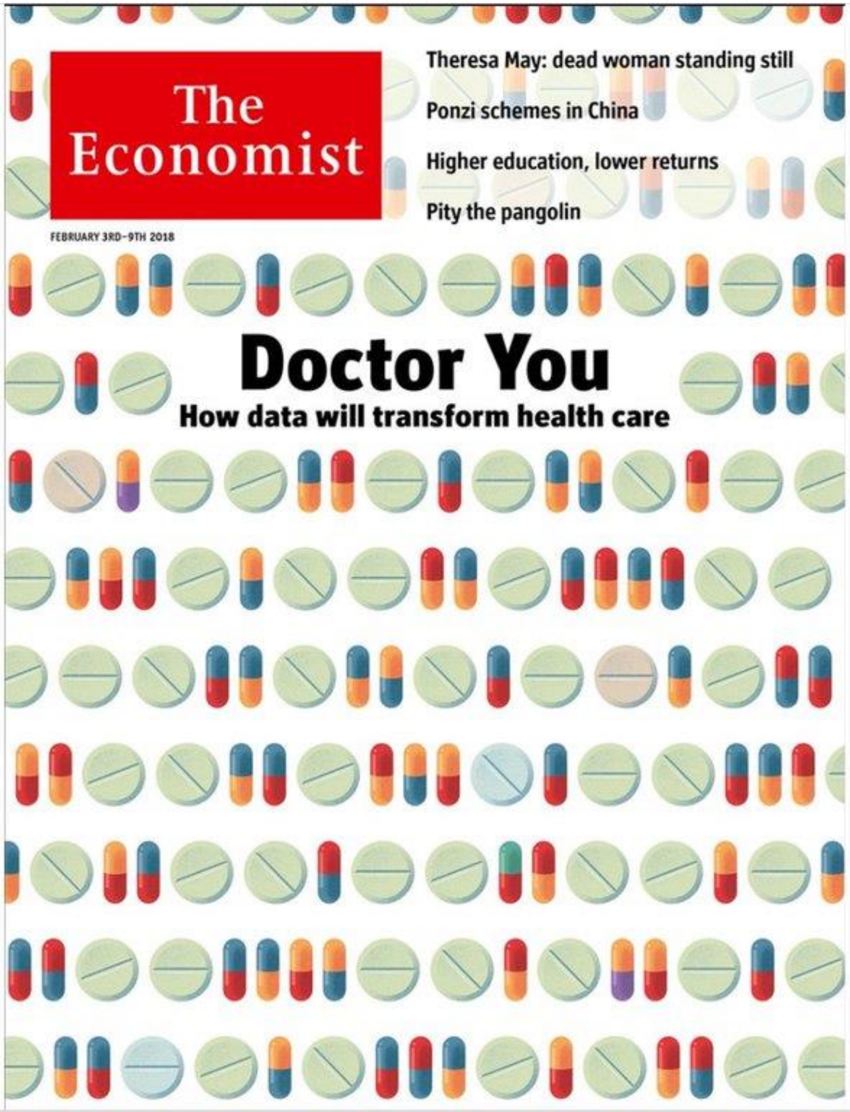
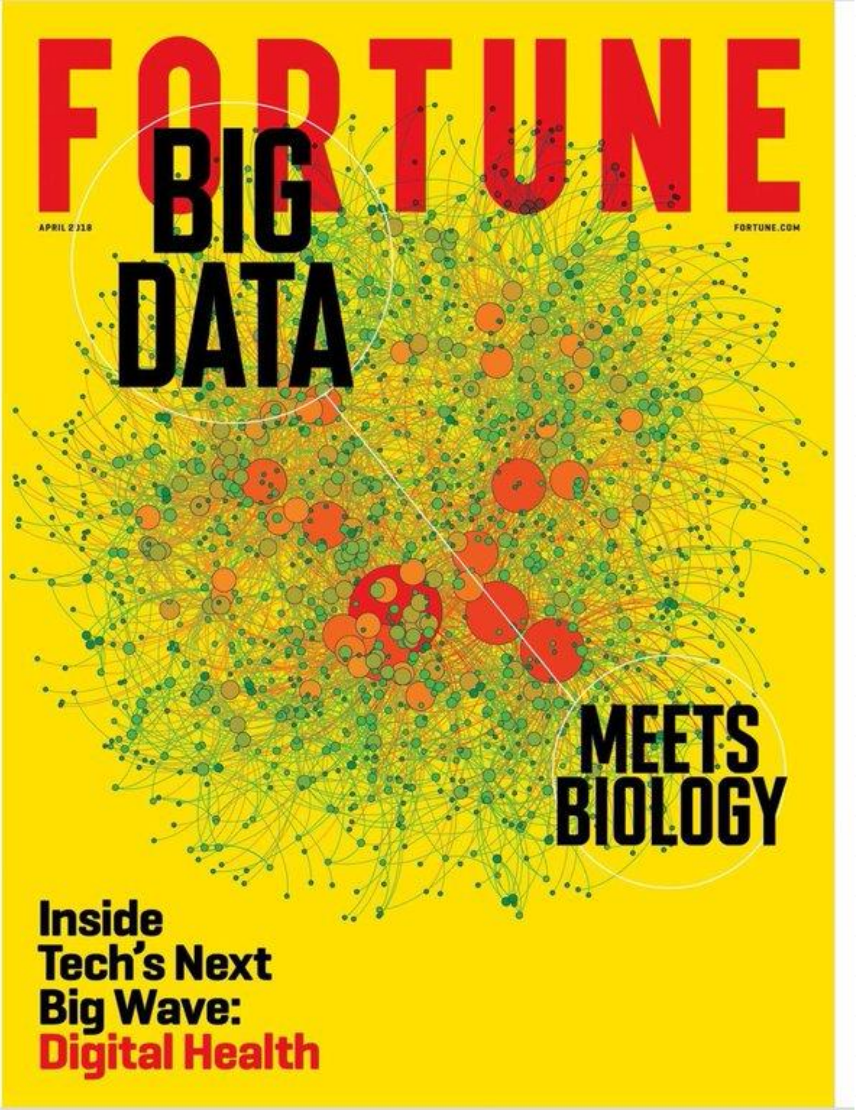
Lecture 1: Introduction to Biomedical Data Science

Course: Biomedical Data Science

Parisa Rashidi

Fall 2018


The Economist
Theresa May: dead woman standing still
Ponzi schemes in China
Higher education, lower returns
Pity the pangolin
FEBRUARY 3RD–9TH 2018
Doctor You
How data will transform health care


FORTUNE
APRIL 2018
FORTUNE.COM
BIG DATA
MEETS BIOLOGY
Inside Tech's Next Big Wave:
Digital Health

Disclaimer

The content, graphics and images in the lecture notes are partially based on:

- David Sontag, Machine Learning for Healthcare 6.S897, HST.S53, MIT, 2017
- Azizi, Palla, Belgrave, ICML Tutorial: Machine Learning for Personalised Health, 2018
- Yan Liu, Jimeng Sun, Deep Learning Models for Health Care - Challenges and Solutions, 2017
- O'Neil, Cathy; Schutt, Rachel. Doing Data Science: Straight Talk from the Frontline, 2016
- Shortliffe, Edward H.; Cimino, James J. (2013). Biomedical Informatics. Springer London.
- Mark Musen, Introduction to Big Data and the Data Lifecycle, The Big Data to Knowledge (BD2K), 2017
- Guide to the Fundamentals of Data Science Computing Overview, Patricia Kovatch, 2017

Agenda

- Introduction to biomedical data science
  - Paper - Thursday
  - What is biomedical data science
    - Why I should care about this
- Thursday
  - Python Programming
  - Paper Reading

Papers

- First paper is available on Canvas
- There are 9 groups with random assignments
  - 3 students per group
- Each Thursday one group will be presenting, starting with group 1
- 15 minutes presentation (background, methods, results)
  - Incorporate your own discussions
  - Each group member will have 5 minutes

Volume 29, Issue 2
February 2018

Article Contents

EDITOR'S CHOICE

Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board FREE

S P Somashekhar, M -J Sepúlveda, S Puglielli, A D Norden, E H Shortliffe, C Rohit Kumar, A Rauthan, N Arun Kumar, P Patil, K Rhee, ... Show more

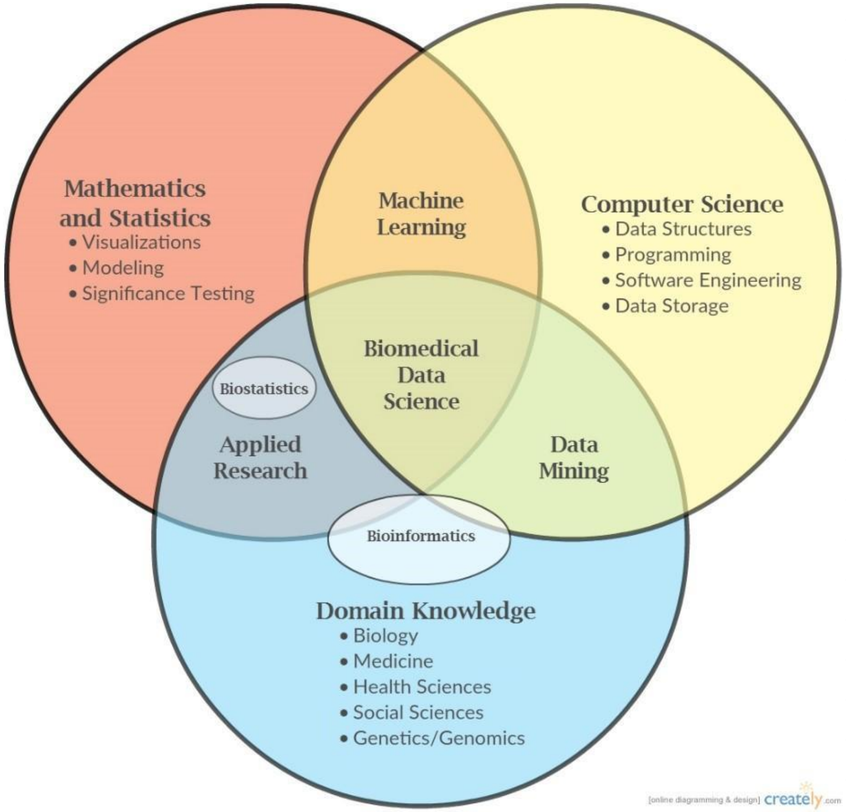
Annals of Oncology, Volume 29, Issue 2, 1 February 2018, Pages 418–423, https://doi.org/10.1093/annonc/mdx781
Published: 09 January 2018

Papers

- Everyone will have to submit a one-paragraph criticism or a question on canvas, to be discussed in class
- Reading grade:
  - Undergrads: 20% discussions -15% presentation
  - Graduate: 20% discussions – 15% presentation (including 10% survey)
- Please do not be afraid to criticize papers! That is part of our goal, to teach you critical thinking.

