All of EHR Informatics in 90 Minutes

Emily Pfaff, PhD, MS
Assistant Professor, Dept of Medicine
Co-Director, TraCS IDSci


Clinical Data Modeling
Basics

Patients

Name: Elena Martinez

Birthdate: April 5, 1956

Race: Other

Ethnicity: Hispanic/Latino

Sex: Female

Gender identity: Female

Preferred Language: English

Patients

Name: Elena Martinez

Birthdate: April 5, 1956

Race: Other

Ethnicity: Hispanic/Latino

Sex: Female

Gender identity: Female

Preferred Language: English

"Problem List"

- Type I diabetes
- Chronic back pain

Location

123 Carolina Street

Durham, NC 27704

Phone Numbers

Work: 919-555-5555

Cell: 919-777-7777

Patients have visits.

January 2021

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
27	28	29	30	31	1	2
3	4	5 Endo. visit, 10AM	6	7	8 Ophth. visit, 2PM	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24 ED visit: Chest pain	25 ----------Hospital Stay----------	26	27	28	29	30
31	1	2	3	4	5	6

Things happen during visits: Diagnoses

- *Diagnosis* roughly = “Why are you here today?”
- Entered in the EHR as an *ICD-10-CM* code. (More on that later.)
- Can be used to:
  - Record a general reason for your visit.
  - Justify an order (e.g., lab test, biopsy, etc.)
  - Document for billing/insurance purposes.
- Not reliable for documenting symptoms
- Not necessarily a surefire way to know a patient has a disease.
- The lack of a diagnosis != lack of disease.

Things happen during visits: Diagnoses

- The “Problem List” is a place to store chronic conditions for future reference.
- Structured, but not necessarily at the visit level.
- Dates should not be interpreted as “onset date of disease,” but rather “entry date of this record.” (Onset date may be separate field, often null.)

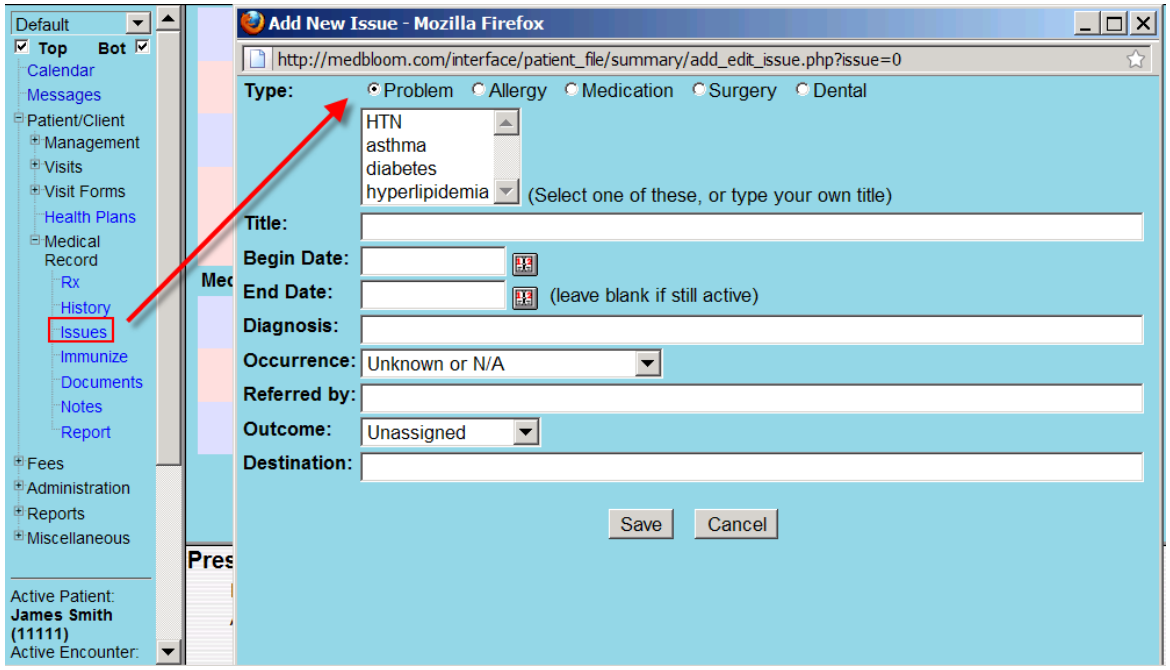


Image from https://www.open-emr.org/wiki/index.php/Problem_List

Things happen during visits: Vitals

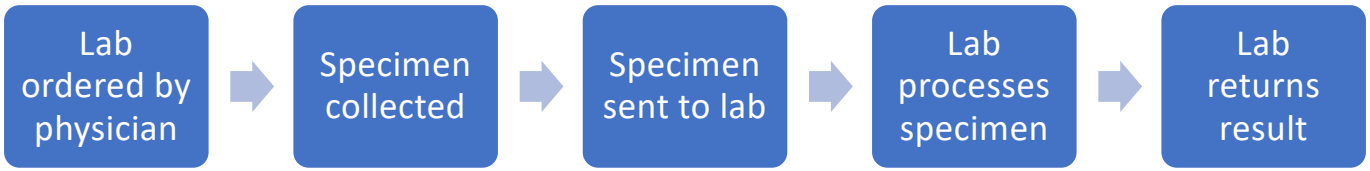
Example Vitals Include:

Blood pressure, heart rate (pulse), oxygen saturation, height, weight, BMI, respiration rate, temperature

- Inpatient vitals are collected very differently from outpatient vitals.
- Mostly quantitative variables that can be plotted over time.
- Not all vitals are collected at every visit.

Things happen during visits: Labs

- Labs are *orders* that have *results*.



