NIH Collaboratory

Health Care Systems Research Collaboratory

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

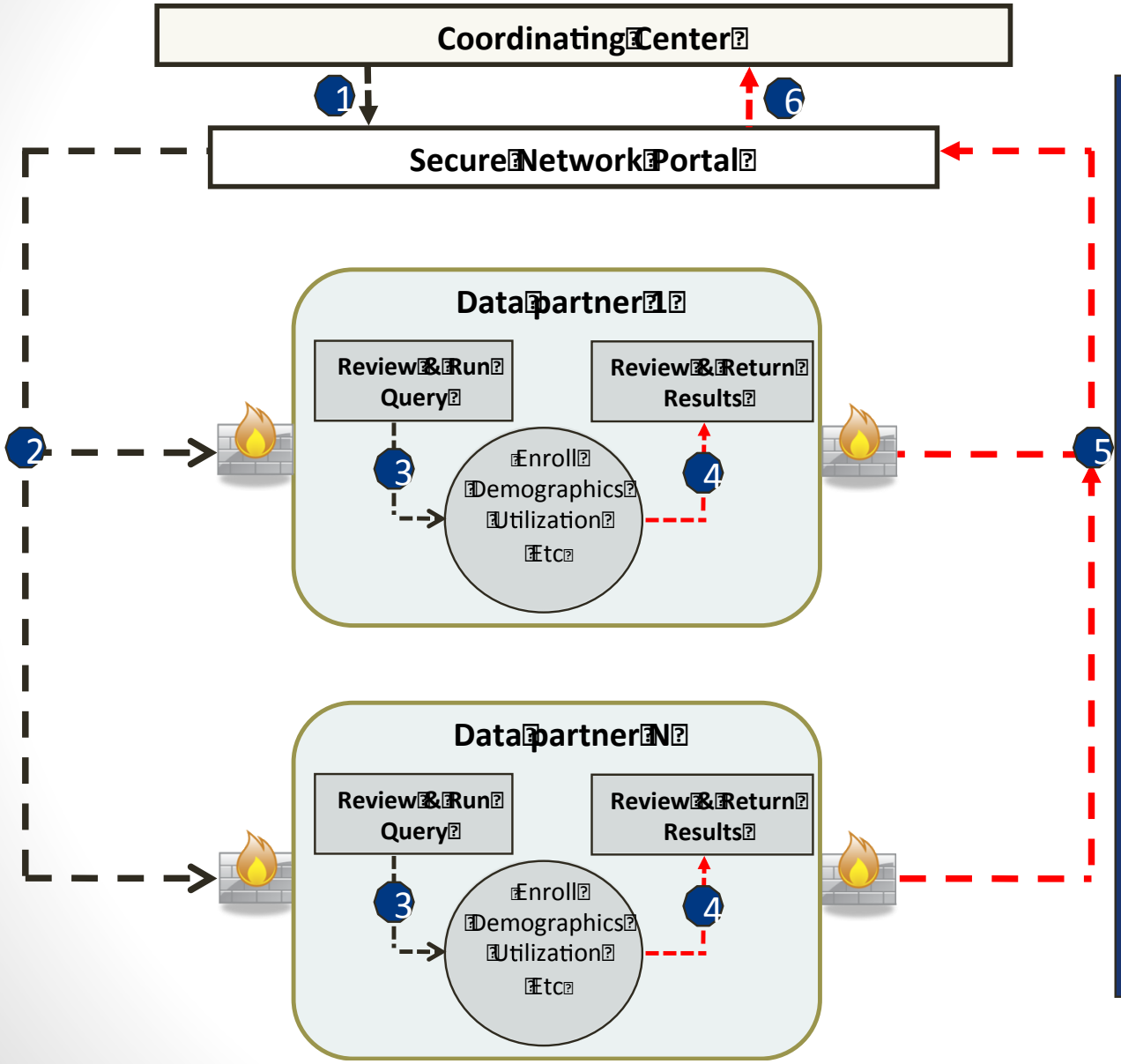
Three critical elements

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

Three critical elements

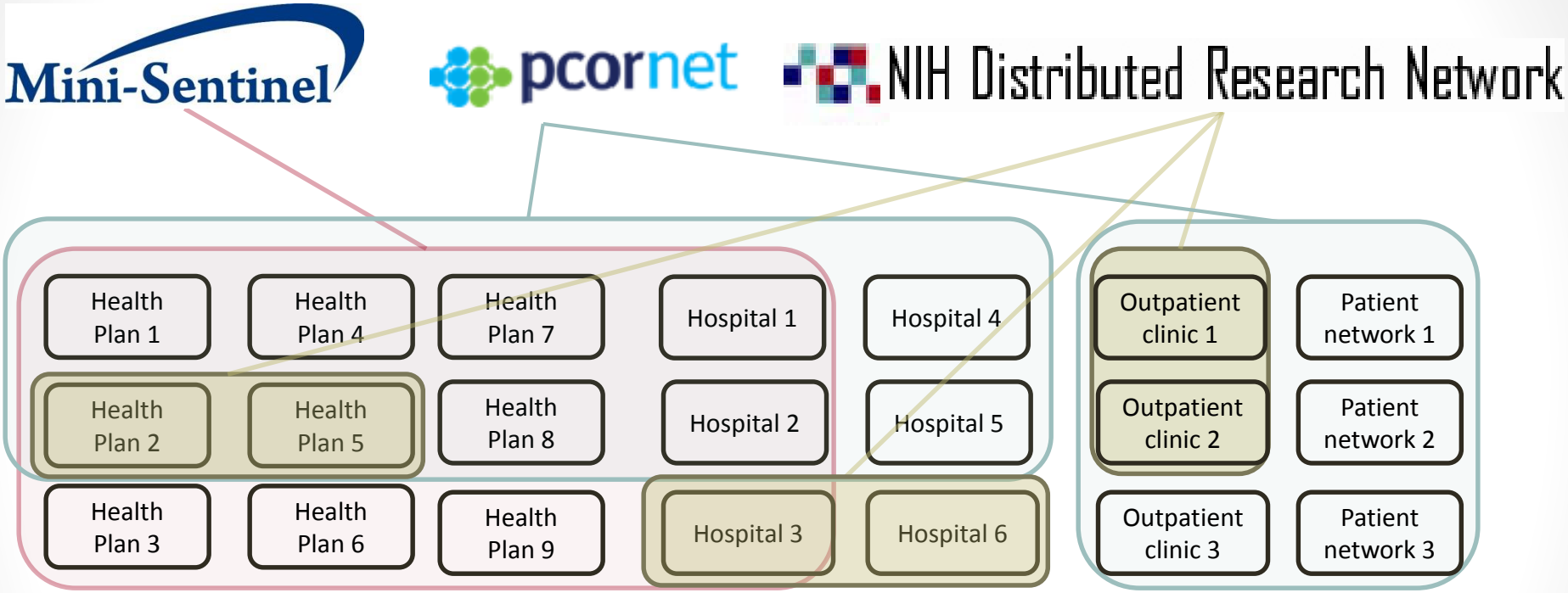
- **Privacy protections**
- Reusable analysis tools
- Analysis-ready data

