Using the NIH Collaboratory's and PCORnet's distributed data networks for clinical trials and observational research - A preview

Millions of people. Strong collaborations. Privacy first.

Jeffrey Brown, PhD, Lesley Curtis, PhD, Richard Platt, MD, MS

Harvard Pilgrim Health Care Institute and Harvard Medical School
Duke University
November 14, 2014
The Collaboratory DRN’s goal

Facilitate multi-site research collaborations between investigators and data partners by creating secure networking capabilities and analysis tools for electronic health data
PCORnet’s goal

Improve the nation’s capacity to conduct rapid, efficient, and economical comparative effectiveness research
Three critical elements

• Privacy protections
• Reusable analysis tools
• Analysis-ready data
Three critical elements

• Privacy protections
• Reusable analysis tools
• Analysis-ready data
Distributed analysis

1. User creates and submits query (a computer program)
2. Individual data partners retrieve query
3. Data partners review and run query against their local data
4. Data partners review results
5. Data partners return results via secure network
6. Results are aggregated
Multiple networks sharing infrastructure

- Each organization can participate in multiple networks
- Each network controls its governance and coordination
- Other networks can participate
- Networks share infrastructure, data curation, analytics, lessons, security, software development
Each organization can participate in multiple networks
Each network controls its governance and coordination
Other networks can participate
Networks share infrastructure, data curation, analytics, lessons, security, software development
Use cases

• Pragmatic clinical trial design
• Observational studies
• Single study private network
• Pragmatic clinical trial follow up
• Reuse of research data
Use cases

- **Pragmatic clinical trial design**
- Observational studies
- Single study private network
- Pragmatic clinical trial follow up
- Reuse of research data
MINI-SENTINEL and CLINICAL TRIALS TRANSFORMATION INITIATIVE

DEVELOPING APPROACHES TO CONDUCTING RANDOMIZED TRIALS USING THE MINI-SENTINEL DISTRIBUTED DATABASE

February 28, 2014
Use Case: IMPACT-AF Cluster Randomized Trial

- Proposed by Christopher Granger, MD, and colleagues
- Primary Aim: Test a multilevel educational intervention to increase the rate of initiation of oral anticoagulants among patients with atrial fibrillation.
- Design: Cluster randomized trial
- Intervention:
  - For patients – Mailed educational material, and recommendation to discuss their anticoagulation status with their clinician
  - For physicians – Notification of eligible patients. Reports regarding their eligible patients’ rate of anticoagulation benchmarked to other providers
- Population: Patients >18 years with atrial fibrillation without anticoagulation AND >1 CHADS₂ (congestive heart failure, hypertension, age > 75 yrs, diabetes, stroke or TIA) risk factor OR >2 CHA₂DS₂ VASc (congestive heart failure, hypertension, age, diabetes, stroke or TIA, vascular disease, female) risk factors
Use Case: IMPACT-AF Cluster Randomized Trial

- Proposed by Christopher Granger, MD, and colleagues
- Primary Aim: Test a multilevel educational intervention to increase the rate of initiation of oral anticoagulants among patients with atrial fibrillation.
- Design: Cluster randomized trial
- Intervention:
  - For patients – Mailed educational material, and recommendation to discuss their anticoagulation status with their clinician.
  - For physicians – Notification of eligible patients. Reports regarding their eligible patients’ rate of anticoagulation benchmarked to other providers.
- Population: Patients >18 years with atrial fibrillation without anticoagulation AND
  >1 CHADS₂ (congestive heart failure, hypertension, age >75 yrs, diabetes, stroke or TIA) risk factor OR
  >2 CHA₂DS₂ VASc (congestive heart failure, hypertension, age, diabetes, stroke or TIA, vascular disease, female) risk factors

Use cases

- Pragmatic clinical trial design
- **Observational studies**
- Single study private network
- Pragmatic clinical trial follow up
- Reuse of research data
Online First

Comparative Risk for Angioedema Associated With the Use of Drugs That Target the Renin-Angiotensin-Aldosterone System