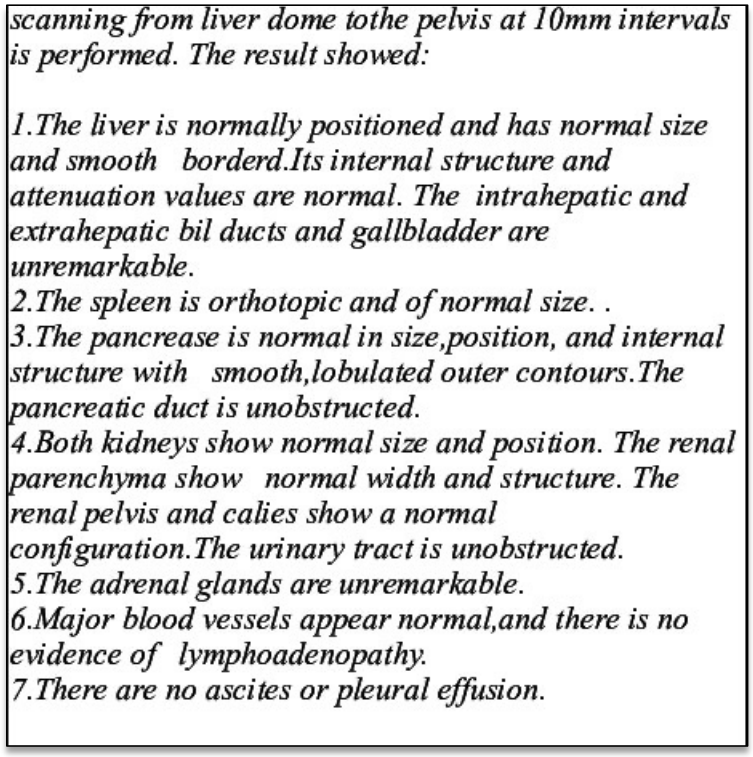
- Results can be quantitative or qualitative.
- There can be one lab per order, or a *panel* of labs.
- Time is a factor (i.e., time between specimen collection and result).

Things happen during visits: Medications

- Medications prescribed at an outpatient visit:
  - Generally sent to a pharmacy for patient pick-up
  - Hard to know if the order was picked up, or if the patient took the drug.
  - Older data (early 2010s and before) is harder to trust because of paper prescriptions.
  - “Stop” dates are often not reliable.
- Medications administered in hospital/clinic:
  - Not necessarily inpatient—there are outpatient infusions, vaccines, etc.
  - More assurance that the drug made it in to the patient.
- Time is a factor in both cases (days supply, number of refills, start/stop date times.)

Things happen during visits: Procedures

- Procedures *sometimes* have results.
  - With results: CT scan, biopsy, stress test
  - Without results: delivering a baby, removing an appendix
- Two methods of coding: ICD-10-PCS (facility) and CPT (physician).
  - Facility charges are things that the hospital gets paid for; physician charges are things that the physician gets paid for.
  - When searching through procedures, always check both.



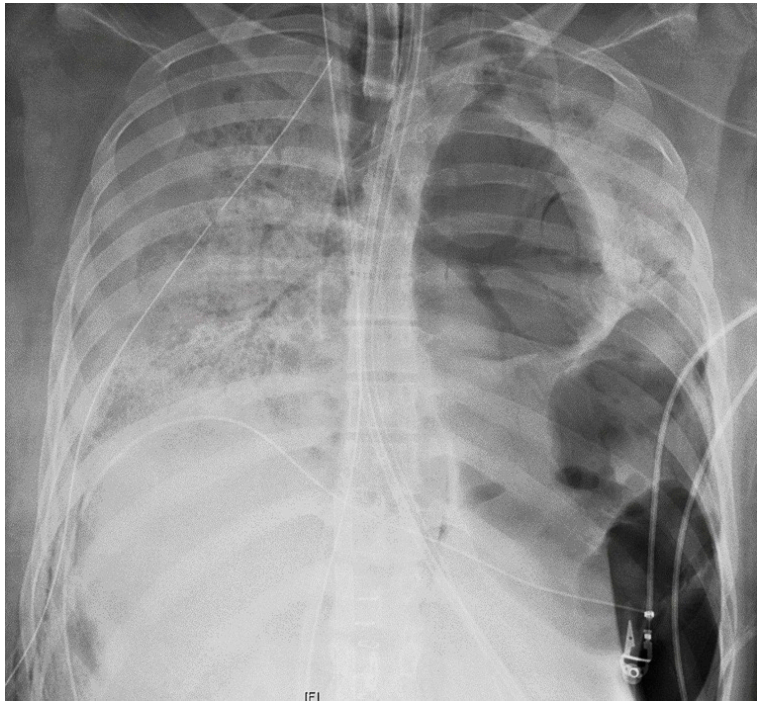
scanning from liver dome tothe pelvis at 10mm intervals is performed. The result showed:

1.The liver is normally positioned and has normal size and smooth borderd.Its internal structure and attenuation values are normal. The intrahepatic and extrahepatic bil ducts and gallbladder are unremarkable.
2.The spleen is orthotopic and of normal size. .
3.The pancrease is normal in size,position, and internal structure with smooth,lobulated outer contours.The pancreatic duct is unobstructed.
4.Both kidneys show normal size and position. The renal parenchyma show normal width and structure. The renal pelvis and calies show a normal configuration.The urinary tract is unobstructed.
5.The adrenal glands are unremarkable.
6.Major blood vessels appear normal,and there is no evidence of lymphoadenopathy.
7.There are no ascites or pleural effusion.

Sample from https://www.researchgate.net/figure/An-example-of-abdomen-CT-scan-report_fig1_220923374

Things happen during visits: Procedures

- Results of procedures are usually in semi-structured or free-text format—or may be an image, PDF, or other "challenging" format.
- Results are usually attached to an order, which is attached to a visit.
- Results in these formats can be analyzed, but require specialized methods.