Distributed analysis



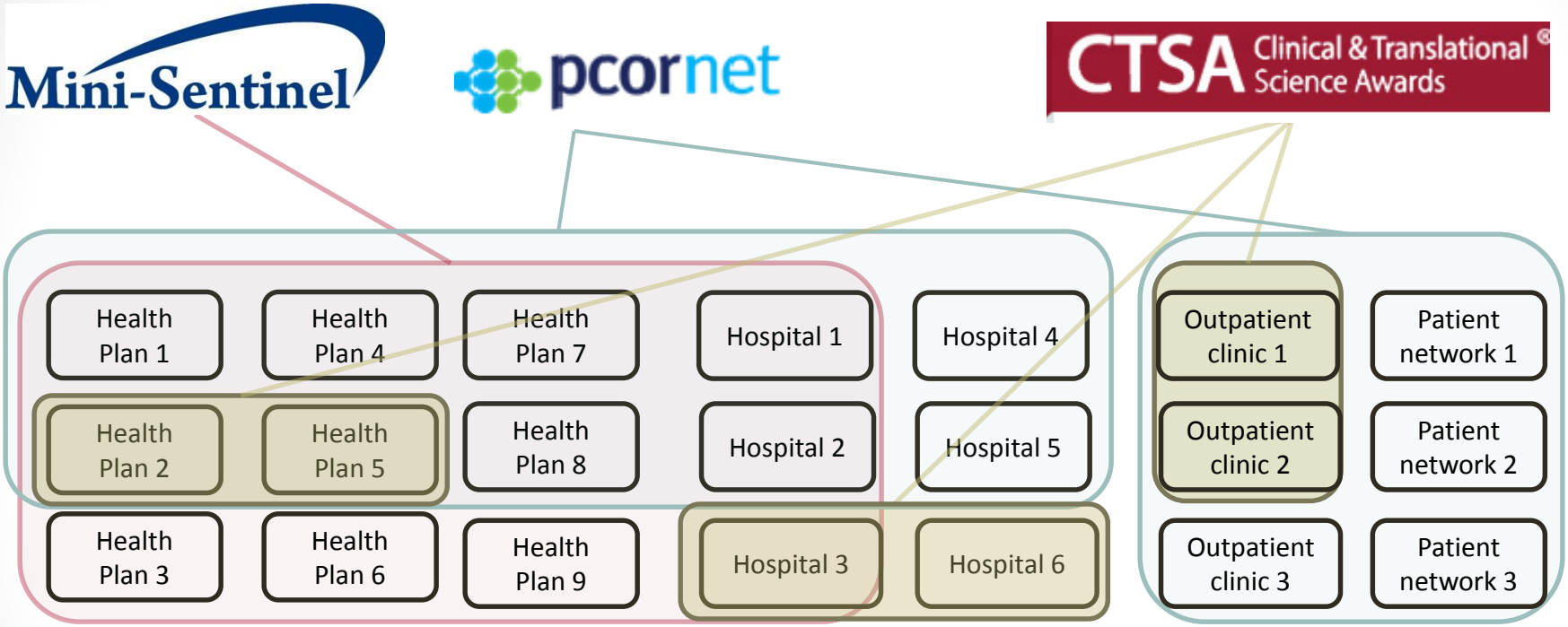
1. User creates and submits query (a computer program)
2. Individual data partners retrieve query
3. Data partners review and run query against their local data
4. Data partners review results
5. Data partners return results via secure network
6. Results are aggregated

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

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

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

CLINICAL TRIALS TRANSFORMATION INITIATIVE

Mini-Sentinel

MINI-SENTINEL and CLINICAL TRIALS TRANSFORMATION INITIATIVE

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

February 28, 2014

www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_CTTI_Developing-Approaches-to-Conducting-Randomized-Trials-Using-MSDD.pdf

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 $CHADS_2$ (congestive heart failure, hypertension, age > 75 yrs, diabetes, stroke or TIA) risk factor OR
  ≥2 CHA_2DS_2 VASc (congestive heart failure, hypertension, age, diabetes, stroke or TIA, vascular disease, female) risk factors

www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_CTTI_Developing-Approaches-to-Conducting-Randomized-Trials-Using-MSDD.pdf

Use Case: IMPACT-AF Cluster Randomized Trial

Can we quickly identify eligible patients?