Biomedical Data Science

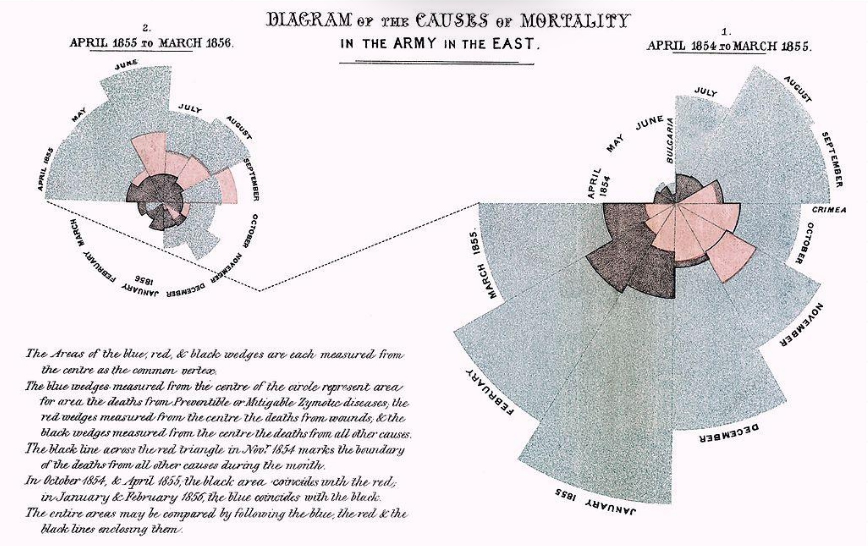
- Data science techniques applied to biomedical science problems



Dawn of Data-Driven Health:1850s

Florence Nightingale

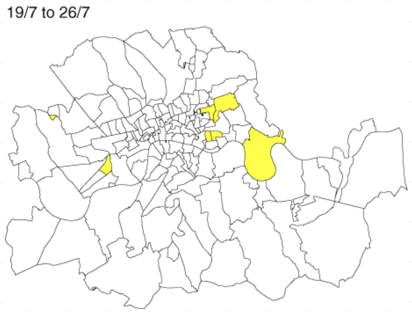
- Studying the causes of mortality in the army
- 16,000 to 18,000 army death due to preventable conditions

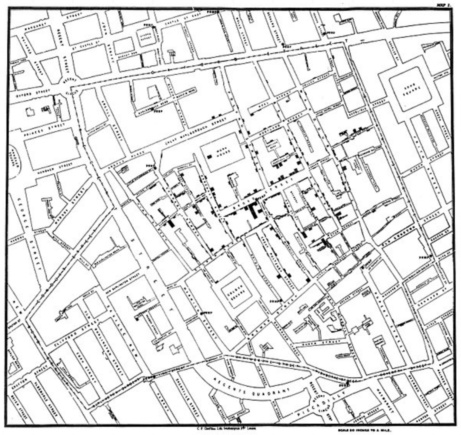


Dawn of Data-Driven Health:1850s

- Tracing the outbreak of Cholera in London in 1854
- Father of modern epidemiology

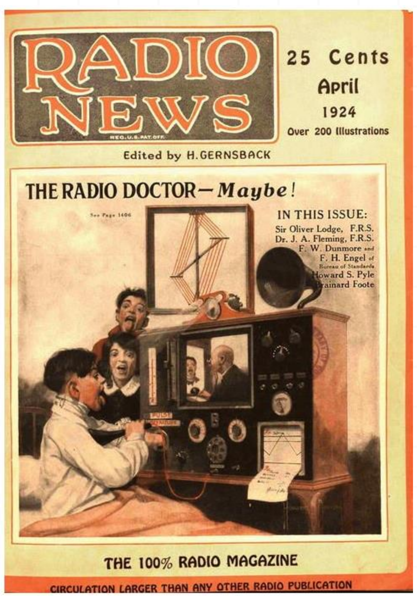
John Snow

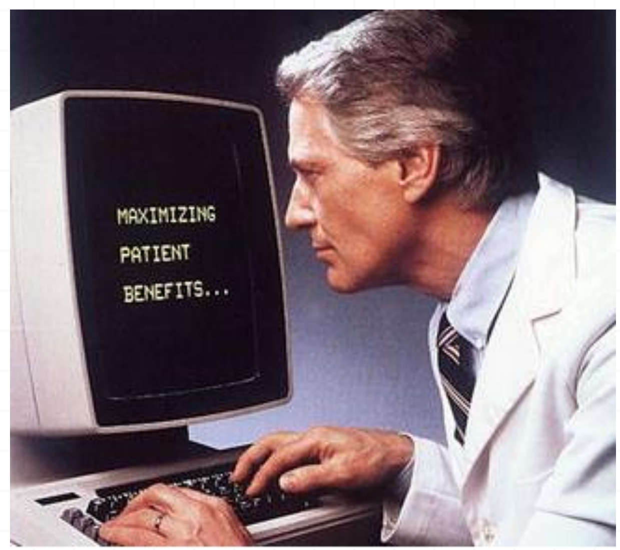




Early Clinical Information Systems: 1960s

- Hospital information systems or HIS (1960s)
  - Mostly a single, large, time-shared computer
  - Distributed HIS started to appear in 1980s





Doctor of the future (Early 1980s).

Expert Systems: 1970

- MYCIN Expert System developed at Stanford in 1972
  - For identifying blood infections based on reported symptoms and medical test results
  - Using about 500 production rules
  - Roughly the same level of competence as blood infection specialists and rather better than general practitioners

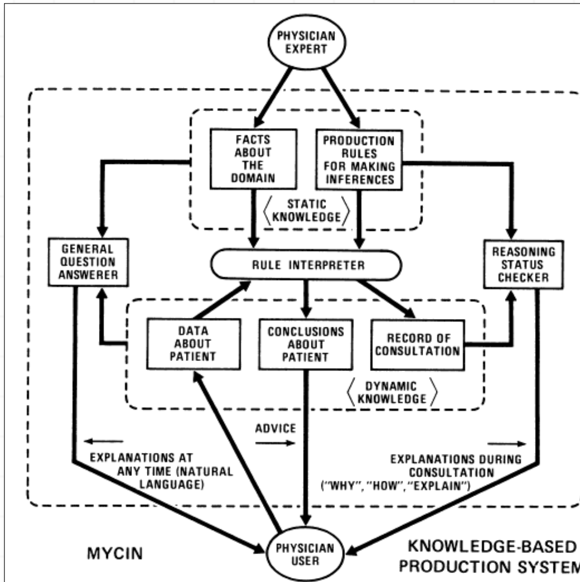


Figure 2 - Diagram summarizing the organization and flow of information within MYCIN. The correlation between this design and the human consultation process depicted in Fig. 1 is discussed in the text. (Figure reproduced from reference 10).

Dialogue interface

I am ready

** THIS IS A 26 YEAR OLD MALE PATIENT

My understanding is:
The age of the patient is 26
The sex of the patient is male

** FIVE DAYS AGO, HE HAD RESPIRATORY-TRACT SYMPTOMS

What is his name?

** JO

My understanding is:
The name of the patient is Jo
Respiratory-tract is one of the symptoms that the patient had

** A COUPLE OF DAYS BEFORE THE ADMISSION, HE HAD A MALAISE

Please give me the date of admission

** MARCH 12, 1979

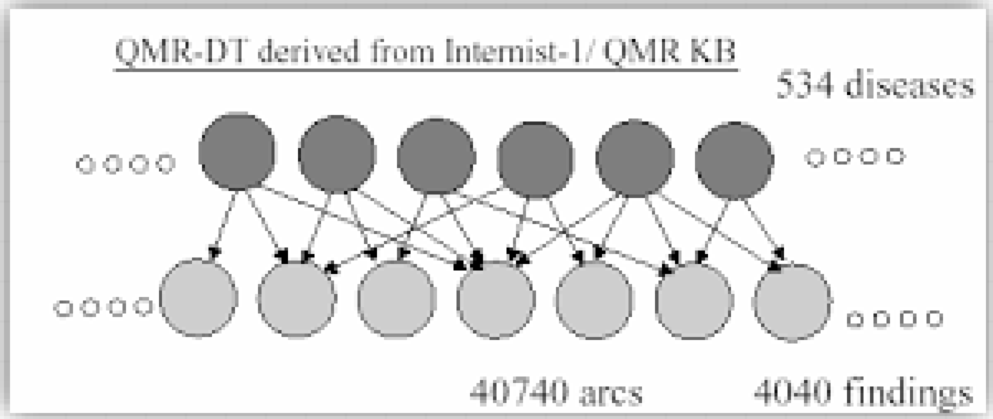
My understanding is:
The patient was admitted at the hospital 3 days ago
Malaise is one of the symptoms that the patient had 5 days ago

FIGURE 33-1 Short sample dialogue. The physician's inputs appear in capital letters after the double asterisks.

