Distributed data network architecture: Lessons from PCORnet and FDA Sentinel

Jeffrey Brown, PhD

Associate Professor

Harvard Pilgrim Health Care Institute and

Harvard Medical School

July 12, 2016

Outline

- Why need to use electronic data for evidence generation
- Why multisite studies
- Examples
  - FDA Sentinel
  - PCORnet
- Thoughts

Medical interventions

- For medical products, at marketing approval we know
  - Within a small, well-defined population in a controlled environment, and short-term exposure, the product is
    - Relatively safe
    - More effective than placebo
- For other interventions we know even less
- For all interventions **we know we don't know**
  - Real-world safety
  - Real-world effectiveness
  - Comparative effectiveness
  - Cost-benefit

Electronic data can help generate real-world evidence

- Electronic health records
- Insurance claims data
- Registries
  - Birth
  - Death
  - Immunization
  - Disease
- Patient-generated data
- Many others

Why data networks

- Rare exposures
- Rare outcomes
- Sample size (speed)
- Sub-group analyses
- Analytic flexibility

Rule: If you can do your work within a single system or institution, you should

Data networks have different goals

- Exchange of patient data for patient care at the point of care (not covered here)
- Public health surveillance
- Research
- Clinical trial enrollment

Some multi-purpose, multi-site data networks I've worked on

- Vaccine Safety Datalink
- Health Care Systems Research Network (previously known as HMO Research Network)
- Cancer Research Network
- FDA Sentinel
- NIH Health Care Systems Research Collbaortory
- Innovation in Medical Evidence Development and Surveillance (IMEDS)
- PCORnet
- Biologics and Biosimilars Collective Intelligence Consortium
- Massachusetts Department of Public Health (MDPHnet)

Some multi-purpose, multi-site data networks I've worked on

- Vaccine Safety Datalink
- Health Care Systems Research Network (previously known as HMO Research Network)
- Cancer Research Network
- **FDA Sentinel**
- NIH Health Care Systems Research Collbaortory
- Innovation in Medical Evidence Development and Surveillance (IMEDS)
- **PCORnet**
- Biologics and Biosimilars Collective Intelligence Consortium
- Massachusetts Department of Public Health (MDPHnet)

FDA Sentinel: Background

- **2007**: FDA Amendments Act
  - A mandate to create an active surveillance system
  - Access data from **25 million** individuals by July 2010
  - Access data from **100 million** individuals by July 2012

- 2008: FDA launched the Sentinel Initiative
- 2009: Mini-Sentinel funded under Sentinel Initiative
- 2014: Funding awarded for Sentinel
- *Operates under FDA's public health authority*

Lead – HPHC Institute

Data and scientific partners

Scientific partners

Why distributed?

- Centralized vs. Distributed

- Distributed data system is preferred because
  - Data sits behind data partner's firewall
  - Data remains under local control
  - Only minimally necessary info is shared in a given analysis
  - Preserve patient privacy & institutional proprietary interests
  - Enables rapid creation of multiple networks that leverage the architecture
  - Avoids complex contracting and institutional agreements

Sentinel distributed architecture

- Standardize data
- Data partners maintain physical control of their data
- Data partners control all uses of their data
- Data partners control all transfer of data
- Computer programs distributed via a secure network supported by PopMedNet™
- Program execute at multiple sites without modification

Sentinel Common Data Model

Administrative

Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure
Person ID	**Person ID**	**Person ID**	**Person ID**	**Person ID**	**Person ID**
Enrollment start & end dates	Birth date	Dispensing date	Service date(s)	Service date(s)	Service date(s)
Drug coverage	Sex	National drug code (NDC)	Encounter ID	Encounter ID	Encounter ID
Medical coverage	ZIP code	Days supply	Encounter type & provider	Encounter type & provider	Encounter type & provider
Medical record availability	Etc.	Amount dispensed	Facility	Diagnosis code & type	Procedure code & type
			Etc.	Principal discharge diagnosis	Etc.

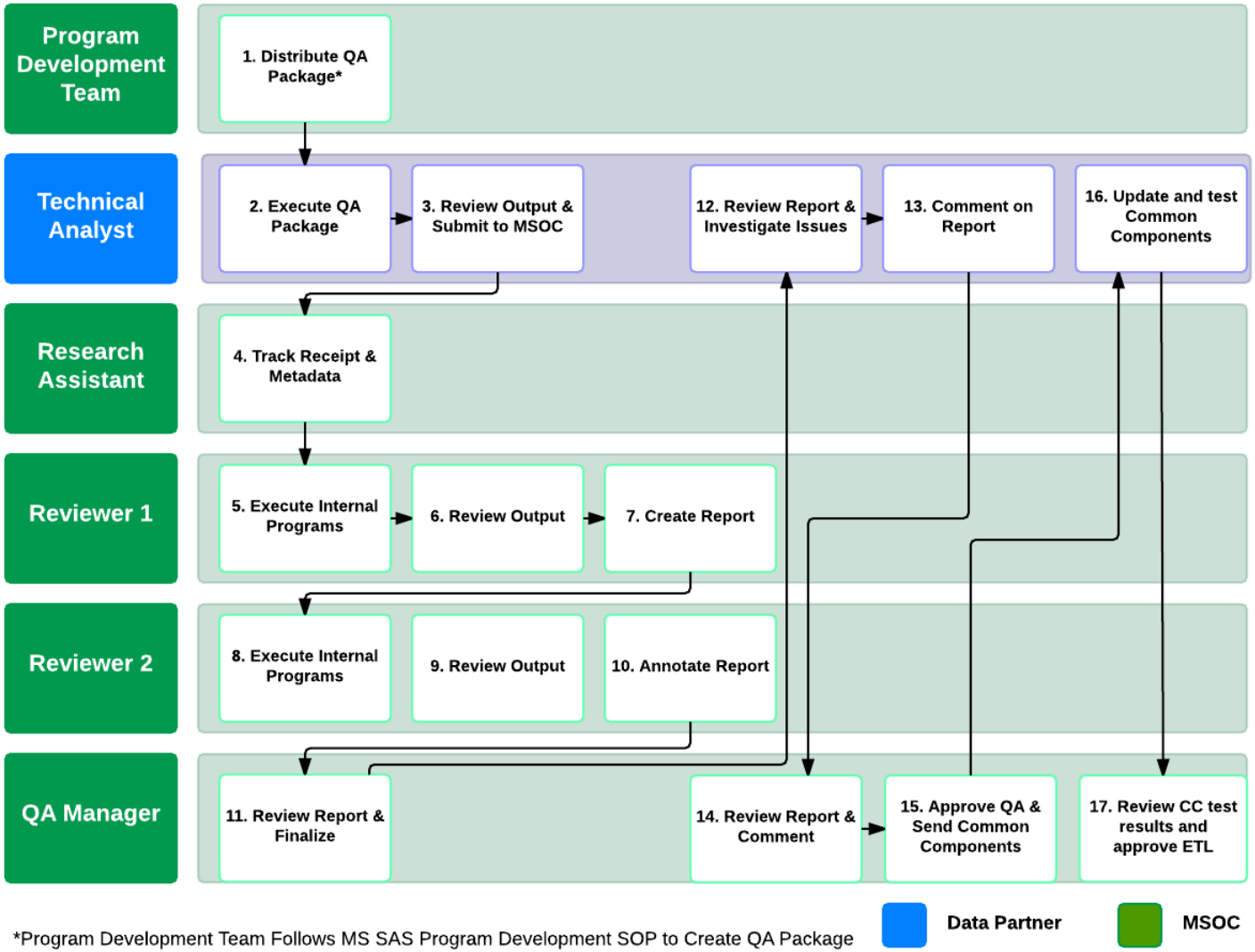
Clinical

Lab Result	Vital Signs
Person ID	**Person ID**
Result and specimen collection dates	Measurement date and time
Test type, immediacy & location	Height and weight
Logical Observation Identifiers Names and Codes (LOINC ®)	Diastolic & systolic BP
Test result & unit	Tobacco use & type
Etc.	Etc.

