

From Hospital Informatics Laboratories to National Data Networks: Positioning Academic Medical Centers to Advance Clinical Research

Lemuel R. Waitman, Professor, Department of Internal Medicine,
Associate Vice Chancellor for Enterprise Analytics
University of Kansas Medical Center

Since the 1960s and 1970s, pioneering academic medical centers (AMC) have been leaders in developing medical informatics systems to improve patient care and aggregate biomedical data to advance research. Since the HITECH Act in 2008 and the subsequent adoption of Electronic Health Records (EHR), the potential to aggregate biomedical data now extends beyond pioneering academic medical centers to all health systems. Led by the National Institutes of Health and the Patient Centered Outcomes Research Institute's creation of PCORnet, federal, nonprofit, and industry sponsors along with clinicians, patients, and investigators are seeking to capitalize on these new clinical data and link them to traditional billing and claims data sources. These institutions are creating local, regional and national data networks that can support prospective and observational research and realize the vision of a learning health system.

I. Developing National Institutes of Health (NIH) Clinical and Translational Science Award (CTSA) Capacity for Biomedical and Informatics Research

As described previously [Merrill Waitman, Lushington, Warren], academic medical center's pursuit of National Institutes of Health Clinical and Translational Science Awards (CTSA) [Zerhouni] catalyzed the development and integration of informatics capabilities to support clinical and translational research. The 2010 proposal for University of Kansas Medical Center's CTSA program, Frontiers, provides an example of a regional vision for biomedical informatics as illustrated in Figure 1. While the precise steps and integration varied from this plan over the five years since award (2011-2016), the program largely

succeeded in achieving these complementary aims. For our central aim, creating the HERON clinical integrated data repository: business agreements, and oversight processes were successfully established between the university and health system leadership, the open source i2b2 software was implemented [Murphy], data was increasingly mapped to national standards aligned with meaningful use standards available from the National Library of Medicine, and over 2 billion facts for 2 million patients were integrated. Over 800 access and data requests from faculty have been approved and investigators have executed over 50,000 queries. Notably, HERON provided a platform for integrating electronic health records with existing national registries (e.g. NAACCR hospital tumor registry, national cardiovascular

research database CathPCI, trauma, and cystic fibrosis) and organizational benchmarking activities (e.g. Visient University Health System Consortium).

In addition to i2b2 for data integration and warehousing, Frontiers biomedical informatics adopted and promoted REDCap (<https://projectredcap.org/>) [Harris] as a common tool for research data capture across the enterprise and our partner institutions: streamlining access to all with a KUMC campus login.

Adoption has been dramatic: with over 2,500 data collection projects in production for over 4,000 users. REDCap use has also extended beyond traditional clinical trial electronic case report forms to also support registries and administrative needs across the medical center and our health system partners; increasing awareness of the CTSA capabilities for the campus and adoption of REDCap at other campuses in the Kansas City region.

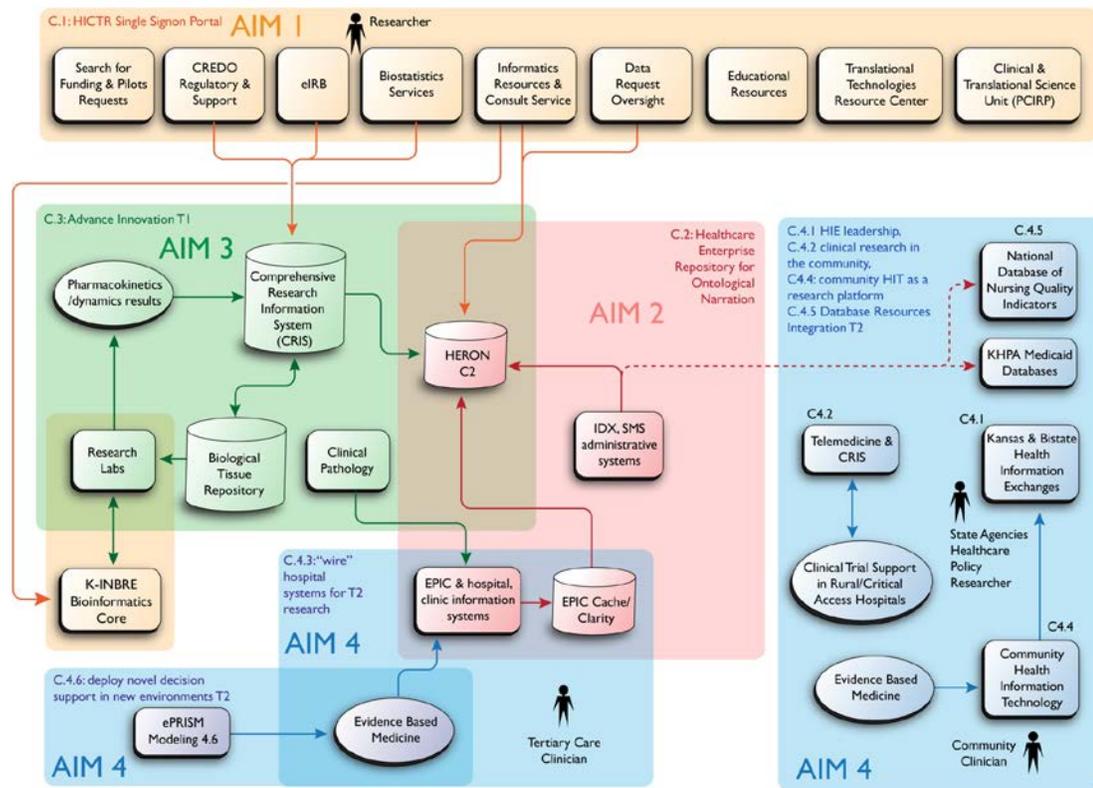


Figure 1. Conceptual Model of Clinical and Translational Science Award Biomedical Informatics and Specific Aims

