

Clinical Integrated Data Repositories and Observations regarding Data Sharing and National Collaboration

Russ Waitman, PhD, University of Kansas Medical Center

Director of Medical Informatics, Associate Vice Chancellor Enterprise Analytics

Professor, Internal Medicine

Greater Plains Collaborative Principal Investigator

May 24, 2017



This work is supported in part by NIH grant UL1TR000001 and the PCORI PCORnet Greater Plains Collaborative CDRN

Acknowledgements and Conflicts

- This presentation brings together slides from many contributors including but not limited PCORnet's master slides introduction and the Greater Plains Collaborative team. Contact me for further information:
rwaitman@kumc.edu

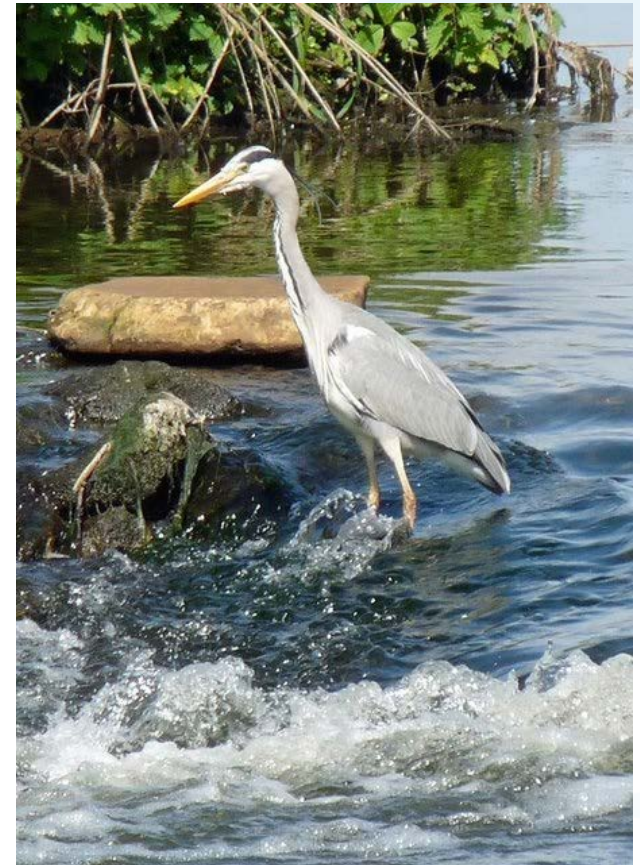
- I am a board member of the i2b2 tranSMART Foundation

Overview of This Talk

- 🌐 **Clinical Integrated Data Repositories common to United States Academic Medical Centers with National Institutes of Health (NIH) funded Clinical and Translational Science Awards (CTSA).**
- 🌐 **Greater Plains Collaborative leveraging NIH CTSA investments to support National Data Sharing: PCORnet**

Central CTSA Informatics Aim: Create a data “fishing” platform: **HERON**, <https://heron.kumc.edu>

- ✚ **Get a License:** Develop business agreements, policies, data use agreements and oversight.
- ✚ **Get a Fishing Rod and Bass Boat:** Implement open source NIH funded (i.e. i2b2 <https://www.i2b2.org/>) initiatives for accessing data.
- ✚ **Know what your catching:** Transform data into information using the NIH National Library of Medicine Unified Medical Language System (UMLS) Metathesaurus as our vocabulary source.
 - <https://www.nlm.nih.gov/research/umls/>
 - Secondary goal; not critical at one site
- ✚ **Stock Different Tasty Fish:** link clinical data sources to enhance their research utility.



FRONTIERS

THE HEARTLAND INSTITUTE FOR
CLINICAL AND TRANSLATIONAL RESEARCH

HERON: Getting a Fishing License

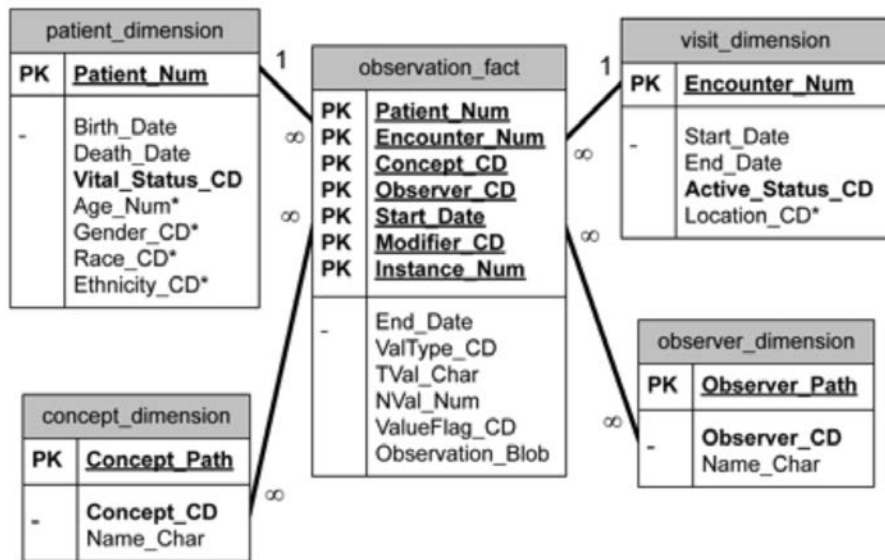
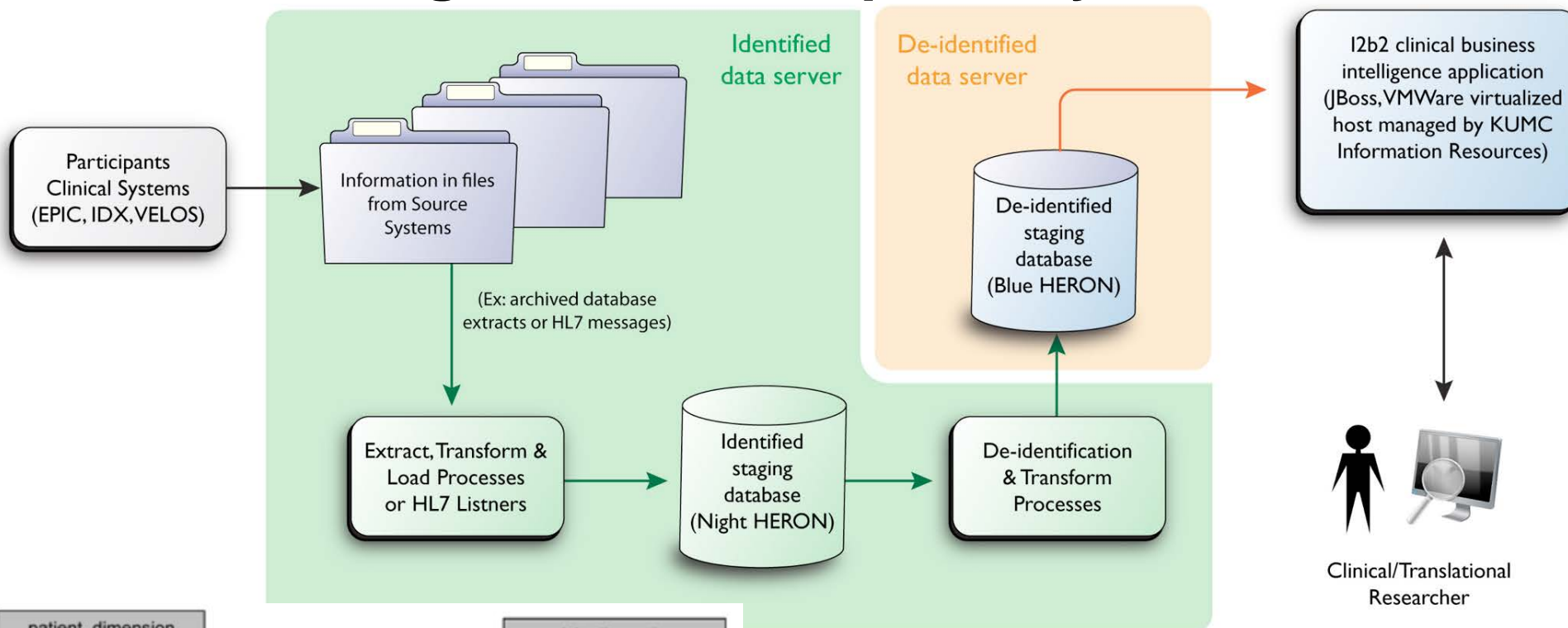