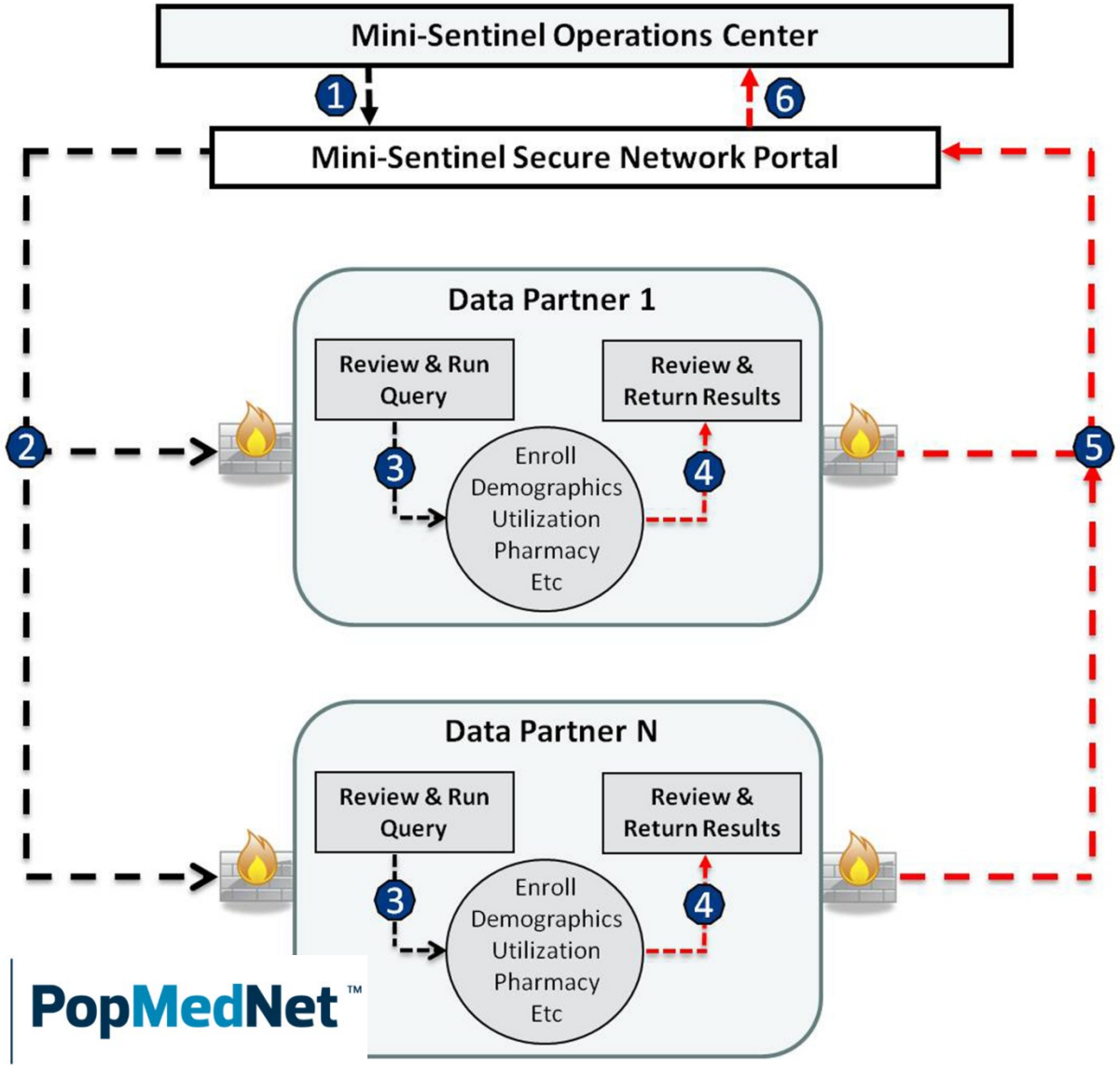
Registry

Death	Cause of Death	State Vaccine
Person ID	**Person ID**	**Person ID**
Death date	Cause of death	Vaccination date
Source	Source	Admission Type
Confidence	Confidence	Vaccine code & type
Etc.	Etc.	Provider
		Etc.

Data Quality Assurance SOP



Distributed Analysis



1- User creates and submits query (a computer program)

2- 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

PopMedNet™

Sentinel Distributed Database

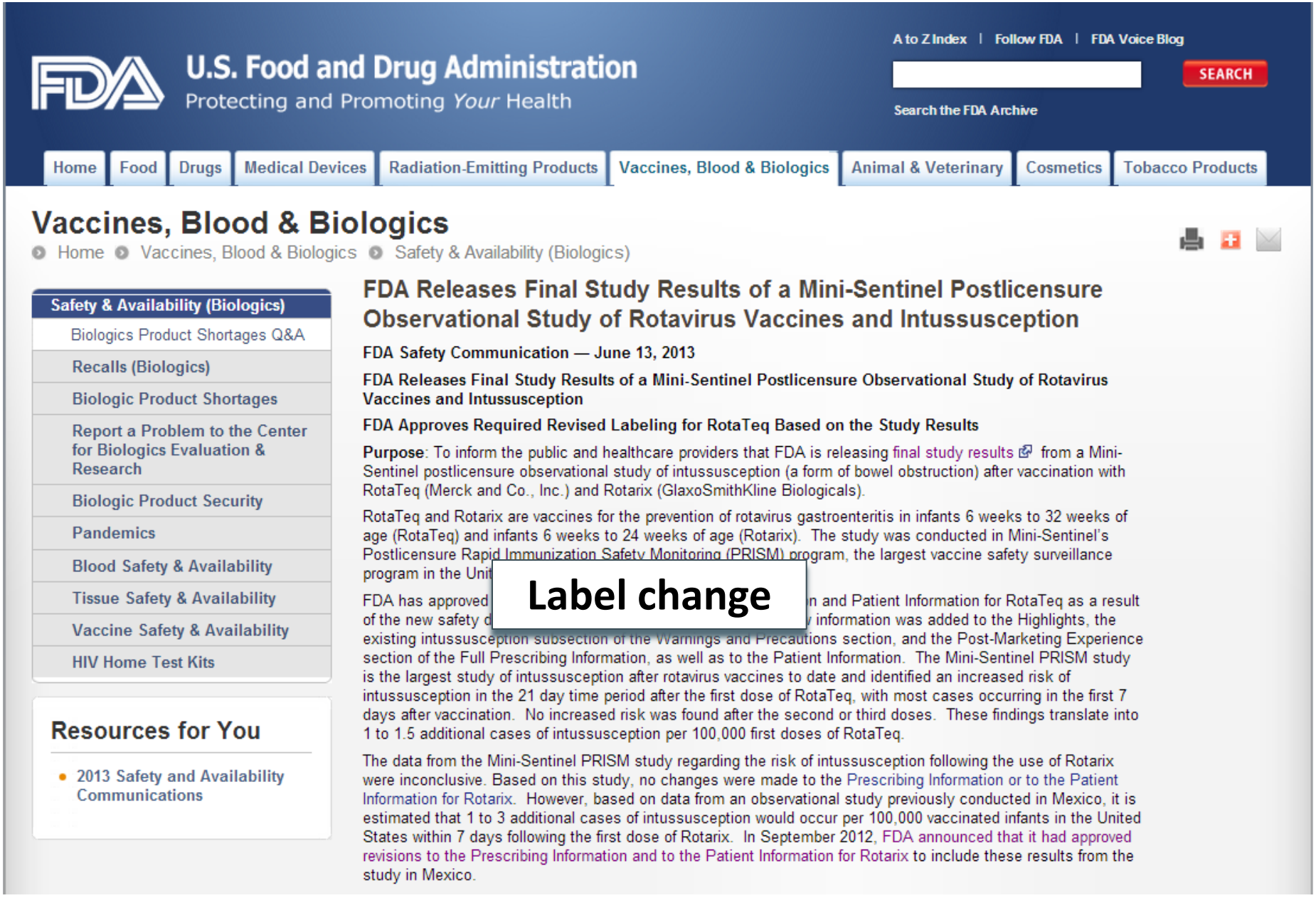
- Populations with well-defined longitudinal person-time for which most medically attended events are known
- **193 million** individuals* with ~351 million person-years of observation time
- **39 million** currently accumulating data
- **4 billion** dispensings, accumulating **46 million**/month
- **5.5 billion** unique encounters, including **51 million** inpatient stays
- **33 million** with at least one laboratory result
- Ability to obtain electronic or paper medical records (redacted and de-identified)

*As of August 2015. Potential for double-counting exists if individuals moved between data partner health plans

Impact / Dissemination

- 4 FDA drug safety communications
  - Tri-valent inactivated flu vaccine and febrile seizures (no increased risk)
  - Rotarix and intussusception (label change)
  - Dabigatran and bleeding (no increased risk)
  - Olmesartan and sprue-like enteropathy (label change)
- Over 70 peer-reviewed articles
- Over 50 methods reports / white papers
- Thousands of unique queries and comparisons contributing to over 140 formal assessments

www.mini-sentinel.org



U.S. Food and Drug Administration
Protecting and Promoting *Your* Health

A to Z Index | Follow FDA | FDA Voice Blog
SEARCH
Search the FDA Archive

Home | Food | Drugs | Medical Devices | Radiation-Emitting Products | Vaccines, Blood & Biologics | Animal & Veterinary | Cosmetics | Tobacco Products

Vaccines, Blood & Biologics

Home › Vaccines, Blood & Biologics › Safety & Availability (Biologics)

- **Safety & Availability (Biologics)**
- Biologics Product Shortages Q&A
- Recalls (Biologics)
- Biologic Product Shortages
- Report a Problem to the Center for Biologics Evaluation & Research
- Biologic Product Security
- Pandemics
- Blood Safety & Availability
- Tissue Safety & Availability
- Vaccine Safety & Availability
- HIV Home Test Kits

Resources for You

- 2013 Safety and Availability Communications

FDA Releases Final Study Results of a Mini-Sentinel Postlicensure Observational Study of Rotavirus Vaccines and Intussusception

FDA Safety Communication — June 13, 2013

FDA Releases Final Study Results of a Mini-Sentinel Postlicensure Observational Study of Rotavirus Vaccines and Intussusception

FDA Approves Required Revised Labeling for RotaTeq Based on the Study Results

Purpose: To inform the public and healthcare providers that FDA is releasing final study results from a Mini-Sentinel postlicensure observational study of intussusception (a form of bowel obstruction) after vaccination with RotaTeq (Merck and Co., Inc.) and Rotarix (GlaxoSmithKline Biologicals).