II. CTSA Informatics Infrastructure’s Potential for National Interoperable Data Research

Frontiers biomedical informatics’ choice of i2b2 and REDCap was fortuitous for supporting broader collaboration nationally. Our campus’ efforts to

make HERON a strong example of extending i2b2 at enterprise scale for a campus and its heightened integration with REDCap was of interest to other academic medical center CTSA programs. Often i2b2 was implemented for specific informatics projects or to only provide

feasibility assessment. These limited scope or reduced functionality implementations of i2b2 would often hinder adoption by the broader research community at a medical center. By 2013, HERON was seen by the broader community as a successful example for data integration which was coincidental with the announcement by the Patient Centered Research Institute (PCORI) that they were creating PCORnet, the National Patient-Centered Clinical Research Network [Fleurence]. This effort was supported by several funding announcements in Spring 2013 that would support the creation of Patient Powered Research Networks, Clinical Data Research Networks, and a Coordinating Center for the national network.

PCORI and its stakeholders from the National Institutes of Health (NIH), the Food and Drug Administration (FDA), patient organizations and insurance plans had a vision to improve our nation's capacity to conduct clinical research. Current state was seen as an environment where a high percentage of decisions made in clinical practice were not supported by the best evidence. Patients health outcomes were not improving and disparities in outcomes were also either stagnant or widening. The machinery of developing prospective or retrospective clinical research studies was slow and expensive or unreliable at creating reproducible research. As a result, clinical research was also unattractive to hospital and health system administrators who often didn't see how this complex additional activity benefited patient care at their institutions. As PCORI's mission is patient centered and more ori-

ented towards pragmatic research (relative to basic science and early translational research support by the NIH) there was a strong emphasis of generating evidence to support daily decision making for doctors, patients, and their families [Tricoci P et al.] PCORI, it's stakeholders of funder and researchers saw that existing networks, recently enabled by the adoption of electronic health records, might provide a platform for conducting research more effectively but also bring together patients, providers and scientists to work as a connected community. PCORnet's goal was "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."

Frontiers biomedical informatics saw high alignment with its work for integrating data in support of our CTSA program and the PCORI funding opportunity to create a Clinical Data Research Network (CDRN) which required the ability to incorporate at least two health systems with over 1 million patients' data and have rich physician and patient engagement. Networks needed to demonstrate governance and the ability to collect patient reported outcomes as well as embed comparative effectiveness trials in the clinical workflow. Some existing data networks were poised to respond while KUMC and related CTSA programs didn't have existing data networks in place. Frontiers worked with other institutions in the Midwest to organize a response and create the Greater Plains Collaborative (GPC) [Waitman, Aaronson] and successfully competed for an initial

Phase 1 CDRN contract (<http://frontiersresearch.org/frontiers/sites/default/files/frontiers/documents/GPC-PCORI-CDRN-Research-Plan-Template-KUMCv44.pdf> ; awarded in 2014 for 18 months) and the subsequent Phase 2 contract (<http://frontiersresearch.org/frontiers/sites/default/files/Phase%20II%20Proposal.pdf> ; awarded in 2015 for 3 years). The Greater Plains Collaborative initially included 10 partner institutions covering an estimated 11.8 million lives, 13 hospitals, 430 clinics, 1800 primary care providers, and 7600 specialists. Each network had to develop their ability to characterize obese patient populations but had flexibility in choosing a common and rare disease focus. GPC worked with community stakeholders to identify breast cancer as its common condition and guided by CTSA leadership choose Amyotrophic Lateral Sclerosis (ALS or Lou Gerhig's disease) as its rare condition. GPC adopted i2b2 and REDCap as common technologies based on their increased adoption across CTSA programs or related programs at other institutions. The Greater Plains Collaborative high level architecture and governance is shown in Figure 2. The initial 10 GPC institutions included: University of Texas Health Sciences Center at San Antonio, University of Texas Southwestern, University of Kansas Medical Center, Children's Mercy Hospital, University of Nebraska Medical Center, University of Iowa, University of Wisconsin, Medical College of Wisconsin, University of Minnesota, and the Marshfield Clinic. In the phase two proposal the Greater Plains

Collaborative expanded to include the University of Missouri and Indiana University (<http://www.gpcnetwork.org/?q=AboutUs>).