Single sign-on using your email username

Real-time check for current human subjects training

- Faculty fill out System Access Agreements to sponsor students/staff/external collaborators
- Data Use Agreements to request data export
- **No Limit!!! IRB Protocol Not Required to view or pull de-identified data.**
 - **Removing all 18 HIPAA identifiers and date shift -> non human subjects determination**
- Check <http://frontiersresearch.org/frontiers/HERON-Introduction> for more information, status, and training videos

HERON Integrated Data Repository Architecture



The i2b2 “Fishing Rod”: build Diabetes cohort

i2b2 Query & Analysis Tool Project: Heron Pawnee (data thru Nov 2011) User: Russ Waitman Find Patients | Analysis Tools | Message Log | Help | Gripe | Logout

Navigate Terms **Find Terms**

Types of “fish” in folders – hierarchical concept paths

HERON Pawnee (data thru Nov 2011)

- [Demographics [14,109,046 facts; 1,882,584 patients] - 0]
- [Diagnoses [17,501,549 facts; 578,213 patients] - 0]
 - [Circulatory system [1,510,529 facts; 120,477 patients] - 0]
 - [Conditions in the perinatal period [202,798 facts; 9,539 patients] - 0]
 - [Congenital anomalies [133,342 facts; 27,927 patients] - 0]
 - [Digestive system [734,498 facts; 100,290 patients] - 0]
 - [Endocrine disorders [578,559 facts; 59,752 patients] - 0]
 - [Other endocrine gland diseases [433,781 facts; 42,725 patients] - 0]
 - [Diabetes mellitus [376,018 facts; 31,783 patients] - 0]
 - [Diseases of thymus gland [58 facts; 39 patients] - 0]

Workplace

- SHARED
- aallen
- ablevins
- achoudhary
- agreiner
- akirk-phillips
- akovac
- alele
- apark2
- ayacoub
- badagarla
- hewstein

Previous Queries

- GI-Ob-00-Ar-10-00-01@10:08:42 [1-25-2012] [mschultz]
- GI-Ob-00-Ar-10-00-01@09:49:12 [1-25-2012] [mschultz]
- GI-Ob-00-Ar-10-00-01@09:48:03 [1-25-2012] [mschultz]
- OSA-no; surg Proc; No colonoscopy; recov score@09:33:33 [1-25-2012] [mschultz]
- OSA; Surg Proc-no colonoscopy; recovery score@09:30:54 [1-25-2012] [mschultz]
- OSA; surgical procedures; post recovery scores@09:26:04 [1-25-2012] [mschultz]

Query Name:

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
[Diabetes mellitus [376,018								

Drag concepts from upper left into panels on the right

Largely ICD9/10 based nationally

- Driven by billing systems
- Hierarchies from UMLS

one or more of these AND drop a term on here

Run Query **Clear** **Print Query** 1 Group **New Group**

Query Status

i2b2 : **AND** in Frontiers Research Registry

i2b2 Query & Analysis Tool Project: Heron Pawnee (data thru Nov 2011) User: Russ Waitman Find Patients | Analysis Tools | Message Log | Help | Gripe | Logout

Navigate Terms Find Terms

- Cancer Cases (KUH Tumor Registry - DRAFT status)
- HERON Pawnee (data thru Nov 2011)
 - [Demographics [14,109,046 facts; 1,882,584 patients] - 0]
 - [Age [1,853,709 facts; 1,853,709 patients] - 0]
 - [Frontiers Research Participant Registry [7,654 facts; 7,654 patients] - 0]
 - [Gender [1,882,584 facts; 1,882,584 patients] - 0]
 - [Income - 0]
 - [Language [1,882,584 facts; 1,882,584 patients] - 0]
 - [Marital Status [1,882,584 facts; 1,882,584 patients] - 0]
 - [Place: School District [210,373 facts; 210,373 patients] - 0]
 - [Place: State [1,240,046 facts; 1,240,046 patients] - 0]

Query Tool

Query Name:

Temporal Constraint:

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
[Diabetes mellitus [376,018			[Frontiers Research Parti					

Workplace

- SHARED
- aallen
- ablevins
- achoudhary
- agreiner
- akirk-phillips
- akovac
- alele
- apark2
- ayacoub
- badagarla
- ...

Previous Queries

- Maligna-Cardiov@11:34:48 [1-25-2012] [cbefort]
- Malig-Evalu-Cardi@11:34:22 [1-25-2012] [cbefort]
- Malig-Evalu-Cardi@11:32:57 [1-25-2012] [cbefort]
- Maligna-Evaluat@11:27:59 [1-25-2012] [cbefort]
- GI-Ob-00-Ar-10-00-01@10:41:09 [1-25-2012] [mschultz]
- GI-Ob-00-Ar-10-00-01@10:08:42 [1-25-2012] [mschultz]

Query Status

one or more of these AND drop a term on here

Run Query Clear Print Query 1 Group New Group

Draggng over the second fact: local term

i2b2: **AND** a high Hemoglobin A1C

The screenshot shows the i2b2 Query & Analysis Tool interface. The top bar includes the project name 'Heron Pawnee (data thru Nov 2011)', the user 'Russ Waitman', and navigation links. The 'Find Terms' panel on the left lists various medical terms, including 'HEMOGLOBIN A1C (#2034)'. The 'Query Tool' panel on the right shows three query groups. A red arrow points from the 'HEMOGLOBIN A1C' term in the 'Find Terms' list to the 'HEMOGLOBIN A1C (#2034)' term in Group 3. A dialog box titled 'Choose value of HEMOGLOBIN A1C (Test:KUH|COMPONENT_ID:2034)' is open, showing options to constrain the search by lab values. The 'By value' option is selected, and the operator 'GREATER THAN (>)' is chosen with a value of '7.5'. A red box highlights the 'AND' operator in the query groups, and a green box highlights the text 'one or more of these'.