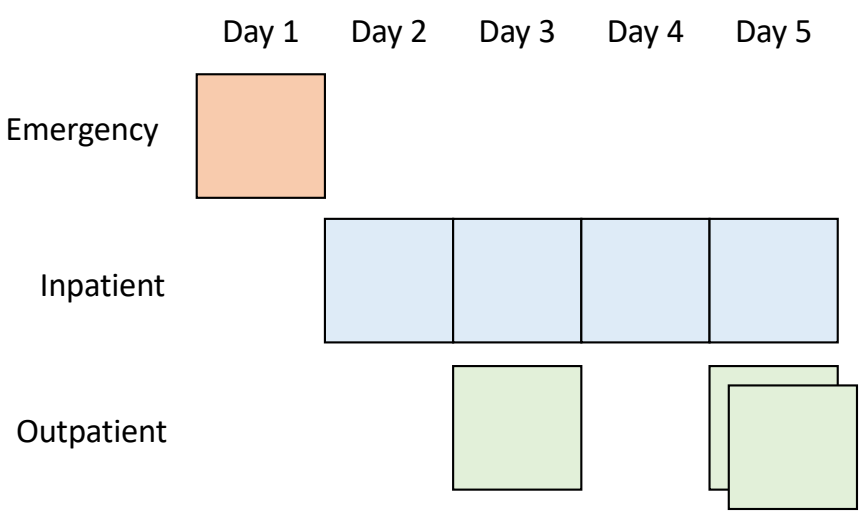
COVID-19 lung image from https://www.nbcnews.com/health/health-news/after-covid-19-destroyed-her-lungs-young-chicago-woman-receives-n1229841

Visits themselves can get complicated.

- Three major visit categories: outpatient, emergency, inpatient.
- Many more granular categories within those types (see right).
- Structure of visits in a healthcare database has more to do with business rules than clinical care.
- Not consistent across institutions.

581478	OMOP4822457	Ambulance Visit
705159	OMOP5160862	Ambulatory long COVID clinic
581479	OMOP4822456	Ambulatory Rehabilitation Visit
262	ERIP	Emergency Room and Inpatient Visit
9203	ER	Emergency Room Visit
32693	HE	Health examination
32759	OMOP4873970	Home isolation
581476	OMOP4822459	Home Visit

Complex visit example



Database Record #

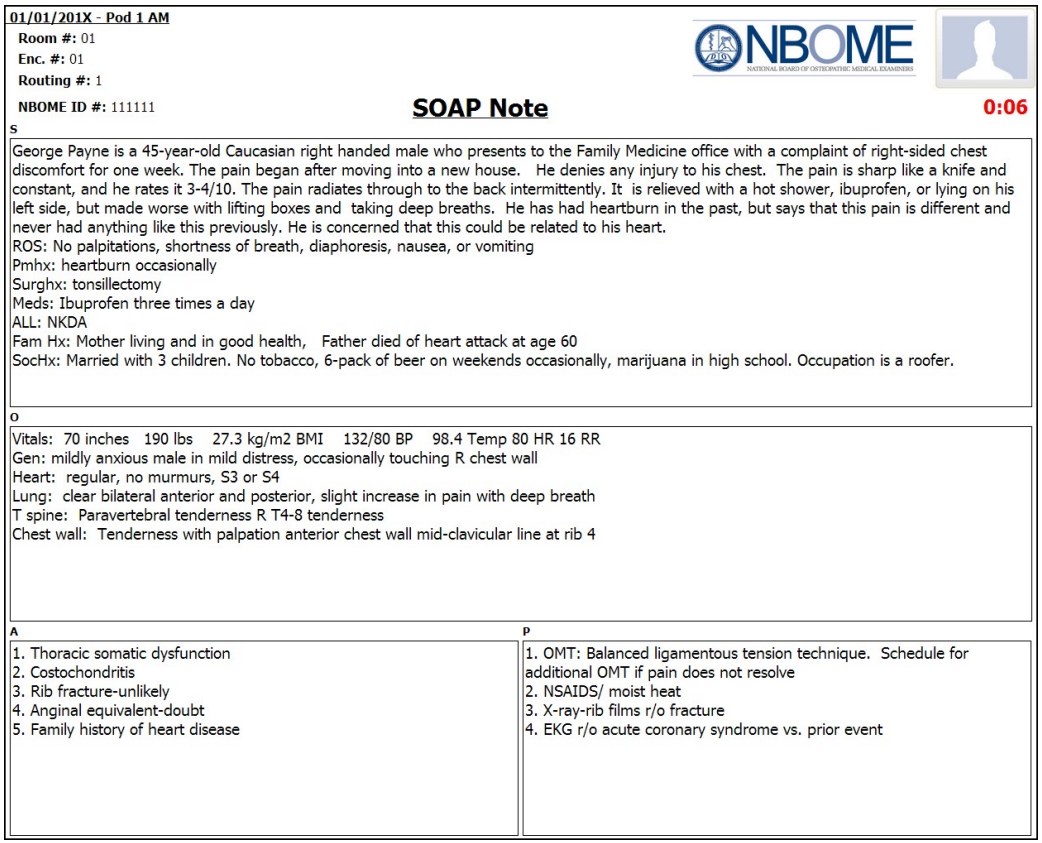
56927934

92834732

10923489
72983234
20983842

Visits of all types have notes & summaries.

- Inpatient: progress notes, discharge summaries, social worker assessments, etc.
- Outpatient: After-Visit Summary, SOAP note
- Challenging to analyze, but good source for symptoms, rare diseases, social determinants, etc.