GPC PCORI Network Components and Levels of Governance

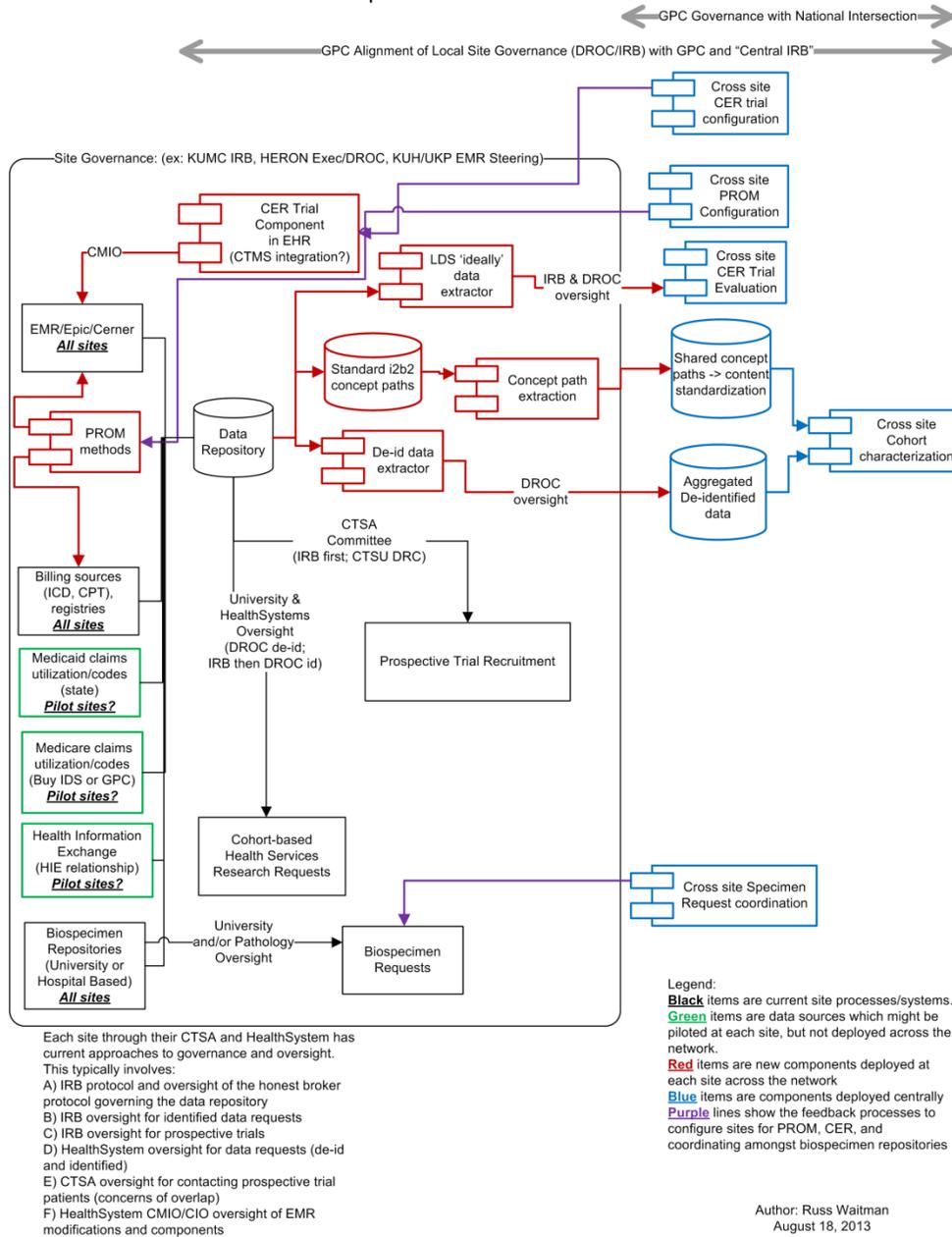


Figure 2. Greater Plains Collaborative Architecture and Governance

III. Managing data expectations amongst differing data constituents: Clinical Researchers (“their data”) and Epidemiologists (“all the data”)

In addition to meeting our contractual milestones outlined in our proposals for becoming a viable Clinical Data Research Network, all network partners were instructed to also shift effort to support adoption and implementation of a new data infrastructure: the PCORnet Common Data Model (CDM) that was to be based on a data model created for the

FDA by a portion of the PCORnet coordinating center at Harvard Pilgrim’s Insurance (the Mini-Sentinel Common Data Model). Mini-Sentinel was created to support adverse drug event surveillance using insurance claims administrative data: classically diagnoses, procedures, and hospitalizations across insured populations who were “covered” by an insurance plan during well defined enrollment periods. The PCORnet Common Data Model version 3.0 is shown in Figure 3.

