

Principles of Health Data Science

Precision medicine, EHR sources and controlled clinical terminologies

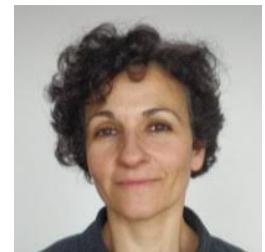
Spiros Denaxas
Institute of Health Informatics
University College London
<http://denaxaslab.org>

About the module

- **Structure**
 - Lecture + practical + tutorials + reflection questions
 - Every Monday
 - 09:00-10:30 (lecture),
 - 11:00-12:30 (practical group 1)
 - 14:00-15:30 (practical group 2)
 - Reading week – November 9th
- **People**
 - **Spiros Denaxas** – s.denaxas@ucl.ac.uk
 - **Holger Kunz** – h.kunz@ucl.ac.uk
- **Assessment**
 - Coursework at end of term 1

About me and our lab

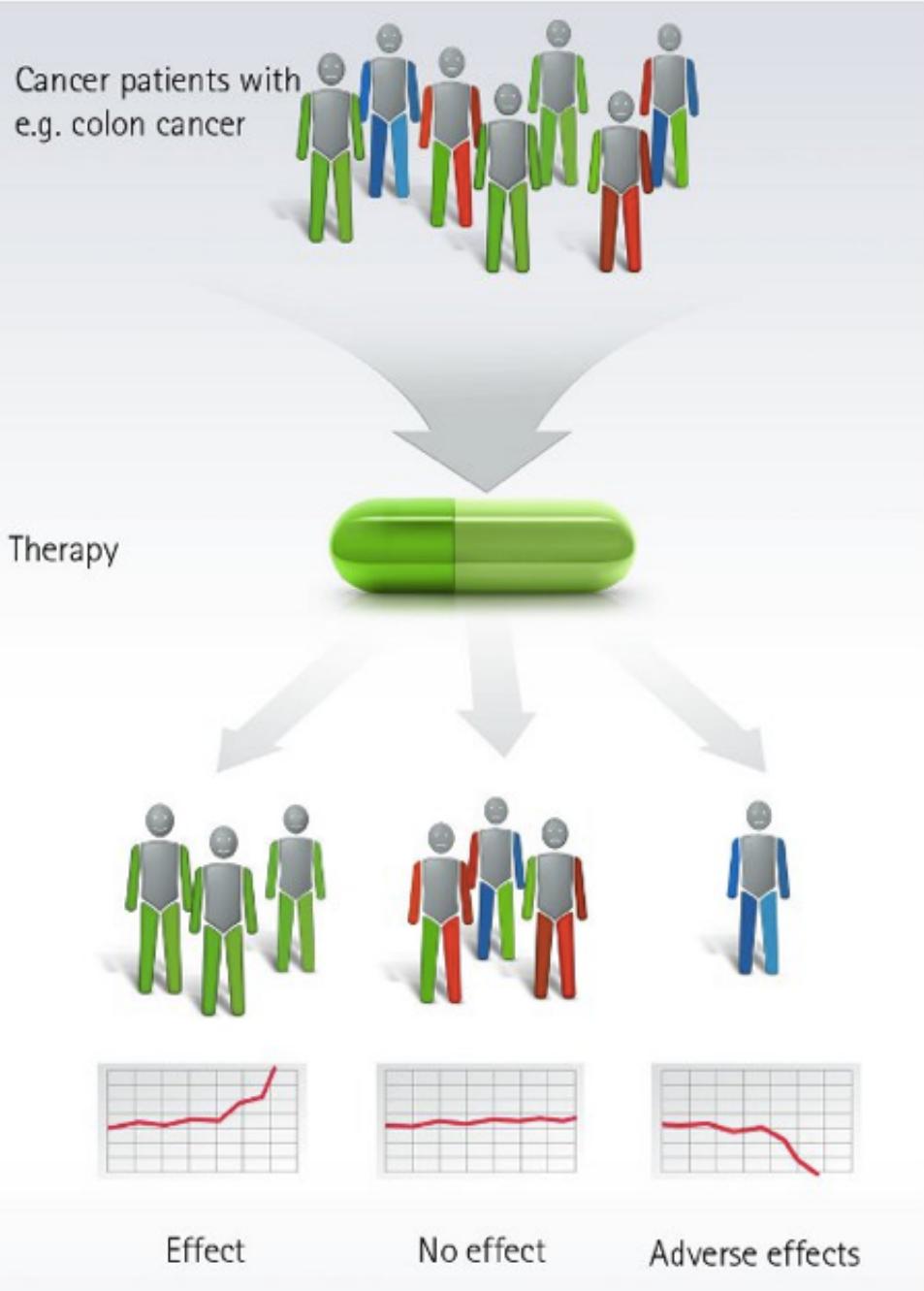
- I am Greek: Σπύρος
 - Pronouns: he/him/his
 - Computer scientist
 - PhD in Bioinformatics
 - Working with EHR, AI and biomedical data
-
- <http://denaxaslab.org>

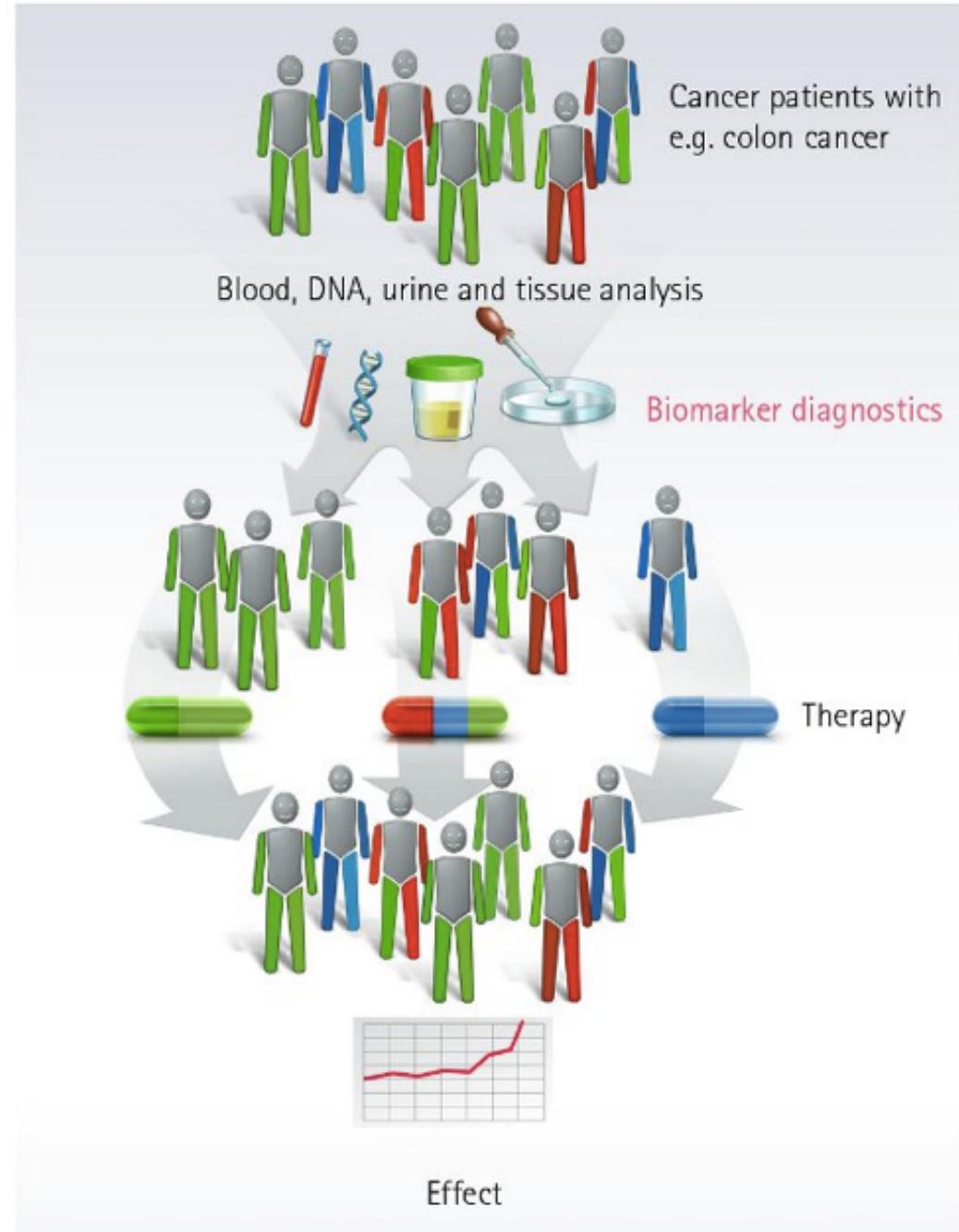
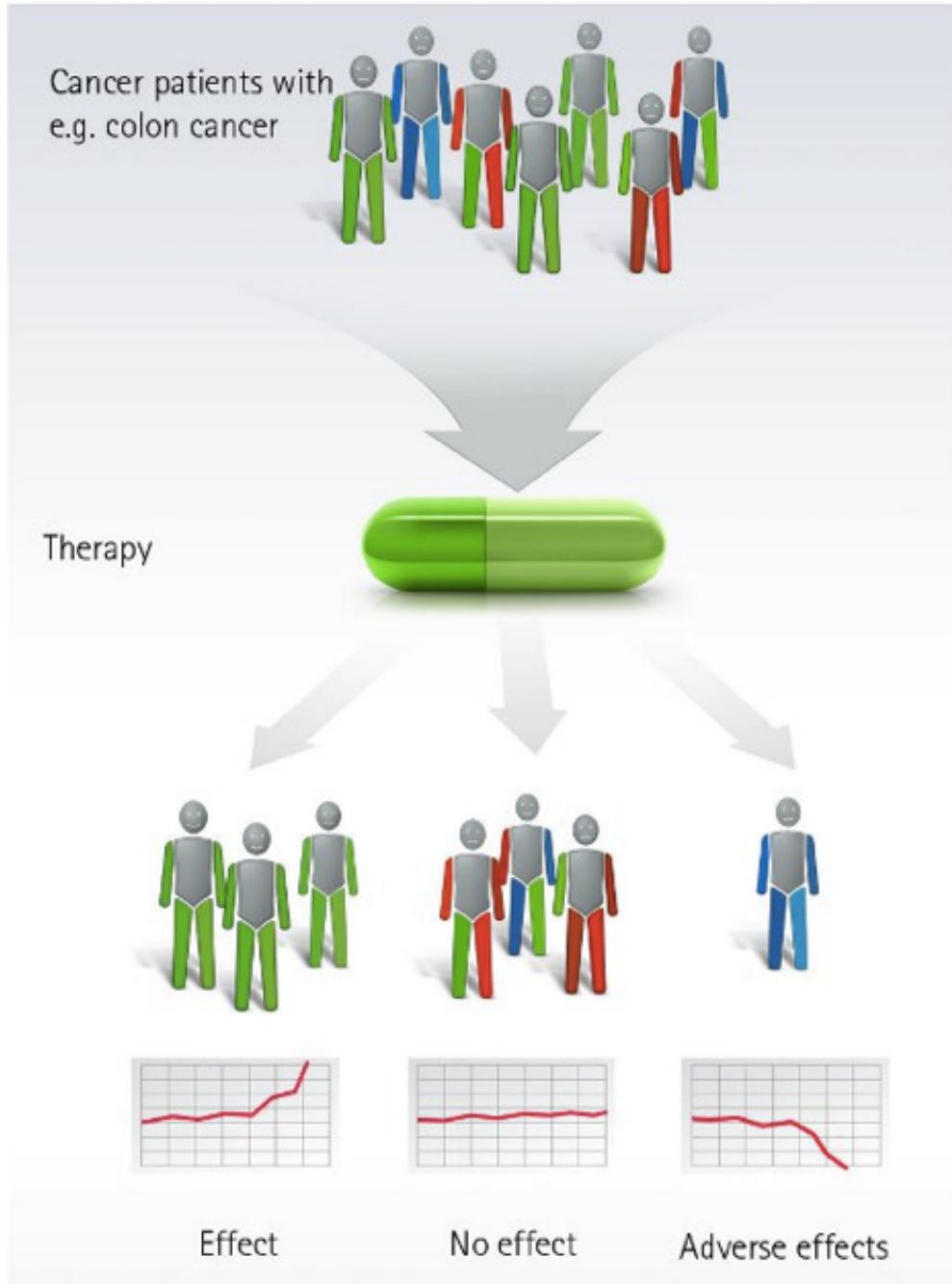


Learning outcomes

- By the end of the lecture, students:
 1. Will be familiar with the motivation of precision/personalized medicine
 2. Will know the main types of electronic health records data and potential applications for research
 3. Will be able to describe what controlled clinical terminologies are, and how they are used, give examples of common systems
 4. Will be exposed to the main principles behind clinical decision support systems

Part 1: Precision medicine





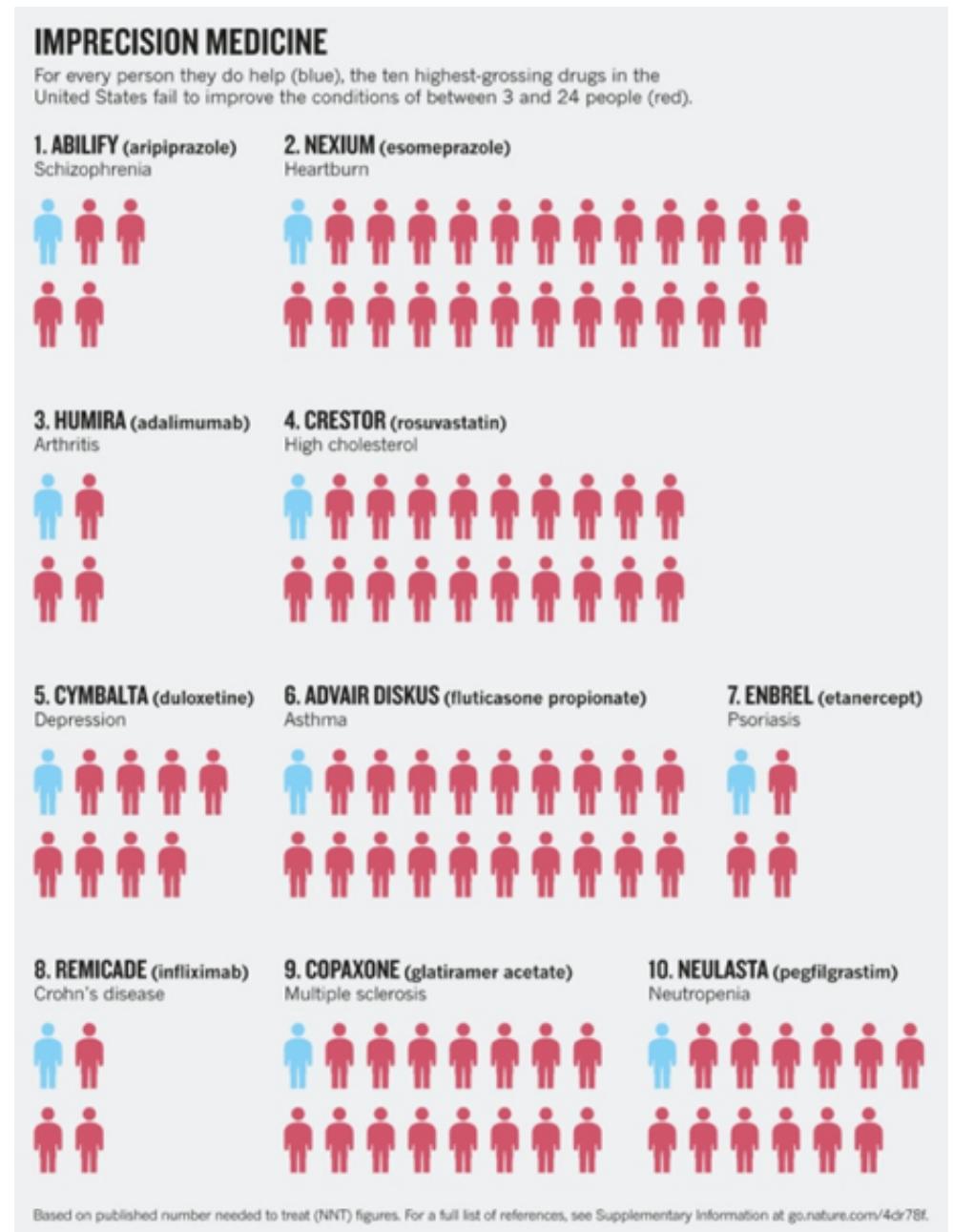
Precision medicine

New drugs:

- Cost: \$2.6-2.8bn
- Time: 10-15 years
- Fail: 97% of R&D fails
- Most top-selling drugs fail to work in 50-70% of patients

Why?

