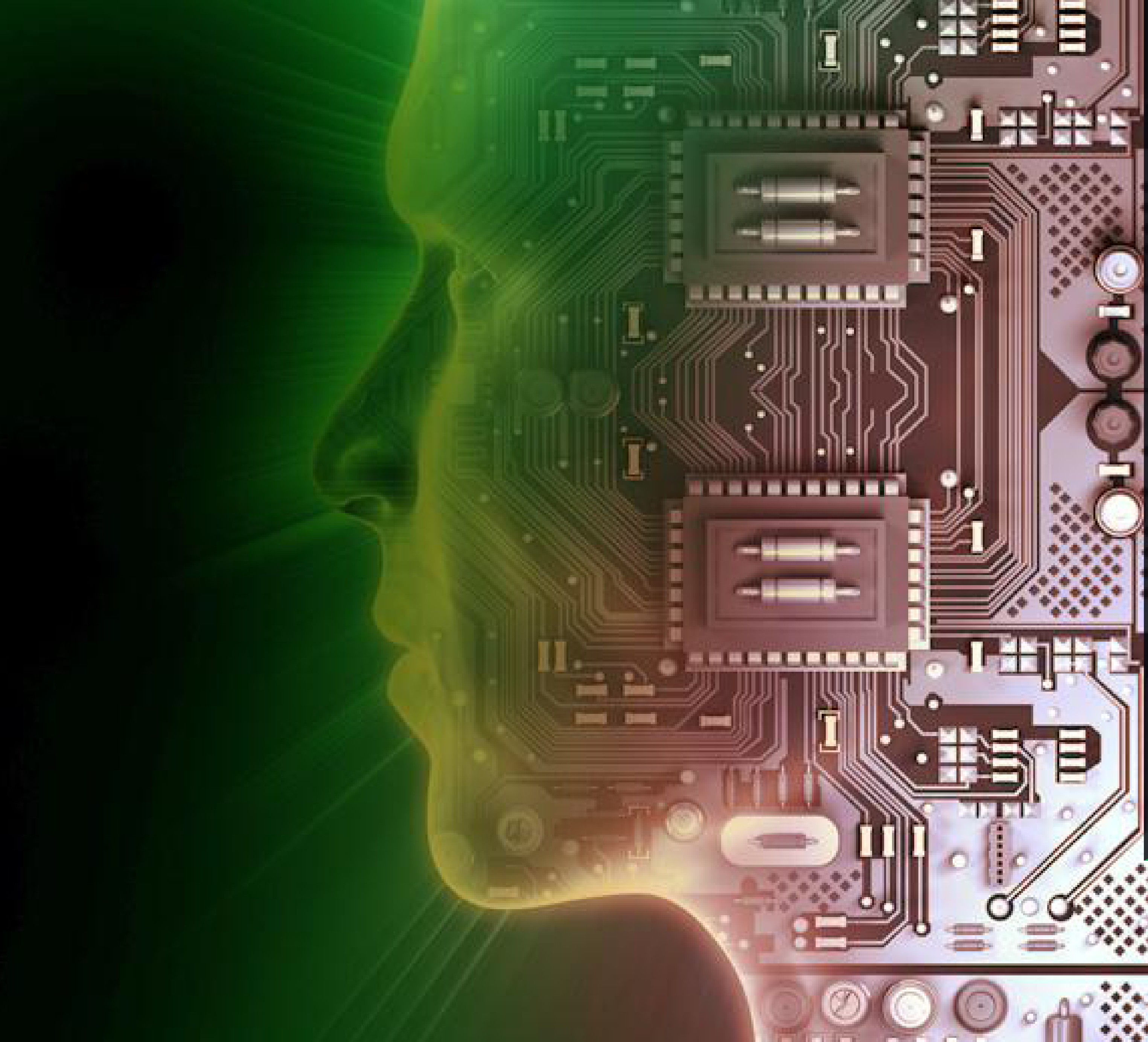


# Big Data in IoT for Healthcare

Pepgra Drives Innovation

**Dr.Radhika ganesan**, ceo pepgra healthcare private limited

http://www.pepgra.com/

# Outline

Value Based Healthcare System – How it is seen today

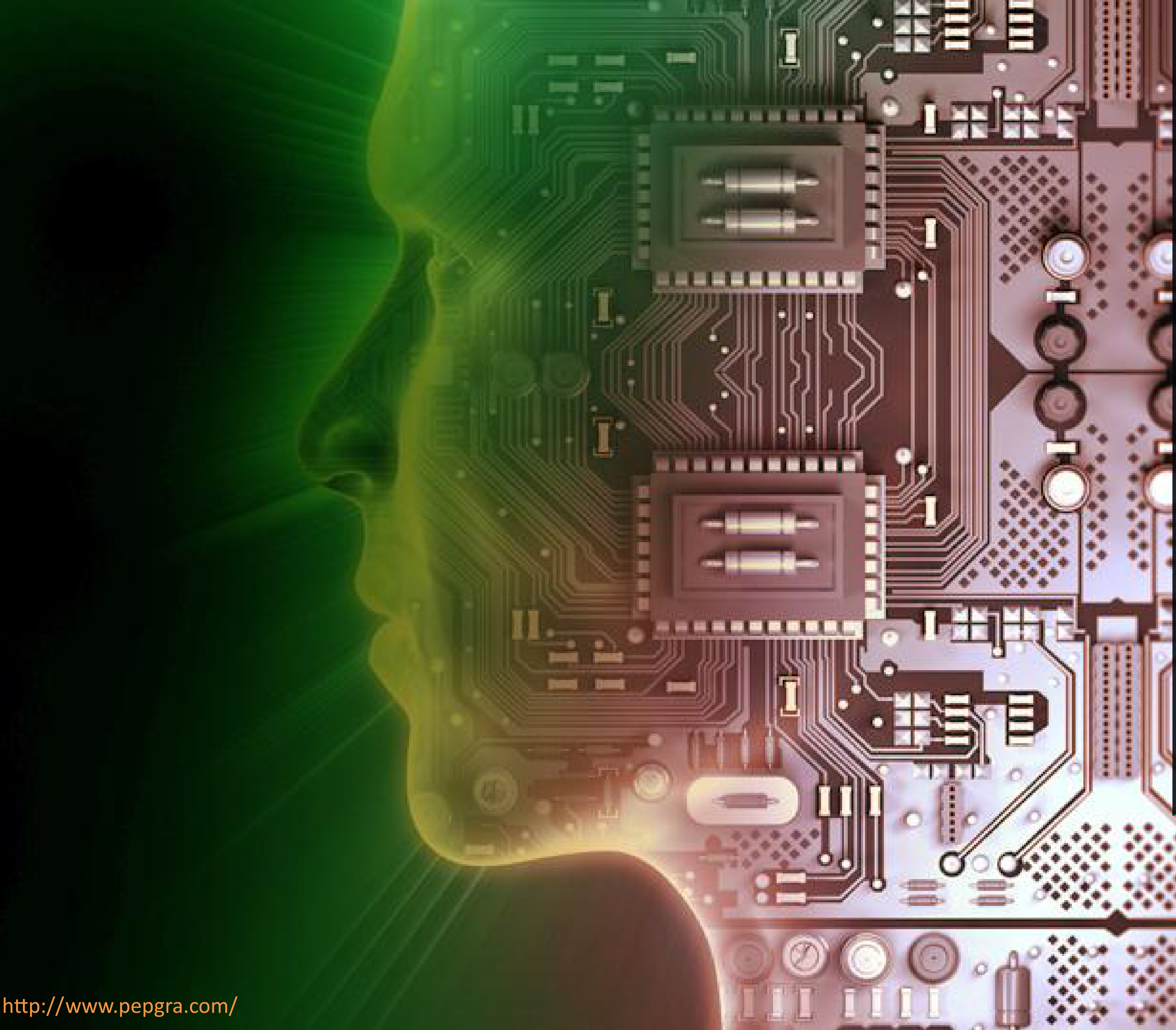
Healthcare Challenge & IoT as a Solution

IoT – Big Data Structure

Recent Trends in IoT Big Data Analytics

Challenges & Our Future


TODAY
What we consume
http://www.pepgra.com/



# HEALTHCARE TODAY

**Where are we**

# Where Are We? – India

India has 2,00,000 centenarians (100+ population)

Percentage of 60+ population expected to increase from: **7.6%** (77 million) in 2000

**20.6 %** (324 million ) in 2050

Over 9 core elderly population in india in 2011- only 12 other countries have a total population higher than that.

In the same period, percentage of the 80+ population will increase from **0.61%** (6.1 million) in 2000

**3.06%** (48.2 million in 2050)

**48.2%** of elderly are women, **58%** of them being widowed, divorced. **75%** of India's elderly live in rural areas and one-third live below the poverty line.

- **5.5 crore** go to sleep on an empty stomach every night—just about the population of the UK.
- An estimated **50 lakh** live alone – more than all of Australia.
- In 2040, within **30 years**, the grey population in India will double again.

Source: Adopted from Insights (2017)

**AGEING IN INDIA**

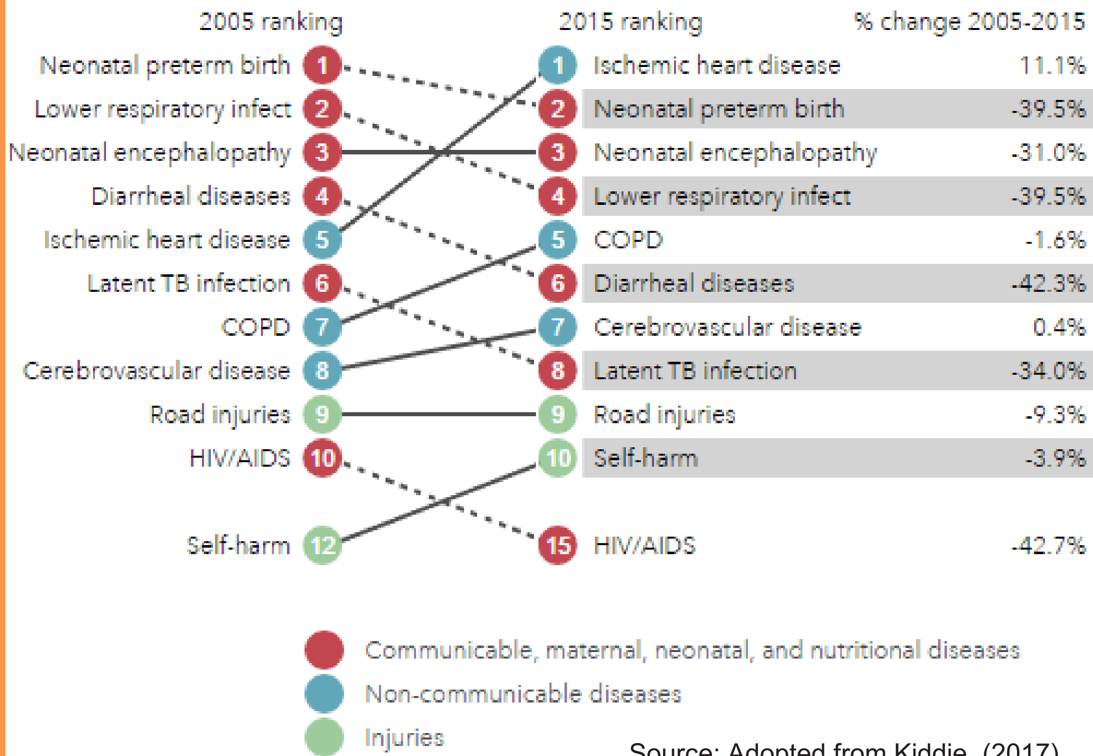
Neglect of Rural population

Import western models and less emphasis on cultural model

Shortage of Medical Personnel

Expensive Health Service (Allopathy Vs, Ayurveda, Unani & Homeopathy)

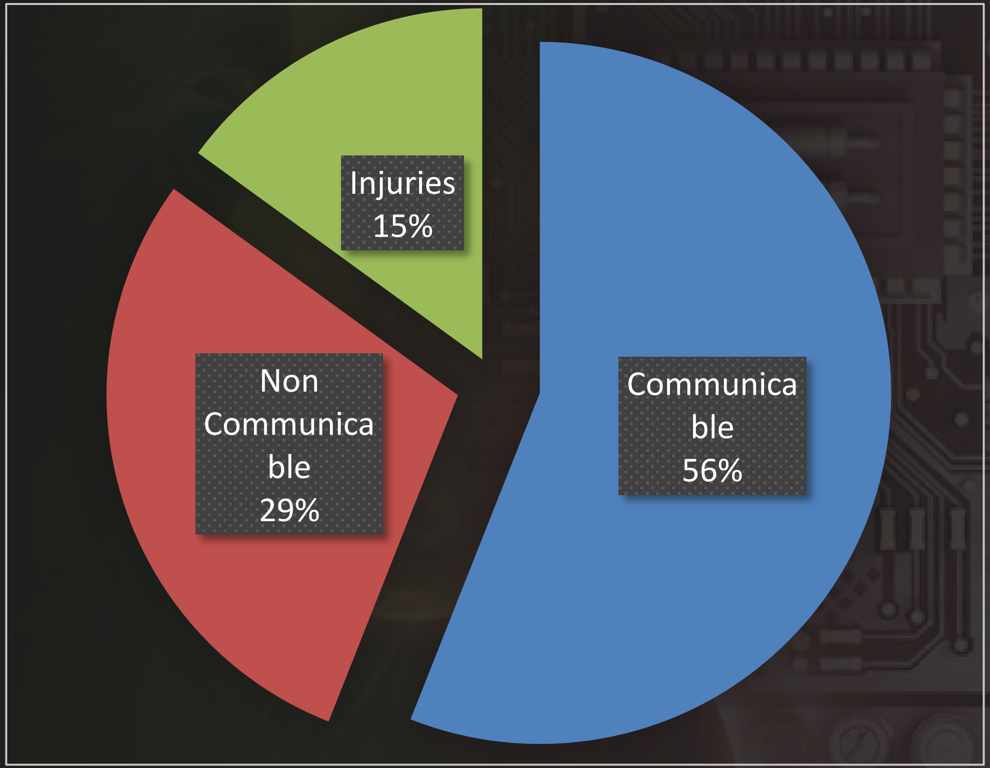
## What Causes The Most Premature Death?



