



OPEN ACCESS



Open Access
Scan to access more
free content

Changing the research landscape: the New York City Clinical Data Research Network

Rainu Kaushal,^{1,2} George Hripcsak,³ Deborah D Ascheim,⁴ Toby Bloom,⁵ Thomas R Campion Jr,¹ Arthur L Caplan,⁶ Brian P Currie,⁷ Thomas Check,¹² Emme Levin Deland,² Marc N Gourevitch,⁶ Raffaella Hart,⁸ Carol R Horowitz,⁴ Isaac Kastenbaum,² Arthur Aaron Levin,⁹ Alexander F H Low,¹ Paul Meissner,⁷ Parsa Mirhaji,⁷ Harold A Pincus,^{2,3} Charles Scaglione,¹³ Donna Shelley,⁶ Jonathan N Tobin,^{10,11} on behalf of the NYC-CDRN

For numbered affiliations see end of article.

Correspondence to

Rainu Kaushal, 425 East 61st Street, Suite 301, New York, NY 10065, USA; rak2007@med.cornell.edu

Received 28 February 2014
Revised 15 March 2014
Accepted 25 March 2014
Published Online First
12 May 2014



- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002864>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002747>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002758>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002744>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002746>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002743>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002827>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002740>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002727>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002751>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002745>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002759>
- ▶ <http://dx.doi.org/10.1136/amiajnl-2014-002756>



CrossMark

To cite: Kaushal R, Hripcsak G, Ascheim DD, et al. *J Am Med Inform Assoc* 2014;**21**:587–590.

ABSTRACT

The New York City Clinical Data Research Network (NYC-CDRN), funded by the Patient-Centered Outcomes Research Institute (PCORI), brings together 22 organizations including seven independent health systems to enable patient-centered clinical research, support a national network, and facilitate learning healthcare systems. The NYC-CDRN includes a robust, collaborative governance and organizational infrastructure, which takes advantage of its participants' experience, expertise, and history of collaboration. The technical design will employ an information model to document and manage the collection and transformation of clinical data, local institutional staging areas to transform and validate data, a centralized data processing facility to aggregate and share data, and use of common standards and tools. We strive to ensure that our project is patient-centered; nurtures collaboration among all stakeholders; develops scalable solutions facilitating growth and connections; chooses simple, elegant solutions wherever possible; and explores ways to streamline the administrative and regulatory approval process across sites.

INTRODUCTION

New York City is home to one of the largest, most diverse urban populations in the USA, including more than 8 million people with a wide range of socioeconomic and health characteristics.¹ Its healthcare system is marked by a concentration of academic medical centers with expertise in clinical care, research, and education. Despite this wealth of resources, healthcare delivery remains fragmented, as patients often receive care from multiple institutions, complicating efforts to conduct research, manage population health, and develop learning healthcare systems.

Funded by the Patient-Centered Outcomes Research Institute (PCORI), the New York City Clinical Data Research Network (NYC-CDRN) was formed to create an accessible, sustainable, scalable clinical data network that will enable patient-centered research, support a national research network, and facilitate the development of learning healthcare systems. This project features a unique collaboration across 22 organizations, including seven independent health systems, which will create unprecedented opportunities for city- and nation-wide population

health management, patient-centered clinical trials, observational studies, and precision medicine. Specific goals include aggregating data on a minimum of 1 million patients, engaging patients and front-line clinicians in all phases of the project, embedding research activity into the delivery of healthcare, aligning regulatory oversight across multiple health systems, and disseminating study results across healthcare systems.

This paper describes the project's goals, governance and organizational structure, and technical approach.

ORGANIZATIONAL AND SCIENTIFIC APPROACH

The NYC-CDRN includes a robust and collaborative governance and organizational infrastructure, which takes advantage of its participants' experience, expertise, and history of collaboration.