RotaTeq and Rotarix are vaccines for the prevention of rotavirus gastroenteritis in infants 6 weeks to 32 weeks of age (RotaTeq) and infants 6 weeks to 24 weeks of age (Rotarix). The study was conducted in Mini-Sentinel's Postlicensure Rapid Immunization Safety Monitoring (PRISM) program, the largest vaccine safety surveillance program in the Unit

Label change

FDA has approved n and Patient Information for RotaTeq as a result of the new safety d information was added to the Highlights, the existing intussusception subsection of the Warnings and Precautions section, and the Post-Marketing Experience section of the Full Prescribing Information, as well as to the Patient Information. The Mini-Sentinel PRISM study is the largest study of intussusception after rotavirus vaccines to date and identified an increased risk of intussusception in the 21 day time period after the first dose of RotaTeq, with most cases occurring in the first 7 days after vaccination. No increased risk was found after the second or third doses. These findings translate into 1 to 1.5 additional cases of intussusception per 100,000 first doses of RotaTeq.

The data from the Mini-Sentinel PRISM study regarding the risk of intussusception following the use of Rotarix were inconclusive. Based on this study, no changes were made to the Prescribing Information or to the Patient Information for Rotarix. However, based on data from an observational study previously conducted in Mexico, it is estimated that 1 to 3 additional cases of intussusception would occur per 100,000 vaccinated infants in the United States within 7 days following the first dose of Rotarix. In September 2012, FDA announced that it had approved revisions to the Prescribing Information and to the Patient Information for Rotarix to include these results from the study in Mexico.

http://www.fda.gov/BiologicsBloodVaccines/SafetyAvailability/ucm356758.htm

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Intussusception Risk after Rotavirus Vaccination in U.S. Infants

W. Katherine Yih, Ph.D., M.P.H., Tracy A. Lieu, M.D., M.P.H., Martin Kulldorff, Ph.D., David Martin, M.D., M.P.H., Cheryl N. McMahill-Walraven, M.S.W., Ph.D., Richard Platt, M.D., Nandini Selvam, Ph.D., M.P.H., Mano Selvan, Ph.D., Grace M. Lee, M.D., M.P.H., and Michael Nguyen, M.D.

Yih, N Engl J Med. 2014;370:503

"This assessment […used…] FDA's Mini-Sentinel pilot..."

www.fda.gov/Drugs/DrugSafety/ucm326580.htm; Nov 2, 2012

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

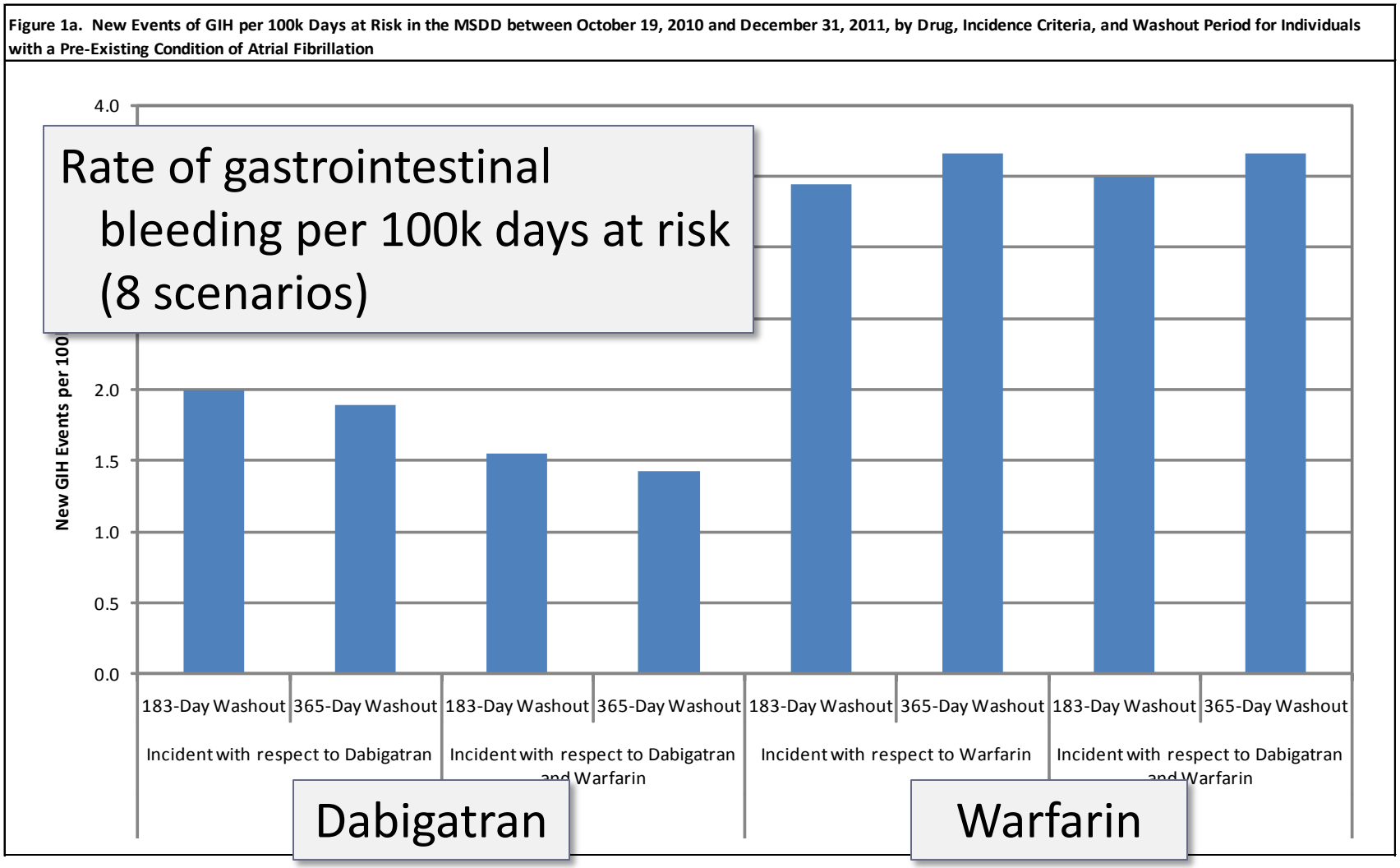
Dabigatran and Postmarketing Reports of Bleeding

Mary Ross Southworth, Pharm.D., Marsha E. Reichman, Ph.D., and Ellis F. Unger, M.D.

"In the months following the approval of the oral anticoagulant dabigatran ... in October, 2010, the FDA received through the FDA Adverse Event Reporting System many reports of serious and fatal bleeding events associated with use of the drug."

N Engl J Med 2013. DOI: 10.1056/NEJMp1302834

Output



http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Modular-Program-Report_%20MSY3_MPR41_Dabigatran-Warfarin-GIH-ICH_Part-1.pdf

The NEW ENGLAND JOURNAL of MEDICINE

Perspective

Dabigatran and Postmarketing Reports of Bleeding

Mary Ross Southworth, Pharm.D., Marsha E. Reichman, Ph.D., and Ellis F. Unger, M.D.

Intracranial and Gastrointestinal Bleeding Events in New Users of Dabigatran and Warfarin from the Mini-Sentinel Distributed Database, October 2010 through December 2011.*

Analysis	Dabigatran			Warfarin		
	No. of Patients	No. of Events	Incidence *no. of events/ 100,000 days at risk*	No. of Patients	No. of Events	Incidence *no. of events/ 100,000 days at risk*
Gastrointestinal hemorrhage						
Analysis with required diagnosis of atrial fibrillation	10,599	16	1.6	43,541	160	3.5
Sensitivity analysis without required diagnosis of atrial fibrillation	12,195	19	1.6	119,940	338	3.1
Intracranial hemorrhage						
Analysis with required diagnosis of atrial fibrillation	10,587	8	0.8	43,594	109	2.4
Sensitivity analysis without required diagnosis of atrial fibrillation	12,182	10	0.9	120,020	204	1.9

