

Biomedical Research and Electronic Health Record

George Hripcsak, MD, MS
Department of Biomedical Informatics
Columbia University

Using EHR/PHR for clinical research

- \$2.5 trillion industry
- not whether, but what and how

Setting

EHR adoption

- Ambulatory
 - 17%
- Hospital
 - 77% laboratory
 - 15% system
- \$19 billion for EHR adoption

EHR at one campus

2,500,000	patient records
62,000,000	laboratory batteries
6,000,000	provider notes
34,000,000	ancillary reports
20,000,000	orders

National data, per year

1,000,000,000	visit notes
35,000,000	admit notes, discharge sum.
46,000,000	procedure notes
3,000,000,000	prescriptions
1,000,000,000	laboratory tests
>50,000,000,000	facts

PHR

- Tethered
 - Allscripts
 - Epic
- Untethered
 - HealthVault

Testdad
[Switch person](#)

MY MEDICAL RECORDS

MY HEALTH EXPLAINED

MY HOSPITAL

MY DOCTORS

Immunization History for **Testdad Healthvault** (Age: 47y ; Born: 1/1/1962)[Español](#)[Options](#)

Series	Type of Vaccination	Date	Source	Details	Status
Hepatitis B (Hep B)	Hep B, adolescent or pediatric	4/7/2009	Parent Entered		Overdue
	Hep B, adolescent or pediatric	4/20/2009	Parent Entered		
Diphtheria, Tetanus, Pertussis (DTaP)					No shots
Inactivated Poliovirus (IPV)					No shots
Haemophilus influenzae type b (Hib)	Hib (HbOC)	4/20/2009	Parent Entered		Overdue
Pneumococcal	pneumococcal conjugate	4/7/2009	Parent Entered		No status
Rotavirus (severe diarrhea)					No shots
Measles, Mumps, Rubella(MMR)	measles	4/20/2009	Parent Entered		Overdue
	MMR	4/20/2009	Parent Entered		
Varicella (chickenpox)					No shots
Hepatitis A (Hep A)					No shots
Human Papillomavirus (HPV)	HPV	4/20/2009	Parent Entered		No status
Meningococcal Conjugate					No shots
Tetanus, Diphtheria, Pertussis (TDaP)					No shots
Influenza	influenza, split	4/20/2009	Parent Entered		Overdue
PPD					No shots
Other					No shots
Unknown					No shots

How the EHR/PHR may help

1. Data
2. Subjects — patient recruitment
3. Knowledge — sample size
4. Continuity — long term follow up
5. Fully EHR-based observational studies
 - without case-specific curation
6. Fully EHR-based interventional trials

Supply data

EHR

- Been using for decades
- Manually curated
 - Read record, enter into research database

PHR

- Information that would not otherwise be captured

“As the new templates are being developed for the EHRs, how is family history being documented? In any consistent way?”

Clinical Mgr DEV-W - Citrix XenApp Plugins for Hosted Apps
 SANDIEGO, CARMEN - Structured Notes Entry Dialog
 Nursing Adult Admission History - Structured Notes

Document Topic: _____
 Authored: Date Now 12-18-2007 Time: 08:07
 Authored By: Me Other _____ Source: _____
 Co-Signatures: _____
 Flag As: Incomplete Results Pending Priority Modify Acronym Modify Template

Allergies
 Problem List
 Advance Directive
 Avascular
 Admission Data
 Immunization History
 Past Medical Hx (include psych) **Past Medical Hx (include psych)**
 Travel Assessment
 Pain Assessment
 Outpatient Medication Profile
 Medications, Herbs & Nutrition
 Self Care History
 Functional Screen - Responses
 Nutritional Screen - Certain resp
 Social History
 Social History
 Belongings
 Unit Orientation
 Education Assessment - Pt/Fam
 Religion / Culture

Past Medical Hx (include psychiatric)

Prior Illness None
 Abdominal Aortic Aneurysm
 Abdominal Trauma
 Acute Renal Failure
 AIDS

Prior Hospitalizations _____
 Prior Surgeries _____

Family Health Hx: None Angina Cancer CHF Coronary Artery Disease CVA Diabetes Hypertension Seizures TIA _____

Alcohol/Drug Use
 Does not drink alcohol Drinks socially Drinks daily (enter amount) Uses cocaine Uses marijuana Uses crack Uses methadone Uses pain killers Uses sedatives Uses amphetamines Uses diet pills
 Uses heroin _____

Tobacco Use/Smoking History No, patient has not used tobacco in the past year. Yes, patient has used tobacco in the past year; smoking cessation advice/material provided. _____
 COMPLETE ONLY IF AN RN HAS IDENTIFIED AS SUICIDE. In the last month, does the patient indicate that (A) things were so bad that he/she thought that life is not worth living or that he/she would be better off dead or (B) he/she had thoughts of suicide.
 A and B A B Neither _____

First Previous Section Previous Next Next Section Last

Section Document
 Copy Forward Calculate Submit Cancel
 Refer to Note Hide Review

Family Health Hx None Angina Cancer CHF Coronary Artery Disease CVA

Alcohol/Drug Use

Diabetes Hypertension Seizures TIA _____

“Although I understand the reason for performing in this way, this won’t provide much helpful information because it is so shallow.”

Family Health History - Welcome - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://familyhistory.hhs.gov/fhh-web/home.action

Get Help En Español

My Family Health Portrait

A tool from the Surgeon General

Using *My Family Health Portrait* you can:

- Enter your family health history.
- Create drawings of your family health history to share with family or health care worker.
- Use the health history of your family to create your own.

Talking with your health care worker about your family health history can help you stay healthy!

[Learn more about My Family Health Portrait](#)

Create a Family History

Open a Saved History File



[Glossary](#) | [FAQ](#) | [Accessibility](#) | [Privacy & Security Policy](#) | [About This Site](#) | [Contact Us](#) | [Site Updates](#)



Done

familyhistory.hhs.gov

Family Health History - Family Health History - Mozilla Firefox

File Edit View History Bookmarks Tools Help

https://familyhistory.hhs.gov/fhh-web/familyHistory/st

Google

My Family He

Home | My Family History

My Family Health

Create A New Family

Click "Create My History" to create your family history. You can add...

Name
My Family

[Glossary](#) | [FAQ](#) | [Accessibility](#)

Personal Information (click to hide) Get Help

We start the family history with you. Enter the required personal information and your health history information. At the bottom of the page (you may need to scroll), click the "Next" button. You will then be asked to tell the system which family members you would like to add to the health history.

*Indicates required information.

Name:

***Gender:** Male Female

***Date of Birth:**
mm/dd/yyyy

Were you born a twin? Yes - Identical (Same) Yes - Not Identical (Fraternal) No

Were you adopted? Yes

Height: Feet Inches

Weight:

Health Information (click to hide) Get Help

Select the diseases and conditions as well as what age the disease came. If a disease or condition is not on the list, select "Other - Add New" to add a new one to the list. (Any disease that you add will appear at the bottom the list so it can be selected for other relatives.) If this person does not have any diseases or conditions, this section may be skipped.