Participating institutions

The NYC-CDRN's participating institutions (table 1) have several notable features that provide an important foundation for the consortium. The NYC-CDRN includes six Clinical and Translational Science Award (CTSA) centers,² which already collaborate on research, data sharing, and patient engagement. Second, the participating health systems—including five medical schools, four affiliated health systems, and one practice-based research network of federally qualified health centers—have robust electronic health records (EHRs) and clinical data warehouses with many years of data. Third, the New York Genome Center (NYGC), an independent non-profit entity, with which all health systems are affiliated, has important expertise in genomic data and acts as a neutral party and 'honest broker'³ for aggregating and hosting data from competing institutions for research purposes. In addition, two regional health information organizations, Healthix¹ and the Bronx RHIO's Bronx Regional Informatics Center (BRIC), provide important expertise in patient matching and record de-duplication. Cornell NYC Tech, a new graduate school emphasizing technology and entrepreneurship, provides access to new methods for collecting patient-generated data. The Biomedical Research Alliance of New York (BRANY) serves as the centralized institutional review board (IRB) process to ensure appropriate regulatory oversight and protocol reviews. Finally, several patients and patient advocacy groups provide important expertise in patient engagement.

Table 1 NYC-CDRN participating institutions

Partner	Organization	EHR/HIE platform	Patients in EHR/HIE*		
Health system	Clinical Directors Network (CDN)	eClinicalWorks, GE Centricity	250k		
	Columbia University College of Physicians and Surgeons (CUCPS)† Montefiore Medical Center and Albert Einstein College of Medicine (MMC)†	Allscripts Enterprise GE Centricity‡	767k 1000k		
	Mount Sinai Health System and Icahn School of Medicine (MSHS)† New York-Presbyterian Hospital (NYPH)	Epic Allscripts Sunrise	4700k 1400k		
	New York University Langone Medical Center and New York University School of Medicine (NYULMC)† Weill Cornell Medical College (WCMC)†	Epic	1800k 560k		
	Research infrastructure	Biomedical Research Alliance of New York Cornell NYC Tech Campus New York Genome Center Rockefeller University†	N/A	N/A	
		HIE	Bronx RHIO (Bronx Regional Informatics Center) Healthix	Optum InterSystems HealthShare	1650k 7000k
			Patient organization	American Diabetes Association Center for Medical Consumers Consumer Reports Cystic Fibrosis Foundation New York Academy of Medicine NYS Department of Health	N/A

*Patients overlap and are for the period 1 August 2008–31 July 2013.

†Denotes CTSA site.

‡Montefiore is replacing existing EHR platforms with Epic.

CTSA, Clinical and Translational Science Award; EHR, electronic health record; HIE, health information exchange; N/A, not applicable; NYC-CDRN, New York City Clinical Data Research Network.

Organizational structure

The NYC-CDRN has created a multi-stakeholder organizational structure (figure 1) that includes leadership and participation from researchers, clinicians, and patients.⁴⁻⁶ We have organized our work according to seven overarching goals. One committee leads each section and liaises with the other committees to collaborate on cross-cutting issues. For example, the Technical Committee cannot develop its data model without input from researchers, clinicians, and patients in the Comparative Effectiveness Research (CER) and Patient and Engagement Committees.

1. *Create a strong governance and business infrastructure:* The NYC-CDRN has a robust, collaborative governance and organizational model that operates the network in the interests of all participants. The Governance Board oversees the entire project, sets policies in consultation with stakeholders and advisors, and ensures that all committees and stakeholders are on track to meet their deliverables. It addresses open issues within and among the committees, ensures common understanding of key network concepts and

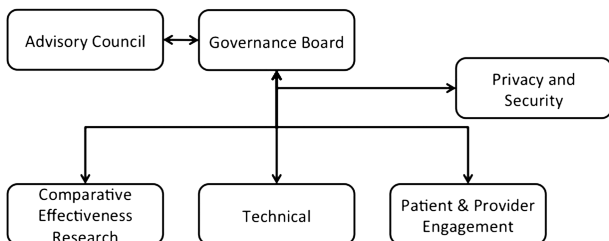


Figure 1 Organizational structure of NYC-CDRN (New York City Clinical Data Research Network).

functions, and facilitates interactions with the healthcare systems among other functions.