Sengwee Toh, ScD; Marsha E. Reichman, PhD; Monika Houstoun, PharmD; Mary Ross Southworth, PharmD; Xiao Ding, PhD; Adrian F. Hernandez, MD; Mark Levenson, PhD; Lingling Li, PhD; Carolyn McCloskey, MD, MPH; Azadeh Shoaiabi, MS, MHS; Eileen Wu, PharmD; Gwen Zornberg, MD, MS, ScD; Sean Hennessy, PharmD, PhD

• Used data for 3.9 million new users of anti-hypertensives in 18 organizations
• Propensity score matched analysis
• **No** person-level data was shared
New program development process

1. Draft functional programming specification
2. Review and approve functional specification
3. Draft technical programming specification
4. Review and approve technical programming specification
5. Develop QC plan and test case scenarios
6. Develop programming package (code, documentation)
7. Submit programming package to Managing Programmer for QC
8. Implement QC plan
9 & 10. Track, resolve and close all QC issues
11. Submit final programming package to Managing Programmer
12. Beta-test programming package
13. Review logs and output from each site

Data Partners

Managing Programmer

Principal Programmer

Auditing Programmer
Comparative Risk for Angioedema Associated With the Use of Drugs That Target the Renin-Angiotensin-Aldosterone System

Sengwee Toh, ScD; Marsha E. Reichman, PhD; Monika Houstoun, PharmD; Mary Ross Southworth, PharmD; Xiao Ding, PhD; Adrian F. Hernandez, MD; Mark Levenson, PhD; Lingling Li, PhD; Carolyn McCloskey, MD, MPH; Azadeh Shoaibi, MS, MHS; Eileen Wu, PharmD; Gwen Zornberg, MD, MS, ScD; Sean Hennessy, PharmD, PhD

- Used data for 3.9 million new users of anti-hypertensives in 18 organizations
- Propensity score matched analysis
- No person-level data was shared
- **Five months and $250,000 required for programming and analysis – compared to 1-2 years and $2 million without analysis-ready distributed dataset**
Comprehensive Risk for Angioedema Associated With the Use of Drugs That Target the Renin-Angiotensin-Aldosterone System

Sengwee Toh, ScD; Marsha E. Kuklina, ScD; Monika Houstoun, PharmD; Mary Ross Southworth, PharmD; Xiao Ding, PhD; Adrian F. Hernandez, MD; Bill G. Emonson, PhD; Lingling Li, PhD; Carolyn McCloskey, MD, MPH; Azadeh Shoabi, MS, MHS; Eileen Wu, PhD; Donald R. B. Harnage, MD, MS, ScD; Sean Hennessy, PharmD, PhD

- Used data for 3.9 million new users of anti-hypertensives in 18 organizations
- Propensity score matched analysis
- No person-level data was shared
- Five months and $250,000 required for programming and analysis – compared to 1-2 years and $2 million without analysis-ready distributed dataset

Can we reduce the effort, time, and cost?
Yes
Three critical elements

- Privacy protections
- **Reusable analysis tools**
- Analysis-ready data
Reusable analysis tools

Two levels of querying complexity and analysis

- **Level 1**: Identify and characterize cohorts (e.g., treatments, outcomes, etc)
- **Level 2**: Comparative analyses with analytic adjustment for confounding using available analytic adjustment tools (e.g., propensity score matching)
Cohort Identification and Descriptive Analysis Tool

- Parameterized program “template” to identify cohorts based on an array of available parameter options
  - Exposure, outcome, inclusion/exclusion criteria, covariate definitions; incidence assessment, age range and groups
- Sample uses
  - Background rates
  - Exposures and follow-up (outcome rates)
  - Concomitant exposure characterization
- Complex exposure and outcome definitions (“combo tool”)
  - Rhabdomyolysis definition example: inpatient diagnosis of rhabdomyolysis AND creatine kinase (CK) total value > 1,000 U/L in the +/- 14 days
- Generates standard output for reporting and for use by additional tools
Patient A (IMPACT-AF example)

