The Digital Architecture for the Learning Health System

Monday, January 27, 2020

Christopher B. Forrest, MD, PhD
Professor of Pediatrics
Children's Hospital of Philadelphia
Director, PEDSnet (pedsnet.org)
Program Director, PEDSnet Scholars

Agenda for Today

1. Review LHS concepts to contextualize applied informatics
2. Set expectations for data science group projects
3. Provide a vision for informatics in a high functioning LHS
4. Define and discuss essential applied informatics concepts
5. Digital architecture for PEDSnet
6. Review how the EHR can be used for research and registries
7. Computable phenotyping and population denominators

LHS Training Competencies Addressed

Domain	Competency
Informatics	4.1: Demonstrate ability to use data derived from electronic health records and other clinical information sources for research and quality improvement.
	4.4: Demonstrate knowledge of population health informatics, including disease surveillance, monitoring of community health, assessment of social and behavioral determinants of health, and geographic information systems.
	4.5: Demonstrate knowledge of clinical information systems, including electronic health records, clinical documentation, computerized physician order entry (CPOE), clinical decision support systems, electronic prescribing, medical imaging, and clinical/population dashboards.

1. Review key LHS concepts to contextualize applied informatics

Learning Health Systems

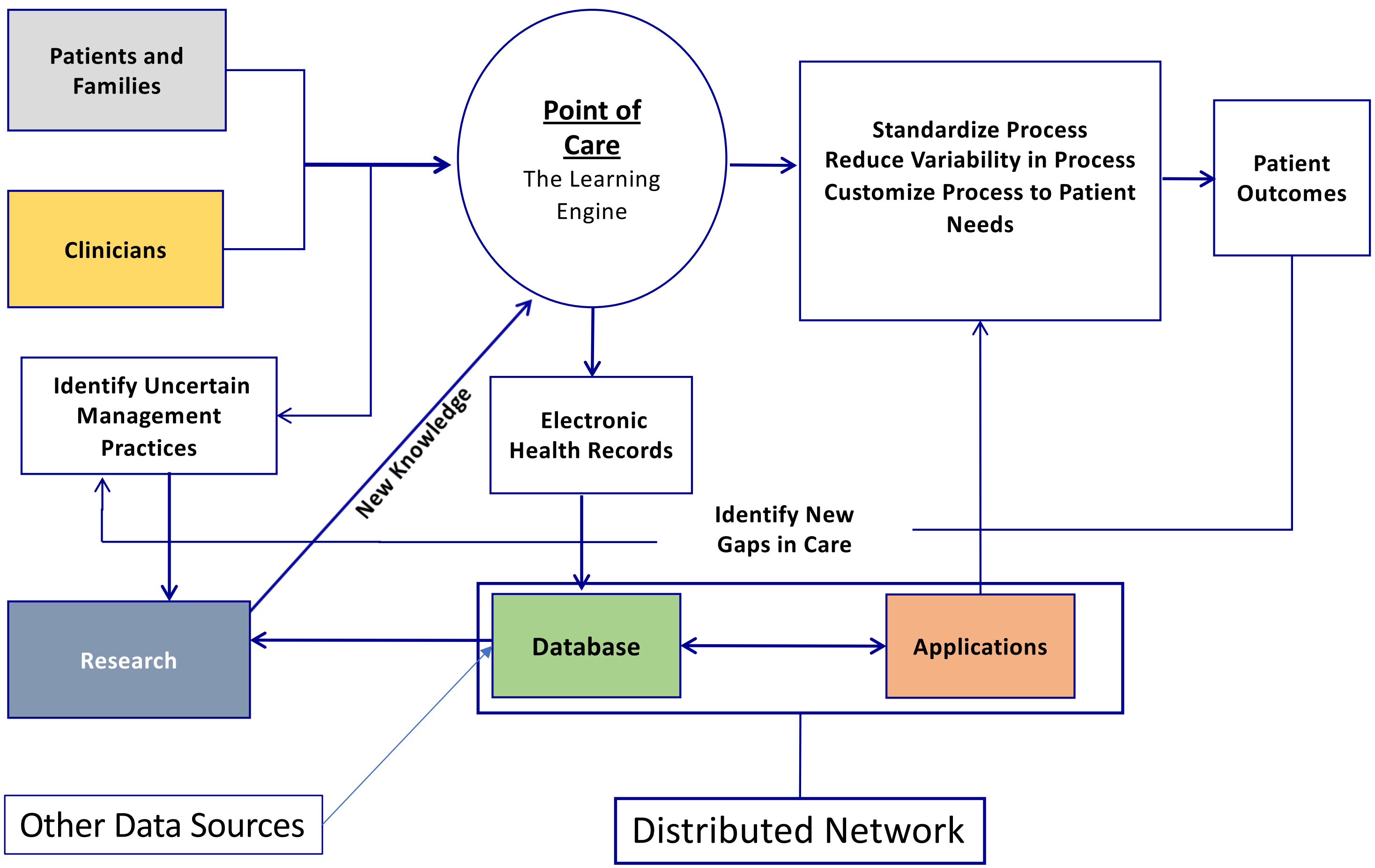
Health systems, at any scale, that can continuously and routinely study and improve themselves.

The NEW ENGLAND JOURNAL of MEDICINE

Perspective: Jan 3, 2013
"Code Red and Blue — Safely Limiting Health Care's GDP Footprint"
Arnold Milstein, M.D., M.P.H.

…U.S. health care needs to adopt new work methods, outlined in the Institute of Medicine's vision for a learning health system…

Improving Outcomes with a Learning Health System



Learning Health Systems

1. Focus on the **outcomes** of people & systems

2. **Research ⇔ Improvement** seamlessly linked

3. Leverage **existing data**, such as EHRs

4. Deeply **engage** stakeholders

5. **Embed** researchers in the system

Source: National Academy of Medicine, 2013: Forrest et al, Health Affairs, 2014

2. Data science group project

Three groups

Group 1: Bourque, Rao, Smith-Parrish, Bunnell

Group 2: Chiotos, Hildenbrand, Ong, Razzaghi

Group 3: Schultz, Shah, Sood, Varnell, Bailey

Tasks

1. Define the research question: novel and feasible
2. Create the scientific specifications
3. Develop the code-sets
4. Implement the specifications
5. Review results
6. Report results

1-3 will be done before FtF meeting

4-6 will be done at the FtF meeting

3. Informatics vision for a high functioning learning health system

2-year old boy with with a three day history of vomiting and one day of lethargy.

The boy has been a patient of one of a PEDSnet primary care practices since birth, so his entire medial history is stored in PEDSnet Data Trust when he presents to the ED. He is seen by the triage nurse within 5 minutes of arrival, because his primary care doctor made an electronic referral indicating a high level of urgency. The nurse enters triage data noting presence of fever and abdominal pain. The constellation of these symptoms is processed by the Data Trust, which integrates his past history with current information. A prediction model is used to assign a risk score, which indicates a high risk for a surgical abdomen and calls up the surgical abdomen practice guideline that was developed by PEDSnet hospitals. The nurse acts promptly on this information and triages the patient to urgent care; the boy is seen by a physician within 10 minutes of arrival to the ED. The ED physician examines the patient's EHR, which indicates no past medical problems and 1 visit to primary care for vomiting two days ago. She speaks with the family and examines the patient; the EHR uses voice recognition to add information from this exchange to the record. The doctor is presented with a differential diagnosis according to the pattern of signs, symptoms, and physical exam findings generated during the interview. A diagnosis of viral gastritis is listed as most likely; bowel obstruction and metabolic derangements are also listed. The resident and medical student working with this doctor see the same differential on their handheld personal medical devices and click on knowledge links to obtain brief consults from the literature. After obtaining permission from the family to link the EHR data to the family's personal health record, the doctor adds the patient's genetic profile, which is downloaded from the PHR, to the other data inputs. Synthesis of all this information by the data trust indicates an 85% chance of intussusception. Furthermore, the chances of an enema reduction are 92%, which is a figure based on an examination of all patients with suspected intussusception seen in PEDSnet hospitals. The patient receives appropriate intravenous fluid resuscitation and the radiologist and surgeon are notified; they access all the ED information on their handheld computers. The patient undergoes a successful enema reduction of an intussusception within 45 minutes of entry to the ED. The Attending Physician receives immediate feedback as a video-based instant message from the radiologist once the procedure is complete.