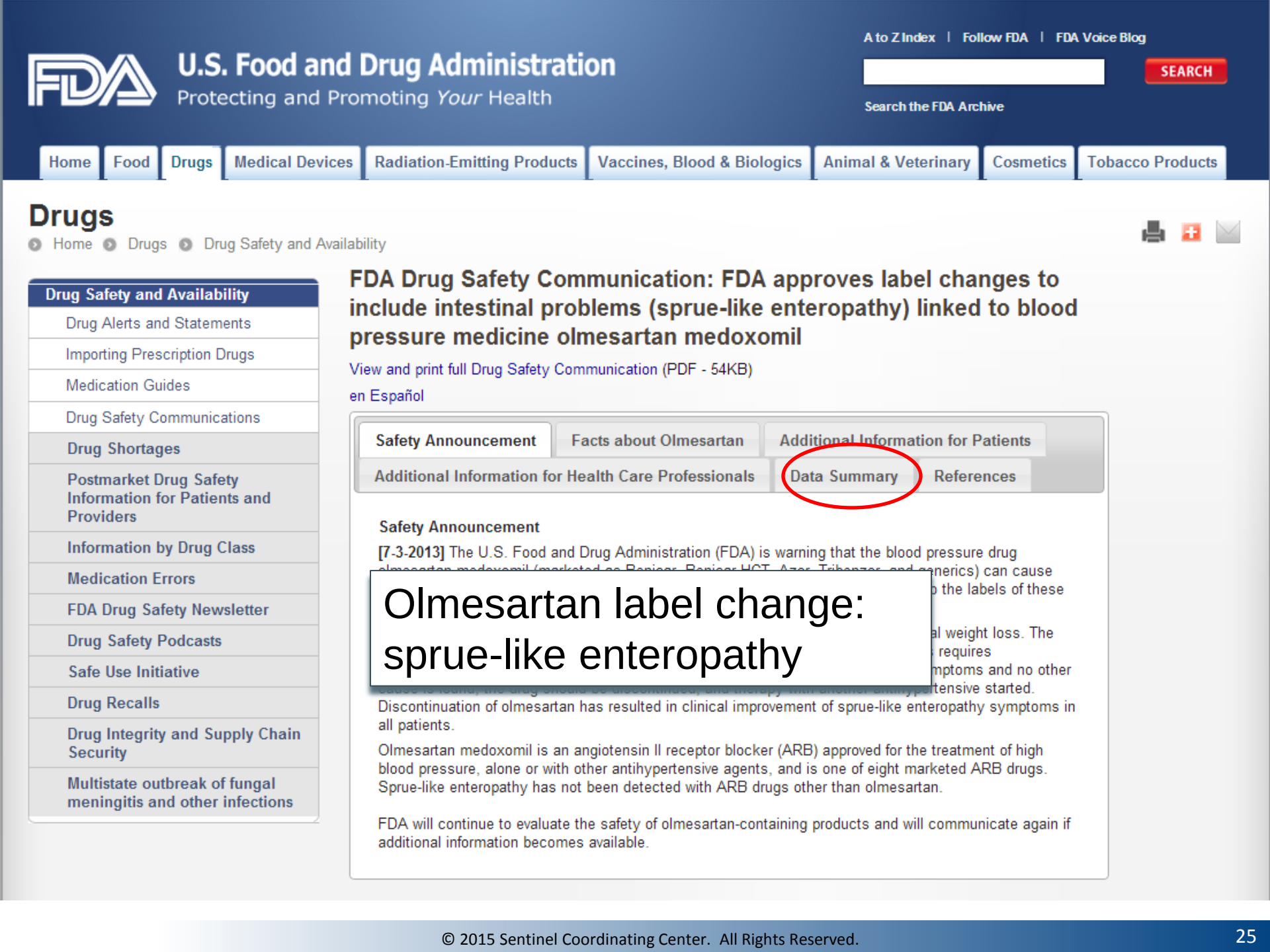
Protocol-based assessment – Dabigatran

MINI-SENTINEL MEDICAL PRODUCT ASSESSMENT

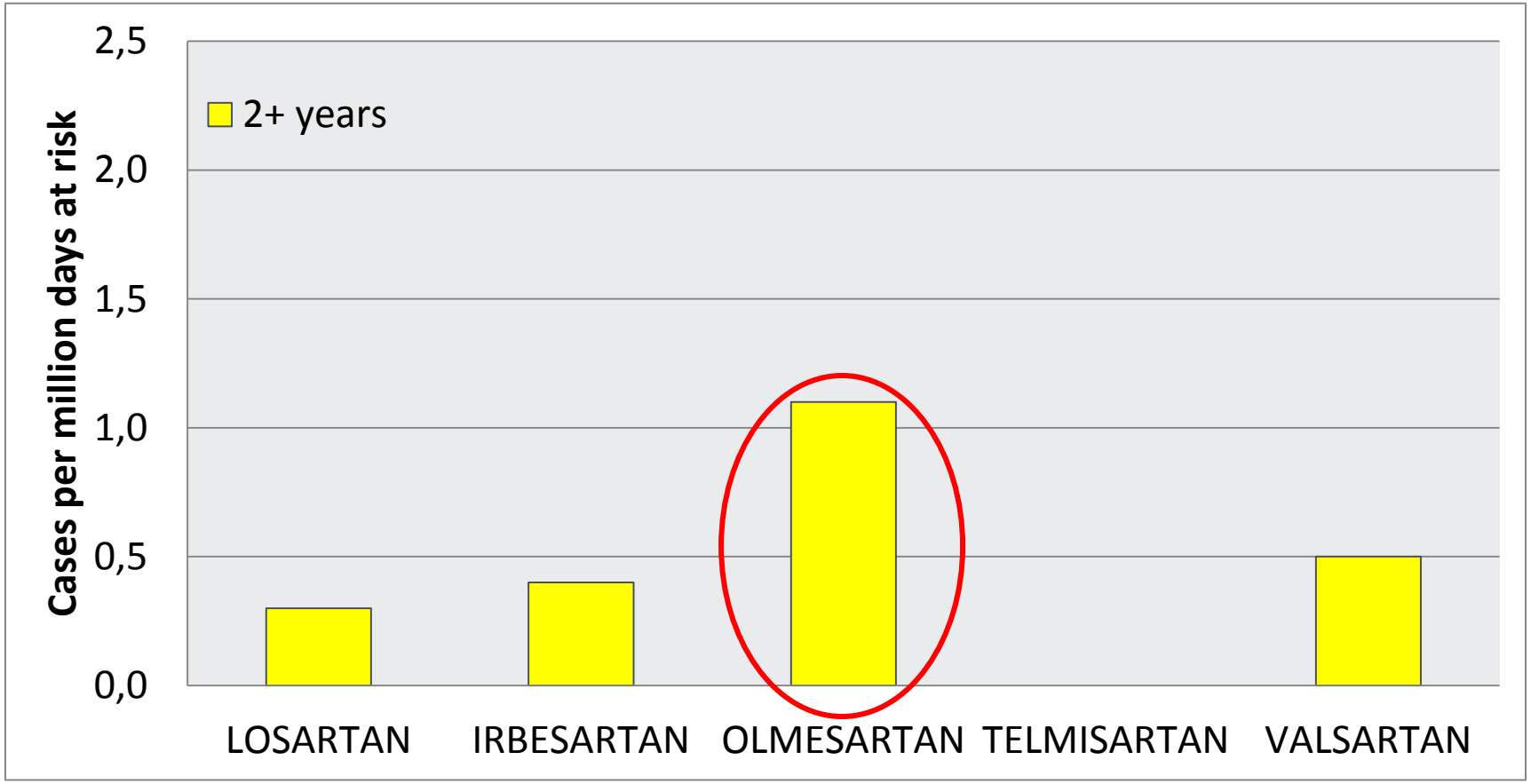
A PROTOCOL FOR ASSESSMENT OF DABIGATRAN

Prepared by: Alan S. Go, MD[1], Daniel Singer, MD[2], T. Craig Cheetham, PharmD MS[3], Darren Toh, ScD[4], Marsha Reichman, PhD[5], David Graham, MD MPH[5], Mary Ross Southworth, PharmD[6], Rongmei Zhang PhD[7], Monika Houstoun, PharmD[5], Yu-te Wu, PhD[7], Katrina Mott, MS[5], Joshua Gagne, PharmD ScD[8]

http://www.mini-sentinel.org/work_products/Assessments/Mini-Sentinel_Protocol-for-Assessment-of-Dabigatran.pdf



ARBs and celiac disease: 2+ years



Cases	9	1	5	0	7
New users	25,045	2,721	4,419	1,124	13,925

The NEW ENGLAND JOURNAL of MEDICINE

Mini-Sentinel and Regulatory Science — Big Data Rendered Fit and Functional

Bruce M. Psaty, M.D., Ph.D., and Alasdair M. Breckenridge, M.D.

In medicine, "big data" come in many forms. With the financial incentives provided by Medicare and Medicaid for the "meaningful

functions of billing and clinical care. If, as Nate Silver suggests in *The Signal and the Noise*, "Most of

tional Heart, Lung, and Blood Institute have produced data sets with millions of genetic variants for each participant, encouraged the development of consortia with hundreds of thousands of study participants, and resulted in discoveries about the genetic origins

that can contribute meaningfully to the health of the public.

One model is the Mini-Sentinel. A pilot project of the Sentinel Initiative of the Food and Drug Administration (FDA), the Mini-Sentinel has created a nationwide system that uses electronic data

"The Mini-Sentinel ' provides an essential public health service. The current configuration — the data model, the methods development, and the investigative team — represents an impressive achievement..