- Highly-heterogeneous underlying biological mechanisms
- Over-simplistic representations of disease phenotypes



Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records



Maria Pikoula^{1,2*†}, Jennifer Kathleen Quint^{2,3,4†}, Francis Nissen^{2,4}, Harry Hemingway^{1,2}, Liam Smeeth^{2,4} and Spiros Denaxas^{1,2}

- Anxiety / Depression
- Not Comorbid
- CVD / Diabetes
- Frail / Severe COPD
- Atopy / Obesity

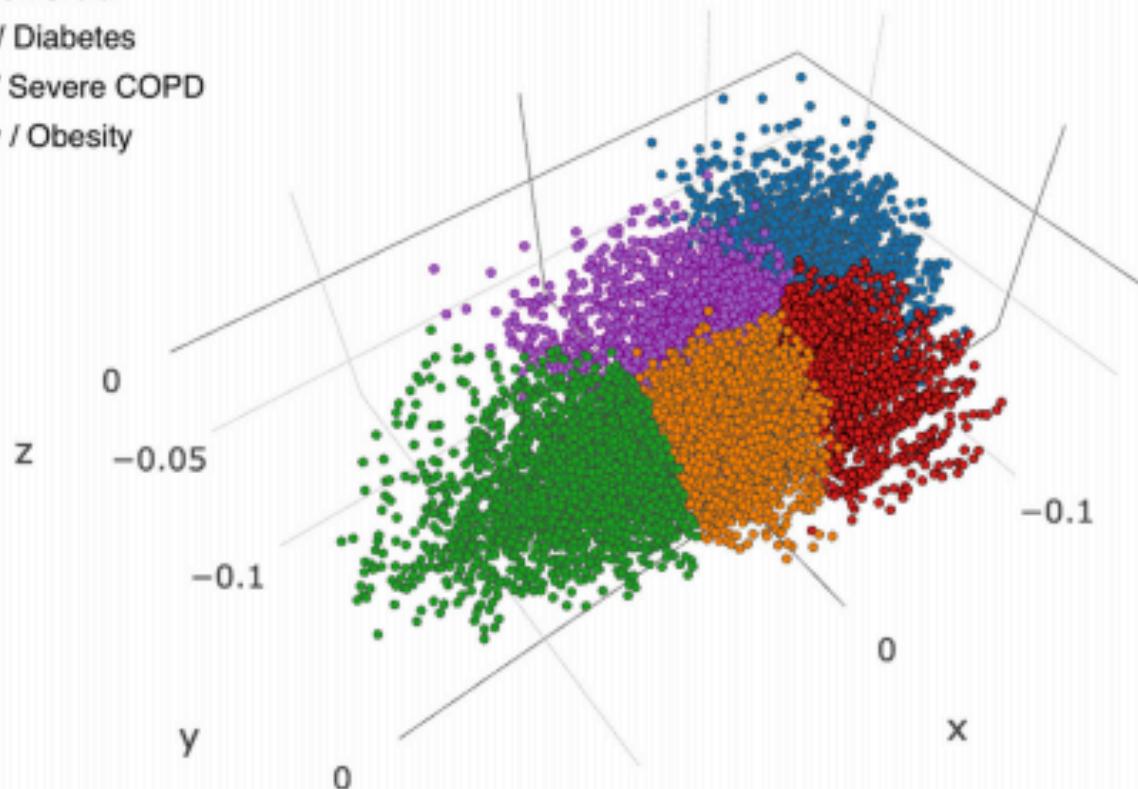


Fig. 4 3D scatter plot of the three MCA Components colour-coded by cluster assignment

Identification of type 2 diabetes subgroups through topological analysis of patient similarity

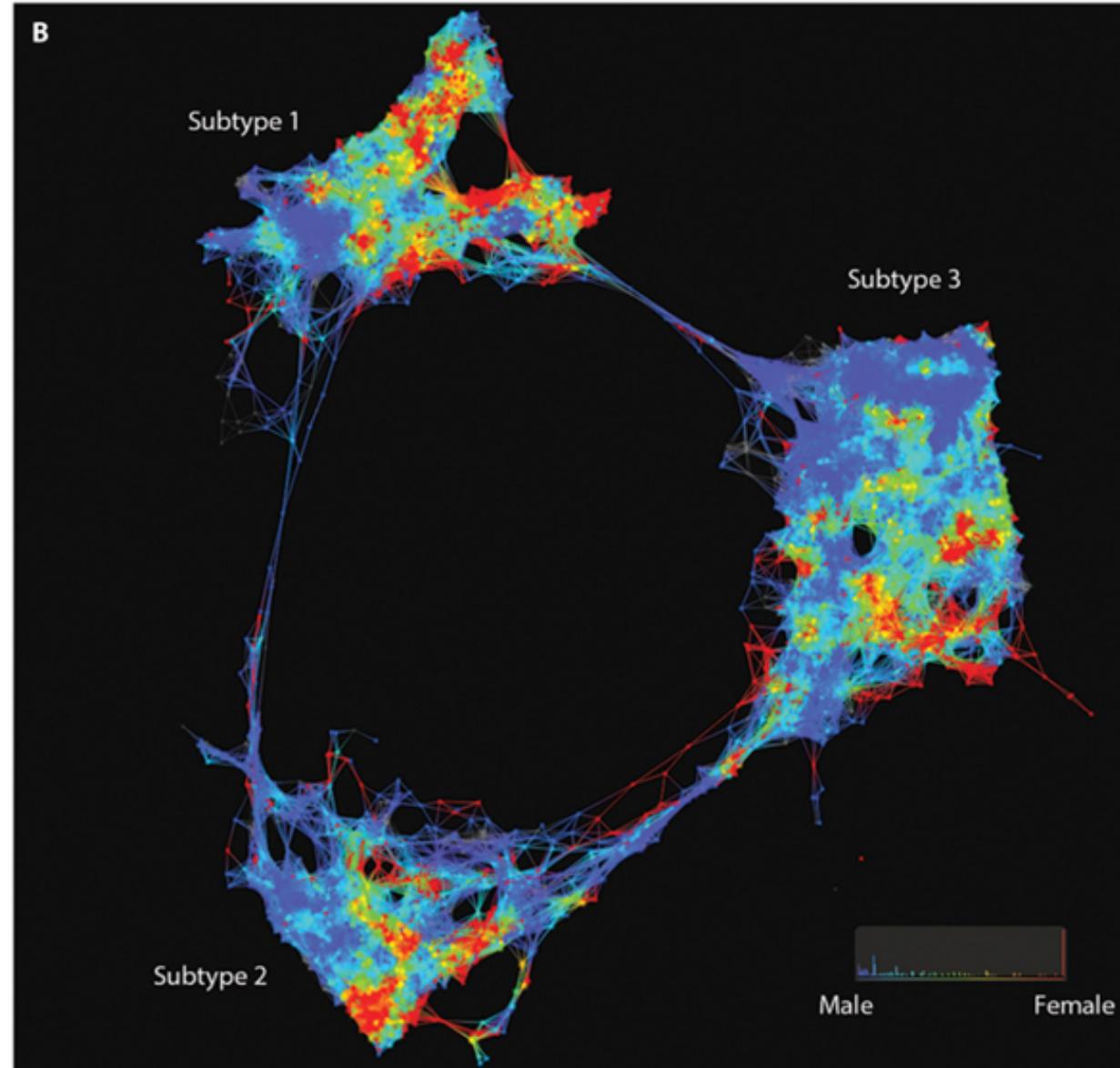
Li Li¹, Wei-Yi Cheng¹, Benjamin S. Glicksberg¹, Omri Gottesman², Ronald Tamler³, Rong Chen¹, Erwin P. Bottinger² and Joel...

+ See all authors and affiliations

Science Translational Medicine 28 Oct 2015:

Vol. 7, Issue 311, pp. 311ra174

DOI: 10.1126/scitranslmed.aaa9364



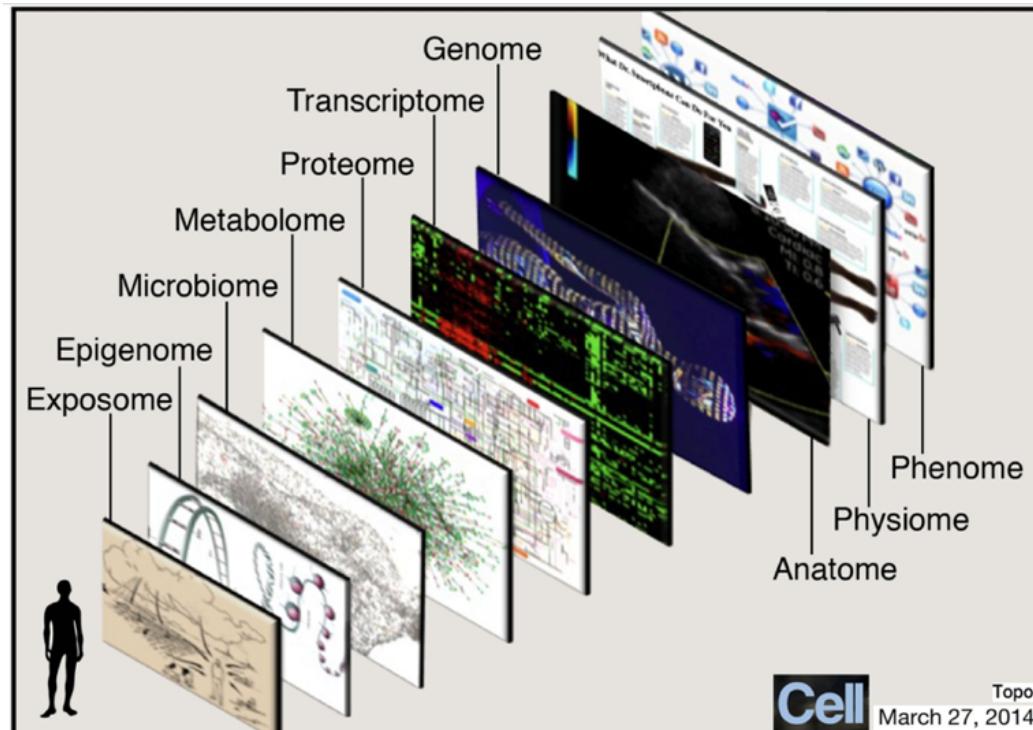
Using big data to guide us

News Feature | Published: 05 December 2019

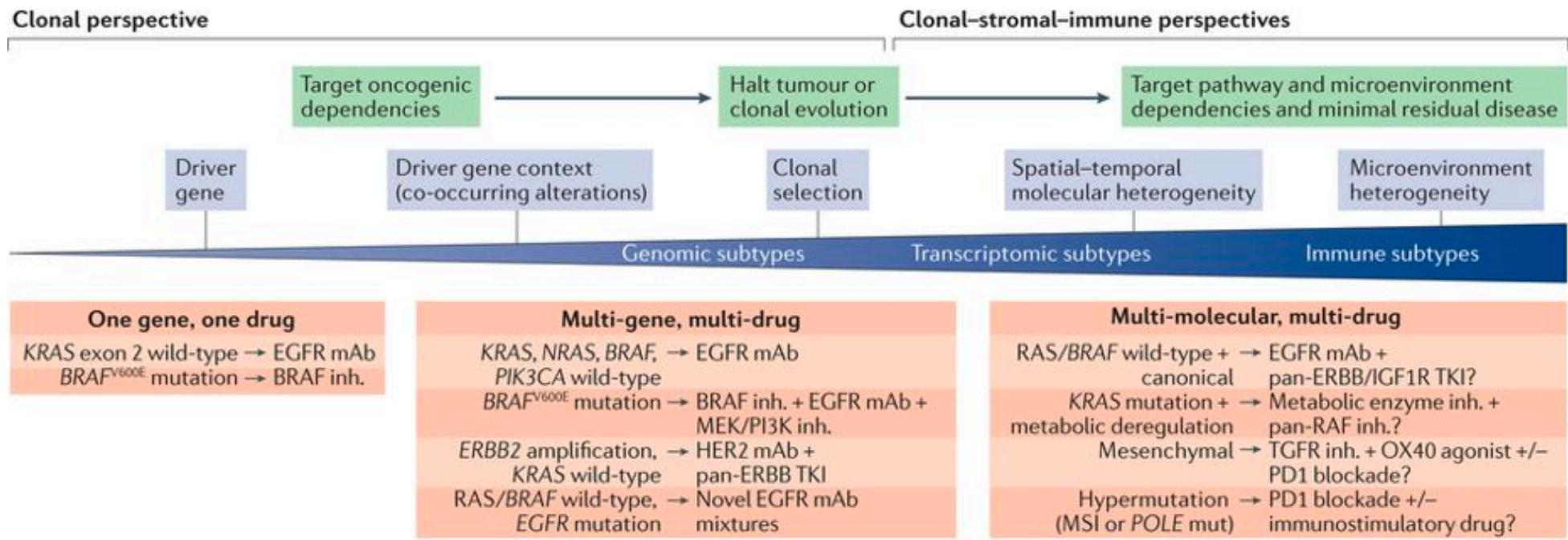
Looking forward 25 years: the future of medicine