01/01/201X - Pod 1 AM

Room #: 01
Enc. #: 01
Routing #: 1
NBOME ID #: 111111

SOAP Note

0:06

S

George Payne is a 45-year-old Caucasian right handed male who presents to the Family Medicine office with a complaint of right-sided chest discomfort for one week. The pain began after moving into a new house. He denies any injury to his chest. The pain is sharp like a knife and constant, and he rates it 3-4/10. The pain radiates through to the back intermittently. It is relieved with a hot shower, ibuprofen, or lying on his left side, but made worse with lifting boxes and taking deep breaths. He has had heartburn in the past, but says that this pain is different and never had anything like this previously. He is concerned that this could be related to his heart.
ROS: No palpitations, shortness of breath, diaphoresis, nausea, or vomiting
Pmhx: heartburn occasionally
Surghx: tonsillectomy
Meds: Ibuprofen three times a day
ALL: NKDA
Fam Hx: Mother living and in good health, Father died of heart attack at age 60
SocHx: Married with 3 children. No tobacco, 6-pack of beer on weekends occasionally, marijuana in high school. Occupation is a roofer.

O

Vitals: 70 inches 190 lbs 27.3 kg/m2 BMI 132/80 BP 98.4 Temp 80 HR 16 RR
Gen: mildly anxious male in mild distress, occasionally touching R chest wall
Heart: regular, no murmurs, S3 or S4
Lung: clear bilateral anterior and posterior, slight increase in pain with deep breath
T spine: Paravertebral tenderness R T4-8 tenderness
Chest wall: Tenderness with palpation anterior chest wall mid-clavicular line at rib 4

A	P
1. Thoracic somatic dysfunction 2. Costochondritis 3. Rib fracture-unlikely 4. Anginal equivalent-doubt 5. Family history of heart disease	1. OMT: Balanced ligamentous tension technique. Schedule for additional OMT if pain does not resolve 2. NSAIDS/ moist heat 3. X-ray-rib films r/o fracture 4. EKG r/o acute coronary syndrome vs. prior event

Sample from https://www.nbome.org/resources/completed-esoap-note-sample/

Insurance

- Insurance coverage is at the visit level, and can change over time.
- May be multiple insurers per visit.
- Plan specifics are rarely used in EHR data—roll-ups are common.
- Guarantor of the account is not necessarily the patient.
- Medicaid enrollment often used as a proxy for SES.
- Insurance data, called “claims” data, is a data source separate from EHR that may be used in research with special permission.
- Some states have an “all-payer claims database” available. (NC does not.)

Death

- EHRs reliably record deaths that occur inside the hospital.
- For an inpatient visit, a death date and a discharge date with a *disposition* of “Expired” are generally recorded.
- State death data may also be available, but usually lags in time by a few (or several) months.
- Death cause, if available, is encoded in ICD-10. (Often not available.)
- A patient’s family may also self-report a death, in which case a date will be available (but may be inaccurate).

Family, Medical, and Surgical History

- Self-reported data; may or may not be structured.
  - Drop-downs, pick lists = likely to be structured.
  - Notes, e.g. “Pt notes paternal grandmother had T1D.” = not structured
- Some EHRs allow family “links”; feature is not always used.
- Can be used in tandem with other data to assist with completeness.
  - Good use: “Find all patients who had a hysterectomy.”
  - Bad use: “Find all patients with siblings.”

Social History and Social Determinants

Social History

- Smoking status
- Alcohol use
- Sexual behaviors

Social Determinants of Health (SDoH)

- Food insecurity
- Housing insecurity
- Financial insecurity
- Unemployment
- May be at patient level or area level, structured or unstructured.

Pulling it all together

- This is not a comprehensive list of all data collected in the EHR.
- Most important: the EHR does not provide a full picture of the patient. It's a useful snapshot.

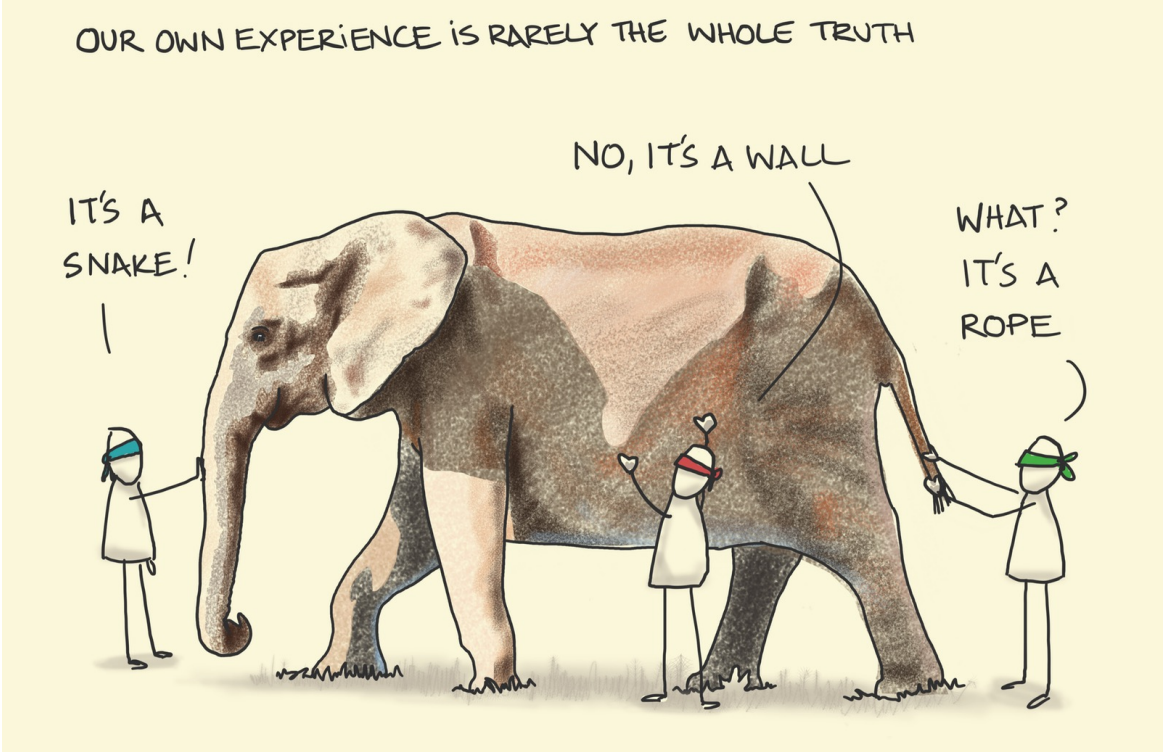


Image from https://sketchplanations.com/the-blind-and-the-elephant


Clinical Data
Standardization

Hospital/clinical data warehousing

Patient_ID	Patient_Nm	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…

Visit_ID	Patient_ID	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…
…	…	…	…	…	…

Other than filing it in a database, how do we make data useful for various analytics use cases?

