Distributed Networking

Millions of people. Strong collaborations. Privacy first.

Jeffrey Brown, Lesley Curtis, Richard Platt
Harvard Pilgrim Health Care Institute and Harvard Medical School
Duke Medical School

March 15, 2013
The goal

• Facilitate multi-site research collaborations between investigators and data stewards by creating secure networking capabilities and analysis tools
**Not the goal**

We will **not** create a new stand-alone network with its own research agenda or content experts.

Investigators will **not** have access to data without data stewards’ active engagement.
Reminder: Mini-Sentinel’s foundation

- Strong collaborations between investigators and data partners
  - Creation of a community of trust with shared goals, backed by clear governance policies
  - Data partners’ participation as collaborators
  - Data partners’ voluntary participation on a case-by-case basis
Developing the Sentinel System — A National Resource for Evidence Development

Rachel E. Behrman, M.D., M.P.H., Joshua S. Benner, Pharm.D., Sc.D., Jeffrey S. Brown, Ph.D.,
Mark McClellan, M.D., Ph.D., Janet Woodcock, M.D., and Richard Platt, M.D.

The Food and Drug Administration (FDA) now has the capacity to “query” the electronic health information of more than 60 million people, posing specific questions in order to monitor the safety of approved medical products. This information to answer additional convening an ongoing series of discussions among stakeholders to address the near- and long-term challenges inherent in implementing the Sentinel System. In 2009, the FDA gave the Harvard Pilgrim Health Care Institute the lead role...
Use case: Assess disease burden/outcomes

- An NIDDK program officer wants to characterize the use and outcomes of insulin pumps for diabetes
- The Collaboratory networking center uses pre-existing ("canned") programs to query electronic data from millions of people to assess:
  - Frequency of use
  - Characteristics of the users (age, sex, prior treatment history)
  - Frequency of selected outcomes before and after initiation of use
Use case: Pragmatic clinical trial design

- Investigators planning a multi-center pragmatic trial of stroke prevention regimens want to assess the feasibility of embedding a clinical trial in care settings.
- The Collaboratory networking center queries electronic health data to:
  - Assess baseline hospitalization rate with a stroke diagnosis
  - Identify organizations with enough potential study participants
  - Identify potential study participants – all identifiable information stays with the host organization
Use case: Pragmatic clinical trial follow up

- Investigators conducting a multi-center pragmatic trial of stroke prevention regimens want to simplify follow up.
- The Collaboratory networking center supports clinical organizations’ periodic scans of their electronic data covering study participants to identify:
  - Dispensing of prescription medications, including dates, names, and amounts dispensed.
  - All inpatient and ambulatory medical encounters, with dates and diagnoses and procedures.
Use case: Reuse of research data

- A clinically rich research dataset of patients with incident hypertension contains longitudinal records of all blood pressure measurements, BMI, medical utilization, diagnoses, treatments, and laboratory test results.
- The data steward uses the Collaboratory’s networking capability to allow an investigator at another organization to submit analytic programs.
- The output does not contain direct identifiers.
Use case: Single study private network

• A multi-center pragmatic trial team wants to create a pooled final analysis data file
• The Collaboratory networking center establishes a private distributed network
  • To distribute programs that create separate analysis files at each site
  • To securely transfer the analysis files to the analyst
Benefits

• Assessing disease burden
  • New capability, speed, low cost, privacy protection
• Trial design / follow-up
  • New capability, speed, low cost, privacy protection
• Reuse of data
  • HIPAA compliance
    • Avoids need to create limited or de-identified datasets
    • In some cases, full datasets are more useful
  • Data sharing
    • Avoids need for some data use or business associate agreements
    • Preserves clinical organizations’ sharing restrictions
• Private network
  • Secure access, auditable procedures
NIH Distributed Networking Coordinating Center

- Leverages existing networks’ data and analysis tools
  - Can use many data types, e.g., EHR, claims, registries
  - Can use many data models, e.g., Mini-Sentinel, i2b2, OMOP
  - Can use existing querying tools, e.g., Mini-Sentinel modular programs
- Every use requires the agreement of the data steward
What is a distributed research network?