Disease or Condition:

Age at Diagnosis:

Disease / Condition	Age at Diagnosis	Remove
---------------------	------------------	--------

Get Help | En Español

many people are in your

Done

familyhistory.hhs.gov

Recruitment

Study recruitment

- Use EHR to find patients eligible for trials, inform doctor
- Patient seeks out a specific study, with the aid of a PHR
- Patient volunteers to be contacted for future studies, select via EHR/PHR

TrialX - Enabling Patients to Find New Treatments through Clinical Trials Recruitment - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.trialx.com/ Google

TrialX Enabling Patients to Find New Treatments. Sign Up as Patient | Sign Up as Investigator | Sign in

Home Take a Tour Patients Investigators Community Help

You don't have to spend hours finding a trial or investigator

[Learn More](#)



TrialX is a free service to enable you to find new treatments in your area >

"TrialX matches patients to relevant clinical trials based on their health information"
- The New York Times

Feedback

What is a Clinical Trial?

Clinical trial is the scientific term for a test or study of a drug or medical device in people. These tests are done to see if the product is safe and effective for people to use.

Why volunteer in a Clinical Trial?

By taking part in a clinical trial, you can try a new treatment that may or may not be better than those that already exist. You can also help others better understand how the treatment works in people of different races and genders.

[Continue reading...](#)

News & Events

- Meet us at the [Microsoft Connected Health Conference 2009](#), June 10-12, Seattle, WA
- TrialX wins first [NYC Entrepreneur Venture Competition!](#)
- Meet us at [ACRP 2009](#). Drop by [Booth #323](#). Apr 25-27, Denver, CO

Start Here! All services are FREE! Your data are kept completely secure and confidential.

Search Trials By Condition or Treatment

Enter conditions:
Example: diabetes, breast cancer

AND/OR

Enter treatments:
Example: glutamine, infrared therapy

[Find Clinical Trials](#)

Recent Searches

- [uveitis anterior](#) 2 minutes ago
- [prostate cancer](#) 3 minutes ago
- [hypoplasia pulmonary](#) 4 minutes ago
- [fracture](#) 4 minutes ago
- [basal cell carcinomas](#) 5 minutes ago

TrialX Facts

- 14,000+ ongoing clinical trials approved by the US FDA
- 8000+ clinical trial investigators
- 3000+ and counting participants matched to clinical trials

Get Matching Trials based on your Personal Health Record

Copy FROM  Microsoft HealthVault

 Google Health BETA

Microsoft HealthVault is the way to collect, store and share. TrialX is a Google Health integrated partner. Start an account

TrialX Services

- [Twitter App](#) **New!**
- [Community](#)
- [Investigator Registration](#)

Done

Fully EHR-based observational studies

Recent examples

- Worsening hypertension and incomplete treatment in Netherlands and Italy
 - Integrated Primary Care Information (IPCI) of Dutch GPs
 - retrospective cohort
 - Sturkenboom ... J Hum Hypertens. 2008;22:704-13
- Hypoglycemic episodes increase risk of dementia
 - Kaiser Permanente
 - EHR and previously surveyed (future PHR?)
 - Whitmer ... JAMA. 2009;301:1565-1572
- Does radiographic contrast material really cause kidney damage
 - non-contrast studies have the same rise in creatinine
 - check accuracy of traditional studies
 - Newhouse ... Am J Roentgenol. 2008;191:376-82

NEJM (2 months)

- 18 randomized clinical trials
 - 1 used a registry to recruit patients
- 4 analyses of biological samples
- 3 large population cohort studies
- 2 questionnaire-based studies
- 1 multicenter observational trial
 - genetic samples & manually curated clinical data

Solvable challenges

- Lack of penetration of EHRs
- Distributed systems
- Inconsistent formats
- Privacy

Data exchange standards

- Clinical trials data standards
 - Clinical Data Interchange Standards Consortium (CDISC)
 - standards for clinical trials data
 - cancer Biomedical Informatics Grid (caBIG)
 - open infrastructure for sharing bioinformatics data and knowledge
- Clinical care data standards
 - Health Level Seven (HL7) version 2
 - Health Level Seven (HL7) version 3
 - Reference Information Model (RIM)
 - Clinical Document Architecture (CDA)
- Interoperability between the electronic health record and electronic data capture
 - CDISC eSDI, eClinical Forum/PhRMA EDC/eSource Taskforce

Terminology standards and tools

- International Classification of Diseases (ICD)
- Current Procedural Terminology (CPT)
- Digital Imaging and Communications in Medicine (DICOM)
- Logical Observation Identifiers Names and Codes (LOINC)
- Systematized Nomenclature for Medicine-Clinical Terms (SNOMED-CT)
- Unified Medical Language System (UMLS)

[The Medical Entities Dictionary \(Top #72145\)](#)

Dynamic model viewer

Hierarchy	Slots
2 Parents	32699 - Serum Potassium Ion Tests
P 32100 - Serum Chemistry Test [156] 32310 - Intravascular Potassium Test [8]	1-UMLS-CODE: C0236573 5-SYNONYMS: 7-HAS-PARTS * 8-PART-OF * 11-DEFINITION: 14-ASSESES-SAMPLE * "2393 - Serum Chemistry Specimen" 16-ENTITY-MEASURED * "32051 - Potassium" 17-UNITS: 23-TEST->RESULT-TYPE * "1178 - Number or String Result" 38-CPMC-NORMAL-VALUE: 39-CPMC-LOW-NORMAL-VALUE: 40-CPMC-HIGH-NORMAL-VALUE: 41-CPMC-MALE-LOW-NORMAL-VALUE: 42-CPMC-MALE-HIGH-NORMAL-VALUE: 43-CPMC-FEMALE-LOW-NORMAL-VALUE: 44-CPMC-FEMALE-HIGH-NORMAL-VALUE: 45-CPMC-NORMAL-RANGES-TEXT: 50-MAIN-MESH: 51-SUPPLEMENTARY-MESH: potassium and blood 125-COLLECTED-BY * "33023 - Specimen Collection" 133-HAS-DEFAULT-DISPLAYS * 136-ELEMENTS-DISPLAYED-BY * 138-IS-DISPLAY-PARAMETER-OF * "40444 - Chem-7 Display" 138-IS-DISPLAY-PARAMETER-OF * "40445 - Chem-20 Display"
32699 - Serum Potassium Ion Tests	
C 1609 - Serum Potassium Ion Measurement 32713 - Serum Potassium Ion Measurement 2 35994 - CPMC Laboratory Test: Potassium(Allen) 56254 - CPMC Laboratory Test: Serum Potassium Measurement	
4 Children	

Select new Medcode: * Submit * Search the MED: * On Slot:

Interoperability projects

- Nationwide Health Information (NHIN)
 - Electronic Health Records/Clinical Research
 - Pharma and health care
- i2b2
 - NCBC open source framework for translational research
 - primary role to bridge clinical research data and basic science knowledge
 - components for reusing EHR data
- ...

