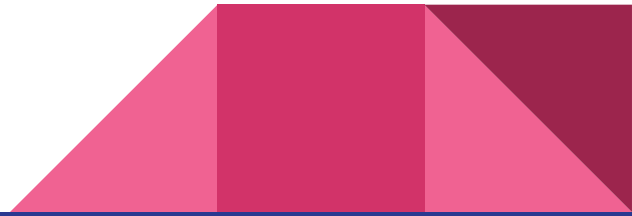
Big Data Applications in Health

Suleyman Akbas

Overview

1. Papers
2. Introduction
3. Big Data in Health
4. Potential Use Areas
5. Current Applications
6. Privacy Issues
7. Discussion

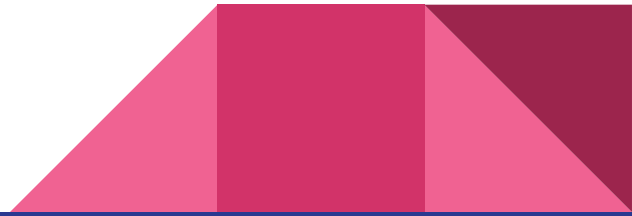


Papers

1. Big Data for Health; Javier Andreu-Perez et.al, IEEE Journal of Biomedical and Health Informatics, 2015
2. Big data analytics in healthcare: promise and potential; Raghupathi et al., Health Information Science and Systems, 2014

Introduction

U.S. Congress in August 2012 defines big data as “large volumes of high velocity, complex, and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management and analysis of the information”

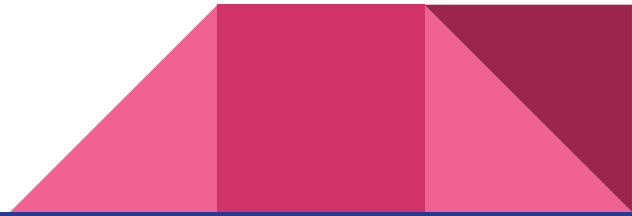


Big Data in Health

Health informatics applications are known to generate datasets that are complicated to store, resolve, organize, process, and, above all, interpret.

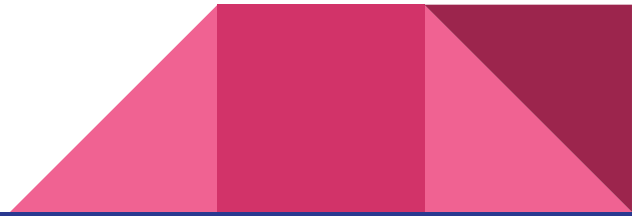
Prevention, early intervention, optimal management, reducing costs.

Explore the trends and propose predictive models.



Large volume of structured and unstructured data to analyze in order to conduct large scale population-based studies.

Healthcare spending in the UK by 2021 will make 6.4% of the gross domestic product (GDP) whilst the total projected healthcare share of the GDP in the United States is expected to reach 19.9% by 2022.

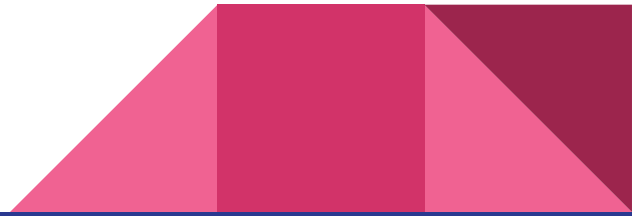


How Big Data Applies to Health (6 Vs)

Volume

US healthcare system alone already reached 150 exabytes five years ago. Before long, we will be dealing with zettabyte and yottabyte data for countries with large populations such as China and India.

Daunting volume of existing healthcare data includes personal medical records, radiology images, clinical trial data, human genetics and population data, genomic sequences, etc. Newer forms of big data, such as 3D imaging, genomics and biometric sensor readings, are also fueling this exponential growth.



Velocity

This data is accumulated in real-time and at a rapid pace.

With the versatility, diversity, and connectivity of data capturing devices, additional data is generated at increasingly high speed.

Regular monitoring, such as multiple daily diabetic glucose measurements, blood pressure readings, and EKGs (Electrocardiography).