Elements of a high-functioning digital architecture for the LHS

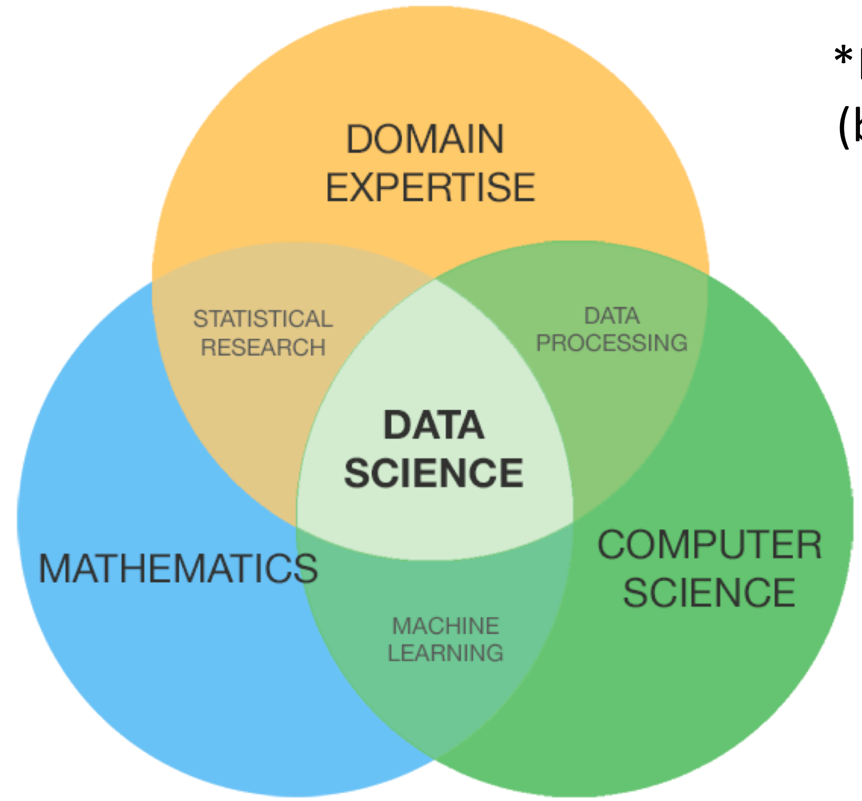
- EHR data linked/linkable to other clinical, social, and personal data
- Worlds information available in real-time
- Knowledge seamlessly shared within and between organizations
- Multi-level: from the microsystem to the national system
- Patient control their APIs
- Multi-stakeholder governance
- Data systems are fully interoperable across institutions
- Appropriate and useful decision support that integrates predictive models with personal data
- Rapid, fit for purpose communication

4. Essential applied informatics concepts

Data Science

Comprises statistical analysis, data mining (i.e., machine learning/AI), database management, and and data retrieval processes on a large amount of data to novel insights (trends, explain events, and (most importantly) predict future outcomes).

Data Science



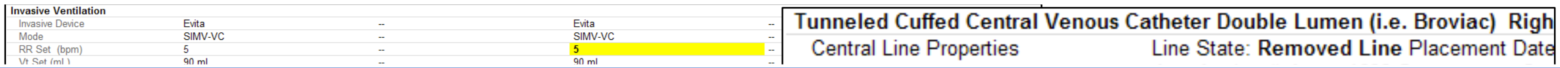
*Big Data not required
(but is often included)

Source: Bob Grundmeir, CHOP

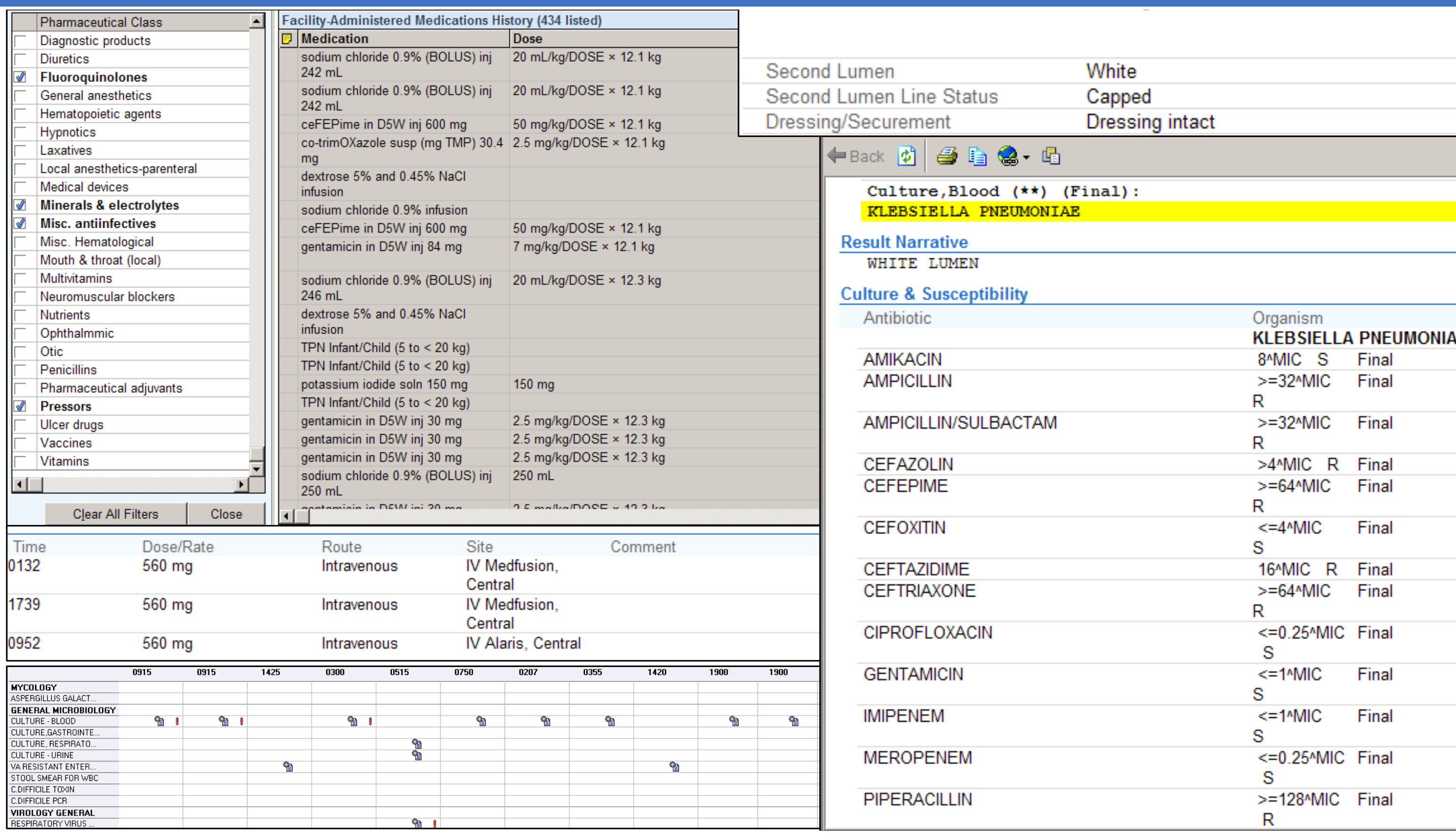
Obtaining Data from your Health System

- For many electronic health records today, data are extracted using structured query language (SQL)
  - Typically data analysts extract data on behalf of research, quality improvement, and operational teams
- Data may be delivered in many formats
  - One or more “spreadsheets”
  - Database backup files
  - eXtensible Markup Language (XML)

Source: Bob Grundmeir, CHOP



The Charts are COMPLEX!



Data Exploration: The Many Forms of Data

The data scientist must readily adapt

- Individual level data (one row per patient)
- Multiple lines of data per patient
- Multiple different levels of data per patient
- Summary information that has already been aggregated across multiple patients
- Numeric, categorical, dates, free text, and mixtures
- Missing and misclassified data