2. *Ensure strong accountability and coordination among project committees and stakeholders:* The NYC-CDRN project is a complex endeavor with many moving, intersecting, and inter-dependent parts. The Operations Group has established a project management infrastructure to guide that activity. It drives, monitors, and reports progress; ensures quality and accountability across all stakeholders; and tracks adherence to milestones and timelines.
3. *Develop an overarching vision and sustainability:* The NYC-CDRN reviews its strategy and vision with an Advisory Council of external healthcare leaders and subject matter experts. The Council ensures that the project benefits from new ideas, stays aligned with local and national developments, and focuses on financial sustainability.
4. *Establish a legal foundation that protects patient privacy and security:* All participating health systems have data sharing policies, IRB processes, and privacy and security policies in place. However, it would be a slow process for researchers interested in multi-site studies to obtain necessary approvals and negotiate separate policies and requirements individually from all IRBs. Thus, the project's Privacy and Security Group works with participants to agree to a common, consistent set of network processes, policies, and data sharing agreements. The participants have agreed to use a central IRB, housed at BRANY.
5. *Engage patients and clinicians:* This project relies on strong leadership and input from patients and clinicians in all its phases. Patients and clinicians participate in governance, inform and develop research questions, and ensure that the network's policies protect patient privacy and security. The Patient and Clinician Engagement Committee ensures that

Table 2 New York City population characteristics

Characteristic	%
Age (years)*	
≤19	24
20–44	39
45–64	24
65+	12
Median	36
Race*	
White	44
Black	26
Hispanic/Latino*	28
Female*	53
% Household income <\$25k†	28
% Publicly insured†	37
% Self-reported diabetes‡	11
% Self-reported high cholesterol‡	31
% Self-reported current cholesterol meds‡	37
% Self-reported high blood pressure‡	29
% Self-reported asthma‡	12
% Overweight and/or obese‡	58
% Receiving mental health medication	4
% Current smoker	15

*2010 Census.
 †2009 American Community Survey.
 ‡2011 NYC Community Health Survey.

all the other committees are identifying key policies and processes needing patient and clinician input. It also focuses on the collection of patient-reported outcomes.

6. *Embed research into practice:* Participating institutions all have expertise and experience in embedding aspects of research into practice while minimizing disruption of healthcare delivery—identifying patients for research, implementing research protocols, monitoring activities, and disseminating research outcomes to improve practice. The CER Committee develops use cases for the network and ensures that the network facilitates different types of research designs, including retrospective studies, observational studies, and randomized clinical trials at the level of the individual and cohort. Community workgroups are being established to identify the best ways to engage patients in those communities and to inform research.
7. *Build the technical infrastructure of the research data network:* In their initial 18 months, all CDRN projects must aggregate comprehensive, longitudinal data for at least 1 million patients for research purposes. Given the number of institutions involved, it is a significant challenge to compile that data in a standard way, match and link patient identities across institutions, de-identify the records, and

make available quality data. The project’s Technical Committee oversees the design of the network architecture, the data model, and the design for the NYC-CDRN Informatics Center. These activities are described in more detail below.

Patient population and selected cohorts

PCORI CDRN awardees must focus on three conditions: a common condition, a rare condition, and obesity. NYC-CDRN has selected diabetes as its common condition and cystic fibrosis as its rare condition. According to the official city data, nearly 60% of New Yorkers are either overweight (34%) or obese (24%), and 11% have diabetes (table 2). Cystic fibrosis is a genetic disease that affects the digestive and respiratory systems, and NYC-CDRN has identified over 5000 patients among its institutions.

TECHNICAL APPROACH

The NYC-CDRN’s technical approach will employ an information model to document and manage the collection and transformation of clinical data, local institutional staging areas to transform and validate data, a centralized data processing facility to aggregate and share data, and use of common standards and tools.

The NYC-CDRN Informatics Center, hosted at NYGC, will aggregate data from all the health systems centrally and make it available for research queries (figure 2). NYC-CDRN is being designed so that it is not constrained to a single technology or platform. It will utilize agile design and development with testing and iterative refinements as well as extensive quality controls.