Psaty. N Engl J Med 2014;370:2165

Sentinel Routine Querying Tools

Distributed querying in Sentinel

- Pre-tabulated summary tables
- Reusable modular programs (SAS)
- Custom SAS code
  - Protocol-Based Assessments
  - Other de novo or ad hoc programs
- Qualitative, survey requests

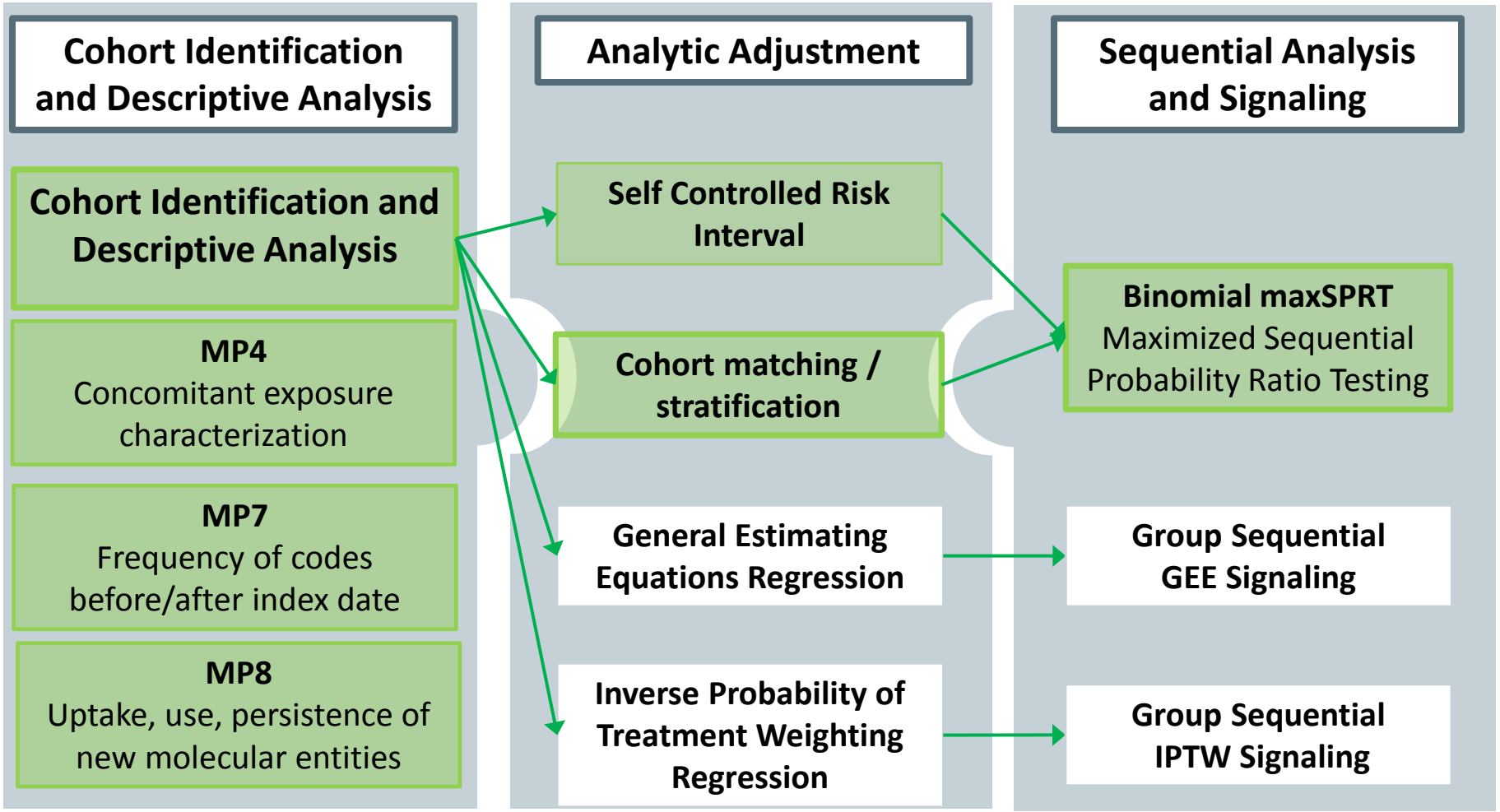
Distributed querying in Sentinel

- Pre-tabulated summary tables
- Reusable modular programs (SAS)

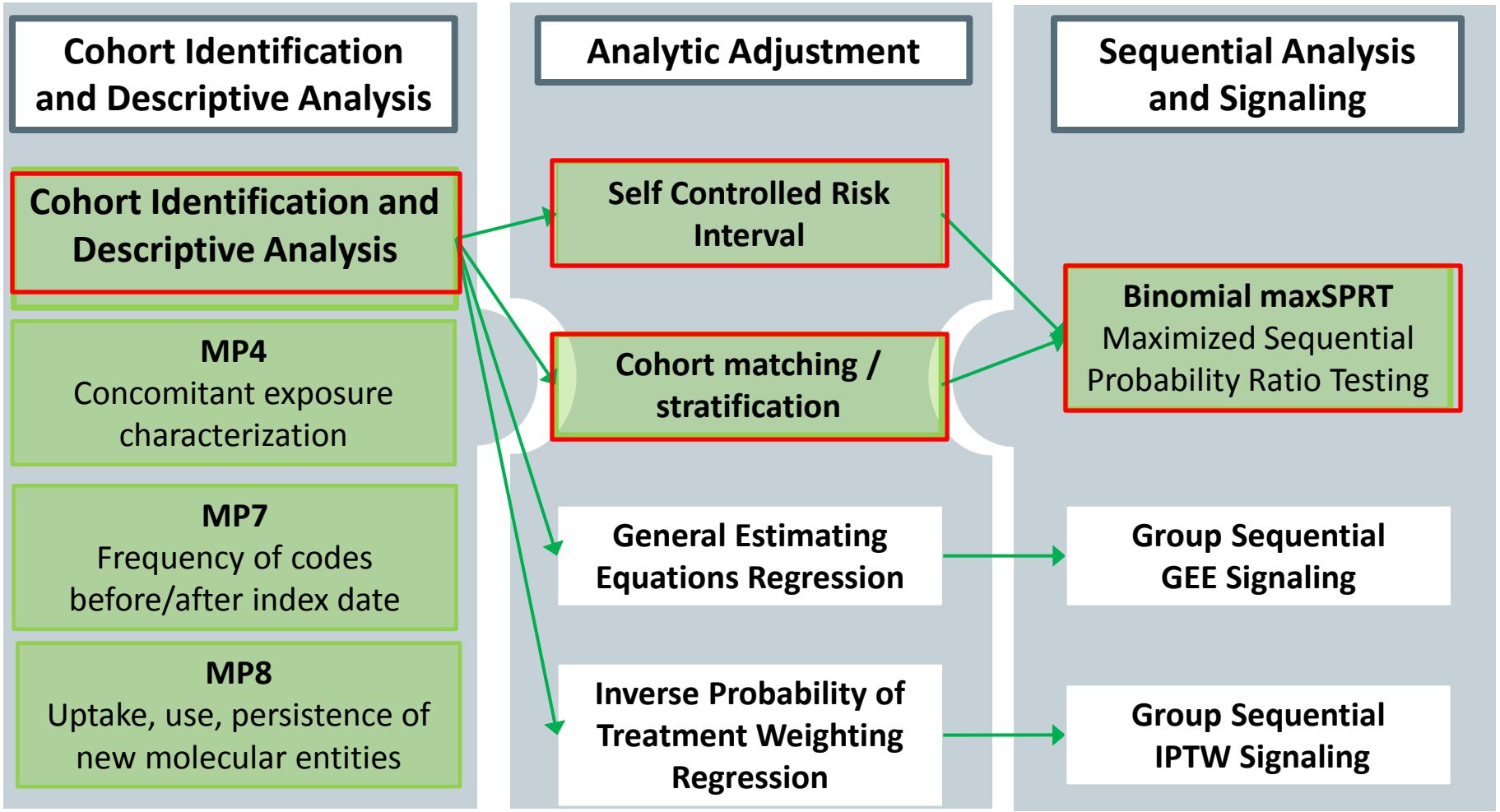
Rapid response queries

- Custom SAS code
  - Protocol-Based Assessments
  - Other de novo or ad hoc programs
- Qualitative, survey requests