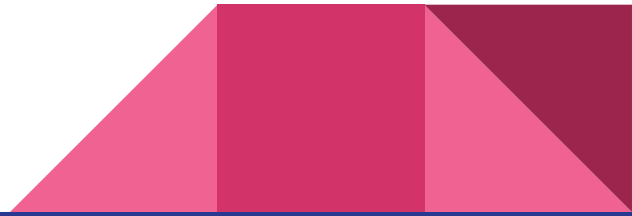
Decision support must be made available near real time.

In many medical situations, constant real-time data can mean the difference between life and death.

Variety

Health data includes quantitative (e.g., sensor data, images, gene arrays, laboratory tests) or qualitative (e.g., free text, demographics) data.

Increasingly, the data is in multimedia format and unstructured.



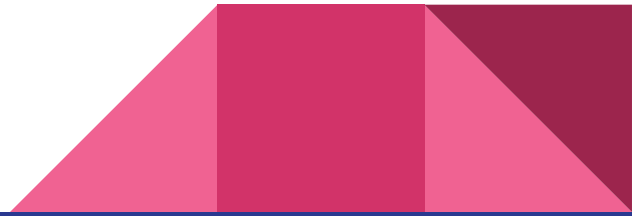
Veracity

Trustworthiness of the data obtained.

Personal health records may contain typographical errors, abbreviations, and cryptic notes.

Portable measurements are sometimes taken within less reliable, uncontrolled environments compared to clinical data.

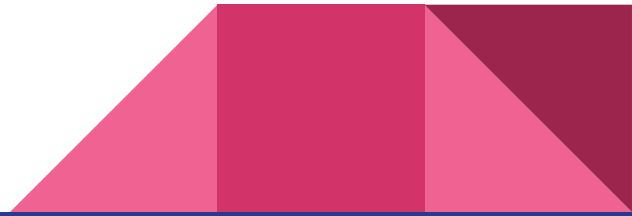
Life or death decisions depend on having the accurate information, and the quality of healthcare data.



Value

Clinical value of data.

Resting blood pressure measurements are more valuable.



Variability

Consistency of data over time.

Seasonal effects on health creates variability.

Value		Clinically relevant data Longitudinal studies
Volume		High-throughput technologies Continuous monitoring of vital signs
Velocity		High-speed processing for fast clinical decision support Increasing data generation rate by the health infrastructure
Variety		Heterogeneous and unstructured data sources Differences in frequencies and taxonomies
Veracity		Data quality is unreliable Data coming from uncontrolled environments
Variability		Seasonal health effects and disease evolution Non-deterministic models of illness and health

Potential Use Areas of Big Data

Electronic Health Records (EHRs)

Mining EHRs is important to improve clinical knowledge and support research.

This data can be used for disease management and predicting health risks.

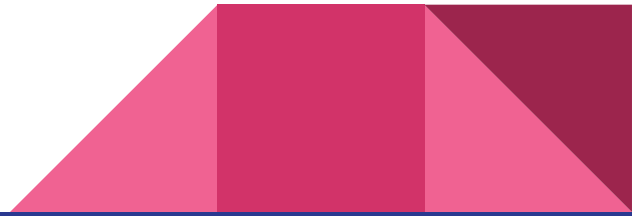
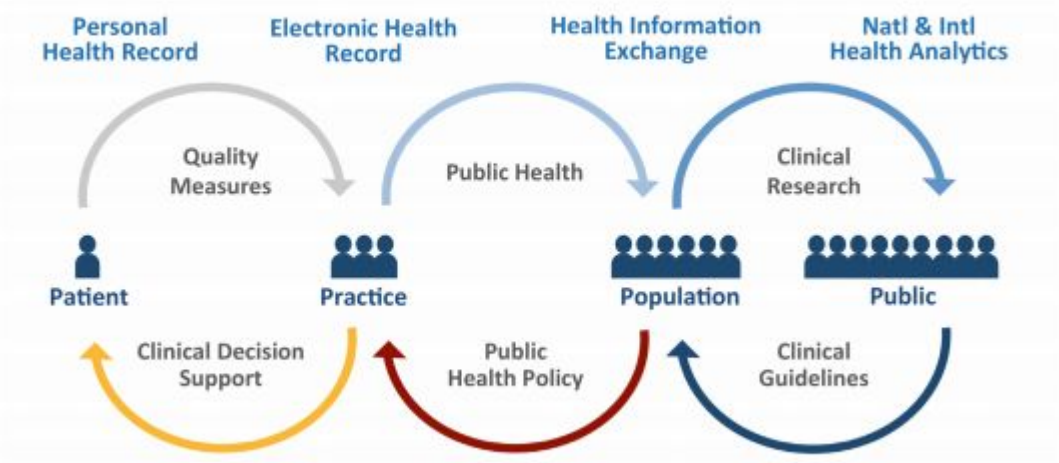


Figure 1. Vision of the Health IT Ecosystem



Source: Office of the National Coordinator

Social Health

Connecting doctors and patients beyond the clinic.