Shortliffe EH. Mycin: A Knowledge-Based Computer Program Applied to Infectious Diseases. Proceedings of the Annual Symposium on Computer Application in Medical Care. October 1977:66-69.

Probabilistic Models: 1980

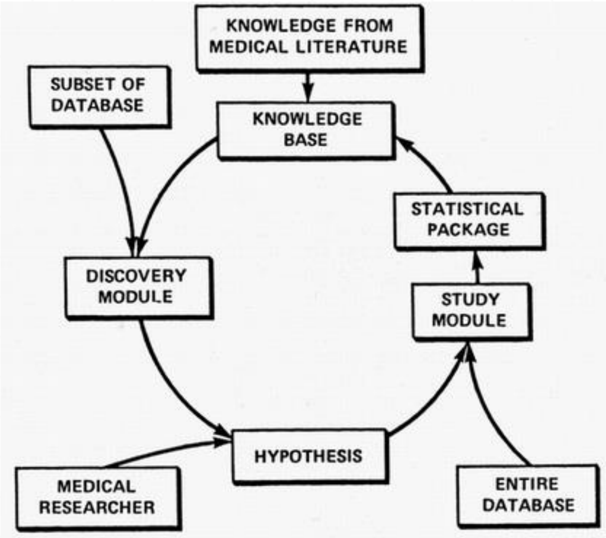
- INTERNIST/QMR was developed at University of Pittsburgh, 15 person-years of work
- A broad-based computer-assisted diagnostic tool
- Probabilistic model with 534 binary disease variables 4,040 binary symptom variables, 45,470 edges



Issues
- Manual Symptom entry by physicians
- Difficulty in maintenance and generalization

Data Mining: 1980

- The RX Project: discovering medical facts
- An early example of data mining under AI control
- Data from 50 severe Lupus patients
  - 52 attributes



Blum, R. L. (1982). Discovery, confirmation, and incorporation of causal relationships from a large time-oriented clinical data base: the RX project. Computers and Biomedical Research, 15(2), 164-187.

Neural Networks: 1990

- Neural networks in clinical applications started to appear in 1990

Small networks, poor generalization,

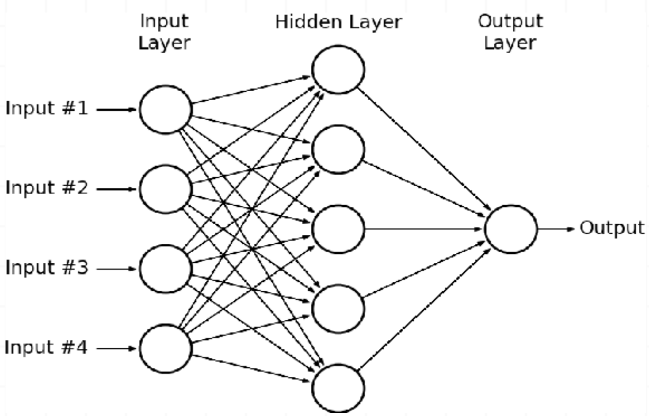


Table 1 • 25 Neural Network Studies in Medical Decision Making*

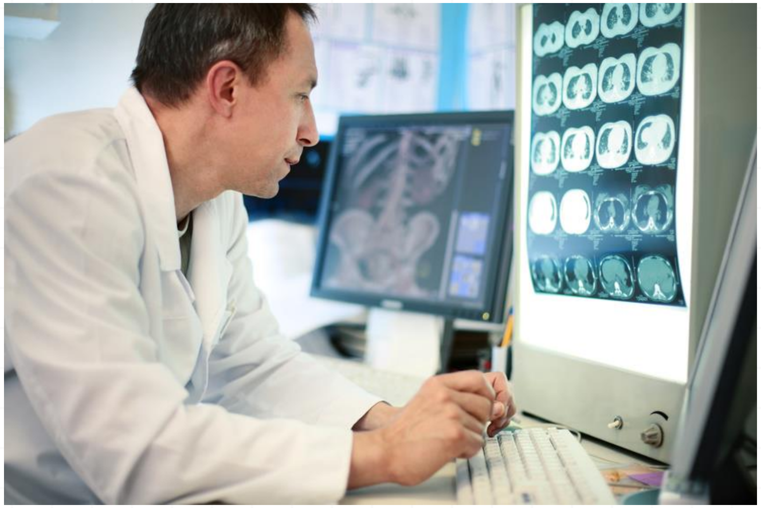
Subject	No. of Examples: Training	No. of Examples: Test	P†	Network	D‡	Accuracy§: Neural	Accuracy§: Other
Breast cancer[4]	57	20	60	9–15–2	0.6	80	75
Vasculitis[2]	404	403	73	8–5–1	8.0	94	—
Myocardial infarction[6]	351	331	89	20–10–10–1	1.1	97	**84**
Myocardial infarction[8]	356	350	87	20–10–10–1	1.1	97	**84**
Low back pain[11]	100	100	25	50–48–2	0.2	90	90
Cancer outcome[13]	5,169	3,102	—	54–40–1	1.4	0.779	0.776
Psychiatric length of stay[17]	957	106	73	48–400–4	0.2	74	76
Intensive care outcome[23]	284	138	91	27–18–1	0.5	0.82	**0.82**
Skin tumor[21]	150	100	80	18	—	80	**90**
Evoked potentials[35]	100	67	52	14–4–3	3.8	77	**77**
Head injury[47]	500	500	50	6–3–3	20	66	**77**
Psychiatric outcome[54]	289	**92**	**60**	41–10–1	0.7	79	—
Tumor classification[55]	53	**6**	**38**	8–9–3	1.4	99	**88**
Dementia[57]	75	18	19	80–10–7–7	0.6	61	—
Pulmonary embolism[59]	607	606	69	50–4–1	2.9	0.82	**0.83**
Heart disease[62]	460	230	54	35–16–8–2	3	**83**	**84**
Thyroid function[62]	3,600	1,800	93	21–16–8–3	22	**98**	**93**
Breast cancer[62]	350	175	66	9–4–4–2	**10**	**97**	**96**
Diabetes[62]	384	192	65	8–4–4–2	**12**	**77**	**75**
Mycardial infarction[63]	2,856	1,429	56	291–1	9.8	85	—
Hepatitis[65]	39	42	38	4–4–3	3.3	74	**79**
Psychiatric admission[76]	319	339	85	53–1–1	6.0	91	—
Cardiac length of stay[83]	713	696	73	15–12–1	3.5	0.70	—
Anti-cancer agents[89]	127	141	25	60–7–6	1.5	91	**86**
Ovarian cancer[91]	75	98	—	6–6–2	2.6	84	**81**
MEDIAN VALUE	350	175	71	20	2.8		

Penny, Will, and David Frost. "Neural networks in clinical medicine." Medical Decision Making 16.4 (1996): 386-398.

Why Biomedical Data Science?

Current Healthcare Themes (1): Electronic Health Records (EHR)

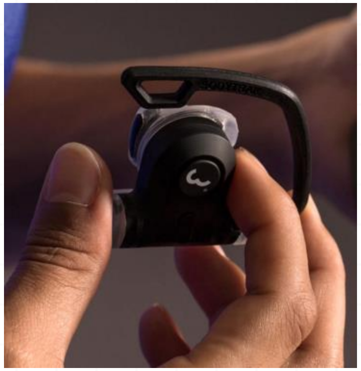
- **Moving beyond paper-based records**
  - Regional and national health data integration facilitating
    - Patient care
    - Administrative & financial
    - Research
    - Scholarly information
    - Office automation



Current Healthcare Themes (2): Smart and Connect Health

- Smart and connected health (Quantified Self)

CES 2017: Polar: embedded with heart rate-monitoring sensors, a motion-tracking sensor to track speed, distance and acceleration

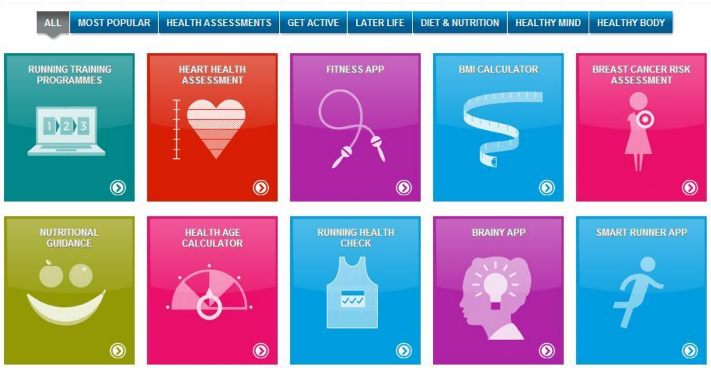


CES 2017: Bodytrak: a pair of earbuds equipped with an in-ear thermometer to measure core body temperature

CES 2017: Flow, a smart air quality tracker

Current Healthcare Themes (3): Information Access

- **Information access**
  - Health-related searches among the most popular
    - Most Web information are anecdotal