When you add a numeric concept, i2b2 asks if you want to set a constraint

Lab results “moving” to LOINC meaningful use standards

i2b2 Result: 497 patients in Cohort

i2b2 Query & Analysis Tool Project: Heron Pawnee (data thru Nov 2011) User: Russ Waitman Find Patients | Analysis Tools | Message Log | Help | Gripe | Logout

Navigate Terms Find Terms

- [GLUCOSE, 6 HR. (#2045) [<10 facts] - 0]
- [GLUCOSE, FASTING (#2035) [5,381 facts; 4,666 patients] - 0]
- [GLUCOSE, PP 2HR-75 GRAM (#515) [<10 facts] - 0]
- [GLYCOHEMOGLOBIN, TOTAL (#695) [<10 facts] - 0]
- [GRAMS OF GLUCOSE (#2037) [2,414 facts; 2,127 patients] - 0]
- [HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients] - 0]
- [IMMUNOGLOBULIN D (#1349) [40 facts; 28 patients] - 0]
- [INSULIN SPECIMEN #1 (#3416) [25 facts; 25 patients] - 0]
- [INSULIN SPECIMEN #2 (#3418) [25 facts; 25 patients] - 0]
- [INSULIN SPECIMEN #3 (#3420) [18 facts; 18 patients] - 0]
- [INSULIN TIME #1 (#3415) [25 facts; 25 patients] - 0]
- [INSULIN TIME #2 (#3417) [25 facts; 25 patients] - 0]

Workplace

- SHARED
- aallen
- ablevins
- achoudhary
- agreiner
- akirk-phillips
- akovac
- alele
- apark2
- ayacoub
- badagarla
- hemsch

Previous Queries

- Diabe-Front-HEMOG@11:40:52 [1-25-2012] [rwaitman]
- Maligna-Cardiov@11:34:48 [1-25-2012] [cbefort]
- Malig-Evalu-Cardi@11:34:22 [1-25-2012] [cbefort]
- Malig-Evalu-Cardi@11:32:57 [1-25-2012] [cbefort]
- Maligna-Evaluat@11:27:59 [1-25-2012] [cbefort]
- GLob-00-Ar-10-00-01@10-11-09 [1-25-2012] [mschultz]

Query Tool

Query Name: Diabe-Front-HEMOG@11:40:52

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
[Diabetes mellitus [376,018			[Frontiers Research Partici			HEMOGLOBIN A1C (#2034		

one or more of these AND one or more of these AND one or more of these

Run Query Clear Print Query 3 Groups New Group

Query Status

Finished Query: "Diabe-Front-HEMOG@11:40:52" [6.8 secs]
Compute Time: 4 secs

Patient Set - 497 Patients

Number of patients
patient_count: 497

Run the Query
Query took 4 seconds
497 patient in cohort

i2b2: Explore Cohort, Visualize

i2b2 Query & Analysis Tool

Project: Heron Pawnee (data thru Nov 2011)

User: Russ Waitman

[Find Patients](#)

[Analysis Tools](#)

[Message Log](#)

[Help](#)

[Gripe](#)

[Logout](#)

Timeline

[Specify Data](#)

[View Results](#)

[Plugin Help](#)

<<< start: 1 size: 10 go >>>

zoom: - + pan: < >

10/10/2000

5/5/2006

11/29/2011

Person_#1774_f_Non-hispanic

006- #301070 BMI (Calculated) [288,218 facts; 80,399 patients]

Diabetes mellitus [376,018 facts; 31,783 patients]

Frontiers Research Participant Registry [7,654 facts; 7,654 patients]

GLUCOSE (#2011) [1,087,668 facts; 118,524 patients]

HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients]

Person_#1809_f_Non-hispanic

006- #301070 BMI (Calculated) [288,218 facts; 80,399 patients]

Diabetes mellitus [376,018 facts; 31,783 patients]

Frontiers Research Participant Registry [7,654 facts; 7,654 patients]

GLUCOSE (#2010) [672,582 facts; 100,192 patients]

GLUCOSE (#2011) [1,087,668 facts; 118,524 patients]

HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients]

Person_#21253_f_Non-hispanic

006- #301070 BMI (Calculated) [288,218 facts; 80,399 patients]

Diabetes mellitus [376,018 facts; 31,783 patients]

Frontiers Research Participant Registry [7,654 facts; 7,654 patients]

GLUCOSE (#2010) [672,582 facts; 100,192 patients]

GLUCOSE (#2011) [1,087,668 facts; 118,524 patients]

HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients]

Person_#13473_f_Non-hispanic

006- #301070 BMI (Calculated) [288,218 facts; 80,399 patients]

Diabetes mellitus [376,018 facts; 31,783 patients]

Frontiers Research Participant Registry [7,654 facts; 7,654 patients]

GLUCOSE (#2010) [672,582 facts; 100,192 patients]

GLUCOSE (#2011) [1,087,668 facts; 118,524 patients]

HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients]

Person_#3138_m_Non-hispanic

006- #301070 BMI (Calculated) [288,218 facts; 80,399 patients]

Diabetes mellitus [376,018 facts; 31,783 patients]

Frontiers Research Participant Registry [7,654 facts; 7,654 patients]

GLUCOSE (#2010) [672,582 facts; 100,192 patients]

GLUCOSE (#2011) [1,087,668 facts; 118,524 patients]

HEMOGLOBIN A1C (#2034) [89,793 facts; 35,269 patients]

Person_#25806_f_Non-hispanic

Diabetes mellitus [376,018 facts; 31,783 patients]