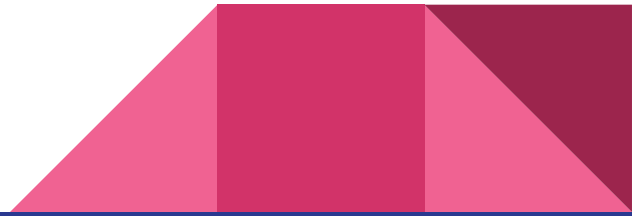
This can be used for disease outbreaks, collective dynamics of smoking, and monitoring mental health.

Correlation between suicide rates and emotional tweets.

Lifestyle Environmental Factors, and Public Health

Climate observance

Monitoring impact of air pollution on human health



Genetics

Use of technology is increased a lot after first human genome mapping.

A single whole human genome sequencing is 3-200 GB depending on the depth of coverage.

The increasing amount of biomedical and genomic data are analyzed to formulate knowledge and medical tools.

Drug Response on Genetic Variations

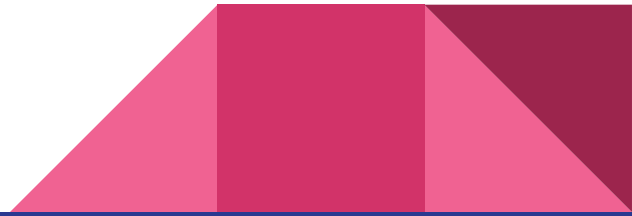
Variations to drug response is studied due to genetic differences to improve medicine precision.

Cancer medicine drugs have different responses even for the same type of cancer.

Gene Networks

Gene networks of different syndromes of the same person are studied to understand how they are interrelated.

Chromosome 21 is studied to understand the relationship between Down’s Syndrome and cardiac abnormalities.



Recent Large-Scale Database Launches

Several large-scale biological databases are launched to facilitate studying disease mechanisms and progressions by research collaborations.

Mostly 3-D structures of molecules, gene, and drug databases.

These are important to understand complex diseases like cancer.

Pattern mining and clustering are used to discover knowledge.

Wearable, Implantable, and Ambient Sensors

They are important to monitor chronic diseases, capture critical events, and stream a continuous health information.

For diabetic patients, these sensors monitor glucose level.

They generate datasets which are currently beyond our capabilities to easily organize and interpret.

From Sensor Data to Patient Management

Not all the sensor data is clinically relevant.

Organizing and interpreting this data is important for early intervention.

Automatic classifications are necessary because manuel interpretation within reasonable time is not possible with this amount of data.

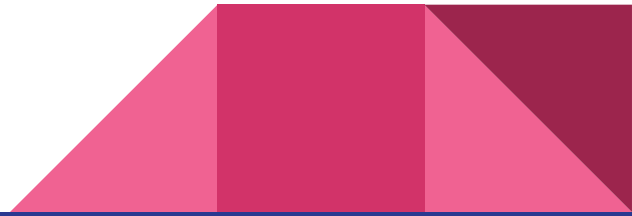
New algorithms needed to reduce false alarms.

Mobile Health

1.75 billion smartphones are in use today.

They are important for personal logging of health information.

Collecting and analyzing heart rate variabilities.

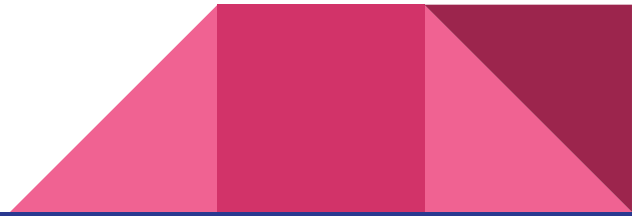


Medical Imaging

Ever-increasing amount of real time medical imaging data.