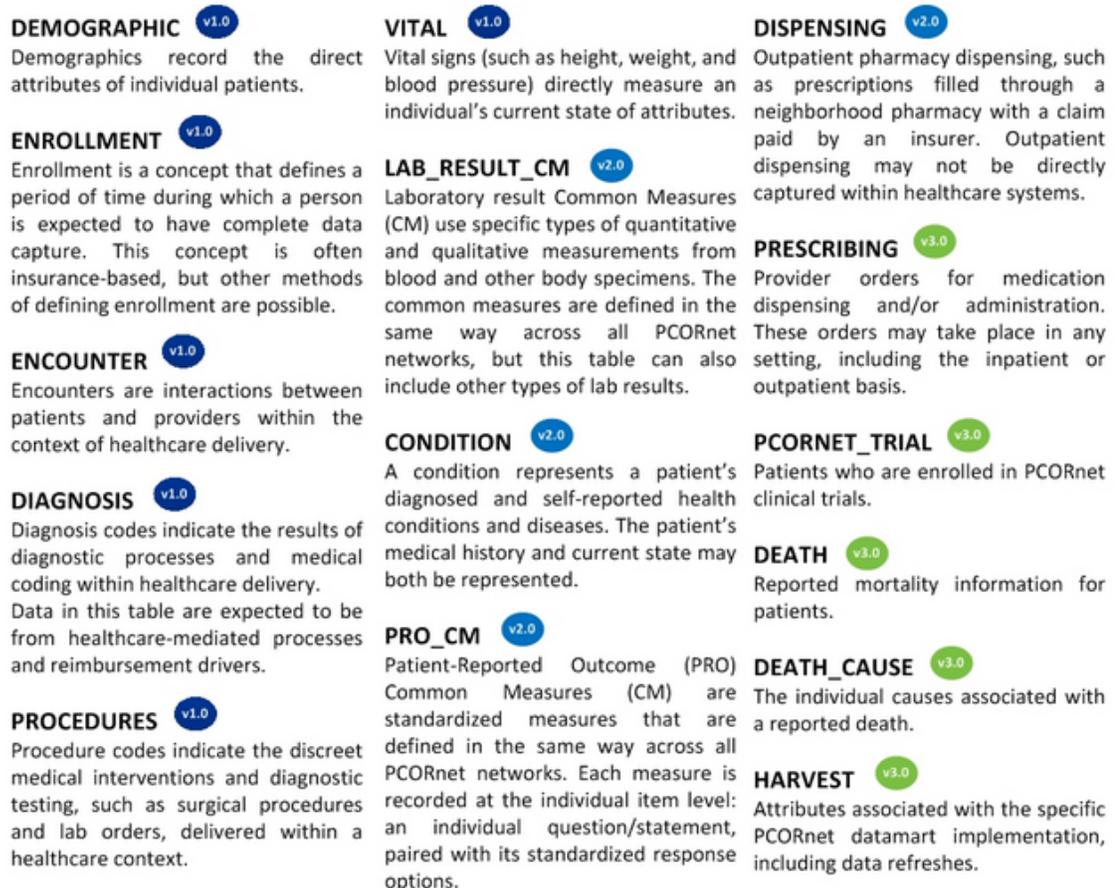


Figure 3. PCORnet Common Data Model version 3.0

As we worked to develop the network and focus on our common condition, breast cancer, the differences in perspective between the epidemiology focused coordinating center data modeling team and those embedded in health systems with rich clinical research goals became apparent. Clinical researchers often are intimately familiar with unique data available to their profession and disease area. For example, Figure 4 illustrates how incorporating standardized hospital tumor registry tables defined by the North American Association of Central Cancer Registries (NAACCR) can act as a standard against which one can compare clinical and billing information from electronic and administrative systems at a health system. Figure 4 also illustrates HERON's ability to link to EHR patient portal usage (MyChart) and incorporate the social security administration death master file (SSDMF) so clinical researchers may exclude patients who have died outside their health system from being contacted for clinical trials or use SSDMF status for outcomes research. This leads to a gross observation that trialists and

clinical researchers want access to "their data". They are familiar with relevant clinical workflows and registries unique to their profession and would like it incorporated in a manner similar to how they are used to seeing this data. Since the majority of their research is at a single site, they are less concerned with aligning data to national standards which may actually hinder their interpretability.

In contrast, PCORnet's data coordination was centered at Harvard Pilgrims which is epidemiology focused. Upon reflection, that was clear from the initial funding announcement which called for participants to: 1) create research-ready datasets that included comprehensive data from EHRs to describe patients' care experience over time and in different care settings. 2) CDRNs were to utilize multiple rich data sources to support research, such as electronic health records, insurance claims data, and data reported directly by patients. This called for CDRNs to establish relationships with external data partners (Centers for Medicare and Medicaid, State, private insurers).

