Design of a Distributed Data Network for Comparative Effectiveness Research

Jeffrey Brown, PhD
Assistant Professor
Department of Population Medicine
Harvard Pilgrim Health Care Institute / Harvard Medical School

HMORN Conference 2010
Austin, TX

March 23, 2010
Developing a Distributed Research Network

**Partners:**
DEcIDE center at the HMO Research Network Center for Education and Research on Therapeutics

DEcIDE center at the University of Pennsylvania

LincolnPeak Partners

**Participating Health Plans:** Geisinger Health System, Group Health Cooperative, Harvard Pilgrim Health Care, HealthPartners, Kaiser Permanente Colorado, and Kaiser Permanente Northern California
Outline of Presentation

- Background
- Design of a network
- Example
- Planned enhancements
- Joining the network
HMORN Distributed Model

- Standardize data
- Data holders maintain physical control of their data
- Data holders control all uses of their data
- Data holders control all transfer of data
- Computer programs should run at multiple sites without modification
Workflow of Typical HMORN Multi-site Query

- SAS program specified, written and tested
- SAS program distributed via e-mail or shared portal (e.g., collaboration website)
- SAS program executed locally, results reviewed and approved for transfer
- Results returned to requestor as appropriate for data (secure e-mail, secure FTP, certified mail)

It can take a long time to get simple counts; multiple queries or revisions are a burden
Project Overview

• Developing a set of tools and procedures to improve efficiency of our typical workflow
  – Give investigators and analysts tools to submit data requests
  – Give site administrators tools to speed the review and response to requests
  – Define (and implement) different levels of access that can permit different levels of authorization to query data
  – Audit trail of all activity
Use Cases Used in Design

• Simple menu-driven querying (summary tables)
  – Counts, rates, usage patterns
• Complex menu-driven querying (patient-level tables)
  – Monitoring and surveillance
  – Comparative evaluations
• Distribution of SAS code
  – All else, including distributed regression
DRN Network Design & Prototype

Access Control Manager
Query Manager
Results Manager
Workflow Manager
User Manager

Audit Manager
Researcher User Interface
Data Mart Admin User Interface
System Administrator User Interface

DRN Hub API

Message Protocol

Hub Database

Portal (hub)

Data Mart Security Manager
Query Execution Manager
Data Source Manager

Data Mart Audit Manager

Data Mart API

Data Mart Database

Data Holder DataMart
Example
Investigator: Login

INSTRUCTIONS: Please complete the form below and read the Terms and Conditions before using the Query Tool.

Username

Password

☐ I agree to the Terms and Conditions of use.

Start Query Tool
Investigator: Query Type

- Pharmacy Dispensings by Generic Name
- Pharmacy Dispensings by Drug Class
- Dispensings by National Drug Code
- ICD-9 Diagnoses
- ICD-9 Procedures
- HCPCS Procedures
- Eligibility and Enrollment
**Investigator: Build Query**

**Sort codes by**: [ ]

Please select one or more **Code**:

- [x] 401 ESSENTIAL HYPERTENSION
- [ ] 402 HYPERTENSIVE HEART DISEASE
- [ ] 403 HYPERTENSIVE RENAL DISEASE
- [ ] 404 HYPERTENSIVE HEART AND RENAL DISEASE
- [x] 405 SECONDARY HYPERTENSION
- [x] 410 ACUTE MYOCARDIAL INFARCTION
- [ ] 411 OTH ACUTE&SUBACUTE FORMS ISCHEMIC HEART DISEASE
- [ ] 412 OLD MYOCARDIAL INFARCTION

Please select one or more **Age Group**:

- [x] 0-4
- [x] 5-9
- [x] 10-14
- [x] 15-19
- [ ] 20-44
- [ ] 45-64
- [ ] 65-74
- [x] 75+

Please select one or more **Gender**:

- [x] Male
- [x] Female

Please select one or more **Periods**:

- [x] 2000
- [x] 2001
- [x] 2002
- [x] 2003
- [x] 2004
- [ ] 2005
- [ ] 2006
- [ ] 2007
- [ ] 2008