- **…d by Christopher Granger, MD, and colleagues**
- **P…** …est a multilevel educational intervention to increase the rate of initiati… …coagulants among patients with atrial fibrillation.
- **Design:** Cluste… …d trial
- **Intervention:**
  - For patients – Mailed edu… …ial, and recommendation to discuss their anticoagulation status with the…
  - For physicians – Notification of eligib… …orts regarding their eligible patients' rate of anticoagulation benchma… …oviders.
- **Population: Patients ≥18 years with atrial fibri… …t anticoagulation AND ≥1 $CHADS_2$ (congestive heart failure, hypertension, age >… …tes, stroke or TIA) risk factor OR ≥2 CHA_2DS_2 VASc (congestive heart failure, hypertension, age, diabe… …, stroke or TIA, vascular disease, female) risk factors**

www.mini-sentinel.org/work_products/Statistical_Methods/Mini-Sentinel_Methods_CTTI_Developing-Approaches-to-Conducting-Randomized-Trials-Using-MSDD.pdf

Use cases

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

ORIGINAL INVESTIGATION

ONLINE FIRST

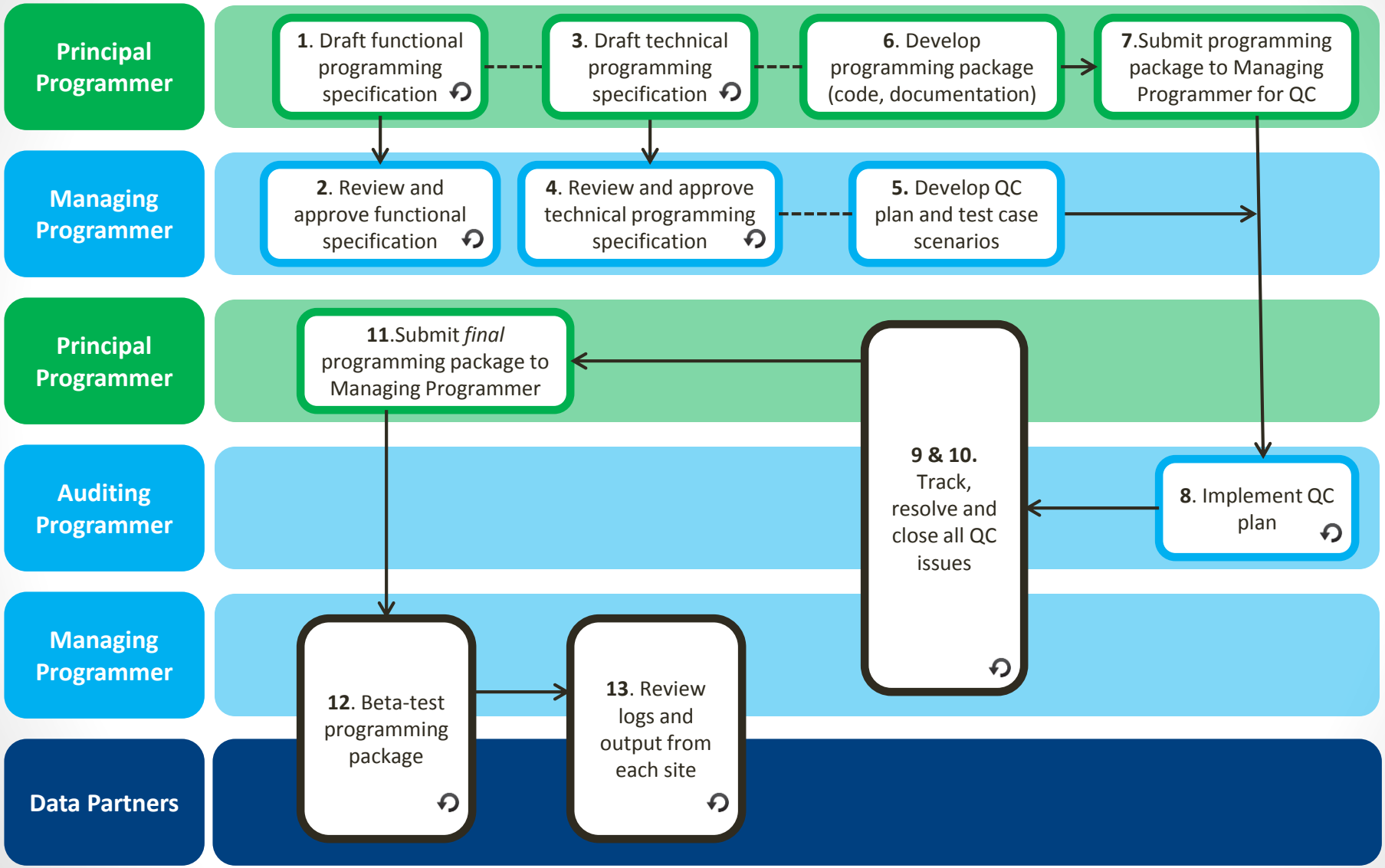
Comparative Risk for Angioedema Associated With the Use of Drugs That Target the Renin-Angiotensin-Aldosterone System

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 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

Toh Arch Intern Med.2012;172:1582-1589.

New program development process



ORIGINAL INVESTIGATION

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 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**

Toh Arch Intern Med.2012;172:1582-1589.

ORIGINAL INVESTIGATION

ONLINE FIRST

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

Sengwee Toh, ScD; Marsha E. [illegible] Monika Houstoun, PharmD; Mary Ross Southworth, PharmD; Xiao Ding, PhD; Adrian F. Hernandez [illegible] nson, PhD; Lingling Li, PhD; Carolyn McCloskey, MD, MPH; Azadeh Shoaibi, MS, MHS; Eileen Wu, Ph[illegible] berg, MD, MS, ScD; Sean Hennessy, PharmD, PhD

Can we reduce the effort, time, and cost?

- Used data for 3.9 million new u[illegible]i-hypertensives in 18 organizations
- Propensity score matched analysis
- No person-level data was shared
- Five months and $250,000 required for programmi[illegible] analysis – compared to 1-2 years and $2 million without [illegible]lysis-ready distributed dataset

NIH Collaboratory | NIH Distributed Research Network

Toh Arch Intern Med.2012;172:1582-1589.