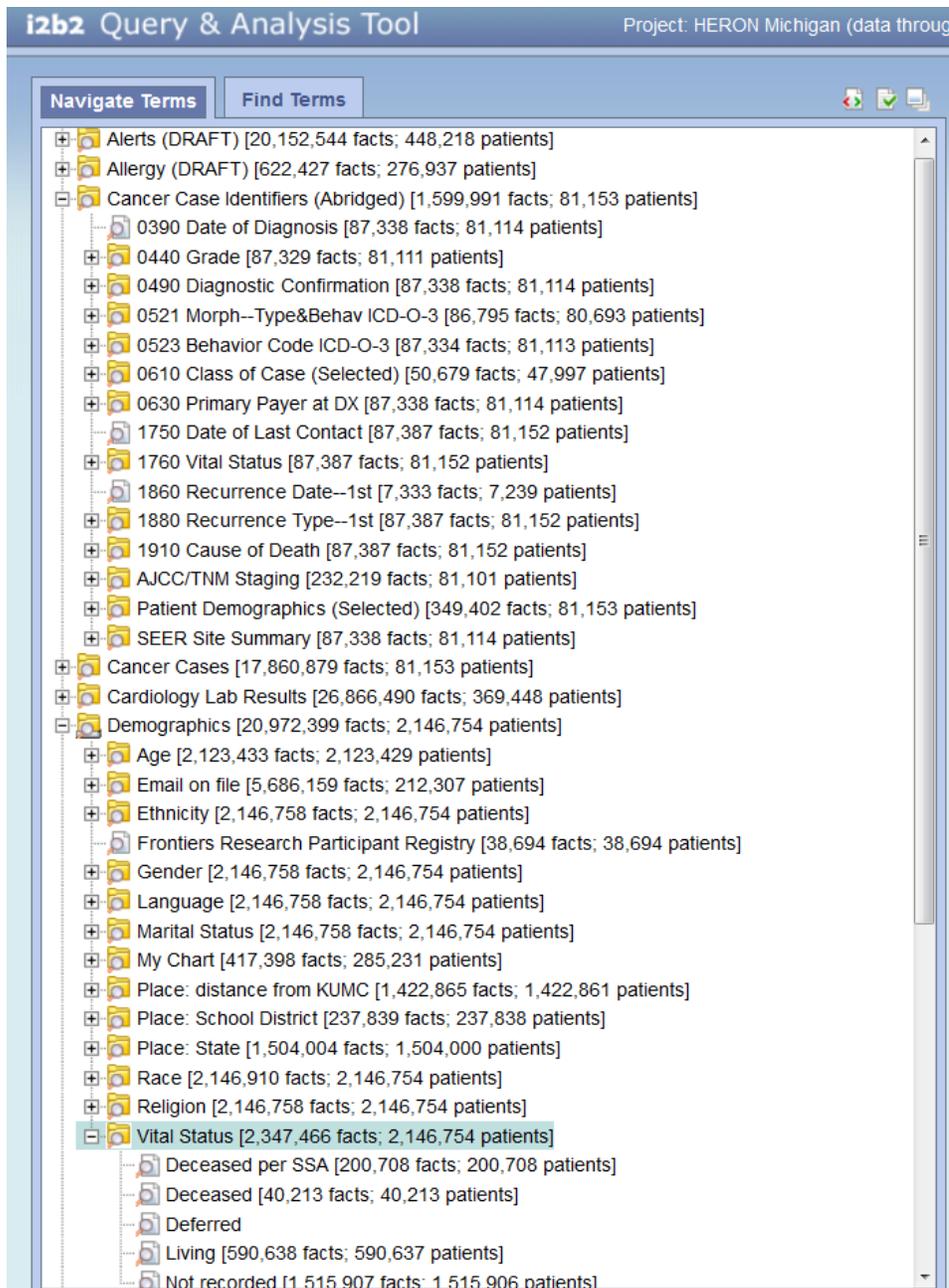


Figure 4: North American Association of Central Cancer Registries (NAACCR) hospital tumor registry data incorporated within the KUMC HERON i2b2 data warehouse.

Figure 5 from the Phase 1 GPC proposal provides a visualization of PCORnet's goal of complete and comprehensive data by revealing gaps in data for a typical academic medical center's cancer center. Data after breast cancer diagnosis

is reasonably complete during treatment but is often missing for common data elements prior to diagnosis (e.g. vital signs, common labs) since the medical center predominantly provides specialty care.

A year after diagnosis, we may see similar decline as the patient's primary care is

provided outside the academic medical center.

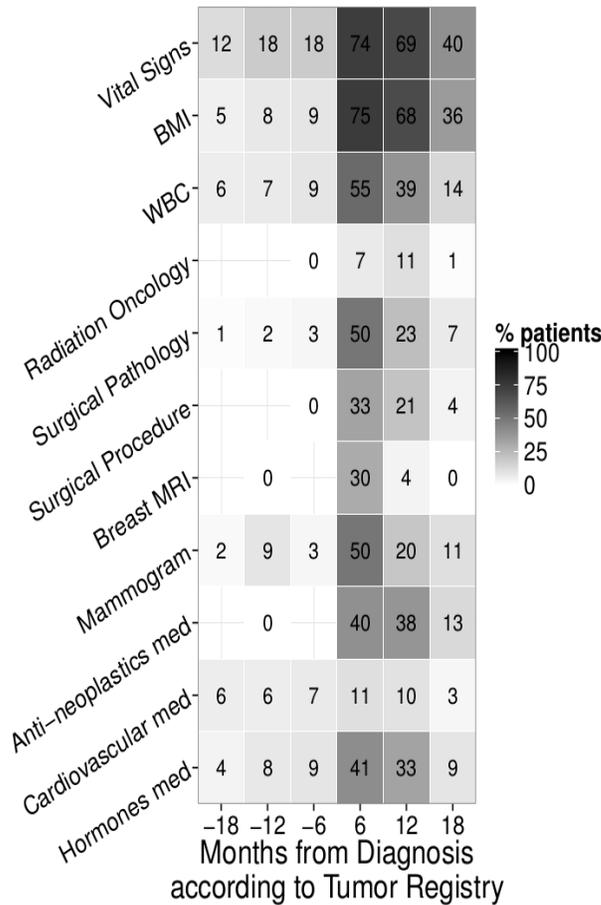


Figure 5. Comprehensive and complete data example from the University of Kansas Cancer Center: heat map of percentage of proposed data elements from the EHR and billing sources recorded in six month intervals surrounding the data of breast cancer diagnosis specified by the hospital tumor registry.

This highlights a challenge for network infrastructure teams: clinical researchers seeking data specific to their professions specific needs and workflow while epidemiologists seeking transforming data out of specific workflows and registries into a unified common data model that harmonizes “all the data”. Medical centers struggle as they participate in multiple national initiatives with how to manage both kinds of customers

(predominately local clinical researchers versus national network epidemiologic driven) and potentially conflicting national common data models required for participation in different efforts (NIH CTSA, All of Us for the National Cancer Institute, PCORnet, Mini-Sentinel for FDA, etc).

GPC has chosen various strategies to promote data exchange and compare terminology alignment across partners. Sites use i2b2 to incorporate rich clinical

data sources such as cancer registries and share their terminologies using a centralized website (<https://babel.gpcnetwork.org>) shown in Figure 6 but also using software developed in partnership with Harvard University and the SCILHS network (<http://scilhs.org/>) to transform data from i2b2 into the PCORnet Common Data Model (<https://github.com/kumc-bmi/i2p-transform>). Sites align their data in i2b2 to use consistent terminologies for common variable such as diagnoses, procedures, demographics and laboratory results. This allows the network to support local

and regional investigators who can directly use i2b2 to determine study feasibility and KUMC developed software to extract data from i2b2 (Data Builder; <https://informatics.gpcnetwork.org/trac/Project/wiki/DataBuilder>) but also have their data quality checked as it's transformed into the PCORnet CDM. The PCORnet CDM in turn allows the medical centers in the GPC to participate in the national research initiatives and also supports GPC level investigators developing studies that will leverage the PCORnet CDM.