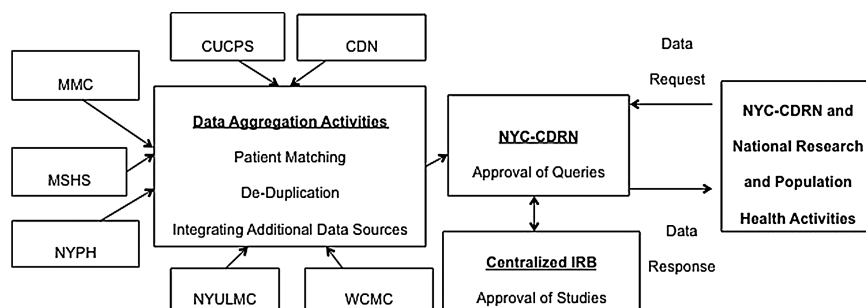
Informational model

The NYC-CDRN will use a centrally defined information model and standardized set of terminologies to form the basis of data integration across institutions and for interoperability with PCORnet. Health systems will extract data from their EHRs or clinical data warehouse platforms according to a common set of vocabularies and then leverage existing models such as Observational Medical Outcomes Partnership (OMOP) to validate mappings between standard vocabularies to integrate demographics, ethnicity and race, diagnoses, procedures, medications, laboratory results, and other clinical elements.⁷ By separating representation of concepts from data storage implementation, the information model will enable use of different technologies for distinct purposes.

Local staging areas

Health systems will host a local staging area for their data feeds. They will follow procedures defined by NYC-CDRN for

Figure 2 NYC-CDRN (New York City Clinical Data Research Network) data flows.



standards-based mapping of health information, quality assurance, data cleaning, and validation prior to sending a limited dataset to the Informatics Center. Systems iteratively will contribute data to the Informatics Center. For example, the first deliverable is for institutions to contribute patient demographics⁸ in a defined format followed by patient encounter data and then clinical observation data such as diagnoses, procedures, medications, and laboratory results as defined in the central information model.

Centralized data processing facility

The Informatics Center will aggregate each system's data into a patient-matched, de-duplicated central database and perform data shifting to preserve anonymity. This de-identified dataset will be available for query by investigators.⁹

It is critically important for a project like NYC-CDRN to match patients while preserving anonymity across multiple EHRs as a way of creating an integrated and complete view of longitudinal clinical data.^{10 11} The Informatics Center will leverage two health information exchanges' existing electronic master patient indices, patient matching algorithms, and patient de-duplication techniques provided by vendors (table 1) to align data contributed by systems to NYC-CDRN.

The central database will link to other sources including public and commercial claims data; patient-reported and patient-generated data, including data actively collected through surveys and passively collected through mobile devices; genomic data allowing for novel links to biologic and molecular disease markers; and other publicly available data.¹²

DISCUSSION

The NYC-CDRN is an ambitious project that has the potential to significantly change the research landscape in New York City and help shape national research efforts through the national PCORnet. To ensure our best chance of success, we abide by several guiding principles.

First, we strive to make the network truly patient-centered. We conduct all our activities in a fashion that is guided by, and accessible and understandable to, patients, caregivers, and their care teams. Patients have a wealth of knowledge about their conditions and healthcare experience that can inform and inspire new research opportunities.

Second, NYC-CDRN depends on the active and successful collaboration of many different institutions and individuals. By nurturing that collaboration effectively, we will have access not only to a great wealth of existing expertise and resources within our participating institutions but also to new ideas and initiatives created by the interaction of those parties, such as innovative research protocols, patient engagement methods, and technical models.

Third, the network needs to scale easily. As NYC-CDRN builds a network of health systems, we must continue to add new partners and link to the national PCORnet network. NYC-CDRN will draw strength from its scale.

Fourth, we strive to not over-complicate an already complex job. We endeavor to choose simple, elegant solutions wherever possible. For example, we are employing an iterative process to develop our data model—starting with small sections of the template, building, testing, and improving before expanding the dataset and moving on to new sections.