Nature Medicine 25, 1804–1807(2019) | Cite this article

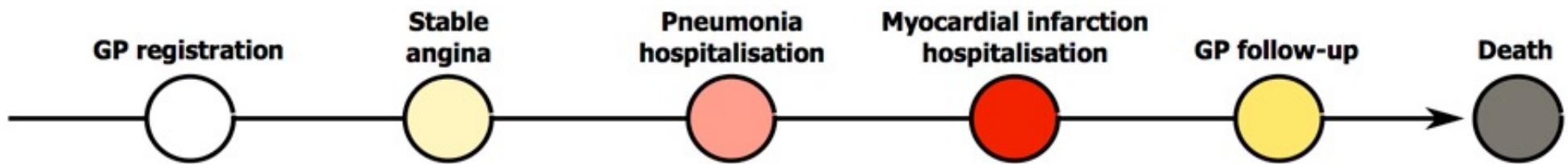
15k Accesses | 326 Altmetric | Metrics



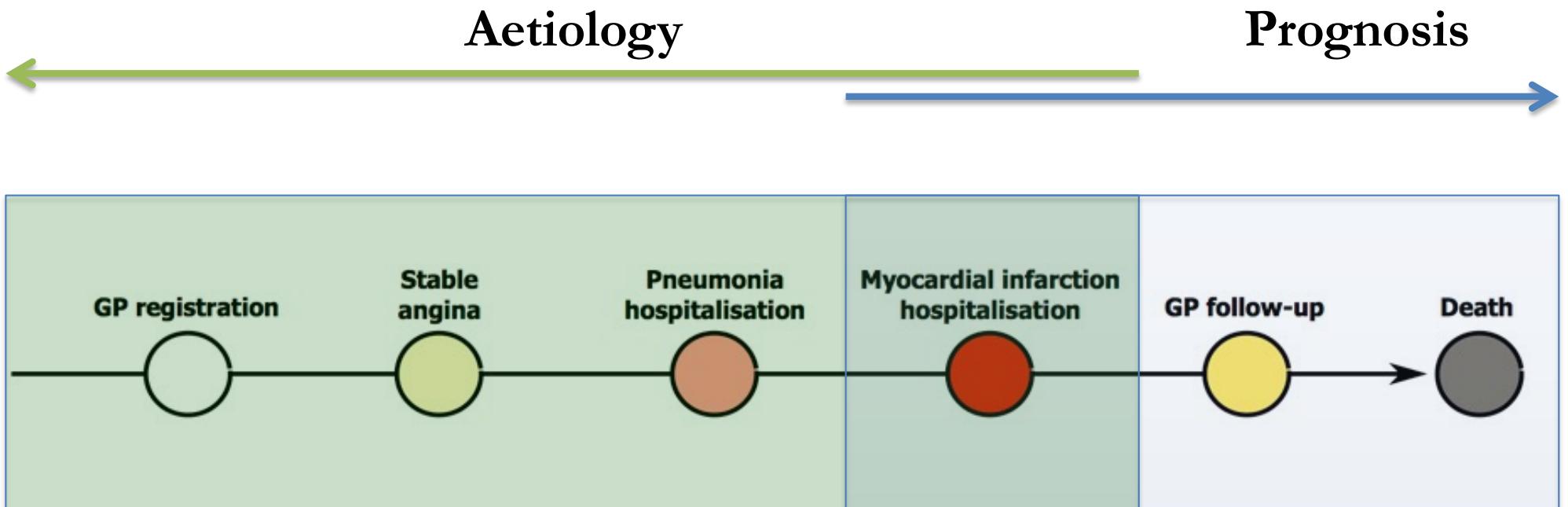
Colorectal cancer treatment evolution



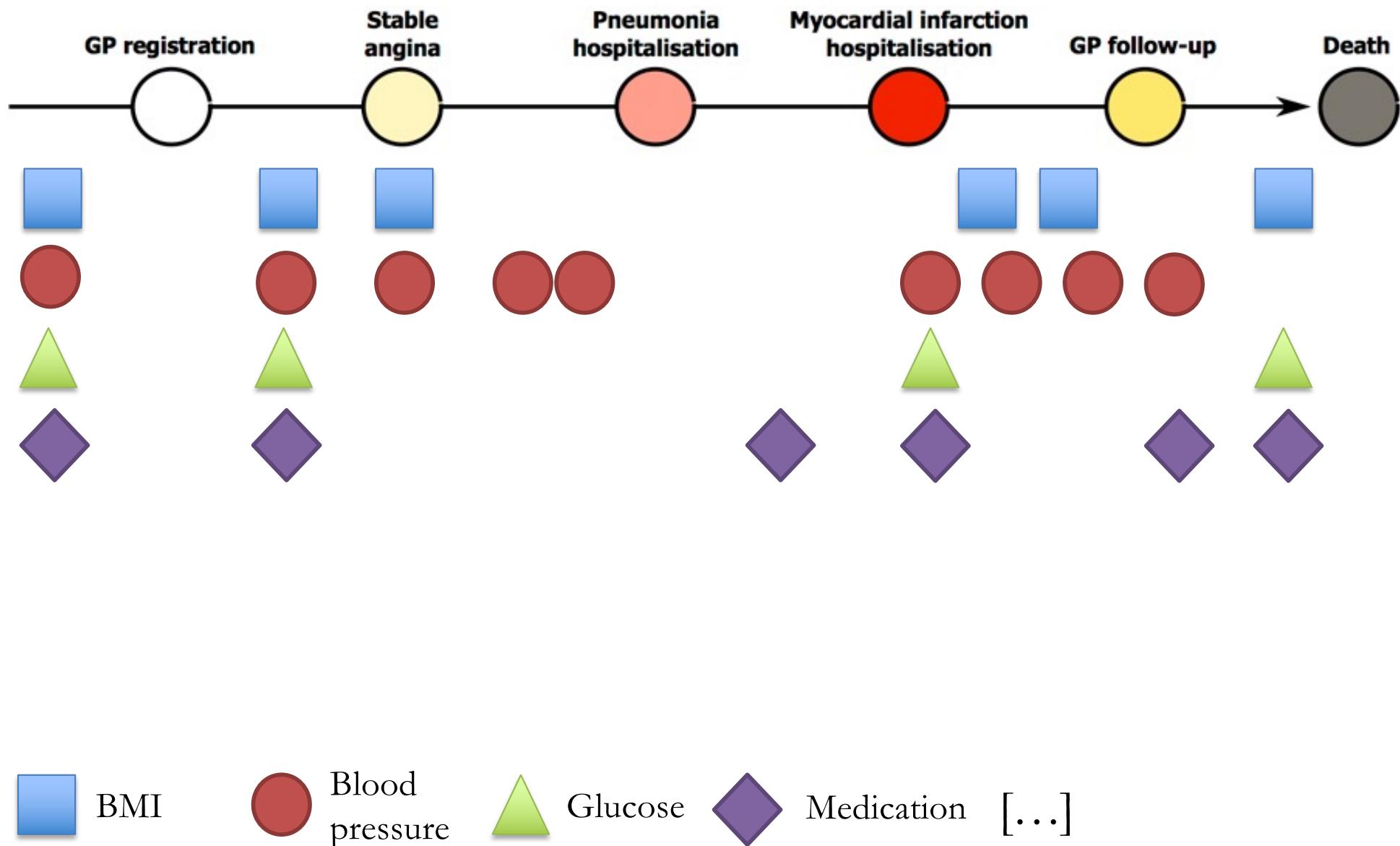
Patient pathway



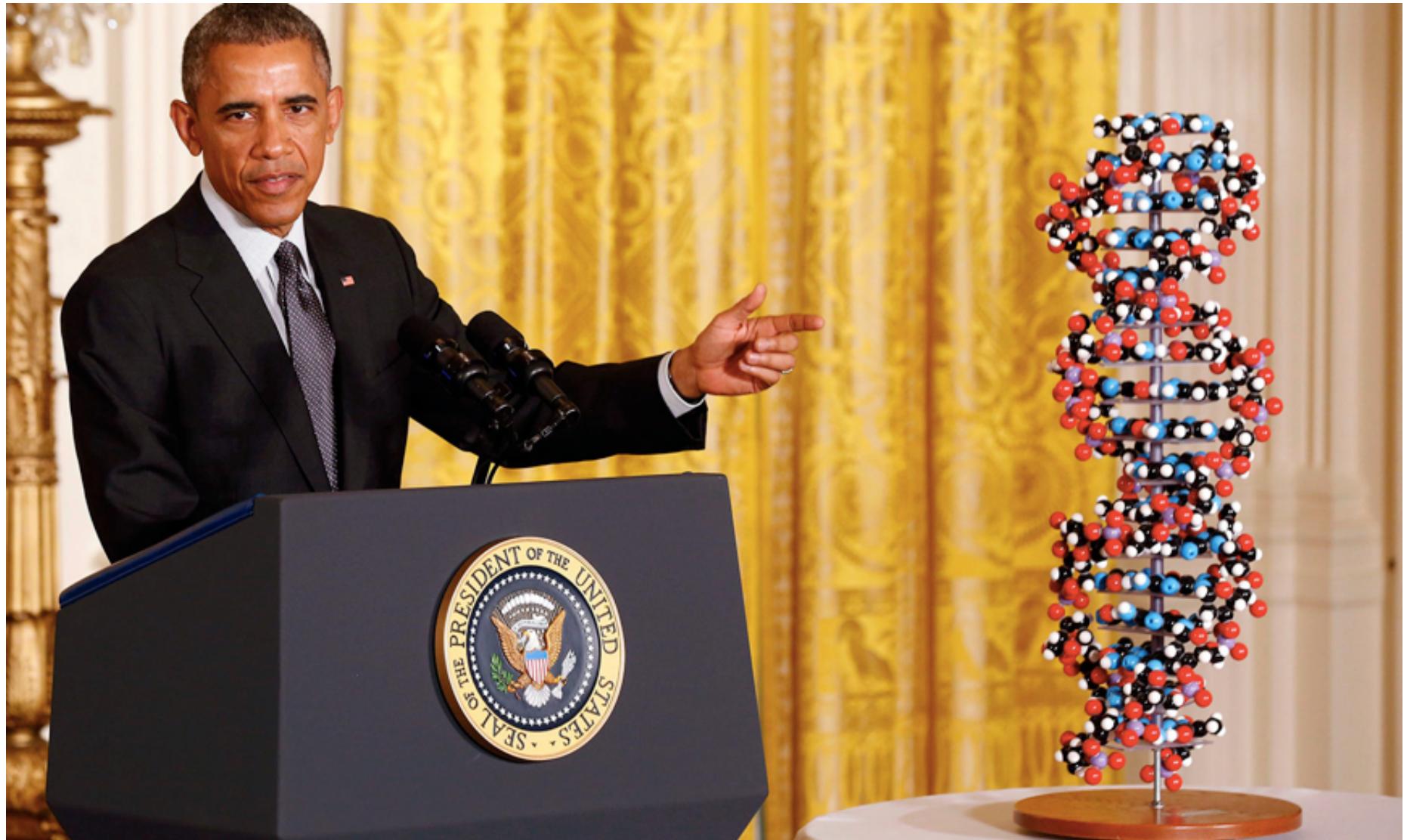
Patient pathway



We need data



Personalized, data-driven medicine



All of Us

National Institutes of Health
All of Us Research Program

[Log In](#) | [JOIN NOW](#) | [Search](#)

About Get Involved Funding and Program Partners Protecting Data and Privacy News and Events

Interested in the All of Us Research Program? [GET INVOLVED](#)

Diversity, transparency, and accessibility are core values that guide All of Us. [VIEW OUR CORE VALUES](#)

Sign up to be notified of announcements, events, funding news and more. [SUBSCRIBE](#)

We are building a research program of 1,000,000+ people.

The mission of the All of Us Research Program is to accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us.

OPPORTUNITIES FOR RESEARCHERS

Research focuses on the intersection of 3 factors

Search Across Data Types

[Keyword Search](#)

Data includes 225,140 participants and is current as of 2/11/2020.

EHR Domains:

Conditions 20,776 medical concepts 113,200 participants in this domain View Top Conditions	Drug Exposures 20,951 medical concepts 104,500 participants in this domain View Top Drug Exposures	Labs & Measurements 10,049 medical concepts 109,100 participants in this domain View Top Labs & Measurements	Procedures 20,546 medical concepts 102,140 participants in this domain View Top Procedures
--	--	--	--

Survey Questions:

The Basics 16 questions available 225,140 participants in this domain This survey includes participant demographic information. View Complete Survey	Overall Health 21 questions available 219,800 participants in this domain Survey includes information about how participants report levels of individual health. View Complete Survey	Lifestyle 26 questions available 218,500 participants in this domain Survey includes information on participant smoking, alcohol, and recreational drug use. View Complete Survey	Personal Medical History 465 questions available 39,320 participants in this domain This survey includes information about past medical history, including medical conditions and approximate age of diagnosis. View Complete Survey
---	--	--	---

<https://allofus.nih.gov/>

<https://www.researchallofus.org/data-snapshots/>

End of Part 1

- **Summary**
 - Diseases are very heterogenous
 - Disease subtypes can affect how well (or not) treatments work
 - Identifying and exploring these subtypes requires data
 - Precision medicine: moving away from one-size-fits-all
 - Data-driven medicine
- **Questions?**

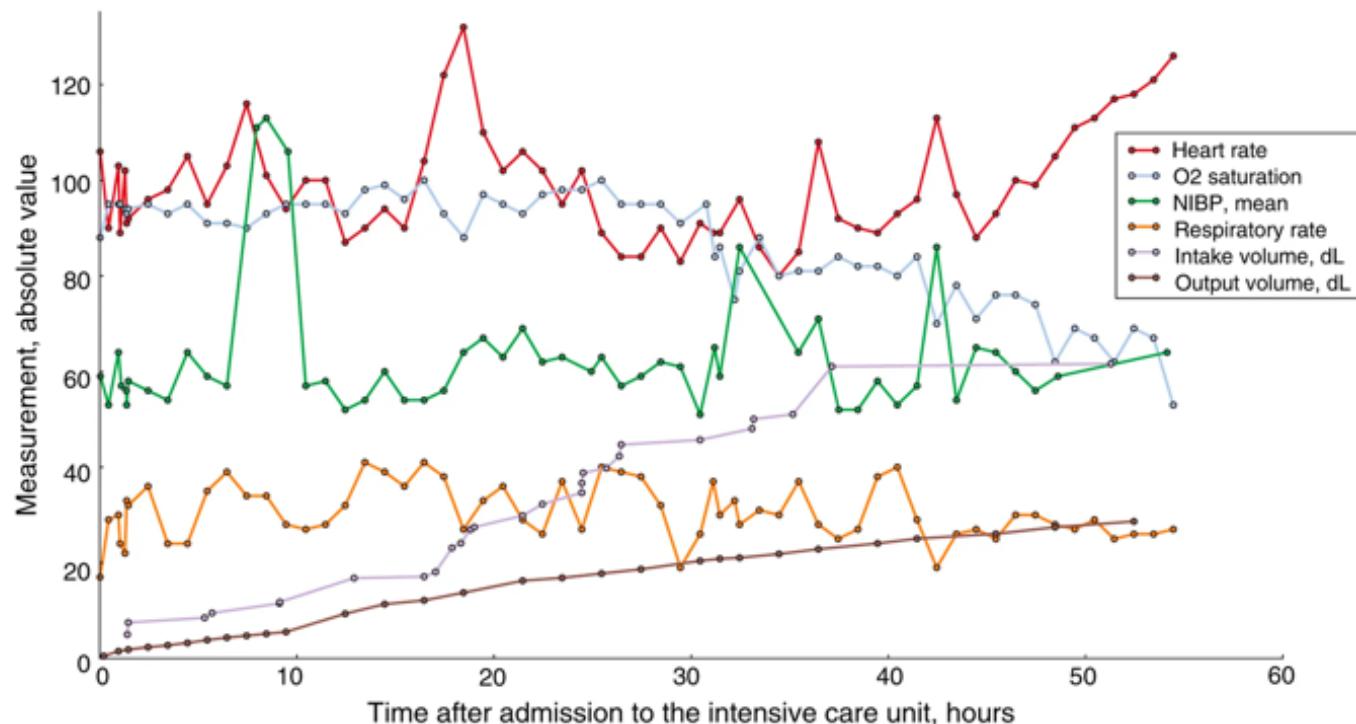


5 minutes break

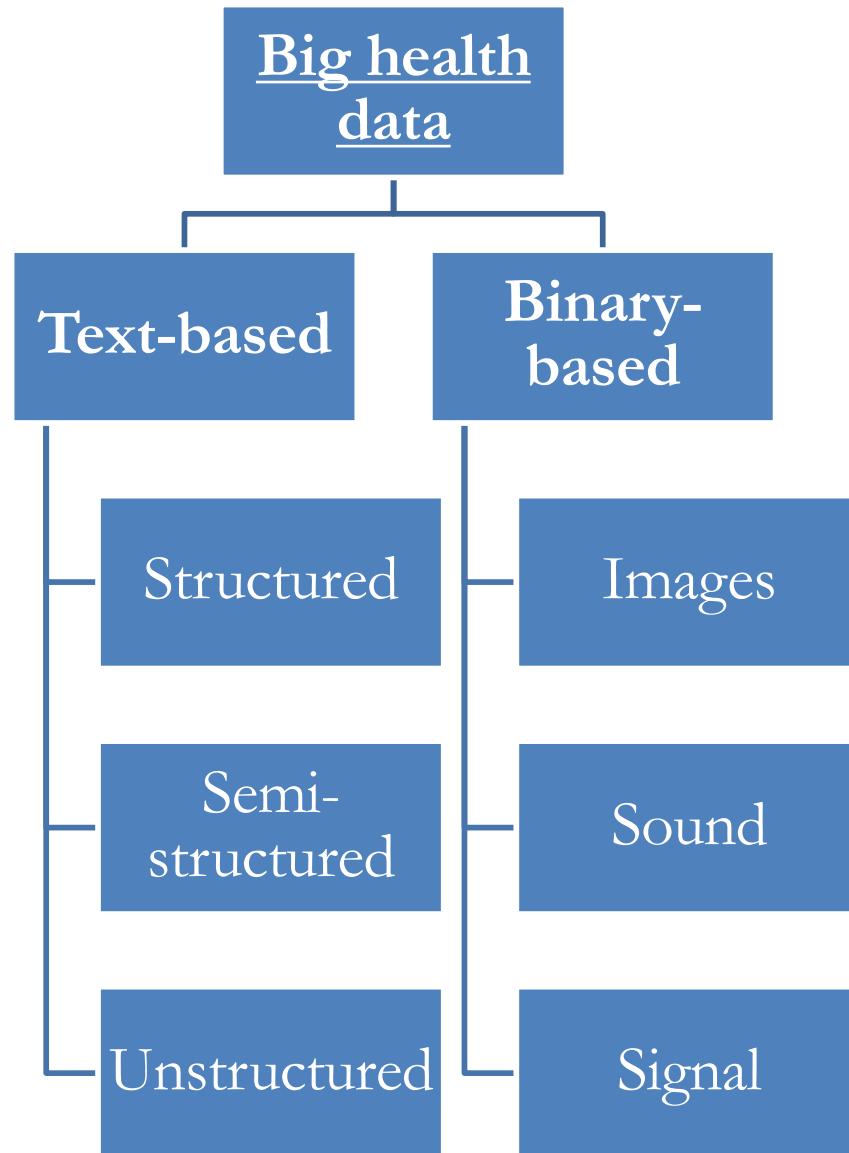
Part 2: Big Data in Healthcare

Medicine is data-intensive!

Code status	Full code						Comfort measures
GCS: Verbal	Oriented		Oriented		Oriented		Incomprehensible sounds
GCS: Motor	Obeys commands		Obeys commands		Obeys commands		Flex-withdraws
GCS: Eye	Spontaneously		Spontaneously		Spontaneously		None
Platelet, K/uL	48	53		46			45
Creatinine, mg/dL	0.7		0.7				0.8
White blood cell, K/uL	9.1	12.4		16.8			23.2
Neutrophil, %	37						
Morphine Sulfate							
Vancomycin (1 dose)							
Piperacillin (1 dose)							
NaCl 0.9%							
Amiodarone							
Dextrose 5%							



Data types



Big data types

Text-based

76 yo man with h/o HTN, DM, and sleep apnea who presented to the ED complaining of chest pain. He states that the pain began the day before and consisted of a sharp pain that lasted around 30 seconds, followed by a dull pain that would last around 2 minutes.

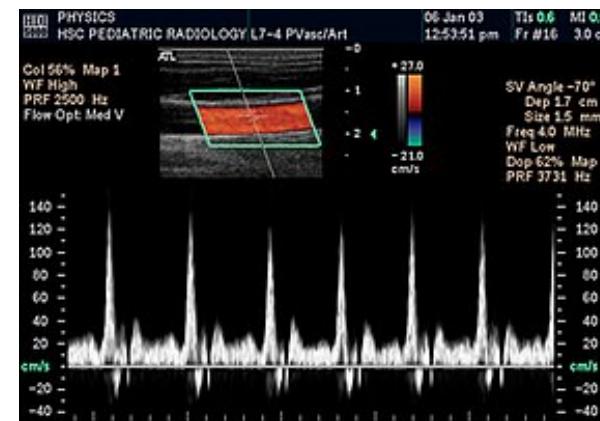
aaaaaaaaaaaaataaaattaaaaaaaaataaaaaataaaataaa
ataaaaaataaaaaaaaaataataaaaaaaaaatattata
aaaaaaaaaa

I21.1 Acute myocardial infarction

Systolic blood pressure (mmHg): 133

Diastolic blood pressure (mmHg): 88

Binary data



- Interpreted as text
- Designed for consumption by humans
- Consist of Unicode characters

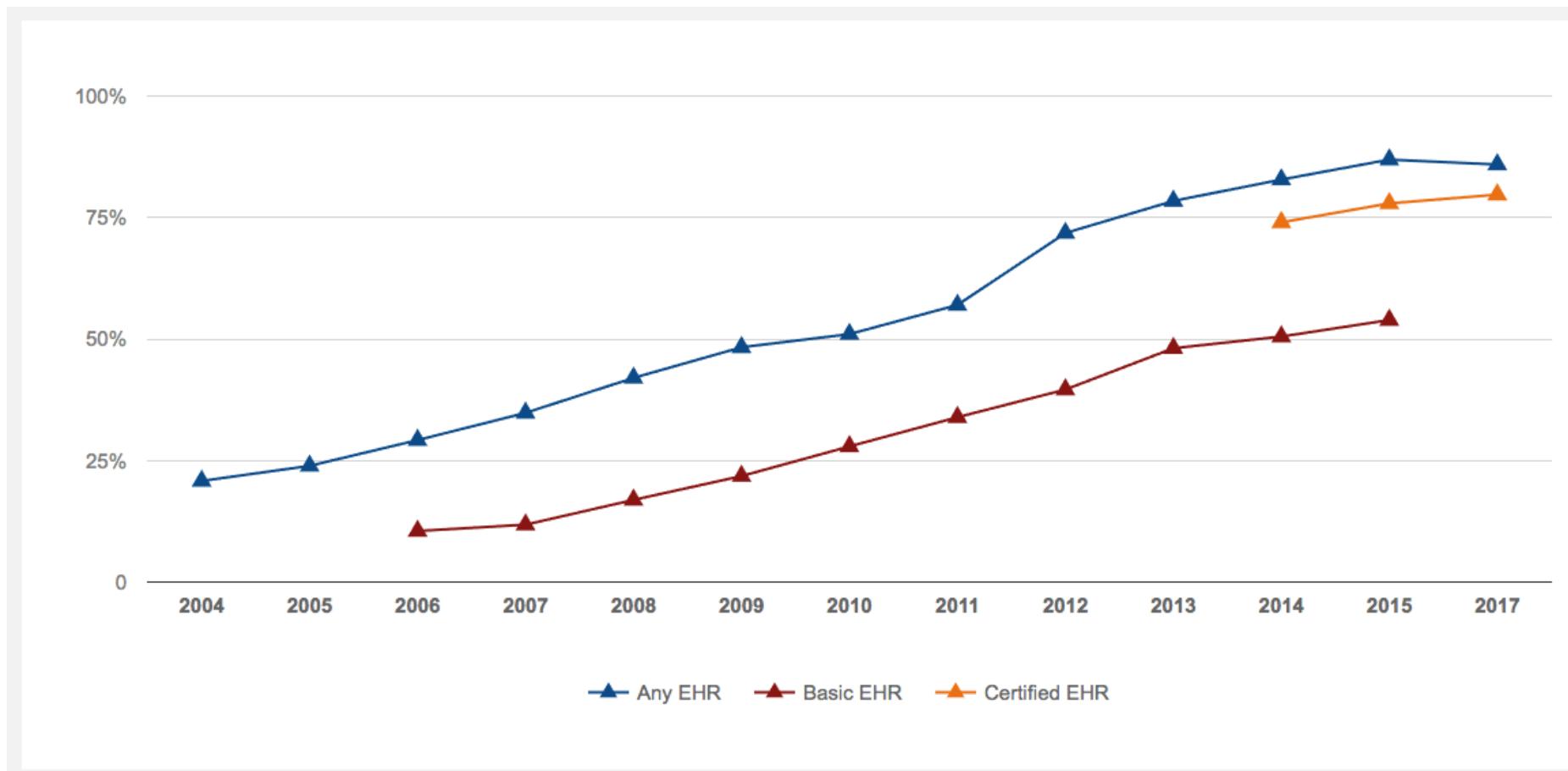
- Cannot be interpreted as text
- Designed for consumption through computer applications
- Consist of Unicode characters

Electronic Health Records (EHR)

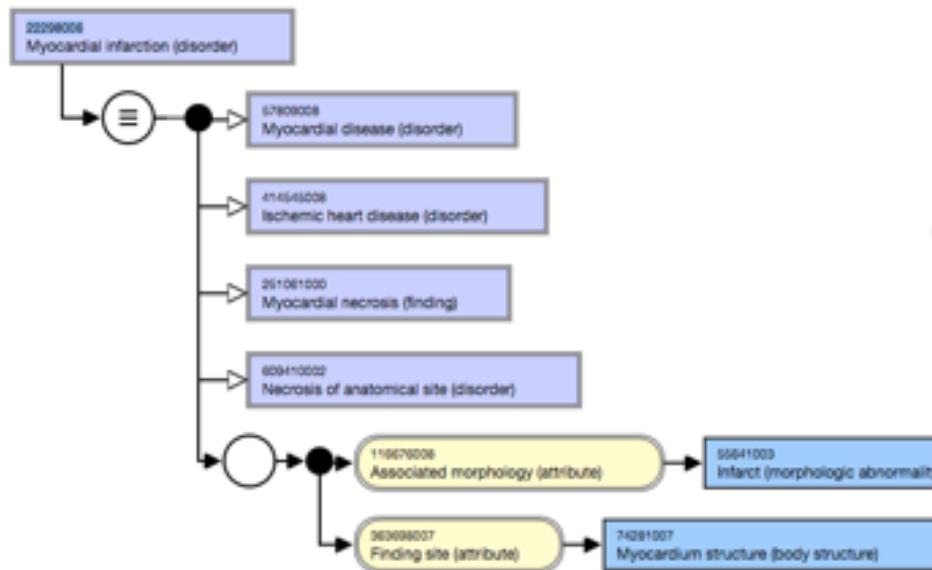


Helen Golden is shown at the Kard-Veyer in the Medical Record Department with the innovative pneumatic tube behind her.