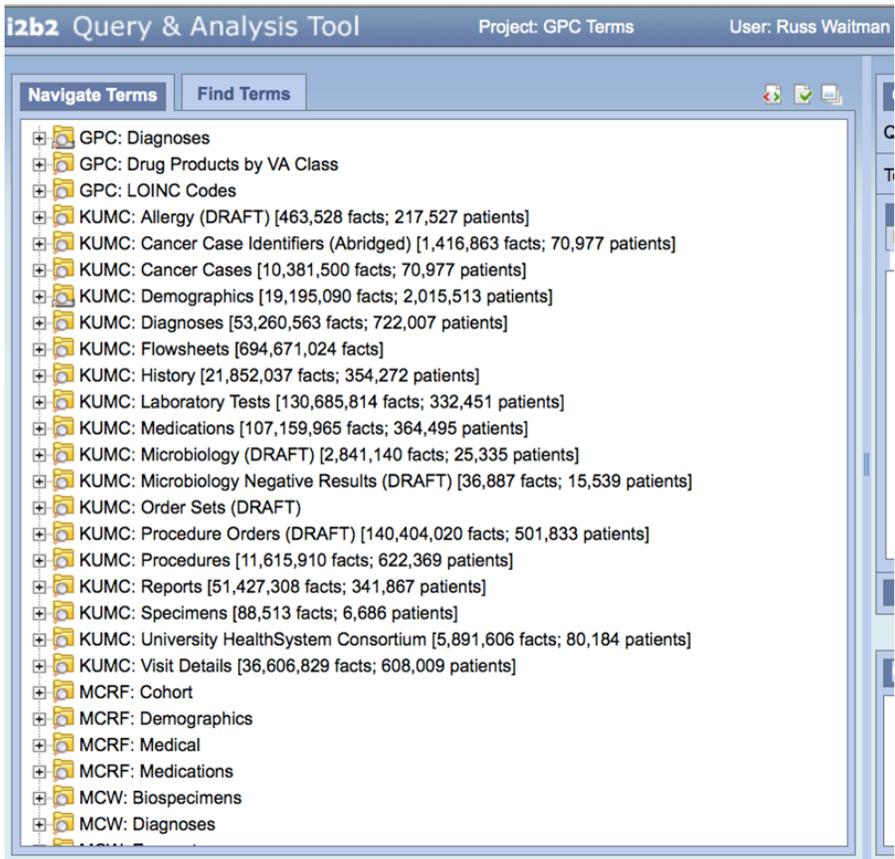


Figure 6. Greater Plains Collaborative Babel terminology service illustrating terminology types from University of Kansas, Marshfield Clinic, and the Medical College of Wisconsin.

IV. From Infrastructure Building to National Studies: prospective interventional and observational PCORnet demonstration projects

While much of PCORnet's activity was establishing governance and data infrastructure, the network also collaboratively prioritized and devised three national demonstration projects: the prospective ADAPTABLE pragmatic trial and two observational studies regarding obesity.

Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) (<http://theaspirinstudy.org/>) is PCORnet's first pragmatic clinical trial. It is not only important clinically but provides the richest test of a medical center's willingness and capability to conduct trials in a novel, more efficient manner. ADAPTABLE seeks to compare the effectiveness and safety of two doses of aspirin (81 mg and 325 mg) in 20,000 high-risk patients with atherosclerotic cardiovascular disease (ASCVD). Its cost per enrolled patient is an order of magnitude lower than

traditional trials (ADAPTABLE = \$850; 3 Simple, NIH Pragmatic Trials: \$2,260 to \$13,269; Industry Trials: \$8,500). Its key innovation is to determine eligible patients by screening electronic health records for defined by a computable phenotype and then approach them via predominantly high volume, low cost channels (email, EHR patient portals, and physical mailers). Patients then visit the ADAPTABLE website and enter their "Golden Ticket" provided by the approach email/letter. After reviewing the trial and online consent videos, the patients are consented and enrolled, typically at their convenience in their home. ADAPTABLE includes researchers from 8 PCORnet CDRNs and 35 health systems along with an "Adaptor Team composed of 8 patients representing each CDRN supported by the Health eHeart Alliance PPRN. While enrollment has now exceed 5000 patients, Figure 7 provides a snapshot of enrollment yield rates across sites in June 2017.

CDRN	Site	Total Number Eligible	Total Number Approached	% of Eligible Approached	Golden Tickets Entered	% Golden Tickets entered per Approached	Total Enrolled	# Non-Internet Enrolled	% Enrolled Per Approached	% Enrolled Per Golden Ticket Entered	Enrolled last week
MidSouth	Vanderbilt	22,271	10,314	46%	1,282	12%	686	20	7%	54%	39
Mid-South	Duke	20,127	1,236	6%	549	44%	413	78	33%	75%	26
PaTH	UPMC	13,879	8,682	63%	1,101	13%	349	0	4%	32%	8
REACHnet	Ochsner	13,560	6,941	51%	651	9%	254	46	4%	39%	3
OneFlorida	U of Florida	29,738	3,110	10%	268	9%	203	36	7%	76%	8
NYC-CDRN	Montefiore	53,151	2,437	5%	214	9%	174	69	7%	81%	8
PaTH	Penn St	5,246	3,885	74%	463	12%	156	0	4%	34%	7
PaTH	Utah	5,219	5,945	114%	306	5%	138	14	2%	45%	6
GPC	Iowa	11,391	5,374	47%	284	5%	136	13	3%	48%	4
GPC	KUMC	4,209	4,024	96%	288	7%	128	0	3%	44%	1
GPC	MCW	12,220	6,108	50%	347	6%	121	0	2%	35%	2
GPC	Marshfield Clinic	8,277	6,083	73%	224	4%	115	0	2%	51%	19
CAPriCORN	Northwestern	6,697	5,754	86%	179	3%	84	2	1%	47%	2
pScanner	UCLA	15,669	5,229	33%	150	3%	75	3	1%	50%	0
CAPriCORN	U of Chicago	5,446	574	11%	77	13%	72	41	13%	94%	0
REACHnet	BSW	2,431	2,220	91%	136	6%	47	6	2%	35%	2
NYC-CDRN	Weill Cornell	5,856	1,282	22%	147	11%	44	1	3%	30%	2
NYC-CDRN	NYU	31,795	1,126	4%	142	13%	32	1	3%	23%	0
PaTH	Temple	6,522	4,989	76%	130	3%	30	6	1%	23%	0
CAPriCORN	Rush	8,826	1,365	15%	57	4%	28	2	2%	49%	0
Mid-South	UNC	5,204	692	13%	54	8%	21	3	3%	39%	5
NYC-CDRN	Mt Sinai	15,832	545	3%	57	10%	20	7	4%	35%	0
GPC	UTSW	2,459	522	21%	30	6%	18	0	3%	60%	1
GPC	Missouri	1,204	617	51%	29	5%	10	0	2%	34%	0
REACHnet	Tulane	771	124	16%	19	15%	4	2	3%	21%	0
GPC	Nebraska	3,475	102	3%	11	11%	4	0	4%	36%	0
PaTH	Johns Hopkins	23,935	5	0%	3	60%	1	0	20%	33%	0
TOTAL		335,410	89,285	27%	7,198	8%	3363	350	4%	47%	143

