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

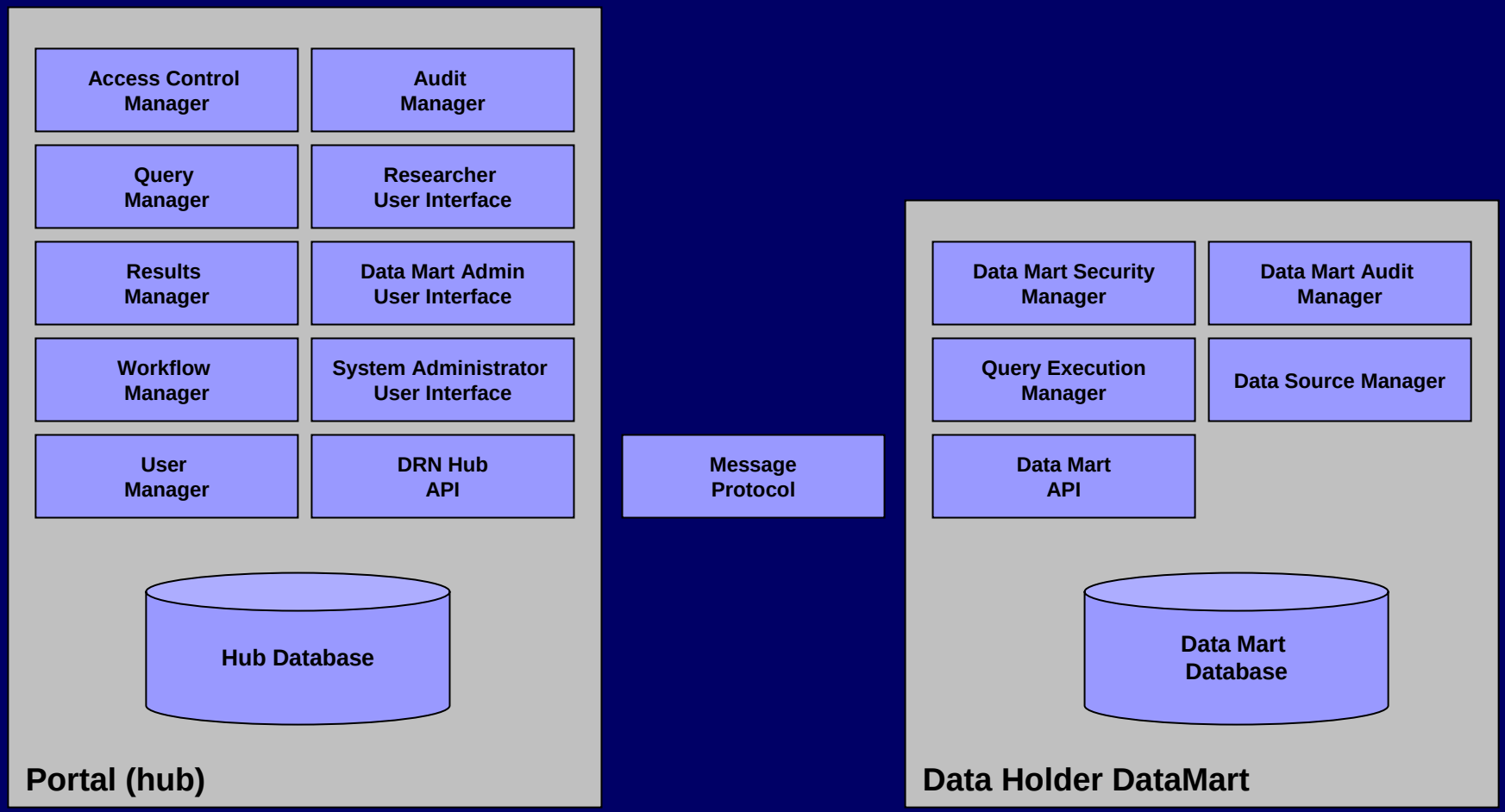
Project Overview

- 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

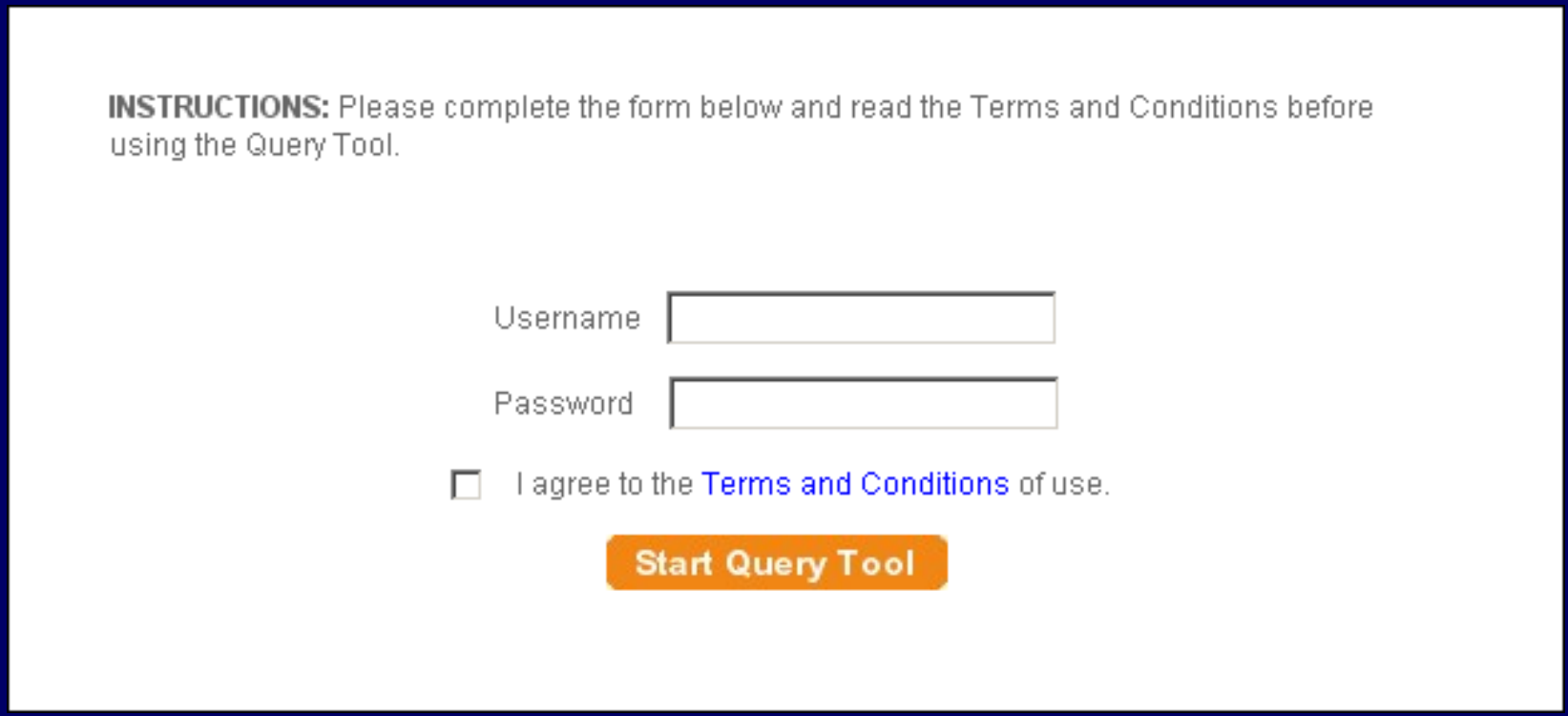
- Simple menu-driven querying (summary tables)
  - Counts, rates, usage patterns
- Complex menu-driven querying (patient-level tables)
  - Monitoring and surveillance
  - Comparative evaluations
- Distribution of SAS code
  - All else, including distributed regression