Source: Bob Grundmeir, CHOP

Cleaning Categorical Gender

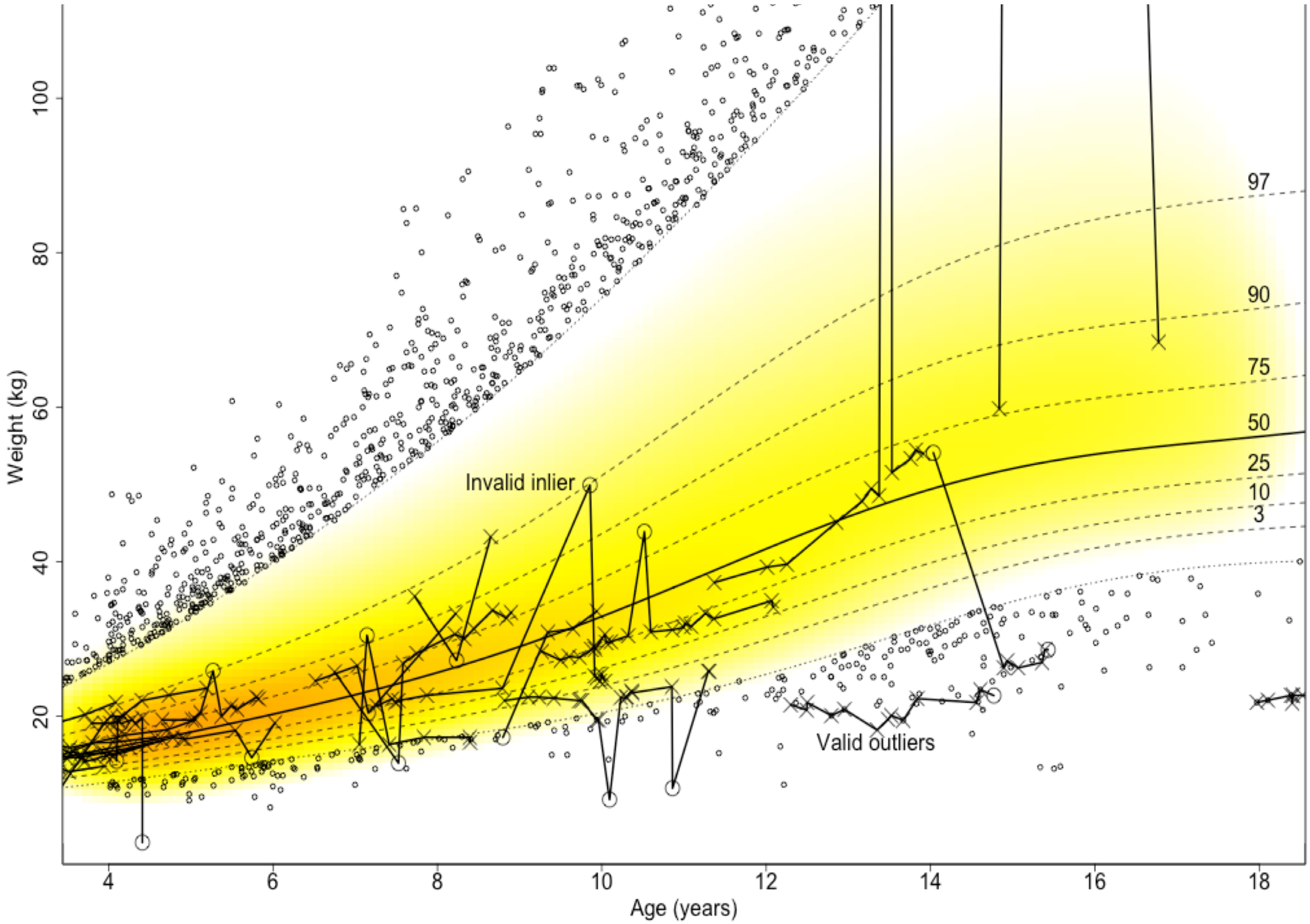
Raw data

source_value	N
M	650268
F	643406
male	36975
female	36955
U	370
unknown	19
O	40

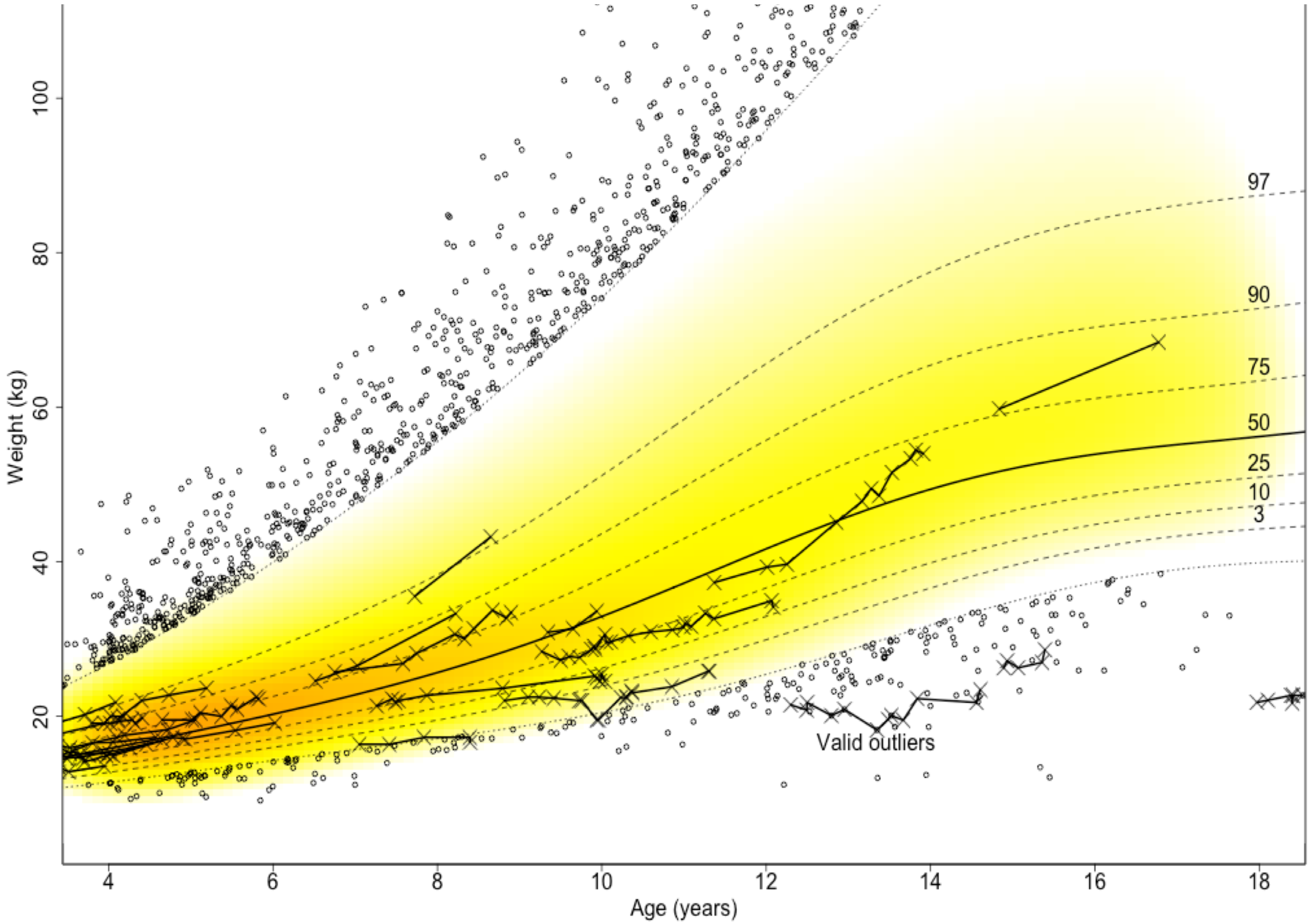
Cleaned data

gender	N
M	687243
F	680361
U	429

Example: Cleaning Numerical Weight Measurements (Before)



Example: Cleaning Numerical Weight Measurements (After)



Electronic health record (EHR)

CMS definition (https://www.cms.gov/Medicare/E-Health/EHealthRecords)

A digital version of a patients medical history, that is maintained by the provider over time, and may include all of the key administrative and clinical data relevant to that persons care under a particular provider, including demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports

EHRs and Administrative Processes

- Better reimbursement
- Fewer chart pulls (no more *library* stacks of charts)
- Improved efficiency of handling telephone messages and medication refills
- Reduced transcription costs
- Increased formulary compliance and clearer prescriptions leading to fewer pharmacy call backs
- Improved coding of visits (better compliance)

EHRs and Quality of Care

- Easier preventive care leading to increased preventive care services (e.g., vaccine reminders)
- Point-of-care decision support
- Rapid and remote access to patient information
- Easier chronic disease management (proactive rather than reactive)
- Integration of evidence-based clinical guidelines
- Better coordination of care (information transfer, integration)

Impact of EHRs on Clinicians

- Less "chart chasing"
- Improved intra-office communication
- Access to patient information while on-call or at home
- Easier compliance with regulations
- Demonstrable high-quality care

But…

- Time consuming data entry
- Hard to retrieve information
- Source of burn-out

EHRs can be linked to…

- Registries
- Health plan claims
- Environmental data-sets (e.g., census, air pollution) via geocodes
- Genomics and molecular diagnostics
- Wearables
- Health-related apps (e.g., sleep, movement)

Registry

“An organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure”

- A registry database is a file (or files) derived from the registry
- Registries can be multipurpose:
  - Study natural history of disease
  - Study disease etiology
  - Assess clinical effectiveness or cost-effectiveness
  - Measure or monitor safety and harm
  - Measure quality of care
- Registries are often defined by the populations they study

Registries can bridge clinical trials and clinical practice by studying effectiveness in real-world populations

EHRs versus Registries

EHRs	Registries
Visit-centered (transactional)	Patient-oriented
Support clinical care, billing, registration, compliance	Purpose-driven
Not designed for research	Designed to derive information on defined exposures and outcomes

Real-world data	Clinical Trials
Results from health interactions or events	Experimental tests of interventions
Real-world evidence – is the intervention effective?	Is the intervention efficacious in controlled trials?
External validity	Internal validity

Common Data Model:
data formats, standard transformations, and definitions/terminologies

SITE 1

- Social Work Visit
- Allied Health
- Office Visit
- Nurse Visit
- Procedure Visit
- Employee Health
- Vascular Lab
- Sleep Study Visit
- Social Work Visit

SITE 2

- Office Visit
- Specimen
- Postpartum Visit
- Clinical Support
- Initial Prenatal