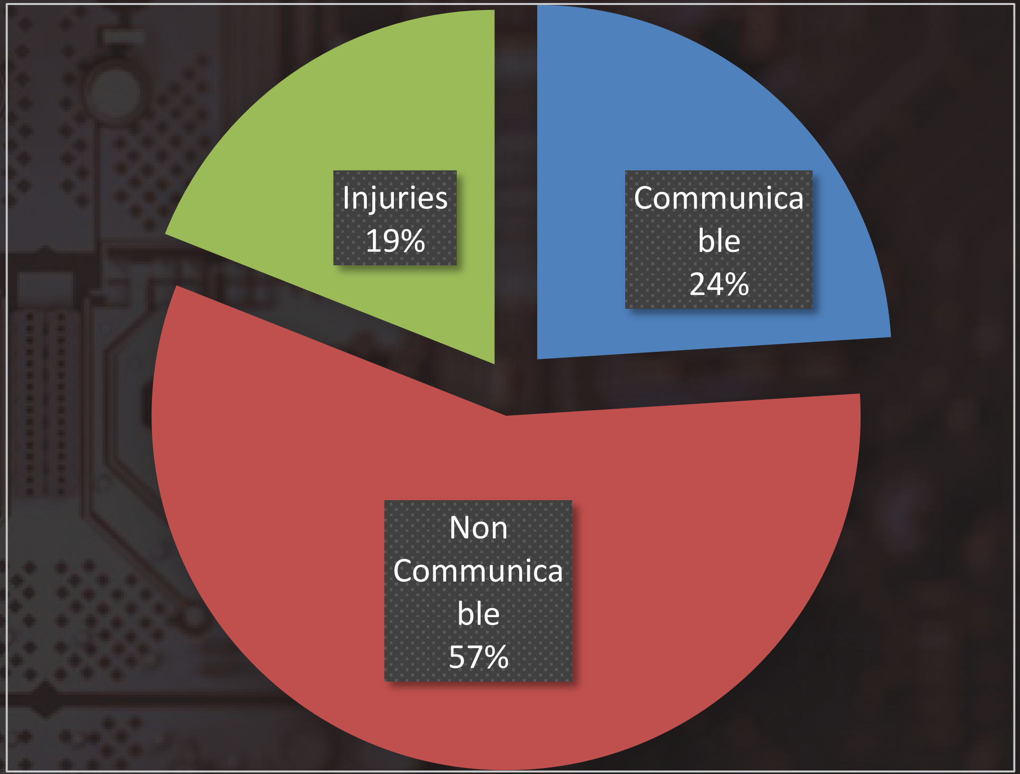

Source: Adopted from Kiddie (2017)

# Distribution of Disease burden – 1990 vs 2020

1990

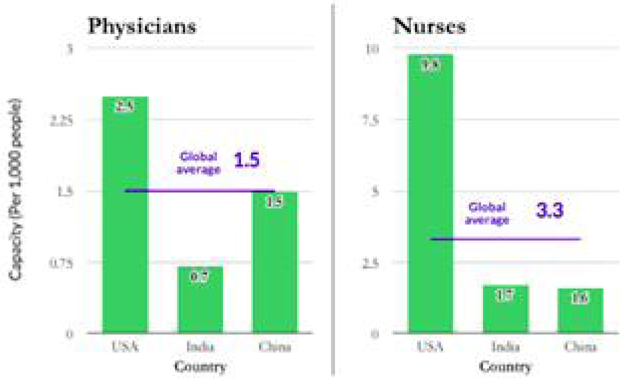



2020



# Operational Challenges

## Capacity And Infrastructure Shortfall




## Infrastructure Shortfall
( India Only )

| | Primary health centers 78% | Community health centers 68% | Sub-Centers 80% |
|---|---|---|---|
| Functional | 25,308 | 5,396 | 153,655 |
| Target | 29,337 | 7,322 | 179,240 |

## Healthcare Spend (2014)

| | USA | UK | INDIA | CHINA |
|---|---|---|---|---|
| As % of GDP | 17.1 | 9.1 | 4.7 | 5.5 |
| Public Spend (as % of GDP) | 8.3 | 7.6 | 1.4 | 3.1 |
| Per Capita (in USD) | $9,403 | $3,935 | $75 | $420 |
| As % of Gov't Spend | 21.3 | 16.5 | 5 | 10.4 |
| Out of Pocket (%) (2004) | 13.4 | 10 | 67.9 | 53.6 |
| Out of Pocket (%) (2014) | 11.1 | 9.7 | 62.4 | 32 |

# 5.2 million medical errors are happening in India annually: Dr Girdhar J. Gyani

TOI

## Medical errors in top 10 killers: WHO

Malathy Iyer | TNN

**Mumbai:** Medicine heals, but this fact doesn't hold true for every 300th patient admitted to hospital. Call it the law of averages or blame human error for it, but the World Health Organization believes that one in 10 hospital admissions leads to an adverse event and one in 300 admissions in death.

An adverse event could range from the patient having to spend an extra day in hospital or missing a dose of medicine, said Dr Nikhil Datar, a gynaecologist and health activist. Unintended medical errors are a big threat to patient safety.

Although there is no Indian data available on this topic, WHO lists it among the top 10 killers in the world. While a British National Health System survey in 2009 reported that 15% of its patients were misdiagnosed, an American study published in the Journal of the American Medical Association in 2000 quantified this problem most effectively. It said that there are 2,000 deaths every year from unnecessary surgery; 7,000 deaths from medication errors in hospitals; 20,000 from other errors in hospitals; 80,000 from infections in hospitals; and 106,000 deaths every year from non-error, adverse effects of medications. In all, 225,000 deaths occur per year in the US due to unintentional medical errors.

It is to create awareness both among doctors and patients about errors dubbed as unintended medical errors that Datar organized a seminar to discuss patient safety at the Indian Medical Associations office on Sunday.

"In the western nations, it is believed that the incidence of unintentional medical errors is between 10% and 17% of all cases," said Datar.

The Indian government has woken up to the concept. It set up the National Initiative on Patient Safety in AIIMS a couple of years back. But the idea, as Dr Akhil Sangal of the Indian Confederation for Healthcare Accreditation, points out is not to apportion blame. "When medical negligence occurs, the first question to be asked is who is to blame. We instead have to evolve to a system in which we ask questions about how, when and where the negligence occurred," he said.

# + HOW ARE WE

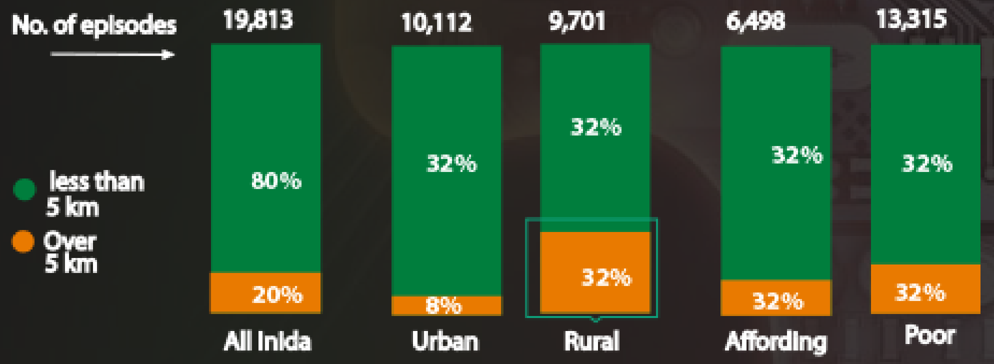
http://www.pepgra.com/

# Challenges in Healthcare

Long Waiting Time

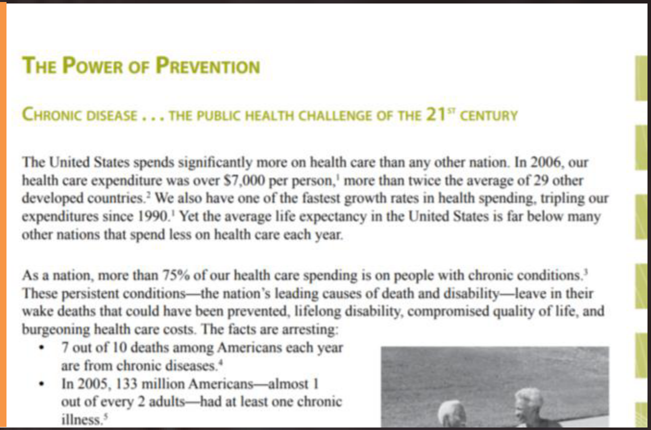
Distance Travelled to OPD

## Distance travelled to seek OPD treatment




Missed Medication

Prevent Chronic disease



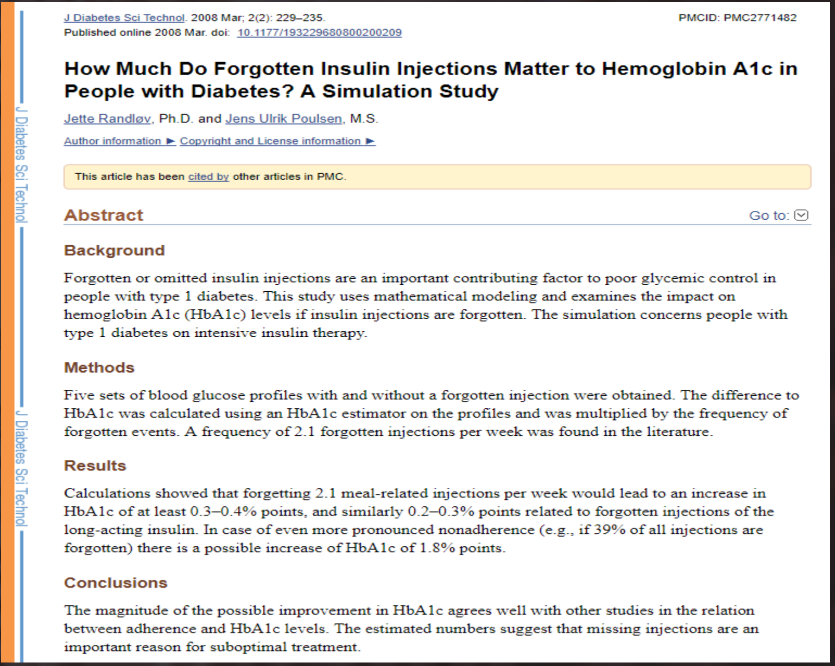
THE POWER OF PREVENTION

CHRONIC DISEASE . . . THE PUBLIC HEALTH CHALLENGE OF THE 21ST CENTURY

The United States spends significantly more on health care than any other nation. In 2006, our health care expenditure was over $7,000 per person,[1] more than twice the average of 29 other developed countries.[2] We also have one of the fastest growth rates in health spending, tripling our expenditures since 1990.[1] Yet the average life expectancy in the United States is far below many other nations that spend less on health care each year.

As a nation, more than 75% of our health care spending is on people with chronic conditions.[3] These persistent conditions—the nation's leading causes of death and disability—leave in their wake deaths that could have been prevented, lifelong disability, compromised quality of life, and burgeoning health care costs. The facts are arresting:

- 7 out of 10 deaths among Americans each year are from chronic diseases.[4]
- In 2005, 133 million Americans—almost 1 out of every 2 adults—had at least one chronic illness.[5]



J Diabetes Sci Technol. 2008 Mar; 2(2): 229–235.
Published online 2008 Mar. doi: 10.1177/193229680800200209
PMCID: PMC2771482

How Much Do Forgotten Insulin Injections Matter to Hemoglobin A1c in People with Diabetes? A Simulation Study

Jette Randløv, Ph.D. and Jens Ulrik Poulsen, M.S.

Author information ► Copyright and License information ►

This article has been cited by other articles in PMC.

Abstract

Go to:

Background

Forgotten or omitted insulin injections are an important contributing factor to poor glycemic control in people with type 1 diabetes. This study uses mathematical modeling and examines the impact on hemoglobin A1c (HbA1c) levels if insulin injections are forgotten. The simulation concerns people with type 1 diabetes on intensive insulin therapy.

Methods

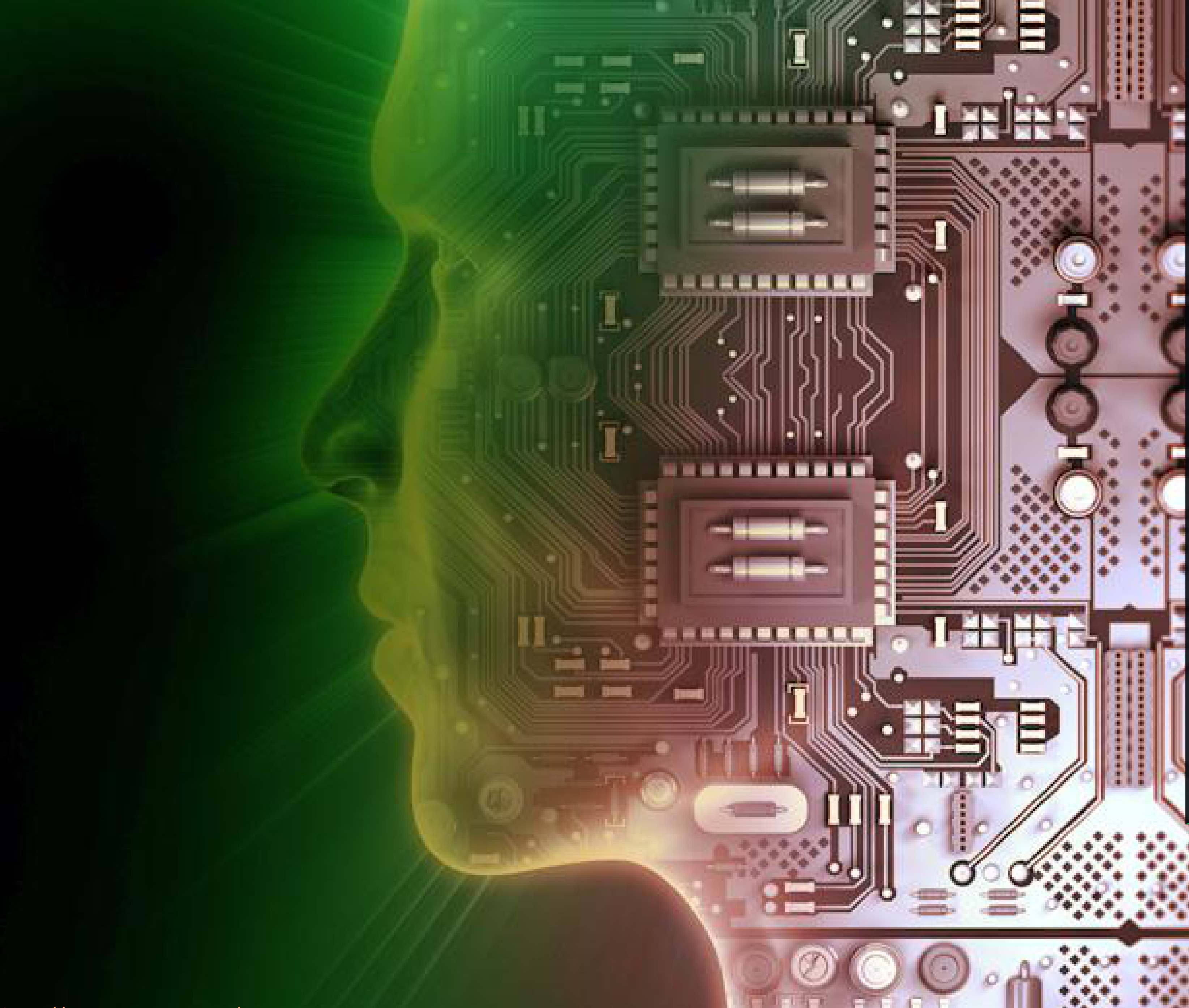
Five sets of blood glucose profiles with and without a forgotten injection were obtained. The difference to HbA1c was calculated using an HbA1c estimator on the profiles and was multiplied by the frequency of forgotten events. A frequency of 2.1 forgotten injections per week was found in the literature.

Results

Calculations showed that forgetting 2.1 meal-related injections per week would lead to an increase in HbA1c of at least 0.3–0.4% points, and similarly 0.2–0.3% points related to forgotten injections of the long-acting insulin. In case of even more pronounced nonadherence (e.g., if 39% of all injections are forgotten) there is a possible increase of HbA1c of 1.8% points.