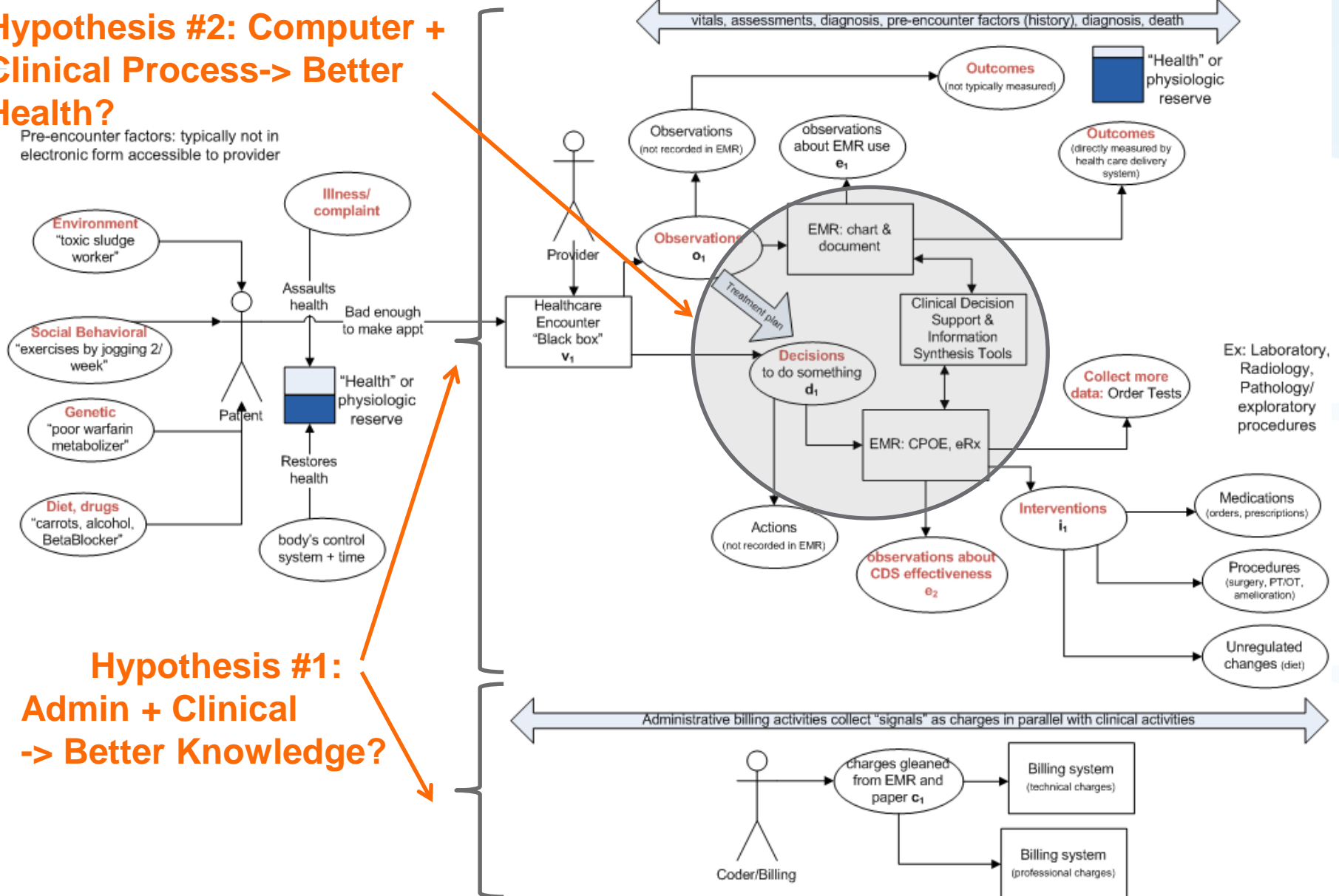
Frontiers Research Participant Registry [7,654 facts; 7,654 patients]

GLUCOSE (#2010) [672,582 facts; 100,192 patients]

Medical Informatics Hypotheses and EMR's role in Research post 2008 (HITECH and ACA laws)

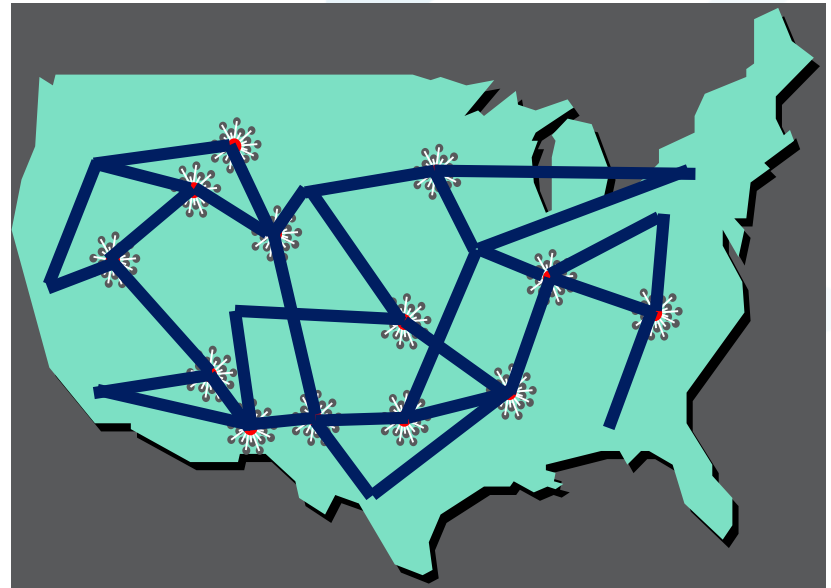
Hypothesis #2: Computer + Clinical Process -> Better Health?

Pre-encounter factors: typically not in electronic form accessible to provider



Both researchers and funders now recognize the value in integrating clinical research networks

- Linking existing networks means clinical research can be conducted more effectively
- Ensures that patients, providers, and scientists form true “communities of research”
- Creates “interoperability” – networks can share sites and data

