

**U.S. Department of Health and Human Services
Office of the National Coordinator for Health Information Technology**

**HIT Policy Committee
Meaningful Use Workgroup**

Panel 4: Finding Solutions, Creating Outcomes

Wednesday, October 5, 2011

Prepared Statement

Richard Elmore

ONC

Coordinator, S&I Framework – Query Health

Thank you for the opportunity to participate in the Finding Solutions, Creating Outcomes panel.

I am serving as the ONC S&I Framework Coordinator for Query Health. Query Health was launched 4 weeks ago to establish the standards and protocols for distributed population queries of EHRs and other clinical records. These are aggregate queries which, with the exception of potentially individual identifiable data for public health permitted uses, solicit responses consisting of aggregated de-identified data or limited data sets. Distributed population queries can be used for a variety of applications – from disease outbreaks, prevention activities, health research, performance measures and others.

The Office of the National Coordinator for Health IT recently published the updated Federal Health IT Strategic Plan. Query Health is an integral part of the ONC strategy directly related to two of these goals:

- improving population health and
- enabling innovation and appropriate use of health information to improve knowledge about health across populations, sometimes called the learning health care system.

In preparation for Query Health, ONC launched the Summer Concert Series which is an environmental scan of work on distributed population queries. Practitioners were asked to illustrate their approach and rationale for using aggregated information from distributed queries to improve healthcare quality, medical research, public health monitoring and prevention, among others. These are vitally important applications to the health of patients and populations. The following summarize some key points from the environmental scan.

PopMedNet uses distributed networks for secondary use such as disease surveillance, comparative effectiveness, and medical product safety. This summer, the FDA identified final specifications for a query to monitor the cardiac outcomes of a particular drug. Rich Platt and Jeff Brown, using PopMedNet, distributed the specs to 17 data partners with a combined 99 million individuals. Within two days the report was delivered back to FDA outlining statistics about new users (including distributions by age, sex, year, and health plan), first treatment exposure duration, and cardiac outcome diagnoses during treatment. PopMedNet's flexibility allows for each network of participants to create, administer and coordinate their own networks. PopMedNet is responsive to non-regulated data requirements where healthcare organizations want the control over their information that distributed queries provide.

i2b2 is a robust model used primarily for research and is deployed at over 60 organizations. i2b2's distributed network capability, **Shrine**, enables distributed population queries in several different networks across the country. I2b2 maintains specific aspects of ontology management, data ownership, and patient privacy while overlaying a generic data architecture. The generic star schema data model and the service oriented architecture of i2b2 lends itself to a natural implementation of a distributed query system. Zak Kohane's research into comorbidities and autism showed the power of this new approach. Prior studies had taken years to complete and reviewed ~200 patients. In 3 weeks, he was able to determine the comorbidities with autism across 14,000+ patients and was able to compare that to a general hospital patient population of millions.

The **Indiana Network for Patient Care (INPC)** is one of the nation's most comprehensive and longest tenured health information exchanges, containing more than 3.1 billion coded standardized clinical observations across 12 million patients, and a physician cohort exceeding 19,000 providers. In addition to providing a single virtual record across dozens of Indiana hospitals, the INPC supports many other aggregate population measure use cases including public health reporting, clinical research, and quality reporting. Sean Grannis demonstrated how Regenstrief's implementation demonstrates the power of distributed population queries at a regional level for public health reporting, clinical research and quality reporting.

Observational Medical Outcomes Partnership or OMOP conducts methodological research to inform the appropriate use of observational healthcare data including administrative data and electronic health records. Patrick Ryan, March Overhage and Tom Scarnecchia reported on how OMOP has established a central research laboratory and a distributed data network of over 150 million lives. OMOP's common clinical information model has been applied with DARTNet and other distributed query networks, and OMOP's tools have been

applied for many use cases identified as priorities within the ONC Query Health initiative.

The **Distribute** project was initiated by the International Society for Disease Surveillance (ISDS) in 2006 to pilot influenza-like illness (ILI) monitoring using data aggregated from syndromic surveillance systems operated by state and local health departments. David Buckeridge and team have demonstrated the value and quality of aggregate data for disease monitoring using Distribute. In 2009, following the first wave of H1N1 influenza pandemic, Distribute was rapidly expanded to enhance ILI surveillance, covering approximately one-third of US emergency department visits.

CDC's **BioSense 2.0** program is targeted at nationwide and regional situational awareness for all-hazard health-related threats and to support national, state, and local responses to those threats. Taha Kass-Hout described how BioSense 2.0 is architecturally distributed in a cloud-based model removing many of the costs and limitations associated with the monolithic BioSense 1.0 physical architecture, while still making the distributed aspects of the system transparent to end users.

DARTNet is a collaborative of seven electronic health record based research networks that have agreed to standardize data codification and underlying analytical data models. With Wilson Pace's leadership, the groups work together to advance research methodology from the macro level, such as study design, to the micro level, such as analytical decisions. All participants share an underlying practice-based research framework on which they are adding the ability to conduct large comparative effectiveness and pragmatic clinical trials.