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 (eg, treatments, outcomes, etc)
- Level 2: Comparative analyses with analytic adjustment for confounding using available analytic adjustment tools (eg, 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

Query Period	1/1/2006- 12/31/2013
Coverage Requirement	Medical and Drug Coverage
Enrollment Requirement	183 days
Enrollment Gap	45 days
Age Groups	18-34, 35-44, 45-64 65-74, 75+

Patient A (IMPACT-AF example)

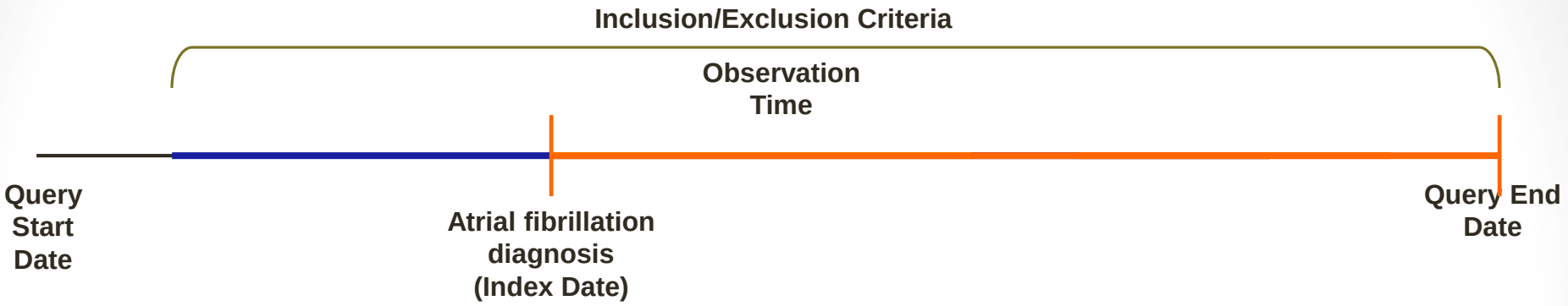
Two cohort definitions

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

Patient A (IMPACT-AF example)



Multiple inclusion/exclusion criteria (n=8)

- At least **one** $CHADS_2$ risk factor OR at least **two** CHA_2DS_2-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding
- At least **two** $CHADS_2$ risk factors OR at least **three** CHA_2DS_2-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding
- At least **one** $CHADS_2$ risk factor, EXCLUDE mechanical prosthetic valve and life-threatening bleeding (only relevant for 75+ group)
- At least **two** $CHADS_2$ risk factors OR at least **one** CHA_2DS_2-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding (only relevant for 75+ group)
- At least **two** $CHADS_2$ risk factors OR at least **two** CHA_2DS_2-VASc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding

Complete specifications

Query Period	1/1/2006- 12/31/2013
Coverage Requirement	Medical and Drug Coverage
Enrollment Requirement	183 days
Enrollment Gap	45 days
Age Groups	18-34, 35-44, 45-64 65-74, 75+

	Incident Diagnosis							Pre-Existing Condition				Post-Diagnosis Treatment				
Scenario	Diagnosis	Virtual Record Date	Care Setting	Incidence Type	Washout (days)	Look-Up Period	Min. Look-up Period	Pre-Existing Inclusion/Exclusion Criteria	Lookback Start	Lookback End	Care Setting/ PDX	Post-Diagnosis Treatment	CodeCount	Incidence Type	Washout (days)	Episode Gap
1	Atrial Fibrillation*	NA	Any	SING	0	6000	0	None	NA	NA	NA	Anticoagulant Use	0	MULT	0	30
2	Two Atrial Fibrillation* codes on different days	First Date	Any	SING	0	6000	0	None	NA	NA	NA	Anticoagulant Use	0	MULT	0	30
3	Atrial Fibrillation*	NA	Any	SING	0	6000	0	At least one CHADS2 risk factor OR at least two CHADS2-vasc risk factors, EXCLUDE mechanical prosthetic valve and life-threatening bleeding**	-6000	6000	Any	Anticoagulant Use	0	MULT	0	30
	Two Atrial Fibrillation*							At least one CHADS2 risk factor OR at least two CHADS2-vasc risk				Anticoagulant				

Specifications | Glossary | Questions | Anticoagulants | CHADS2 | CHADS2-Vasc | MechProsthValv | Bleeding

- **16 different cohorts with different definitions for diagnosis and pre-existing condition requirements**
- **Once specifications are complete, results available within weeks**

ORIGINAL INVESTIGATION

ONLINE FIRST

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

Sengwee Toh, ScD; Marsha E. [illegible] Monika Houstoun, PharmD; Mary Ross Southworth, PharmD; Xiao Ding, PhD; Adrian F. Hernandez [illegible] nson, PhD; Lingling Li, PhD; Carolyn McCloskey, MD, MPH; Azadeh Shoaibi, MS, MHS; Eileen Wu, Ph[illegible] berg, MD, MS, ScD; Sean Hennessy, PharmD, PhD

Can we reduce the effort, time, and cost?

- Used data for 3.9 million new u[illegible]i-hypertensives in 18 organizations
- Propensity score matched analysis
- No person-level data was shared
- Five months and $250,000 required for programmi[illegible] analysis – compared to 1-2 years and $2 million without [illegible]lysis-ready distributed dataset

NIH Collaboratory | NIH Distributed Research Network

Toh Arch Intern Med.2012;172:1582-1589.

Propensity Score Matched tool

- 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

Mini-Sentinel

Overview

Request Description	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
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 = .025)
Table 3	Table displaying Sequential Estimates for Angioedema Events by Analysis Type, and
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/Excluions 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

Overview | Specifications | Glossary | PSM_Glossary | Monitoring Period | Table 1 | Table 2 | Table 3 | Appendix A

Specifications

Mini-Sentinel

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

Enrollment Gap	45 days
Age Range	18-125
Query Period	01/01/2008 -09/30/2013
Coverage Requirement	Medical and Drug Coverage
Propensity Score Matching Ratio	1:1
Propensity Score Matching Caliper	0.025
Enrollment Requirement	183 days

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

Overview | Specifications | Glossary | PSM_Glossary | Monitoring Period | Table 1 | Table 2 | Table 3 | Appendix A | Appen

Table 1 Unmatched cohorts

Mini-Sentinel

Table 1. Cohort of New Initiators of ACE Inhibitors and Beta Blockers (Unmatched)