Conclusions

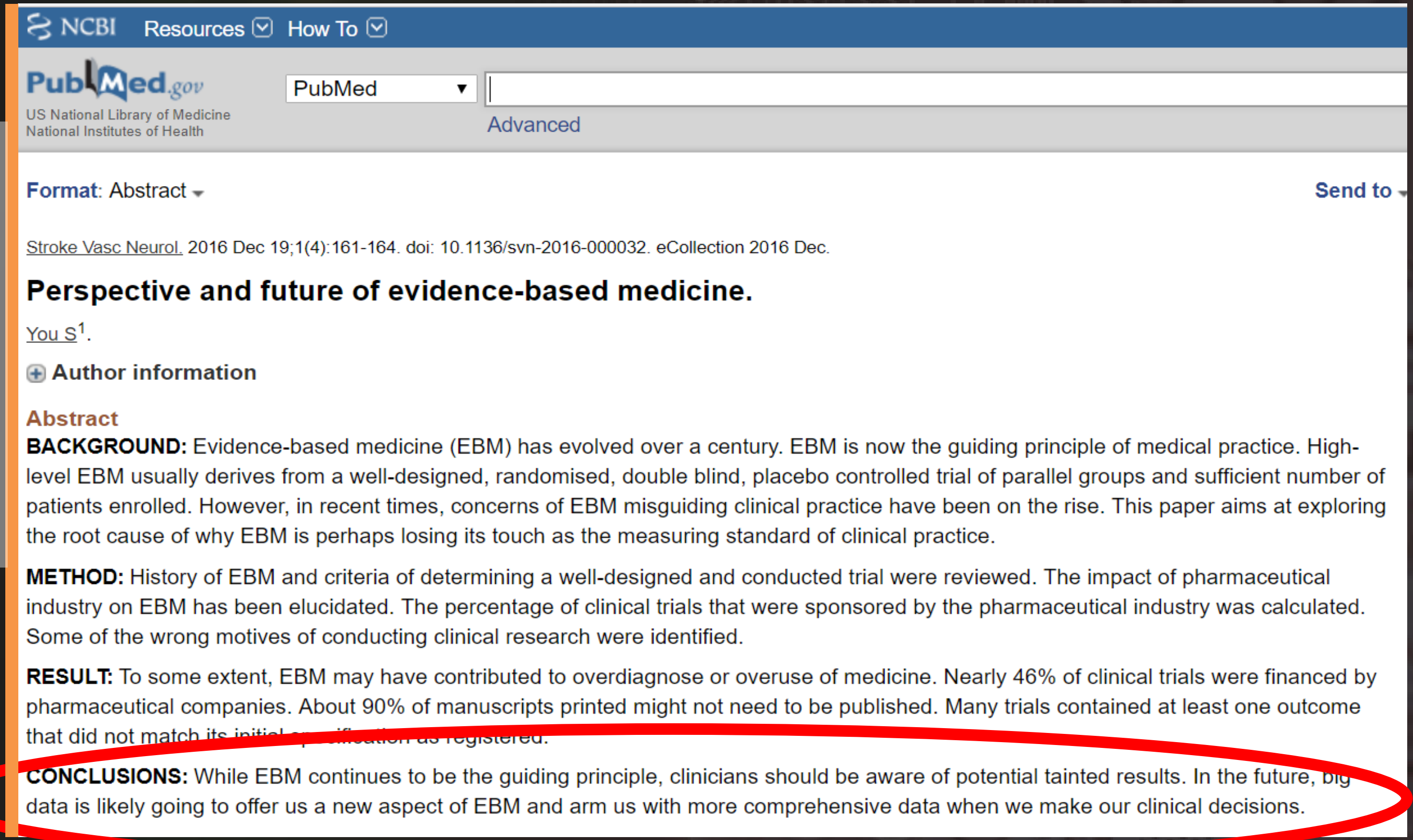
The magnitude of the possible improvement in HbA1c agrees well with other studies in the relation between adherence and HbA1c levels. The estimated numbers suggest that missing injections are an important reason for suboptimal treatment.



# + HEALTHCARE TOMORROW

# Evidence Based Medicine

Source: Adopted from You (2016)



NCBI Resources How To

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed

Advanced

**Format**: Abstract

Send to

Stroke Vasc Neurol. 2016 Dec 19;1(4):161-164. doi: 10.1136/svn-2016-000032. eCollection 2016 Dec.

**Perspective and future of evidence-based medicine.**

You S[1].

**Author information**

**Abstract**

**BACKGROUND:** Evidence-based medicine (EBM) has evolved over a century. EBM is now the guiding principle of medical practice. High-level EBM usually derives from a well-designed, randomised, double blind, placebo controlled trial of parallel groups and sufficient number of patients enrolled. However, in recent times, concerns of EBM misguiding clinical practice have been on the rise. This paper aims at exploring the root cause of why EBM is perhaps losing its touch as the measuring standard of clinical practice.

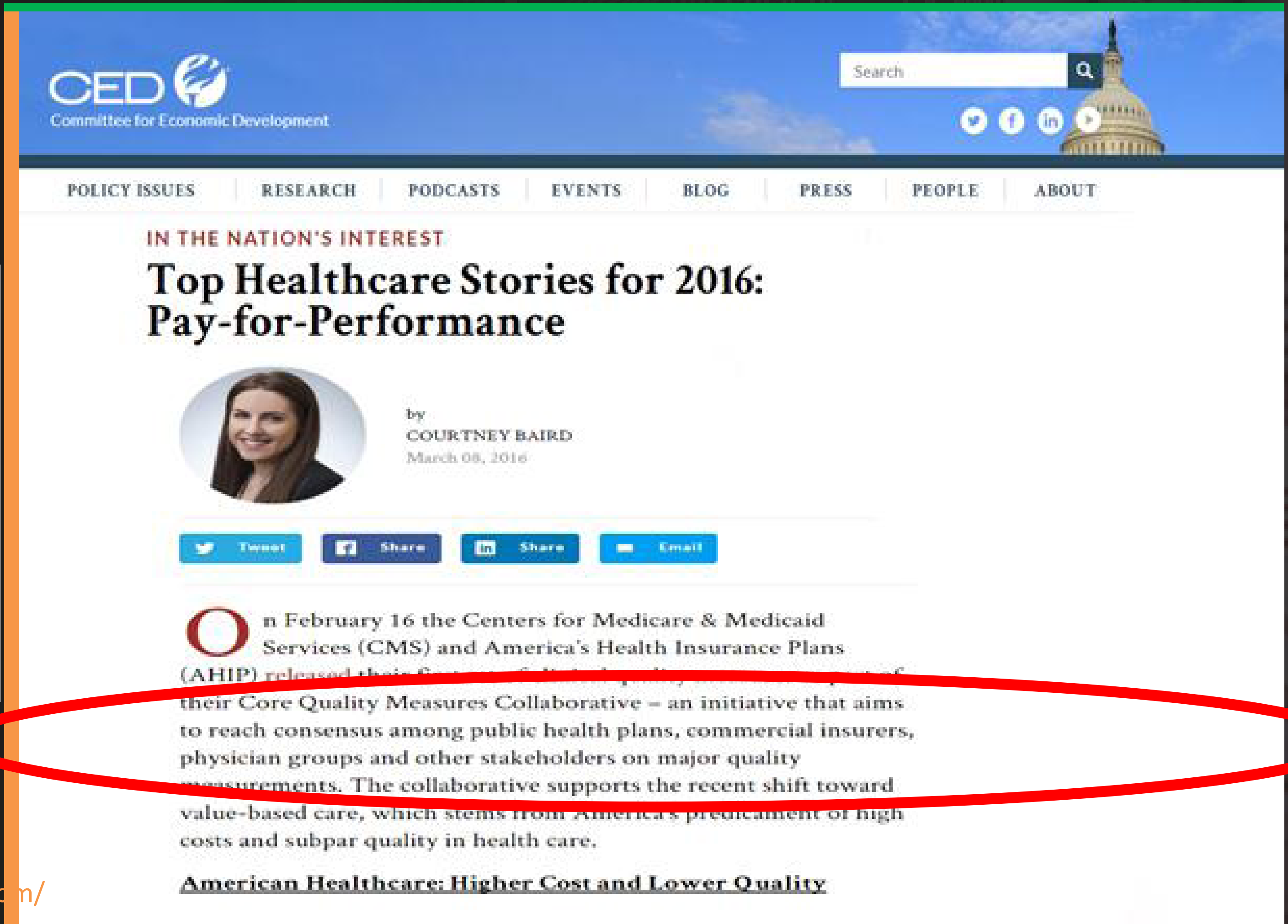
**METHOD:** History of EBM and criteria of determining a well-designed and conducted trial were reviewed. The impact of pharmaceutical industry on EBM has been elucidated. The percentage of clinical trials that were sponsored by the pharmaceutical industry was calculated. Some of the wrong motives of conducting clinical research were identified.

**RESULT:** To some extent, EBM may have contributed to overdiagnose or overuse of medicine. Nearly 46% of clinical trials were financed by pharmaceutical companies. About 90% of manuscripts printed might not need to be published. Many trials contained at least one outcome that did not match its initial specification as registered.

**CONCLUSIONS:** While EBM continues to be the guiding principle, clinicians should be aware of potential tainted results. In the future, big data is likely going to offer us a new aspect of EBM and arm us with more comprehensive data when we make our clinical decisions.

Focusing On Prevention rather than Wait and See Approach

# Shift : fee-for-service to a fee-for outcome


CED Committee for Economic Development

Search

POLICY ISSUES | RESEARCH | PODCASTS | EVENTS | BLOG | PRESS | PEOPLE | ABOUT

IN THE NATION'S INTEREST

## Top Healthcare Stories for 2016: Pay-for-Performance

by
COURTNEY BAIRD
March 08, 2016

Tweet | Share | Share | Email

On February 16 the Centers for Medicare & Medicaid Services (CMS) and America's Health Insurance Plans (AHIP) released their first [illegible] their Core Quality Measures Collaborative – an initiative that aims to reach consensus among public health plans, commercial insurers, physician groups and other stakeholders on major quality measurements. The collaborative supports the recent shift toward value-based care, which stems from America's predicament of high costs and subpar quality in health care.

American Healthcare: Higher Cost and Lower Quality

Source: Adopted from Baird (2016)

**Treatment Today**

- Led to Change the Model from Fee-for-service to Value Based Payments
- Both Incentives & Penalties