Query parameters

<table>
<thead>
<tr>
<th>Query Period</th>
<th>1/1/2006- 12/31/2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coverage Requirement</td>
<td>Medical and Drug Coverage</td>
</tr>
<tr>
<td>Enrollment Requirement</td>
<td>183 days</td>
</tr>
<tr>
<td>Enrollment Gap</td>
<td>45 days</td>
</tr>
<tr>
<td>Age Groups</td>
<td>18-34, 35-44, 45-64 65-74, 75+</td>
</tr>
</tbody>
</table>
Patient A (IMPACT-AF example)

Two cohort definitions

Atrial Fibrillation diagnosis in any care setting at any time in observation period

Two Atrial Fibrillation diagnosis codes on different days in any care setting at any time in observation period; index is first observation
Patient A (IMPACT-AF example)

Observation time: Identify anticoagulant use at any time after index date
Multiple inclusion/exclusion criteria (n=8)

- At least one CHADS\textsubscript{2} risk factor OR at least two CHA\textsubscript{2}DS\textsubscript{2}-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding
- At least two CHADS\textsubscript{2} risk factors OR at least three CHA\textsubscript{2}DS\textsubscript{2}-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding
- At least one CHADS\textsubscript{2} risk factor, EXCLUDE mechanical prosthetic valve and life-threatening bleeding (only relevant for 75+ group)
- At least two CHADS\textsubscript{2} risk factors OR at least one CHA\textsubscript{2}DS\textsubscript{2}-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding (only relevant for 75+ group)
- At least two CHADS\textsubscript{2} risk factors OR at least two CHA\textsubscript{2}DS\textsubscript{2}-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding
Complete specifications

- 16 different cohorts with different definitions for diagnosis and pre-existing condition requirements
- Once specifications are complete, results available within weeks
Comparative Risk for Angioedema Associated With the Use of Drugs That Target the Renin-Angiotensin-Aldosterone System

Sengwee Toh, ScD; Marsha E. Kowluru, PharmD; Jomika Houstoun, PharmD; Mary Ross Southworth, PharmD; Xiao Ding, PhD; Adrian F. Hernandez, MPH; Laura W. Johnson, PhD; Lingling Li, PhD; Carolyn McCloskey, MD, MPH; Azadeh Shoabibi, MS, MHS; Eileen Wu, PhD; Robert H. Kusniberger, MD, MS, ScD; Sean Hennessy, PharmD, PhD

- Used data for 3.9 million new users of anti-hypertensives in 18 organizations
- Propensity score matched analysis
- No person-level data was shared
- Five months and $250,000 required for programming and analysis – compared to 1-2 years and $2 million without a ready distributed dataset
Output of the “Cohort Identification and Descriptive Analysis Tool (CIDA)” is the input for the propensity score matched tool.

Effect estimation based on exposure propensity-score matched parallel new user cohorts defined using the “CIDA” tool.

Three Propensity Score (PS) estimation options:
- **Predefined**: requesters specify code lists.
- **Empirically identified (through high-dimensional PS)**: empirically selected covariates.
- **Predefined + empirically identified (through high-dimensional PS)**: all predefined and empirically selected covariates included in the model.

Two matching options:
- 1:1; 1:100 variable.

Three caliper options:
- .01, .025, .05.
Propensity Score Matched tool

- High-dimensional propensity score options
  - Ranking algorithm
  - Number of covariates considered by data dimension
  - Number of covariates to select for hdPS model
- Subgroup analysis
  - Using any predefined covariate
- Decile stratification
- Output
  - Diagnostics, effect estimates, confidence intervals
Overview

The Protocol Core and FDA has requested execution of the Cohort Identification and Descriptive Analysis (CIDA) and Propensity Score Matching (PSM) tools to investigate exposure to angiotensin-converting-enzyme (ACE) inhibitors and beta blockers and angioedema events in the Mini-Sentinel Distributed Database (MSDD). To be included in the cohort, members must have had no evidence of a prescription for any ACE inhibitor, beta-blocker, angiotensin receptor blocker (ARB), or aliskiren in the 183 days prior to incident drug use. This package was distributed to seven Data Partners on September 23rd, 2014 and an additional ten data partners on September 30th, 2013. This report includes results from 13 data partners. The query period for this request was January 1st, 2008-September 30th, 2013. Please see Appendix A for a list of NDCs used to define ACE inhibitors and beta blockers in this request. Please see Appendix B for a list of codes used to define the outcomes in this request. Please see Appendix C for monitoring period for this request.