DRN Network Design & Prototype

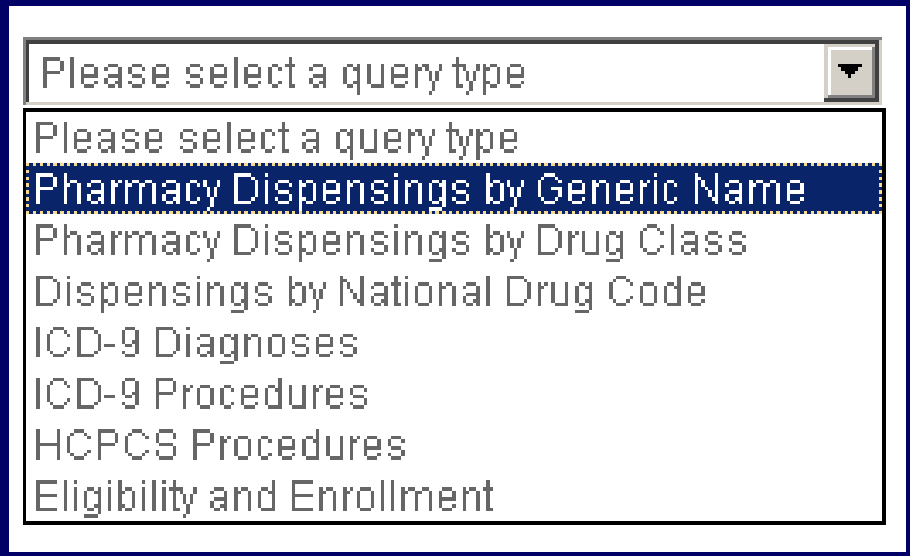


Example

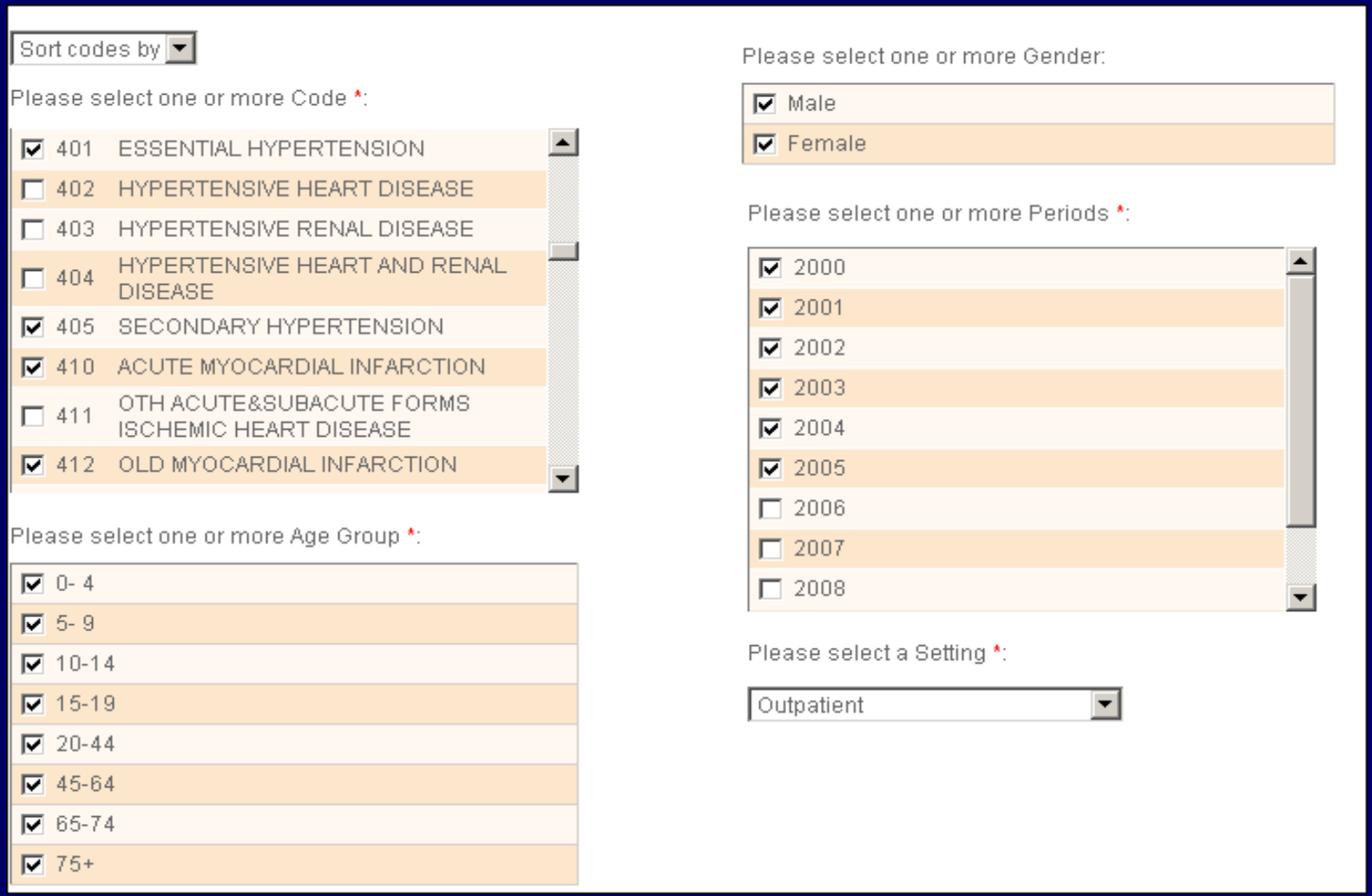
Investigator: Login



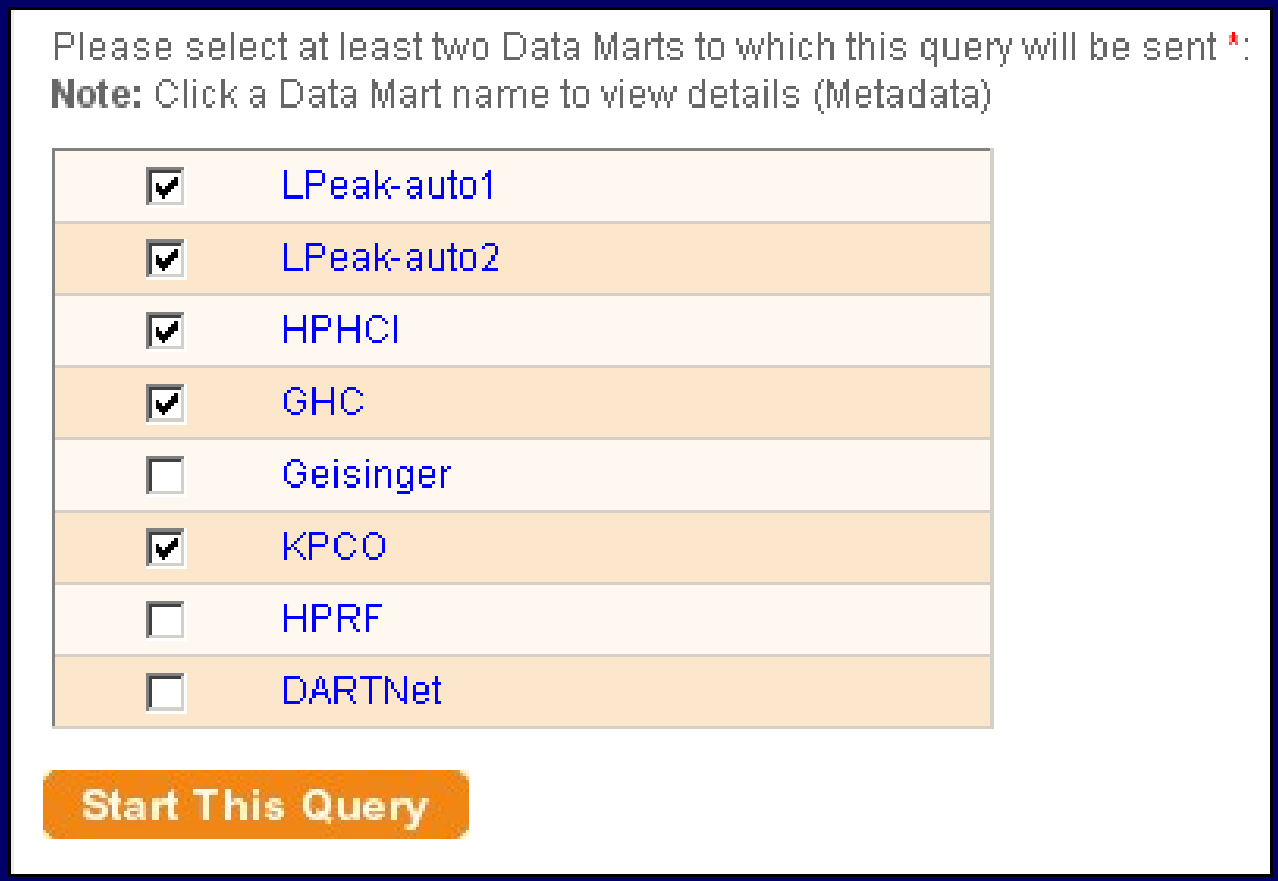
Investigator: Query Type



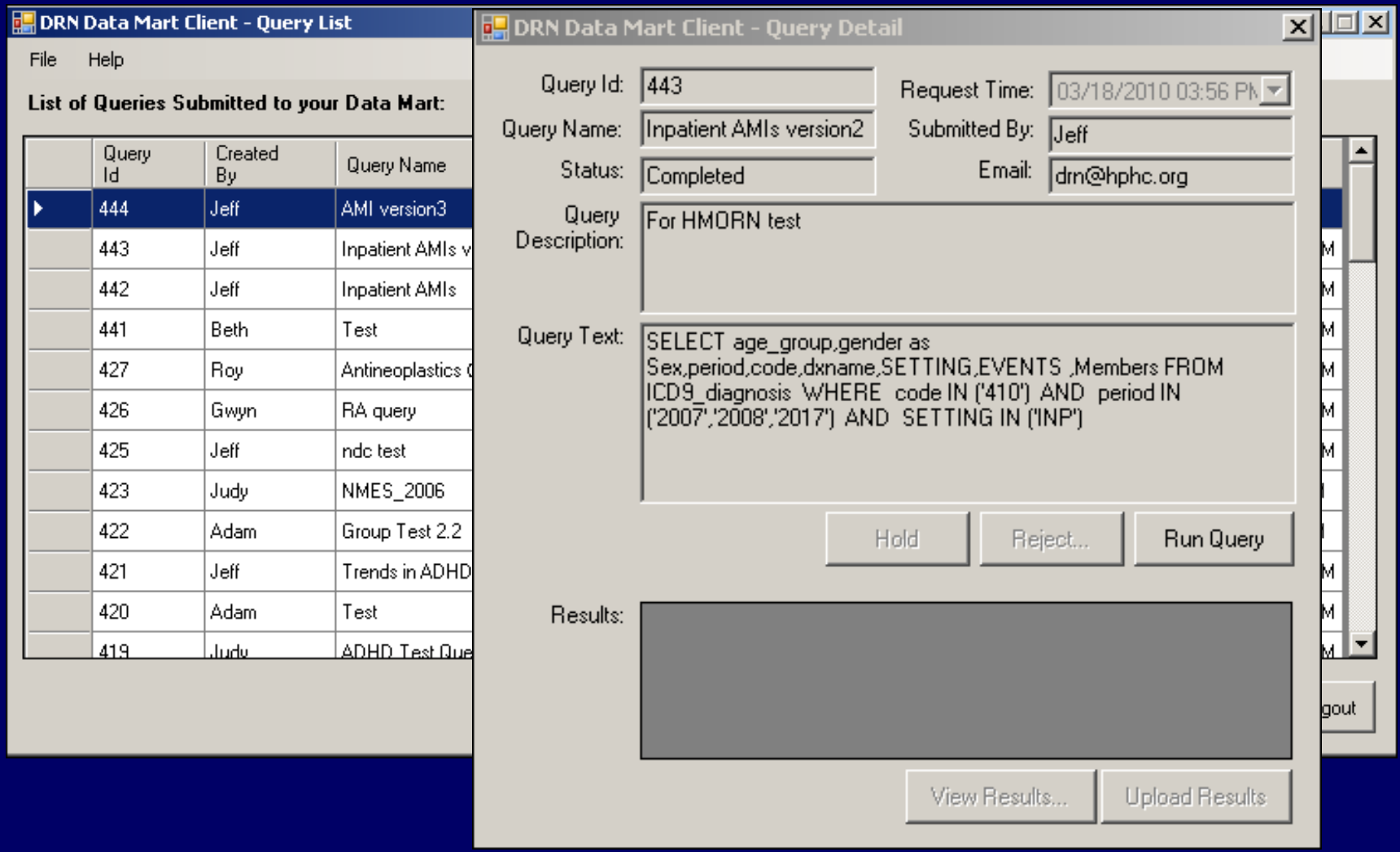