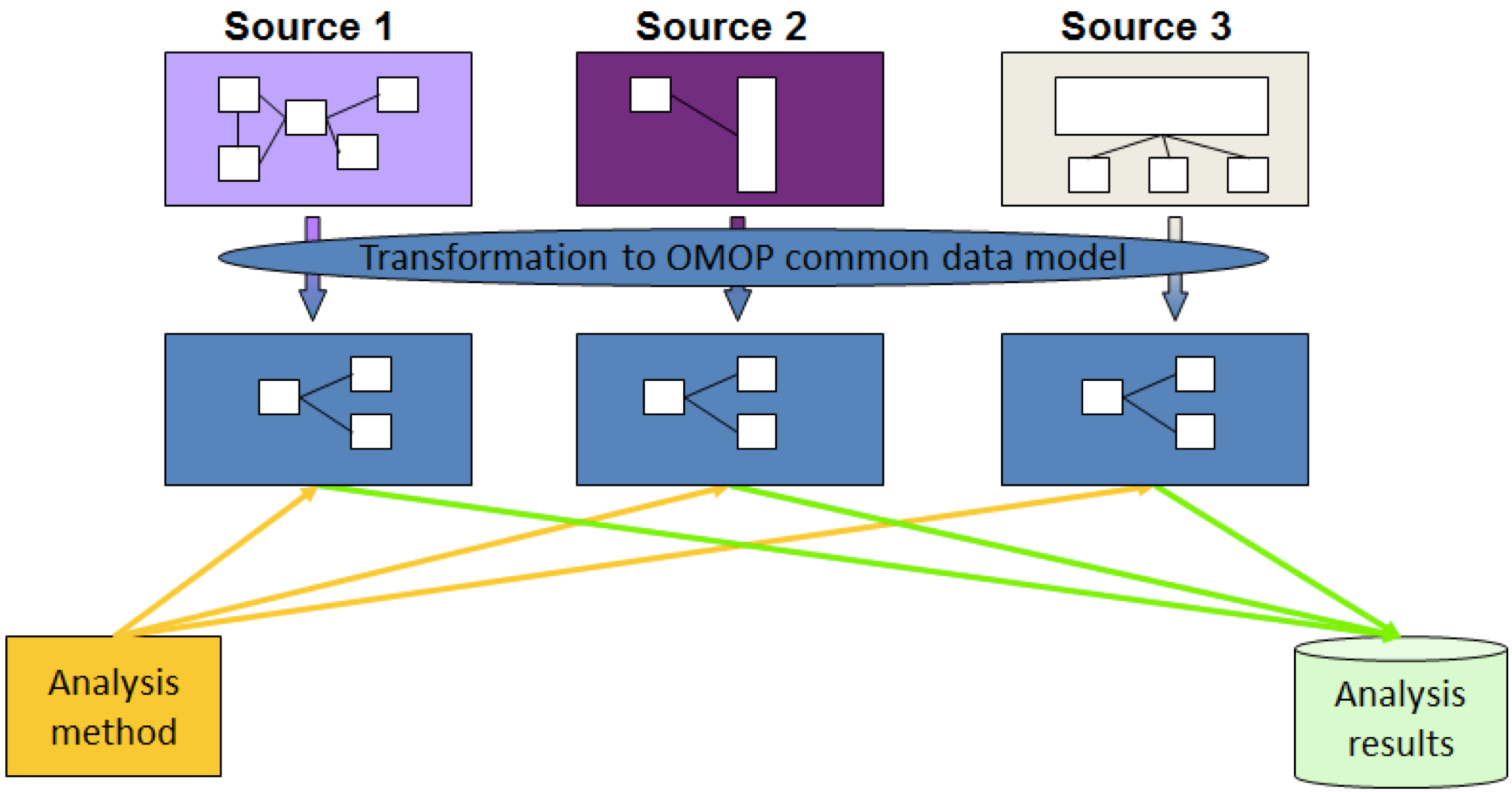
SITE 3

- Home Care Visit
- Office Visit
- Therapy Visit
- Orders Only
- Cardiology Testing
- Hospital Encounter

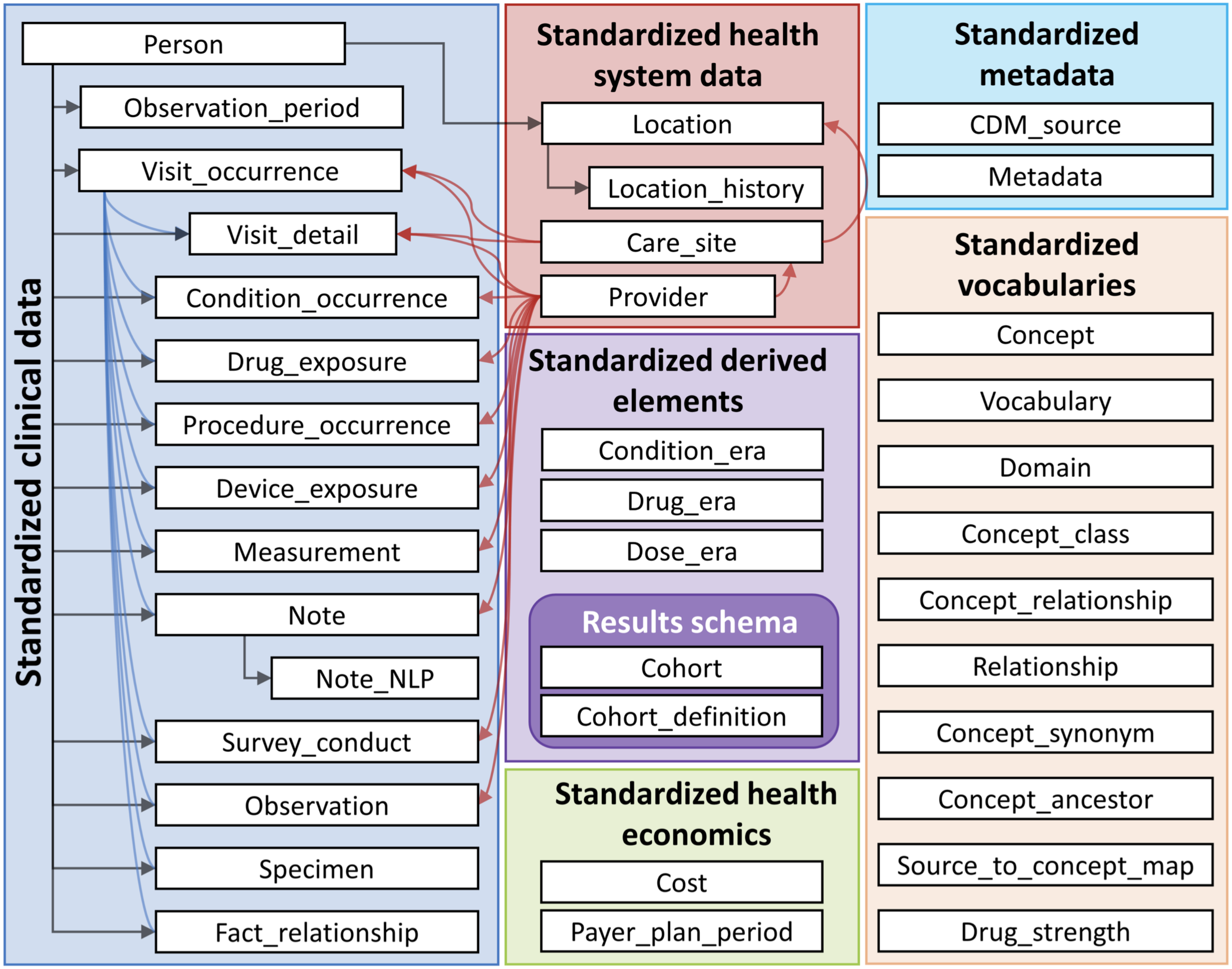
Common Data Model

- Ambulatory Visit (AV)
- Emergency Department (ED)
- ED Admit to Inpatient (EI)
- Inpatient Hospital (IP)
- Non-Acute Inst. Stay (IS)
- Other Ambulatory (OA)
- Other (OT)
- Unknown (UN)
- No Information (NI)
- (null)
- Ambulatory Visit (AV)

Data are represented differently across institutions, even if the same EHR software is used


Source 1
Source 2
Source 3
Transformation to OMOP common data model
Analysis method
Analysis results