# Future Healthcare

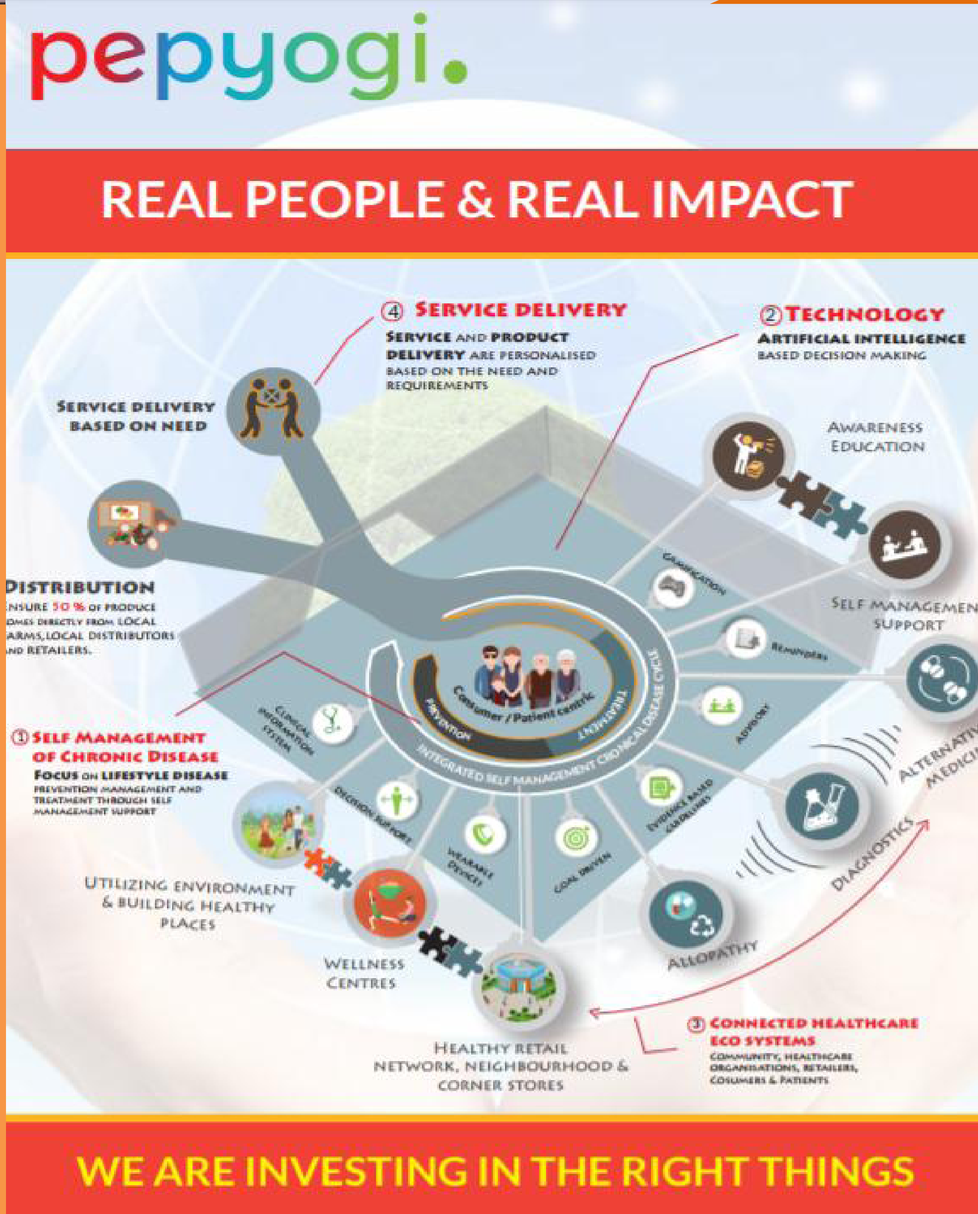
**Everything is Connected**

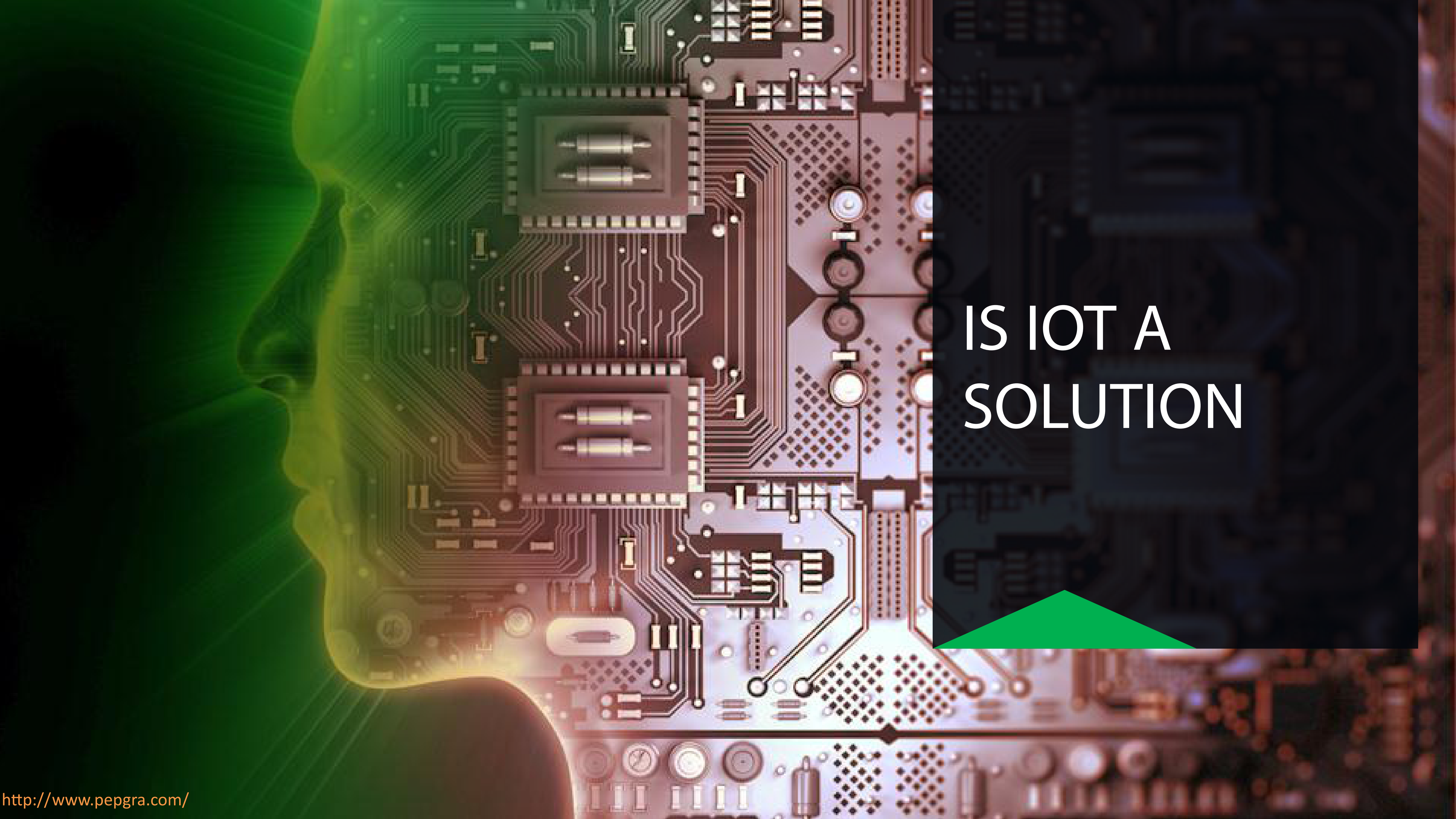
Self Management of Chronic Disease

Technology

Connected Healthcare Ecosystem

Service Delivery





# IS IOT A SOLUTION

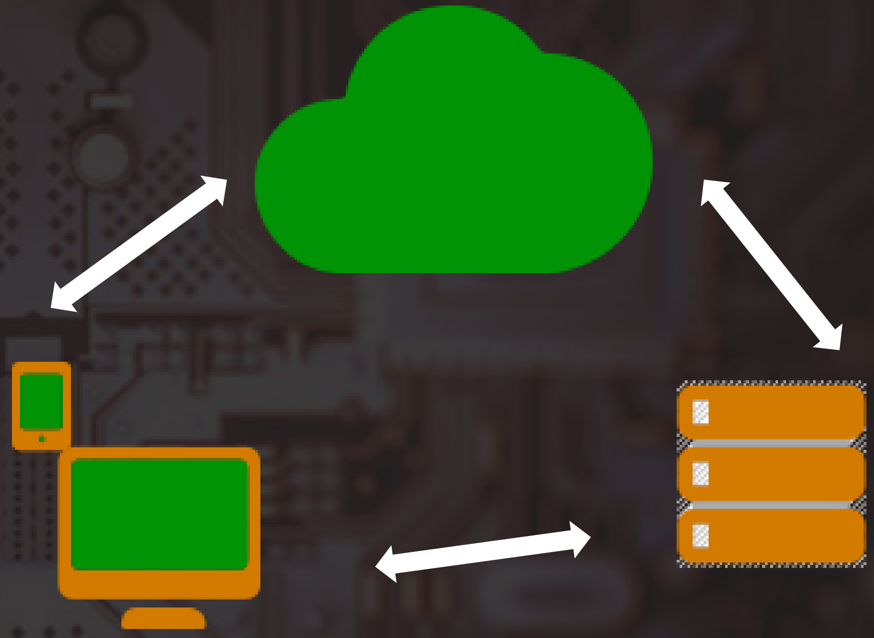
http://www.pepgra.com/

# IoT – Machine Talking to Machine

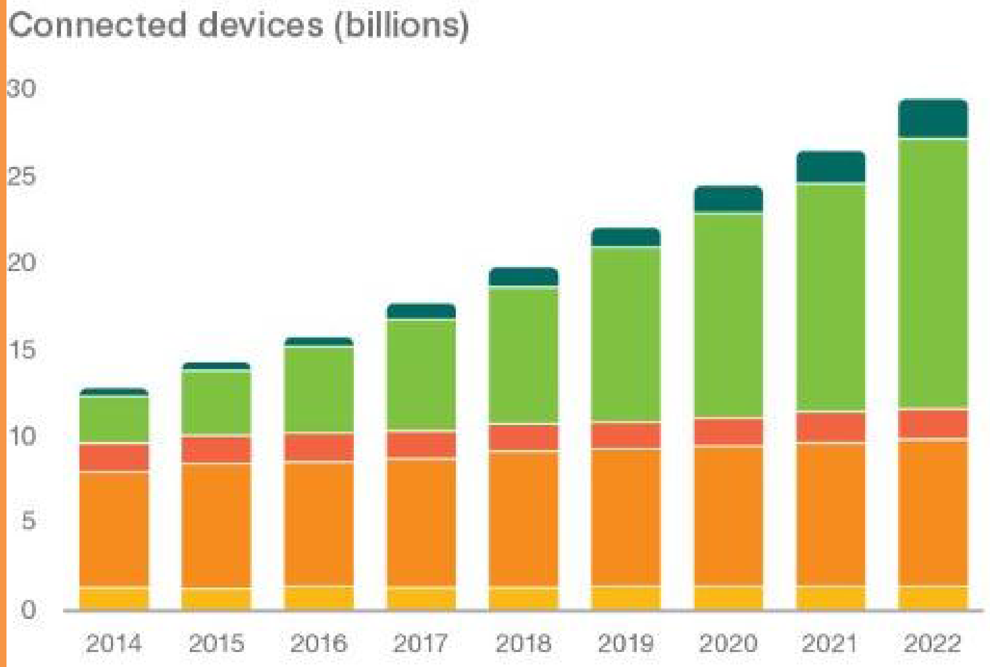
A global Network Infrastructure linking Physical & Virtual Objects

- Infrastructure – Internet and Network developments
- Specific object identification, sensor, and connection capability

Internet of Medical Things, a network devices, connect directly with each other to capture, share and monitor vital data automatically through a SSL that connects a central command and control server in the cloud.



# Prediction of IoT Usage




| | 2016 | 2022 | CAGR |
|---|---|---|---|
| Wide-area IoT | 0.4 | 2.1 | 30% |
| Short-range IoT | 5.2 | 16 | 20% |
| PC/laptop/tablet | 1.6 | 1.7 | 0% |
| Mobile phones | 7.3 | 8.6 | 3% |
| Fixed phones | 1.4 | 1.3 | 0% |
| | 16 billion | 29 billion | 10% |

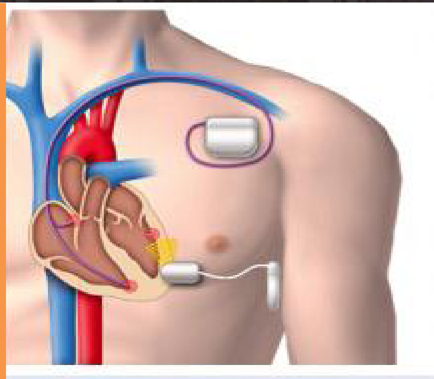
# IoTs for the Challenges We face Today


Blood sampling sensors

Telemedicine

Ingestible sensors (for example, in the form of a pill and eventually dissolved)


External sensors that connect to the body



Tissue-embedded sensors (for example, a pacemaker or implantable cardio defibrillator)

# What it all Delivers?

Data…Data…Data

IoT Generated

> “Data is changing, and it shows no sign *of stopping. Along with that change, the scope and scale of data are* continuously increasing”.

# BIG DATA

# The Model has changed

**Old Model:** Few companies are generating data, all others are consuming data

**New Model:** all of us are generating data, and all of us are consuming data

# Data environment is Rapidly changing

**Healthcare organizations are facing a deluge of rich data that is enabling them to become more efficient, operate with greater insight and effectiveness, and deliver better service**

Sources of the data deluge

Mobile

EMRs

Paper / Text Documents

Social Media

Images

Videos

Sensors / Devices

**40-50%**
**Annual growth in digital data volume***

**62%**
**Annual growth in unstructured data***

**~9X**
**of unstructured data vs. structured data by 2020****

Advances in computing power and techniques

Smarter Algorithms

Faster Processing Speed

Improved Visualization

Retail Consumerism

Value Based Care

Personalized Digital Health

Preventative Medicine

Drug Trials and Discovery

Advances analytical and computing techniques coupled with the explosion of data in healthcare organizations can help uncover leading clinical practices, shrink research discovery time, streamline administration, and offer new personalized engagement paradigms at an industrial scale that align people's decisions and actions in ways that improve outcomes and add value

Customer Centric

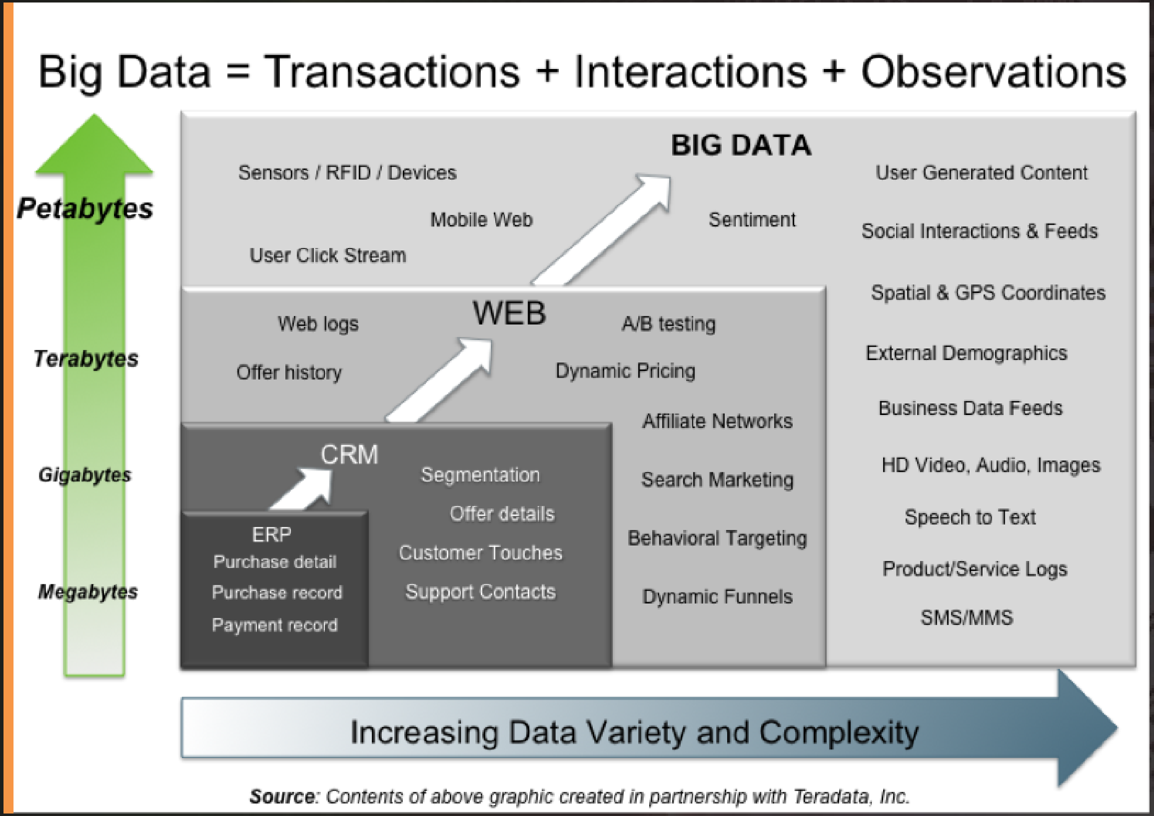
Optimal Cost Structure

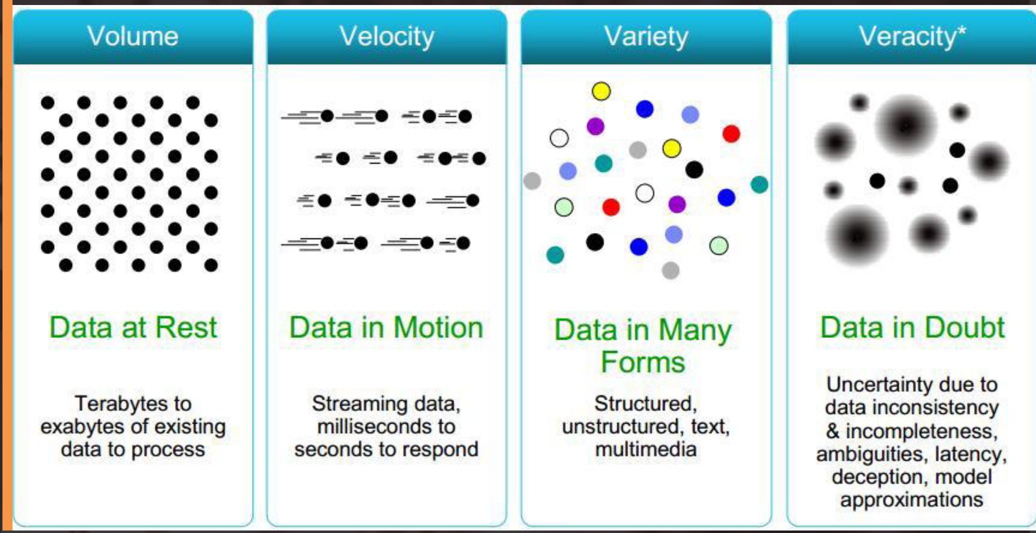
Adaptive Organization

* HP Autonomy, *Transitioning to a new era of human information*, 2013

** Steve Hagan, *Big data, cloud computing, spatial databases*, 2012

# Big Data – 3Vs, 4V, now 6Vs ++ Value, Variability



# What Data is generated?

Log files

EHR data

Social Media sentiments

Clickstream information

Temperature, Pressure, Position, Speed, a Switch that's on or off.

Activity Tracking: date, time, GPS coordinates and Biometrics

Health Activity: Size of a step taken,

Blood pressure, respiration rate, oxygen saturation, heart rate, hydration, galvanic skin response, EKG, Distance, Speed, Step count, fall detection, calories burned, cadence, acceleration, location and altitude,

# Type of Generated

- Non-textual
- Textual
- Audio
- Video
- Presentation
- Pictures
- .rar files

# HealthCare data

**Table 1.** Recent biomedical sensing research works.

| Research | Biomedical Signals | Devices |
|---|---|---|
| Real-time streaming data in healthcare applications [34] | Generic Biomedical signals | Generic Biomedical sensors |
| Recognition of activities and health monitoring [28] | Heart biomedical signals | Smartphones & wearable devices |
| Long-term monitoring of respiration and pulse [26] | Respiration and pulse | Non-contact sensors textile-integrated |
| Diabetes monitoring [29] | Daily activity data | Smartphone & smartwatch |
| Active assistance [30] | Activity and environment data | Wearable sensors and smartphone |
| Detect and prevent venous stasis [27] | Pulse and blood flow data | Multi-sensor plethysmography device |
| Physiological data of elderly patients [33] | Oxygen saturation level, Heart Rate | Biomedical sensors & smartphone |
| ECG Smart Healthcare monitoring [31] | ECG signals | Wearable ECG sensors and Cloud for processing |
| Mobile medical computing systems [32] | Medical signal and context information | Different sensors and actuators |
| Applications in the pervasive environment [35] | Pulse rate, blood pressure, level of alcohol, etc. | Mobile healthcare |

# HealthCare Data - transformed into meaningful insights, which explain the value in 6Vs.

**01**

In healthcare big data environment, the physiological data, EHR, 3D imaging, radiology images, genomic sequencing, clinical, and billing data are the sources of big data, which describe the *volume.*

**02**

Real-time and emergency patient monitoring such as BAN patients, heart beat monitoring and ICU patient monitoring are the sources of streaming data, data arrival rate from the patients escribe the **velocity**.