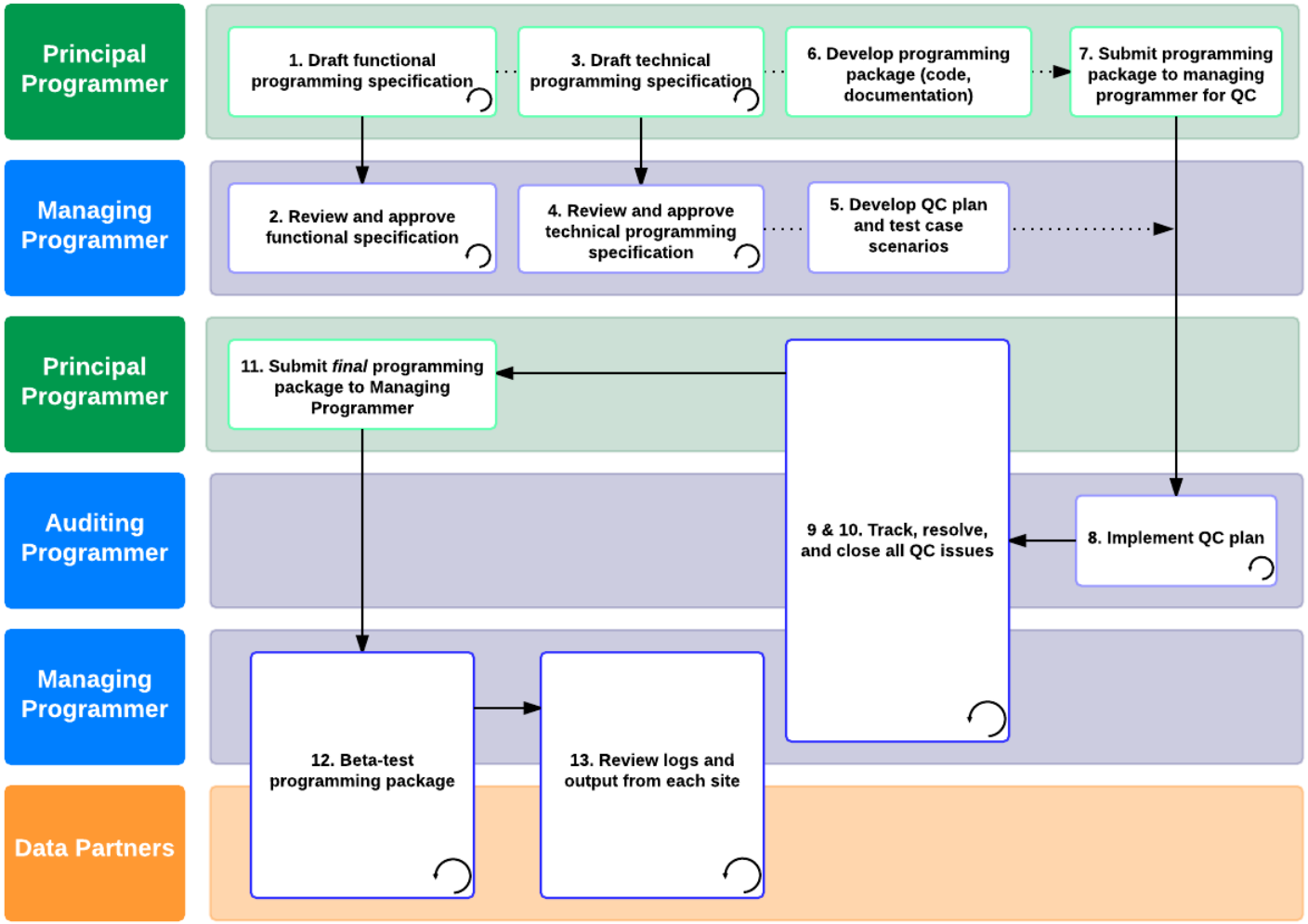
Query tool architecture



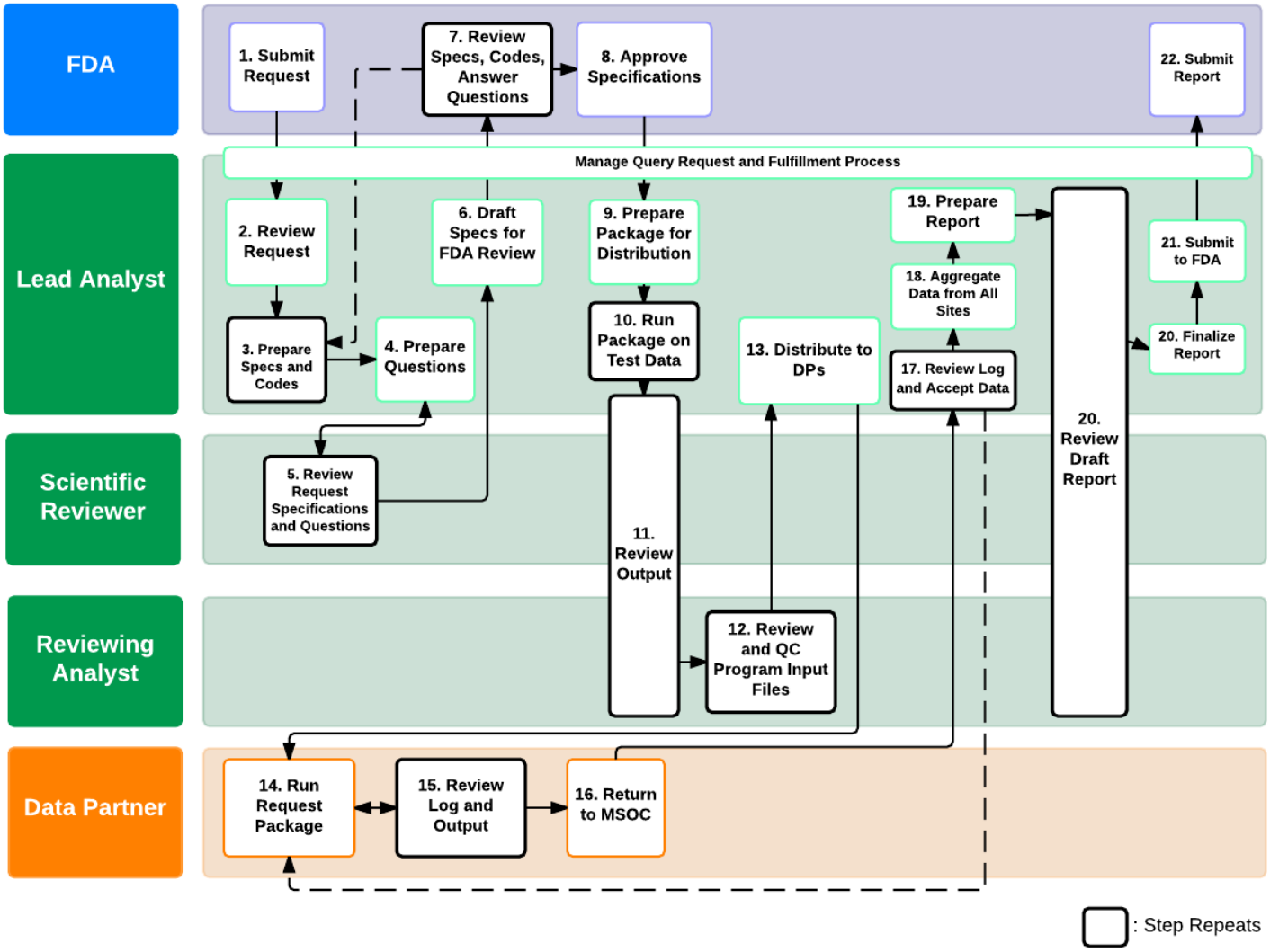
Query tool architecture



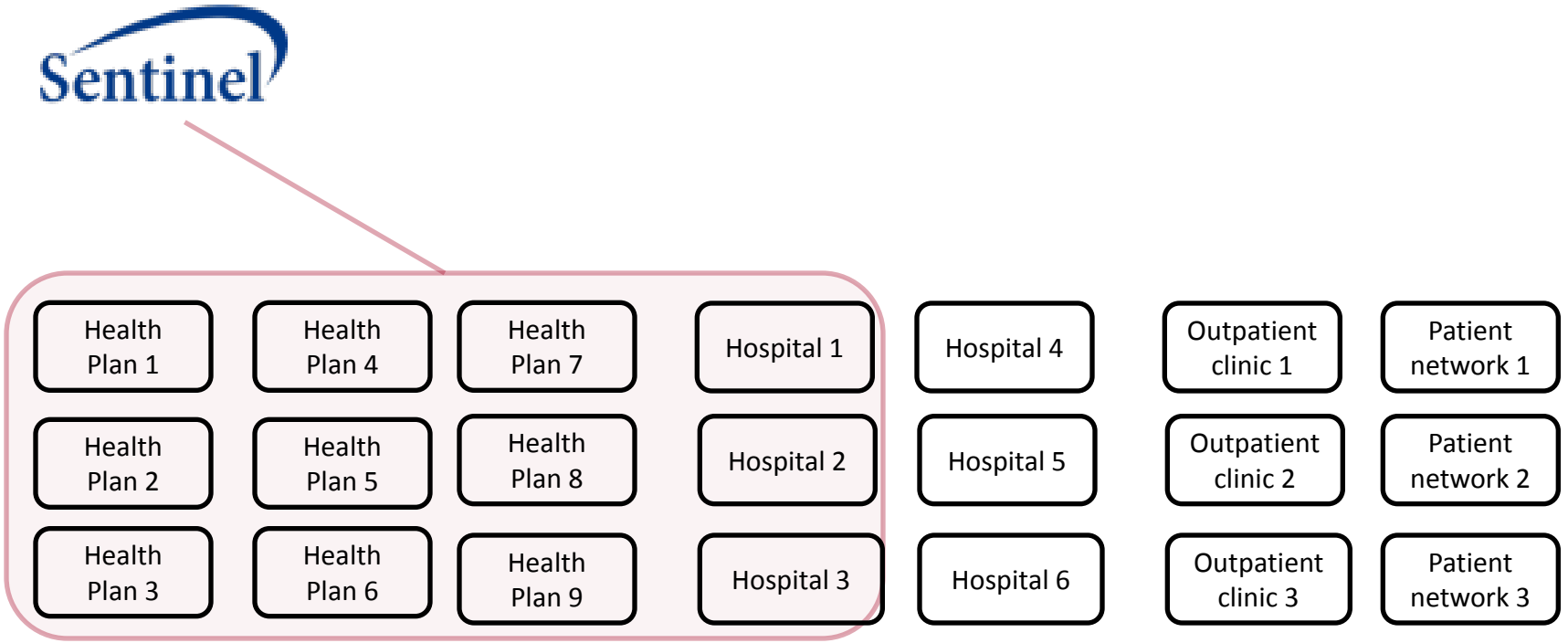
New Program Development SOP



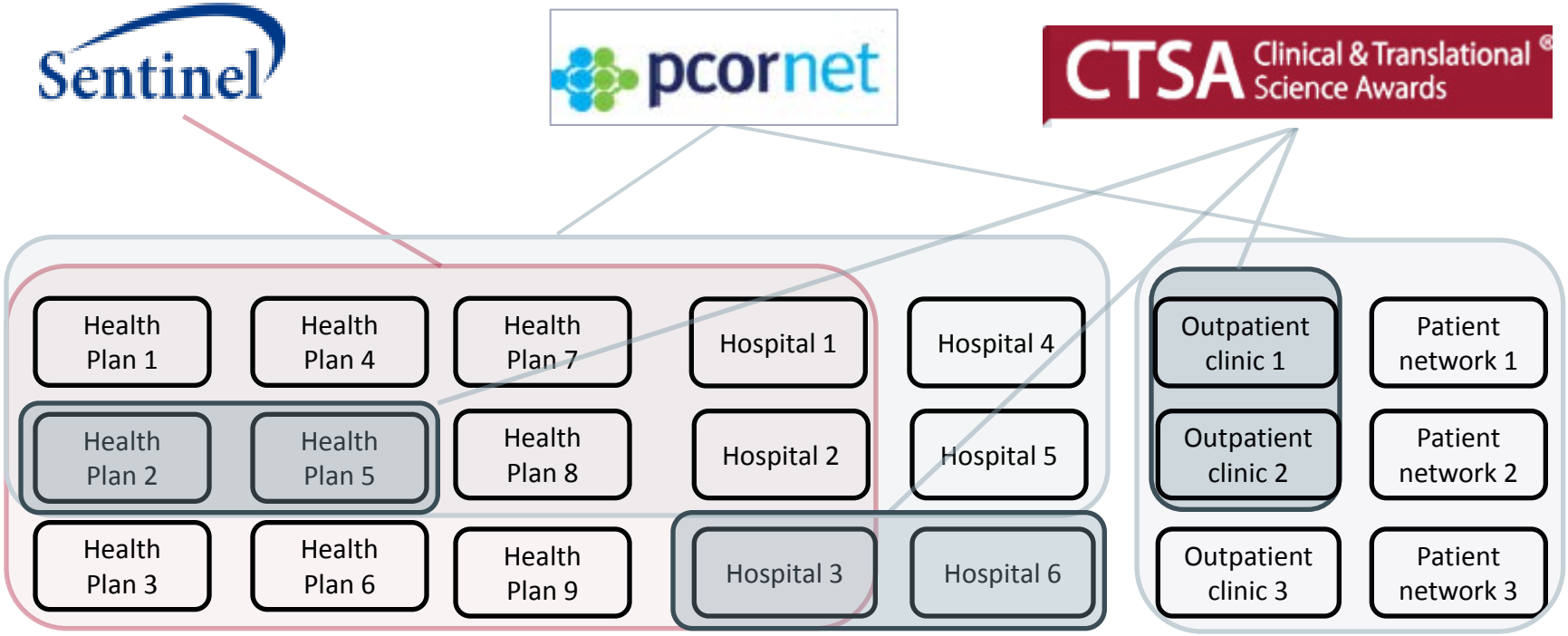
Query Fulfillment SOP



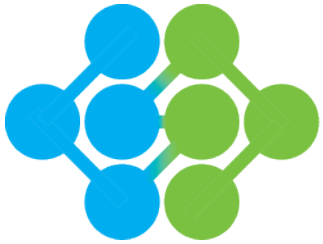
Multiple networks sharing infrastructure



Critical Partners in a National Infrastructure



- Each organization can participate in multiple networks
- Each network controls its governance and coordination
- Networks share infrastructure, data curation, analytics, lessons, security, software development
- Other potential partners: disease or treatment-specific networks



pcornet℠