Privacy

- HIPAA
- IRB
- Perception
- Value

Hard challenges

- Quality of the data
 - Ambiguous or unknown meaning
 - Accuracy
 - 50-100% accuracy [Hogan JAMIA 1997]
 - Completeness
 - mostly missing
 - Complexity
 - disease ontologies
- Bias

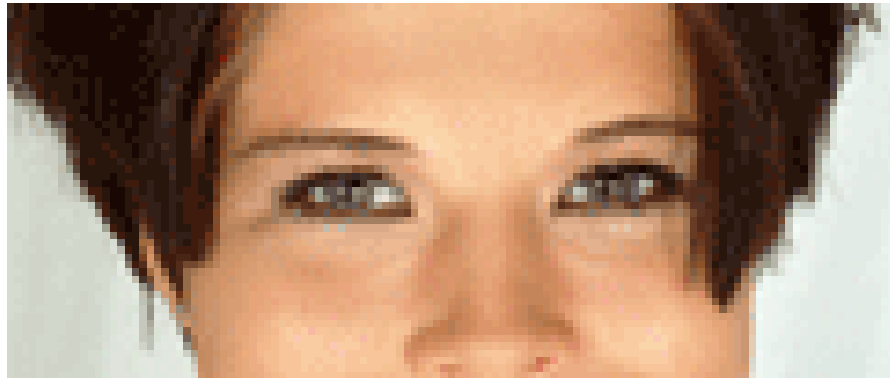
Data quality

- All medical record information should be regarded as suspect; much of it is fiction.
 - Burnum JF ... Ann Intern Med 1989
- Data shall be used only for the purpose for which they were collected. If no purpose was defined prior to the collection of the data, then the data should not be used.
 - van der Lei J ... Method Inform Med 1991

No more and no less

- **PERRLA**

Pupils equal, round, reactive to light and accommodation



Medicine resident daily progress note:

1. Events overnight

Medicine Resident Daily Progress Note - Structured Notes

Document Topic:

Authored: Date Now 09-Oct-2008 Time: 14:53

Authored By: Me Other Source:

Co-Signatures:

Flag As: Incomplete Results Pending Priority

Medicine Resident Daily Progress Note

Events Overnight

Events Overnight: None Events Free text option

Fever: cultures drawn cultures not drawn antibiotics initiated/changed antibiotics continued

Hypotension: symptomatic asymptomatic responsive to IVF unresponsive to IVF anti-hypertensives held

Chest Pain: no EKG changes EKG changes cardiac enzymes sent

Management: Management: responsive to nitroglycerin not responsive to nitroglycerin beta-blocker given beta-blocker not given asa given asa not given

Arrhythmia: Afib Aflutter SVT Vtach Vfib other hemodynamically stable hemodynamically unstable responsive to adenosine

responsive to AV nodal blocking agents responsive to cardioversion

Desaturation: Desaturation: nasal cannula face mask non-rebreather non-invasive ventilation intubation

Management: Management: suctioning Lasix nebs antibiotics CXR ordered

Hemoglobin Drop: repeat hgb stable repeat hgb lower transfusion required

Electrolyte abnormalities: Electrolyte abnormalities: hyperkalemia hypokalemia hypernatremia hyponatremia hypercalcemia hypocalcemia

Disorientation: agitated hallucinating ABO x 2 ABO x 1 ABO x 0 antipsychotics administered restraints utilized 1:1 ordered security called

Comment:

Auto Enter

Medicine resident daily progress note: 13. Compliance

The screenshot shows a software interface for a structured medical note. On the left, a navigation pane lists various sections. Three sections are highlighted with arrows and labels: 'Inform Pt edu' points to the 'Patient Education' section, 'Smoke' points to the 'Tobacco' section, and 'Pain' points to the 'Pain' section. The main area contains the following sections:

- Document Topic:** [Text field]
- Authored:** Date Now 09-Oct-2008 Time: 15:08
- Authored By:** Me Other [Text field] Source: [Text field]
- Co-Signatures:** [Text field]
- Flag As:** Incomplete Results Pending Priority
- Modify Acronyms** **Modify Template**
- Problem List** (History, Allergy, Medications, Flowcharts, Physical Exam, Lab, Radiology)
- Attending's Note (if applicable)**
 - Discussed with resident, agree with above: [Text field]
- Informant/Interpreter**
 - Informant: Patient Other [Text field]
 - Interpreter Used: No Yes [Text field]
- Patient Education**
 - Learner: Patient Refer to teaching protocol/Interdisciplinary Family Education Record(if applicable) Other [Text field]
 - Barriers to Learning: None Hearing Vision Cognitive ability Physical ability Cultural or religious considerations [Text field]
 - Topics taught: Medication Use Reviewed Diet/Wt Management Injury Prevention/Safety Safe Sex/STD Substance Use Other [Text field]
 - Methods of teaching: Explained Demonstrated Handout Audiovisual Other [Text field]
 - Outcome: Verified/demonstrated understanding Needs reinforcement [Text field]
- Tobacco**
 - Tobacco Use: Historical Current Never
 - Historical Tobacco Use: [Text field]
 - Smoking Cessation Treatment(for current users): MD Advice Not interested in quitting RX given Referral Made
- Pain**
 - Patient reported pain: [Text field]
- Medication Reconciliation:** Done

At the bottom, there are buttons for 'Section', 'Document', 'Copy Forward', 'Calculate', 'Submit', 'Cancel', 'Refer to Note', 'Hide', and 'Review'.

“I don’t read notes anymore; I just write them.”

Narrative text entry

- Expressive
- Cut and paste
 - ... 36 year old man ... 27 year old woman ...
- Misspellings and interesting abbreviations
 - text messaging
 - s/p LURT 1998 c/b 1A rejection 7/07 back on HD
- Not computable