**03**

Healthcare data such as ECG, EMG and clinical reports are the unstructured data, whereas the patients visits, personal records are the structured data, which describe the **variety**

**04**

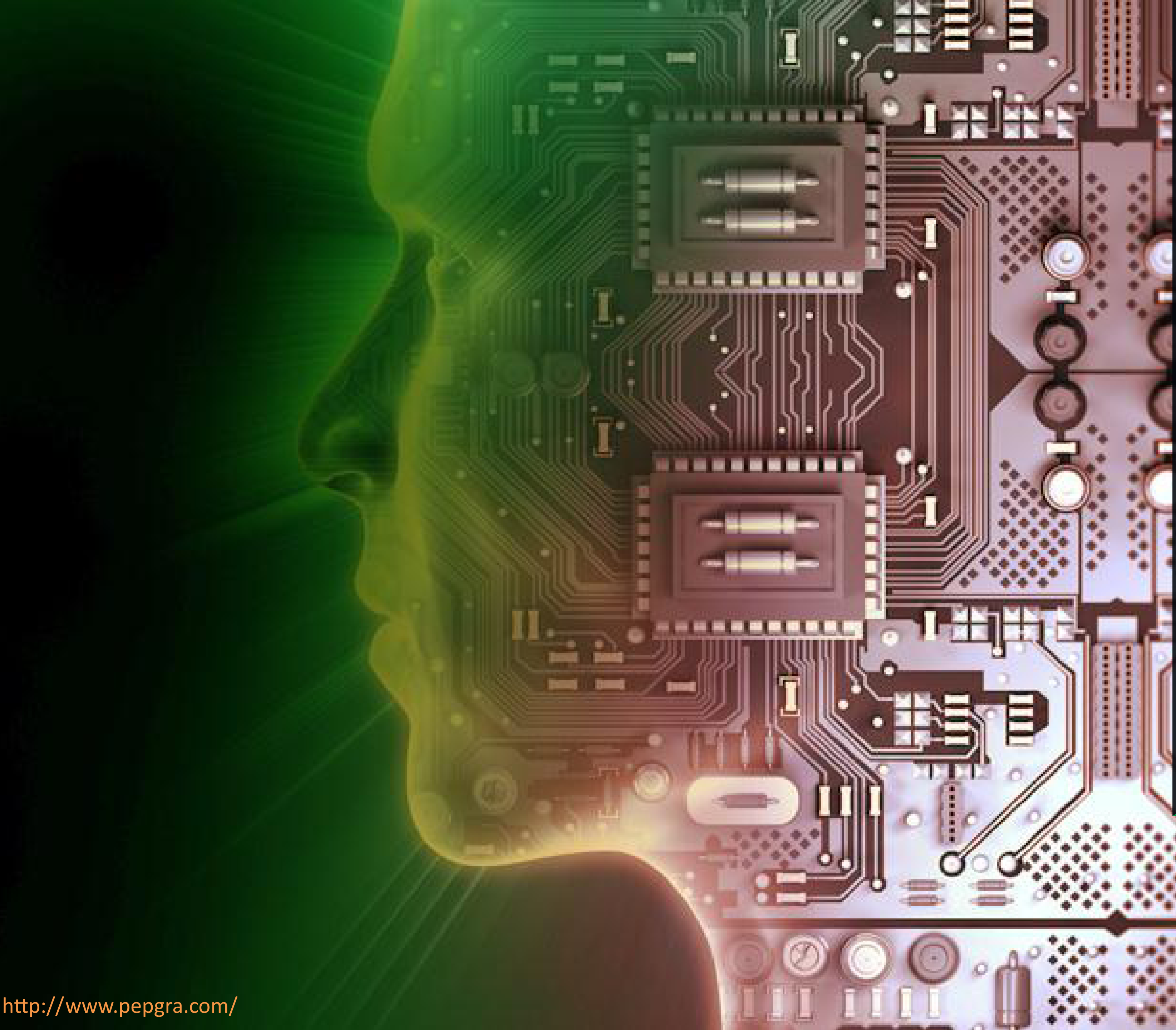
veracity explains the truthfulness of the data sets with respect to data availability and authenticity.

**05**

Variability It deals with inconsistencies in big data flow. Data loads become hard to maintain, especially with the increasing usage of social media that generally causes peaks in data loads when certain events occur

**06**

Value How do large amounts of data influence the value of insight, benefits, and business processes? The



# SO MUCH DATA? WHY, WHAT & HOW?

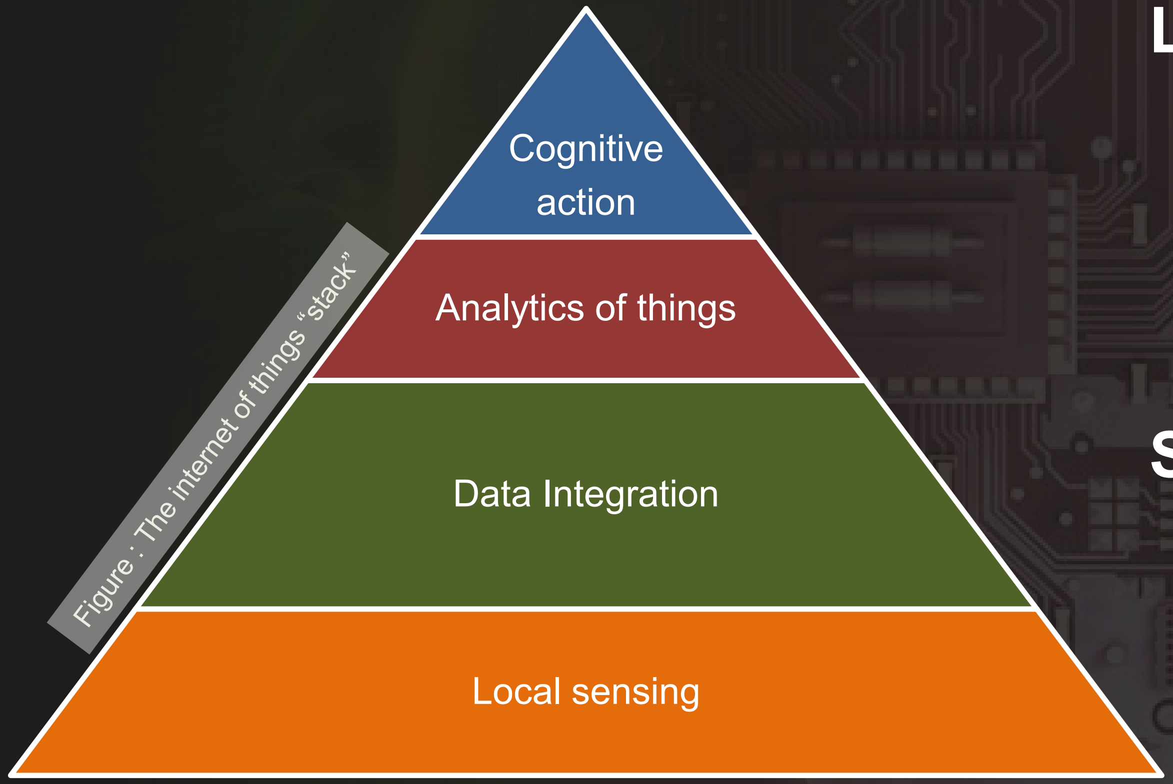
# Why – Prevention, Treatment & Management

- **Descriptive**
  - What Happened?
- **Diagnostic**
  - Why it is Happened?
- **Predictive**
  - What will happen?
- **Prescriptive**
  - How we can make it Happen

# What - other types of analytics of things are there?

- Understanding patterns and reasons for variation—developing statistical models that explain variation
- Anomaly detection—identifying situations that are outside of identified boundary conditions, such as a temperature that is too high or an image depicting someone in an area that should be uninhabited
- Predictive asset maintenance—using sensor data to detect potential problems before they actually occur, risk of classifying patients
- Optimization—using sensor data and analysis to optimize a process, as dosage adjustment, food/diet adjustments,
- Prescription—employing sensor and other types of data to tell patients what to do next, surgery, or diet or medications
- Situational awareness—piecing together seemingly disconnected events and putting together an explanation, as with socioeconomic condition, walking, diet, medication adherence, will lead to less Hba1c

# How ?




## Layer 1

Sensor layer – Integrated smart objects along with the sensor. These sensors empower the interconnections of the real worked and the physical measurements for real time information process.

## Sensors

Measures – quality, temp, electricity and movement

Sensors entails connectivity to the senor gateways in the form of personal area networks PAN such as Bluetooth, ZigBee, Ultra-wideband, LAN, WiFi, ethernet connections

# Layer 2, & 3: Data Integration & Analytics

Figure 1 shows a very general IoT scheme, which is the approach shown in most of the works reviewed in the state of the art. There are many tasks throughout the IoT process that can be divided more efficiently.

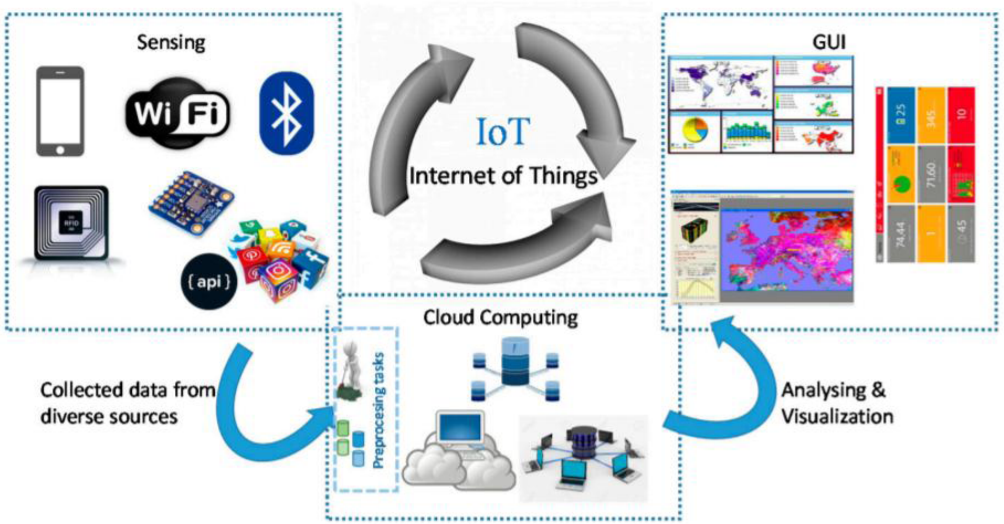


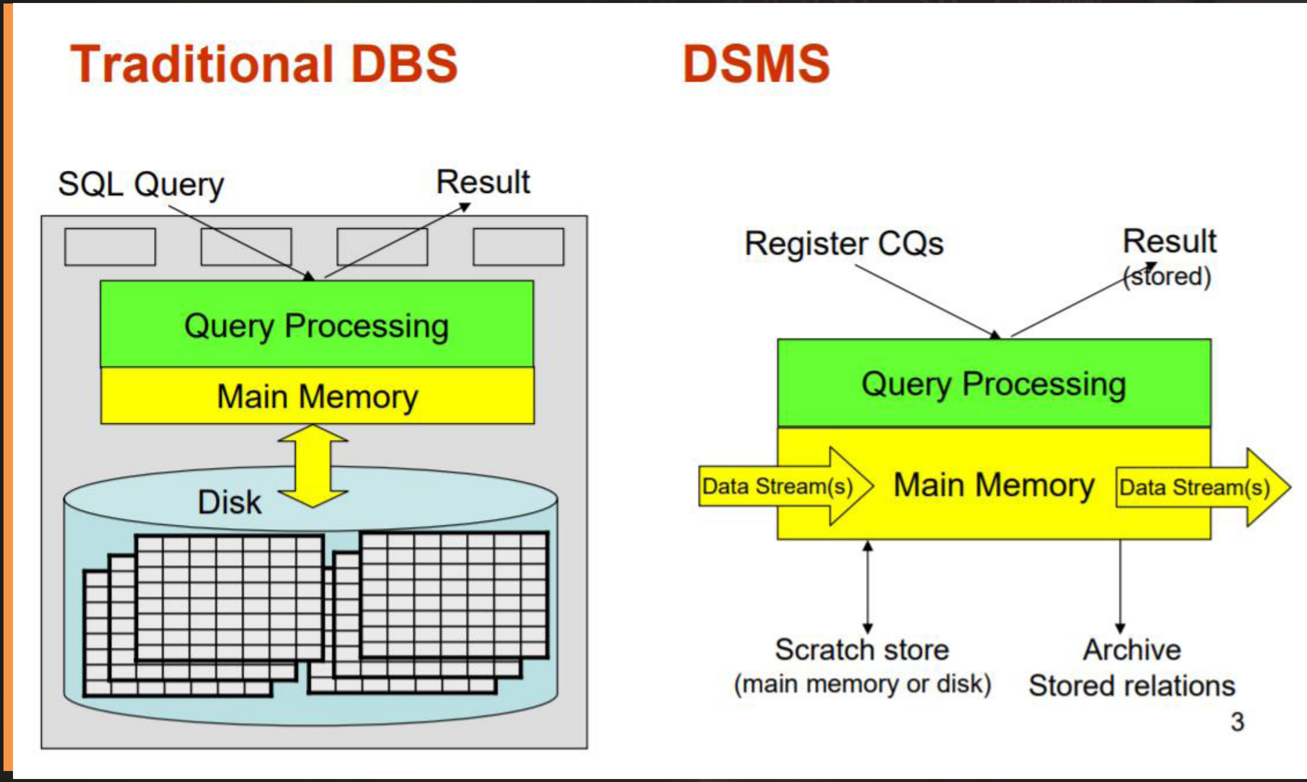

**Figure 1.** General schema of IoT.

Figure : General Schema of IoT

Source; Adopted from Mora et al. (2017)

Figure shows a very general IoT scheme, Which is the approach shown in most of the words reviewed in the states of the art . There are many tasks throughout the IoT process that can be divided more efficiency

# Database Management System (DBMS)

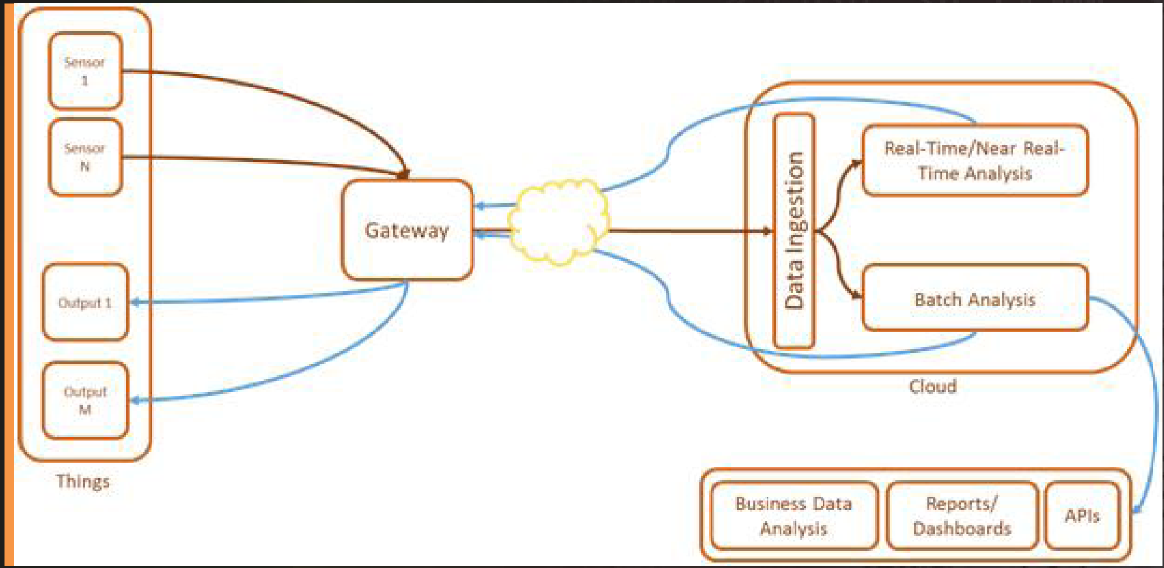


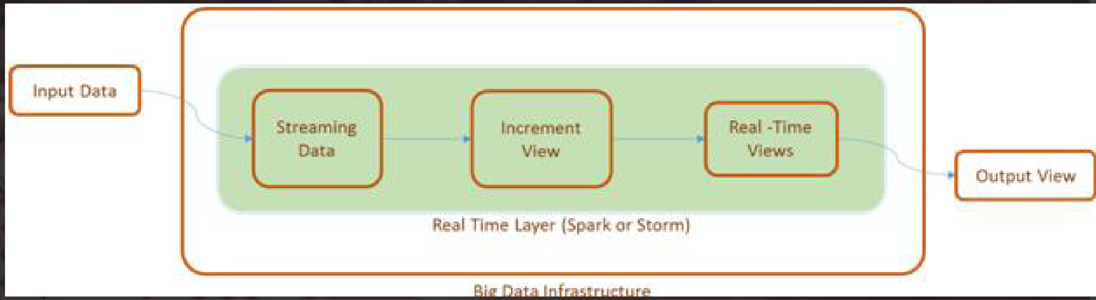

Conventional DBMSs are designed to process queries over finite stored datasets.