Electronic Health Record (EHR) adoption



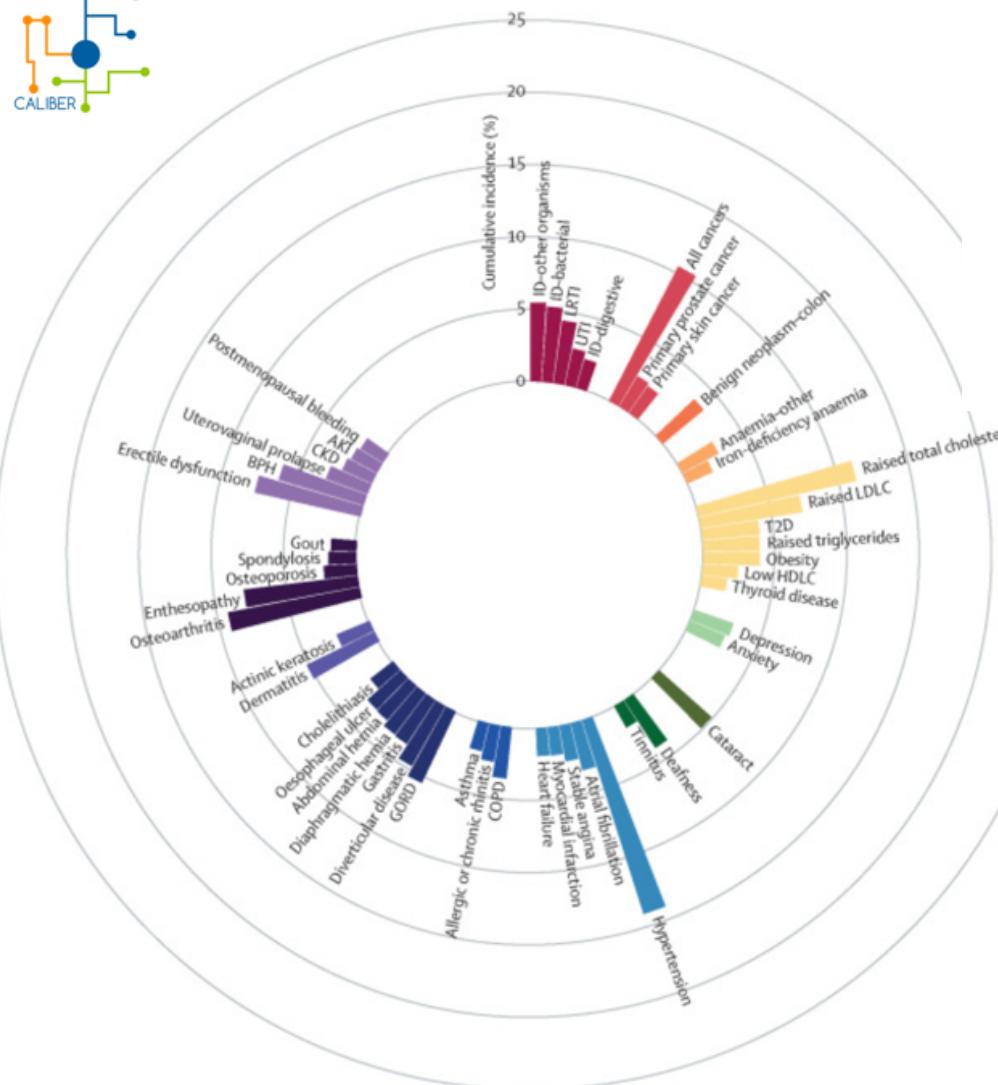
Structured Electronic Health Records



SNOMED-CT

ICD-10 →

I50	Heart failure
	<i>Excl.r complicating:</i>
	<ul style="list-style-type: none">• abortion or ectopic or molar pregnancy (Q00-Q07, Q08.8)• obstetric surgery and procedures (O75.4)
	due to hypertension (I11.0)
	<ul style="list-style-type: none">• with renal disease (I13.-)
	following cardiac surgery or due to presence of cardiac prosthesis (I97.1)
	neonatal cardiac failure (P29.0)
I50.0	Congestive heart failure
	Congestive heart disease
	Right ventricular failure (secondary to left heart failure)
I50.1	Left ventricular failure
	Cardiac asthma
	Left heart failure
	Oedema of lung
	Pulmonary oedema
	with mention of heart disease NOS or heart failure
I50.9	Heart failure, unspecified
	Cardiac, heart or myocardial failure NOS



The Lancet Digital Health

Available online 20 May 2019



Articles

A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service

Life course epidemiology ($n=4M$)

- Longitudinal view of disease, from birth to death
- 371 disease, biomarker and lifestyle risk factor EHR-derived phenotypes
 - Implementation details and logic
 - Validation evidence
 - Use-cases

¹ <https://www.caliberresearch.org/portal/phenotypes/chronological-map>

Using recurrent neural network models for early detection of heart failure onset ⚡

Edward Choi, Andy Schuetz, Walter F Stewart, Jimeng Sun ✉

Journal of the American Medical Informatics Association, Volume 24, Issue 2, March 2017

Pages 361–370, <https://doi.org/10.1093/jamia/ocw112>

Published: 13 August 2016 Article history ▾

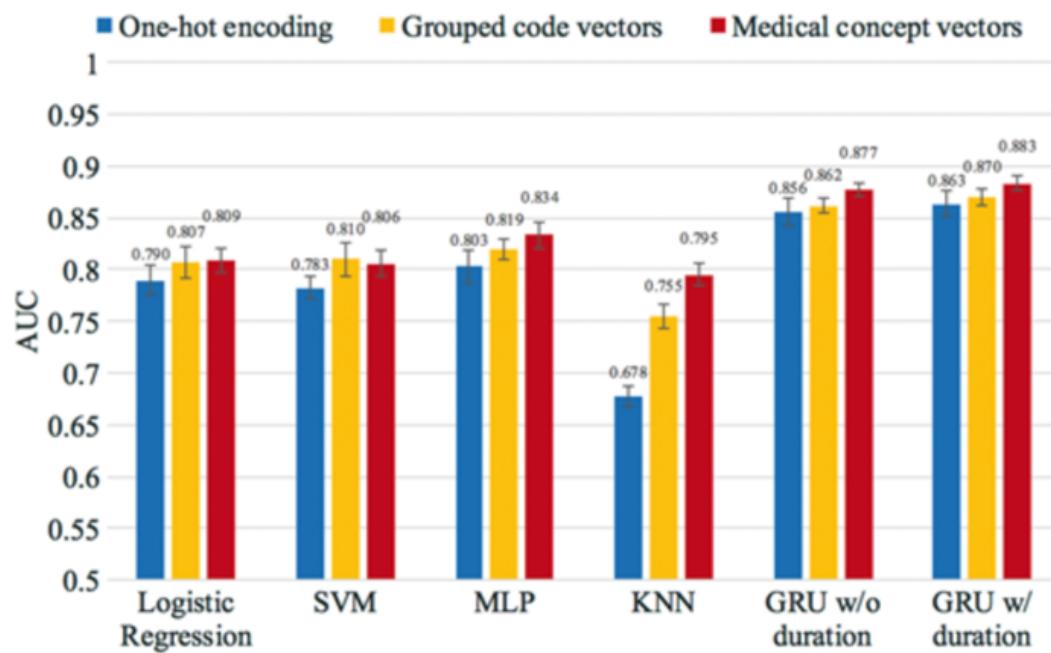


Figure 1.

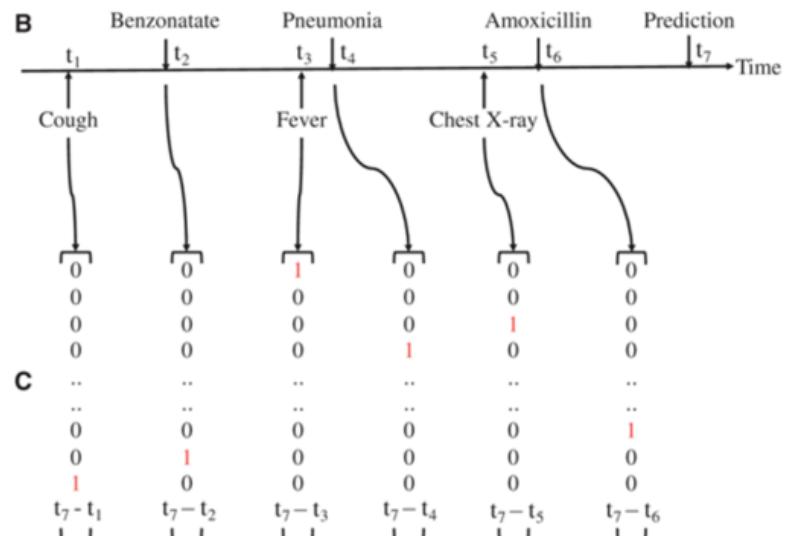
A

N -dimensional vector

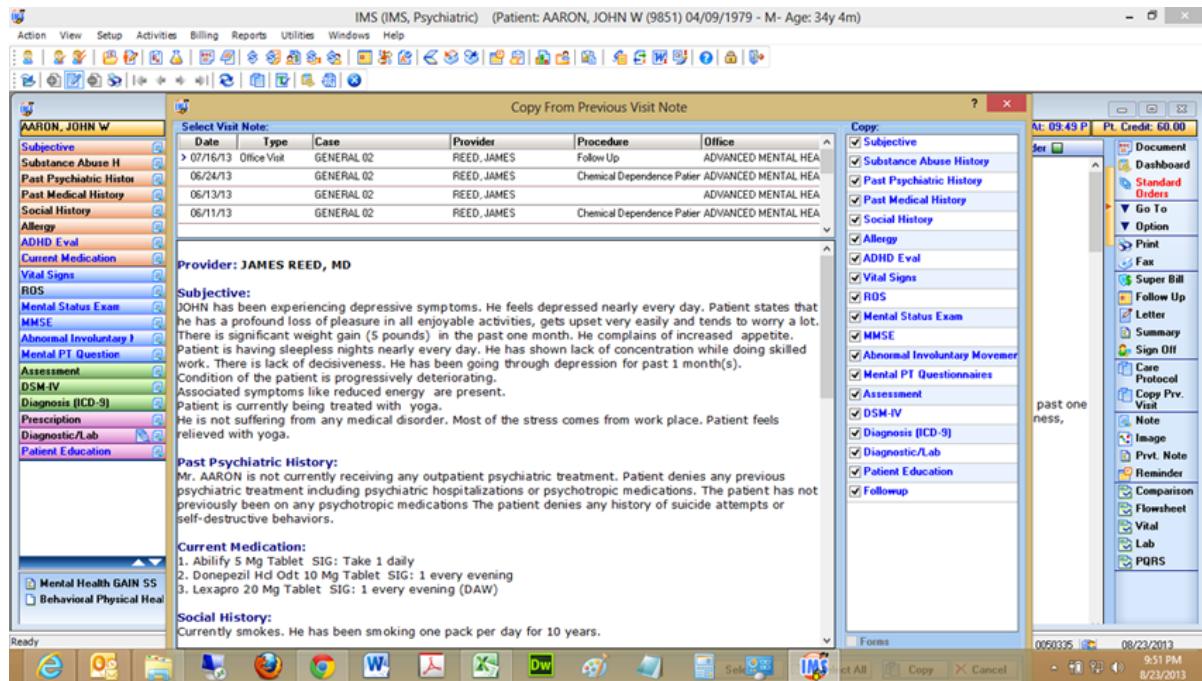
N unique events

Fever: [1, 0, 0, 0, ..., 0, 0, 0]
Lorazepam: [0, 1, 0, 0, ..., 0, 0, 0]
Chest X-ray: [0, 0, 1, 0, ..., 0, 0, 0]
Pneumonia: [0, 0, 0, 1, 0, ..., 0, 0, 0]
⋮
Benzonatate: [0, 0, 0, 0, 0, ..., 0, 1, 0]
Cough: [0, 0, 0, 0, 0, ..., 0, 0, 1]

B



Unstructured Electronic Health Records



76 yo man with h/o HTN, DM, and sleep apnea who presented to the ED complaining of chest pain. He states that the pain began the day before and consisted of a sharp pain that lasted around 30 seconds, followed by a dull pain that would last around 2 minutes. The pain was located over his left chest area somewhat near his shoulder. The onset of pain came while the patient was walking in his home. He did not sit and rest during the pain, but continued to do household chores. Later on in the afternoon he went to the gym where he walked 1 mile on the treadmill, rode the bike for 5 minutes, and swam in the pool. After returning from the gym he did some work out in the yard, cutting back some vines. He did not have any reoccurrences of chest pain while at the gym or later in the evening. The following morning (of his presentation to the ED) he noticed the pain as he was getting out of bed. Once again it was a dull pain, preceded by a short interval of a sharp pain. The patient did experience some tingling in his right arm after the pain ceased. He continued to have several episodes of the pain throughout the morning, so his daughter-in-law decided to take him to the ED around 12:30pm. The painful episodes did not increase in intensity or severity during this time.