Hospital/clinical data warehousing

Patient_ID	Patient_Nm	Race	Ethn
444	Smith	White	Non-Hisp
333	Ramirez	Wh/Cau	Hispanic
222	Walters	Caucasian	Hisp

Visit_ID	Patient_ID	Diagnosis
A224	444	Common Cold
A298	444	Cold with sneezing
A090	222	Sneezing

NOT STANDARDIZED = NOT USEFUL

Hospital/clinical data warehousing

Patient_ID	Patient_Nm	Race	Ethn
444	Smith	02	NH
333	Ramirez	02	H
222	Walters	02	H

Visit_ID	Patient_ID	Diagnosis
A224	444	J00
A298	444	J00
A090	222	J00

That's more like it!

Hospital/Clinical Data Standards

- To move toward “nationally interoperable” health care data, we need to be consistent not only within an organization, but across the country.
- An abbreviated list:

Purpose	Standard
Provider	National Provider Identifier (NPI)
Race/Ethnicity	OMB standards
Lab tests	LOINC

Purpose	Standard
Diagnosis	ICD-10
Medications	RxNorm
Procedures	CPT/HCPCS, ICD-10

Hospital/Clinical Data Standards

Note: when every thing that can happen to a human being has to be “codeable,” things can take a turn for the ridiculous. Take ICD-10, for example…

Hospital/Clinical Data Standards

Note: when every thing that can happen to a human being has to be "codeable," things can take a turn for the ridiculous. Take ICD-10, for example…

W61.62XD: Struck by duck, subsequent encounter

Hospital/Clinical Data Standards

Note: when every thing that can happen to a human being has to be “codeable,” things can take a turn for the ridiculous. Take ICD-10, for example…

W61.62XD: Struck by duck, subsequent encounter

V91.07XD: Burn due to water skis on fire, subsequent encounter

Hospital/Clinical Data Standards

Note: when every thing that can happen to a human being has to be “codeable,” things can take a turn for the ridiculous. Take ICD-10, for example…

W61.62XD: Struck by duck, subsequent encounter

V91.07XD: Burn due to water skis on fire, subsequent encounter

V95.43XS: Spacecraft collision injuring occupant, sequela

Where flat files fall short

- In your career, you've probably heard someone talk about their "Excel database."
  - This isn't inaccurate! A database is an structured collection of data, stored electronically—and *flat files*, like Excel sheets, CSVs, and other types of *delimited files*, fit that bill.
  - Here's an example of a flat file used to track patient visits.

Patient Name	Patient Address	Patient Date of Birth	Visit Date	Visit Provider	Clinic Name
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	11/1/2016	JL KLEIN	UNC FAM MED

  - Seems straightforward and well-organized!

Where flat files fall short

- In your career, you've probably heard someone talk about their "Excel database."
  - This isn't inaccurate! A database is an structured collection of data, stored electronically—and *flat files*, like Excel sheets, CSVs, and other types of *delimited files*, fit that bill.
  - Here's an example of a flat file used to track patient visits.

Patient Name	Patient Address	Patient Date of Birth	Visit Date	Visit Provider	Clinic Name
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	11/1/2016	JL KLEIN	UNC FAM MED
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	5/3/2016	J LARRY KLEIN	UNC FAM MED
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	8/4/2015	PHILIP RENSLER	Internal Medicine

  - Now that our patient has more than one visit, and some information changes over time and some doesn't, we start to run into some problems.