	Primary Analysis				Covariate Balance	
Characteristic	**ACE Inhibitors**		**Beta Blockers**			
	N	**%**	**N**	**%**	**Absolute Difference**	**Standardized Difference**
Patients						
Events while on therapy						
Person-time at risk (days)						
Patient Characteristics						
Gender (F)						
Mean age (std dev)						
Recorded History of:						
Allergic reactions						
Diabetes						
Heart failure						
Ischemic heart diseases						
NSAID use						
Health Service Utilization Intensity:	**Mean**	**Std Dev**	**Mean**	**Std Dev**		
Number of generics						
Number of filled prescriptions						
Number of inpatient hospital encounters (IP)						
Number of non-acute institutional encounters (IS)						
Number of emergency room encounters (ED)						
Number of ambulatory encounters (AV)						
Number of other ambulatory encounters (OA)						

Overview | Specifications | Glossary | PSM_Glossary | Monitoring Period | **Table 1** | Table 2 | Table 3 | Appendix A | Appendix

Table 2 Matched cohorts

Mini-Sentinel

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

	Primary Analysis				Covariate Balance	
Characteristic	**ACE Inhibitors**		**Beta Blockers**			
	N	**%**	**N**	**%**	**Absolute Difference**	**Standardized Difference**
Patients						
Events while on therapy						
Person-time at risk (days)						
Patient Characteristics						
Gender (F)						
Mean age (std dev)						
Recorded History of:						
Allergic reactions						
Diabetes						
Heart failure						
Ischemic heart diseases						
NSAID use						
Health Service Utilization Intensity:	**Mean**	**Std Dev**	**Mean**	**Std Dev**		
Number of generics						
Number of filled prescriptions						
Number of inpatient hospital encounters (IP)						
Number of non-acute institutional encounters (IS)						
Number of emergency room encounters (ED)						
Number of ambulatory encounters (AV)						
Number of other ambulatory encounters (OA)						

Table 3 Rates, differences, hazard ratios

Mini-Sentinel

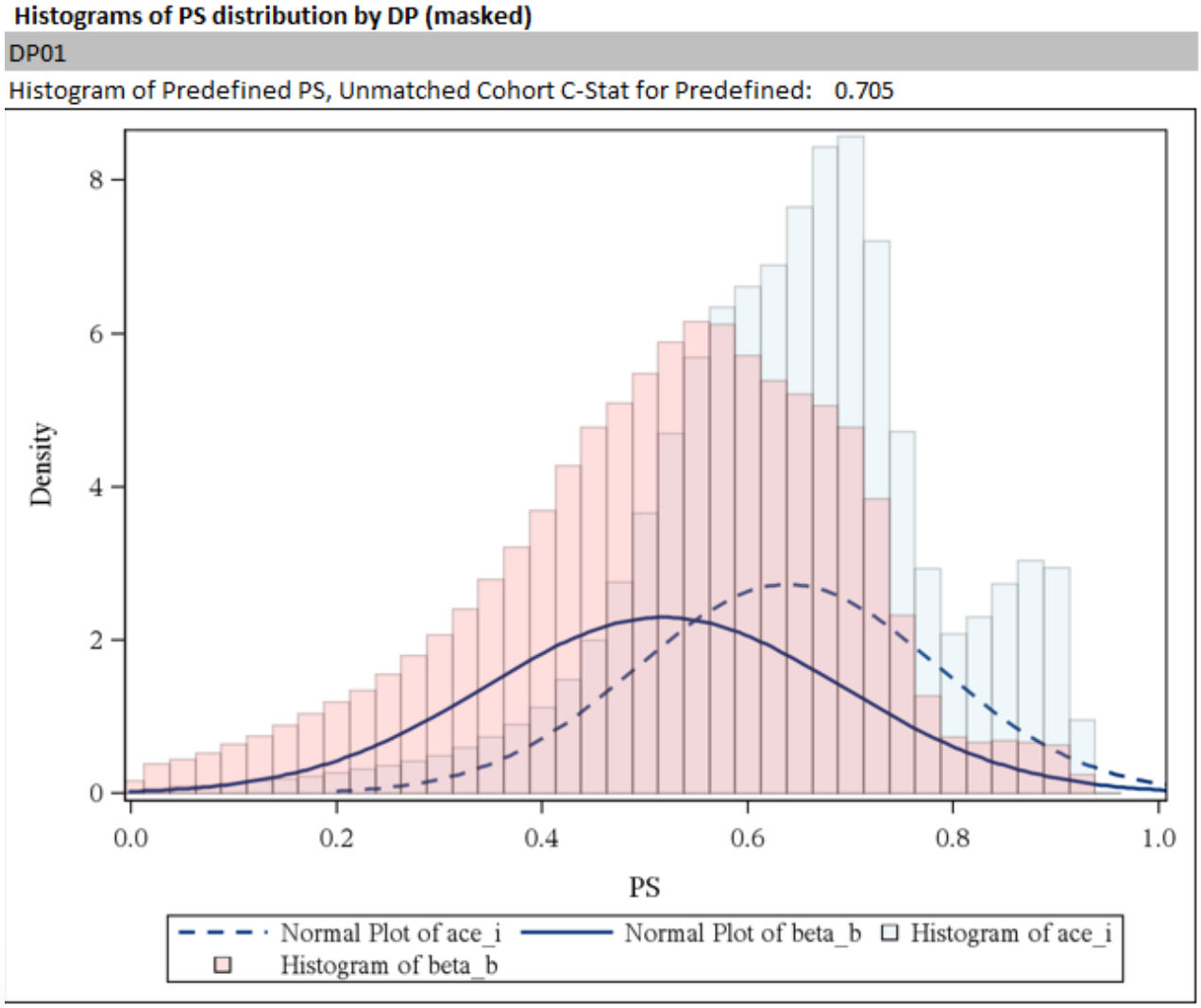
Table 3: Sequential Estimates for Angioedema Events by Analysis Type, and Drug Pair

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

Subsequent workbook sheets show histograms of unmatched and matched propensity scores for each of 13 data partners