Figure 7. PCORnet’s ADAPTABLE trial enrollment and recruitment yield rates across sites circa June 2017.

PCORnet also undertook two observational studies addressing controversial subjects with the largest sample sizes to date. The bariatric surgery study (largely adult patients with seven GPC sites participating) includes 48 institutions, 11 CDRNs, 3 PPRNs, and 65,000 people (1,000 of who are adolescents) and studies which surgical approach is best for treating severe obesity between Roux-en-y gastric bypass, Adjustable gastric banding, or Sleeve gastrectomy shown in Figure 8.

This study focuses on one, three and five year outcome that matter to obese patients: weight loss, improvement in diabetes, and risk of adverse events.

The pediatric obesity survey studies whether antibiotics given to children increase risk for obesity and includes 10

Participating CDRNs, 4 Participating PPRNs, 41 Institutions, and 650,000 children. Its main effect analyses evaluates antibiotics use during the first 24 months and weight outcomes at five and ten years of age.

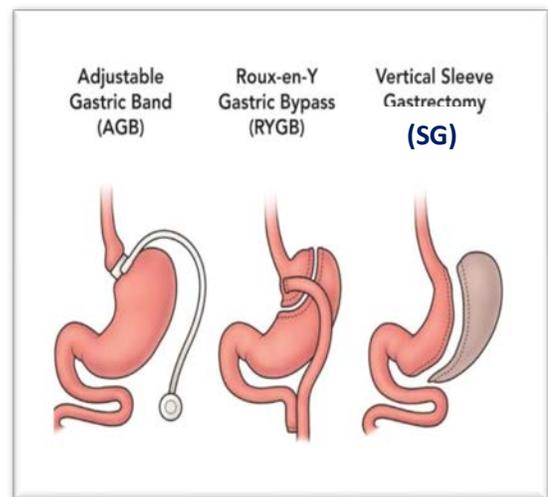


Figure 8. PCORnet bariatric study surgical approaches (disseminated by the PCORnet coordinating center).

V. How does a medical center (KUMC) and the GPC (peer AMCs) fit into the evolving national landscape?

As the University of Kansas Medical Center and our peers in the Greater Plains Collaborative complete four years of building PCORnet, we reflect upon how our participation has impacted our campuses. The majority of our campuses are involved in all three demonstration projects. Our network, led by Dr. Elizabeth Chrischilles at Iowa, leads the national collaborative research group for advancing PCORnet's cancer research and Dr. Russ Waitman has served as the national chair for the PCORnet data committee. GPC sites have been responsive to national common data model queries issued from the national coordinating center to each participating site and all GPC sites adopted the SmartIRB national reliance model before any other network. The GPC strategy to merge Medicare and Medicaid Claims was successful and has created a centralized claims repository for integrating EHR and CDM data (<https://informatics.gpcnetwork.org/trac/Project/wiki/GROUSE>). While the national structure of PCORnet is evolving, in many ways the GPC has served as a data fitness camp for academic medical centers to participate in national data intensive research. PCORI announced in 2017 that it would transition infrastructure support for PCORnet to a newly created non-profit: the Patient Centered Research Foundation (<http://www.pcrfoundation.org/>) which will in turn contract with Clinical Data Research Networks instead of networks contracting with PCORI. This will give the networks and foundation flexibility

in seeking varied sponsors and funders who may seek to use the network. Questions arise though as to how structure informs network design and collaboration. Currently, PCORnet is coordinated in largely a traditional model where recruiting sites serve the central coordinating center. But new trends, such as reciprocal IRB and the Greater Plains Collaborative's use of complementary reciprocal data sharing shift the model to allow interoperable data exchange and coordination so that trials made led by each participating site or medical center. GPC currently sees its role as shifting to an intermediary, member governed collaborative providing the following services and roles: 1) helping member improve their regulatory, patient and clinician engagement to complement data infrastructure for participating in national research, 2) contracting with Patient Center Research Foundation (PCRF) PCORnet 2.0, 3) governing peer to peer data Sharing to complement SMART IRB (http://www.gpcnetwork.org/sites/default/files/GPC%20Resource%20Guide_Pilot%20Program%20Supplement.pdf), 4) providing a forum for members to share technology and be accountable to one another for data quality and capability, and 5) consolidating data assets as needed such as CMS claims via GROUSE. This structure, shown in Figure 9 would allow grants to be awarded at the member site and contracting for services if needed at higher organizational levels (GPC or PCRF – formerly referred to as NewCo). National collaborative opportunities would flow-

down through GPC who coordinates regional quality and capability as well as supports centralized resources.