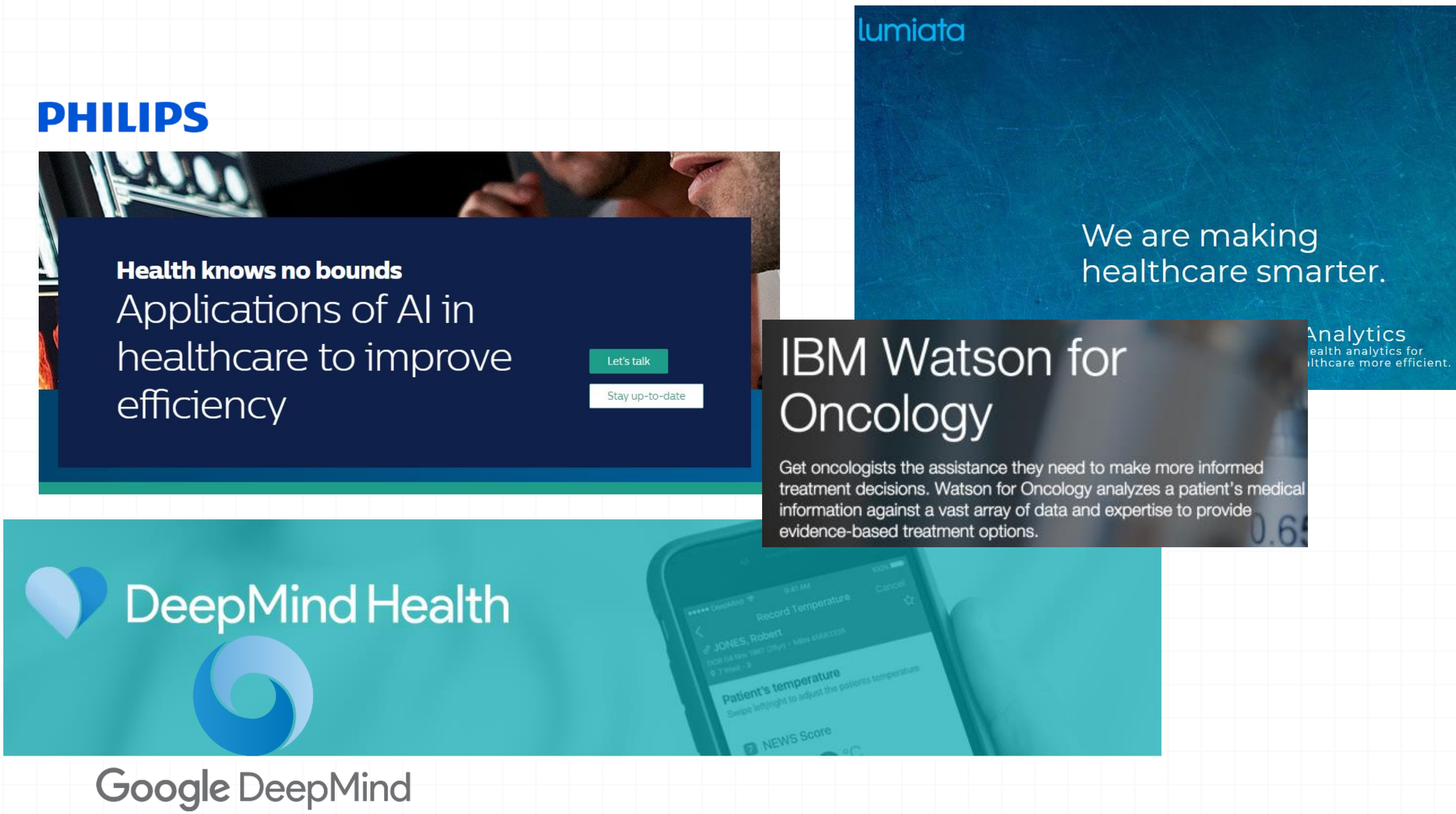
http://wholeheartedlyhealthy.com/2012/06/helping-you-find-healthy-2.html

Current Healthcare Themes (4): Omics, Imaging, …

- Complex heterogeneous types of data

Link

Industry & Healthcare



Healthcare Costs

- In 2017, reaching $3.5 trillion or $10K+ per person

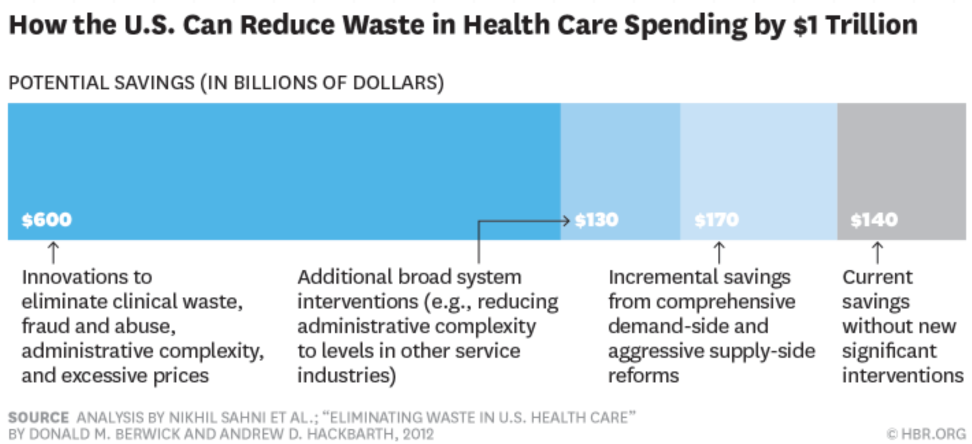
Top 10 most valuable companies combined

Net worth of

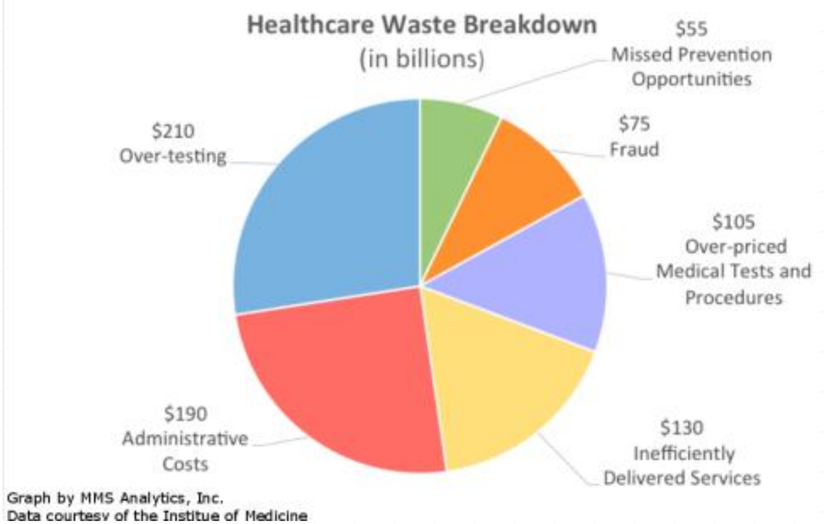
$72 billion $58 billion

Yan Liu, Jimeng Sun, Deep Learning Models for Health Care - Challenges and Solutions, 2017

Healthcare is Broken



$765 billions = 50 years budget



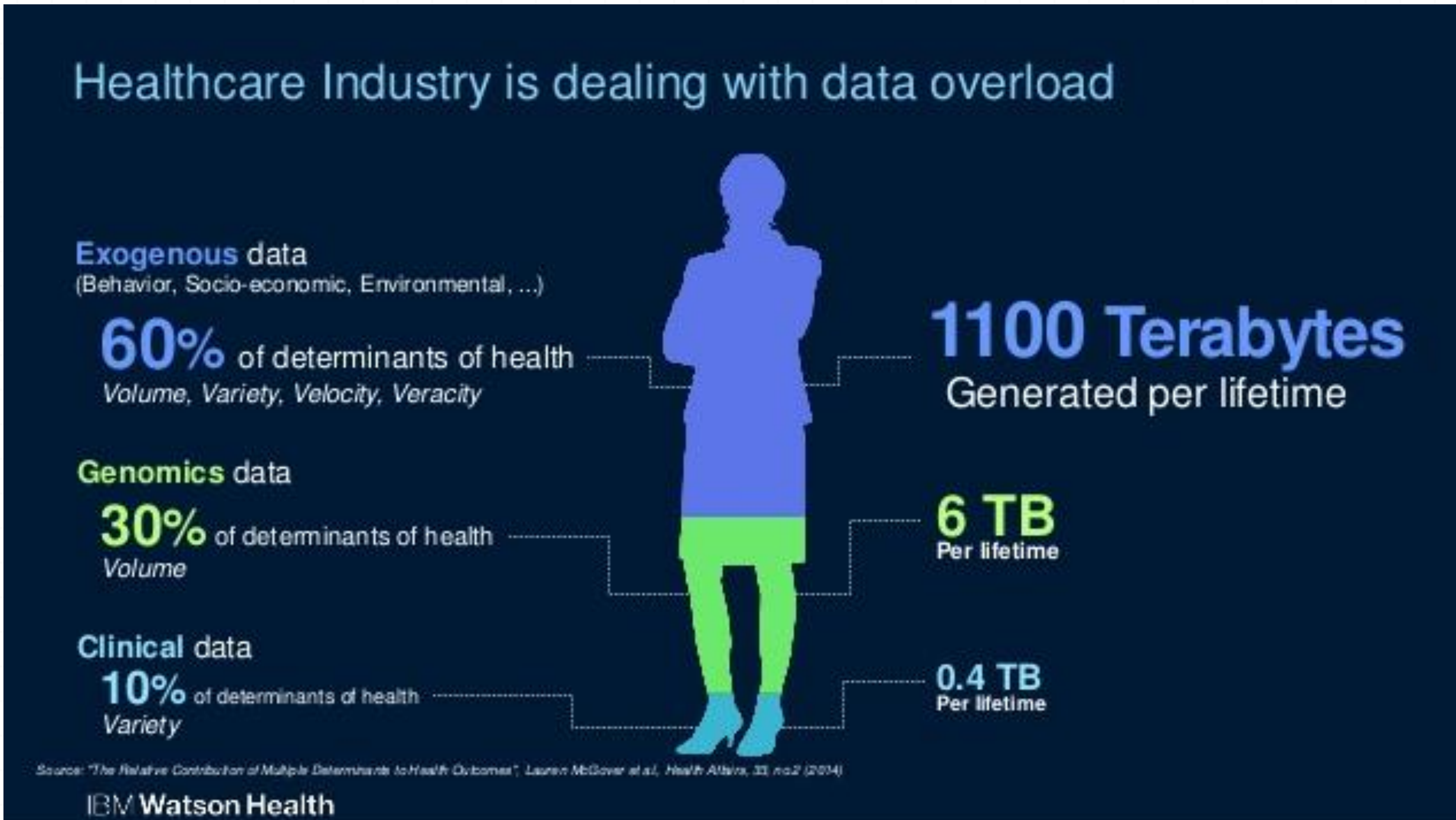
Yan Liu, Jimeng Sun, Deep Learning Models for Health Care - Challenges and Solutions, 2017