Query Semnatics: One time query that logically do not change while a query runs vs. Continuous queries

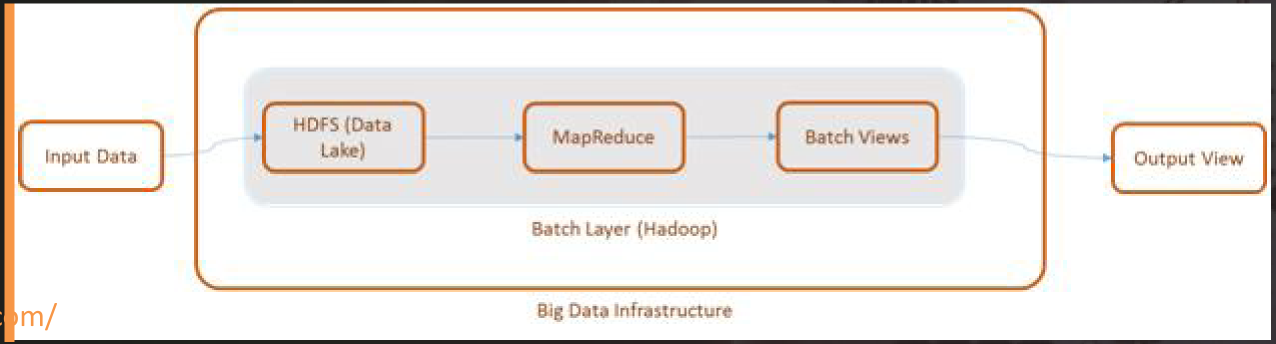
Query Plan chosen – one per query using statistics available vs. adaptive execution plan based on stream and system conditions as query runs

# Data Stream




Pseudo real time analytics: Following are different options for implementing the real-time layer




Huge volumes of data handled by batch operations. & processed from permanent distributed storages using Hadoop MapReduce or in-memory computations using Apache Spark. Apache Pig and Hive are used for data querying and analyses. Since these run on cheap commodity servers on a distributed manner, they are the best bet for processing historical data and deriving insights and predictive models out of it.

# Real Time data stream

These types of analytics refer to the system that depends on instantaneous feedback based on the data received from the sensors.

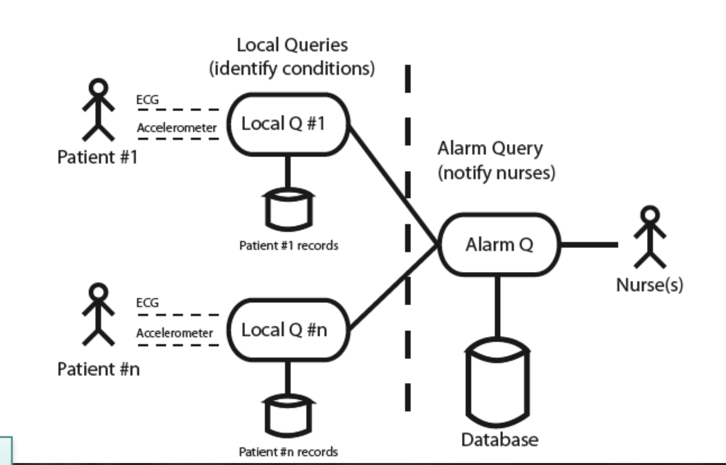
For example, IoT receives data from numerous sensors on a patient's body. Need to aggregate real time data & run algorithms to detect situations that need immediate medical attention.

E.g. A medical provider or an emergency response system should be notified immediately or who needs 24/ 7 health status observation.

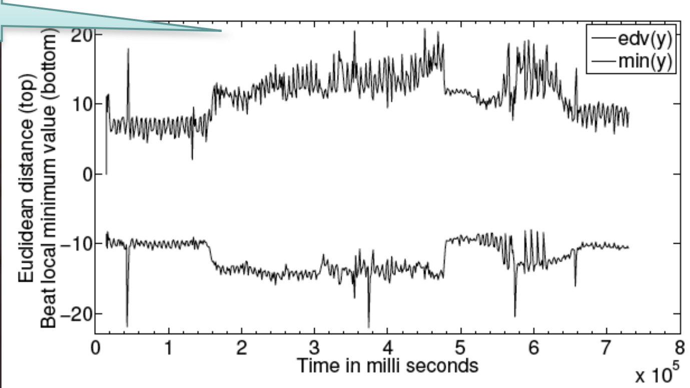
Analysis-response cycle should only take few seconds as every second would be a matter of life and death.

An Example of Network Data Stream

| Timestamp | Source | Destination | Duration | Bytes | Protocol |
|---|---|---|---|---|---|
| 11001 | 10.0.0.1 | 14.2.4.1 | 14 | 16K | http |
| 11002 | 17.3.0.2 | 12.1.4.3 | 18 | 46K | http |
| 11003 | 12.4.8.5 | 16.2.8.7 | 30 | 70K | ftp |
| 11004 | 19.7.0.1 | 10.1.0.1 | 14 | 28K | sftp |
| ... | .... | ... | ... | ... | ... |




Heartattack!

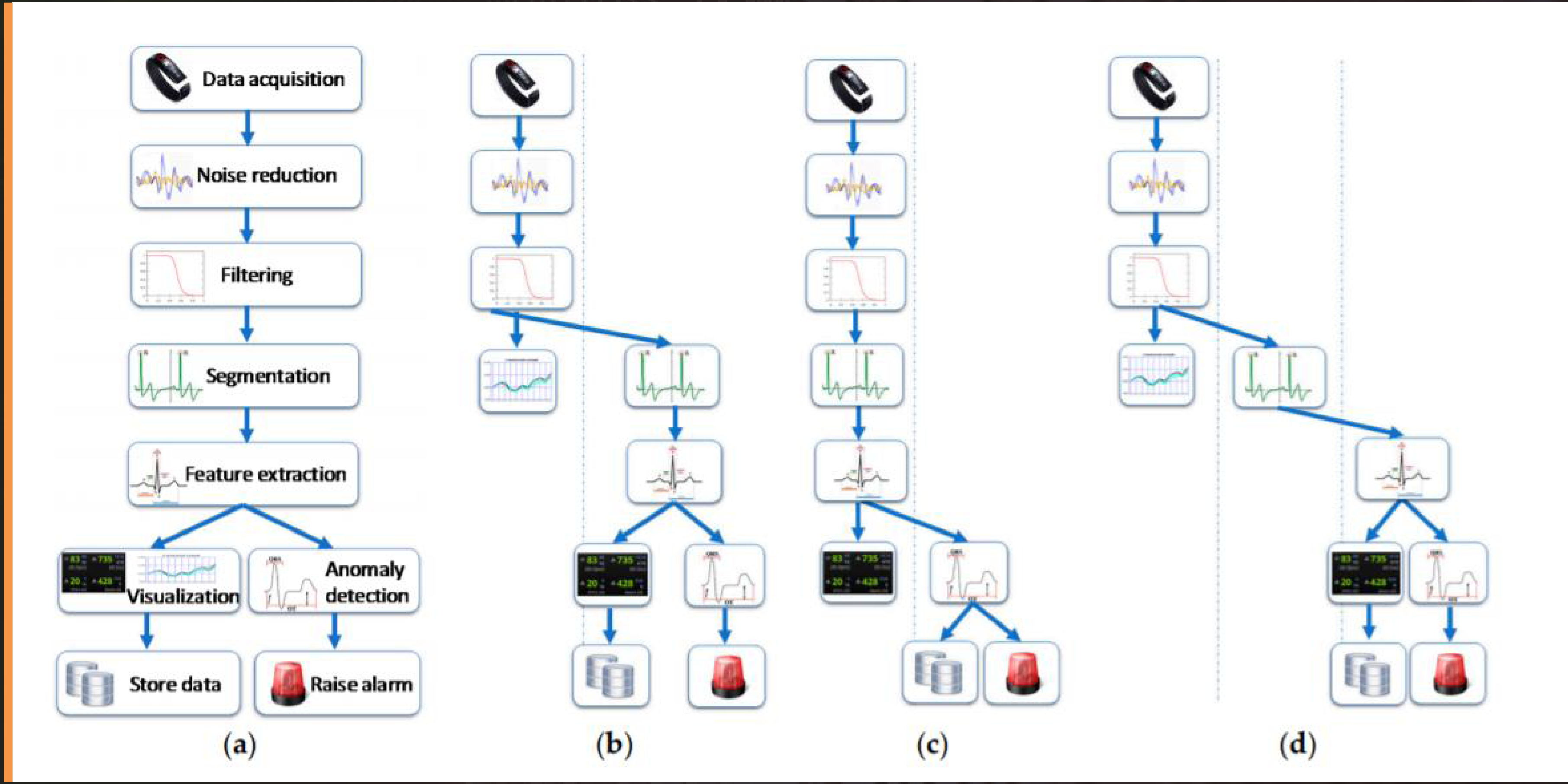


# Solution

A classic Hadoop based solution might not work in the above cases because of the fact that it relies on MapReduce which is considerable slow involving costly IO operations.

The solution is to augment Hadoop ecosystem with a faster real-time engine like Spark, Storm, Kafka, Trident – Scalable, reliable, distributed, scalable, high throughput, fault tolerant, fast and real time computing to process high velocity data – to process high velocity data stream

# Application of IoT Based Framework



# IoT Framework

Data acquisition: by the biosensors. A digital system, before processing, the analog ECG signal converted to a digital form at a particular sampling rate or frequency

→ Noise reduction – biosignals noisy due to weak signals.

→ Filtering – remove freq. or attenuate from the input data (linear, nonlinear, polynomial, high pas filters, finite impulse

→ Segmentation – accuracy of QRS complex

→ Feature extraction – discrimination – e.g. wavelet

→ Anomaly detection – heart failure, myocardial infarction – accuracy and computational costs

→ visualization

→ Store data

→ Raise alarms

# Data Acquisition

The first step is to be able to acquire and filter the massive input stream generated by millions of sources from the IoT at an application-defined frequency.

- To define online filters in order to discard redundant data without loss of useful information (at the source level, or at a higher level).
- when a jogger stops to take a rest her sensor reads the same value at regular intervals. These values could be locally filtered in order to compress the input data set. We showed that the input workload is continuous but that the flow rate varies over time.
- A key challenge is to design and implement a scalable way of supporting a variable number of connected objects in order to handle peaks of workload.

# Data Cleaning

**Sensor data from smartphones is inherently erroneous and uncertain.**

The main factors are battery life, imprecision, and transmission failures. This problem is especially challenging when we consider stream processing.

For instance, a smartphone can exhaust its battery life in the middle of the route or its GPS sensor can position it outside the route, which corrupts the resulting GPS trace.

Addressing this problem requires detection and correction of this kind of data by performing online data cleaning.

# Data Processing

Data processing requires faster speed, and in many areas data have been requested to carry out in real-time processing such as disease risk prediction and requirement of surgery or not

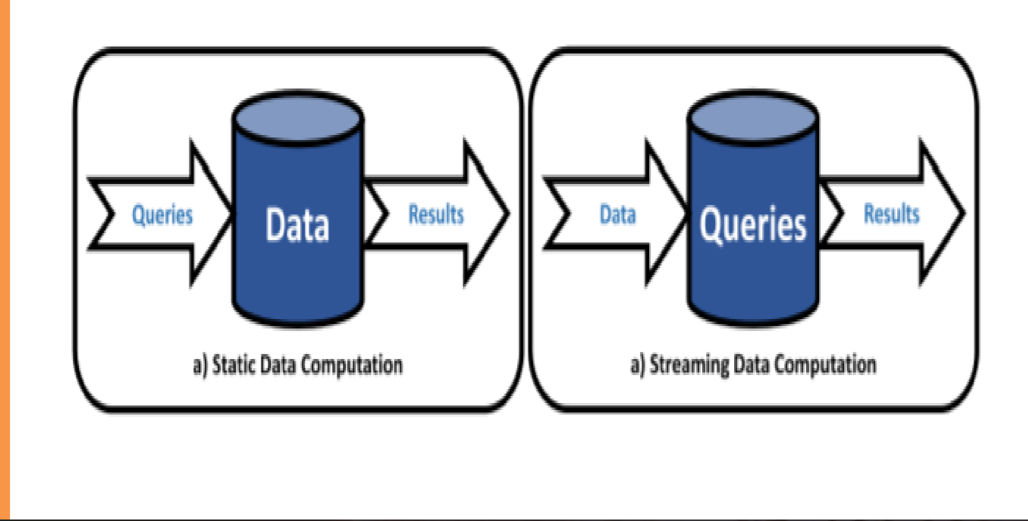



Figure : static data computation versus streaming data computation

Source: Adopted from IBM (2017)

# Data Processing

| Name | Release Year | Description | DBMS | DSMS | CEP | Distributed |
|---|---|---|---|---|---|---|
| Google Photon [3] | 2013 | Distributed stream processing system | | • | | • |
| Walmart Muppet [14] | 2012 | Distributed event processing system | | • | | • |
| StreamDrill [23] | 2012 | Stream processing system | | | • | |
| SAP HANA [8] | 2011 | In-memory database ‡ | • | • | | • |
| Apache Storm [15] | 2011 | Distributed stream processing system | | • | | • |
| Apache YARN [25] | 2011 | Distributed general-purpose processing system † | • | • | • | • |
| Apache Flume [9] | 2011 | Distributed stream processing system ‡ | | • | | • |
| Apache Kafka [13] | 2011 | Distributed stream processing system ‡ | | • | | • |
| Apache S4 [18] | 2011 | Distributed event processing system. | | • | | • |
| Apache Chukwa [20] | 2010 | Distributed stream processing system † | | • | | • |
| HStreaming [11] | 2010 | Distributed stream processing system † | | • | | • |
| AMPLab Spark [26] | 2010 | Distributed general-purpose processing system ‡ | • | • | | • |
| VoltDB [24] | 2010 | In-memory distributed database ‡ | • | • | | • |
| Esper [7] | 2006 | Complex Event Processing System | | | • | |
| StreamBase CEP [22] | 2003 | Distributed complex event processing system ‡ | | | • | • |
| SQLstream [21] | 2003 | Distributed stream processing system ‡ | | • | | • |

† Tools based on Hadoop's infrastructure
‡ Tools that can interact with Hadoop's infrastructure

Source: Adopted from Carvalho et al. (2013)

Table: List of event processing tools and his main characteristics

# Query Processing Challenges

Query processing in the data stream model of computation comes with its own unique challenges

Unbounded in size, the amount of storage required to compute an exact answer to a data stream query may also grow without bound. While external memory algorithms for handling data sets larger than main memory have been studied, such algorithms are not well suited to data stream applications since they do not support continuous queries and are typically too slow for real-time response.

Approximation algorithms for problems defined over data streams has been a fruitful research area in the algorithms community - for data reduction and synopsis construction, including: sketches, random sampling , histograms , and wavelets .

Window Sliding: One technique for producing an approximate answer to a data stream query is to evaluate the query not over the entire past history of the data streams, but rather only over sliding windows of recent data from the streams. For example, only data from the last week could be considered in producing query answers, with data older than one week being discarded.