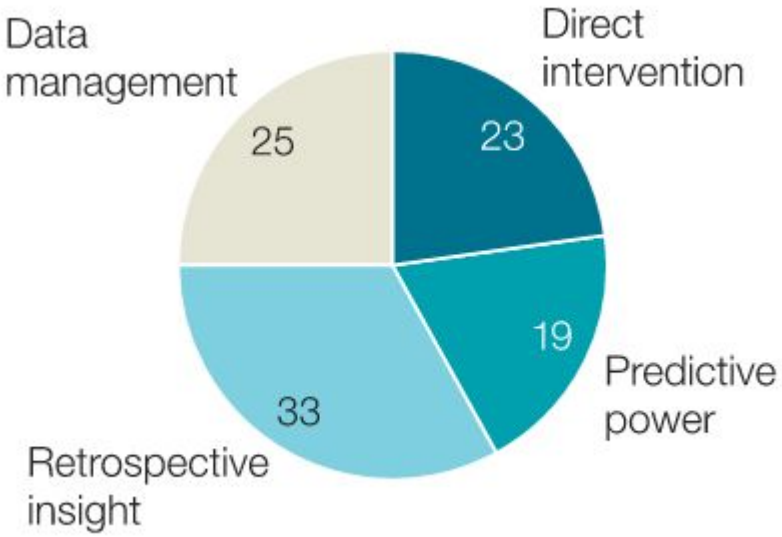
Getting faster, higher resolution, more versatile.

Hard to organize, mine, and harvest knowledge.



Current Applications

U.S. Healthcare Applications 2010-12



US healthcare data apps by type of data/analytic capability, 2010-12, excludes ideas that didn't involve big data.

Premier

The U.S. healthcare alliance network

2.700 member hospitals

Data of 25% of the patients discharged from hospitals

Comprehensive and comparable clinical outcome measures, resource utilization reports and transaction level cost data.

Improved decision-making and healthcare processes at 330 hospitals

Saving an estimated 29,000 lives and reducing spendings by $7 billion.

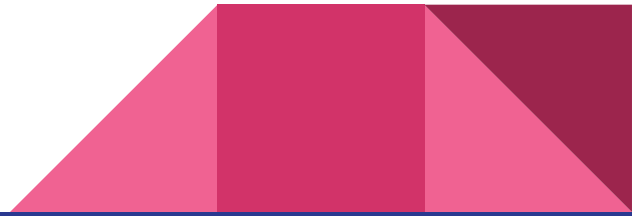
Columbia University

Columbia University Medical Center

Analyzing complex correlations of streams of data related to brain injuries.

To provide medical professionals with critical and timely information to aggressively treat complications.

Diagnoses serious complications in patients who have suffered a bleeding stroke as much as 48 hours sooner than previously used methods.



Project Artemis

Toronto’s Hospital for Sick Children, the University of Ontario Institute of Technology (UOIT) in Oshawa, Ont., and IBM Canada.

To detect life-threatening infections for newborns.

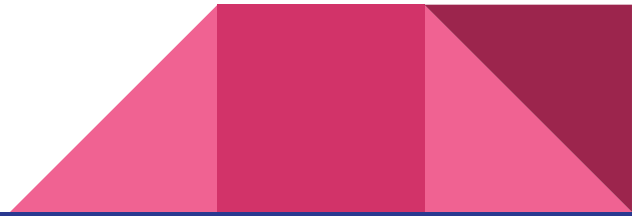
The hospital was able to predict the onset of nosocomial infections (a disease originating in a hospital) 24 hours before symptoms appeared.