OMOP CDM

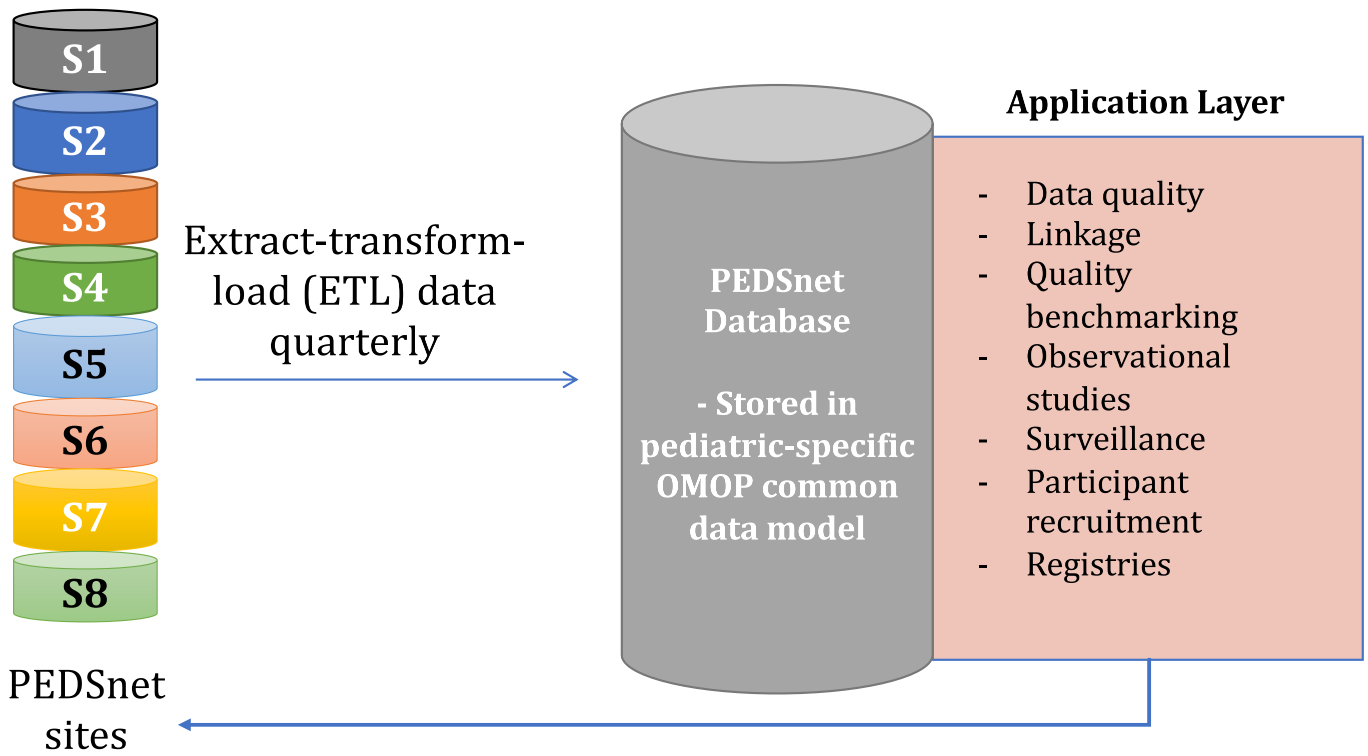

Standardized clinical data
Person
Observation_period
Visit_occurrence
Visit_detail
Condition_occurrence
Drug_exposure
Procedure_occurrence
Device_exposure
Measurement
Note
Note_NLP
Survey_conduct
Observation
Specimen
Fact_relationship
Standardized health system data
Location
Location_history
Care_site
Provider
Standardized derived elements
Condition_era
Drug_era
Dose_era
Results schema
Cohort
Cohort_definition
Standardized health economics
Cost
Payer_plan_period
Standardized metadata
CDM_source
Metadata
Standardized vocabularies
Concept
Vocabulary
Domain
Concept_class
Concept_relationship
Relationship
Concept_synonym
Concept_ancestor
Source_to_concept_map
Drug_strength

Learning More…

- Coursera (online learning – more offerings appear almost daily)
  - “Data Science,” 9 course introduction (Johns Hopkins)
  - “Applied Data Science with Python,” 5 course series (University of Michigan)
- Textbooks
  - “Data Science for Dummies,” Lillian Pierson
  - “Data Science from Scratch,” Joel Grus
  - “Practical Data Science with R,” Nina Zumel

5. Digital architecture for PEDSnet

PEDSnet Data Pipeline



Data Quality Reports
Studies

Structured data types in PEDSnet data

Data Type	Example
Identifiers	Encounter, patient, provider, hospital
Geocodes	Census tract/block based on lat/long; 5-digit zip; county
Diagnoses	SNOMED CT codes; chronic conditions; computable phenotypes
Immunizations	DTaP, HepB, IPV
Medications	RxNorm codes indicating ingredient and route; dosage
Procedures	Surgeries, imaging studies, physical therapy (CPT, HCPC, ICD code-sets)
Laboratories	Ordered and results; LOINC codes
Vital signs	Heart rate, temperature, respiratory rate
Anthropometrics	Height, weight, body mass index, age-sex standardized z-scores
Specialty	General pediatrics, urology, psychiatry, audiology
Utilization	Primary care, outpatient specialty, hospitalization, length of stay, ED visits

Unstructured Data: progress notes, lab reports, imaging reports, operative notes, path reports
Future Data Types: genomics, environmental (social and physical), patient-reported outcomes, patient-entered data

Clinical Terminologies

Provide consistent approach for documenting digital heath data supporting semantic interoperability

Examples/

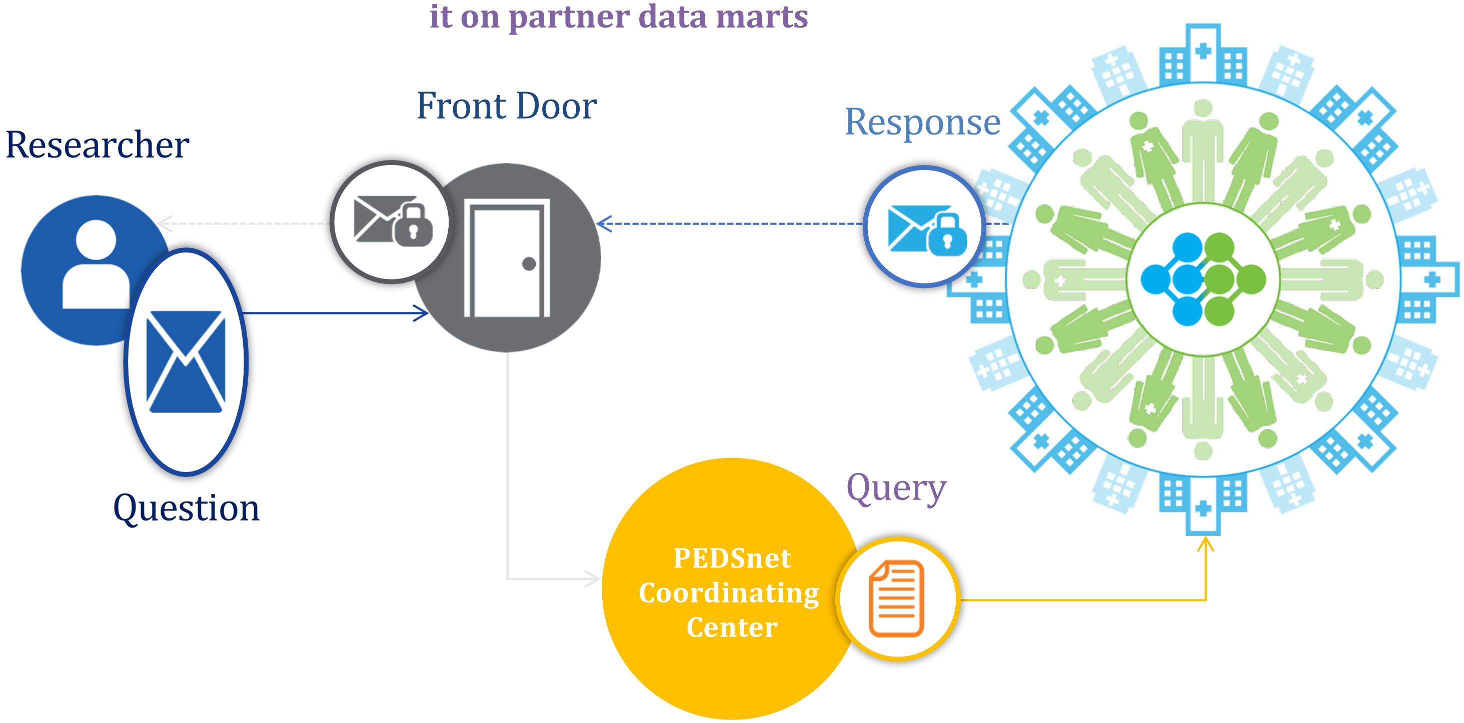
- Diagnosis: SNOMED CT and ICD
- Medications: RxNorm
- Procedures: CPT-4, ICD, HCPC
- Labs: Loinc

Here's how PEDSnet's data network works for researchers

The Researcher sends a question to the PEDSnet Coordinating Center through the Front Door

The Coordinating Center converts the question into a query with an underlying executable code, and executes it on partner data marts

PEDSnet reviews the query results and provides a response, which is sent back through the Front Door to the Researcher