The New York City Department of Health's **Hub Population Health System** ("The Hub") enables the distribution of electronic health record queries, targeted clinical decision support, and secure messages to more than 350 small practices and large community health centers in New York City, covering 1.1 million patients. Jesse Singer and Michael Buck have demonstrated that the power and value of distributed queries applied to small practices. The hub can also dynamically send new queries to its network.

Both popCCR and hQuery re-use standardized data exchange as a Rosetta stone for addressing a long-standing challenge– the need to be able to relate local clinical information to a standard information model.

populationCCR is an approach for simplifying and integrating performance reporting and decision support. Steven Waldren's strategy was to decouple analytics from current and future health IT systems. popCCR eases the creation of rules by non-programmers and harmonizes the representation of clinical data

across patient records, performance reporting, and decision support - all by using widely accepted standard technologies.

hQuery is built on the lessons learned from the popHealth and Laika open source initiatives. Marc Hadley and Andy Gregorowicz used pervasive internet-based technologies to deliver flexible easy-to-understand queries and distributed ultra-large scale query execution capabilities. It uses a simplified standard information model, based on nationally recognized data elements, to enable generic queries against and summarized responses from various clinical data sources. hQuery integrates with certified EHR systems by leveraging their ability to generate HITSP C32 or ASTM CCR patient summaries.

caBIG – the Cancer biomedical informatics grid is an NCI sponsored collaborative information network that accelerates the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer. Ken Buetow describes caGrid as an open-source framework that addresses the needs of a heterogeneous community to access, integrate, and analyze biomedical data of diverse nature and source. caGRID allows diverse, authorized applications, running on multiple platforms, to access appropriate data (both individual level and aggregated) from distributed sources.

New York's **Universal Public Health Node** (UPHN) is an informatics approach and technical infrastructure for public health practice through health information exchanges in New York. Ivan Gotham, Leroy Jones and Vince Lewis demonstrated the power of regional coordination and the use of a publish and subscribe model for distributed queries to keep control with the disclosing entities.

What have we learned from these efforts, apart from the somewhat common distaste in capitalizing project names? Out of the summer concert series, we heard a consistent set of value statements related to distributed population queries. Distributed population queries today can enable:

- Comparative effectiveness research
- Speed, adaptability and broad patient coverage in response to a clinical question
- Understanding of condition counts and disease prevalence
- Monitoring of disease outbreaks
- Proactive and preventative care
- Performance and quality reporting
- Control by data stewards
- Better protected patient data

Those are powerful and compelling benefits. With a few notable exceptions, today these benefits accrue only to organizations with larger research or IT budgets.

The Institute of Medicine, sponsored by ONC, recently convened experts that documented the requirements for a digital infrastructure for a learning health system. These include:

- Build a shared learning environment
- Engage health and health care, population and patient
- Leverage existing programs and policies
- Embed services and research in a continuous learning loop
- Anchor in an ultra-large-scale systems approach
- Emphasize decentralization and specifications parsimony
- Keep use barriers low and complexity incremental
- Foster a socio-technical perspective, focused on the population
- Weave a strong and secure trust fabric among stakeholders
- Provide continuous evaluation and improvement

Distributed population queries are a key step in enabling this digital infrastructure to improve the health of the nation. Query Health's goal is to create standards for distributed population queries that everyone can leverage – including enabling national scaling of distributed query networks such as the examples already cited.

Panel Question: What are the key data challenges to improving America's health system from the perspective you represent (e.g., quality measure development, certification of healthcare professionals, consumer use of comparison data, coordination of care, payer)?

For any of the healthcare experts that have dedicated themselves in service to the nation through work on registries or distributed population queries, Job 1 is to have a consistent expression of clinical concepts. Today we don't have a consistently applied computable definition of angina, nor acute hepatitis B, nor type 2 diabetes. Not nationally, not regionally and not even within the same practice using the same EHR. These are not obscure concepts. These are examples of conditions that are directly related to the trillions spent on healthcare each year.

To reduce costs, we need better data to enable an improved understanding of performance, quality and comparative efficacy. There are many groups that have worked on mapping concepts to codes. The Meaningful Use working group can go a long way by establishing the consistent uses of concepts/codes and distributed population queries.

We need to shrink the time it takes to produce observational studies and introduce flexibility in the questions that are asked in response to a better understanding of clinical questions. As h1N1 became better understood two years ago and as an e-coli outbreak this year became better understood, it was important to be able to refine the questions being asked in near-time to understand more about each situation.

Most Americans trust that the clinicians have the data needed to make informed decisions about each patient’s care. The public also trusts that clinicians have the data to ensure that there are appropriate proactive responses to protect population health. Yet there is so much more that we could know about the nation’s health, treatment and outcomes.