Data Revolution

- All the data processing we did in the last 2 years is more than all the data processing we did in the last three thousand years
- We are now being exposed to as much information in a single day as our 15th century ancestors were exposed to in their entire lifetime
- Every two days the human race is now generating as much data as were generated from the dawn of humanity through the year 2003

http://pressroom.pbs.org/Programs/h/The-Human-Face-of-Big-Data.aspx7

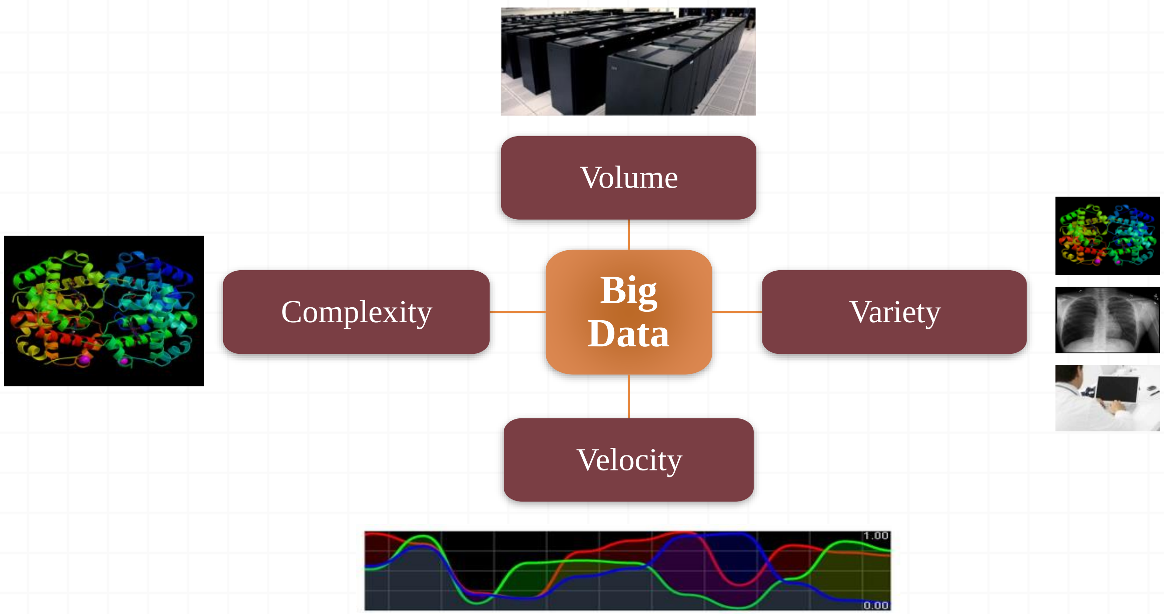
Big Data



Healthcare Industry is dealing with data overload
Exogenous data
(Behavior, Socio-economic, Environmental, ...)
60% of determinants of health
Volume, Variety, Velocity, Veracity
Genomics data
30% of determinants of health
Volume
Clinical data
10% of determinants of health
Variety
1100 Terabytes
Generated per lifetime
6 TB
Per lifetime
0.4 TB
Per lifetime
IBM Watson Health

What is Big Data Really?

- Big data has many aspects besides volume



What makes healthcare different?

- Life or death decisions
  - Need **robust** algorithms
  - Checks and balances built into ML deployment
  - (Also arises in other applications of AI such as autonomous driving)
  - Need **fair** and **accountable** algorithms
- Many questions are about unsupervised learning
  - Discovering disease subtypes, or answering question such as “characterize the types of people that are highly likely to be readmitted to the hospital”?
- Many of the questions we want to answer are *causal*
  - Naïve use of supervised machine learning is insufficient

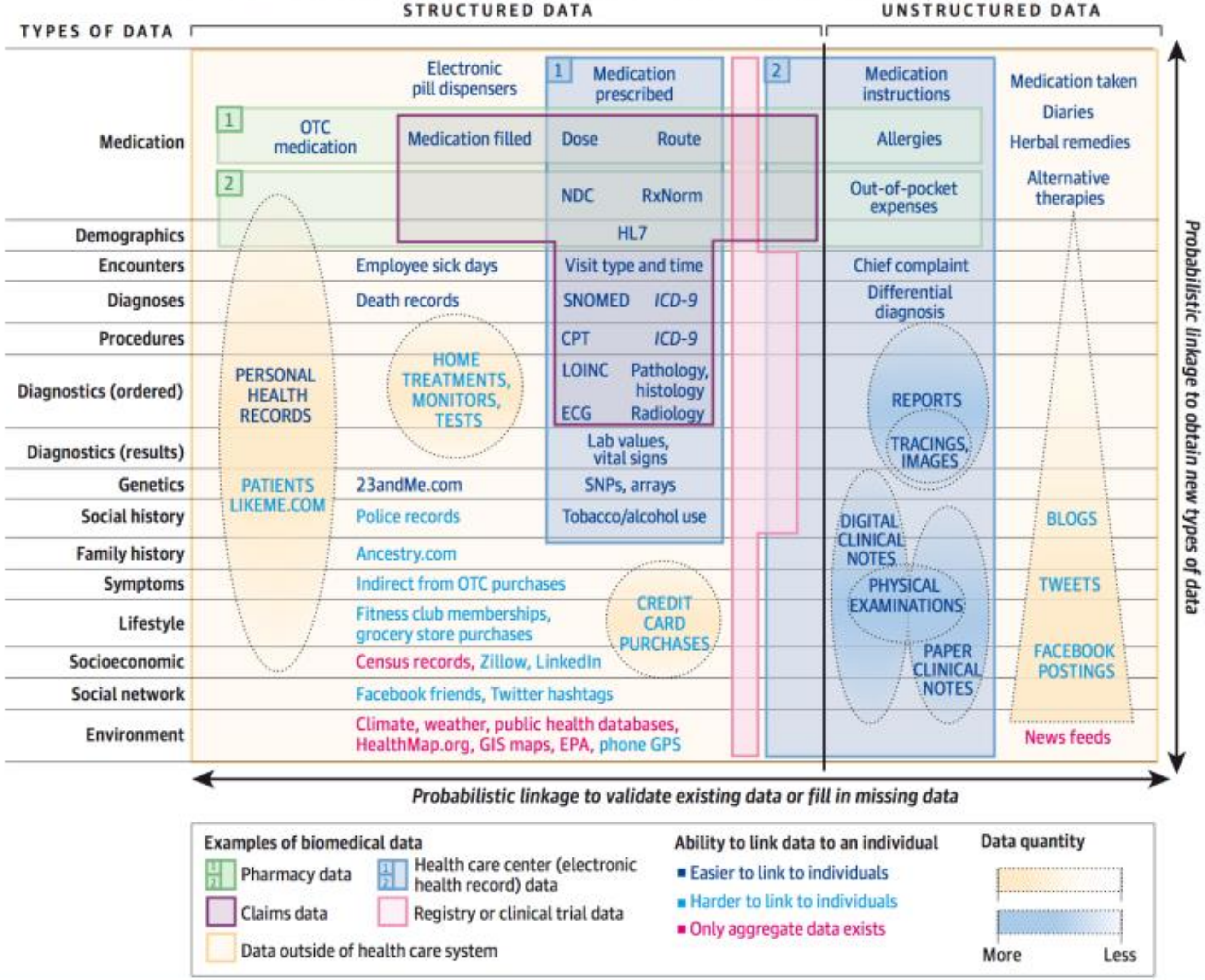
What makes healthcare different?

- Often very little labeled data (e.g., for clinical NLP)
  - Motivates semi-supervised learning algorithms
- Sometimes small numbers of samples (e.g., a rare disease)
  - Learn as much as possible from other data (e.g. healthy patients)
  - Model the problem carefully
- Lots of missing data, varying time intervals, censored labels

What makes healthcare different?

- Difficulty of de-identifying data
  - Need for data sharing agreements and sensitivity
- Difficulty of deploying ML
  - Commercial electronic health record software is difficult to modify
  - Data is often in silos; everyone recognizes need for interoperability, but slow progress
  - Careful testing and iteration is needed

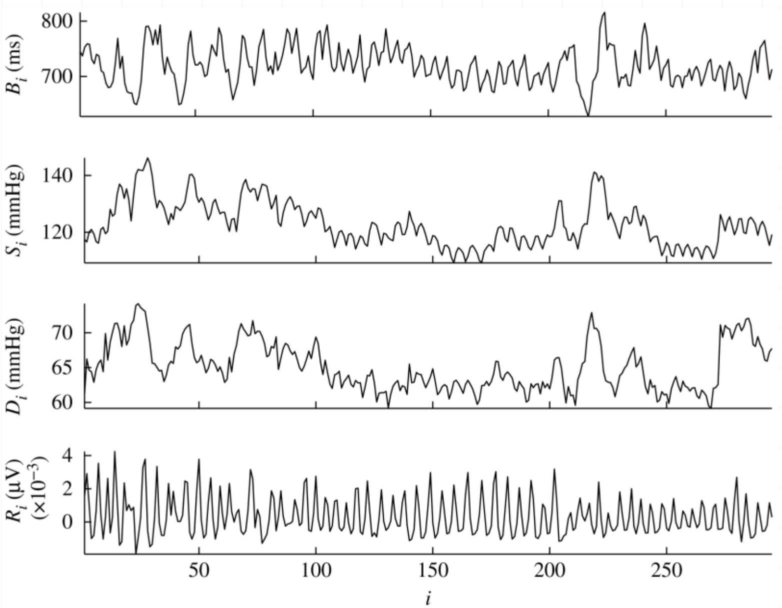
Healthcare data sources



Weber, Griffin M., Kenneth D. Mandl, and Isaac S. Kohane. 2014. "Finding the Missing Link for Big Biomedical Data." *JAMA: The Journal of the American Medical Association* 311 (24): 2479–80.