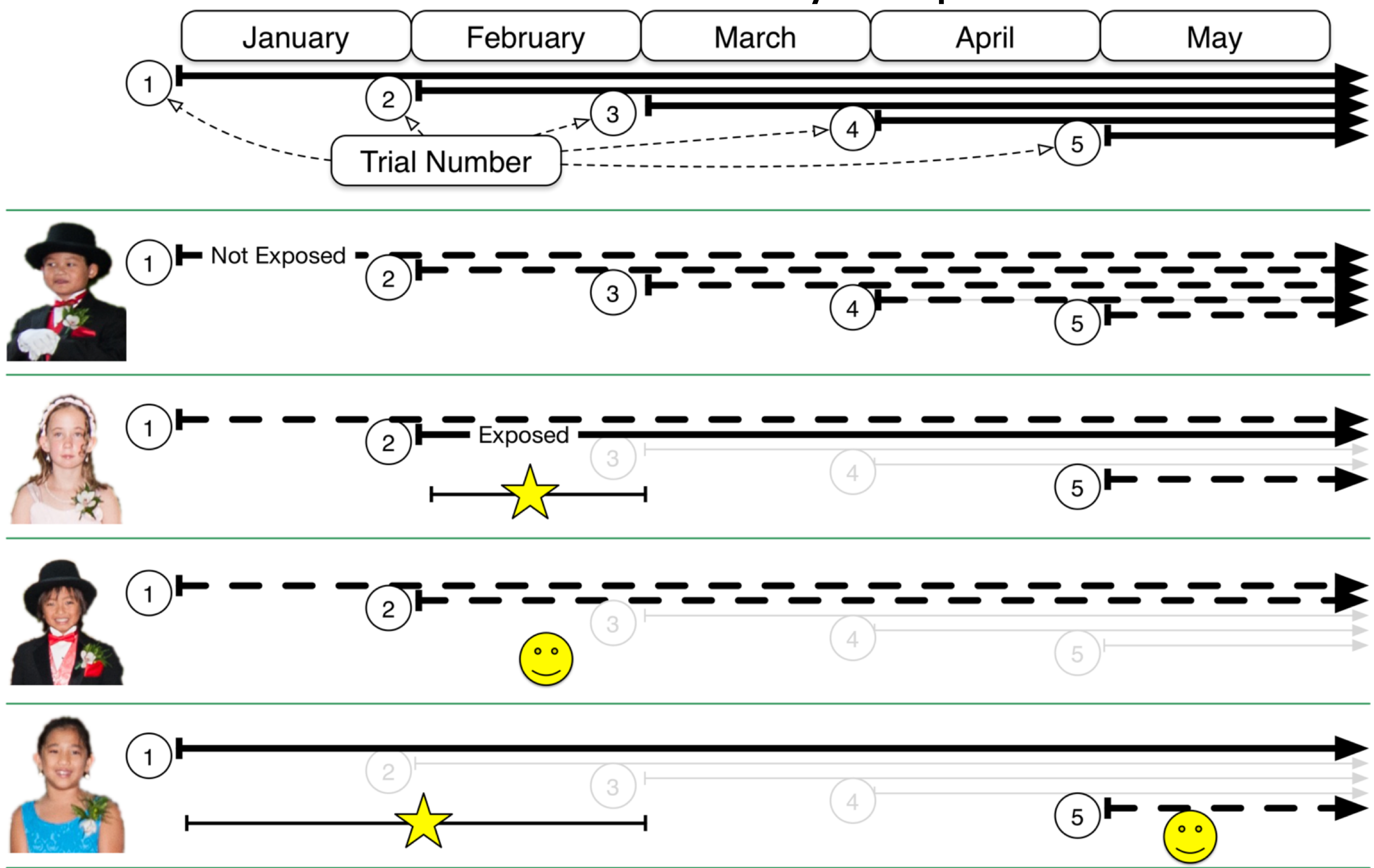
6. Review how the EHR can be used for research and registries

Pseudo-Trials

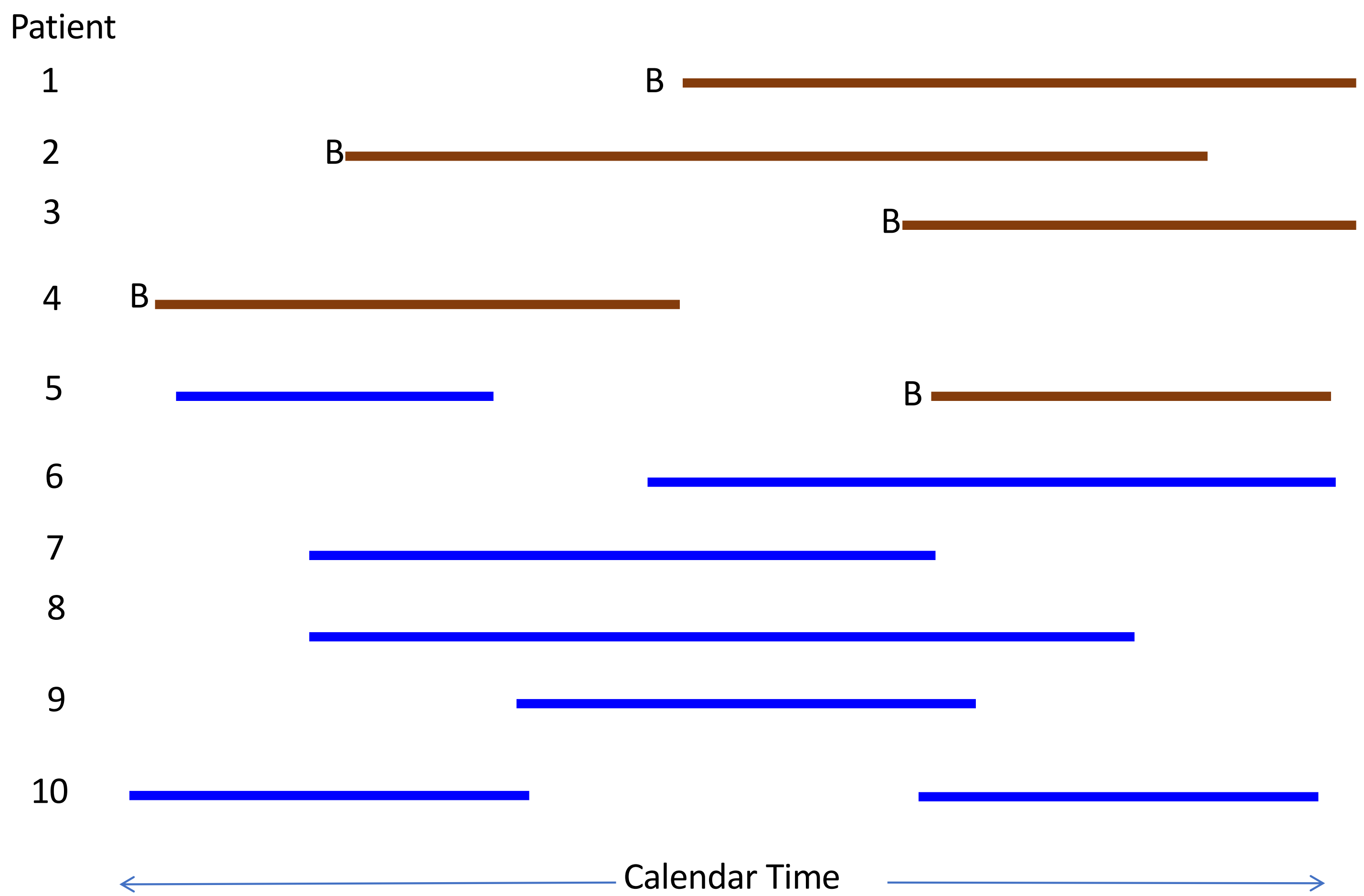
- Method for analyzing observational data like randomized experiments
- Helps in situations where you have a lot of data, but the exposures or outcomes of interest are rare

Ross ME, Kreider AR, Huang YS, Matone M, Rubin DM, Localio AR. Propensity Score Methods for Analyzing Observational Data Like Randomized Experiments: Challenges and Solutions for Rare Outcomes and Exposures. *American Journal of Epidemiology*. 2015 Jun 15;181(12):989-95

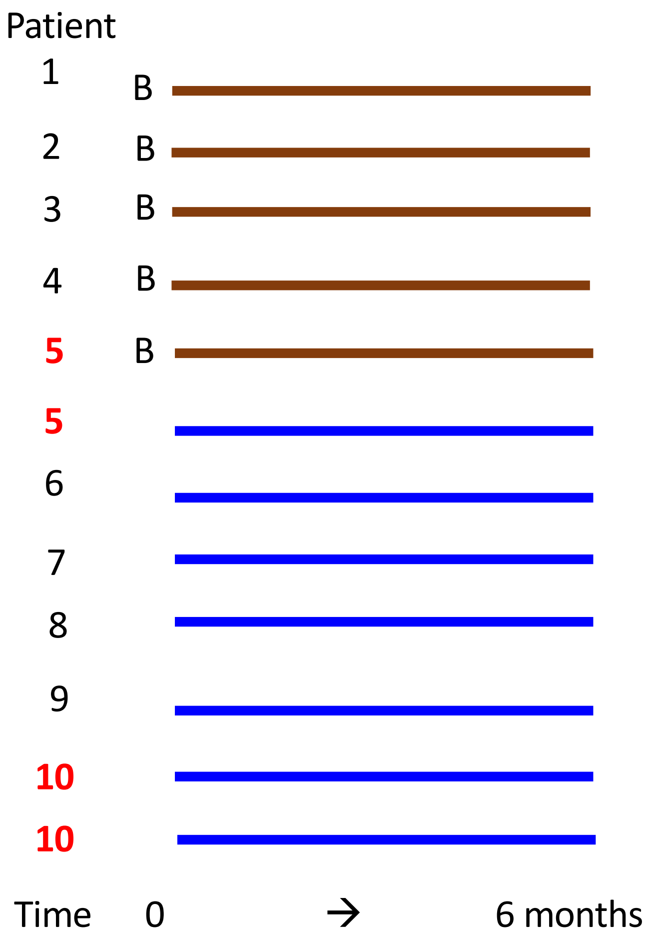
Now enrolling… pseudo trials to see if outcome 🙂 is caused by exposure ⭐