Natural language processing

“Slight increase of
pulmonary vascular
congestion with new
left pleural effusion,
question mild
congestive changes”

pulmonary vascular congestion

change: increase

degree: low

pleural effusion

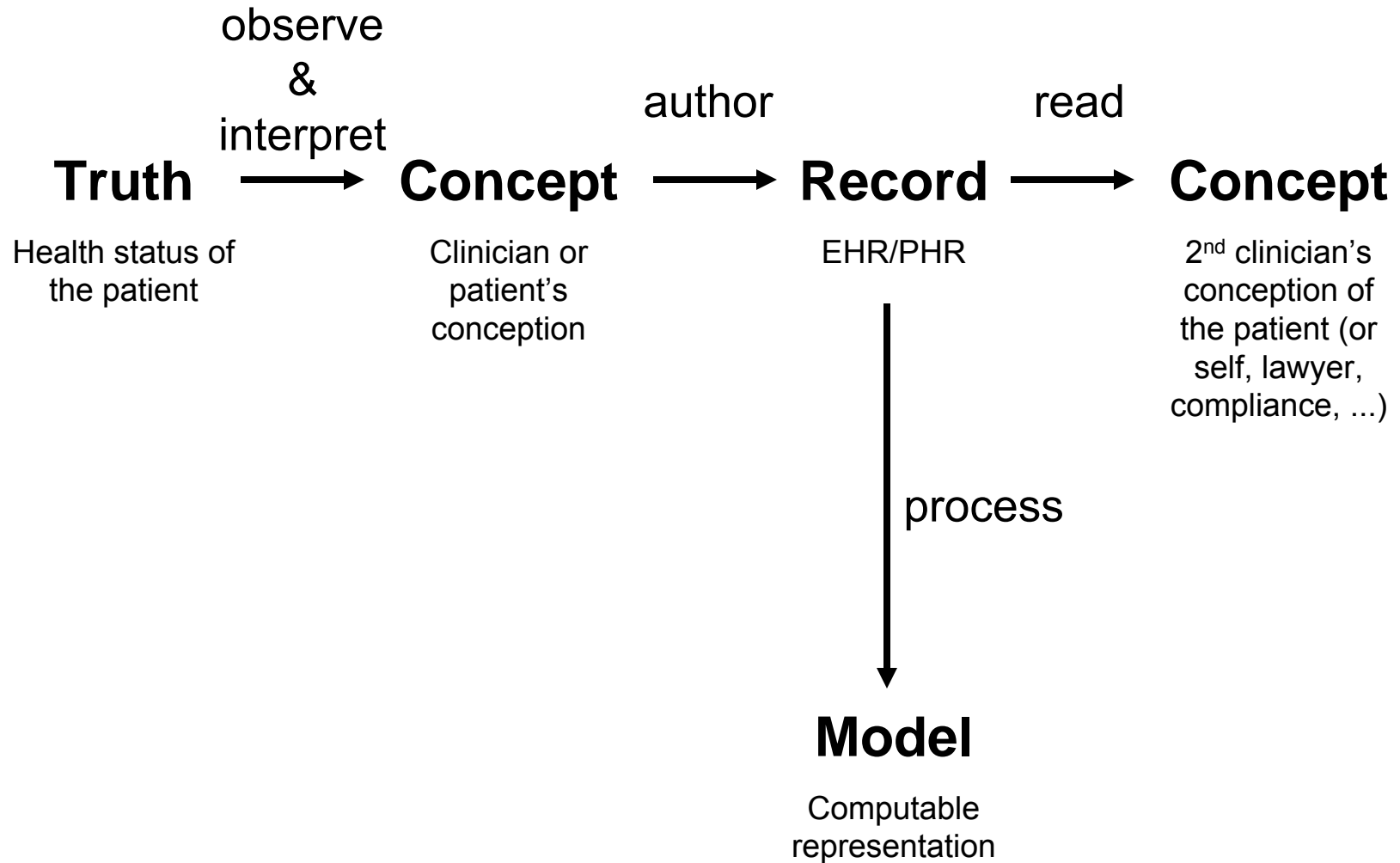
region: left

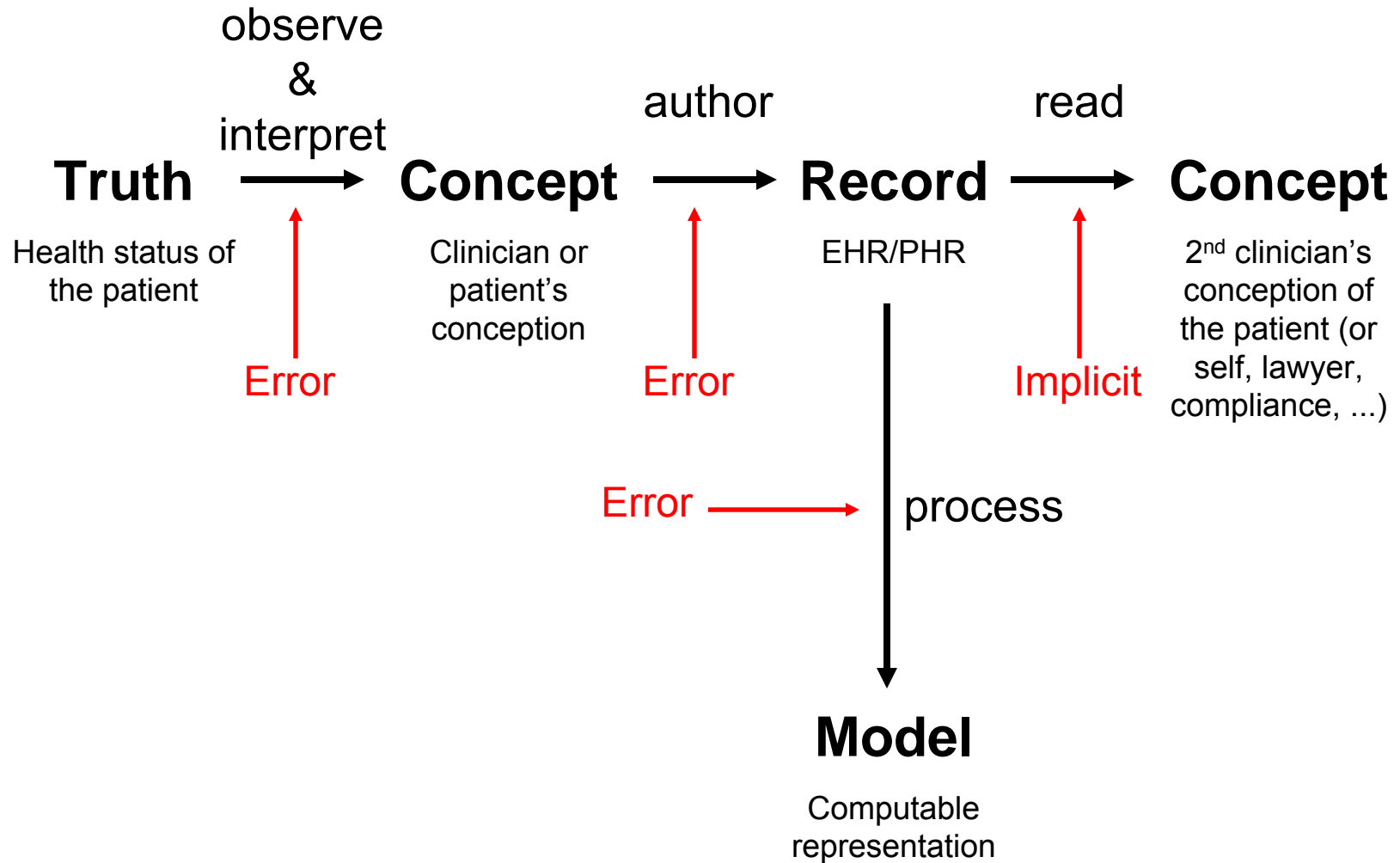
status: new

congestive changes

certainty: moderate

degree: low





“State of the art”