The National Patient-Centered Clinical Research Network

Our national clinical research system is well-intentioned but flawed

We are not generating the evidence we need to support the healthcare decisions that patients and their doctors have to make every day.

- High percentage of decisions are not supported by evidence
- Health outcomes and disparities are not improving
- Current clinical research system faces several problems:

Too slow

Too expensive

Doesn't answer questions that matter most to people

pcornet

PCORnet embodies a “community of research” by uniting people, clinicians & systems

20
Patient-Powered Research Networks (**PPRNs**)

+

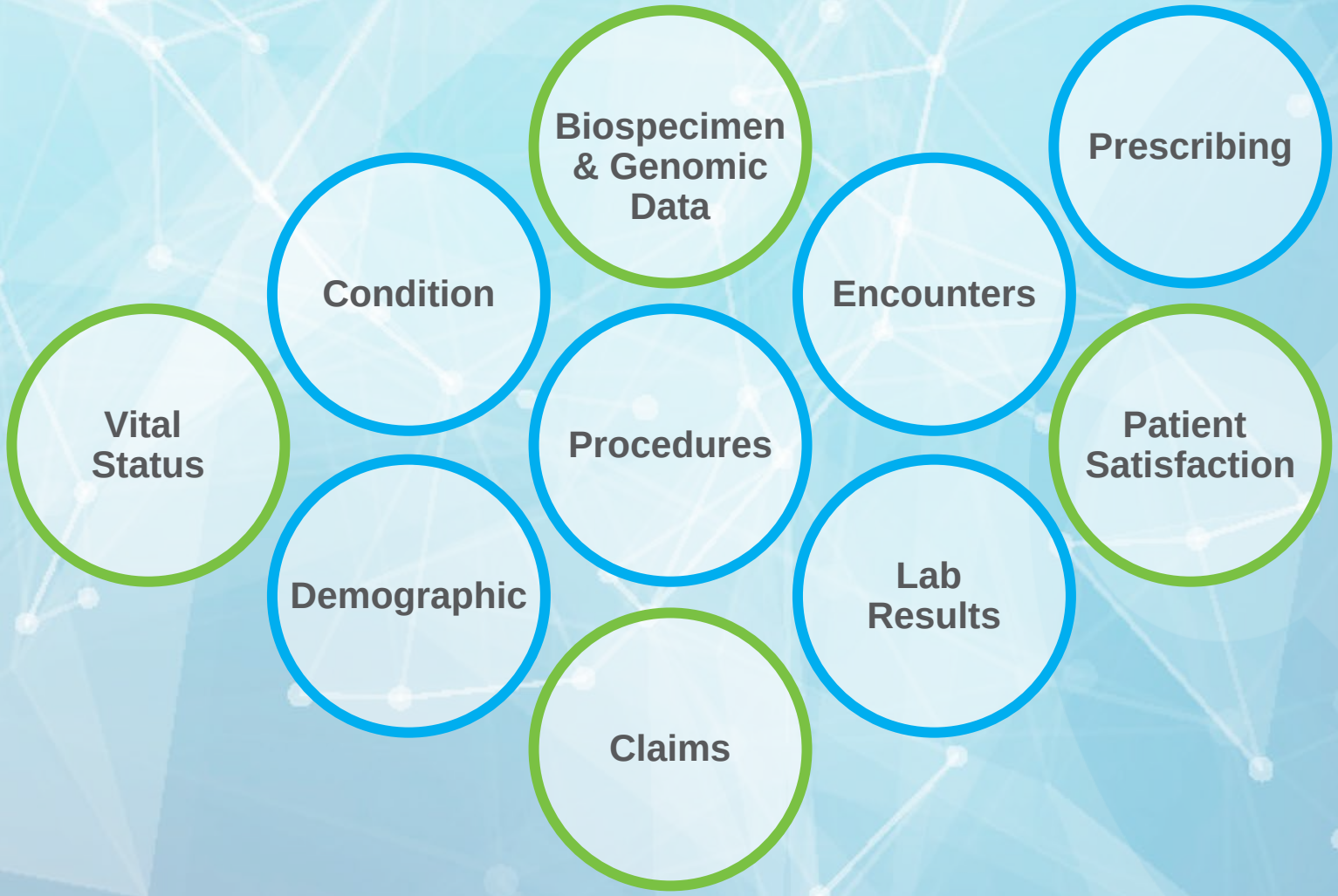
13
Clinical Data Research Networks (**CDRNs**)

