Clinical Data

What is Biomedical & Health Informatics?
William Hersh
Copyright 2023
Oregon Health & Science University

What are clinical data?

- A datum is a single observation
- Clinical data are the collection of observations about a patient
  - Example from John Halamka of Geek Doctor blog fame (https://dmice.ohsu.edu/hersh/halamka-record.pdf), part of Personal Genomes Project (https://www.personalgenomes.org/us)
- Each datum about a patient has a minimum of four elements:
  - the patient (Bill Hersh)
  - the attribute (heart rate)
  - the value of the attribute (50 beats per minute)
  - the time of the observation (1:00 pm on 7/1/1990 – many ways to record dates!)

Types of clinical data

- Structured or discrete
  - Numerical measurements – blood pressure, temperature, lab values
  - Coded data – selection from a controlled terminology system
  - Recorded signals – ECG, EEG
  - Images – radiographs, photographs, etc.
- Unstructured or documents
  - Narrative – documents typically generated by clinicians
  - Other textual documents

Uses of clinical data

- Form basis of historical record
- Support communication among providers
- Anticipate future health problems
- Record standard preventive measures
- Coding and billing
- Provide a legal record
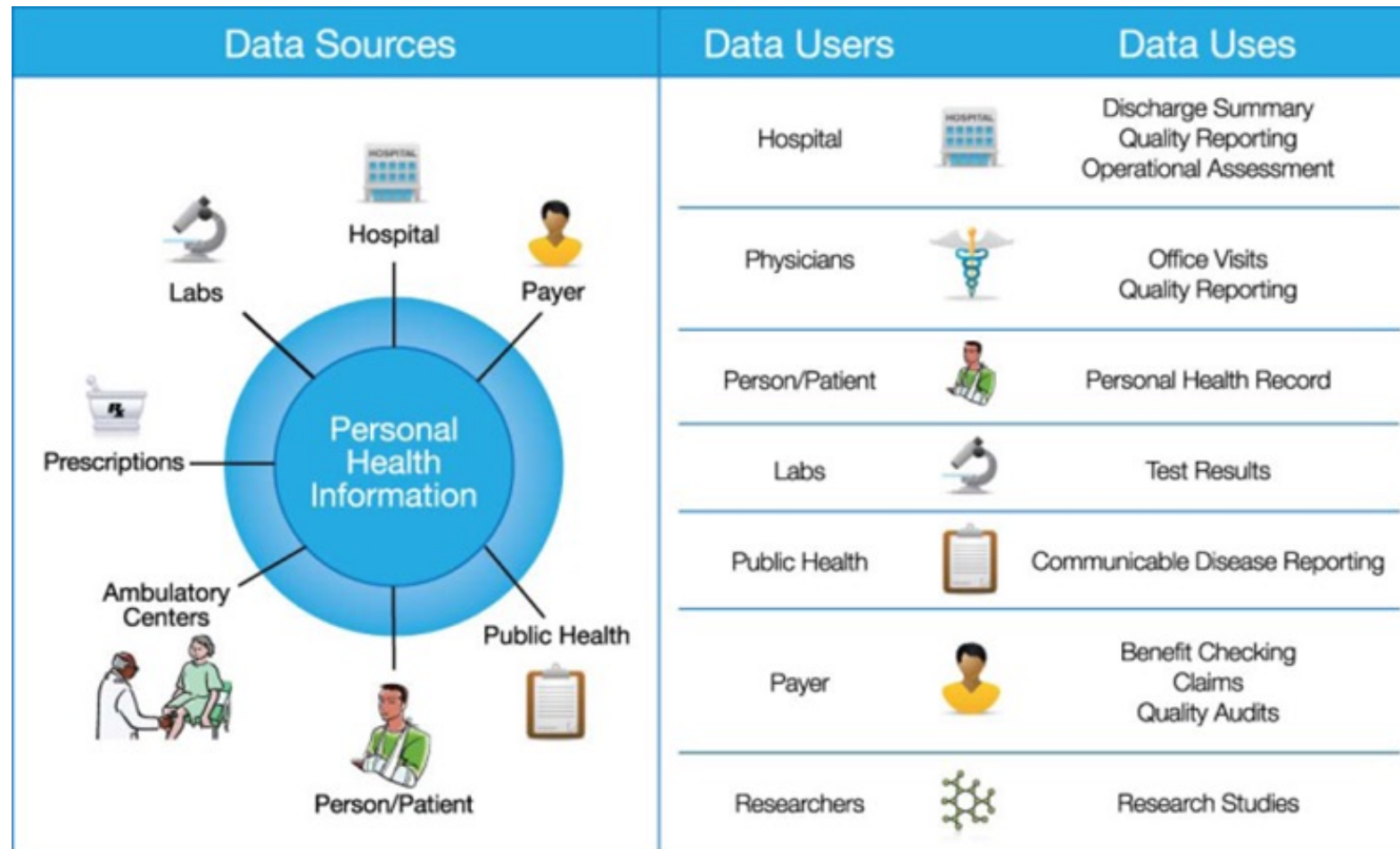- Support clinical research