Please select a **Setting**:

- [ ] Outpatient

---

12
Investigator: Distribute

Please select at least two Data Marts to which this query will be sent *

Note: Click a Data Mart name to view details (Metadata)

- [x] LPeak-auto1
- [x] LPeak-auto2
- [x] HPHCI
- [x] GHC
- [ ] Geisinger
- [x] KPCO
- [ ] HPRF
- [ ] DARTNet

Start This Query
Data Holder: Review Query

Query Details:
- Query Id: 443
- Query Name: Inpatient AMIs version2
- Status: Completed
- Query Description: For HMORN test
- Query Text:
  ```sql
  SELECT age_group, gender as Sex, period, code, dname, SETTING, EVENTS, Members FROM ICD9_diagnosis WHERE code IN ('410') AND period IN ('2007', '2008', '2017') AND SETTING IN ('INP')
  ```

Results:
Data Holder: Review Results

DrN Data Mart Client - Query Detail

- Query Id: 443
- Request Time: 03/18
- Query Name: Inpatient AMIs version2
- Submitted By: Jeff
- Status: Submitted
- Query Description: For HMORN test
- Query Text:

```
SELECT age_group, gender, sex, period, code, dname, setting, events, members
FROM inpatient
WHERE age_group IN ('0-4', '10-14', '15-19', '20-44', '45-64', '65-74', '75+)
AND period IN ('2007', '2008', '2017')
AND setting IN ('INP')
```

DrN Data Mart Client - View Results

<table>
<thead>
<tr>
<th>age_group</th>
<th>Sex</th>
<th>period</th>
<th>code</th>
<th>dname</th>
<th>SETTING</th>
<th>EVENTS</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10-14</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15-19</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>20-44</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>157</td>
<td>46</td>
</tr>
<tr>
<td>20-44</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>339</td>
<td>107</td>
</tr>
<tr>
<td>45-64</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>914</td>
<td>284</td>
</tr>
<tr>
<td>45-64</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>2500</td>
<td>628</td>
</tr>
<tr>
<td>65-74</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>187</td>
<td>69</td>
</tr>
<tr>
<td>65-74</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>454</td>
<td>123</td>
</tr>
<tr>
<td>75+</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDI...</td>
<td>INP</td>
<td>243</td>
<td>74</td>
</tr>
</tbody>
</table>

Upload Results
Investigator: Results Sets

Mar 19, 2010

**VIEW A QUERY:** This page allows you to look at individual query **Inpatient AMIs version2 ID 443** of type **ICD-9 Diagnoses** submitted 3/18/2010 3:56:37 PM.

The first (top) table shows the status of this query against various DATA MARTS. Requests are either Complete, Failed or In Progress. To see the results of a completed request, select (single click) on it and its details will appear in the second (bottom) table. You can also filter and sort either table by column where appropriate.

<table>
<thead>
<tr>
<th>Data Mart</th>
<th>Last Response</th>
<th>Message</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPeak-auto1</td>
<td>Completed</td>
<td>3/18/2010 3:56:40 PM</td>
</tr>
<tr>
<td>LPeak-auto2</td>
<td>Completed</td>
<td>3/18/2010 3:56:40 PM</td>
</tr>
<tr>
<td>HPHCI</td>
<td>Completed</td>
<td>3/18/2010 5:12:50 PM</td>
</tr>
</tbody>
</table>
# Investigator: View Results