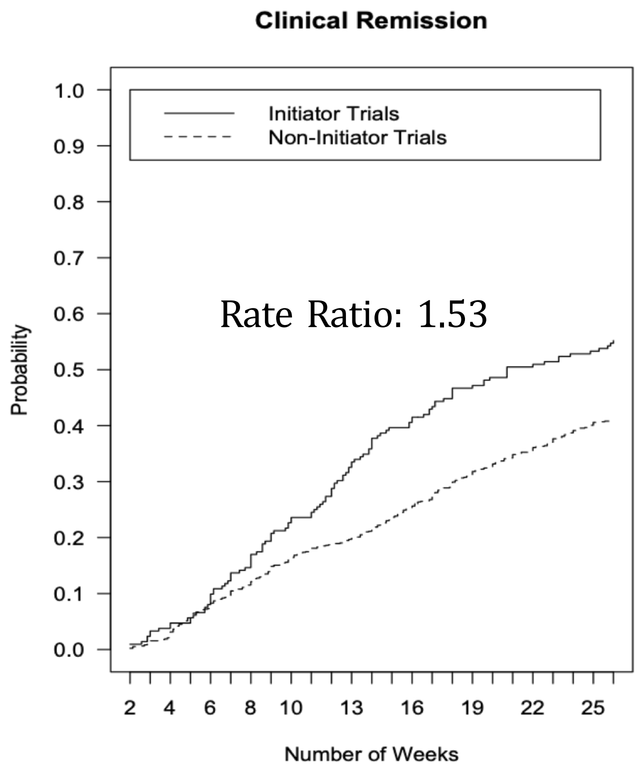
Episodes of Care for new-onset moderate/severe IBD

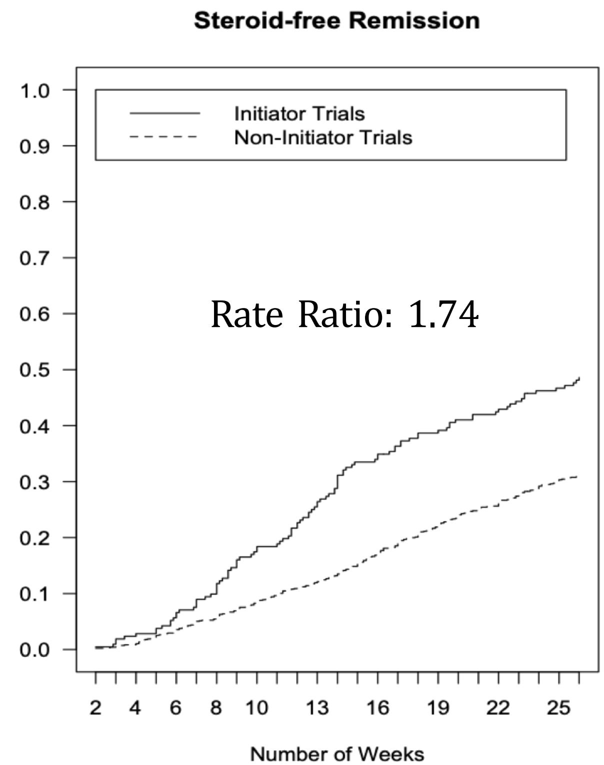


Create ‘Pseudo-Trials’ from extant data



Infliximab/Biologics initiators versus those receiving standard care (thiopurines/steroids) for Crohn's disease patients with new-onset moderate/severe episodes of illness.





Source: Forrest, Crandall, Bailey, et al. Pediatrics, 2014

Epidemiologic Study On Obesity

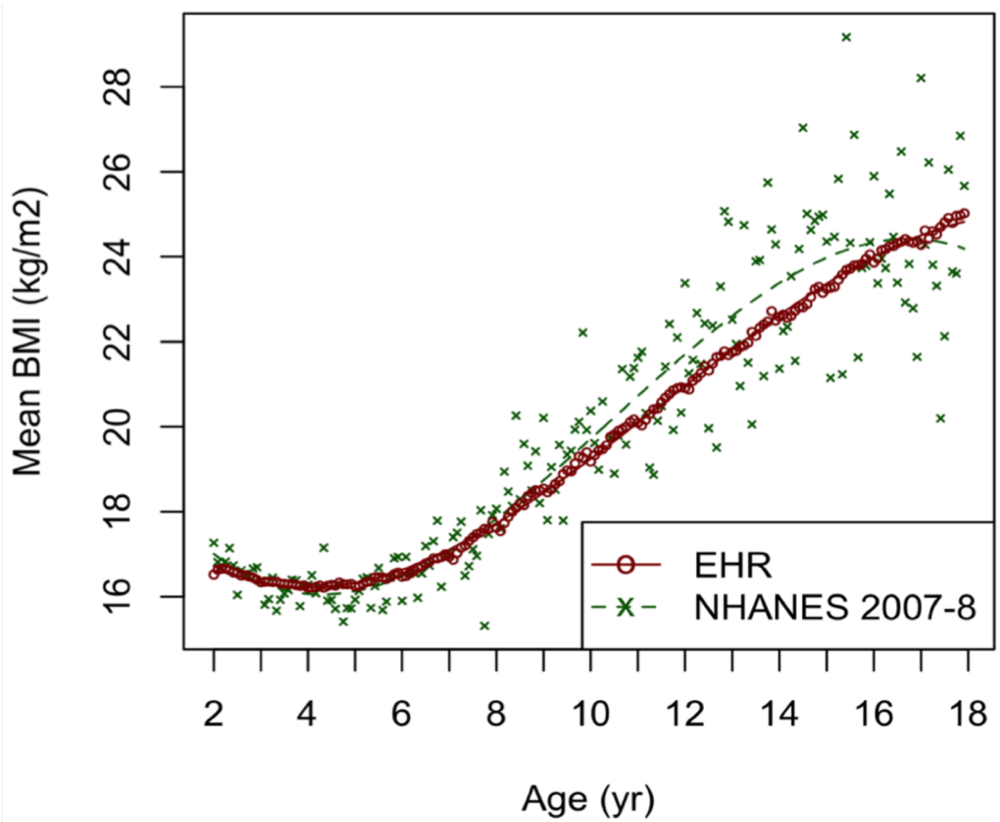
Prevalence of Obesity and Overweight in EHR-Derived Data and NHANES Data.

	Fraction of sample[b]	% Obese	% Overweight, never obese
NHANES 2007–8[a]			
2–17 years	1.000	18	16
2–4 years	0.194	11	12
5–10 years	0.349	19	15
11–17 years	0.457	20	17
Multi-site EHR Data			
2–17 years	1.000	18	17
2–4 years	0.280[c]	14	16
5–10 years	0.418[c]	18	17
11–17 years	0.374[c]	20	17

PEDSnet

Source: Bailey, Milov, Kelleher, et al. *PLoS One*, 2013

BMI by age, PEDSnet v NHANES; PEDSnet estimates are more stable because 1000s of children per month of age are analyzed versus 10-20 per month of age for NHANES.



One plotting point for each age in months

PEDSnet = red
NHANES = green

Source: Bailey, Milov, Kelleher, et al. *PLoS One*, 2013

7. Computable phenotyping and population denominators

If you wanted to compute prevalence and incidence rates from PEDSnet EHR data, what would you use for your denominator?