Investigator: Build Query



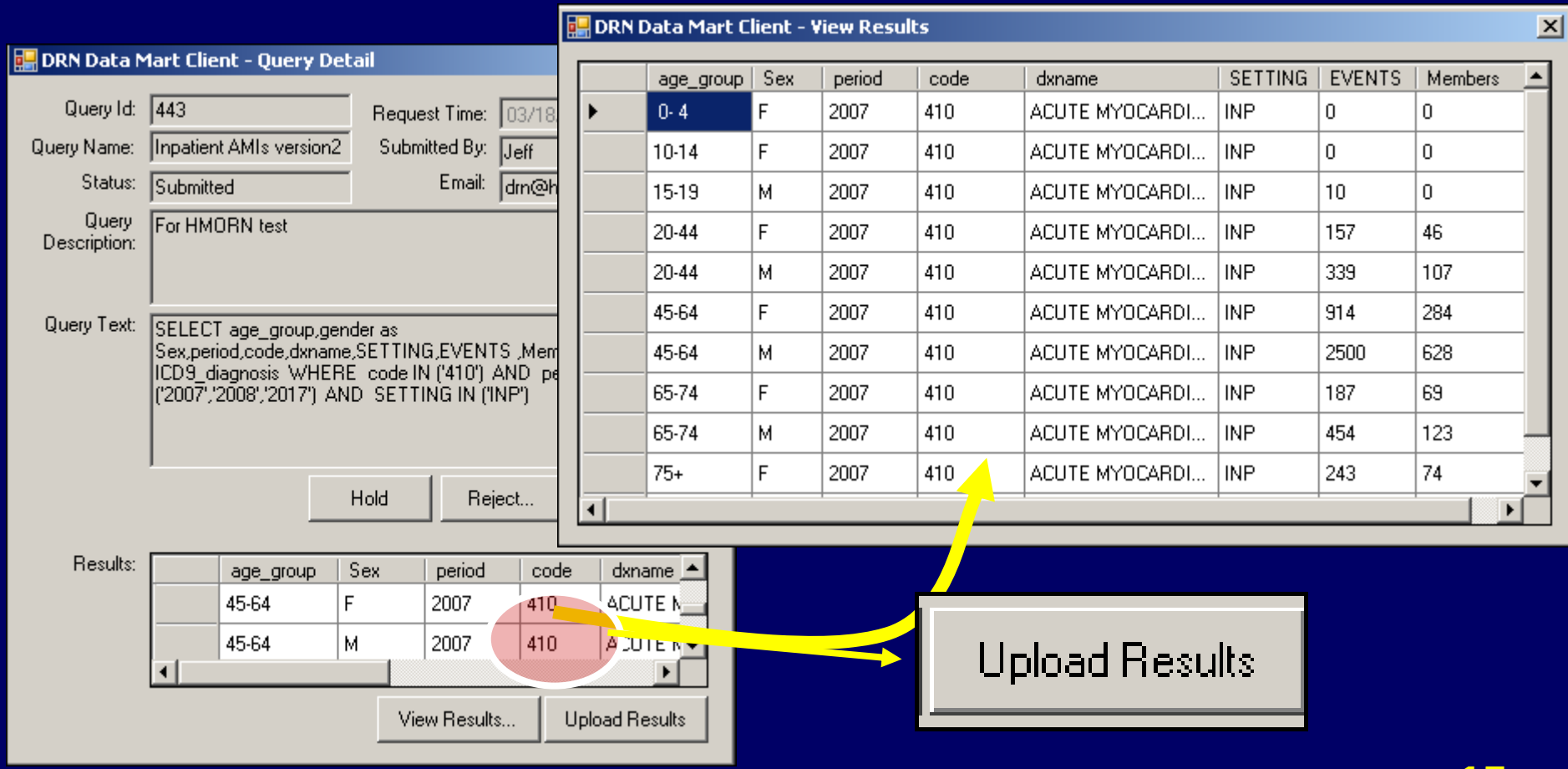
Investigator: Distribute



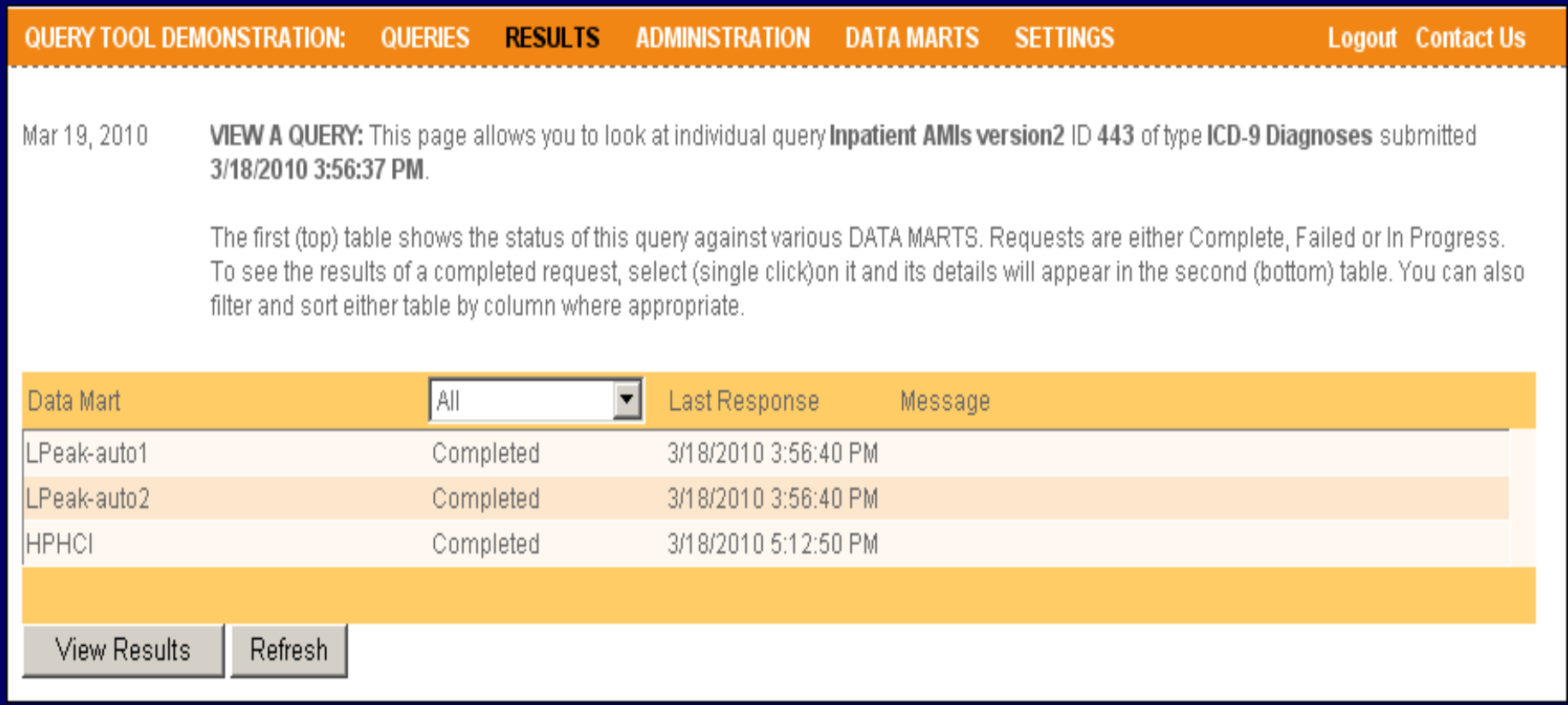
Data Holder: Review Query



Data Holder: Review Results



Investigator: Results Sets



Investigator: View Results

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

CSV Version

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)

Annals of Internal Medicine | ACADEMIA AND CLINIC

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.

Ann Intern Med. 2009;151. www.annals.org
For author affiliations, see end of text.
This article was published at www.annals.org on 28 July 2009.

Confidential Do Not Quote Medical Care 2010 (in press)

Title: Distributed health data networks: a practical and preferred approach to multi-institutional evaluations of comparative effectiveness, safety, and quality of care

Authors: Jeffrey S. Brown, PhD[a]; John H. Holmes, PhD[b]; Kiran Shah, BA[c]; Ken Hall, [MDIV][d]; Ross Lazarus, MBBS, MPH[e]; Richard Platt, MD, MSc[f]

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