Request ID: to09y05_dev_mpd_wp07_b01, to09y05_dev_mpd_wp07_b02

Requester: Protocol Core Work Group / FDA

Specifications: Program parameter inputs and scenarios

Glossary: List of Terms found in this Report and their Definitions

Monitoring Period: Monitoring Period for this request

Table 1: Table displaying Cohort of New Initiators of ACE Inhibitors and Beta Blockers

Table 2: Table displaying Cohort of New Initiators of ACE Inhibitors and Beta Blockers (Matched Predefined PS, Caliper = 0.025)

Table 3: Table displaying Sequential Estimates for Angioedema Events by Analysis Type, and Histograms of PS distribution by DP (masked)

DP01-DP13 Histograms: Histograms of PS distribution by DP (masked)

Appendix A: Table of National Drug Codes (NDCs) used to Define Exposures in this Request

Appendix B: Table of Diagnosis Codes used to Define Outcomes in this Request

Appendix C: Table of Codes used to Define Pre-Existing Inclusions/Exclusions in this Request

Appendix D: Table of Codes used to Define Covariate Codes in this Request

Notes: Please contact the Mini-Sentinel Operations Center (MSOC_Requests@harvardpilgrim.org) for questions and to provide comments/suggestions for future enhancements to this document.
## Specifications

### Specifications for to09y05_dev_mpd_wp07_b01 and to09y05_dev_mpd_wp07_b02

**Purpose:** To assess the ability of Mini-Sentinel prospective surveillance tools to reproduce the known association between ACE inhibitors and angioedema, compared to beta blockers

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollment Gap</td>
<td>45 days</td>
</tr>
<tr>
<td>Age Range</td>
<td>18-125</td>
</tr>
<tr>
<td>Query Period</td>
<td>01/01/2008 - 09/30/2013</td>
</tr>
<tr>
<td>Coverage Requirement</td>
<td>Medical and Drug Coverage</td>
</tr>
<tr>
<td>Propensity Score Matching Ratio</td>
<td>1:1</td>
</tr>
<tr>
<td>Propensity Score Matching Caliper</td>
<td>0.025</td>
</tr>
<tr>
<td>Enrollment Requirement</td>
<td>183 days</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure of Interest</th>
<th>Comparator of Interest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ACE Inhibitors</strong></td>
<td><strong>Beta Blockers</strong></td>
</tr>
<tr>
<td>Drug/Exposure:</td>
<td></td>
</tr>
<tr>
<td>Incident w/ respect to:</td>
<td>Beta Blockers, Aliskiren, ARBs</td>
</tr>
<tr>
<td>Washout (days)</td>
<td>183</td>
</tr>
<tr>
<td>Cohort Definition</td>
<td>01</td>
</tr>
<tr>
<td>Episode Gap</td>
<td>14</td>
</tr>
<tr>
<td>Exposure Extension Period</td>
<td>14</td>
</tr>
<tr>
<td>Minimum Episode Duration</td>
<td>0</td>
</tr>
<tr>
<td>Minimum Days Supplied</td>
<td>0</td>
</tr>
<tr>
<td>Episode Truncation by Incident Exposure</td>
<td>Yes</td>
</tr>
<tr>
<td>Inclusion/Exclusion:</td>
<td></td>
</tr>
<tr>
<td>Criterion</td>
<td>Prescription for Aliskiren or any ARB</td>
</tr>
<tr>
<td>Include or Exclude</td>
<td>Exclude</td>
</tr>
</tbody>
</table>

| **Beta Blockers**                     | **ACE Inhibitors, Aliskiren, ARBs** |
| Drug/Exposure:                        |                        |
| Incident w/ respect to:               |                        |
| Washout (days)                        | 183                    |
| Cohort Definition                     | 01                     |
| Episode Gap                           | 14                     |
| Exposure Extension Period             | 14                     |
| Minimum Episode Duration              | 0                      |
| Minimum Days Supplied                 | 0                      |
| Episode Truncation by Incident Exposure | Yes                   |
| Inclusion/Exclusion:                  |                        |
| Criterion                             | Prescription for Aliskiren or any ARB |
| Include or Exclude                    | Exclude                |
Table 1 Unmatched cohorts