Types of clinical data documents

- History and physical (H&P) – initial assessment by a clinician

- Progress notes – update of progress by primary, consulting, and ancillary providers

- Reports – by specialists, ancillary providers

- Typical paper chart maintained all patient notes in chronological order, sometimes separated into different components
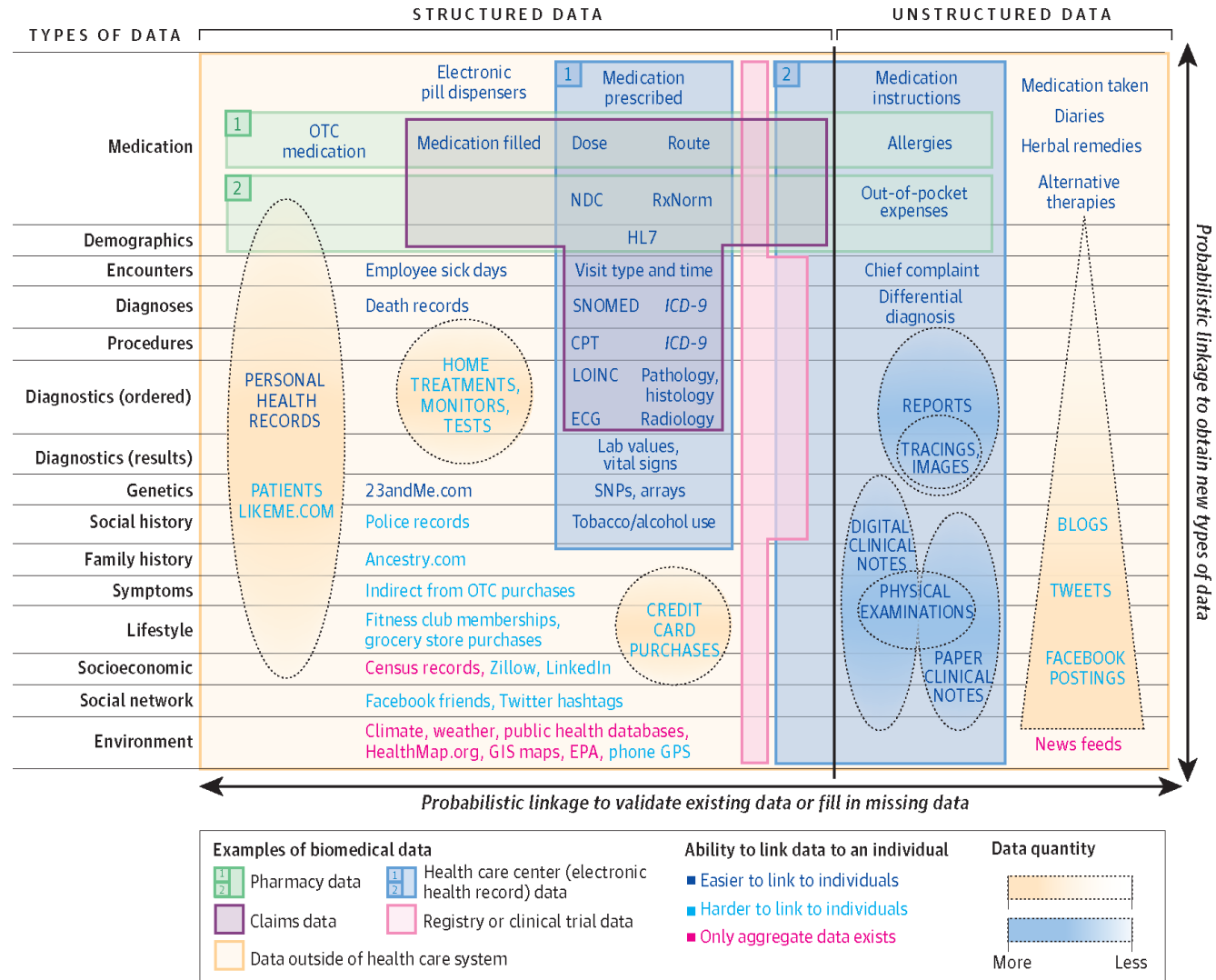
Assessment of a stable patient