# Stream Data Mining

Traditional data clustering algorithms such as K-means Self Organizing Maps, density based clustering techniques such as DBScan and CLIQUE, are applied on finite static data

because data streams are infinite, data stream mining algorithms need to process the data in single pass

Anytime data mining algorithms such as K processing, anytime learning, anytime classification

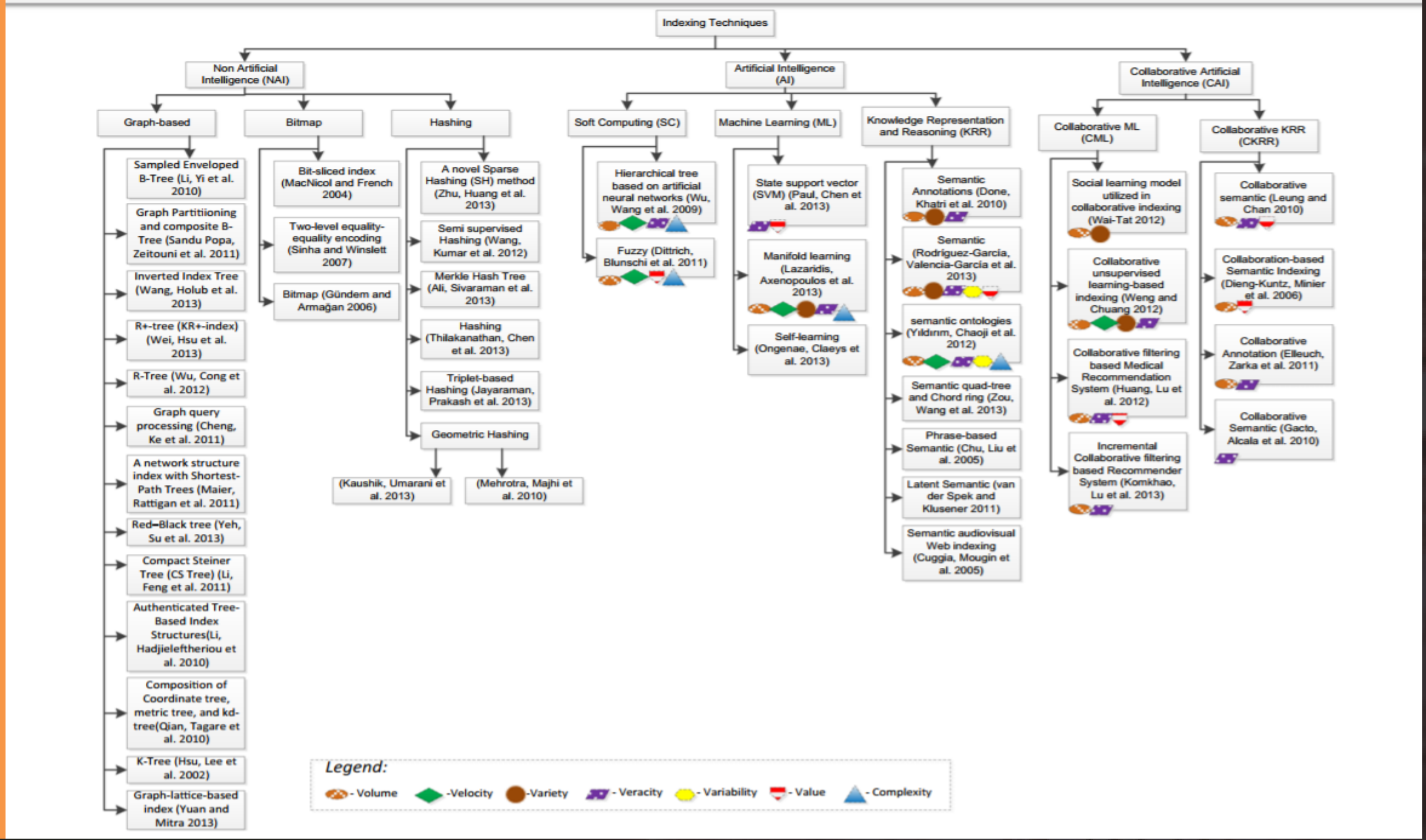
# Stream Data Mining

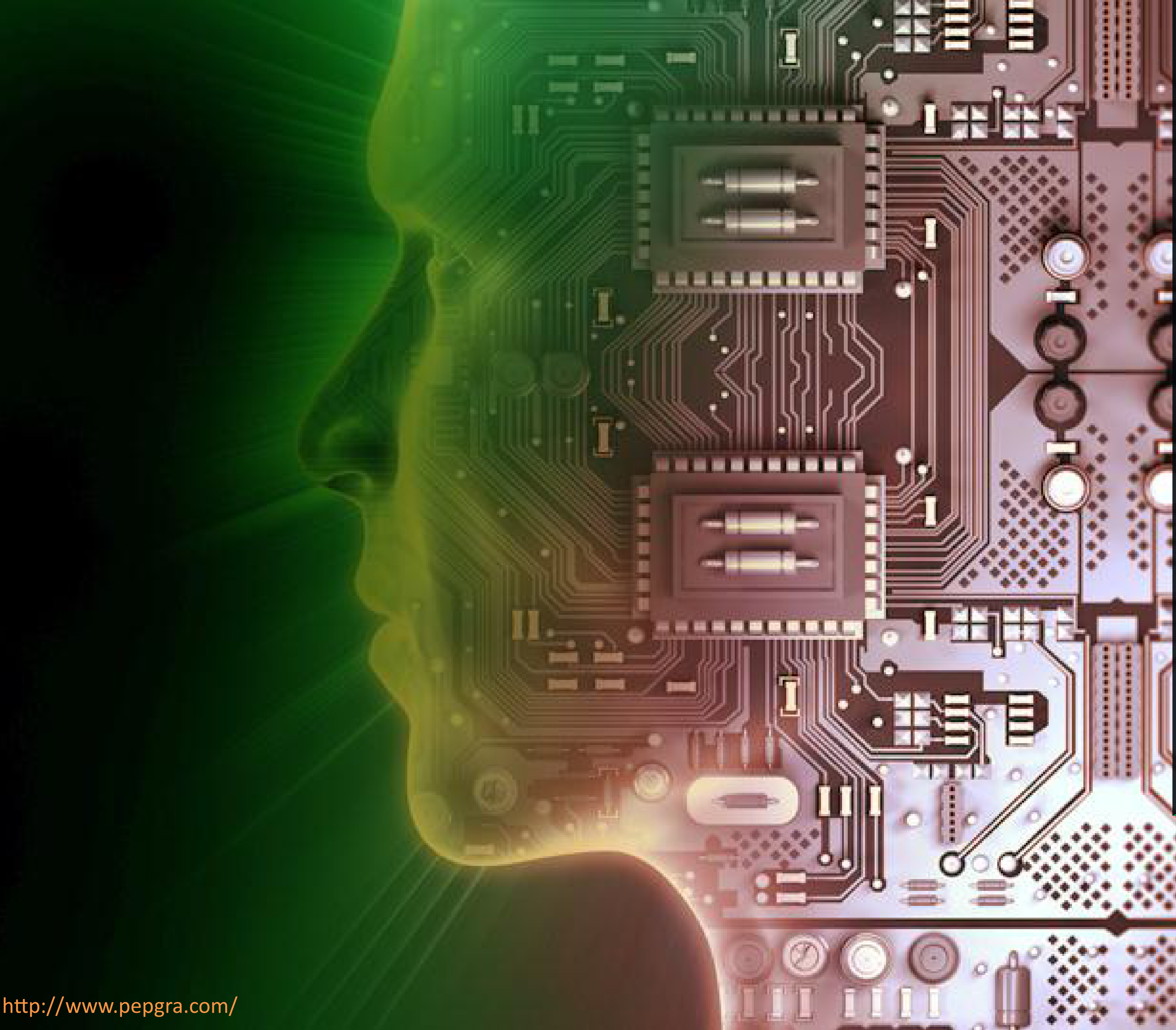
Traditional data clustering algorithms such as K-means Self Organizing Maps, density based clustering techniques such as DBScan and CLIQUE, are applied on finite static data

because data streams are infinite, data stream mining algorithms need to process the data in single pass

Anytime data mining algorithms such as K processing, anytime learning, anytime classification