- How to circumvent messy data
 - For the past several decades, use iteratively derived heuristic rules generated by knowledge engineers and domain experts
 - Optimize sensitivity and specificity

TB detection

R1. Positive Mycobacterium tuberculosis culture

R2. Positive acid fast bacillus smear or any positive Mycobacterium culture

R3. Chest radiograph report whose description or impression section is either:

**(a) highly suspicious for tuberculosis (radiologist mentions tuberculosis or cavitory disease)
OR**

(b) moderately suspicious (various forms of infiltrate, granuloma, effusion, or pleurisy) with immunocompromise (implied by use of medications that treat complications of immunocompromise or by tests that demonstrate immunocompromise)

R4. Inpatient use of anti-tuberculosis medications

R5. R1 or R2 (approximates laboratory criteria)

R6. R3 and R4 (approximates clinical criteria)

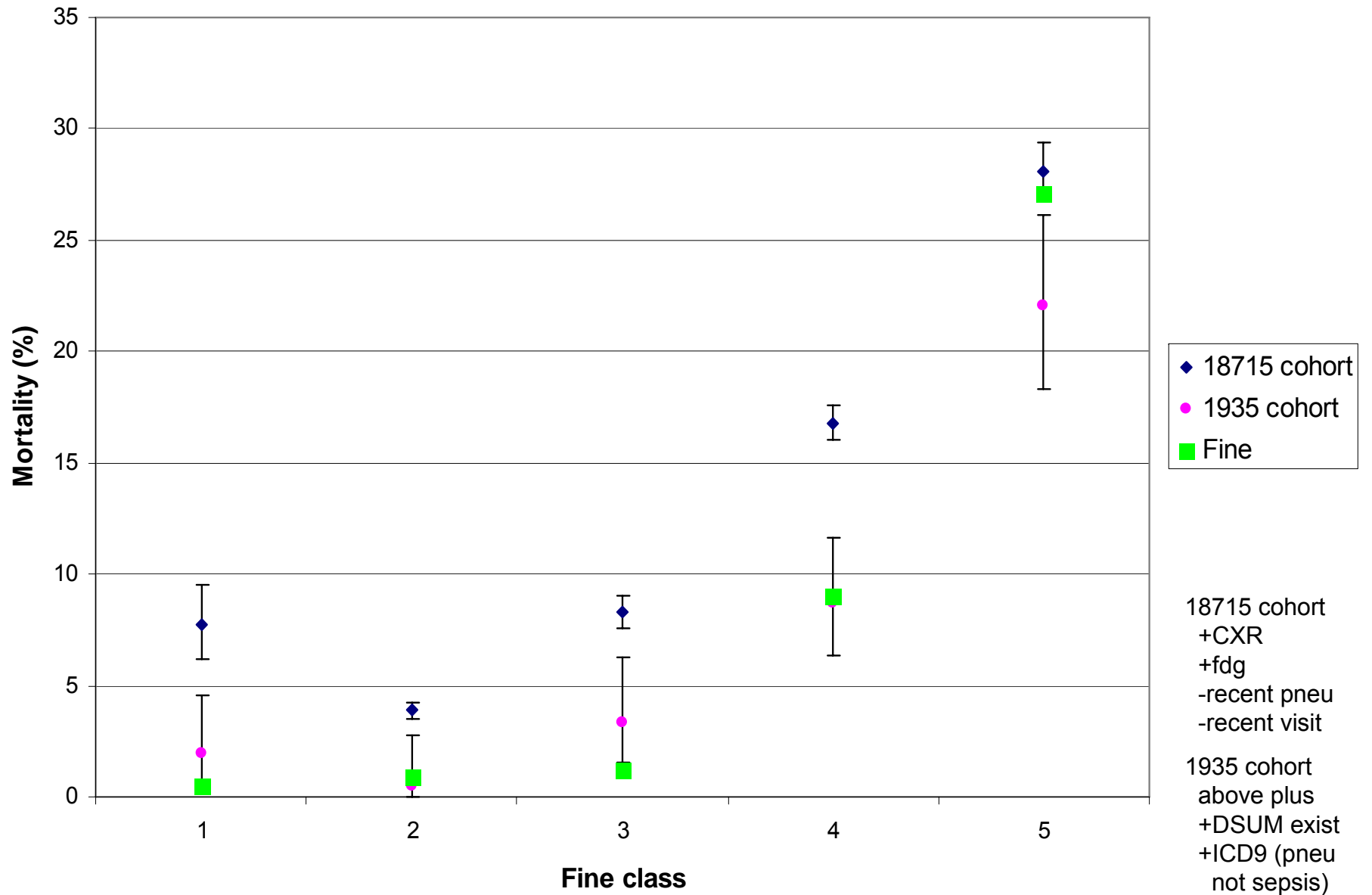
R7. R1 or R2 or {R3 and R4} (approximates overall criteria)

R8. R1 or R2 or R3 or R4 (most sensitive rule)