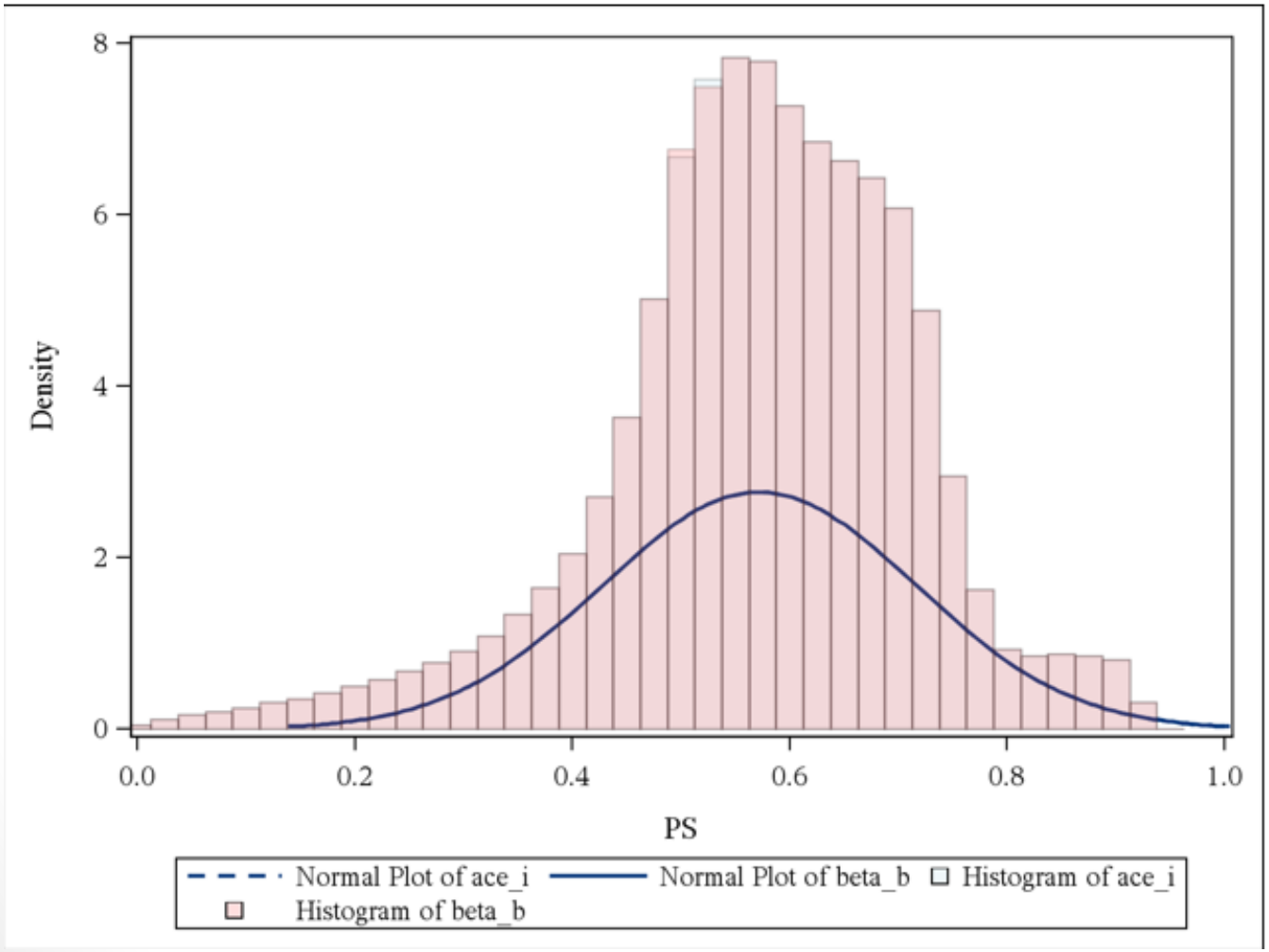
erview | Specifications | Glossary | PSM_Glossary | Monitoring Period | Table 1 | Table 2 | Table 3 | Appendix A | Appendix B | Appendix

Propensity scores before match: One site



Propensity scores before match: One site

Histogram of Predefined PS among Predefined PS Matched Cohort, Matched Cal = .025 C-Stat for Predefined: 0.705