Finally, we strive to streamline the administrative and regulatory process to ensure that researchers can embark on critical research studies in a timely fashion, while ensuring the highest standards of patient safety and privacy.

CONCLUSION

With funding from PCORI, the NYC-CDRN is creating an accessible, sustainable, scalable clinical data network that will enable patient-centered research embedded within the functioning healthcare system, support a national research network, and facilitate the development of learning healthcare systems. The NYC-CDRN is well positioned to transform the research landscape in New York City and create new opportunities for wide-scale collaborations to design, conduct, and disseminate innovative clinical trials, CER, and population health management.

Author affiliations

- ¹Weill Cornell Medical College, New York, New York, USA
- ²NewYork-Presbyterian, New York, New York, USA
- ³Columbia University, New York, New York, USA
- ⁴Icahn School of Medicine at Mount Sinai, New York, New York, USA
- ⁵New York Genome Center, New York, New York, USA
- ⁶New York University Langone Medical Center, New York, New York, USA
- ⁷Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, New York, USA
- ⁸Biomedical Research Alliance of New York Institutional Review Board, Lake Success, New York, USA
- ⁹Center for Medical Consumers, New York, New York, USA
- ¹⁰Clinical Directors Network, New York, New York, USA
- ¹¹The Rockefeller University, New York, New York, USA
- ¹²Healthix, New York, New York, USA
- ¹³Bronx RHIO (Bronx Regional Informatics Center), Bronx, New York, USA

Funding Patient-Centered Outcomes Research Institute (PCORI); contract number CDRN-1306-03961.

Contributors RK, GH, TRC, AFHL prepared the concept and body of the manuscript. DDA, TB, ALC, TC, ELD, RH, CRH, IK, AAL, PMeissner, PMirhaji, HAP, CS, DS, and JNT participated in conceptualization and provided critical feedback to the manuscript.

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- 1 Onyile A, Vaidya SR, Kuperman G, *et al*. Geographical distribution of patients visiting a health information exchange in New York City. *J Am Med Inform Assoc* 2013;20:e125–30.
- 2 Collins FS. Reengineering translational science: the time is right. *Sci Transl Med* 2011;3:90cm17.
- 3 Boyd AD, Saxman PR, Hunscher DA, *et al*. The University of Michigan Honest Broker: a web-based service for clinical and translational research and practice. *J Am Med Inform Assoc* 16:784–91.
- 4 Hripscak G, Bloomrosen M, Flatlybrennan P, *et al*. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. *J Am Med Inform Assoc* 2014;21:204–11.
- 5 Kim KK, Browe DK, Logan HC, *et al*. Data governance requirements for distributed clinical research networks: triangulating perspectives of diverse stakeholders. *J Am Med Inform Assoc* 2014;21:714–9.
- 6 Lorenzi NM, Riley RT, Blyth AJ, *et al*. Antecedents of the people and organizational aspects of medical informatics: review of the literature. *J Am Med Inform Assoc* 1997;4:79–93.
- 7 Overhage JM, Ryan PB, Reich CG, *et al*. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* 2012;19:54–60.
- 8 Weber GM, Murphy SN, McMurry AJ, *et al*. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. *J Am Med Inform Assoc* 2009;16:624–30.
- 9 Murphy SN, Gainer V, Mendis M, *et al*. Strategies for maintaining patient privacy in i2b2. *J Am Med Inform Assoc* 2011;18(Suppl 1):i103–8.
- 10 DuVall SL, Fraser AM, Rowe K, *et al*. Evaluation of record linkage between a large healthcare provider and the Utah Population Database. *J Am Med Inform Assoc* 2012;19:e54–9.
- 11 Weber GM. Federated queries of clinical data repositories: the sum of the parts does not equal the whole. *J Am Med Inform Assoc* 2013;20:e155–61.
- 12 SPARCS Overview [Internet]. [cited 2013 Sep 21]. <https://www.health.ny.gov/statistics/sparcs/operations/overview.htm>