- Chief complaint
- History of the present illness
- Past medical history
- Social history
- Family history
- Review of systems
- Physical examination
- Testing – lab, x-ray, other
- Assessment and plan

Many sources, users, and uses of clinical data (NCVHS, 2009)

Data Sources	Data Users	Data Uses
Hospital, Labs, Payer, Prescriptions, Ambulatory Centers, Public Health, Person/Patient → Personal Health Information	Hospital	Discharge Summary, Quality Reporting, Operational Assessment
	Physicians	Office Visits, Quality Reporting
	Person/Patient	Personal Health Record
	Labs	Test Results
	Public Health	Communicable Disease Reporting
	Payer	Benefit Checking, Claims, Quality Audits
	Researchers	Research Studies

OHSU

And many other data sources, including beyond EHR (Weber, 2014)

S T R U C T U R E D   D A T A   U N S T R U C T U R E D   D A T A

TYPES OF DATA

Medication — Electronic pill dispensers; OTC medication; Medication prescribed; Medication filled; Dose; Route; NDC; RxNorm; Medication instructions; Allergies; Out-of-pocket expenses; Medication taken; Diaries; Herbal remedies; Alternative therapies; HL7

Demographics

Encounters — Employee sick days; Visit type and time; Chief complaint

Diagnoses — Death records; SNOMED; ICD-9; Differential diagnosis

Procedures — CPT; ICD-9

Diagnostics (ordered) — PERSONAL HEALTH RECORDS; HOME TREATMENTS, MONITORS, TESTS; LOINC; Pathology, histology; ECG; Radiology; REPORTS

Diagnostics (results) — Lab values, vital signs; TRACINGS, IMAGES

Genetics — 23andMe.com; SNPs, arrays

Social history — PATIENTS LIKEME.COM; Police records; Tobacco/alcohol use; DIGITAL CLINICAL NOTES; BLOGS

Family history — Ancestry.com

Symptoms — Indirect from OTC purchases; PHYSICAL EXAMINATIONS; TWEETS

Lifestyle — Fitness club memberships, grocery store purchases; CREDIT CARD PURCHASES

Socioeconomic — Census records, Zillow, LinkedIn; PAPER CLINICAL NOTES; FACEBOOK POSTINGS

Social network — Facebook friends, Twitter hashtags

Environment — Climate, weather, public health databases, HealthMap.org, GIS maps, EPA, phone GPS; News feeds

Probabilistic linkage to obtain new types of data

Probabilistic linkage to validate existing data or fill in missing data

Examples of biomedical data
- Pharmacy data
- Claims data
- Data outside of health care system
- Health care center (electronic health record) data
- Registry or clinical trial data

Ability to link data to an individual
- Easier to link to individuals
- Harder to link to individuals
- Only aggregate data exists

Data quantity
More — Less

WhatIs05

OHSU

Another important part of data: social determinants of health (SDOH)

(Artiga, 2018)