Active Computerized Pharmacovigilance Using Natural Language Processing, Statistics, and Electronic Health Records: A Feasibility Study

Xiaoyan Wang, MPhi, George Hripcsak, MD, MS, Marianthi Markatou, PhD,
Carol Friedman, PhD 

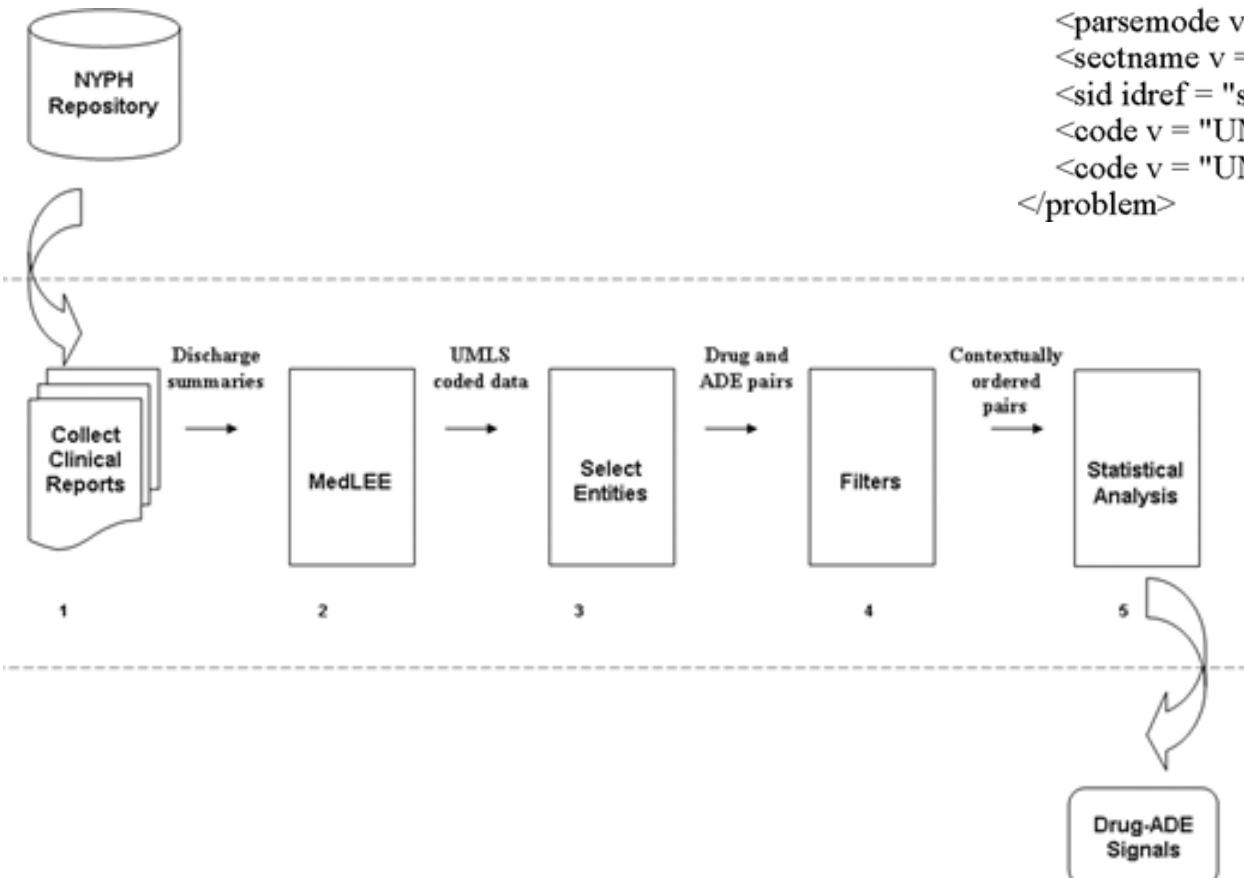
Journal of the American Medical Informatics Association, Volume 16, Issue 3, May 2009,

Pages 328–337, <https://doi.org/10.1197/jamia.M3028>

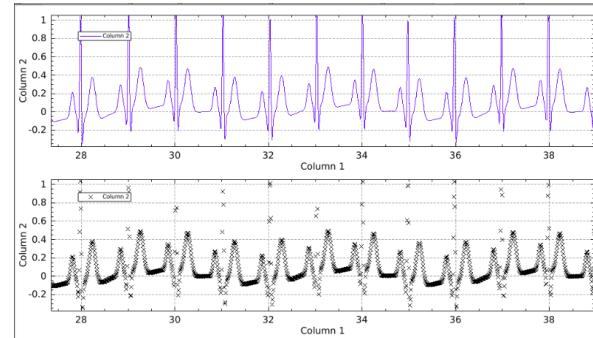
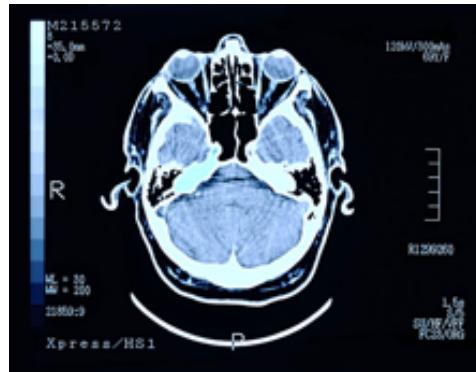
Published: 01 May 2009 Article history ▾

"She has recurring frontal headaches."

```
<problem v = "headache" code = "UMLS:C0018681_Headache" idref
  = "p10">
  <certainty v = "high certainty"></certainty>
  <region v = "front"></region>
  <status v = "recurrence"></status>
  <parsemode v = "model1"></parsemode>
  <sectname v = "report hospital course item"></sectname>
  <sid idref = "s1"></sid>
  <code v = "UMLS:C0239888_Headache recurrent"></code>
  <code v = "UMLS:C0239886_Frontal headache"></code>
</problem>
```



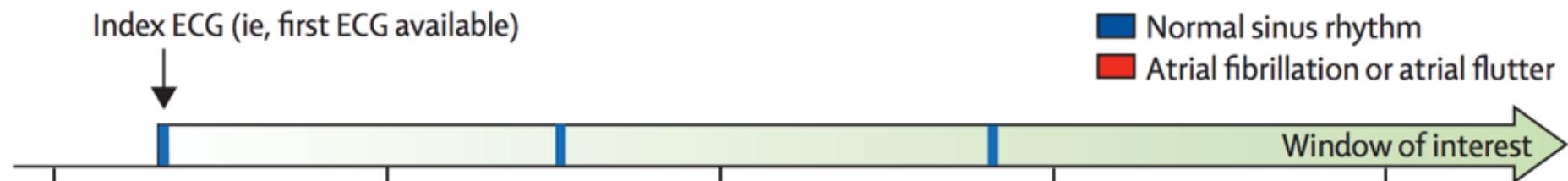
Binary data in Electronic Health Records



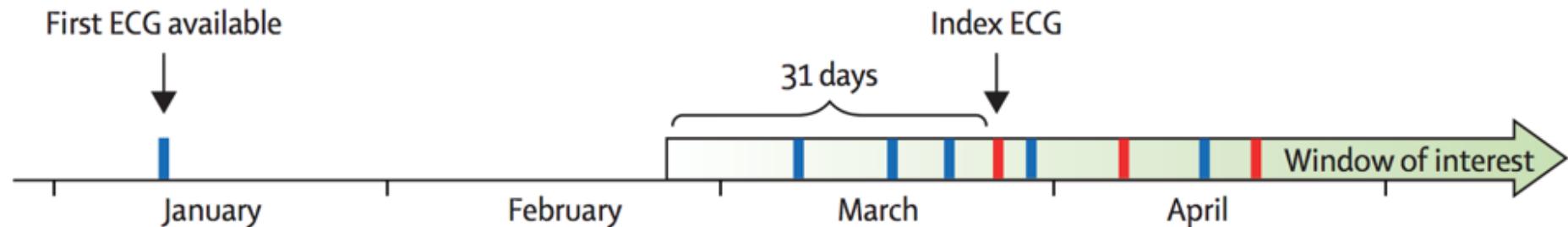
An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction

Zachi I Attia*, Peter A Noseworthy*, Francisco Lopez-Jimenez, Samuel J Asirvatham, Abhishek J Deshmukh, Bernard J Gersh, Rickey E Carter, Xiaoxi Yao, Alejandro A Rabinstein, Brad J Erickson, Suraj Kapa, Paul A Friedman

Patient with no atrial fibrillation rhythms recorded



Patient with at least one atrial fibrillation rhythm recorded



Clinically applicable deep learning for diagnosis and referral in retinal disease

Jeffrey De Fauw, Joseph R. Ledsam, [...] Olaf Ronneberger 

Nature Medicine **24**, 1342–1350(2018) | Cite this article

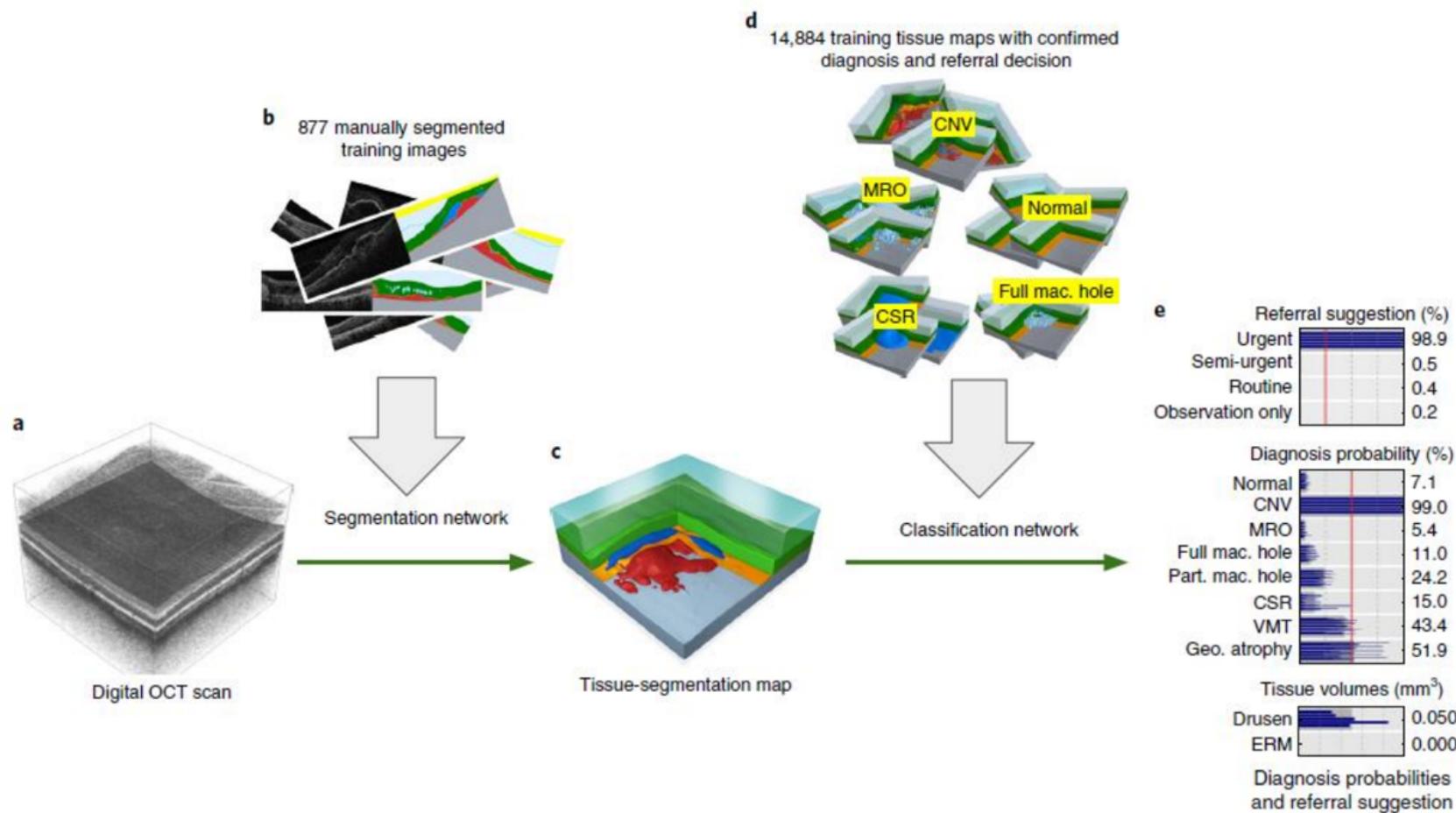
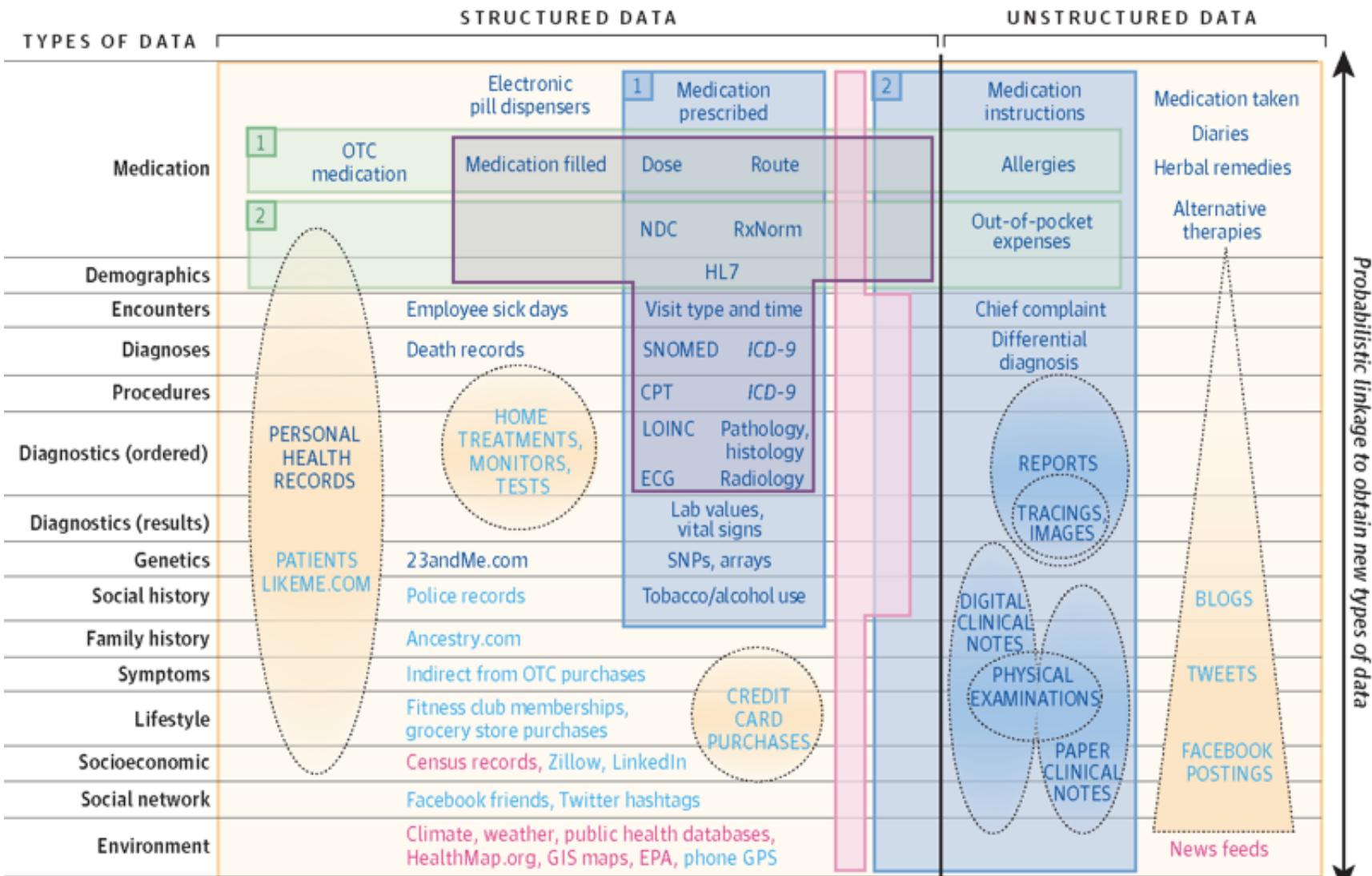


Fig. 1 | Our proposed AI framework. **a**, Raw retinal OCT scan ($6 \times 6 \times 2.3 \text{ mm}^3$ around the macula). **b**, Deep segmentation network, trained with manually segmented OCT scans. **c**, Resulting tissue segmentation map. **d**, Deep classification network, trained with tissue maps with confirmed diagnoses and optimal referral decisions. **e**, Predicted diagnosis probabilities and referral suggestions.



Examples of biomedical data



Pharmacy data



Health care center (electronic health record) data



Claims data



Registry or clinical trial data

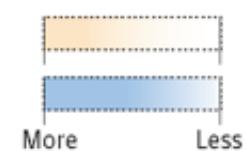


Data outside of health care system

Ability to link data to an individual

- Easier to link to individuals
- Harder to link to individuals
- Only aggregate data exists

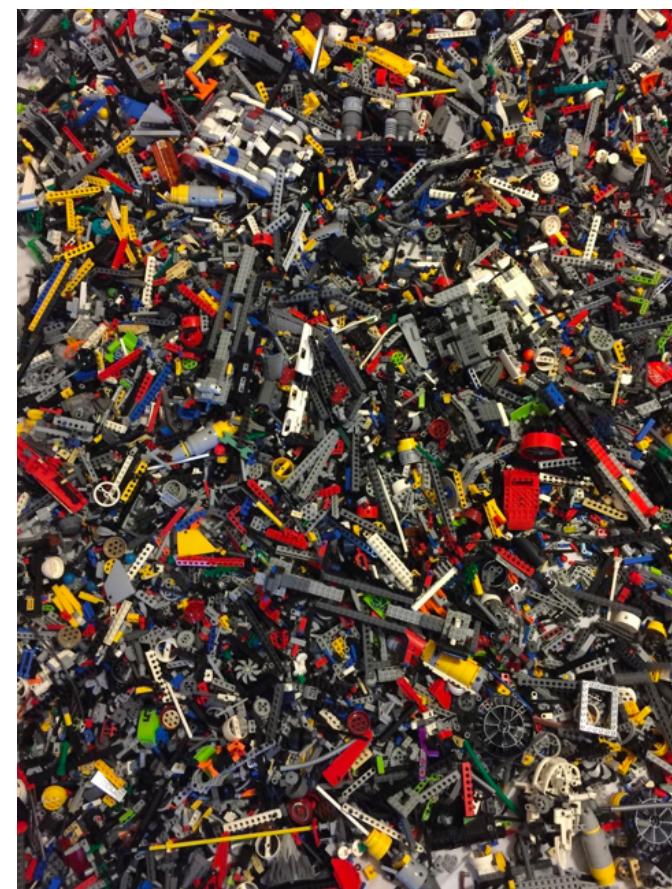
Data quantity



Research data



Electronic Health Records



EHR phenotyping



1001, 2000-01-01, 23,1,NULL,I48
1121, 2013-05-04, 7,1,3,5,14AN.00
1121, 2011-05-21, 81,1,9, G573100
1511, 1993-01-11, 91,1,6,9hF1.00
1511, 199-03-11, 91,1,6, G573100
9913, 2012-05-21, 81,1,9, G573100
67222, 1994-11-01,1234,1,3,7L1H300
1001, 1994-08-11,1234,1,3,7L1H300
1001, 1993-01-01, 253,1,1,793Mz00
1231, 2012-03-03, 23,1,123,K65
1121, 2013-05-04, 7,1,3,5,14AN.00
1121, 2011-05-21, 81,1,9, G573100
1511, 1993-01-11, 91,1,6,9hF1.00
1511, 199-03-11, 91,1,6, G573100
9913, 2012-05-21, 81,1,9, G573100
67222, 1994-11-01,1234,1,3,7L1H300
67222, 1995-12-21,1234,1,3,7L1H300
67222, 1991-03-03,1234,1,3,7L1H310
682444, 1993-01-01, 253,1,1,793Mz00
1121, 2013-05-04, 7,1,3,5,14AN.00
1121, 2011-05-21, 81,1,9, G573100
1511, 1993-01-11, 91,1,6,9hF1.00
1511, 199-03-11, 91,1,6, G573100
67222, 1995-12-21,1234,1,3,7L1H300
67222, 1994-11-01,1234,1,3,7L1H300