1- User creates and submits query (a computer program)
2- Data stewards retrieve query
3- Data stewards review and run query against their local data
4- Data stewards review results
5- Data stewards return results via secure network
6- Results are aggregated
Mini-Sentinel’s Common Data Model

<table>
<thead>
<tr>
<th>Enrollment</th>
<th>Demographic</th>
<th>Dispensing</th>
<th>Encounter</th>
<th>Lab Result</th>
<th>Vital Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
</tr>
<tr>
<td></td>
<td>Birth date</td>
<td>Dispensing date</td>
<td>Dates of service</td>
<td>Dates of order, collection &amp; result</td>
<td>Date &amp; time of measurement</td>
</tr>
<tr>
<td></td>
<td>Sex</td>
<td>National drug code (NDC)</td>
<td>Provider seen</td>
<td>Test type, immediacy &amp; location</td>
<td>Height</td>
</tr>
<tr>
<td></td>
<td>Race</td>
<td>Days supply</td>
<td>Type of encounter</td>
<td>Procedure code &amp; type</td>
<td>Weight</td>
</tr>
<tr>
<td></td>
<td>Etc.</td>
<td>Amount dispensed</td>
<td>Facility</td>
<td>Test result &amp; unit</td>
<td>Diastolic &amp; systolic BP</td>
</tr>
<tr>
<td></td>
<td>Etc.</td>
<td>Etc.</td>
<td>Etc.</td>
<td>Abnormal result indicator</td>
<td>Tobacco use &amp; type</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Death</th>
<th>Cause of Death</th>
<th>Diagnosis</th>
<th>Procedure</th>
<th>Etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
<td>Person ID</td>
</tr>
<tr>
<td>Date of death</td>
<td>Cause of death</td>
<td>Date</td>
<td>Dates of service</td>
<td>Date &amp; time of measurement</td>
</tr>
<tr>
<td>Source</td>
<td>Diagnosis code &amp; code type</td>
<td>Principle diagnosis flag</td>
<td>Procedure code &amp; type</td>
<td>Height</td>
</tr>
<tr>
<td>Confidence</td>
<td>Source</td>
<td>Encounter type &amp; provider</td>
<td>Encounter type &amp; provider</td>
<td>Weight</td>
</tr>
<tr>
<td>Etc.</td>
<td>Confidence</td>
<td>Etc.</td>
<td>Etc.</td>
<td>Diastolic &amp; systolic BP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Vital Signs</th>
<th>Etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person ID</td>
<td>Person ID</td>
</tr>
<tr>
<td>Date &amp; time of measurement</td>
<td>Date &amp; time of measurement</td>
</tr>
<tr>
<td>Height</td>
<td>Height</td>
</tr>
<tr>
<td>Weight</td>
<td>Weight</td>
</tr>
<tr>
<td>Diastolic &amp; systolic BP</td>
<td>Diastolic &amp; systolic BP</td>
</tr>
<tr>
<td>Tobacco use &amp; type</td>
<td>Tobacco use &amp; type</td>
</tr>
<tr>
<td>BP type &amp; position</td>
<td>BP type &amp; position</td>
</tr>
</tbody>
</table>

Mini-Sentinel’s Common Data Model

info@mini-sentinel.org
Mini-Sentinel’s distributed dataset data checks

- ~400 data checks per refresh
- 100+ tables per data partner per refresh
Ready to use tools for common data model

Data Activities

Mini-Sentinel uses a distributed data approach in which Data Partners maintain physical and operational control over electronic data in their existing environments. The Mini-Sentinel Common Data Model standardizes administrative and clinical information across Data Partners. Data Partners execute, within their own institutions’ firewalls, standardized computer programs (e.g., modular programs) provided by the Operations Center or project workgroups. Data Partners then share the output of these programs with the Operations Center and project workgroups, typically in aggregated form.