Time Series Data

- Time series data everywhere
  - Heart rate, blood pressure, EEG, accelerometer data …



Clinical Data

- Patient (ID)
- Parameter being observed (e.g. liver size, urine sugar,…)
- Value of the parameter
- Time of observation
- Method observation (self-report, lab results, …)

Patient No.	Last name	First name	Sex	Date of birth	Ward No.
454	Smith	John	M	14.08.58	6
223	Jones	Peter	M	07.12.65	8
597	Brown	Brenda	F	17.06.61	3
234	Jenkins	Alan	M	29.01.67	7
244	Wells	Christopher	M	25.02.55	6

Ward No.	Ward name	Type	No. of Beds
3	Carey	Medical	8
6	Bracken	Medical	16
7	Brent	Surgical	12
8	Meavy	Surgical	10

*http://www.technologyuk.net/computing/sad/relational_data_analysis.shtml

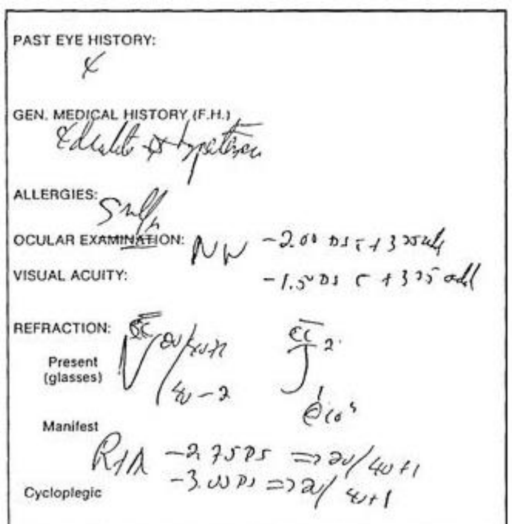
Complications with Clinical Data

- Frequency of data recording
  - Annual checkups versus continuous measurements of mean arterial blood pressure in cardiogenic shock
- Circumstances (context)
  - Was blood pressure taken in the leg or arm? Standing or sitting? What kind of device?
- Uncertainty
  - A radiologist looking at a shadow on a chest X-ray film is not sure whether it represents overlapping blood vessels or a lung tumor.
  - A confused patient is able to respond to simple questions about his or her illness, but under the circumstances the physician is uncertain how much of the history being reported is reliable.

Narrative Clinical Data

- Lots of data is in narrative form
  - Patient description of illness, responses to physician’s questions, physician's evaluations, pathologic examination, surgical procedures, …

An ophthalmologist’s notes



PAST EYE HISTORY:

GEN. MEDICAL HISTORY (F.H.)

ALLERGIES:

OCULAR EXAMINATION:

VISUAL ACUITY:

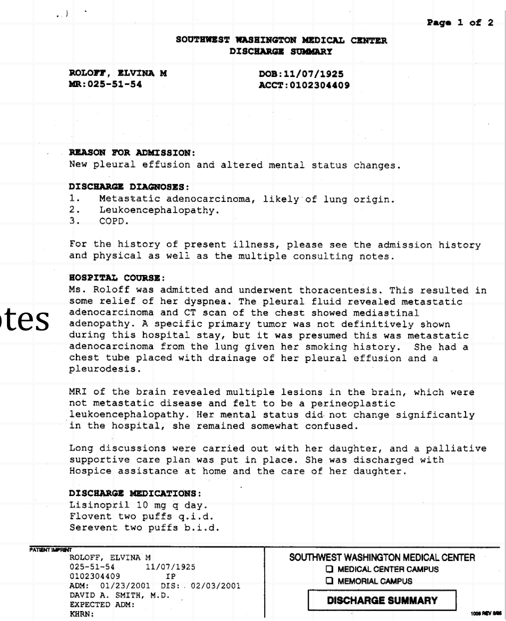
REFRACTION:

Present (glasses)

Manifest

Cycloplegic

Patient Discharge Notes



Page 1 of 2

SOUTHWEST WASHINGTON MEDICAL CENTER
DISCHARGE SUMMARY

ROLOFF, ELVINA M DOB:11/07/1925
MR:025-51-54 ACCT:0102304409

REASON FOR ADMISSION:
New pleural effusion and altered mental status changes.

DISCHARGE DIAGNOSES:
1. Metastatic adenocarcinoma, likely of lung origin.
2. Leukoencephalopathy.
3. COPD.

For the history of present illness, please see the admission history and physical as well as the multiple consulting notes.

HOSPITAL COURSE:
Ms. Roloff was admitted and underwent thoracentesis. This resulted in some relief of her dyspnea. The pleural fluid revealed metastatic adenocarcinoma and CT scan of the chest showed mediastinal adenopathy. A specific primary tumor was not definitively shown during this hospital stay, but it was presumed this was metastatic adenocarcinoma from the lung given her smoking history. She had a chest tube placed with drainage of her pleural effusion and a pleurodesis.

MRI of the brain revealed multiple lesions in the brain, which were not metastatic disease and felt to be a perineoplastic leukoencephalopathy. Her mental status did not change significantly in the hospital, she remained somewhat confused.

Long discussions were carried out with her daughter, and a palliative supportive care plan was put in place. She was discharged with Hospice assistance at home and the care of her daughter.

DISCHARGE MEDICATIONS:
Lisinopril 10 mg q day.
Flovent two puffs q.i.d.
Serevent two puffs b.i.d.

PATIENT IMPRINT
ROLOFF, ELVINA M
025-51-54 11/07/1925
0102304409 IP
ADM: 01/23/2001 DIS: 02/03/2001
DAVID A. SMITH, M.D.
EXPECTED ADM:
KHRN:

SOUTHWEST WASHINGTON MEDICAL CENTER
☐ MEDICAL CENTER CAMPUS
☐ MEMORIAL CAMPUS

DISCHARGE SUMMARY

1006 REV 8/96

Sample Clinical Documents

Pt recently hospitalized 7/19/06 for chf exacerbation (diastolic dysfunction) 2nd to dietary and medicine noncompliance (salty foods , stopped her HCTZ) and continued to smoke. Pt diuresed and sent home on new lasix 60qam 40qpm regimen. Pt noticed steady decline in functional status during the last 3 weeks because of SOB. at baseline should sat 85% on ra , 95% on 6L02NC at rest and ambulation. (on home o2) but now , can't ambulate , sating 83-89% on 6l at rest. also notes pnd , orthopnea. Pt notes intermif ent chest pain on and off lasting 5 minutes not associated with exertion or any other cardiac sx. 8/15 dobuta mibi-> ischemia in d1 territory. 11/19 :echo-->ef 60% , Pa pressure 48 + RA. no valve dz. rv enlarged and hypokinetic. A/P: pump: decompesated CHF (diastolic dysfxn , cor pulmonale component) 2nd to diet/med non-compliance. uptitrate captopril , continue iv lasix 60 qd with goal net neg 2 liters , daily weights , strict Iand O. check cxray. Switched to po lasix 10/06 , back to lisinopril for d/c Fri. ischemia: has + mibi in past , but no further workup to d1 lesion. can't get ecasa 2nd to vWD. continue BB , will hold off on statin since not hyperlipidemic. rate:tele. …

Issues with Narrative Data

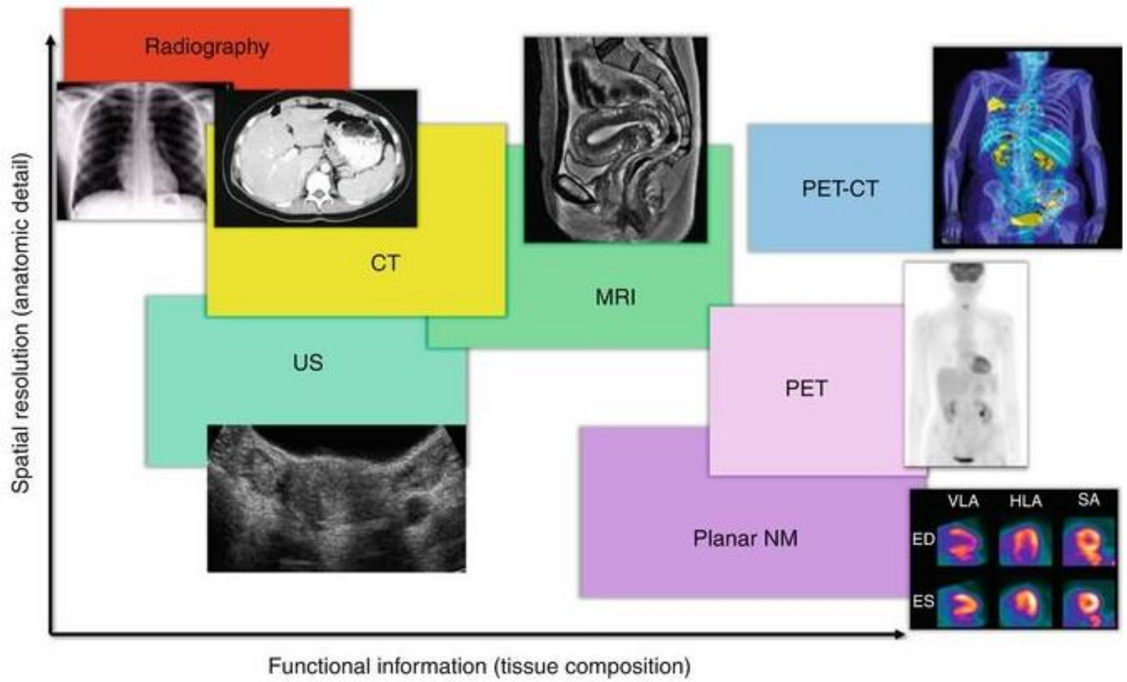
- Physician-specific
  - Each physician writes reflecting his/her thought process
    - Shortness of breath vs. dyspnea
- Domain-specific shorthand conventions
  - In eye examination: PERRLA ("Pupils are Equal in size, Round, and Reactive to Light and Accommodation)
- Context-specific
  - MI: Mitral Insufficiency (leakage in one of the heart's valves)
  - MI: Myocardial infarction (the medical term for what is commonly called a heart attack).
- Sentences changed into phrases
  - Pain relieved by antiacid