Three critical elements

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

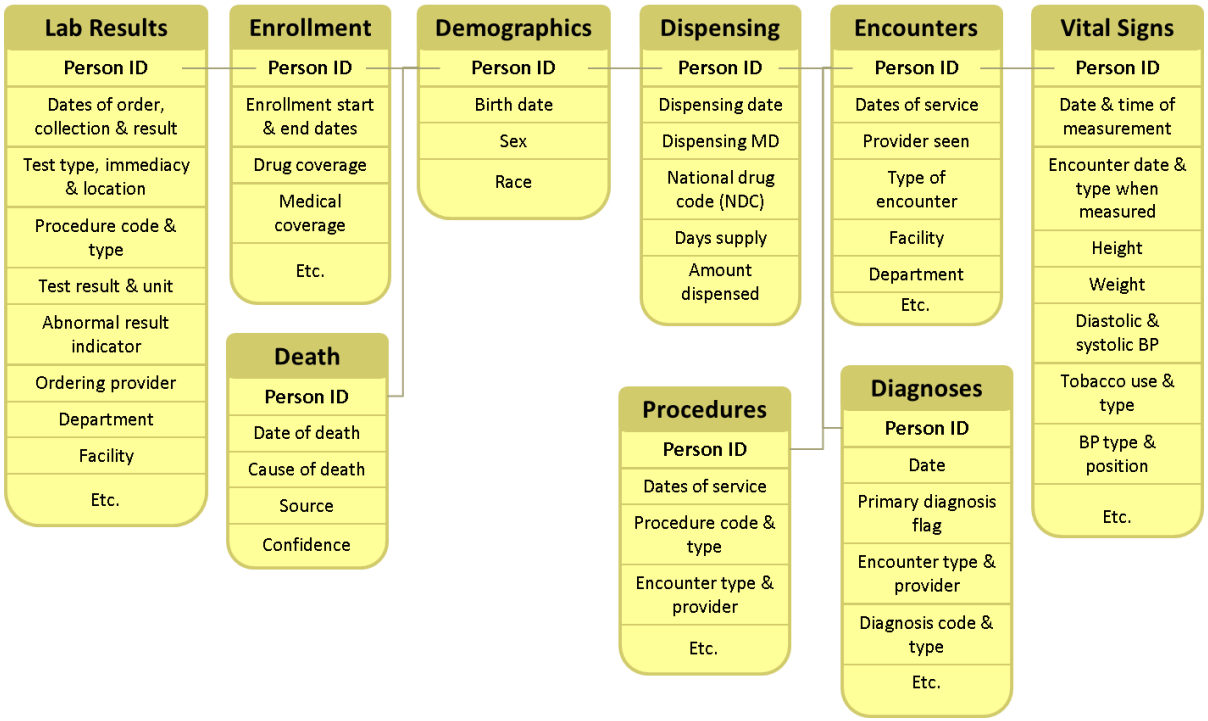
Common data model—guiding principles

- 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

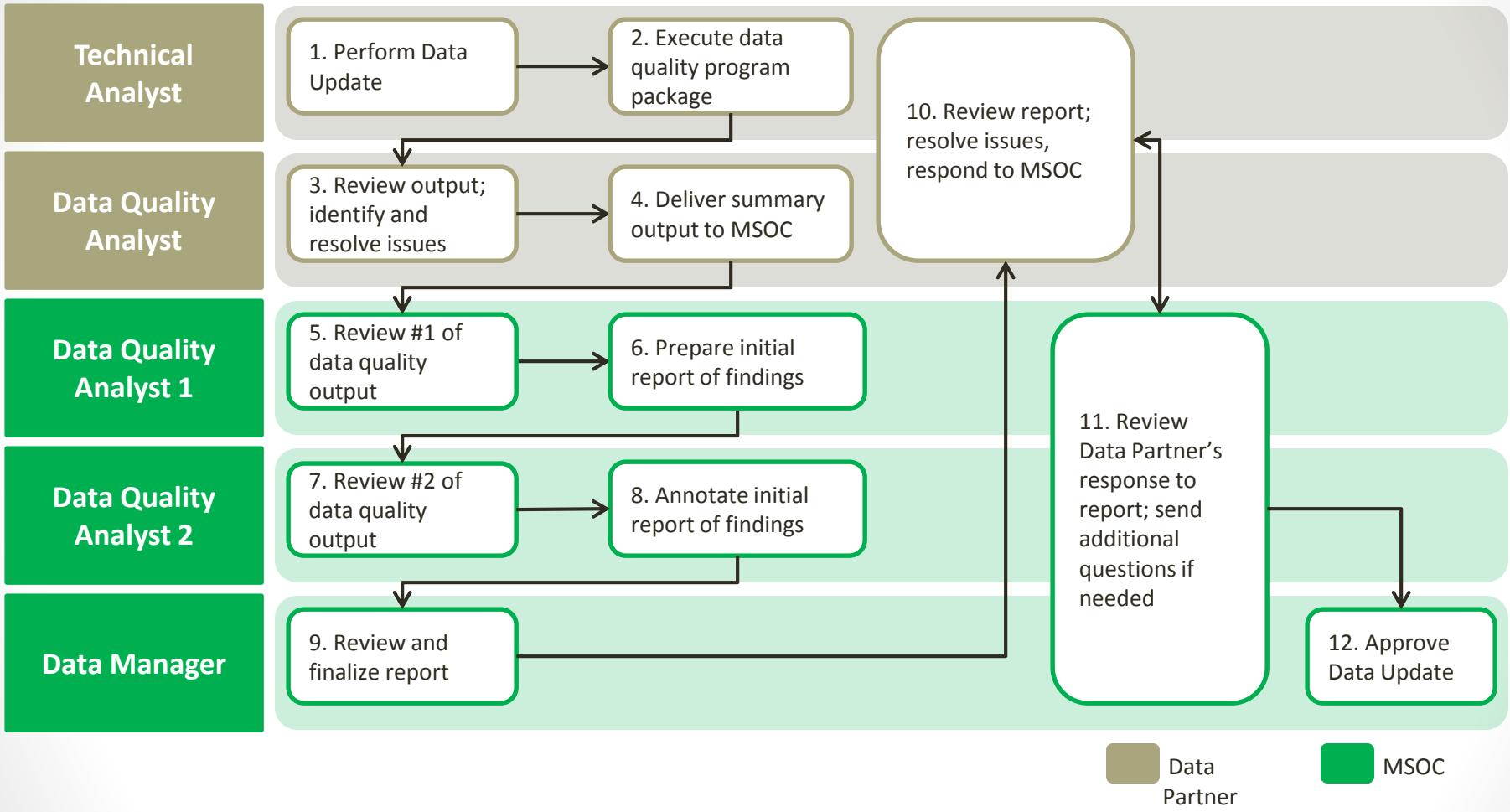
- Transparent, intuitive design that is easily understood by analysts and investigators
- Local implementation may include “site-specific” variables

Common data model



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

Data QA review process



Rigorous data checking and characterization

- ~1500 data checks per refresh

Obs	ENCTYPE	ADATE	COUNT	PERCENT
1	AV	2000	7030952	5.1370
2	AV	2001	7454699	5.4466
3	AV	2002	8014346	5.8555
4	AV	2003	8261199	6.0358
5	AV	2004	8251011	6.0284
6	AV	2005	8857635	6.4716
7	AV	2006	9576674	6.9969
8	AV	2007	10240959	7.4823
9	AV	2008	11831682	8.6445
10	AV	2009	13785025	10.0716
11	AV	2010	14499322	10.5935
12	AV	2011	14988289	10.9508
13	ED	2000	193108	0.1411
14	ED	2001	213180	0.1558
15	ED	2002	231296	0.1690
16	ED	2003	232122	0.1696
17	ED	2004	230756	0.1686
18	ED	2005	266406	0.1946
19	ED	2006	291381	0.2129
20	ED	2007	314060	0.2295
21	ED	2008	343936	0.2513
22	ED	2009	400500	0.2926
23	ED	2010	414312	0.3027
24	ED	2011	451881	0.3302
25	IP	2000	432504	0.3
26	IP	2001	477466	0.3
27	IP	2002	517710	0.3
28	IP	2003	543660	0.3
29	IP	2004	543692	0.3
30	IP	2005	587863	0.4