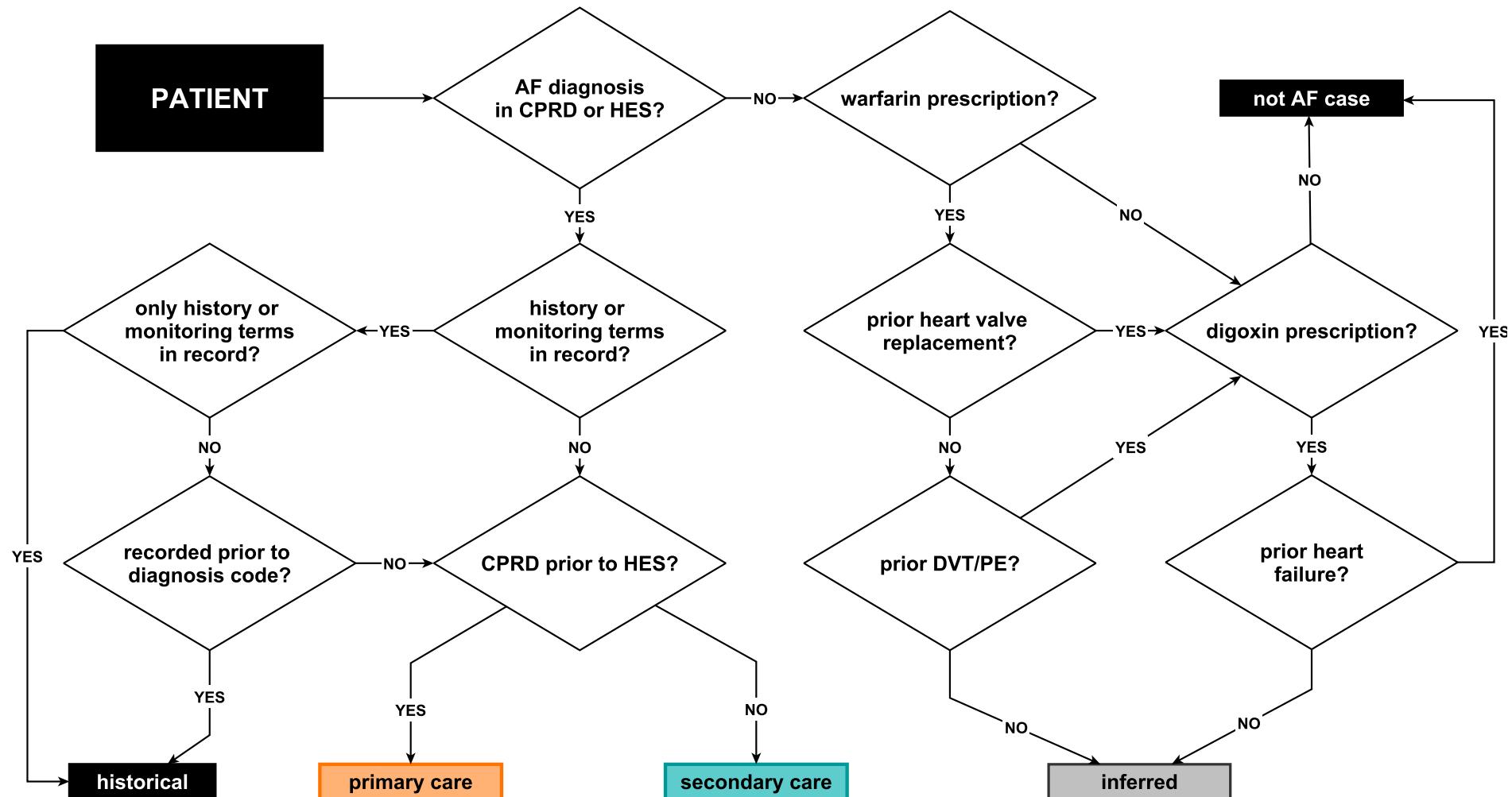


Validation

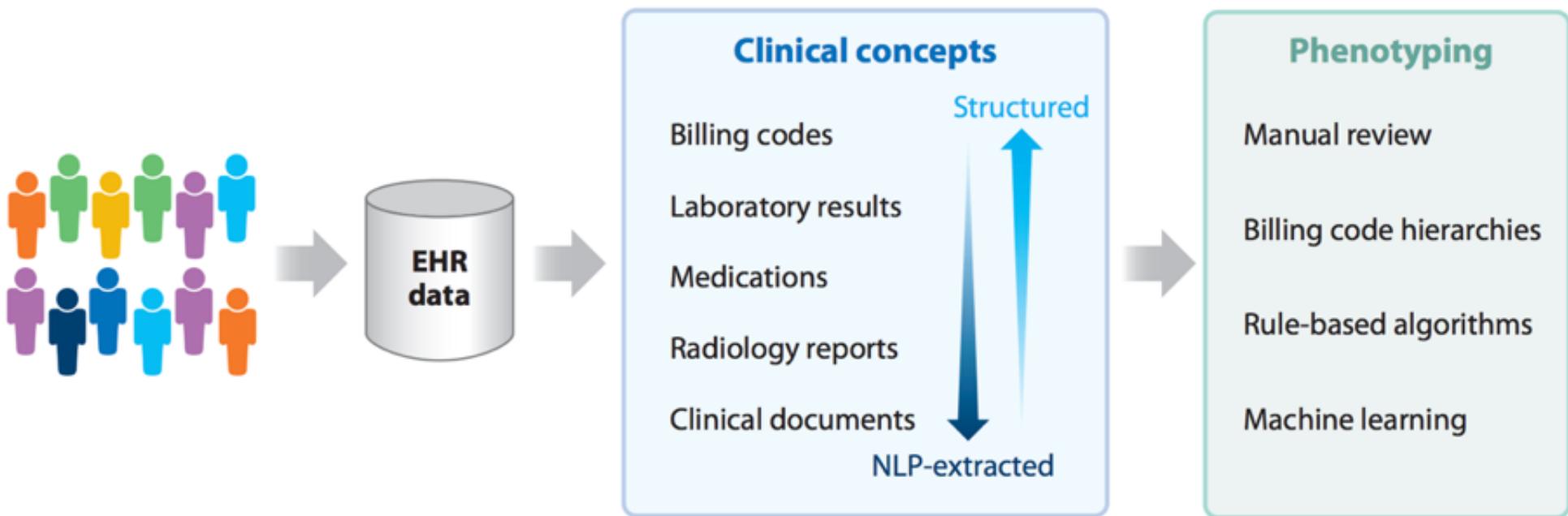
- Prognosis
- Aetiology
- Genetic
- Cross-source
- External replication
- Case-note review

**Disease status: severity,
onset, certainty
Lifestyle risk factor
Biomarker**

EHR phenotyping



EHR Phenotyping



End of Part 2

- **Summary**
 - Text data vs binary data
 - Data types
 - Structured
 - Unstructured
 - imaging, sound, etc
 - Data linkage
 - Phenotyping
- **Questions?**



5 minutes break

Part 3: Controlled clinical terminologies

Health terminology

- Greek->Arabic->Latin->modern languages
- **Complex and multi-faceted**
 - Acronyms (MI, TIA, DM)
 - Synonyms (heart attack, myocardial infarction, AMI)
 - Polysemy (knee -> joint or anatomical angle)
 - Contextual (infarction -> myocardial or cerebral)
 - Absence of generally agreed definitions

Health terminology



1. *Malignant neoplasm of breast*
2. *Breast carcinoma*
3. *Female breast carcinoma*
4. *Noninfiltrating intraductal carcinoma*
5. *Mammary ductal carcinoma*
6. *Carcinoma breast*
7. *Breast tumour*
8. *Breast tumor*
9. *Breast malignant neoplasm*
10. *Breast malignant tumor*
11. *Cancer of breast*
12. *Mammary cancer [...]*

How do we describe human disease?

- Endocrine and metabolic diseases
- Infectious diseases
- Cancers
- Behavioral diseases
- Eye disease
- Circulatory disease
- Respiratory disease
- Skin disease
- [...]

How do we describe human disease?

- Endocrine and metabolic diseases 
 - Thyroid diseases
 - Diabetes
 - Obesity
 - Hypoglycaemia
 - [...]
- Infectious diseases
- Cancers
- Behavioral diseases
- Eye disease
- Circulatory disease
- Respiratory disease
- Skin disease
- [...]

How do we describe human disease?

- Endocrine and metabolic diseases 
 - Thyroid diseases
 - Diabetes 
 - Obesity
 - Hypoglycaemia
 - [...]
- Infectious diseases
- Cancers
- Behavioral diseases
- Eye disease
- Circulatory disease
- Respiratory disease
- Skin disease
- [...]

How do we describe human disease?

- Endocrine and metabolic diseases 

 - Thyroid diseases
 - Diabetes 
 - Type 1
 - Type 2
 - Diabetic retinopathy
 - Obesity
 - Hypoglycaemia
 - [...]

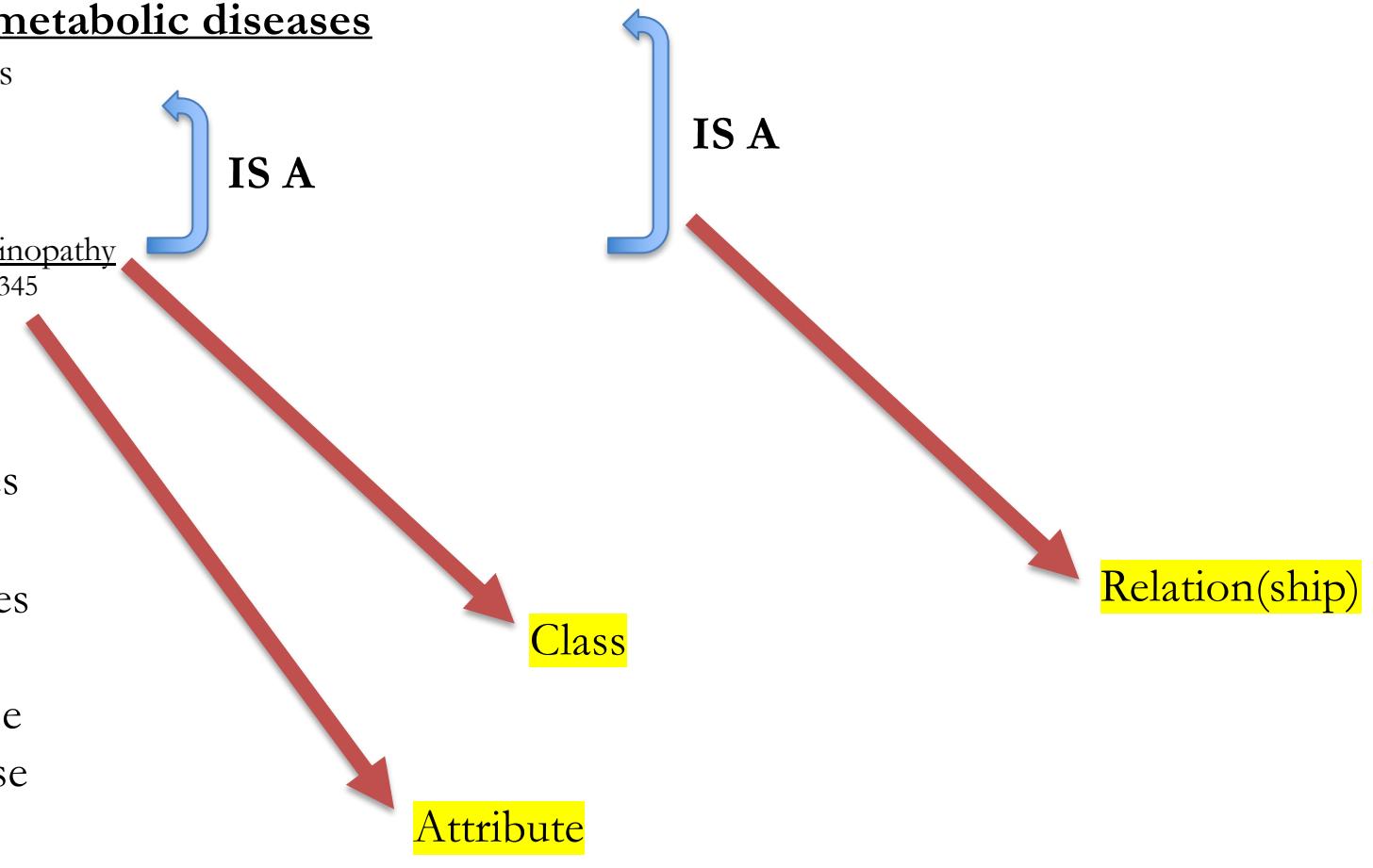
- Infectious diseases
- Cancers
- Behavioral diseases
- Eye disease
- Circulatory disease
- Respiratory disease
- Skin disease
- [...]

How do we describe human disease?

- Endocrine and metabolic diseases 
 - Thyroid diseases
 - Diabetes 
 - Type 1
 - Type 2
 - Diabetic retinopathy
 - Obesity
 - Hypoglycaemia
 - [...]
- Infectious diseases
- Cancers
- Behavioral diseases
- Eye disease
- Circulatory disease
- Respiratory disease
- Skin disease
- [...]

How do we describe human disease?

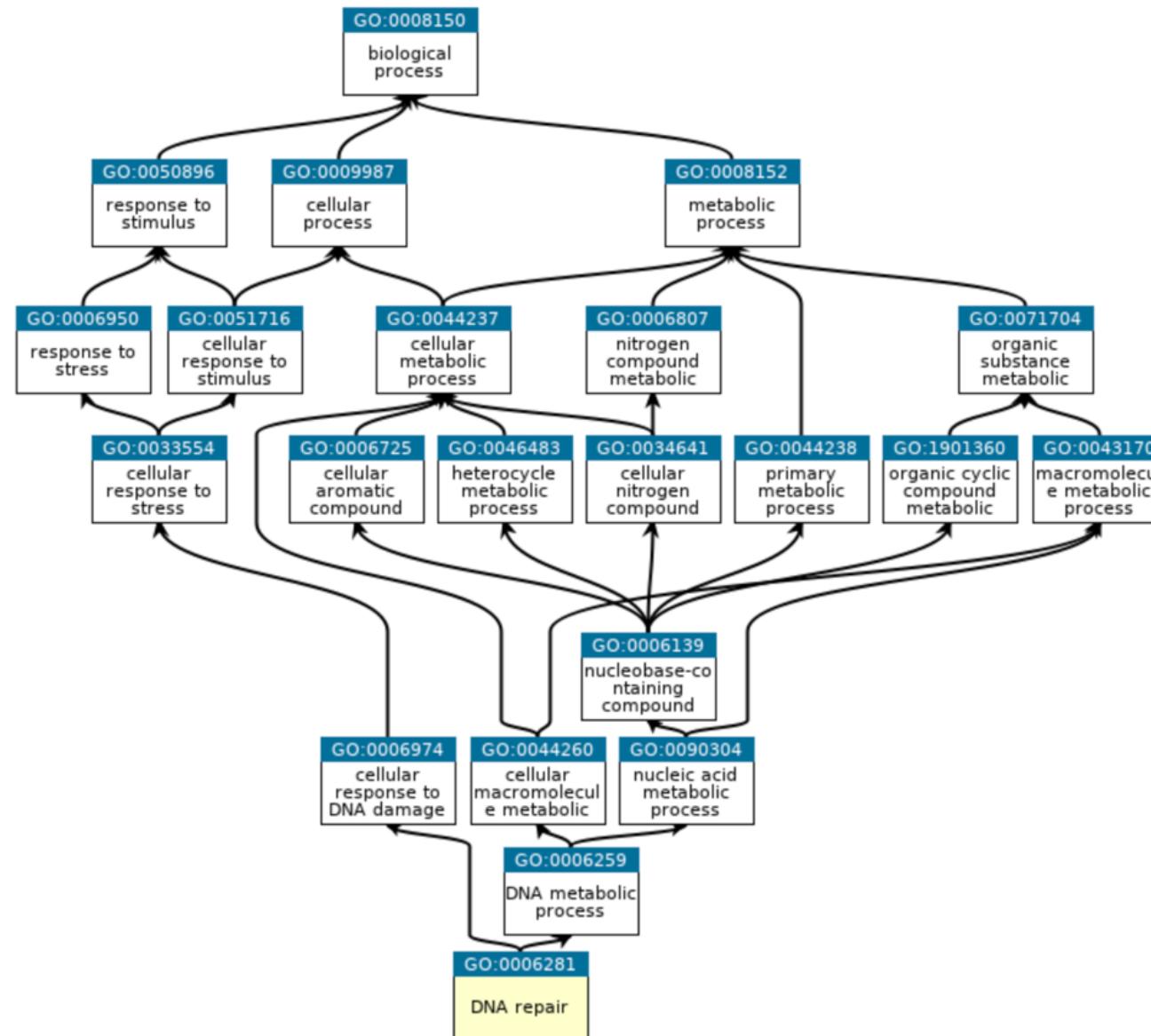
- Endocrine and metabolic diseases
 - Thyroid diseases
 - Diabetes
 - Type 1
 - Type 2
 - Diabetic retinopathy
 - ID12345
 - Obesity
 - Hypoglycaemia
 - [...]
- Infectious diseases
- Cancers
- Behavioral diseases
- Eye disease
- Circulatory disease
- Respiratory disease
- Skin disease
- [...]