Genomic Data

- Human Genome Project started in 1990s
- An example of big data
  - How to integrate EHR and genomic data
  - Store the entire genome or just genetic markers?
  - Will lead to personalized medicine

Imaging Data

1. Structural or anatomic: size and shape of organs
2. Functional



Biomedical & Clinical Standards

The Need for Standards

- Excessive diversity in communication among many units
  - Primary physician
  - Specialists
  - Reimbursement & Insurance
  - …
- Pooling data together for research

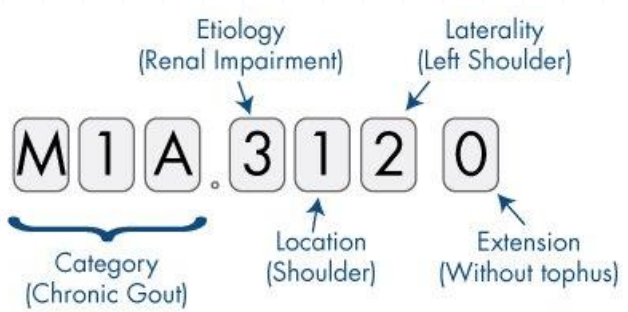
- Early interest in standards were driven by the need for exchanging data between clinical laboratories and clinical systems.

Some Important Standards

- Diagnosis: ICD-10
- Procedures: CPT
- Laboratory tests: LOINC
- Drugs: National Drug Codes (NDC), RxNorm
- Data Exchange: HL7
- Unified Medical Language System: UMLS

Examples of Standards

- There are many clinical standards



ICD-10 Example

Schema	Number of Codes	Examples
ICD-10 *(Diagnosis)*	68,000	- J9600: Acute respiratory failure - I509: Heart failure - I5020: Systolic heart failure
CPT *(Procedures)*	9,641	- 72146: MRI Thoracic Spine - 67810: Eyelid skin biopsy - 19301: Partial mastectomy
LOINC *(Laboratory)*	80,868	- 4024-6: Salicylate, Serum - 56478-1: Ethanol, Blood - 3414-0: Buprenorphine Screen
RxNorm *(Medications)*	116,075	- 161: Acetaminophen - 7052: Morphine - 1819: Buprenorphine

Shickel, Benjamin, Patrick James Tighe, Azra Bihorac, and Parisa Rashidi. "Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis." IEEE journal of biomedical and health informatics 22, no. 5 (2018): 1589-1604.

Patient Identifier

- Should include a check digit to ensure accuracy
- Mechanism for issuing such identifiers
  - National Provider Identifier by Center for Medicare and Medicaid Services (CMS)
    - 9 digits + check digit
  - Individual Identifier
    - Law passed in 1996, but pushed back by privacy advocates
    - US is among the few developed countries without such an identifier

Diagnosis Codes: ICD

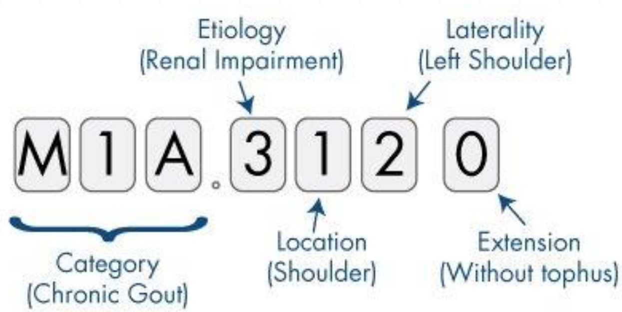
- World Health Organization (WHO) publishes diagnostic coding schema called the International Classification of Diseases (ICD)
- Developed for epidemiologic reporting
  - ICD-10 used in much of the world,
  - ICD-11 has been released in May
- ICD-10 has roughly 68,000 available codes (with flexibility for adding new ones) in comparison to ICD-9's 13,000.

ICD-10

- Codes can be up to seven digits.
- The first digit is always alpha
- The second digit is always numeric.
- The remaining five digits can be any combination.

Example ICD-10 Codes

A01 Typhoid and paratyphoid fevers
    A01.0 Typhoid Fever
        A01.03 Typhoid Pneumonia *
A02 Other salmonella infection
    A02.2 Localized salmonella infections
        A02.22 Salmonella pneumonia *
A20 Plague
    A20.2 Pneumonic plague
A22 Anthrax
    A22.1 Pulmonary anthrax
A37 Whooping cough
    A37.0 Whooping cough due to *Bordetella pertussis*
        A37.01 Whooping cough due to *Bordetella pertussis* with pneumonia *
    A37.1 Whooping cough due to *Bordetella parapertussis*
        A37.11 Whooping cough due to *Bordetella parapertussis* with pneumonia *
    A37.8 Whooping cough due to other Bordetella species
        A37.81 Whooping cough due to other Bordetella species with pneumonia *
    A37.9 Whooping cough, unspecified
        A37.91 Whooping cough, unspecified species with pneumonia *
A50 Congenital syphilis
    A50.0 Early congenital syphilis, symptomatic
        A50.04 Early congenital syphilitic pneumonia *



Procedures: CPT

- Current Procedural Terminology (CPT)
- Widely used in producing bills and reimbursement
  - Specifies information that differentiates the codes based on the cost
    - E.g., there are different codes for pacemaker insertions, depending on whether the leads are "epicardial, by thoracotomy" (33200), "epicardial, by xiphoid approach" (33201) ...
  - Also provides information about the reasons for a procedure

CPT: Example

CPT Code	CPT Code Descriptor	Medicare Physician Fee Schedule - National Average*		
		Global Payment	Professional Payment	Technical Payment
76536	Ultrasound of soft tissues of head and neck (e.g., thyroid, parathyroid, parotid), real time with image documentation	$125.54	$27.22	$98.33
76705	Ultrasound, abdominal, real time with image documentation); limited (e.g., single organ, quadrant, follow-up)	$111.26	$28.58	$82.68
76815	Ultrasound, pregnant uterus, real time with image documentation, limited (eg, fetal heart beat, placental location, fetal position and/or qualitative amniotic fluid volume), one or more fetuses	$92.20	$30.96	$61.24
76817	Ultrasound, pregnant uterus, real time with image documentation, transvaginal	$101.31‡	$36.74	$64.57‡
76818	Fetal biophysical profile; with non-stress testing	$125.89	$51.71	$74.17
76881	Ultrasound, extremity, nonvascular, real-time with image documentation; complete	$124.52	$30.96	$93.56

*http://www.idealmed.com/blog/billing-information-primary-care/

CPT Codes

TABLE 15–4 Examples of CPT Codes

CPT Code	Description
Level 1: 80000–89398	Pathology and laboratory tests
Level 2: 81000–81099	Urinalysis procedures
81015	Urinalysis; microscopic only
81005	Urinalysis; qualitative or semiquantitative, except immunoassays
Level 2: 80100–80103	Drug testing
80100	Drug screen, qualitative; multiple drug classes chromatographic method, each procedure

http://www.csun.edu/~dn58412/IS531/IS531_Lecture-5.ppt

Drugs: NDC & RxNorm

- Drug Codes
  - US: National Drug Codes (NDC): not as comprehensive as ATC, no uniform class hierarchy
  - US: RxNorm: to address the above problems, part of UMLS

Example National Drug Code (NDC)

Different Coding Systems

- Each has idiosyncrasies and limitations
  - ICD9-CM: more than 500 separate codes for Tuberculosis
- None can be completely satisfactory
  - Yet, if there is no structure to data and physicians treat EHR as a blank page, there is little use!
- Tension between the need for a system
  - General enough to cover many different patients
  - Precise and unique enough to cover a specific patient

UMLS

- Researchers have worked for two decades to develop a unified language
  - Unified Medical Language System (UMLS)
    - Metathesaurus: contains over 8.9 million terms from 160 sources, has semantic relations tying concepts

http://www.nlm.nih.gov/research/umls/

UMLS

- Example

Bacterial pneumonia

Source: CSP93/PT/2596-5280; DOR27/DT/U000523; ICD91/PT/482.9; ICD91/IT/482.9
Parent: Bacterial Infections; Pneumonia; Influenza with Pneumonia
Child: Pneumonia, Mycoplasma
Narrower: Pneumonia, Lobar; Pneumonia, Rickettsial; Pneumonia, Staphylococcal; Pneumonia due to *Klebsiella pneumoniae;* Pneumonia due to Pseudomonas; Pneumonia due to *Hemophilus influenzae*
Other: *Klebsiella pneumoniae, Streptococcus pneumoniae*