Economic Stability	Neighborhood and Physical Environment	Education	Food	Community and Social Context	Health Care System
Employment	Housing	Literacy	Hunger	Social integration	Health coverage
Income	Transportation	Language	Access to healthy options	Support systems	Provider availability
Expenses	Safety	Early childhood education		Community engagement	Provider linguistic and cultural competency
Debt	Parks	Vocational training		Discrimination	
Medical bills	Playgrounds	Higher education		Stress	Quality of care
Support	Walkability				
	Zip code / geography				

Health Outcomes
Mortality, Morbidity, Life Expectancy, Health Care Expenditures, Health Status, Functional Limitations

OHSU

Some complications of data

- Circumstances of observation – e.g., how was heart rate taken? pulse? ECG?
- Uncertainty – how accurate is patient reporting, measurement, device?
- Time – what level of specificity do we need?
- Imprecision vs. inaccuracy

Special populations and types of data

- Children – critical functions include (Dufendach, 2015; O'Donnell, 2020)
  - Tracking of well-child visits
  - Support for anthropometric analysis
  - Immunization tracking and forecasting
  - Weight-based drug dosing
- Social and behavioral domains (Fraze, 2019)
  - Only 24% of hospitals and 16% of physician practices reported screening for food insecurity, housing instability, utility needs, transportation needs, and interpersonal violence

Coding of clinical data

- Historically performed by a Clinical Coding Specialist (CCS)
- Major purpose has historically been for reimbursement (Scott, 2013; Basch, 2018)
- A core issue in biomedical informatics has been how to generate and use coded data for other purposes
- Trade-offs
  - Standardization of language vs. freedom of expression
  - Time to narrate vs. code
- Other difficulties
  - Creating and maintaining coding systems
  - Structuring coding systems to capture meaning

OHSU

Other complications, especially for re-use of clinical data

- May be (Hersh, 2013)
  - Inaccurate
  - Incomplete
  - Transformed in ways that undermine meaning
  - Unrecoverable
  - Of unknown provenance
  - Of insufficient granularity
  - Incompatible with research protocols
- Many steps in capture and transformation for analytics and research, all potentially subject to error (Verheij, 2018)
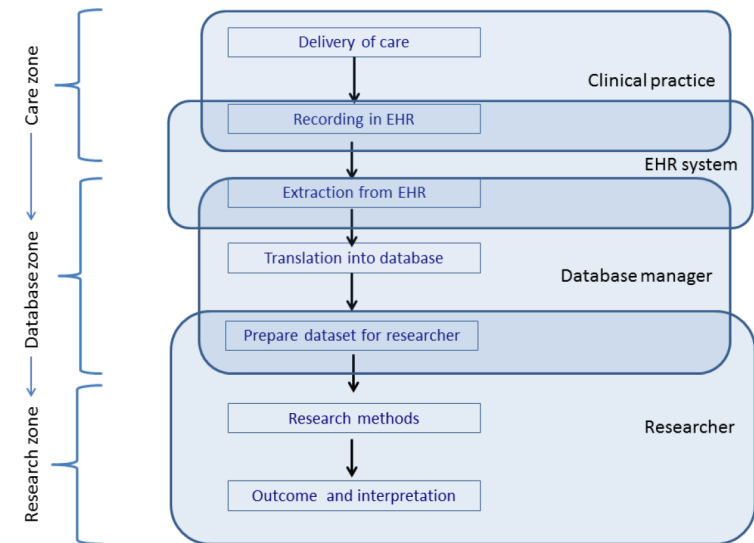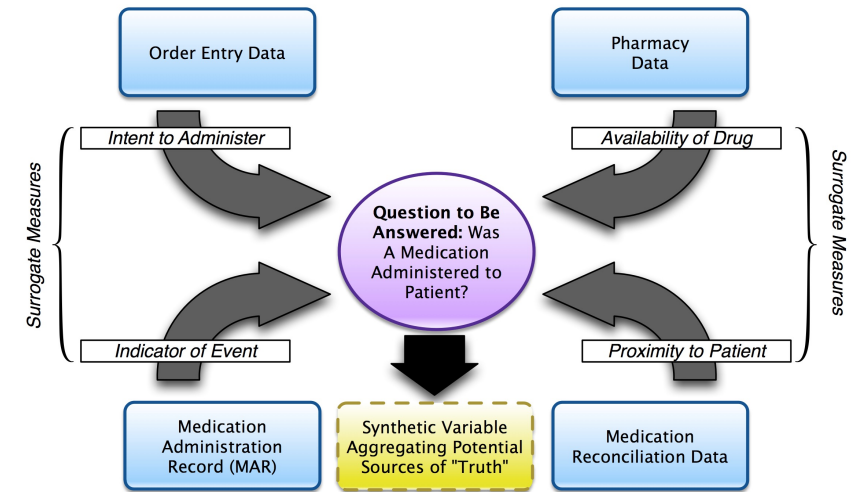
Ambiguity in verbal probabilities (Andreadis, 2021)