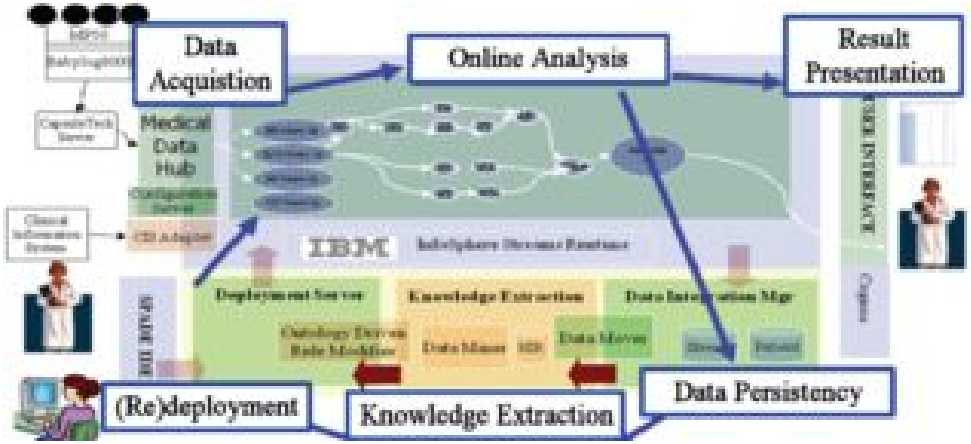
Processing the 1256 readings per second it currently receives per patient.

A note of a single reading every 30 or 60 minutes.

Changes in heart rate, breathing and blood pressure in the earliest stages of infection are flagged, allowing physicians to spot early warnings and administer antibiotics sooner.

Provide real-time analysis to help clinicians to predict more quickly potential negative changes in an infant’s condition.




Data Acquistion
Online Analysis
Result Presentation
Medical Data Hub
IBM
(Re)deployment
Knowledge Extraction
Data Persistency

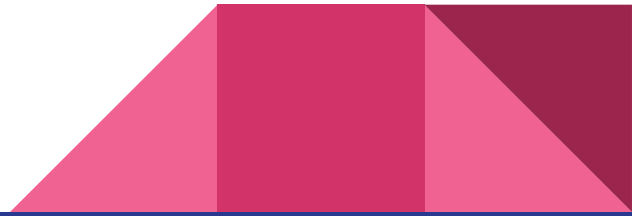

IBM InfoSphere BigInsights
Bring the power of Hadoop to the enterprise.

Data stored to allow doctors to look for common factors to make more accurate predictions about a baby’s condition.

The team is looking at studying the electrical activity of the brain.

Recently, went live with a 200-bed NICU at a hospital in Shanghai, China, a 100-bed neonatal unit in Shenzhen, China.

In the planning stages for another NICU in Canada, as well as one in Australia.



Rizzoli Institute

The Rizzoli Orthopedic Institute in Bologna, Italy

Using advanced analytics to gain more insights of diseases.

Reported to have reduced annual hospitalizations by 30% and the number of imaging tests by 60%.

In the long term, the Institute expects to gain insight into the role of genetic factors to develop treatments.