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>ACE Inhibitors</th>
<th>Beta Blockers</th>
<th>Absolute Difference</th>
<th>Standardized Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Events while on therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person-time at risk (days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (F)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (std dev)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded History of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergic reactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemic heart diseases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSAID use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Service Utilization Intensity:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of generics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of filled prescriptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of inpatient hospital encounters (IP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of non-acute institutional encounters (IS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of emergency room encounters (ED)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of ambulatory encounters (AV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of other ambulatory encounters (OA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2 Matched cohorts

**Table 2. Cohort of New Initiators of ACE Inhibitors and Beta Blockers (Matched Predefined PS, Caliper = .025)**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>ACE Inhibitors</th>
<th>Beta Blockers</th>
<th>Absolute Difference</th>
<th>Standardized Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Events while on therapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Person-time at risk (days)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient Characteristics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (F)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean age (std dev)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recorded History of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allergic reactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heart failure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemic heart diseases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSAID use</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health Service Utilization Intensity:</td>
<td>Mean</td>
<td>Std Dev</td>
<td>Mean</td>
<td>Std Dev</td>
</tr>
<tr>
<td>Number of generics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of filled prescriptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of inpatient hospital</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>encounters (IP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of non-acute</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>institutional encounters (IS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of emergency room encounters (ED)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of ambulatory encounters (AV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of other ambulatory encounters (OA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Sequential Estimates for Angioedema Events by Analysis Type, and Drug Pair

<table>
<thead>
<tr>
<th>Exposure Definition</th>
<th>Monitoring Period</th>
<th>New Users</th>
<th>Person Years at Risk</th>
<th>Average Person Years at Risk</th>
<th>Number of Events</th>
<th>Incidence Rate per 1000 Person Years</th>
<th>Risk per 1000 New Users</th>
<th>Difference per 1000 Person Years</th>
<th>Difference in Risk per 1000 New Users</th>
<th>Hazard Ratio (95% CI)</th>
<th>Wald P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmatched Analysis (Site-adjusted only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE Inhibitors</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta Blockers</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1:1 Matched Analysis; Caliper=0.025</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE Inhibitors</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beta Blockers</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Subsequent workbook sheets show histograms of unmatched and matched propensity scores for each of 13 data partners.
Propensity scores before match: One site
Propensity scores before match: One site
Three critical elements

- Privacy protections
- Reusable analysis tools
- Analysis-ready data
Common data model—guiding principles

• Accommodates project requirements and can evolve to meet expanded objectives
• Able to incorporate new data types and data elements as needs change
• Leverages existing and evolving data standards
  • Uses existing native coding systems and minimizes ontology mapping
• Captures values found in source data
Common data model—guiding principles

• Transparent, intuitive design that is easily understood by analysts and investigators
• Local implementation may include “site-specific” variables
Common data model

- Relational structure provides analysis-ready platform
- Encounter basis incorporates EHR and claims-type data
Rigorous data checking and characterization

• ~1500 data checks per refresh
Why QA after every refresh?

- Underlying data sources are dynamic
- Verify compliance with data model
- Identify changes in data sources or transformation processes
- Identify problems and/or differences in data transformation methods
Why QA after every refresh?

**Green**: records from prior refresh  
**Red**: record from new refresh under review

**Problem:**  
Enrollment data from 2010 was archived between refreshes and not included in latest refresh.

**Outcome:**  
Data Partner was asked to recreate the refresh including 2010 data.
The DRN is ready for NIH to use

- Assess disease burden/outcomes
- Pragmatic clinical trial design
- Single study private network
- Pragmatic clinical trial follow up
- Reuse of research data
Thank You

For more information

- nihcollaboratory.org/Pages/distributed-research-network.aspx
- PopMedNet.org
- info@nihquery.org
- Jeff_brown@harvardpilgrim.org

Prior Grand Rounds on the NIH Collaboratory Distributed Research Network