Verbal probability term	Number of studies	Average numeric estimate, random effects model (%)	95% CI (%)	Minimum sample average (%)*	Maximum sample average (%)*	Range of individual estimates (%)†
Rare(ly)	7	10.00	[7.99, 12.01]	7.0	21	0–80
Rare-severe event	3	10.06	[5.45, 14.68]	6.3	34.8	–
Rare-mild event	3	14.14	[7.88, 20.40]	9.6	39.3	–
Uncommon	4	17.64	[13.19, 22.09]	13.3	22.9	0–90
Unlikely	6	17.71	[14.86, 20.55]	13.3	27	0–85
Common-severe event	3	43.08	[40.27, 45.88]	41.9	45.6	–
Possible(ly)	6	43.28	[36.66, 49.89]	36.9	62	–
Common-mild event	3	50.47	[45.59, 55.34]	48	58	–
Common	6	58.73	[50.40, 67.06]	34.2	70.5	10–100
Very common	3	60.10	[42.36, 77.85]	38.5	71.6	5–100
Probable(ly)	5	69.87	[67.07, 72.67]	66	73.9	20–100
Likely	6	71.87	[69.90, 73.84]	66	94	–
Usual(ly)	3	75.38	[71.53, 79.23]	72	78	–
Very likely	3	84.30	[79.43, 89.17]	75.2	93	20–100

Use of stigmatizing language

- Negative and positive categories (Park, 2021)
- Found in 2.5% of all H&P notes, higher for (Himmelstein, 2022)
  - Patients with diabetes, substance abuse, and chronic pain
  - Non-Hispanic black patients
- Negative descriptors 2.5 times more likely for black patients in H&P notes (Sun, 2022)

Table 1. Negative Language Categories[a]

Categories	Definitions	Examples[b]
Questioning credibility	Implication of physician disbelief of patient reports of their own experience or behaviors	• He insists the pain is behind his knee. • He claims that nicotine patches don't work for him. • I listed several fictitious medication names and she reported she was taking them, and that she takes "whatever is written there"
Disapproval	Highlights poor reasoning, decision-making, or self-care, usually in a way that conveys the patient is unreasonable	• Reports that if she were to fall, she would just "lay there" until someone found her • He was adamant that he does not have prostate cancer because his "bowels are working fine." • Counseled that there is no evidence for this, but patient has strong beliefs. • She is adamant that she cannot perform any kind of exercise due to pain and will not change her diet.
Stereotyping	Quoting African American Vernacular English	• Chief complaint - "I stay tired" • Reports that the bandage got "a li'l wet"
	Quoting incorrect grammar or unsophisticated terms	• States that the lesion "busted open" • Reports she was unable to fill prescription for the "sugar pill"
Difficult patient	Inclusion of details with questionable clinical significance that depict the patient as belligerent or otherwise suggests that the physician is annoyed	• She persevered on the fact that "a lot of stuff is going on at home with my family" but that "you wouldn't understand." • I informed her that this is unlikely to be helped by antibiotics and talked about smoking cessation with her. She said she will ask her 'sinus doctor' for antibiotics.
Unilateral Decisions	Language that emphasizes physician authority over patient	• She was told to discontinue... • I have instructed him to...

We also need effective "stewardship" of data

- Ensuring the "knowledgeable and appropriate" use of data from individuals' personal health information (NCVHS, 2009)

- Maintaining the patient's story in the record (Vigilante, 2018; Gantzer, 2020) – even if EHRs can make that difficult (Kommer, 2018)