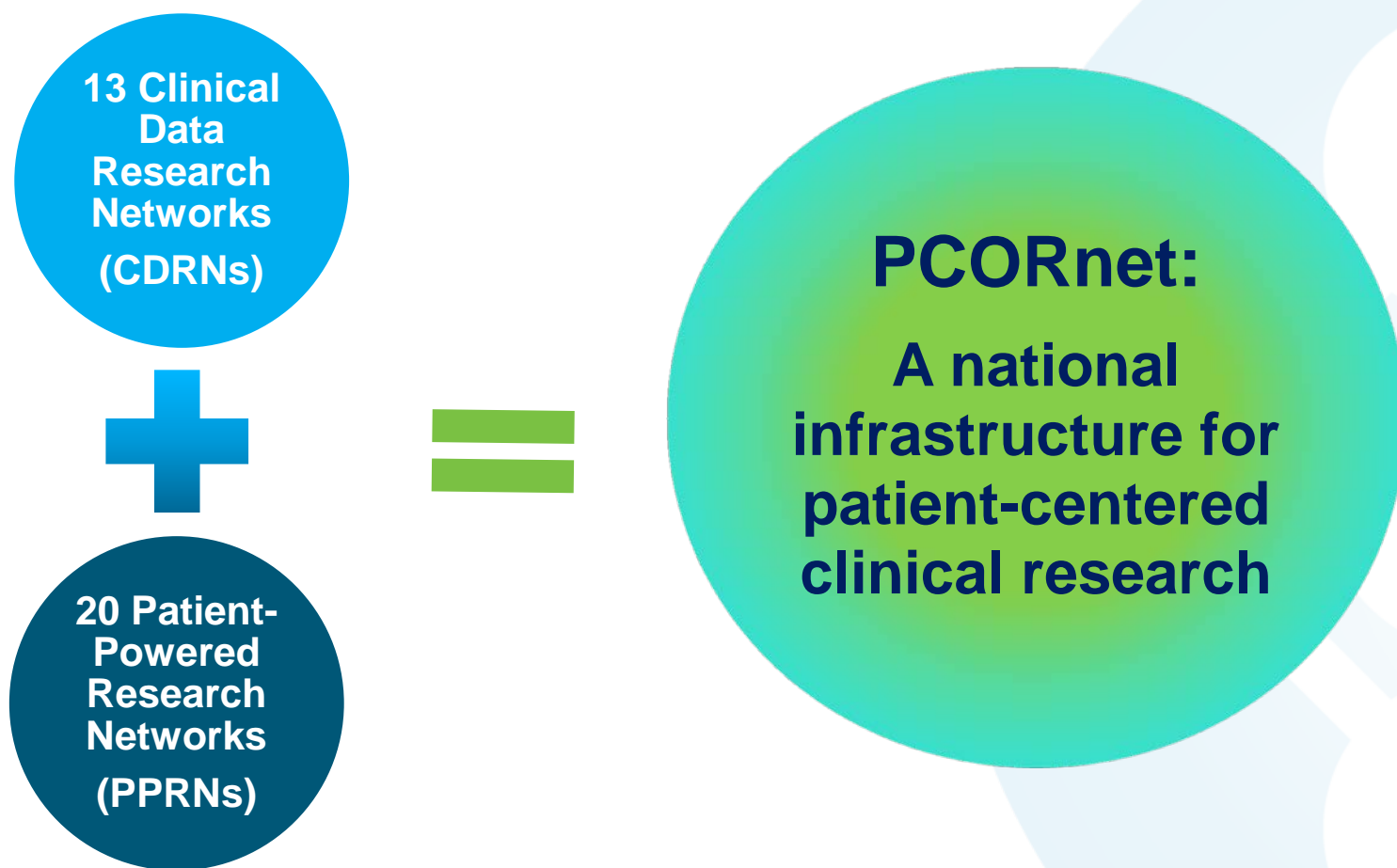


Patient Centered Outcomes Research Institute PCORnet's goal



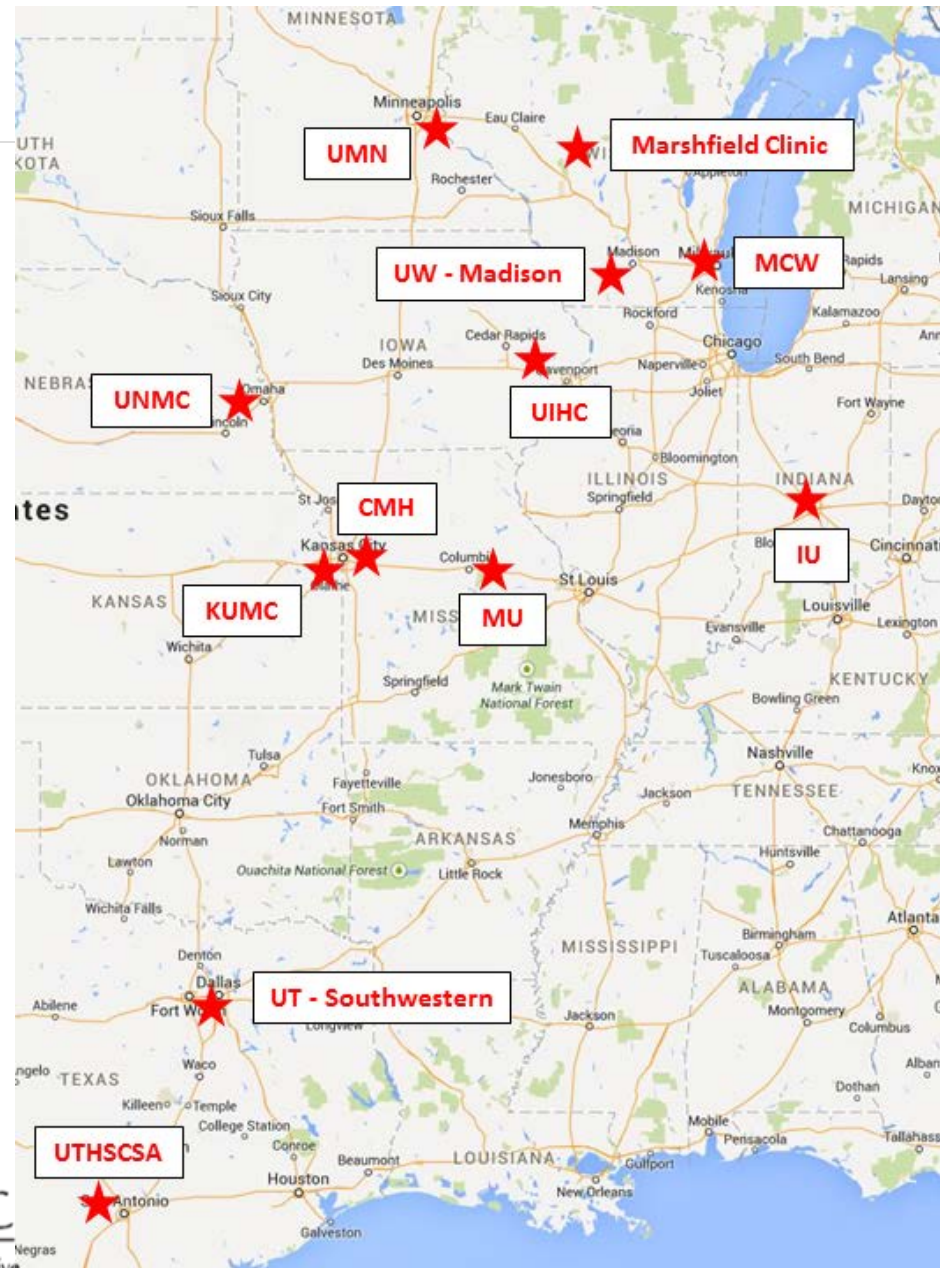
PCORnet seeks to improve the nation's capacity to conduct clinical research by creating a large, highly representative, national patient-centered network that supports more efficient clinical trials and observational studies.

PCORnet embodies a “community of research” by uniting systems, patients & clinicians



The Greater Plains Collaborative CDRN Partners

- KS, the **University of Kansas Medical Center (KUMC)**
- MO, **Children's Mercy Hospital, University of Missouri**
- IA, **University of Iowa Healthcare**
- IN, **Indiana University**
- WI, the **University of Wisconsin-Madison, the Medical College of Wisconsin, and Marshfield Clinic**
- MN, the **University of Minnesota Medical Center**
- NE, the **University of Nebraska Medical Center**
- TX, the **University of Texas Health Sciences Center at San Antonio and the University of Texas Southwestern Medical Center.**



Gross Observation:

**Trialists/Clinicians
Want “Their” Rich
Data:**

**Cancer Registry
Morphology,**

**Patient Portal Usage
Death Index,**



Navigation Terms Find Terms

- Alerts (DRAFT) [20,152,544 facts; 448,218 patients]
- Allergy (DRAFT) [622,427 facts; 276,937 patients]
- Cancer Case Identifiers (Abridged) [1,599,991 facts; 81,153 patients]
 - 0390 Date of Diagnosis [87,338 facts; 81,114 patients]
 - 0440 Grade [87,329 facts; 81,111 patients]
 - 0490 Diagnostic Confirmation [87,338 facts; 81,114 patients]
 - 0521 Morph--Type&Behav ICD-O-3 [86,795 facts; 80,693 patients]
 - 0523 Behavior Code ICD-O-3 [87,334 facts; 81,113 patients]
 - 0610 Class of Case (Selected) [50,679 facts; 47,997 patients]
 - 0630 Primary Payer at DX [87,338 facts; 81,114 patients]
 - 1750 Date of Last Contact [87,387 facts; 81,152 patients]
 - 1760 Vital Status [87,387 facts; 81,152 patients]
 - 1860 Recurrence Date--1st [7,333 facts; 7,239 patients]
 - 1880 Recurrence Type--1st [87,387 facts; 81,152 patients]
 - 1910 Cause of Death [87,387 facts; 81,152 patients]
 - AJCC/TNM Staging [232,219 facts; 81,101 patients]
 - Patient Demographics (Selected) [349,402 facts; 81,153 patients]
 - SEER Site Summary [87,338 facts; 81,114 patients]
- Cancer Cases [17,860,879 facts; 81,153 patients]
- Cardiology Lab Results [26,866,490 facts; 369,448 patients]
- Demographics [20,972,399 facts; 2,146,754 patients]
 - Age [2,123,433 facts; 2,123,429 patients]
 - Email on file [5,686,159 facts; 212,307 patients]
 - Ethnicity [2,146,758 facts; 2,146,754 patients]
 - Frontiers Research Participant Registry [38,694 facts; 38,694 patients]
 - Gender [2,146,758 facts; 2,146,754 patients]
 - Language [2,146,758 facts; 2,146,754 patients]
 - Marital Status [2,146,758 facts; 2,146,754 patients]
 - My Chart [417,398 facts; 285,231 patients]
 - Place: distance from KUMC [1,422,865 facts; 1,422,861 patients]
 - Place: School District [237,839 facts; 237,838 patients]
 - Place: State [1,504,004 facts; 1,504,000 patients]
 - Race [2,146,910 facts; 2,146,754 patients]
 - Religion [2,146,758 facts; 2,146,754 patients]
 - Vital Status [2,347,466 facts; 2,146,754 patients]
 - Deceased per SSA [200,708 facts; 200,708 patients]
 - Deceased [40,213 facts; 40,213 patients]
 - Deferred
 - Living [590,638 facts; 590,637 patients]
 - Not recorded [1,515,907 facts; 1,515,906 patients]

EMR Decision Support Responses

i2b2 Query & Analysis Tool Project: HERON Wiese Slough (data through January 2017)

Navigate Terms

Find

Help

MAR Admin On Discontinued Med Warning [44,716 facts; 14,032 patients] - 14032

MAR Admin On Expired Med Warning [818 facts; 525 patients] - 525

MAR Barcode Unexpected Error Warning [<10 facts] - 7

MAR Data Integrity Warning - Hard Stop [410 facts; 81 patients] - 81

MAR Dose Warning [629,668 facts; 93,136 patients] - 93136

MAR Manual Edit Warning [68,472 facts; 9,532 patients] - 9532

MAR Med Barcode Not Recognized Warning [211,167 facts; 38,944 patients] - 38,944

MAR Medication Not Scanned Warning [2,562,779 facts; 93,163 patients] - 93,163

MAR Missing Components Warning [22,852 facts; 6,520 patients] - 6,520

MAR Multiple Required Scans Warning [173 facts; 131 patients] - 131

MAR NDC No Order For Patient Warning [55,578 facts; 17,230 patients] - 17,230

MAR NDC Order Is Not Active Warning [58,136 facts; 19,134 patients] - 19,134

MAR No Matching Orders Warning [17 facts; 15 patients] - 15

MAR Off Schedule Warning [1,908,423 facts; 70,462 patients] - 70,462

MAR ORD ID Order For Wrong Patient Warning [5,774 facts; 3,609 patients] - 3,609

MAR ORD ID Order Is Not Active Warning [15,136 facts; 6,081 patients] - 6,081

MAR ORD ID Order Not For Current Contact Warning [1,074 facts; 406 patients] - 406

MAR Patient Not Scanned Warning [4,235,335 facts; 100,664 patients] - 100,664

MAR Scanned Concentration Different Warning [22 facts] - 9

MAR Scanning Medications Out of Order Warning [19,986 facts; 9,998 patients] - 9,998

Med Interaction Checking [13,883,967 facts; 338,705 patients] - 338705

Alert Status

Dose [3,774,551 facts; 231,494 patients] - 231494

Drug-Allergy (Active and Inactive Ingredients) [870,444 facts; 127,408 patients] - 127408

Drug-Drug [6,125,077 facts; 248,386 patients] - 248386

Duplicate Medication Orders [1,622,722 facts; 231,225 patients] - 231225

Drug-Drug [6,125,077 facts; 248,386 patients] - 248386

Duplicate Medication Orders [1,622,722 facts; 231,225 patients] - 231225

Drug-Drug [6,125,077 facts; 248,386 patients] - 248386

Alert Status

0

1

2

5 [20,081 facts; 3,634 patients] - 3634

7

A [1,440,648 facts; 120,824 patients] - 120824

Alert Status

ABATACEPT / TUMOR NECROSIS FACTOR (TNF) BLOCKING AGENTS

[66 facts; 30 patients] - 30

ABCIXIMAB / ANTICOAGULANTS [251 facts; 162 patients] - 162

Alert Status

Canceled

Deferred

Filtered

Held

Overridden

Removed

Unknown

Viewed

Observationalists (Epidemiologists) Want “The” Data

🌐 Create a research-ready dataset of at least 1 million patients that is:

- “**Comprehensive**, using data from EHRs to describe patients’ care experience over time and in different care settings”
- “**Utilize** multiple rich data sources to support research, such as electronic health records, insurance claims data, and data reported directly by patients”

🌐 Execution and Governance:

- Developing relationship with external data partners (CMS, State, private insurers)

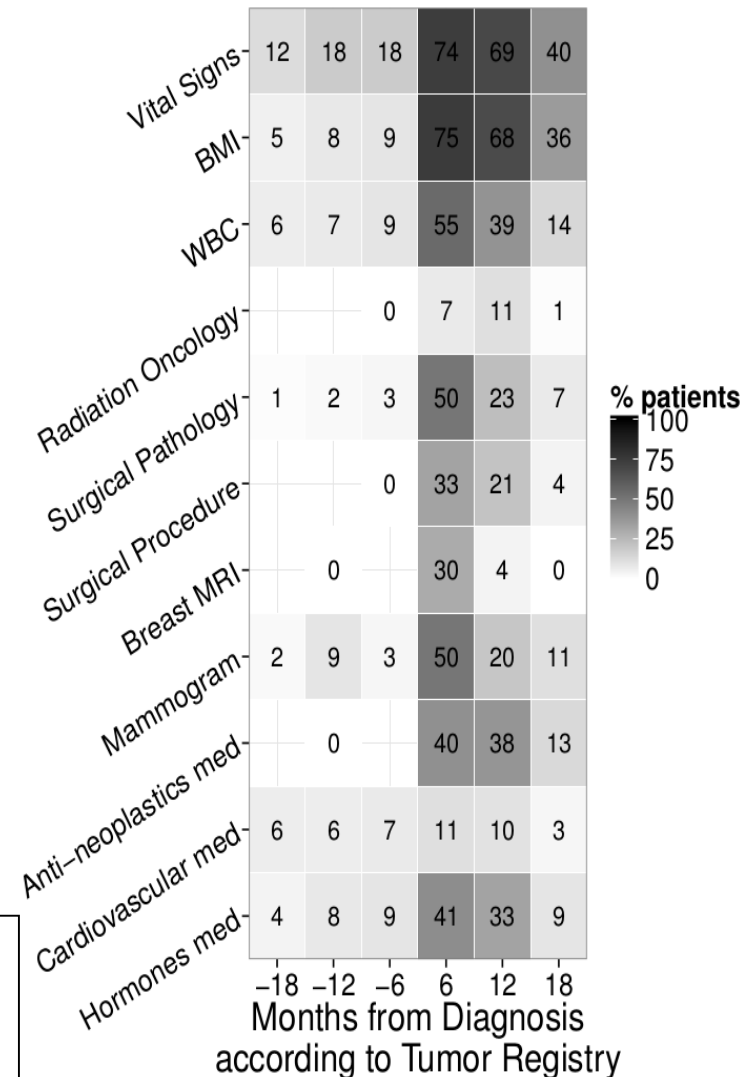


Figure 3.1. Comprehensive and complete data example from KUMC: heat map of percentage of proposed data elements from the HER and billing sources recorded in six month intervals surrounding the data of breast cancer diagnosis specified by the hospital tumor registry.