Pneumonia, Lobar

Source: ICD91/IT/481; MSH94/PM/D011018; MSH94/MH/D011018; SNM2/RT/M-40000; ICD91/PT/481; SNM2/PT/D-0164; DXP92/PT/U000473; MSH94/EP/D011018; INS94/MH/D011018;INS94/SY/D011018
Synonym: Pneumonia, diplococcal
Parent: Bacterial Infections; Influenza with Pneumonia
Broader: Bacterial Pneumonia; Inflammation
Other: *Streptococcus pneumoniae*
Semantic: inverse-is-a: *Pneumonia*
has-result: *Pneumococcal Infections*

Interchange Standards: HL7

- For transfer of clinical and administrative data between hospital information systems
- Version 2: XML capabilities makes it Web-enabled(latest version 3.0)

Interchange Standards: HL7

- Example

```
MSH|^~&\|DHIS|OR|TMR|SICU|199212071425|password|ADT|16603529|P|2.1<cr>

EVN|A02|199212071425||<cr>

PID|||Z99999^5^M11||GUNCH^MODINE^SUE|RILEY|19430704 |F||C|RT. 1, BOX
97^ZIRCONIA^NC^27401 |HEND|(704)982-1234|(704)983-1822||S|C||245-33-
9999<cr>

PV1|1|I|N22^2204|||OR^03|0940^DOCTOR^HOSPITAL^A||| SUR|||||A3<cr>

OBR|7|||93000^EKG REPORT|R|199401111000|199401111330|||RMT||||19940111
11330|?|P030||||||199401120930||||||88-126666|A111|VIRANYI^ANDREW<cr>

OBX|1|ST|93000.1^VENTRICULAR RATE(EKG)||91|/MIN|60-100<cr>

OBX|2|ST|93000.2^ATRIAL RATE(EKG)||150|/MIN|60-100<cr>

. . .

OBX|8|ST|93000&IMP^EKG DIAGNOSIS|1|^ATRIAL FIBRILATION<cr>
```

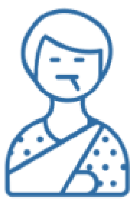
An example of an HL7 ADT transaction message. This message includes the Message Heading segment, the EVN trigger definition segment, the PID patient-identification segment, the PV1 patient-visit segment, the OBR general-order segment, and several OBX results segments.

Interchange Standards: IEEE 1073

- Standard for medical device communication
  - Bedside devices in intensive care unit, operating room, and emergency room

Some Example Applications

The Person at the Centre of Healthcare



Patient/Person

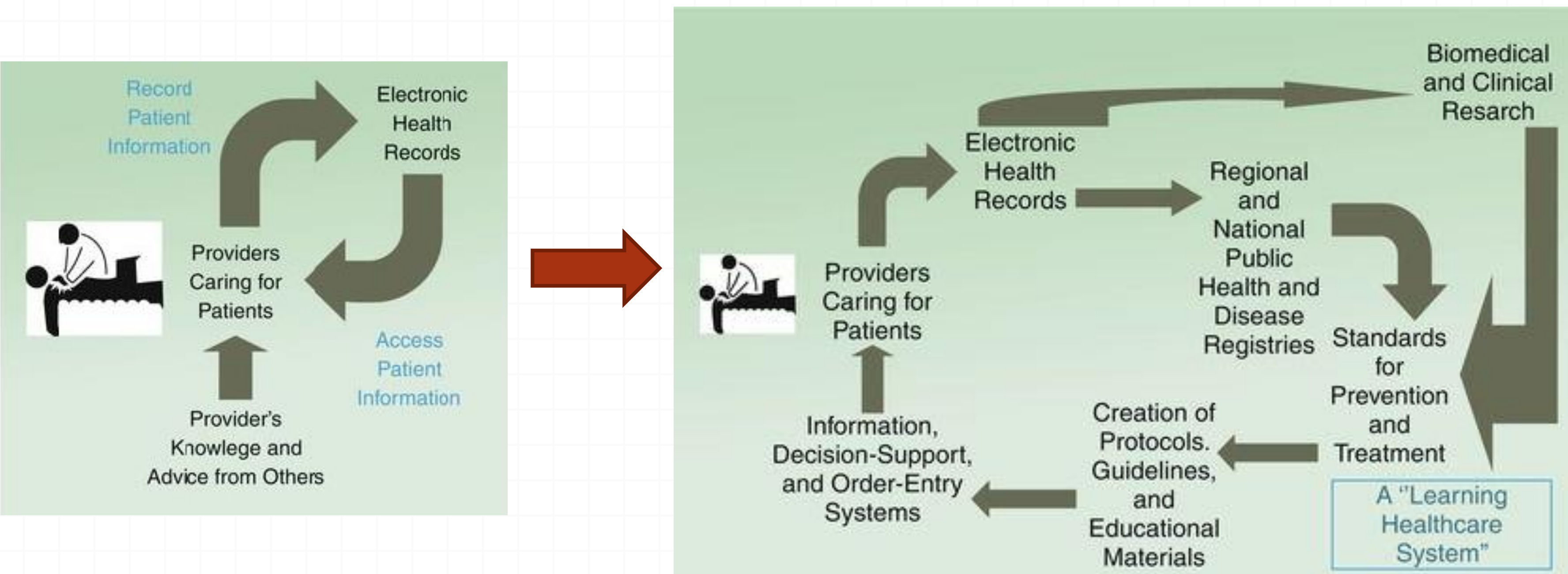
ML has the capacity to transform healthcare

- Understanding physiological changes over time
- Forecasting of progression or onset of disease
- Personalising treatment strategies

Azizi, Palla, Belgrave, ICML Tutorial: Machine Learning for Personalised Health, 2018

A Learning Health Care System

- A cycle of information flow in a distributed form

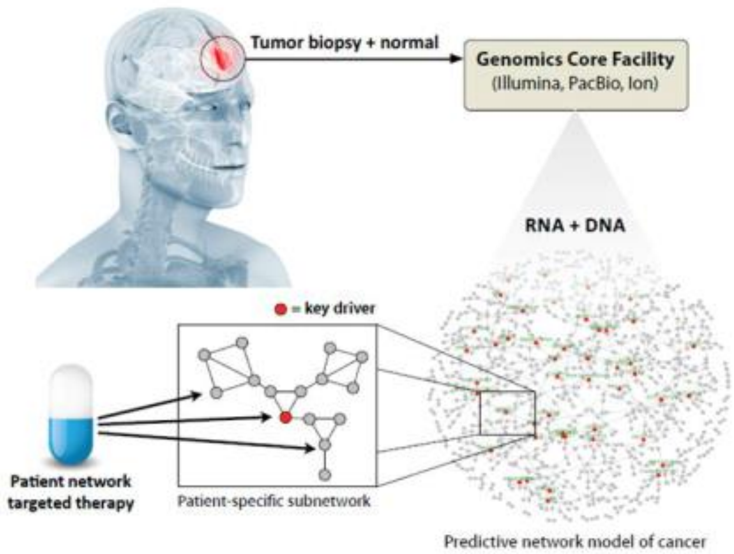


Driving Personalized Cancer Therapy

- New program creates predictive network models for major types of cancer, along with **patient-specific** subnetwork models
- By analyzing patient-specific mutations, the program aims to pinpoint the most **appropriate therapy** for each individual patient

"We can uncover what is happening specifically in a patient's tumor, what key driver genes are mutated, what signaling pathways are altered, and then whether we can better match a given patient's condition to a more appropriate treatment"

Eric Schadt, PhD
Professor and Chair for Genetics and Genomic Sciences
Director of the Icahn Institute

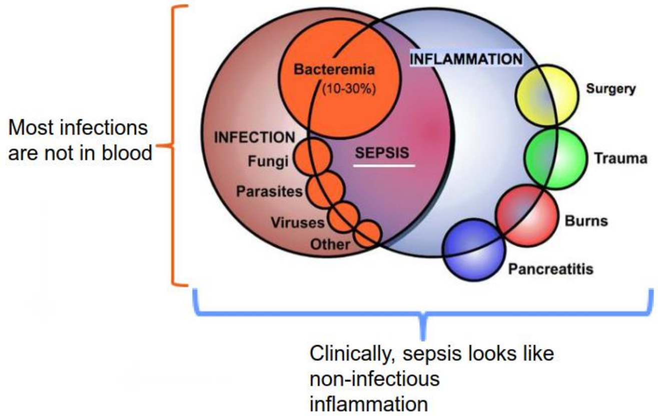


Real-time accurate imaging diagnosis

- Given the task of identifying lung nodules on screening chest CTs
  - An experienced radiologist reads an image correctly 65% of the time
  - Two experienced radiologists read an image correctly about 71% of the time
  - One experienced radiologist with computer-aided diagnosis software reads the image correctly about 79% of the time
- Need clinical decision support/computer-aided diagnosis software to reduce patient stress and healthcare cost from false positives
  - One ultrasound exam typically produces between 3,000-5,000 images
  - Usefulness of ultrasound images highly dependent on the operator
  - Software to detect tumors could assist operator and radiologist

Haygood, T.M., et. al. Consistency of response and image recognition, pulmonary nodules. Br J. Radiol. June 2014 87(1038): 20130767. doi: 10.1259/bjr.20130767

“Sepsis” is a difficult diagnosis



Publicly available data offer

- Alternative experimental conditions
- Real-world heterogeneity
- No cost to perform experiments
- Human samples—
  with built-in IRB approval

- Searching for data sets
- Elucidating genomic signals
- Confirming those signals in “validation” data sets
- Making discoveries without ever performing a primary experiment