<table>
<thead>
<tr>
<th>age_group</th>
<th>Sex</th>
<th>period</th>
<th>code</th>
<th>dxname</th>
<th>SETTING</th>
<th>EVENTS</th>
<th>Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0-4</td>
<td>F</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>0-4</td>
<td>M</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>10-14</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15-19</td>
<td>F</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>32</td>
<td>10</td>
</tr>
<tr>
<td>15-19</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>15-19</td>
<td>M</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>40</td>
<td>16</td>
</tr>
<tr>
<td>20-44</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>157</td>
<td>46</td>
</tr>
<tr>
<td>20-44</td>
<td>F</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>208</td>
<td>22</td>
</tr>
<tr>
<td>20-44</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>339</td>
<td>107</td>
</tr>
<tr>
<td>20-44</td>
<td>M</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>766</td>
<td>138</td>
</tr>
<tr>
<td>45-64</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>914</td>
<td>284</td>
</tr>
<tr>
<td>45-64</td>
<td>F</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>2120</td>
<td>642</td>
</tr>
<tr>
<td>45-64</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>2500</td>
<td>628</td>
</tr>
<tr>
<td>45-64</td>
<td>M</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>5182</td>
<td>1248</td>
</tr>
<tr>
<td>65-74</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>187</td>
<td>69</td>
</tr>
<tr>
<td>65-74</td>
<td>F</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>1514</td>
<td>400</td>
</tr>
<tr>
<td>65-74</td>
<td>M</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>454</td>
<td>123</td>
</tr>
<tr>
<td>65-74</td>
<td>M</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>2270</td>
<td>630</td>
</tr>
<tr>
<td>75+</td>
<td>F</td>
<td>2007</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>243</td>
<td>74</td>
</tr>
<tr>
<td>75+</td>
<td>F</td>
<td>2017</td>
<td>410</td>
<td>ACUTE MYOCARDIAL INFARCTION</td>
<td>INP</td>
<td>2528</td>
<td>724</td>
</tr>
</tbody>
</table>
Ongoing Work

- Enhancing the security features
- Adding the ability to securely transfer files
- General usability enhancements to the application and portal
- Creating a public website
- Secure hosting and go-live
- Prototype illustrating submission and execution of SAS code against local SAS dataset
Why This Approach?

- Practical approach with our health plans’ social, regulatory, and business environment
  - Lowers barriers to acceptance and implementation
  - Small IT footprint and limited risk
  - Focus on things we do well: data manipulation
  - Minimize need for extensive database expertise & ongoing maintenance of complex data structures
- Allows automation of any step via roll based access control
  - Require manual execution if submitted by a, b, or c
  - Allow automated execution of all queries from x, and y
    » Unless topic is mental health
Joining the DRN Network

• Actively seeking additional data partners
• Steps for joining the network as a data mart:
  1. Contact HPHCI for access to the portal and network documentation
  2. Create summary table database in accepted format
  3. Download, install, and setup the desktop client
  4. Log into portal to set DataMart preferences
• DRN Governance Committee currently refining access policies and terms of use
• 7-8am Demonstration/ Q and A (Hill Country A)
Design of a National Distributed Health Data Network

Judith C. Maro, MS; Richard Platt, MD, MSc; John H. Holmes, PhD; Brian L. Strom, MD, MPH; Sean Hennessy, PharmD, PhD; Ross Lazarus, MBBS, MPH; and Jeffrey S. Brown, PhD

A distributed health data network is a system that allows secure remote analysis of separate data sets, each comprising a different medical organization’s or health plan’s records. Distributed health data networks are currently being planned that could cover millions of people, permitting studies of comparative clinical effectiveness, best practices, diffusion of medical technologies, and quality of care. These networks could also support assessment of medical product safety and other public health needs. Distributed network technologies allow data holders to control all uses of their data, which overcomes many practical obstacles related to confidentiality, regulation, and proprietary interests. Some of the challenges and potential methods of operation of a multipurpose, multi-institutional, distributed health data network are described.


For author affiliations, see end of text.
This article was published at www.annals.org on 28 July 2009.
Note

These slides are based on research conducted by the DEcIDE centers at the HMO Research Network Center for Education and Research on Therapeutics and the University of Pennsylvania under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HHSA29020050033I). The findings and conclusions in this document are those of the authors who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this document should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

Financial disclosure: None of the authors have a financial interest in any information discussed in these slides.
Design of a Distributed Data Network for Comparative Effectiveness Research

Jeffrey Brown, PhD
Assistant Professor
Department of Population Medicine
Harvard Pilgrim Health Care Institute / Harvard Medical School

HMORN Conference 2010
Austin, TX

March 23, 2010