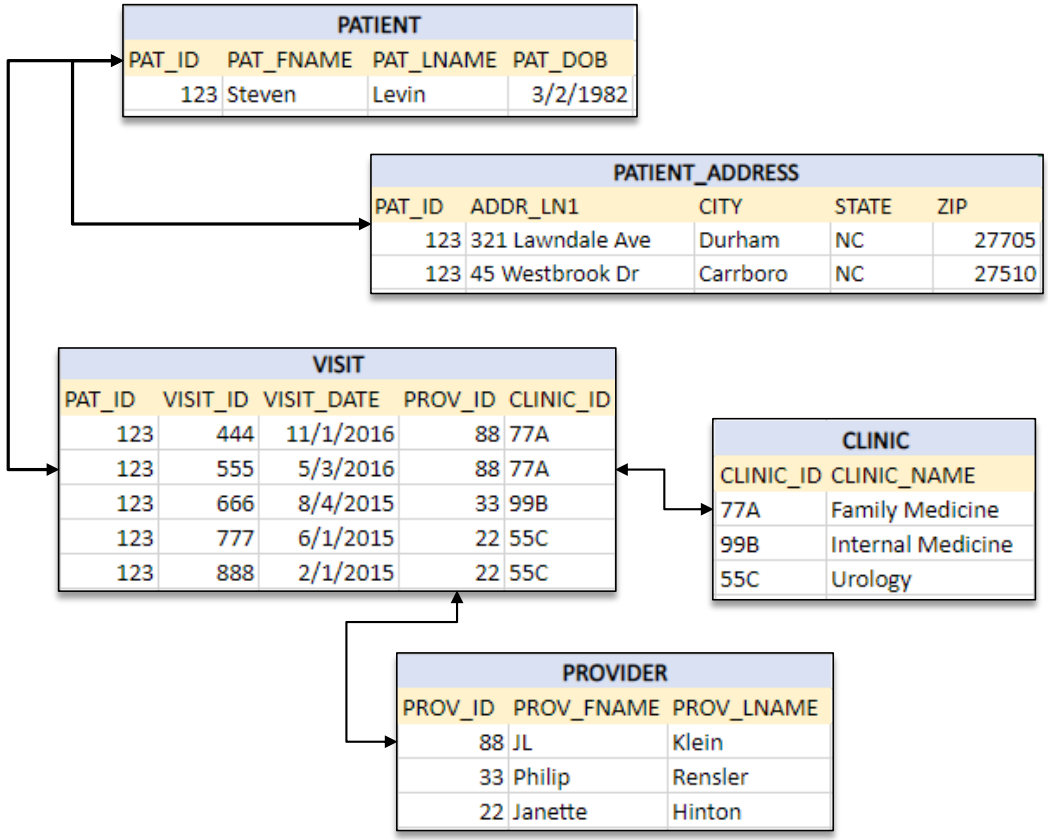
Where flat files fall short

- In your career, you've probably heard someone talk about their "Excel database."
  - This isn't inaccurate! A database is an structured collection of data, stored electronically—and *flat files*, like Excel sheets, CSVs, and other types of *delimited files*, fit that bill.

Patient Name	Patient Address	Patient Date of Birth	Visit Date	Visit Provider	Clinic Name
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	11/1/2016	JL KLEIN	UNC FAM MED
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	5/3/2016	J LARRY KLEIN	UNC FAM MED
Steven Levin	321 Lawndale Ave., Durham, NC 27705	3/2/1982	8/4/2015	PHILIP RENSLER	Internal Medicine
Steven Levin	321 Lawndale Ave., Durham, NC 27705	Mar-2-82	6/1/2015	JANETTE HINTON	UROLOGY
Steven Levin	45 Westbrook Dr., Carrboro NC, 27510	Mar-2-82	2/1/2015	JANETTE HINTON	urol.

  - With even more data, we see more problems—inconsistent formatting, lots of duplication, and human error. This is the point where flat files fall short, and a relational database can help.

From flat file to relational database



PATIENT

PAT_ID	PAT_FNAME	PAT_LNAME	PAT_DOB
123	Steven	Levin	3/2/1982

PATIENT_ADDRESS

PAT_ID	ADDR_LN1	CITY	STATE	ZIP
123	321 Lawndale Ave	Durham	NC	27705
123	45 Westbrook Dr	Carrboro	NC	27510

VISIT

PAT_ID	VISIT_ID	VISIT_DATE	PROV_ID	CLINIC_ID
123	444	11/1/2016	88	77A
123	555	5/3/2016	88	77A
123	666	8/4/2015	33	99B
123	777	6/1/2015	22	55C
123	888	2/1/2015	22	55C

CLINIC

CLINIC_ID	CLINIC_NAME
77A	Family Medicine
99B	Internal Medicine
55C	Urology

PROVIDER

PROV_ID	PROV_FNAME	PROV_LNAME
88	JL	Klein
33	Philip	Rensler
22	Janette	Hinton

Source System Data

- Source system data is much more difficult to control.
- Data warehouses can enforce common codesets to link data from different source systems and make them cross-queryable.

Patient_ID	Race
999	Black

EHR System

Patient_ID	Race
992378	Blk

Radiology System

Patient_ID	Race
A7TR3	AfAm

Research System

Patient_ID	Race
1020	02

Billing System

Source System Data

- Source system data is much more difficult to control.
- Data warehouses can enforce common codesets to link data from different source systems and make them cross-queryable.

Code_Type	Source	Source_Val	Common_Cd
RACE	EHR	Black	003
RACE	RAD	Blk	003
RACE	RSCH	AfAm	003
RACE	BLNG	02	003

Patient_ID	Race
999	Black

EHR System

Patient_ID	Race
992378	Blk

Radiology System

Patient_ID	Race
A7TR3	AfAm

Research System

Patient_ID	Race
1020	02

Billing System

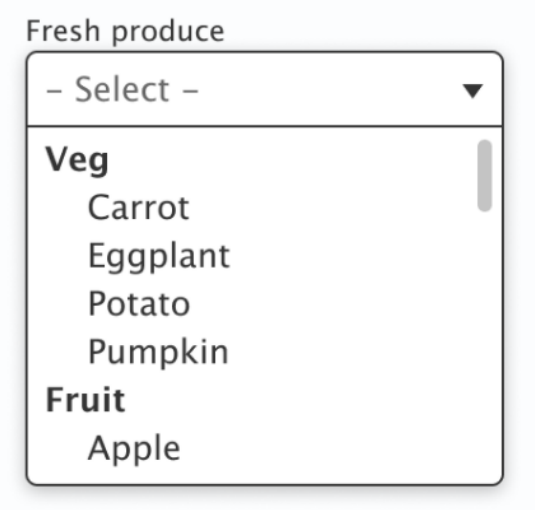

Clinical Terminologies

Ontologies and Terminologies

- Ontology: systematic representation of knowledge; a way to consistently classify and define data across users, databases, institutions, and even countries.
- Terminology: controlled vocabulary in a given domain
- Both provide common terms and definitions for, in this case, clinical concepts
- Ontologies also allow for computational logical reasoning on the relationships between concepts.