A key benefit of the distributed approach is that it minimizes the need to share identifiable patient information. Additionally, each health care data system has unique characteristics, and use of a distributed system better enables the Data Partner’s involvement in running analyses to ensure an informed approach to interpreting results.

Mini-Sentinel data activities fall into the following general categories. Additional information can be found by clicking on the link to each section.

- Distributed Database and Common Data Model
- Distributed Query Tool & Summary Tables
- Modular Programs
- Toolkit Library
- Complementary Data Sources
# Current Networks

<table>
<thead>
<tr>
<th>Data Steward</th>
<th>AHRQ</th>
<th>FDA</th>
<th>ONC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPAN</td>
<td>PEAL</td>
<td>Mini-Sentinel</td>
</tr>
<tr>
<td>HMO Research Network (# sites in each network)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>(11)</td>
<td>(4)</td>
<td>(13)</td>
</tr>
<tr>
<td>Vanderbilt</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aetna</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humana</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optum (United Healthcare)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WellPoint (HealthCore)</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massachusetts League of Community Health Centers</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AtriusHealth</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beth Israel Deaconess Medical Center</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Distributed Data / Distributed Analysis

• Data stewards keep and analyze their own data
• Standardize the data using a common data model
• Distribute code to stewards for local execution
• Provide results, not data, to requestor
• All activities audited and secure
PMN Software – Supports multiple deployment models
- Agnostic to data center infrastructure and complements existing network infrastructure
- VM based deployments enabling ease of disaster recovery and planning
- Seamless overlay of VPN Connections (Remote Access, Site to Site, Two Factor User Authentication)
- Supports consolidation of remote sites into the data center for central management (Data Steward Components can be hosted in a central data center similar to the PMN Portal)
- Secure End to End connection (Encrypted Transport using X.509 certificates)
- Supports industry standard RBAC configuration for users
- Supports Data Source provisioning based on RBAC and additional data source specific metadata
- Queries distributed using a PULL model instead of PUSH model
Design Features

- Any data model from any source
- Flexible and secure distributed querying
  - Execution of custom analytic code
  - Menu-driven queries
- Role-based access control
- Data steward autonomy
- Query execution options range from fully automated to manual
- Auditing
- Software-enabled governance
Implementation Features

- Secure, private multi-center research network
- Open source application
- Data stewards maintain control of their data
- Flexible governance, access control, permissions, auditing
- Mature documentation and set-up procedures
- Scalable: easy to add new data, new partners
- Interoperable with other networks using same networking platform (PopMedNet)
Security Features

• FISMA compliant tier III data center
• 3rd-party secure audit completed
• Passed multiple independent security audits and penetration tests
National Standards

• The networking platform (PopMedNet) is a key component of the ONC’s QueryHealth Initiative

• ONC national standard for distributed querying
  • QueryHealth Initiative uses PMN as the distributed querying platform for policy and governance

• Standards & Interoperability (S&I) Framework: http://wiki.siframework.org/Home
Governance (proposed)

- Data stewards retain control of their data
- All activities are opt-in
- Data stewards can choose to be full partners in the design and implementation of research
- Data steward costs must be reimbursed
  - Includes amortizing cost of maintaining data in query-able form
- TBD: A board of representatives to engage NIH leadership
Operations

- Each data steward designates a single contact for new queries
- Each data steward uses its own process for deciding whether to participate in any activity
Fine print

• Current resources will support ~20 sites
• Using existing data resources is fast; developing new ones is slow
  • Most current resources have extensive claims data, and limited EHR data
• Using existing analysis tools is fast; developing new ones is slow
• Ability to query multiple sites requires
  • Each site’s data to be in the same format
  • Consistent definitions of variables
Timeline

• General querying capability begins July 2013 for organizations participating in existing networks
Thank you!