Consider Pediatric Diabetes Mellitus as the target population

PEDSnet Patients' Locations, n=6.5 Million

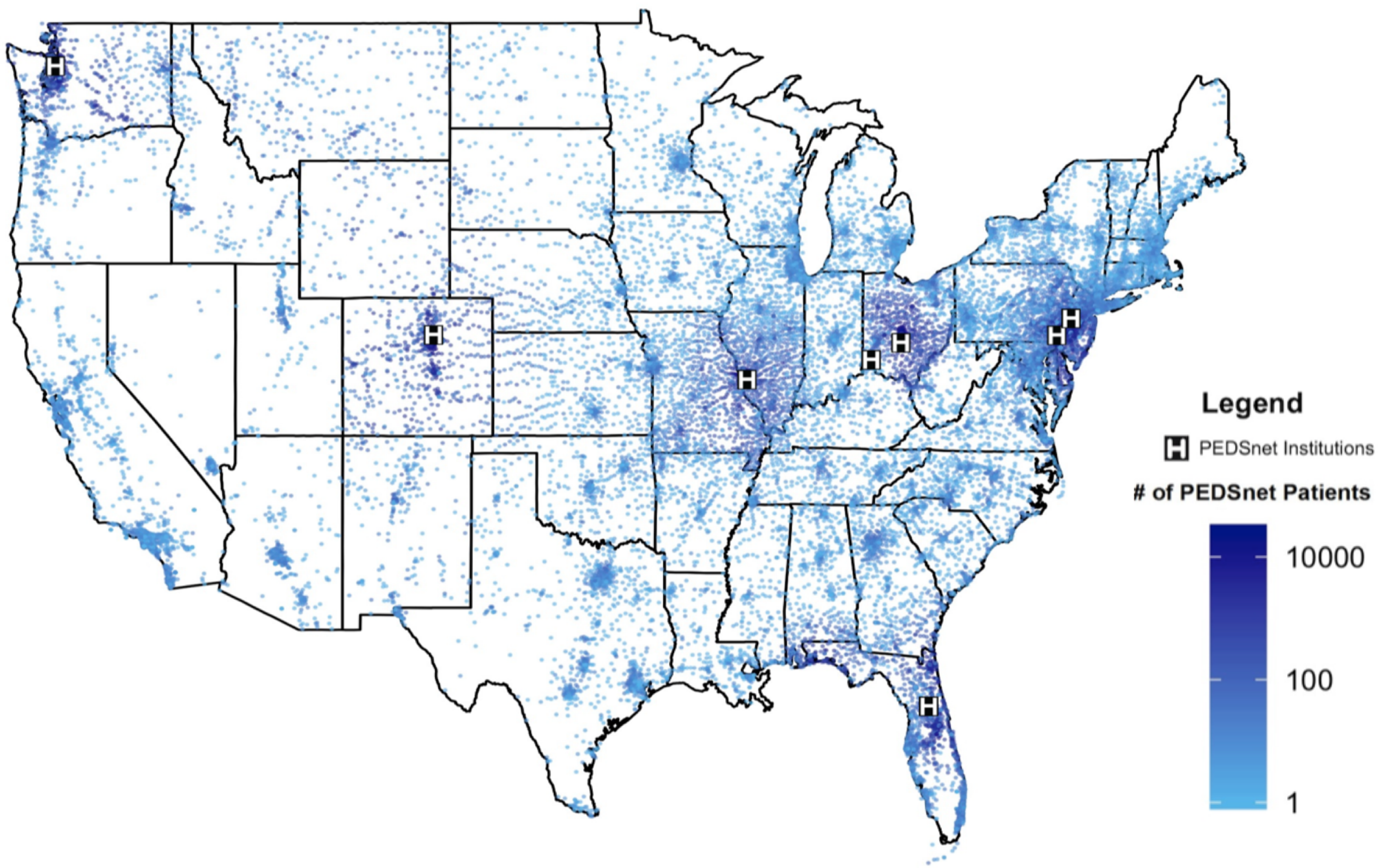


Table 1. Primary Market Counties Used for the Population Denominators by PEDSnet Institution.		
Institution	State: Counties	2018 Denominator Population by Institution and State
Children's Hospital Colorado	CO: Arapahoe, Adams, Broomfield, Denver, Douglas, Jefferson, Summit, El Paso	CO: 176,486
Children's Hospital of Philadelphia	PA: Bucks, Chester, Delaware, Montgomery, Philadelphia NJ: Atlantic, Burlington, Cape May	PA: 296,050 NJ: 44,644
Cincinnati Children's Hospital Medical Center (CCHMC)	IN: Dearborn KY: Boone, Campbell, Kenton OH: Butler, Clermont, Hamilton, Warren	Data not available; Total population size estimated based on prior work to be: 250,000
Nationwide Children's Hospital	OH: Delaware, Franklin, Madison, Union, Fairfield, Pickaway, Perry, Licking, Fayette, Knox, Logan, Hocking, Muskingum, Marion, Ross, Champaign, Morrow, Athens, Pike	OH: 268,172
Nemours Children's Health System	DE: Kent, New Castle, Sussex FL: Indian River, Osceola, Nassau PA: Chester, Delaware	DE: 89,059 FL: 29,362 NJ: 8,239 PA: 62,978
Seattle Children's Hospital	WA: King, Snohomish	WA: 110,211
St. Louis Children's Hospital	MO: Crawford, Iron, Madison, St. Francois, St. Louis City, St. Charles, St. Louis, Washington IL: Calhoun, Jersey, Madison	MO: 110,161 IL: 14,178
Totals	Unique States: 11 Unique Counties: 62	Populations: Without CCHMC: 1,209,540 With CCHMC: 1,459,540

Table 2. Sociodemographic Characteristics of the PEDSnet versus Representative Populations

	0-17 Year-Old Populations		
Characteristic	PEDSnet: Primary Market Counties	Full Population: Primary Market	US Population
N	1,209,540	4,028,663	73,399,342
Age, years, %			
0-4	30	27	27
5-9	27	28	28
10-14	26	28	28
15-17	17	17	17
% Female	48	49	49
Race/Ethnicity, %			
Hispanic	15	17	26
White, non-Hispanic	56	58	52
Black, non-Hispanic	23	18	15
Asian/Pacific Islander	6	7	6
Native American	<1	<1	1
Payer, %			
Private	54	58	55
Public	44	38	40
None	2	4	5
Family Income, $, mean	$77,493	$73,688	$70,544

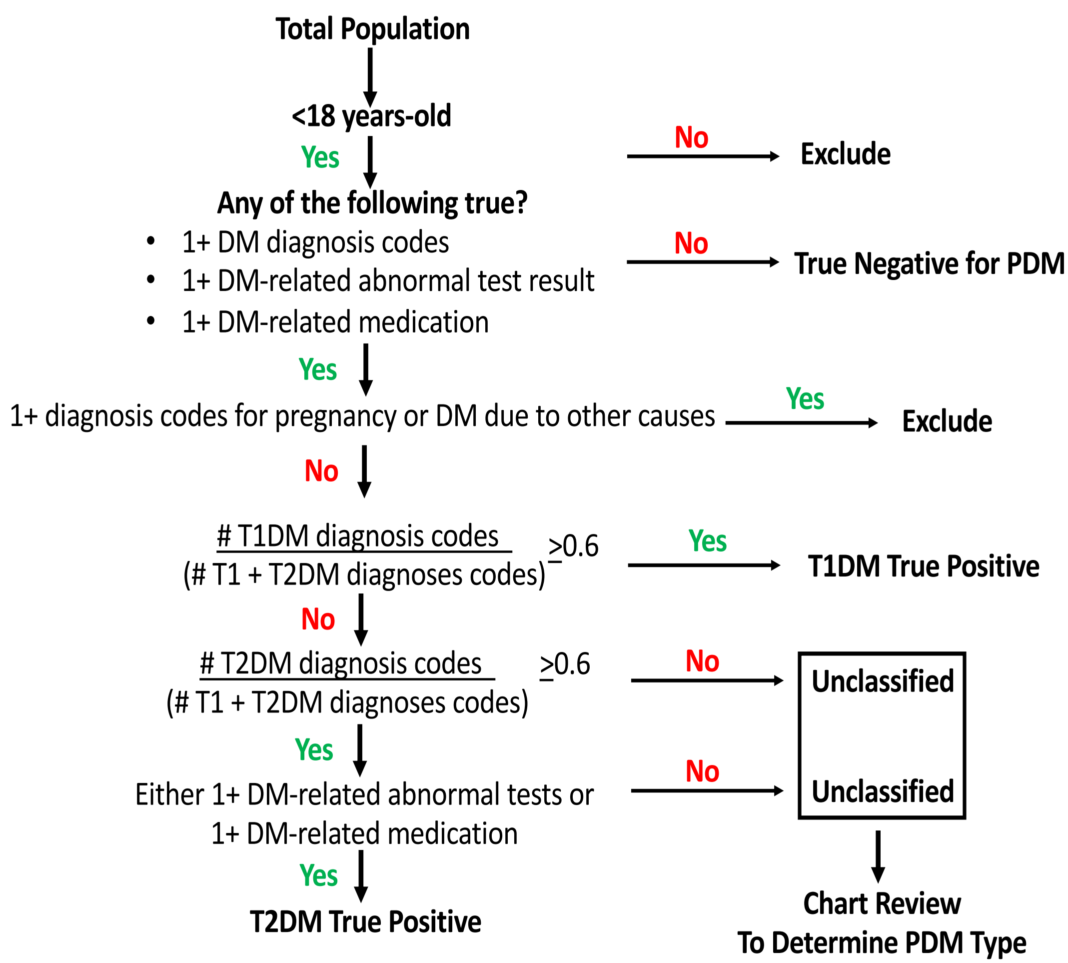


Figure 5. PDM Phenotyping Approach

Table 3. PDM CY2018 Incidence Rates, Overall and by State.					
	Incident Cases			Incidence per 100,000	
State	T1DM	T2DM	Population Denominator	T1DM	T2DM
CO	55	42	176,486	31	24
DE	21	23	89,059	24	26
FL	10	6	29,362	34	20
IL	6	3	14,178	42	21
MO	22	20	110,161	20	18
NJ	19	5	52,883	36	9
OH	88	46	268,172	33	17
PA	79	51	359,028	22	14
WA	37	29	110,211	34	26
Totals	337	225	1,209,540	**28**	**19**

Next Session: Informatics to support network-based learning health systems (L. Charles Bailey, MD, PhD: PEDSnet Director of Data Network)

Competencies he will address:

4.1: Demonstrate ability to use data derived from electronic health records and other clinical information sources for research and quality improvement.

4.2: Demonstrate knowledge about additional data sources that can be linked to health system clinical data in order to augment exposure and outcome ascertainment.

4.3: Demonstrate ability to assess data quality and apply data quality assurance processes, including error prevention, data cleaning, data monitoring, documentation, and relevant data standards.

4.5: Demonstrate knowledge of clinical information systems, including electronic health records, clinical documentation, computerized physician order entry (CPOE), clinical decision support systems, electronic prescribing, medical imaging, and clinical/population dashboards.

5.3: Demonstrate knowledge of specific Health Insurance Portability and Accountability Act (HIPAA) requirements associated with varied data sources used in health systems research activities and seek appropriate approvals.

Key Topics will include:

- Interoperability of data systems (security, vocabularies and coding systems, common data models, data format, data exchange and linkage, introduction to Achilles and Atlas)
- Data quality (structural, how to address missing data, semantic)
- Privacy, HIPAA, and new changes to the Common Rule affecting LHS science