Ontologies and Terminologies

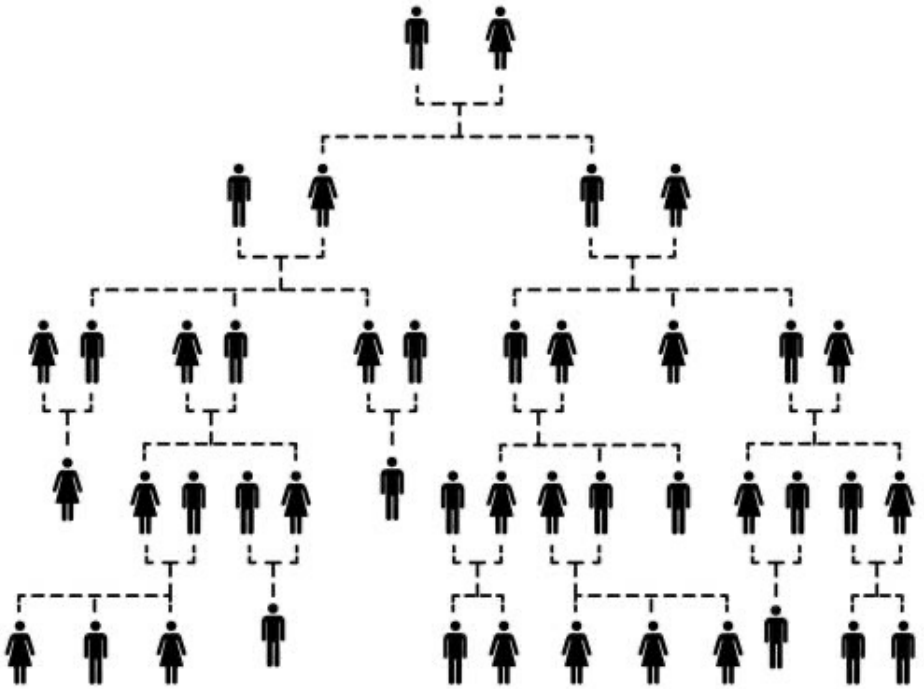
- In order to “find patients with disease X,” or “identify visits where Y occurred,” you need a way to consistently name X and Y.
- In many (but not all) cases, EHRs (and sometimes regulations) force clinicians to use consistent terminology to name things.



← You can try all you want, it still won’t let you pick “chocolate cake.”

Ontologies and Terminologies

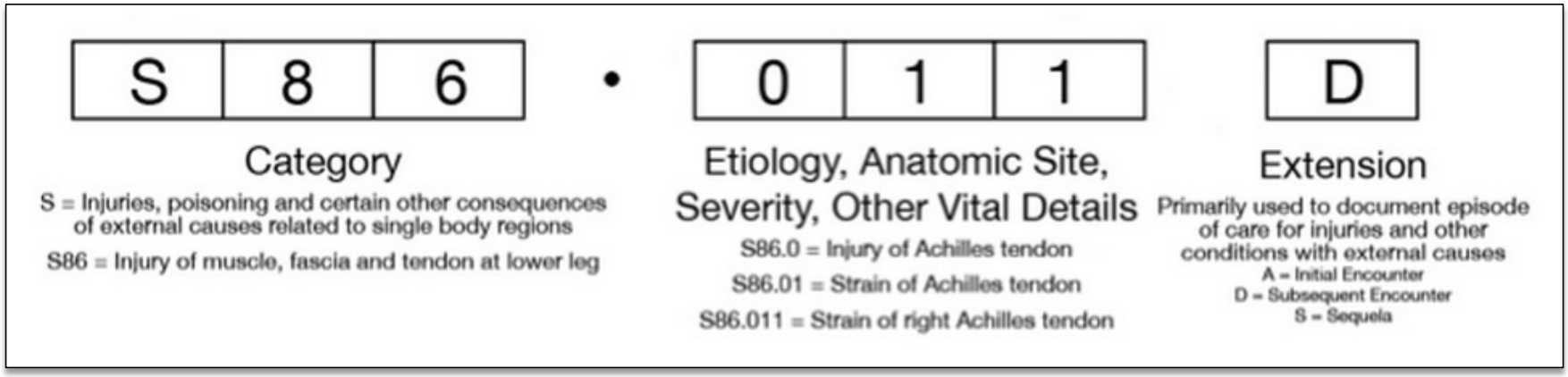
- **Family Terminology**:
- (Grand)father
- (Grand)mother
- (Grand)parent
- Sister
- Brother
- Sibling
- Uncle
- Aunt
- Cousin



Diagnoses: ICD-10-CM

- ICD: International Statistical Classification of Diseases and Related Health Problems; maintained by the World Health Organization (WHO).
- Contains “diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases”
- ICD-10-CM is a subset of ICD-10. Developed by Medicare/Medicaid and the National Center for Health Statistics. ~70,000 codes.
- Used for disease reporting, billing. *Use is required* for Medicare/Medicaid reimbursement.

Anatomy of an ICD-10-CM code



Source: https://www.webpt.com/blog/understanding-icd-10-code-structure/

Note: UNC's i2b2 instance is a great place to explore the ICD-10-CM hierarchy.

ICD-9

- Used until October 1, 2015. Needed when working with pre-2015 data.

ICD-9-CM	ICD-10-CM
• 3 - 5 characters • First character is numeric or alpha (E or V) • Characters 2- 5 are numeric • Always at least 3 characters • Use of decimal after 3 characters	• 3 - 7 characters • Character 1 is alpha (all letters except U are used) • Character 2 is numeric • Characters 3 - 7 are alpha or numeric • Use of decimal after 3 characters • Use of dummy placeholder "x" • Alpha characters are not case-sensitive

Source: https://www.cms.gov/medicare/coding/icd10/downloads/032310_icd10_slides.pdf

Procedures: ICD-10-PCS and HCPCS/CPT

- Facility-billed procedures are generally coded with ICD-10-PCS. Provider-billed procedures are generally coded with HCPCS/CPT.
- ICD-9 had its own method of procedure coding, prior to 10/1/2015.
- CPT: Current Procedural Terminology. Requires a license to use.
  - Tip: Due to the license restrictions, use UNC's i2b2 to browse the codes.
- CPT is a subset of HCPCS (Healthcare Common Procedure Coding System).
- Good idea to use both terminologies when searching for procedures.