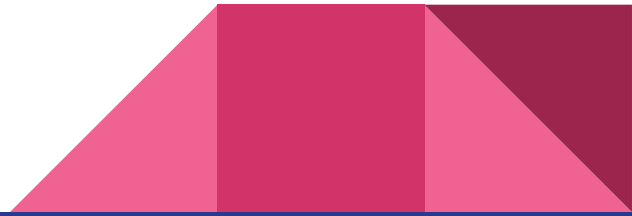
Children’s Cancer Care Initiatives

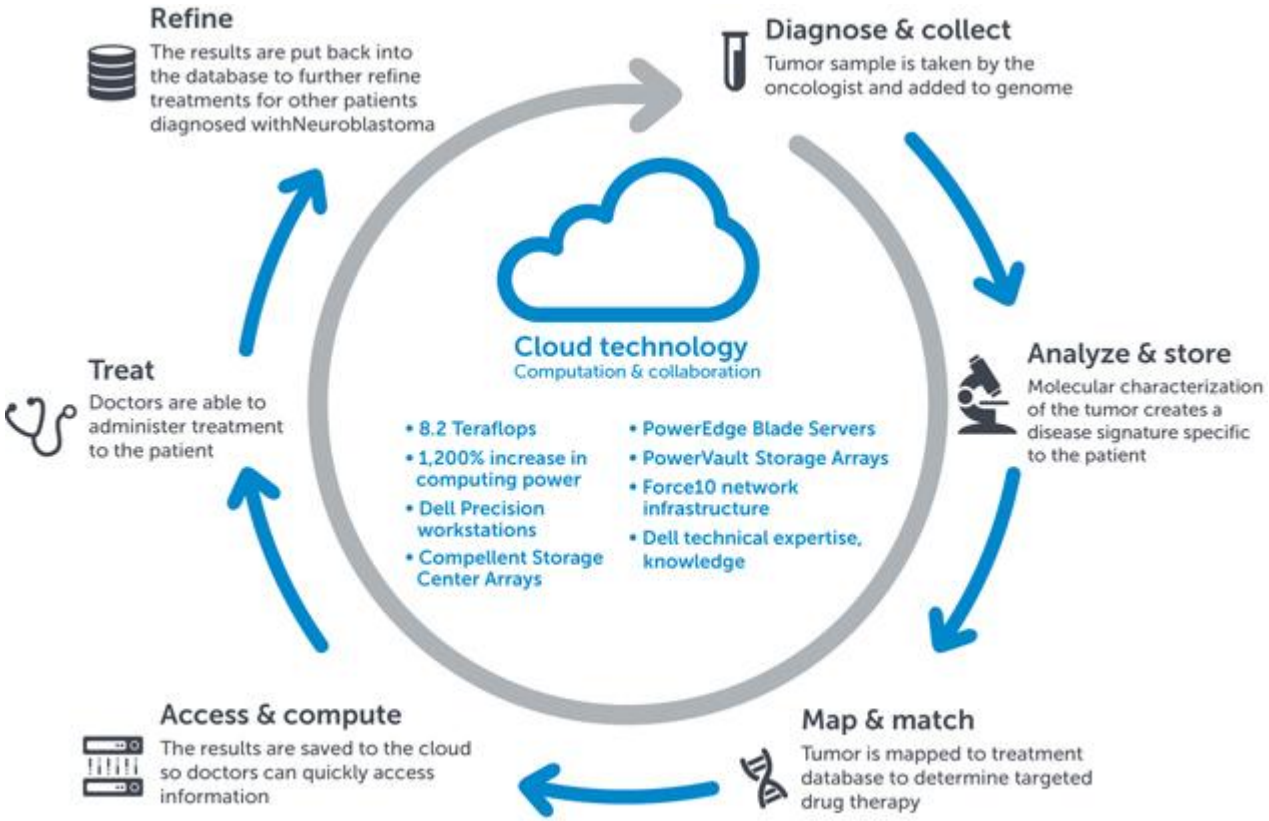
Initiatives by Dell and Intel

In 2011, they focus on children with neuroblastoma, one of the deadliest forms of cancer

Based on each child’s specific genetic profile

More than 27.000 children




Refine
The results are put back into the database to further refine treatments for other patients diagnosed withNeuroblastoma
Diagnose & collect
Tumor sample is taken by the oncologist and added to genome
Cloud technology
Computation & collaboration
• 8.2 Teraflops
• 1,200% increase in computing power
• Dell Precision workstations
• Compellent Storage Center Arrays
• PowerEdge Blade Servers
• PowerVault Storage Arrays
• Force10 network infrastructure
• Dell technical expertise, knowledge
Analyze & store
Molecular characterization of the tumor creates a disease signature specific to the patient
Treat
Doctors are able to administer treatment to the patient
Access & compute
The results are saved to the cloud so doctors can quickly access information
Map & match
Tumor is mapped to treatment database to determine targeted drug therapy

Results

Stopping the progression of cancer in 60% of patients

Achieving slowing down in patients previously deemed incurable

Increasing scientific understanding of cancer and genes

Gathering big data that can facilitate medical breakthroughs

Vioxx

The California-based Kaiser Permanente

Connected clinical and cost data of 1.4 million patients to provide a crucial dataset

Analytics of which led to the discovery of adverse drug effects Vioxx

50% more heart attacks

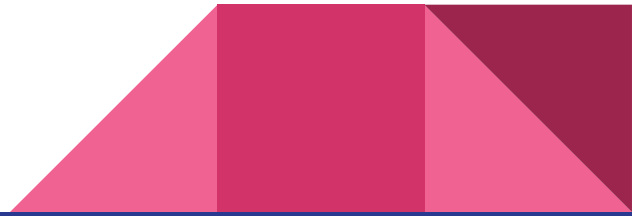
Withdrawal of from the market

Revenue of $2.5 billion

Privacy Issues

In the healthcare domain, data privacy have utmost importance as regulated by laws in most countries.

- Anonymization
- Cryptology
- Data Perturbation (adding noise)



Discussion

Applying big data to health is not only important to biological and physical sciences, but equivalently important to “soft” sciences, such as behavior and social sciences.

Human effect on global warming.

Questions?