# Data Indexing

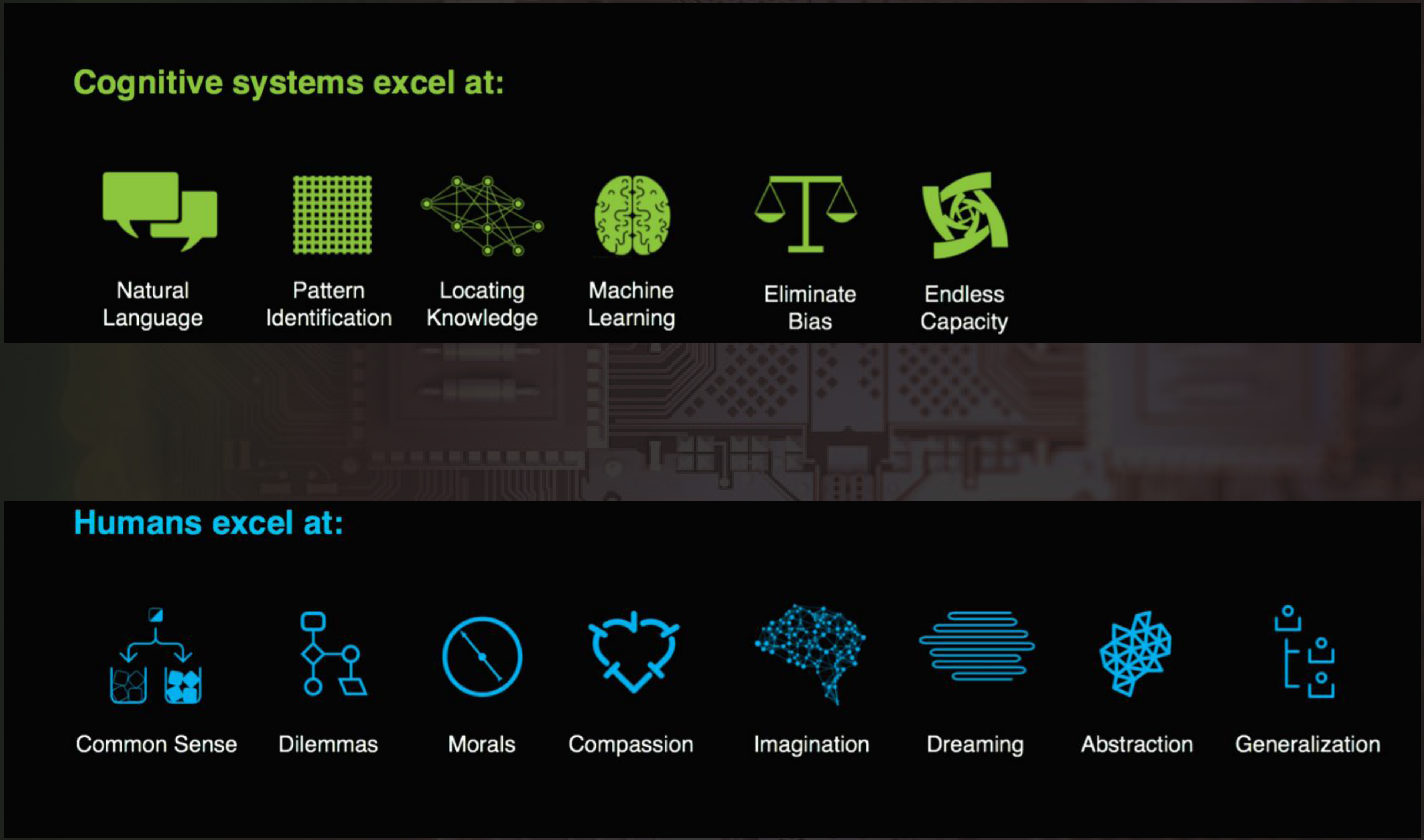


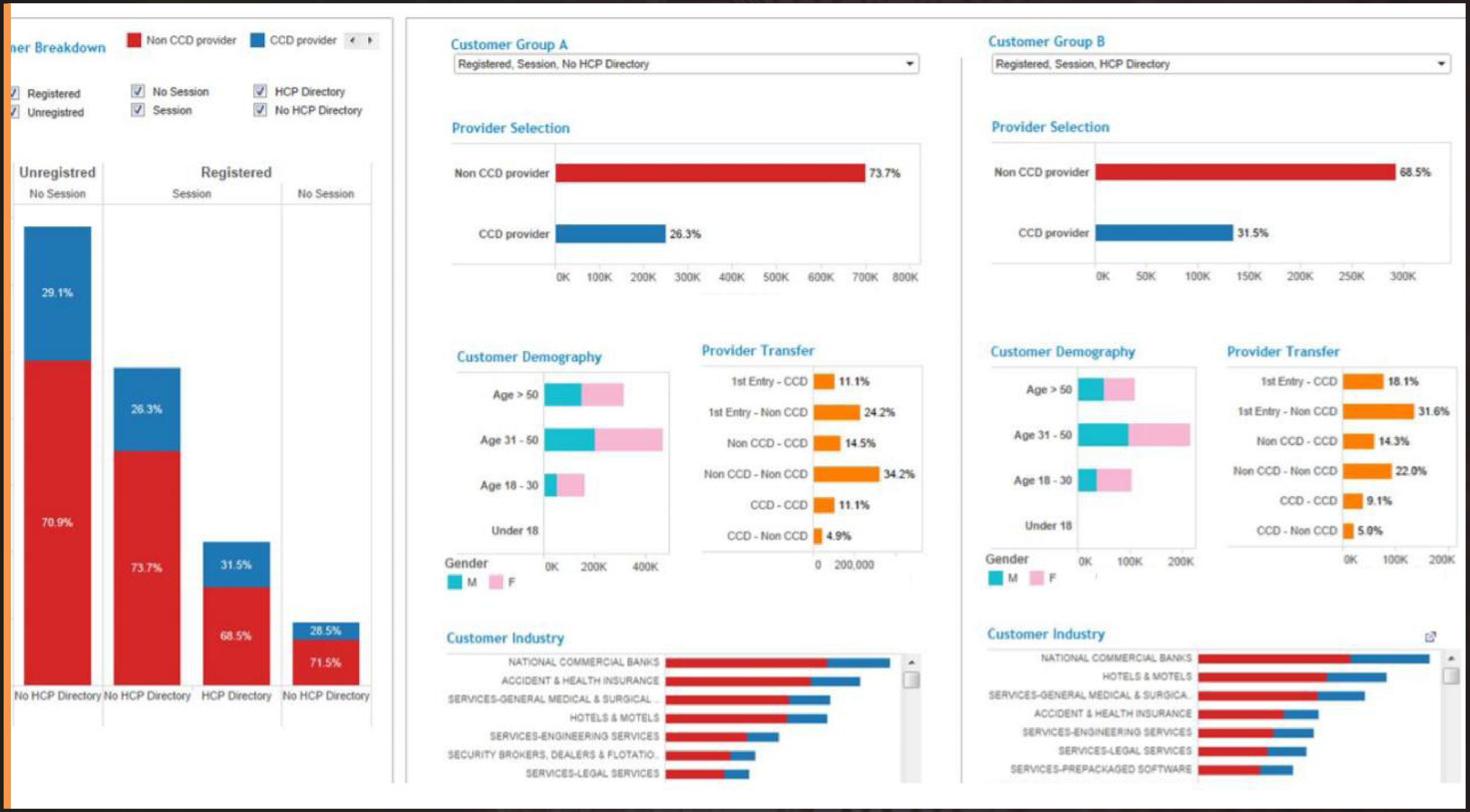


# VISUALIZATION

**Most forgotten areas**

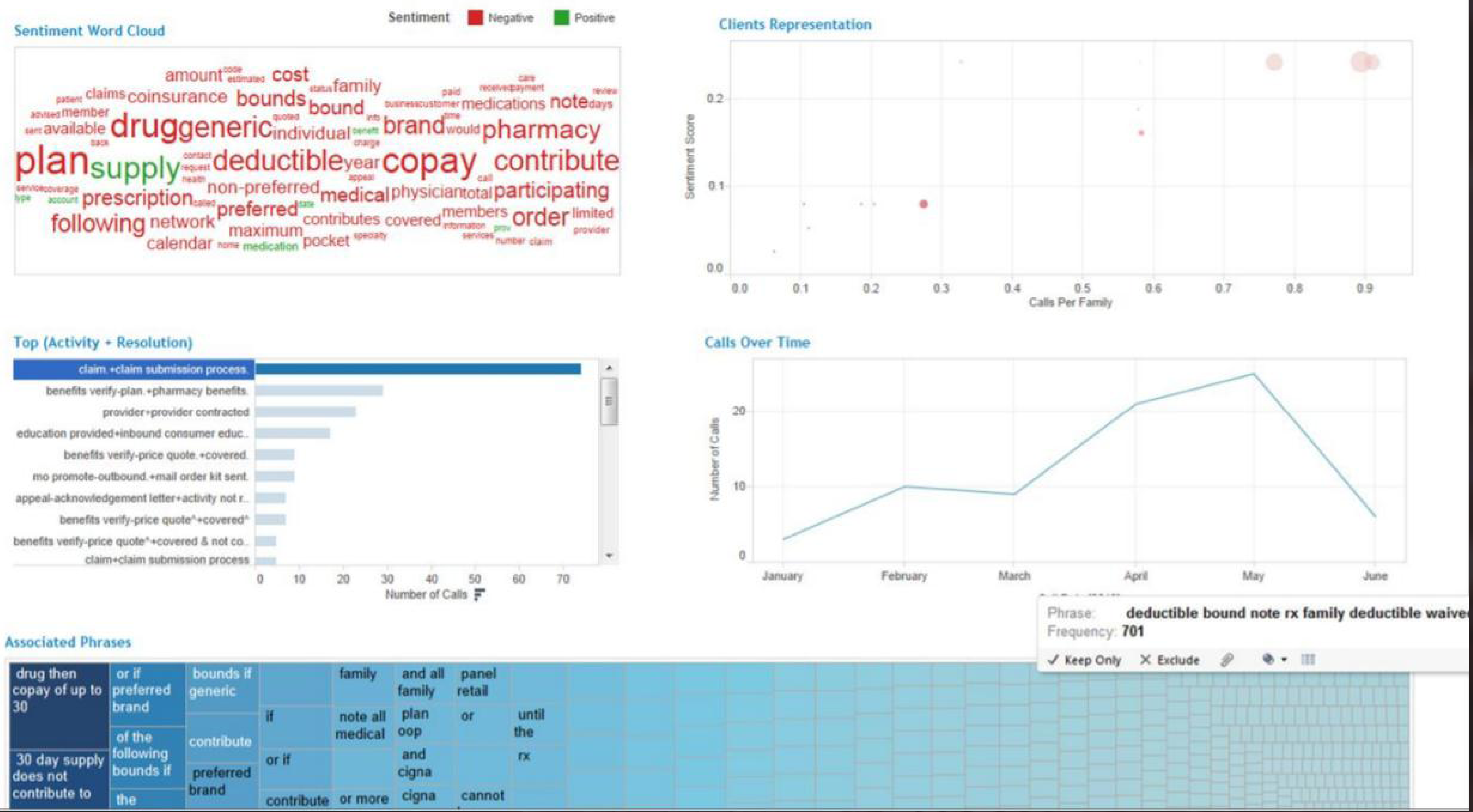
http://www.pepgra.com/


Cognitive systems excel at:
Natural Language
Pattern Identification
Locating Knowledge
Machine Learning
Eliminate Bias
Endless Capacity
Humans excel at:
Common Sense
Dilemmas
Morals
Compassion
Imagination
Dreaming
Abstraction
Generalization

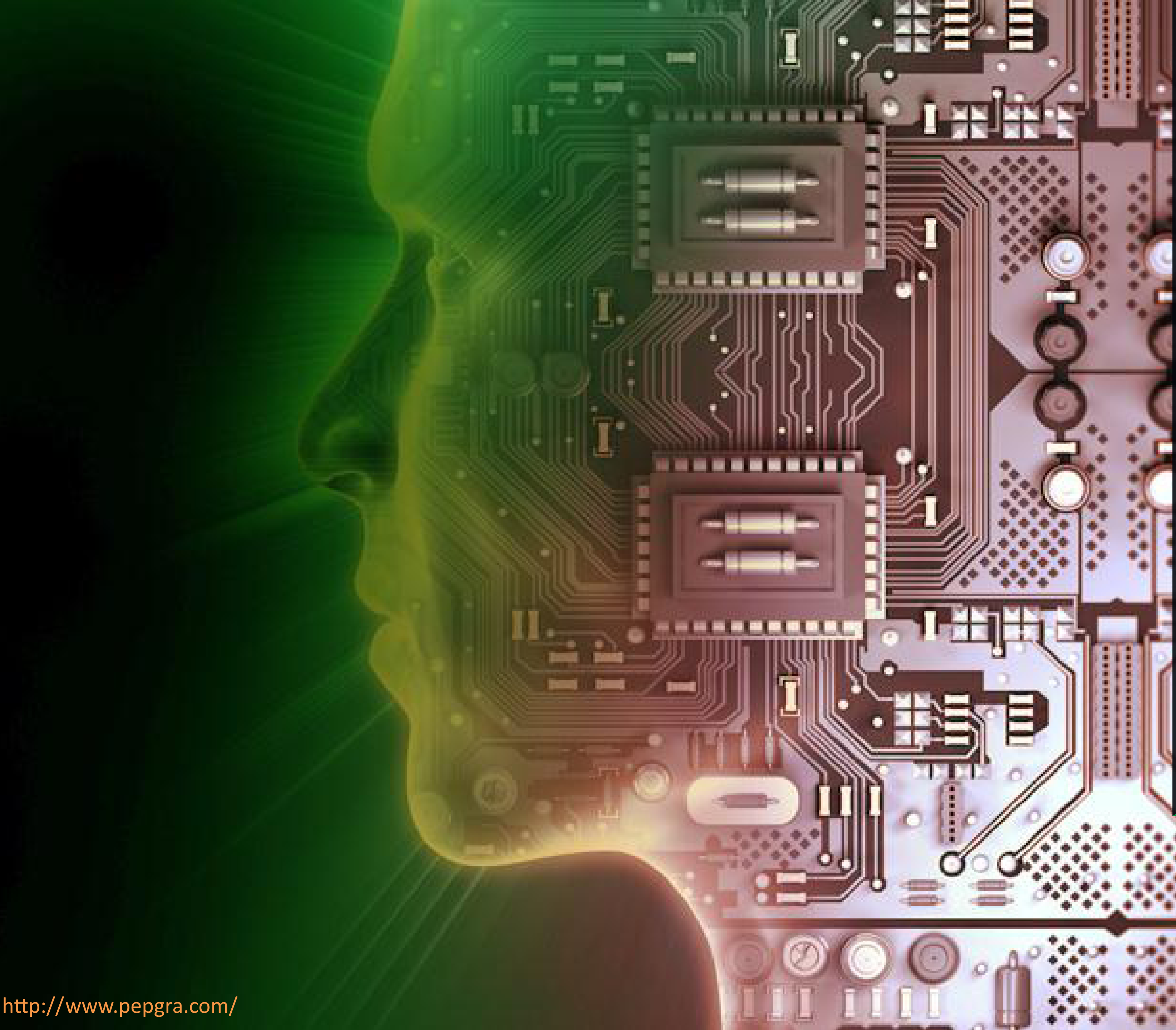


ner Breakdown
Non CCD provider
CCD provider
Registered
Unregistred
No Session
Session
HCP Directory
No HCP Directory
Unregistred
Registered
No Session
Session
No Session
29.1%
70.9%
26.3%
73.7%
31.5%
68.5%
28.5%
71.5%
No HCP Directory
No HCP Directory
HCP Directory
No HCP Directory
Customer Group A
Registered, Session, No HCP Directory
Provider Selection
Non CCD provider
73.7%
CCD provider
26.3%
0K 100K 200K 300K 400K 500K 600K 700K 800K
Customer Demography
Age > 50
Age 31 - 50
Age 18 - 30
Under 18
0K 200K 400K
Gender
M
F
Provider Transfer
1st Entry - CCD 11.1%
1st Entry - Non CCD 24.2%
Non CCD - CCD 14.5%
Non CCD - Non CCD 34.2%
CCD - CCD 11.1%
CCD - Non CCD 4.9%
0 200,000
Customer Industry
NATIONAL COMMERCIAL BANKS
ACCIDENT & HEALTH INSURANCE
SERVICES-GENERAL MEDICAL & SURGICAL
HOTELS & MOTELS
SERVICES-ENGINEERING SERVICES
SECURITY BROKERS, DEALERS & FLOTATIO..
SERVICES-LEGAL SERVICES
Customer Group B
Registered, Session, HCP Directory
Provider Selection
Non CCD provider
68.5%
CCD provider
31.5%
0K 50K 100K 150K 200K 250K 300K
Customer Demography
Age > 50
Age 31 - 50
Age 18 - 30
Under 18
0K 100K 200K
Gender
M
F
Provider Transfer
1st Entry - CCD 18.1%
1st Entry - Non CCD 31.6%
Non CCD - CCD 14.3%
Non CCD - Non CCD 22.0%
CCD - CCD 9.1%
CCD - Non CCD 5.0%
0K 100K 200K
Customer Industry
NATIONAL COMMERCIAL BANKS
HOTELS & MOTELS
SERVICES-GENERAL MEDICAL & SURGICA..
ACCIDENT & HEALTH INSURANCE
SERVICES-ENGINEERING SERVICES
SERVICES-LEGAL SERVICES
SERVICES-PREPACKAGED SOFTWARE

# Use Case 2



Sentiment Word Cloud
Sentiment Negative Positive
amount cost family claims coinsurance bounds bound medications note member available drug generic individual brand pharmacy plan supply deductible year copay contribute non-preferred medical physician total participating prescription preferred following network maximum contributes covered members order limited calendar medication pocket
Clients Representation
Sentiment Score
Calls Per Family
Top (Activity + Resolution)
claim.+claim submission process.
benefits verify-plan.+pharmacy benefits.
provider+provider contracted
education provided+inbound consumer educ..
benefits verify-price quote.+covered.
mo promote-outbound.+mail order kit sent.
appeal-acknowledgement letter+activity not r..
benefits verify-price quote^+covered^
benefits verify-price quote^+covered & not co..
claim+claim submission process
Number of Calls
Calls Over Time
Number of Calls
January
February
March
April
May
June
Phrase: deductible bound note rx family deductible waive
Frequency: 701
Keep Only
Exclude
Associated Phrases
drug then copay of up to 30
30 day supply does not contribute to
or if preferred brand
of the following bounds if
the
bounds if generic
contribute
preferred brand
if
or if
contribute
family
note all medical
or more
and all family
plan oop
and cigna
cigna
panel retail
or
cannot
until the
rx



# WHERE I WILL GET THE DATA

For researchers /academicians

# Data Repositories

1. Data.gov
2. US Census Bureau
3. European Union Open Data Portal
4. Data.gov.uk
5. The CIA World Factbook
6. Healthdata.gov
7. NHS Health and Social Care Information Centre
8. Amazon Web Services public datasets
9. Facebook Graph
10. Gapminder
11. Google Trends
12. Google Finance
13. Google Books Ngrams
14. National Climatic Data Center
15. DBPedia
16. Topsy
17. Likebutton
18. New York Times
19. Freebase
20. Million Song Data Set

Read original article with description for each data repository.

https://www.datasciencecentral.com/profiles/blogs/great-sensor-datasets-to-prepare-your-next-career-move-in-iot-int

https://archive.ics.uci.edu/ml/datasets.html

https://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free

https://www.datasciencecentral.com/profiles/blogs/20-free-big-data-sources-everyone-should-check-out

http://www.conduitlab.org/data-sources/

https://github.com/TomLous/coursera-getting-cleaning-data-project

http://efavdb.com/machine-learning-with-wearable-sensors/

https://awearable.tumblr.com/post/115835279113/the-datasets

https://zenodo.org/record/14996#.WfqnKGiCw2w

https://old.datahub.io/dataset/knoesis-linked-sensor-data

https://www.healthdata.gov/search/type/dataset

# Challenges for "At Risk" Patient Identification & Intervention

## Data Challenges

- **Large Scale**: Up to 10s of millions of patients
- **High Dimensionality:** Thousands of dimensions spanning many years
- **Semi-Structured:** Clinical notes, imaging, medical codes
- **Distributed:** Multiple providers and representations
- **Sparse and Irregular:** Periodic visits, different for each person
- **Uncertain:** Subjective, data entry errors, bias for billing
- **Incomplete:** Many items missing from the medical record

## Task Challenges

- **Critical decisions:** May literally mean life or death
- **No clear right answer:** Evidence is often ambiguous
- **Limited time**: Manage complexity, multiple granularity
- Domain experts are people…
- Data analytics and statistics
- Visualization and user interaction
- Systems
  - Too much (or too little) trust in numbers
  - "But my patients are different…"
  - Users resistant to technology change

# Conclusion

Better treatments…..

**More efficient**
care…………………

# References


Baird, C. (2016) Top Healthcare Stories for 2016: Pay-for-Performance. Available at: https://www.ced.org/blog/entry/top-healthcare-stories-for-2016-pay-for-performance

Carvalho, O.M. de, Roloff, E. & Navaux, P.O.A. (2013). A Survey of the State-of-the-art in Event Processing. 11th Workshop on Parallel and Distributed Processing (WSPPD).

IBM (2017). An introduction to InfoSphere Streams. [Online]. 2017. IBM. Available from: https://www.ibm.com/developerworks/library/bd-streamsintro/index.html. [Accessed: 21 November 2017].

Insights (2017) Insights into Editorial: Ageing with dignity. Available at: http://www.insightsonindia.com/2017/02/24/insights-editorial-ageing-dignity/

Kiddie, J. Y. (2017) New Obesity Study Sheds Light on Dietary Recommendations. Available at: http://www.bbdnutrition.com/2017/06/14/new-obesity-study-sheds-light-on-dietary-recommendations/

Mora, H., Gil, D., Terol, R.M., Azorín, J. & Szymanski, J. (2017). An IoT-Based Computational Framework for Healthcare Monitoring in Mobile Environments. Sensors. [Online]. 17 (10). p.p. 2302. Available from: http://www.mdpi.com/1424-8220/17/10/2302.

Randløv, J. & Poulsen, J.U. (2008). How much do forgotten insulin injections matter to hemoglobin a1c in people with diabetes? A simulation study. Journal of diabetes science and technology. [Online]. 2 (2). pp. 229–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19885347.

Vint (2012). Creating clarity with Big Data. [Online]. 2012. vint. Available from: http://blog.vint.sogeti.com/wp-content/uploads/2012/06/Creating-clarity-with-Big-Data-VINT-Research-Report.pdf. [Accessed: 21 November 2017].

You, S. (2016). Perspective and future of evidence-based medicine. Stroke and vascular neurology. [Online]. 1 (4). p.pp. 161–164. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28959479.

**More efficient**
care……………….

Pepgra
Drives Innovation

# Contact Us

## India

#10, Kutty Street
Nungambakkam,
Chennai, TN.

**+91-8754446690**

## UK

The Portergate
Ecclesall Road
Sheffield,
S11 8NX

**+44-1143520021**

## US

The Portergate
Ecclesall Road
Sheffield,
S11 8NX

**+1-972-502-9262**

**info@pepgra.com**

**www.pepgra.com**