PCORnet Common Data Model v3.0

DEMOGRAPHIC v1.0

Demographics record the direct attributes of individual patients.

ENROLLMENT v1.0

Enrollment is a concept that defines a period of time during which a person is expected to have complete data capture. This concept is often insurance-based, but other methods of defining enrollment are possible.

ENCOUNTER v1.0

Encounters are interactions between patients and providers within the context of healthcare delivery.

DIAGNOSIS v1.0

Diagnosis codes indicate the results of diagnostic processes and medical coding within healthcare delivery. Data in this table are expected to be from healthcare-mediated processes and reimbursement drivers.

PROCEDURES v1.0

Procedure codes indicate the discrete medical interventions and diagnostic testing, such as surgical procedures and lab orders, delivered within a healthcare context.

VITAL v1.0

Vital signs (such as height, weight, and blood pressure) directly measure an individual's current state of attributes.

LAB_RESULT_CM v2.0

Laboratory result Common Measures (CM) use specific types of quantitative and qualitative measurements from blood and other body specimens. The common measures are defined in the same way across all PCORnet networks, but this table can also include other types of lab results.

CONDITION v2.0

A condition represents a patient's diagnosed and self-reported health conditions and diseases. The patient's medical history and current state may both be represented.

PRO_CM v2.0

Patient-Reported Outcome (PRO) Common Measures (CM) are standardized measures that are defined in the same way across all PCORnet networks. Each measure is recorded at the individual item level: an individual question/statement, paired with its standardized response options.

DISPENSING v2.0

Outpatient pharmacy dispensing, such as prescriptions filled through a neighborhood pharmacy with a claim paid by an insurer. Outpatient dispensing may not be directly captured within healthcare systems.

PRESCRIBING v3.0

Provider orders for medication dispensing and/or administration. These orders may take place in any setting, including the inpatient or outpatient basis.

PCORNET_TRIAL v3.0

Patients who are enrolled in PCORnet clinical trials.

DEATH v3.0

Reported mortality information for patients.

DEATH_CAUSE v3.0

The individual causes associated with a reported death.

HARVEST v3.0

Attributes associated with the specific PCORnet datamart implementation, including data refreshes.

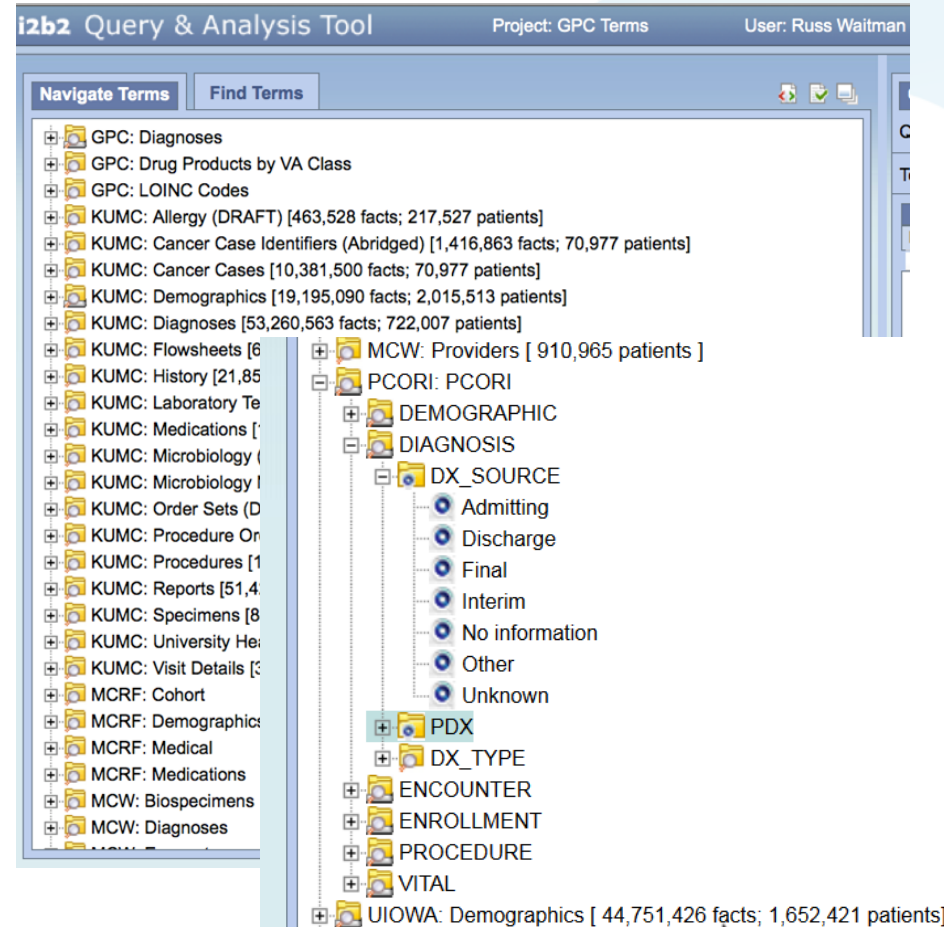
GPC Ontologies -> PCORnet Common Data Model (CDM)

🏥 GPC Progress

- All sites have counts on Babel
 - Study population – 8.4M
 - <https://babel.gpcnetwork.org>
- Annotated data dictionaries have been completed by all sites
- Software has been developed that executes against a GPC i2b2 ontology (containing CDM and additional domains such as labs, meds) to populate the CDM
- <https://github.com/kumc-bmi/i2p-transform>

🏥 Examples on the road to data standardization

- Specification for CDM is billing diagnosis (vs. clinical encounter diagnosis); we can supply both
- GPC can augment CDM with Cancer Registry through i2b2 flexible data model and shared ETL code against NAACCR
- <https://www.naaccr.org/>
- <https://informatics.kumc.edu/work/wiki/TumorRegistry>
- <https://github.com/kumc-bmi/heron>



PCORnet Capabilities and Observational Obesity Studies: Both Address Controversial Topics with Largest N to Date

PCORnet total population ~ 110 million

Bariatric Study: what is best surgical approach?