CPT Code	**ICD-10-PCS Code**
78608	C030MZZ
Brain PET/CT	Brain PET

Medications: RxNorm and NDC

- There are many, many codes for each drug in both RxNorm and NDC.
- Nice overview of RxNorm
- Browse RxNorm codes in RxNav
- The National Drug Code (NDC) is another system that allows manufacturers to code their own products. It can be a very confusing system.
  - If drugs are coded with NDCs, I find it easier to search by name. (Yikes.)
  - RxNav includes an NDC crosswalk.

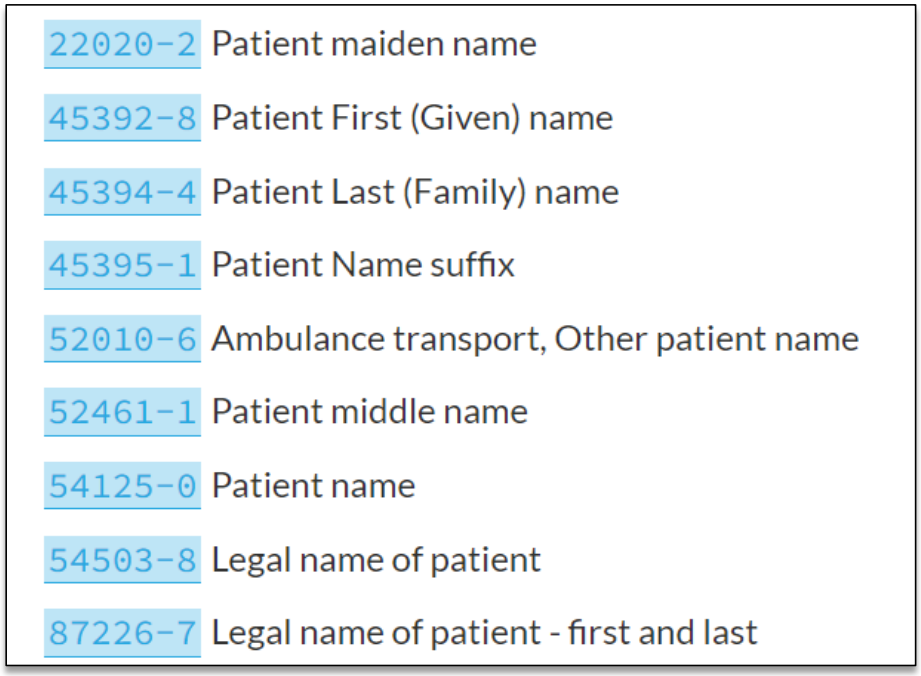
RxNorm	**NDC**
26225	00409112011
Ondansetron	2 ML ondansetron 2 MG/ML
(ingredient)	Prefilled Syringe

Labs (and other stuff): LOINC

- Logical Observation Identifiers Names and Codes.
- Not required for billing, but “preferred” (i.e., considered a best practice for interoperability)
- Two parts: (1) Laboratory LOINC, (2) Clinical LOINC. At present, lab LOINC is much more commonly used in EHRs
- Browse through LOINC-coded COVID tests

LOINC is not very user friendly.

- No rhyme or reason to the LOINC codes.
- On the Clinicial side, LOINC allows synonymous concepts to have different codes.
- Additionally, synonyms are not grouped together, and naming is not consistent.
- Very, very hard to identify “all LOINC codes that may contain patient name.”



LOINC codes are not often used for *results*.

- Tests can have quantitative or qualitative results.
- Qualitative results have LOINC codes, but they are not often used. Tests are standardized, results are not.

Example Answer List LL2021-5

Source: Gen-Probe

Answer	Code	Score	Answer ID
Positive © http://snomed.info/sct ID:10828004 Positive (qualifier value)			LA6576-8
Negative © http://snomed.info/sct ID:260385009 Negative (qualifier value)			LA6577-6
Invalid			LA15841-2

- Looking for test results is an exercise in string matching and synonym finding.

SNOMED

- SNOMED is intended to be the one vocab to rule them all—but isn't, at least in the US.
- Multi-domain (diagnoses, procedures, meds, etc.) and multi-lingual. Used extensively across the world—less so in the US.
- It's an ontology! You can computationally traverse the SNOMED graph.

71388002 |Procedure| :
405815000 |Procedure device| = 122456005 |Laser device| ,
260686004 |Method| = 129304002 |Excision – action| ,
405813007 |Procedure site – direct| = 15497006 |Ovarian structure|

- This example equates to: "the removal of an ovarian structure using a laser device."

SNOMED

Being able to traverse the graph *should* make it easier to gather concepts.

Why care about gathering concepts?

ICD-10-CM: exact code matches, from all over the hierarchy

SNOMED/HPO: conceptual matches, powered by ontologies

Finding the right codes

- There is rarely a single source of truth for the “right” codes.
- UMLS example
- Detective work is almost always necessary: UMLS, Athena, Google

Reminder: All coding systems have the same flaw—human error

Coding Guidance

ICD-10 coding guidance specifies that a code for pre-existing diabetes in pregnancy (O24) should be accompanied by a diabetes code from the E08-E13 range.

≠

Physician Judgment

Jane is pregnant and has Type I diabetes. When she sees her OB for the first time, her OB applies the code O24.019 (Pre-existing type 1 diabetes mellitus, in pregnancy), but does not think to add an E08-E13 code.*

≠

Downstream Researcher

Dr. Felton is running a study on Type I diabetes, and wants to identify all potentially qualifying patients using the EHR. She does a Google search for "Type I diabetes ICD-10," and comes up with E10, which she uses in her search.