Obs	RXDATE	N
1	2000JAN	75816
2	2000FEB	68872
3	2000MAR	240058
4	2000APR	248527
5	2000MAY	261254
6	2000JUN	258289
7	2000JUL	241145
8	2000AUG	260316
9	2000SEP	252799
10	2000OCT	260813
11	2000NOV	254161
12	2000DEC	259611
13	2001JAN	275314
14	2001FEB	242270
15	2001MAR	278558
16	2001APR	260591
17	2001MAY	268647
18	2001JUN	267520
19	2001JUL	257699
20	2001AUG	279320
21	2001SEP	251170

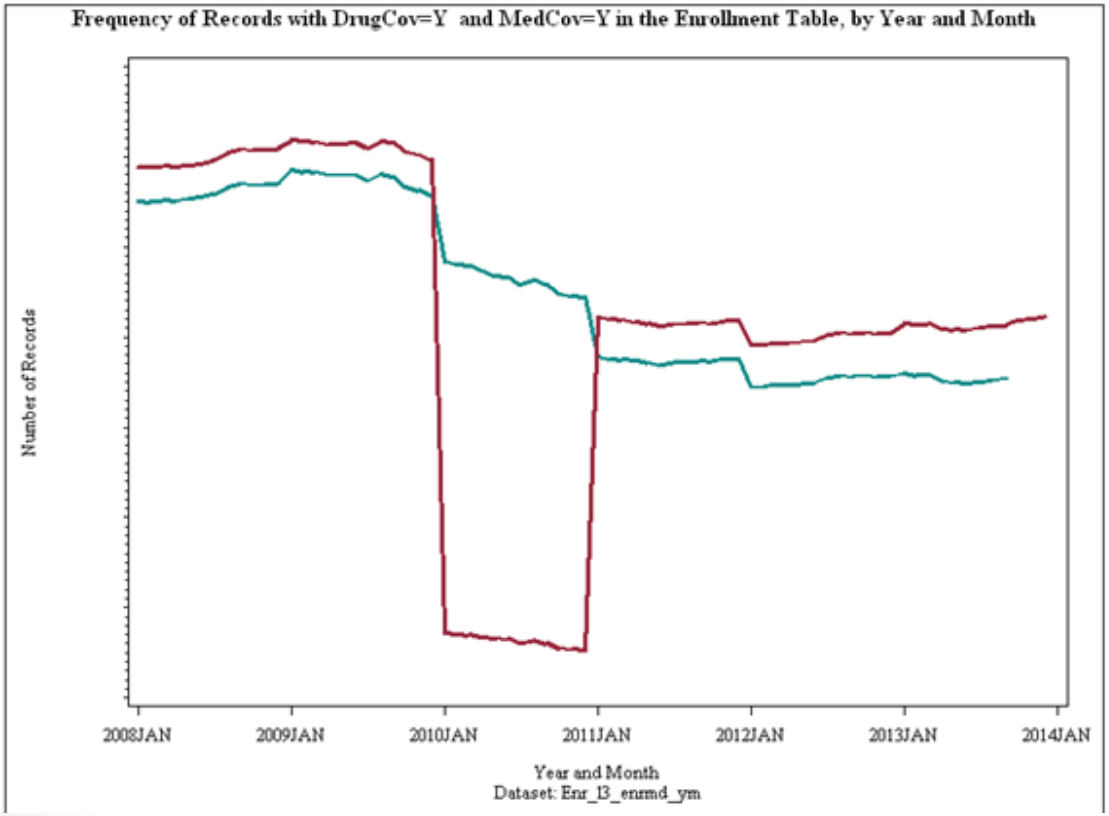
Obs	Age_group	COUNT	PERCENT
1	0.1 0-1 Yrs	602059	1.4996
2	02. 2-4 Yrs	1376997	3.4298
3	03. 5-9 Yrs	2553188	6.3595
4	04. 10-14 Yrs	2638462	6.5719
5	05. 15-18 Yrs	2135457	5.3190
6	06. 19-21 Yrs	1670742	4.1615
7	07. 22-44 Yrs	14770481	36.7906
8	08. 45-64 Yrs	11221814	27.9515
9	09. 65-74 Yrs	1854092	4.6182
10	10. 75+ Yrs	1324163	3.2982

Obs	px_codetype	enctype	COUNT	PERCENT
1	09	AV	3891384	0.2061
2	09	ED	940211	0.0498
3	09	IP	7716848	0.4088
4	09	IS	168596	0.0089
5	09	OA	510196	0.0270
6	C2	AV	4906255	0.2599
7	C2	ED	325738	0.0173
8	C2	IP	392155	0.0208
9	C2	IS	18219	0.0010
10	C2	OA	222605	0.0118
11	C3	AV	212648	0.0113
12	C3	ED	5276	0.0003
13	C3	IP	7755	0.0004
14	C3	IS	269	0.0000
15	C3	OA	2030	0.0001
16	C4	AV	1364119936	72.2580
17	C4	ED	95271865	5.0466
18	C4	IP	50242438	2.6614
19	C4	IS	3914519	0.2074
20	C4	OA	27959691	1.4810
21	HC	AV	252901204	13.3963
22	HC	ED	14811325	0.7846
23	HC	IP	8125355	0.4304
24	HC	IS	1600478	0.0848
	HC	OA	31067795	1.6457
	ND	AV	16692216	0.8842
	ND	ED	639229	0.0339
	ND	IP	147970	0.0078
	ND	IS	12924	0.0007
	ND	OA	819916	0.0434
	OT	AV	194765	0.0103
	OT	ED	374	0.0000
	OT	IP	2607	0.0001
	OT	IS	1367	0.0001
	OT	OA	348	0.0000

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

https://www.nihcollaboratory.org/Pages/Grand-Rounds-03-15-13.aspx

https://www.nihcollaboratory.org/Pages/Grand-Rounds-09-13-13.aspx

https://www.nihcollaboratory.org/Pages/Grand-Rounds-06-13-14.aspx