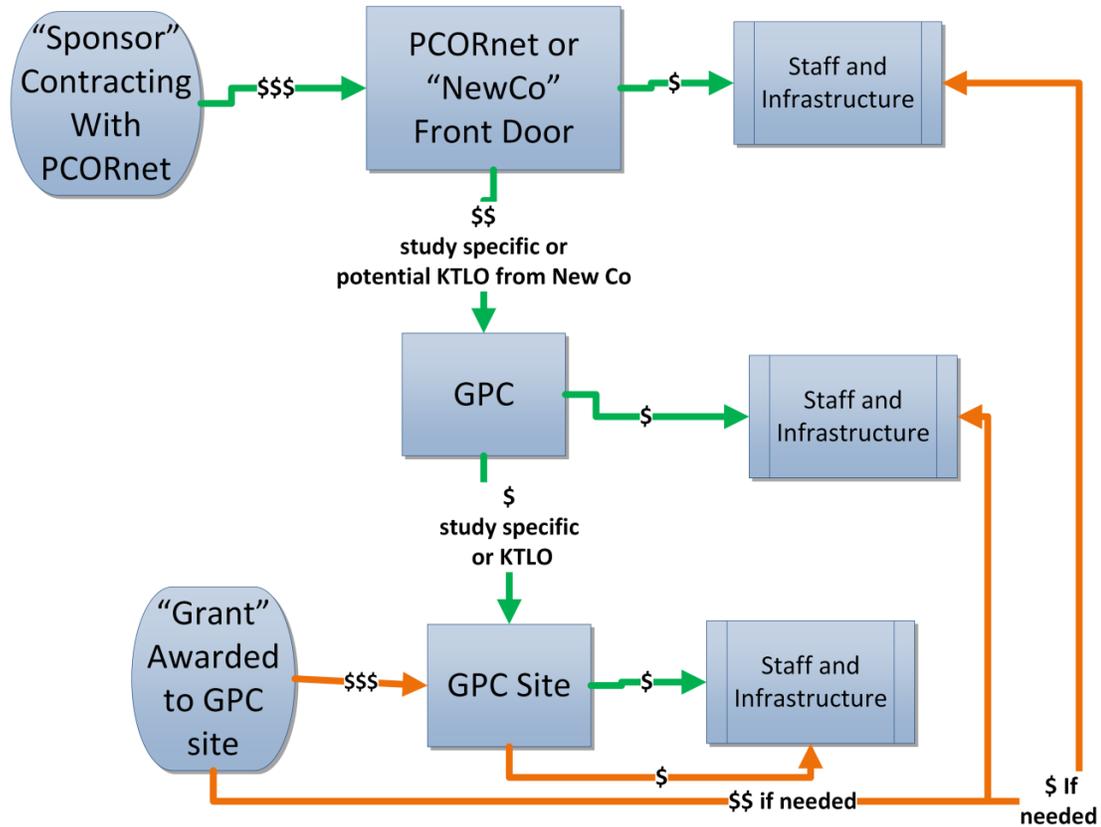


Figure 9. Proposed Greater Plains Collaborative sustainability model post Phase 2 PCORI contract

PCORnet and the Greater Plains Collaborative have catalyzed dramatic changes in regulatory, patient engagement, and data infrastructure to support research at the point of care. But, data, technology, and organizational relationship are very fluid so we are constructing these capabilities in a very dynamic time. Medical centers will continue to demonstrate their desire to lead by providing responsive regulatory and contracting activities, flexible data infrastructure, and the ability to deploy informatics interventions at the point of care with integrated

and responsive approaches to patient, clinician and researcher engagement for both research and care delivery.

Acknowledgements

Portions of this text and figures were previously used in the university CTSA application biomedical informatics sections in 2010 (authors: Lemuel Waitman, Judith Warren, and Gerald Lushington) and 2016 (Lemuel Waitman) and in the PCORI PCORnet 2013 phase 1 and 2015 phase 2 applications for the Greater Plains Collaborative (authors: Lemuel

Waitman, Gary Rosenthal, Lauren Aaronson, Prakash Nadkarni, James Campbell, Daniel Connolly). Multiple references to PCORnet and especially the three demonstration projects were summarized from PCORnet slides developed during 2015-2018 by multiple individuals from PCORI and the PCORnet networks and coordinating centers; many are available at the PCORnet website (<http://pcor-net.org/>) and the Commons (<http://pcor-netcommons.org/>). We would like to acknowledge the participation of Frontiers, Greater Plains Collaborative, and PCORnet leadership and numerous investigators who have contributed to the development and ongoing progress of informatics and pragmatic research infrastructure in the Midwest and nationally.

References

Waitman LR, Lushington G, Warren JJ. Advancing Clinical and Transformational Research with Informatics at the University of Kansas Medical Center. Information Systems as Infrastructure for University Research Now and in the Future. Merrill Series on the Research Mission of Public Universities Mabel L. Rice, Editor. MASC Report No. 116 The University of Kansas 2012. 90-106.

Zerhouni EA. Translational and clinical science—time for a new vision. *NEJM* 2005, 353:1621-23.

Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc.* 2010 Mar-Apr;17(2):124-30.

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009 Apr;42(2):377-81. doi: 10.1016/j.jbi.2008.08.010. Epub 2008 Sep 30.

Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific evidence underlying the ACC/AHA clinical practice guidelines. *JAMA.* 2009 Feb 25;301(8):831-41. doi: 10.1001/jama.2009.205. Erratum in: *JAMA.* 2009 Apr 15;301(15):1544. PubMed PMID: 19244190.

Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. *J Am Med Inform Assoc.* 2014 Jul-Aug;21(4):578-82. doi: 10.1136/amiajnl-2014-02747. Epub 2014 May 12.

Waitman LR, Aaronson LS, Nadkarni PM, Connolly DW, Campbell JR. The Greater Plains Collaborative: a PCORNet Clinical Research Data Network. *J Am Med Inform Assoc.* 2014 May 12.