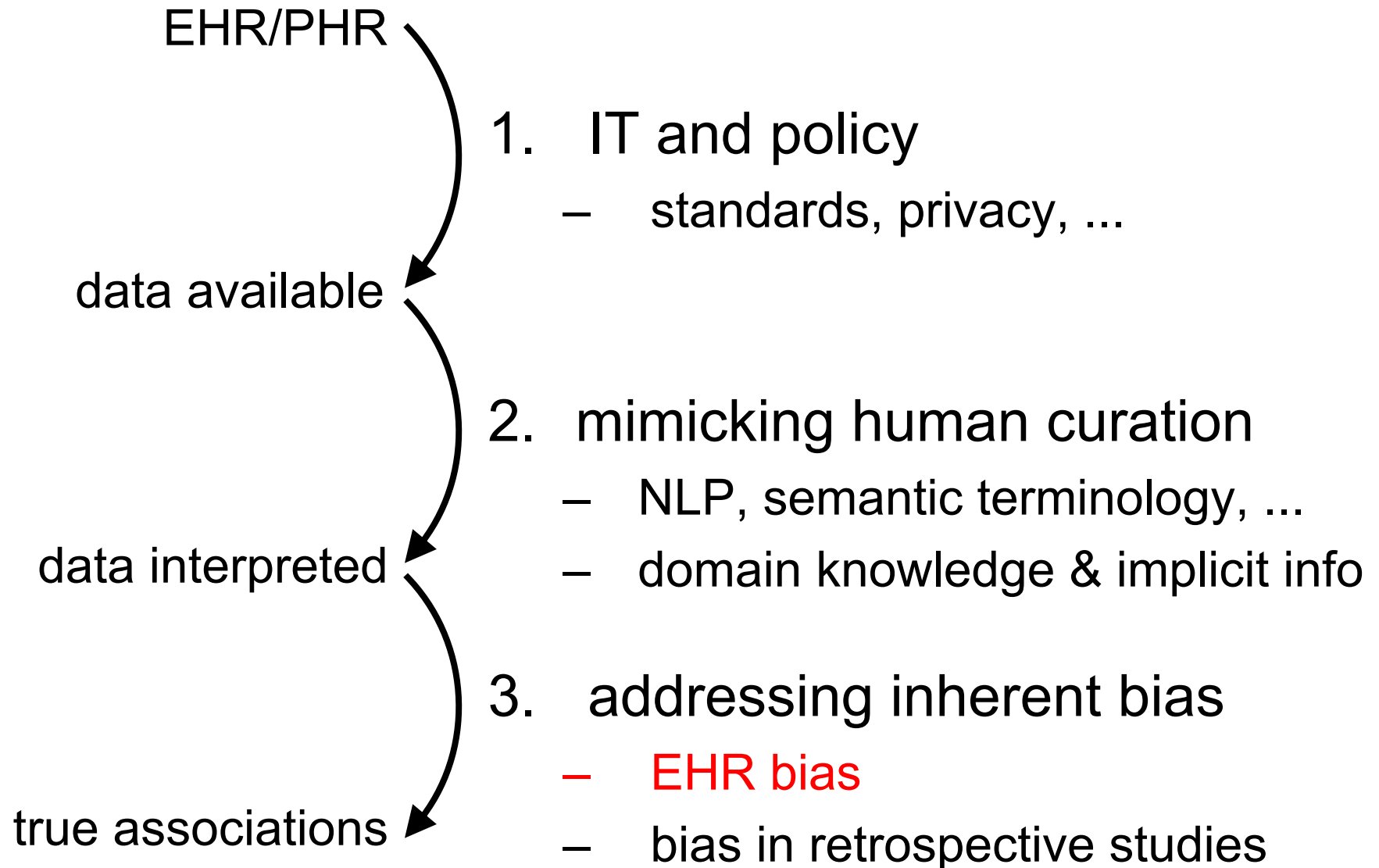
- Tedious
- Imperfect results
- Must be a better way to do this

Population issues

Inpatient mortality for community acquired pneumonia



Using EHR/PHR for research



10 Maastricht rules for reuse

1. State the research purpose before collecting the data
2. Define the population denominator; the unique identifier used to link patients to their data should be identified
3. Record the characteristics of the practices involved in the study
4. Describe the context of data: personal, cultural, technical, health system, financial
5. ...

Maastricht rules

- Pretty much renders magical EHR-based research infeasible

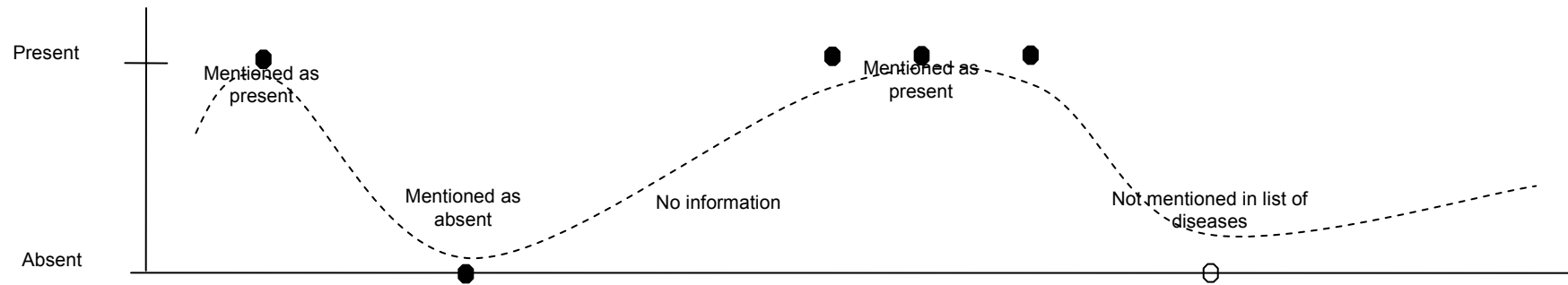
Can we learn from the data?

Physics of medical records

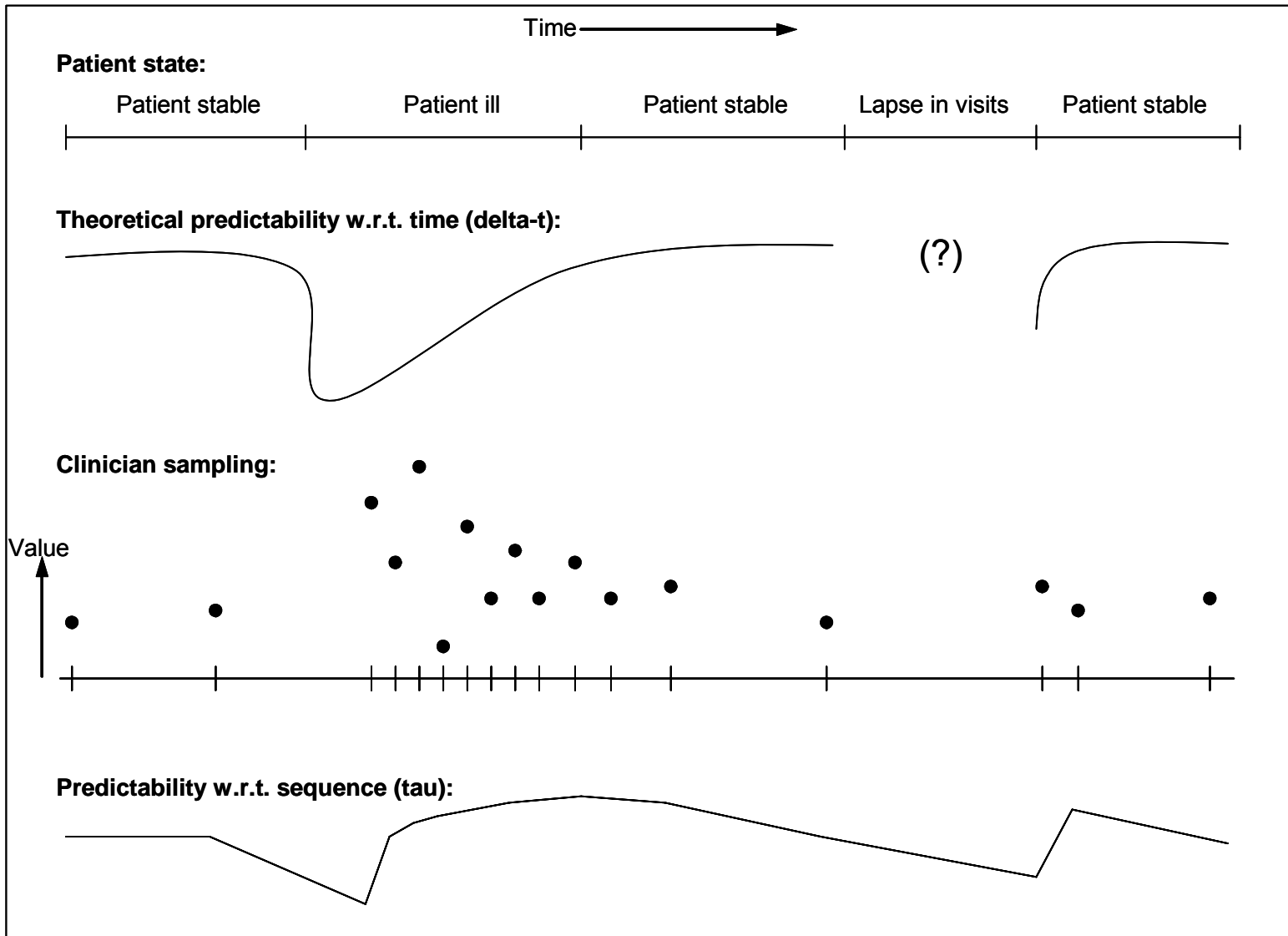
- Study the EHR as an object of interest in itself
 - model mention of pneumonia instead of presence of pneumonia
- Use the EHR to learn about the EHR
- Apply methods from nonlinear time-series analysis