- 11 Participating CDRNs
- 3 Participating PPRNs
- 48 Institutions
- N=65,000 People (1,000 of whom are adolescents)

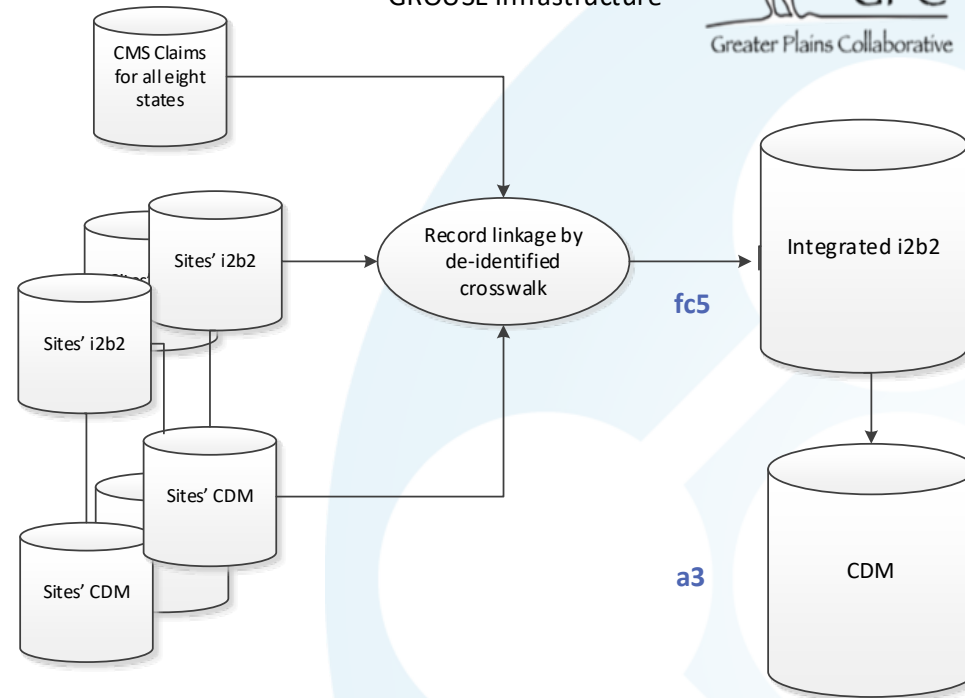
Antibiotics given to young children: do they increase risk for obesity?

- 10 Participating CDRNs
- 4 Participating PPRNs
- 41 Institutions
- N=650,000 children

GPC Reusable Observable Study Environment (GROUSE) generally tertiary care sites for their regions...how biased are we...what do we not know?



Figure 2
GROUSE Infrastructure



- GPC centers in eight states
 - less than 20% US population
- Measure the gain by integrating claims for our 3 cohorts and serve future projects (with CMS approval).
- 3 years of data for Medicare (2011-2013) and one year Medicaid (2011). Adding 2014-2015 now.
- Received and staged claims (~1.2 terabytes) 3.8 billion rows, 19 million beneficiaries
- Have body mass index linked to claims on ~150,000 Medicare/ 50,000 Medicaid beneficiaries at KUMC;
- Support distributed analysis. Must keep data at KUMC per CMS cell size policy



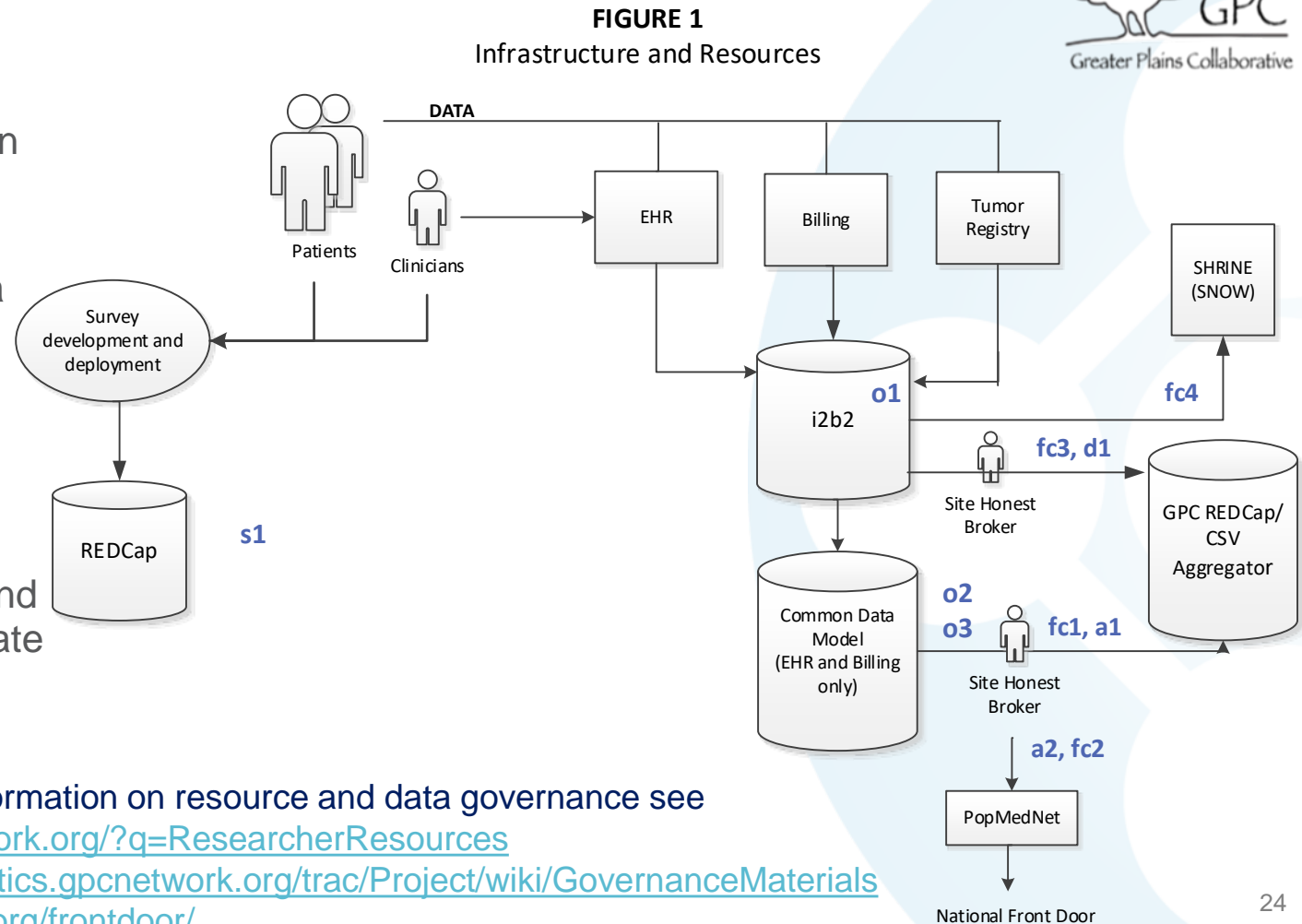
Take home: GPC integrates, queries, analyzes, and ships data among partners and external parties through faculty collaborators and institutional oversight

- i2b2/REDCap for sending data to investigators and feasibility querying with SHRINE

- National intersection with distributed research network

- Common Data Model
- Front Door

- GROUSE: claims integration across states for complete data assessment and comparison with state population



For further information on resource and data governance see

<http://gpcnetwork.org/?q=ResearcherResources>

<https://informatics.gpcnetwork.org/trac/Project/wiki/GovernanceMaterials>

<http://pcornet.org/frontdoor/>