Ontology

- “A **formal specification** of a conceptualization” (Tom Gruber)
 - Formal representation of classes, their attributes and the relationships between them
- **The main purpose is knowledge**
 - Specification mechanism
 - Knowledge sharing
 - Knowledge re-use
 - Machine-readable storage format

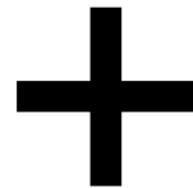
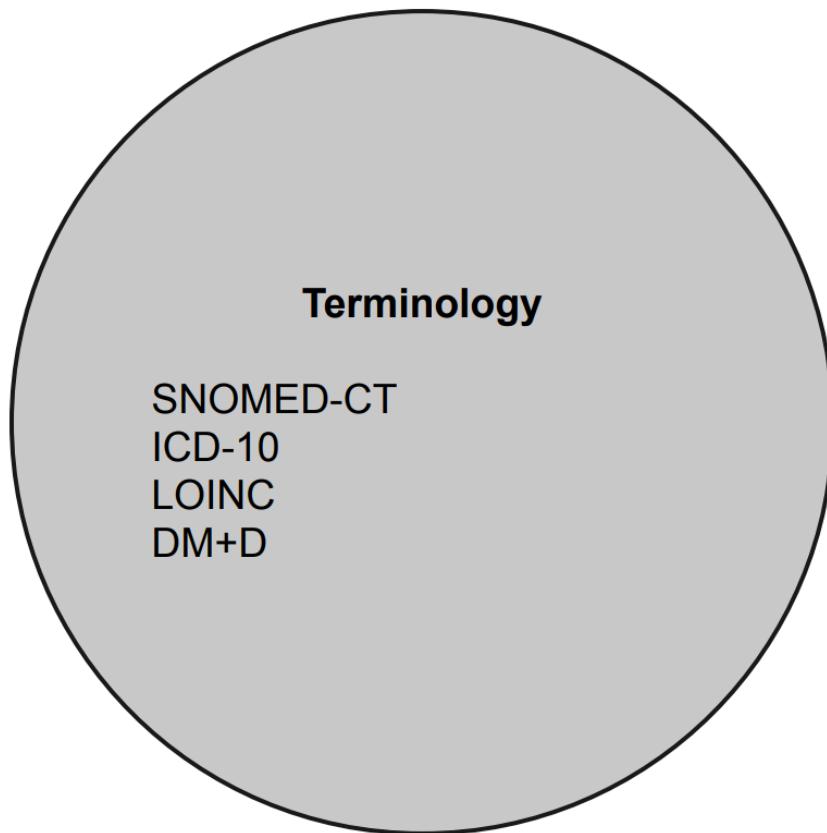
Example: The Gene Ontology



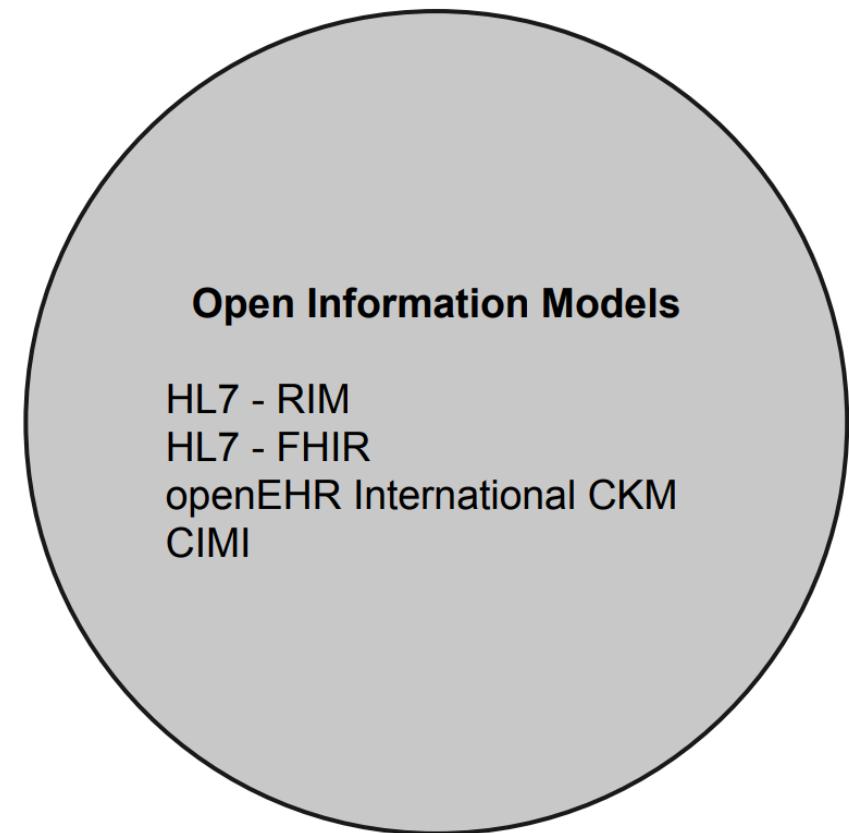
Clinical ontologies

- **Clinical ontologies** concerns the meaning, expression and use of concepts in statements in the medical record or other clinical information system
- A **controlled clinical ontologies aka controlled clinical terminology** is a system which allows the regulated and constrained expression of clinical concepts

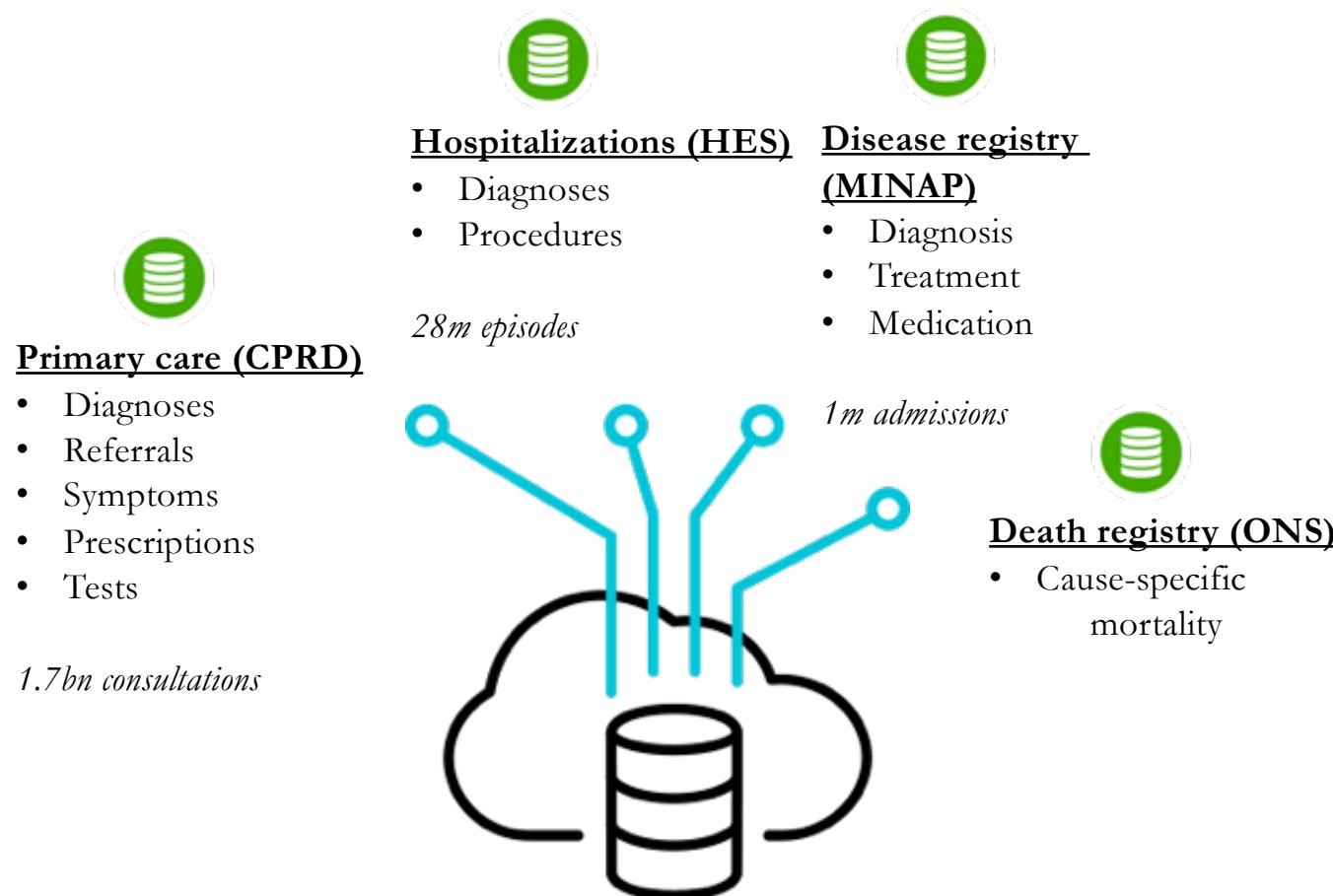
Terminologies are
used to
define healthcare
information



Information
models are **used to**
transport
healthcare
information



CALIBER: four national EHR sources, 10m patients



Primary care EHR

- **Clinical Practice Research Datalink (CPRD)**
- 19 million patients from 738 practices
- Longitudinal data recorded in GP practices
 - Diagnoses
 - Medication
 - Test results (e.g. white blood cell count)
 - Referrals
 - Examinations (e.g. blood pressure)



patid	eventdate	readcode	data1	data2
1001	2011-12-01	246..00	67	67
1001	2012-06-05	G30..14		
1001	2012-06-05	8H2Z.00		
2221	2011-12-01	M211.12		
2221	2011-12-01	246..00	80	110
2221	2013-11-02	246..00	69	128
3241	2009-11-03	H26..00		
7810	2014-01-12	171..00		
7810	2014-01-12	H06z011		

G30..14
Heart
Attack

246.00
Blood
pressure



G30..14



80

110

69

128

Case study: Read codes

	Read 2	Clinical Terms 3
Length	5 bytes	5 bytes
Classes	~82,000	~298,000
Relationships	Encoded in the ontology Single parent	Encoded using separate relationship table Polyhierarchy

- [D] Symptoms, signs and ill-defined conditions
- [X] External causes of morbidity and mortality
- Administration
- Causes of injury and poisoning
- Circulatory system diseases
 - Acute rheumatic fever
 - Arterial, arteriole and capillary disease
 - Cerebrovascular disease
 - Chronic rheumatic heart disease
 - Circulatory system disease NOS
 - Herdity angpthy wth nphrpthy, aneurysm, mscl cramps syndr
 - Hypertensive disease
 - Ischaemic heart disease
 - Acute myocardial infarction
 - Angina pectoris
 - Angina decubitus
 - Angina pectoris NOS
 - Angina on effort
 - Angina pectoris NOS
 - Ischaemic chest pain
 - New onset angina
 - Post infarct angina
 - Stable angina**
 - Status anginosus
 - Stenocardia
 - Syncope anginosa
 - Coronary artery spasm
 - Prinzmetal's angina
 - Cardiac syndrome X
 - Certain current complication follow acute myocardial infarct
 - Coronary microvascular disease
 - Ischaemic heart disease NOS
 - Old myocardial infarction
 - Other acute and subacute ischaemic heart disease
 - Other chronic ischaemic heart disease
 - Other specified ischaemic heart disease
 - Postoperative myocardial infarction
 - Subsequent myocardial infarction
 - Other forms of heart disease
 - Other specified diseases of circulatory system
 - Pulmonary circulation diseases
 - Rheumatic heart disease
 - Vascular calcification
 - Vein, lymphatic and circulatory diseases NOS

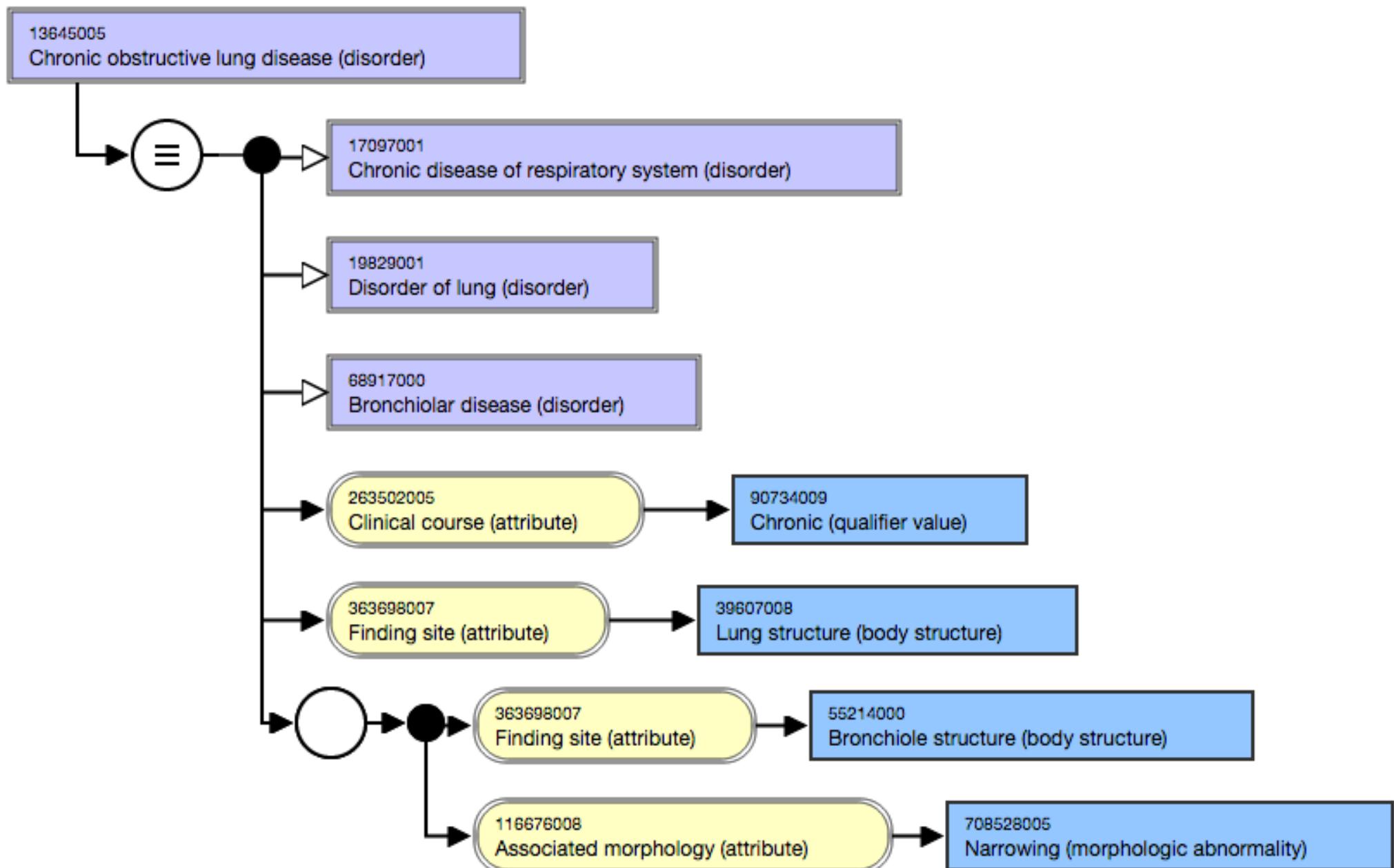
Preferred Name	Stable angina
ID	http://purl.bioontology.org/ontology/RCTV2/G33z700
label	Stable angina
prefLabel	Stable angina
subClassOf	Angina pectoris NOS

Additional values
Administration
Anatomical concepts
Appliances+equipment
Attribute
Causes of injury and poisoning
Clinical findings
Disorders
Accidents, poisoning and injury
Cardiovascular disorder
Arteriovenous disorder
Blood vessel tumour
Cardiovascular infection
Congenital cardiovascular disorder
Disorder of artery
Disorder of heart
Athlete's heart
Cardiac arrest
Cardiac transplant disorder
Cardiomegaly
Chronic cor pulmonale
Disorder of endocardium and heart val
Disorder of heart conduction
Disorder of heart rhythm
Disorder of myocardium
Disorder of pericardium
Heart failure
Hypertensive heart disease
Injury of heart
Innocent murmur
Ischaemic heart disease
Aborted myocardial infarction
Aneurysm of heart
Angina
Angina at rest
Exercise-induced angina
Stable angina
Unstable angina
Cardiac rupture after acute myocardi
Cardiac syndrome X
Coronary spasm
Double coronary vessel disease
Myocardial infarction
Myocardial ischaemia
Post-infarction ventricular septal def
Single coronary vessel disease
Triple vessel disease of the heart

Preferred Name	Stable angina
ID	http://purl.bioontology.org/ontology/RCD/X2008
cui	C0340288
	Unstable angina
Inverse of SIB	Exercise-induced angina Angina at rest
notation	X2008
prefLabel	Stable angina
RID	Y201y
tui	T047
subClassOf	Angina

Case study: SNOMED-CT

- **Myocardial infarction (disorder)**
 - ID: 22298006
 - Synonyms: Cardiac infarction, heart attack, infarction of heart, MI – Myocardial infarction, Myocardial infarct, Myocardial infarction
- Acyclic hierarchy, multiple parent nodes:
 - Ischemic heart disease (disorder)
 - Myocardial disease (disorder)
 - Myocardial necrosis (finding)
 - Necrosis of anatomical site (disorder)



SNOMED-CT descriptions

- The descriptions provided for concept id. 22298006 are:
 - Fully specified name: Myocardial Infarction (disorder)
 - Preferred term: Myocardial Infarction
 - Synonym: Cardiac infarction
 - Synonym: Heart attack
 - Synonym: Infarction of heart
 - Synonym: Infarto de miocardio