$$I(X; Y) = \mathbb{E}_{X, Y}[SI(x, y)] = \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}$$

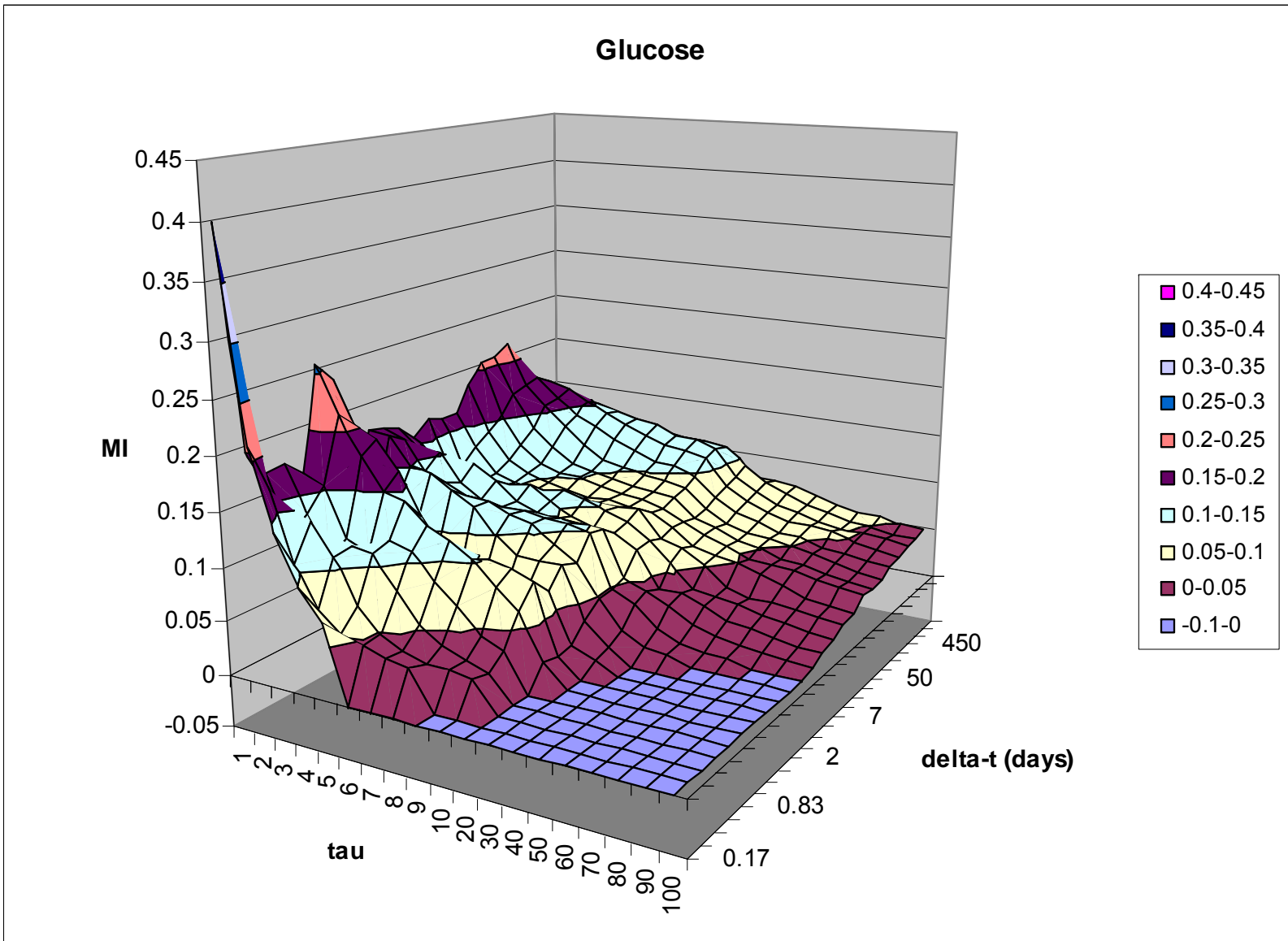
Mention of pneumonia



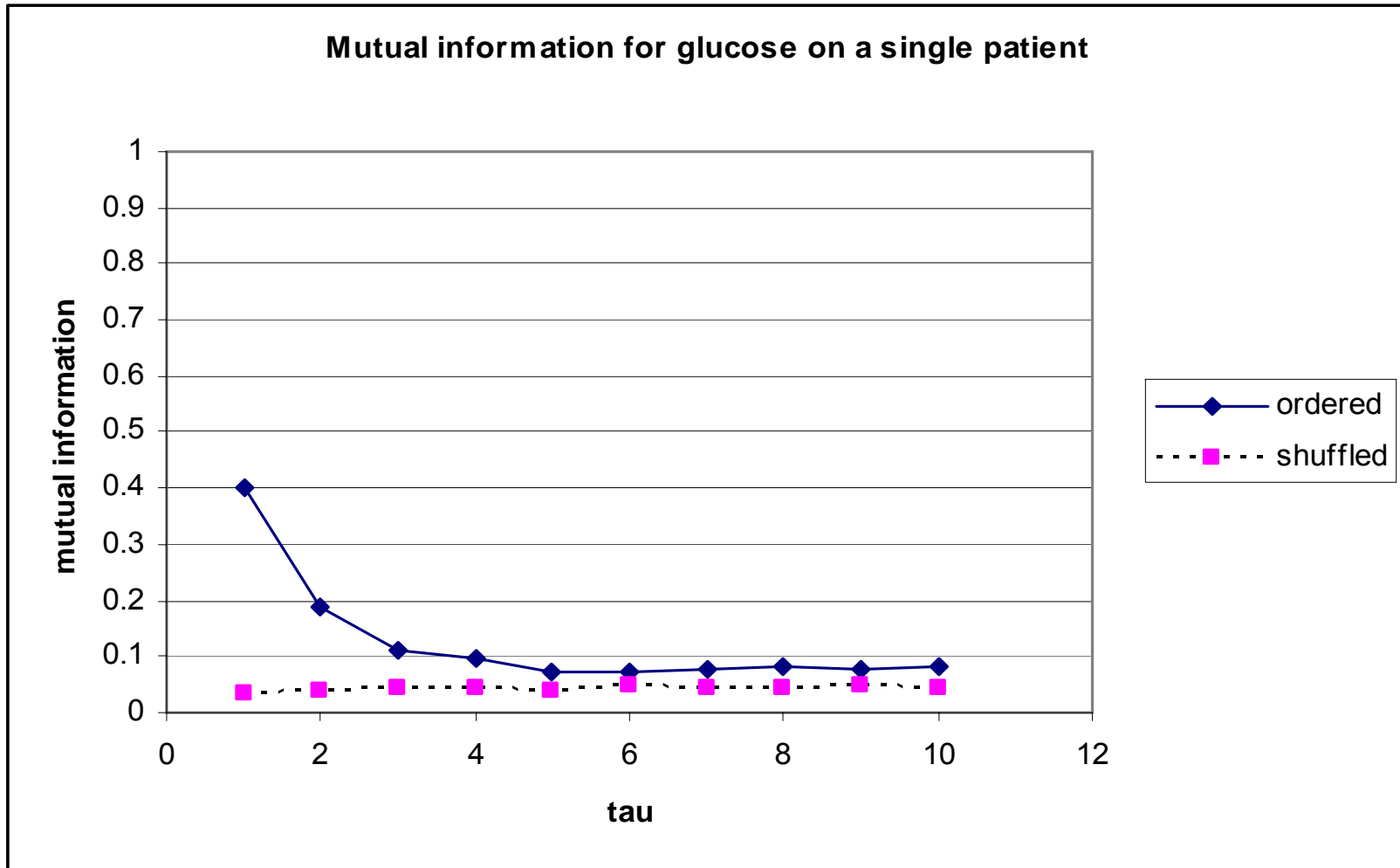
Completeness, sampling bias



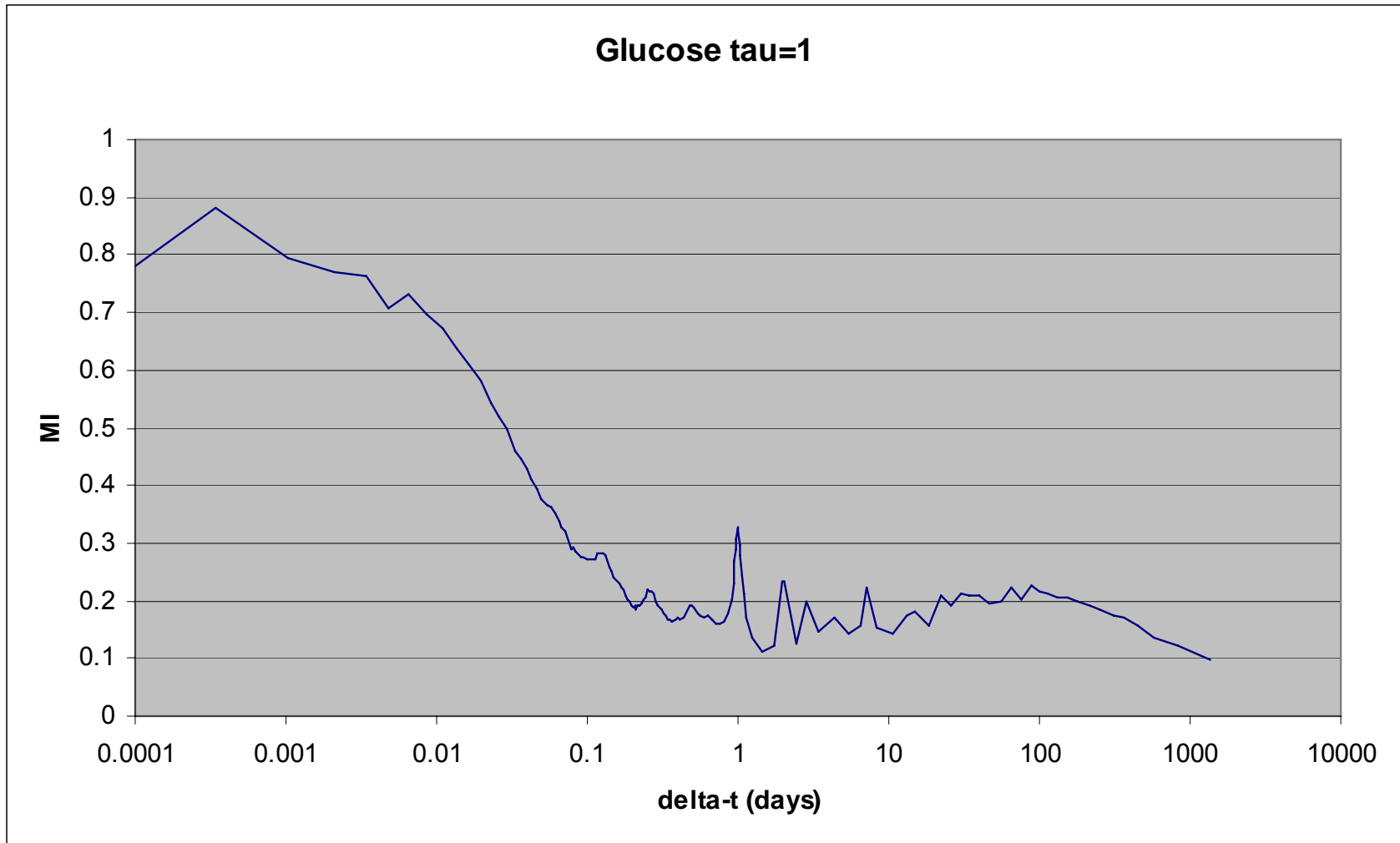
Glucose by Δt and tau



Glucose for one patient



Glucose at tau=1



EHR-based interventional trials

EHR-based interventional trials

- Netherlands Integrated Primary Care Information
- “Randomized database studies”
- GI tolerance of 2 NSAIDs for osteoarthritis

600,000	total population covered
141,395	MD participate
7,127	osteoarthritis and NSAID
170	eligible (visit timing, indication)
20	consented (GP busy, pt refuse)

Role of EHR-based research

- Would like to infer causality
 - randomized clinical trials
 - comparative effectiveness
- Hypothesis
 - generation
 - refinement
 - testing

Summary

- EHRs and PHRs are and will be useful for clinical research
- Working on the solvable challenges
- The EHR and PHR are insufficiently studied and deserve more attention