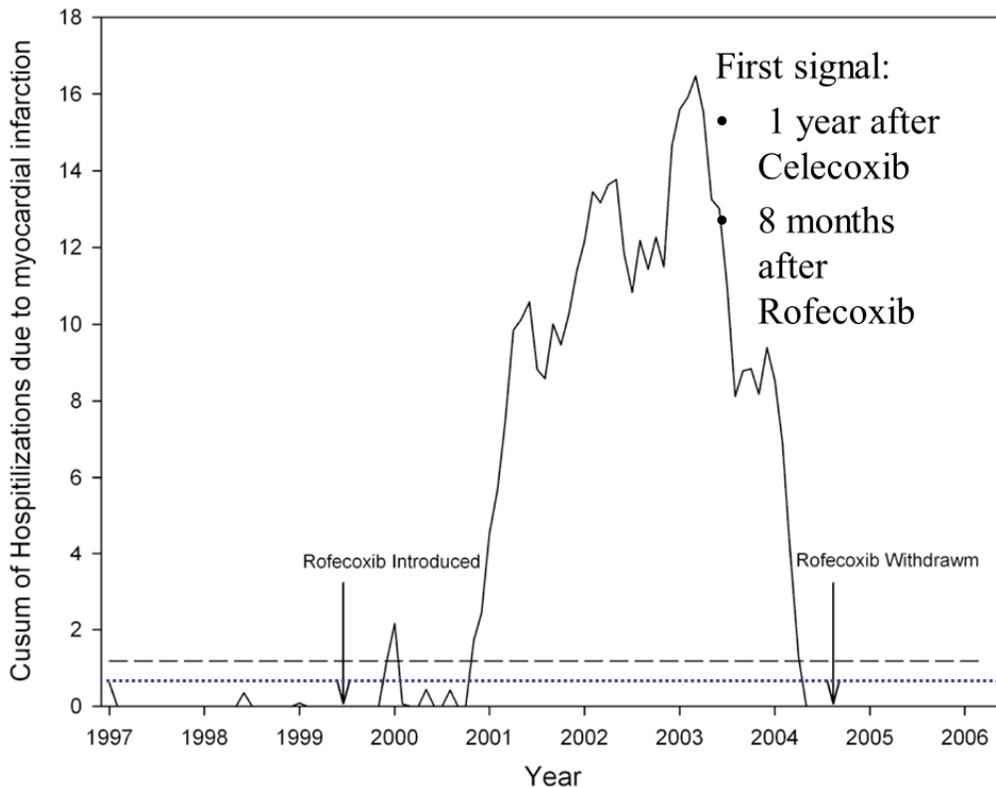


Figure 1: From Bedside to Bench and Back, Zak Kohane, https://www.i2b2.org/events/slides/i2b2_AUG_201106_Kohane.pptx

Observational studies enabled through health IT could save many lives. For example, population health monitoring “as an adjunct to pharmacovigilance methods might have helped confirm the suspected association” between MI and rofecoxib “providing earlier support for the market withdrawal of rofecoxib” (see [The tell-tale heart: population-based surveillance reveals an association of rofecoxib and celecoxib with myocardial infarction](#) by Brownstein, Sordo, Kohane and Mandl). Much of this can be accomplished with distributed population queries that return de-identified aggregated data

The PCAST report ([Report to the President: Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward](#)) identified deficits in the national strategy relative to de-identified aggregated near-time access to health data and exchange based on minimal standards for metadata-tagged data elements. In describing the “potential for real-time, real-world, and comprehensive data”, PCAST described the questions that could be answered related to syndromic surveillance and public health monitoring, safety and adverse event monitoring, assessing quality and use of evidence-based approaches and comparative effectiveness research. This list is almost identical to what’s been found in the Summer Concert Series environmental scan.

Panel Question: What approaches or solution alternatives (e.g., standards, architectural approaches, workflow changes, policy changes) would you recommend to make the acquisition, analysis, and use of health data more effective and efficient from the perspective you represent?

1. Establish the use of consistent computable expressions of clinical concepts. Ensuring extensibility of mappings of concepts and codes in response to new knowledge about diseases and treatments.
2. Include standard use of distributed population queries. Today, there is no standard use of distributed population queries so the plumbing costs are extremely high.
3. Include the standard use of metadata tagged data elements to:
 - clearly specify patient consent coding
 - establish semantic taxonomies in tagged data elements
 - enable submission of meaningful use clinical measures calculated from computable data
 - encourage or require that quality measures under all of the CMS reporting programs can be collected in a tagged data element model.

Should the Meaningful Use Work Group decide to include distributed population queries in Stage 3, there are a few considerations related to implementation.

1. The implementation and deployment by health IT vendors and in health care organizations will take time, both technically and in establishing appropriate processes, controls and data use arrangements.
2. The Meaningful Use Work Group should consider perhaps one national priority for initial application of the technology, ideally one that can return benefits one community or region at a time.

3. While aggregated health information reporting to CMS is one application, the approach should enable a variety of uses, not all top-down and not all government related. The approach should enable multiple collaborative distributed query networks, with voluntary participation through data use arrangements, no central planning and data disclosure fully under the control of the data stewards.