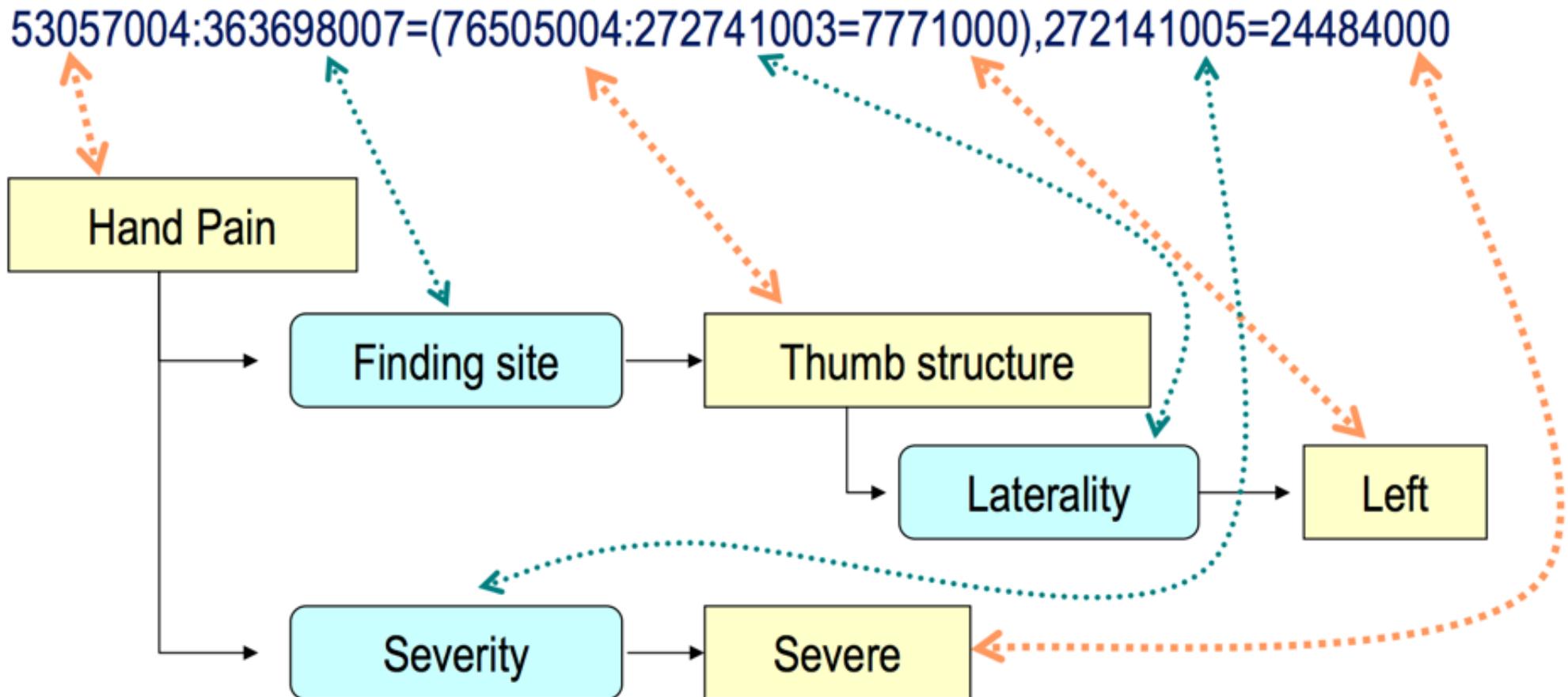
SNOMED-CT concept definition

- Allergic rhinitis due to pollen
 - IS-A
 - allergic rhinitis
 - HAS-FINDING-SITE
 - nose
 - HAS-ASSOCIATED-MORPHOLOGY
 - inflammation
 - HAS-CAUSATIVE-AGENT
 - pollen

Post-coordination

- Having a term for every single possible concept is not feasible...!
- Grouping concepts in abstract concepts is not optimal either -> information loss
- Post-coordination
 - Defined using a robust language (SNOMED-CT Compositional Grammar)
 - Combine multiple existing defined concepts into more complex concepts

New concepts: severe pain of the left thumb



Secondary care EHR

- **Hospital Episode Statistics (HES)**
- Data collected from all inpatient, outpatient, and A&E encounters in England
 - Diagnoses, procedures
 - Demographics
 - Administrative (waiting time, admission method, discharge destination)
 - Geographical information (home vs care)
 - Clinical specialties involved
 - Outcomes



patid	admi_date	disc_date	icd
1001	2011-12-01	2011-12-03	I210*
1001	2012-06-05	2012-06-07	E119
1001	2012-06-05	2012-06-07	S729
2221	2011-12-01	2011-12-01	S322
2221	2011-12-01	2011-12-01	I259
2221	2013-11-02	2011-12-01	I48X
3241	2009-11-03	2009-11-13	I210*
7810	2014-01-12	2014-01-19	C189*

I21.0 Acute Myocardial Infarction

Case study: ICD

- International Statistical Classification of Diseases and Related Health Problems (ICD)
- Statistical classification system maintained by the WHO
 - Used payments, mortality and morbidity statistics
- Most widely used system in the world
 - Currently 10th revision (ICD-10)
 - Many countries develop their own extensions and modifications (ICD-9-CM)

A generall Bill for this present year,
ending the 19 of December 1665. according to.
the Report made to the KINGs most Excellent Majestie.

By the Company of Parish Clerks of London, &c.

<i>Buried Pds.</i>	<i>Buried Pds.</i>	<i>Buried Pds.</i>	<i>Buried Pds.</i>
51 Albany Woodstreet	121	51 Clements Eastcheap	120
51 Althamew Earling	124	51 Dennis Back-church	120
51 Althamew Tordis	16	51 Dunstans East	165
51 Althamew Great	455	51 Edwards Lombard	70
51 Althamew Hosua	10	51 Ethelburgh	105
51 Althamew Leafe	119	51 Faisks	104
51 Althamew Lambardis	62	51 Follers	105
51 Althamew Staining	155	51 Gabiel Fen-church	144
51 Althamew the Wall	100	51 George Boleynland	41
51 Alphege	171	51 Goronys by Pauls	176
51 Andrew Hubbard	71	51 Hellens	108
51 Andrew Wadethan	174	51 Janet Dukes place	190
51 Andrew Wardene	426	51 James Gudchiche	189
51 Anne Aldesgate	153	51 John Bayhill	19
51 Anne Blache-fours	652	51 John Fawcett	9
51 Anthonis Paish	53	51 John Fawcett	15
51 Auline Parish	41	51 John Zacheus	81
51 Bathsheba Coleman	10	51 Katharine Coleman	199
51 Barthol. Exchange	71	51 Katharine Creech	135
51 Bennet Fyack	47	51 Lawrence Jewell	94
51 Benn. Grace-Church	57	51 Lawrence Poultney	140
51 Beanes Paul Wharf	155	51 Leonard Radcliffe	43
51 Bennet Sherehing	11	51 Leonard Radcliffe	171
51 Bessop Billinggate	51	51 Maudlin Milkstreet	44
51 Chiche Church	653	51 Magnus Paish	101
51 Christopher	66	51 Margaret Lothrop	100
	17	51 Michael Bassifull	66

Buried in the 97 Parishes within the walls, 15207 Whereof, of the Plague 9857

51 Andrew Holborn	1913	3103	Bridewell Precinct	130	179	51 Dunstans Well	—	98	166	51 Savories Southward	4216	1446
51 Bartholomew Great	491	344	51 Bessop Billinggate	997	715	51 George Southward	1013	1200	51 Sepulchres Paish	4509	3746	
51 Bartholomew Leffe	93	159	51 Bessop Algate	4936	4051	51 Godes Cripplegate	2069	4275	51 Thomas Southward	375	371	
51 Bridget	211	147	51 Bessop Bishopsgate	3464	1500	51 Olares Southward	4793	3785	51 Trinity Minories	268	133	

Buried in the 16 Parishes without the walls, 41351 Whereof, of the Plague 28885 At the Pesthouse 359 356

51 Godes in the Field	447	73165	Katherine Tower	956	601	51 Magdalen Benyon	794	116	51 Mary Vinenchappel	4766	555
Hatton Parish	113	151	Lambeth Paish	798	537	51 Mary Newington	121	100	51 Redriffle Paish	104	110
51 James Clarkeworth	863	1377	51 Leonard Shoreditch	1669	1949	51 Mary Ellington	895	593	51 Roperty Paish	598	583

Buried in the 9 Parishes in the City and Liberties of Westminster, 21194 Whereof, of the Plague 8403 Whereof, of the Plague 68596

51 Clement Danse	1669	11110	51 Mary Saway	103	1987	The Total of all the Christnings	—	9967
51 Paul Cowne Gardin	405	151	51 Margaret Wellmire	1470	1741	The Total of all the Burials thi year	—	97306
51 Martinis in the Field	4784	2588	51 Servis at the Pesthouse	150		Whereof, of the Plague	—	68596

Buried in the 9 Parishes in the City and Liberties of Westminster, 21194 Whereof, of the Plague 8403 Whereof, of the Plague 68596

A Bonvive and Stilboorne	617	Executed	—	21	Palsie	—	30
Aged	1545	Flox and Small Pox	—	655	Plague	—	68596
Ague and Feaver	5157	Found dead in streets, fields, &c. &c.	20	Plannet	—	6	
Appoplex and Suddenly	116	French Pox	—	86	Plunzie	—	15
Bedeid	10	Frighted	—	23	Poysoned	—	1
Blaſted	5	Gout and Sciatica	—	27	Quinsie	—	35
Bleeding	16	Grief	—	46	Ricketts	—	557
Bloody Flux, Scowring & Flux	185	Griping in the Guts	—	1288	Rising of the Lighes	—	397
Burnt and Scalded	8	Hangd & made away themselfes	7	27	Rupture	—	34
Calenture	3	Headmouldiohe & Mouldfallen	14	Scurvy	—	105	
Cancer, Gangrene and Fistula	56	Jaundies	—	110	Shingles and Swine pox	—	2
Canker, and Thrush	111	Impostume	—	227	Sores, Ulcers, broken and bruised	—	
Childbed	625	Kild by severall accideoes	—	46	Limbs	—	82
Chrifomes and Infants	1258	Kings Evill	—	86	Spleen	—	14
Cold and Cough	68	Leprosie	—	2	Spotted Feaver and Purples	1929	
Collick and Winde	134	Lethargy	—	14	Stopping of the stomack	—	332
Consumption and Tiflick	4808	Livergrown	—	26	Stone and Strangury	—	98
Convallion and Mother	2036	Meagron and Headach	—	27	Surfer	—	1252
Distracted	5	Mealies	—	7	Teeth and Wonns	—	264
Dropie and Timpany	1478	Murthered and Shoe	—	9	Vomiting	—	51
Drowned	50	Overlaid and Starved	—	45	Vvnh	—	5
<i>Males</i>	51142	<i>Males</i>	—	48569			
<i>Christned Females</i>	48533	<i>Buried Females</i>	—	8737	<i>Of the Plague</i>	—	68596
<i>In all</i>	99671	<i>In all</i>	—	97306			

*Increased in the Burials in the 130 Parishes under the Pest-house this year 79009
Increased of the Plague in the 130 Parishes and at the Pest-house this year 68596*

Case study: ICD

- **Classifying disease**
 - C50 Malignant neoplasm of breast
- **Signs and symptoms**
 - R23.2 Flushing
- **Abnormal investigation findings**
 - R78.6 Finding of steroid agent in blood
- **Interactions with the healthcare system**
 - Z29.0 Isolation
- **External causes of injury or disease**
 - X84 Intentional self-harm by unspecified means

Case study: ICD

- Chapter II –
Neoplasms (chapter)

- ▶ I Certain infectious and parasitic diseases
- ▶ II Neoplasms
- ▶ III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
- ▶ IV Endocrine, nutritional and metabolic diseases
- ▶ V Mental and behavioural disorders
- ▶ VI Diseases of the nervous system
- ▶ VII Diseases of the eye and adnexa
- ▶ VIII Diseases of the ear and mastoid process
- ▶ IX Diseases of the circulatory system
- ▶ X Diseases of the respiratory system
- ▶ XI Diseases of the digestive system
- ▶ XII Diseases of the skin and subcutaneous tissue
- ▶ XIII Diseases of the musculoskeletal system and connective tissue
- ▶ XIV Diseases of the genitourinary system
- ▶ XV Pregnancy, childbirth and the puerperium
- ▶ XVI Certain conditions originating in the perinatal period
- ▶ XVII Congenital malformations, deformations and chromosomal abnormalities
- ▶ XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
- ▶ XIX Injury, poisoning and certain other consequences of external causes
- ▶ XX External causes of morbidity and mortality
- ▶ XXI Factors influencing health status and contact with health services
- ▶ XXII Codes for special purposes

Case study: ICD

- Chapter II – Neoplasms (**chapter**)
 - C00-C97 Malignant neoplasms (**block**)
 - C50 Malignant neoplasm of breast
 - C50.0 Nipple and areola
 - C50.1 Central portion of breast
 - **C50.2 Upper-inner quadrant of breast (term)**
 - C50.3 Lower-inner quadrant of breast
 - C50.4 Upper-outer quadrant of breast
 - C50.5 Lower-outer quadrant of breast
 - C50.6 Axillary tail of breast
 - C50.8 Overlapping lesion of breast
 - C50.9 Breast, unspecified

Case study: ICD

- F84 – Pervasive Developmental Disorders
- 3 character code
 - Definition
 - Diagnostic guidelines
 - Inclusion criteria
 - Exclusion criteria

ICD10 F84.5 Asperger Syndrome

- A disorder of uncertain nosological validity, characterized by the same type of qualitative abnormalities of reciprocal social interaction that typify autism, together with a restricted, stereotyped, repetitive repertoire of interests and activities. It differs from autism primarily in the fact that there is no general delay or retardation in language or in cognitive development. This disorder is often associated with marked clumsiness. There is a strong tendency for the abnormalities to persist into adolescence and adult life. Psychotic episodes occasionally occur in early adult life.
- Autistic psychopathy
- Schizoid disorder of childhood

More complex structures

S	8	6
---	---	---

Category

S = Injuries, poisoning and certain other consequences of external causes related to single body regions

S86 = Injury of muscle, fascia and tendon at lower leg

0	1	1
---	---	---

Etiology, Anatomic Site,
Severity, Other Vital Details

S86.0 = Injury of Achilles tendon

S86.01 = Strain of Achilles tendon

S86.011 = Strain of right Achilles tendon

D

Extension

Primarily used to document episode of care for injuries and other conditions with external causes

A = Initial Encounter

D = Subsequent Encounter

S = Sequela

S52 Fracture of forearm

S52.5 Fracture of lower end of radius

S52.52 Torus fracture of lower end of radius

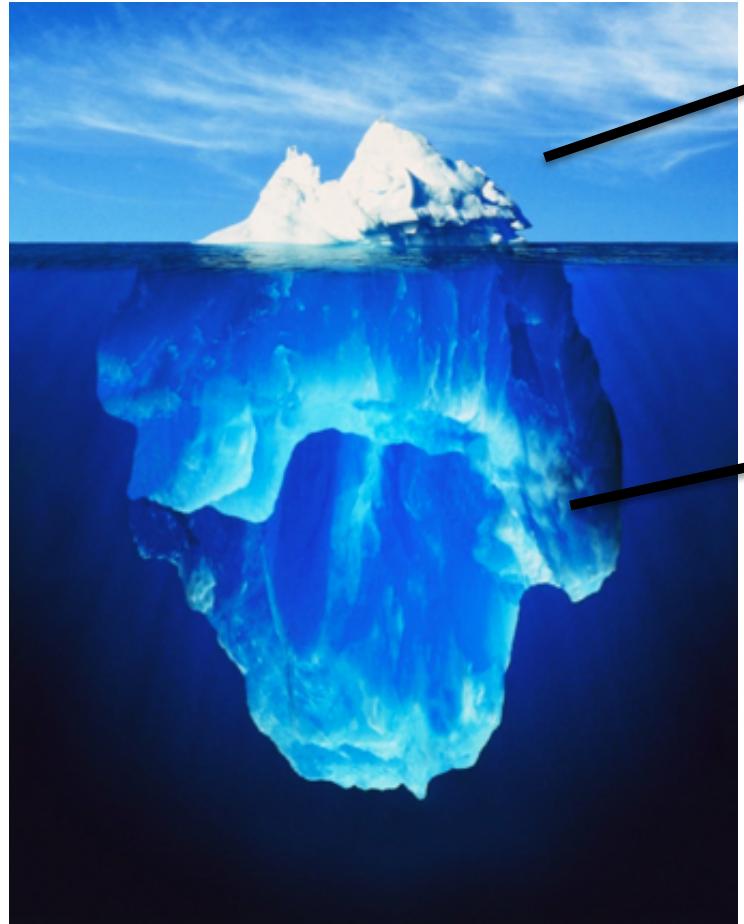
S52.521 Torus fracture of lower end of right radius

S52.521A Torus fracture of lower end of right radius, initial encounter, closed fracture

ICD-10: issues

- **Limited scope**
 - Classification of diseases
- **Constrained** structure makes it hard to update
 - Fixed-depth
- Use of “Not Otherwise Specified”/”Not Elsewhere Classified”
 - J16 - Pneumonia due to other infectious organisms, not elsewhere classified
 - R52 – Pain not elsewhere classified

ICD-10 codes, great for admin but not so great for research



I21.9 Acute myocardial infarction,
unspecified

ECG
Blood pressure
Laboratory measurements
Medication
Diagnostic imaging
Interventions
[...]

Clinical coding

- The **process** by which clinical notes generated during patient care are analyzed, abstracted and assigned standard codes using a medical classification system
- Usually performed by non-clinically trained staff
 - Examine, sort, analyze and summarize evidence
- Process is dictated by guidelines and policies
 - and confounded by other factors

Clinical coding

DCS.IX.4: Myocardial infarction (I21, I22, I25.8)

The time reference of four weeks (28 days) stated in categories **I21-I25** signifies the interval elapsing between the onset of the ischaemic episode and admission to hospital; this time reference must be observed by coders to ensure consistency in recording myocardial infarctions.

Acute myocardial infarction

A code from category **I21.- Acute myocardial infarction** must be assigned every time a patient has an acute myocardial infarction (MI), except when a subsequent MI occurs within 4 weeks of the onset of a previous infarction, in which case a code from category **I22.- Subsequent myocardial infarction** must be used as described below.

Where a new acute MI is diagnosed more than four weeks (28 days) after a previous MI, a code from category **I21.- Acute myocardial infarction** must be assigned.

STEMI and NSTEMI

Non-ST segment elevation myocardial infarction (NSTEMI) must be coded to **I21.4 Acute subendocardial myocardial infarction**.

ST segment elevation myocardial infarction (STEMI) must be classified using a code in the range **I21.0 - I21.3** depending on the site/extent of damage to the heart documented in the medical record. Where the site/extent of damage is not known, code **I21.3 Acute transmural myocardial infarction of unspecified site** must be assigned.

Other examples of ontologies

- **LOINC** - lab values
- **GO** – gene functions and biological processes
- **HPO** – phenotypic abnormalities
- **UMLS** – unified medical language system
- **MeSH** – Medical Subject Headings
- and many more at the Open Biological and Biomedical Ontology Foundry
<http://obofoundry.org/>

Not all terms are useful :-)



W61.62XD Struck by duck, subsequent encounter

W55.29XA Other contact with cow, subsequent encounter

R46.1 Bizarre personal appearance.

V91.07 Burn due to water skis on fire

9N46.00 Doctor walked out

Z6V4100 Running in circles

1F87.00 Eats junk food ++

Summary: Clinical terminologies enable you to

- **Record, store and manipulate** clinical data in electronic health records with the appropriate level of detail
- **Exchange** clinical data between independently developed clinical information systems
- **Combine** clinical data from multiple sources (care/research/audit/planning)
- **Share and re-use** clinical decision support rules