=

PCORnet
A national infrastructure for people-centered clinical research

pcornet

Underpinned by a Common Data Model

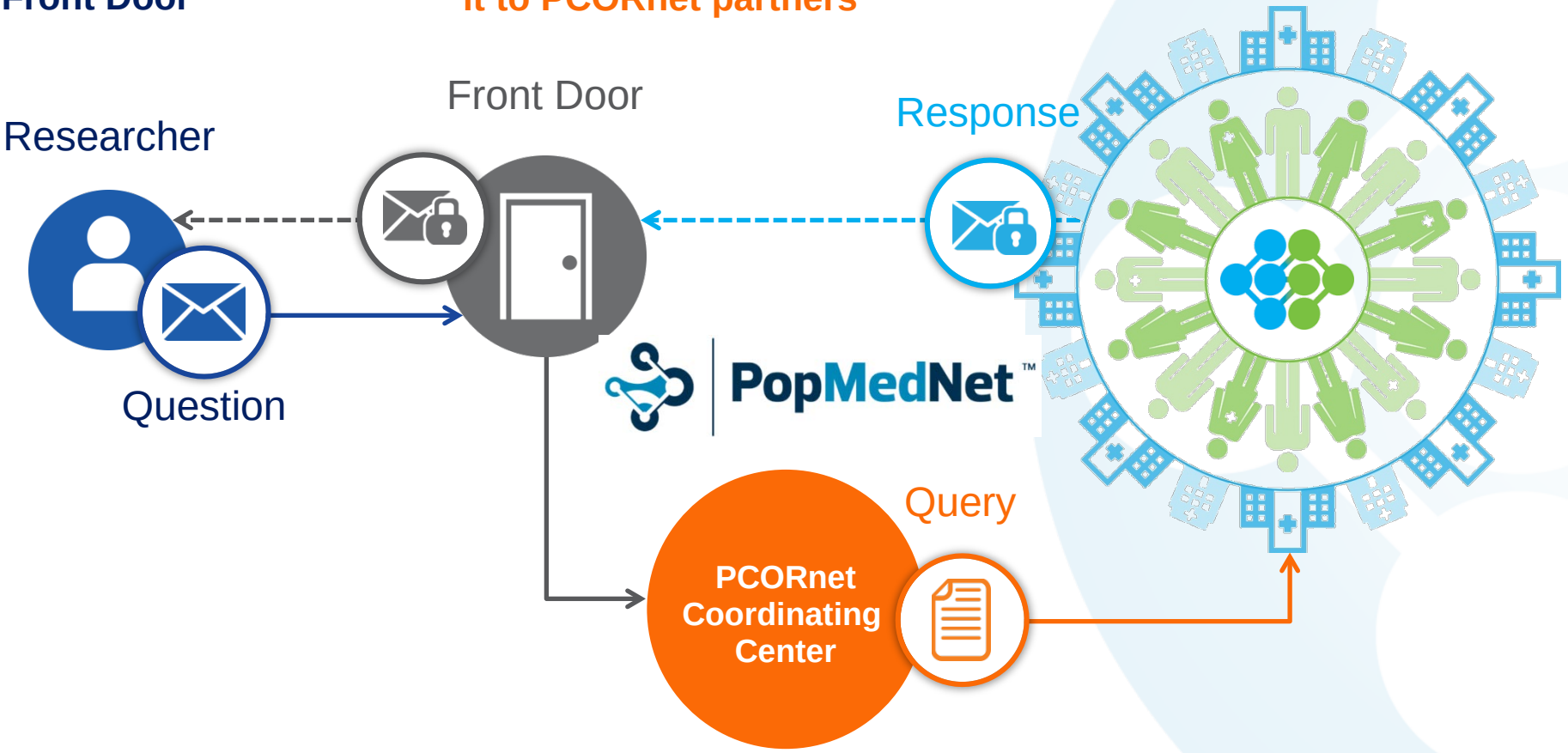


PCORnet distributed querying

The Researcher sends a question to the PCORnet Coordinating Center through the Front Door

The Coordinating Center converts the question into a query with an underlying executable code, and sends it to PCORnet partners

PCORnet partners review the query and provide a response, which is sent back through the Front Door to the Researcher



PCORnet could be for many kinds of research

Pre-research

- ☐ Feasibility queries
- ☐ Engagement
- ☐ Match-making

Observational studies

- ☐ Cross-sectional
- ☐ Epidemiology
- ☐ Health services
- ☐ Comparative effectiveness or safety

Interventional studies

- ☐ Clinical trials
- ☐ Pragmatic randomized clinical trials
  - e-Identification
  - e-Consent
  - e-Randomization
  - e-Follow-up
- ☐ Cluster randomization

pcornet

Issues to Consider for Distributed Querying

- Complexities to consider when developing a stable, distributed querying infrastructure
  - Common Data Model conformance
  - Local system implementation variability
    - Software and hardware
    - Computing environments
    - IT environments and configuration
    - Local expertise
  - Source data size
  - Programming efficiency
  - Transparency

pcornet

PCORnet Query Types

- Menu-Driven Queries
- SAS Program Package Queries

pcornet

PCORnet Query Types: Menu-Driven Queries

- Simple interface for query creation incorporated into the PCORnet Query Tool
- Query securely distributed to sites for execution against the local RDBMS in PCORnet CDM format
- Query generates aggregate data for local review and secure return
- Asynchronous by design

pcornet

Menu-Driven Query: From Distribution to Receipt

- Menu-Driven Queries undergo a number of conversions throughout a full query cycle
- Json is the file type used by PopMedNet to enforce a standardized query and response format.
  - Scalability across different data models and database management systems
  - Easier data visualization
- Request information always transmitted between the web portal and the DataMart Client through encrypted TLS pathways.

pcornet

Technical Menu-Driven Query Details

1. A request.json file is generated when a query is composed in the web portal
2. Request.json is pulled to the DataMart Client by DataMart Administrator
3. DataMart Administrator clicks Run
4. Request.json is converted to a Linq statement
5. Linq is fed through an implementation of Microsoft Entity Framework
6. Entity Framework translates Linq to one of three SQL flavors (SQL Server, Oracle, Postgres)
7. SQL is executed against the DataMart's relational database

Technical Query Request Details: Query Response

1. Result set is converted to a response.json file in the DataMart Client
2. Response.json is sent back to the web portal when DataMart Administrator clicks Upload
3. Response.json is converted back to the raw result set each time a user views results on the web portal

PCORnet Query Types: SAS Queries

- SAS code package designed to execute against SAS datasets in PCORnet CDM format
- Distributed via the PCORnet Query Tool
- Downloaded and executed locally
- Response uploaded and returned via Query Tool

pcornet

Recent PCORnet Query (SAS)

- Distributed to at least one site at **13 out of 13 CDRNs**
- Output received from at least one site at **9 out of 13 CDRNs**

pcornet

Different ecosystems responding

DM	Operating System	SAS Version	SAS Views/Datasets	Run Time	# Days to Response	Median Patient Count (GT or LT)
1	Windows Server 2012	9.4	Views	56 min.	2	Greater
2		9.4	Datasets	10 min.	9	Less
3			Views	6 min.	8	Less
4		9.3	Views	5.5 hr.	3	Greater
5	Windows 7 Enterprise		Datasets	6 min.	10	Less
6	Windows 2008	9.4	Datasets	35 min.	9	Greater
7		9.3	Datasets	4 hr.	11	Greater
8			Views	11 min.	4	Less
9	Linux	9.4	Views	21 min.	11	Less

Different ecosystems responding

DM	Operating System	SAS Version	SAS Views/Datasets	Run Time	# Days to Response	Median Patient Count (GT or LT)
10	Windows 7 Enterprise	9.4	Datasets	34 min.	2	Less
11		9.4	Views	40 min.	7	Less
12	Windows 7 Enterprise	9.4	Views	18 min.	24	Greater
13		9.3	Datasets	21 min.	1	Less
14	Windows 8.1	9.3	Datasets	41 min.	2	Greater
15		9.4	Datasets	41 min.	10	Greater
16		9.3	Datasets	2.5 hr.	4	Greater
17		9.2	Views	39 min.	1	Greater
18			Datasets	<1 min.	3	Less

Concluding thoughts

- Start small and keep it simple and base work on focused use cases
- Research networks are mostly about governance and trust
- Bridge the **Informatics-Research** divide
- Training Data Scientists is critical – especially in research methods
- Extract critical data elements, standardize format, and leave data where it was collected
  - Don't map data elements unless you have to
  - Analytic team should make analytic decisions **at the time of analysis**
- Do analysis behind firewalls, move as little information as possible
- Be careful not to create a data sharing network versus a research network (unless that is your goal)

pcornet

Concluding thoughts

- Match the data to the research question
- Honestly assess what types of research the data can support
  - Public health surveillance of infectious disease?
  - Epidemiology?
  - Comparative safety?
  - Comparative effectiveness?
  - Clinical trial enrollment? Outcomes?
- Think creatively about research designs
  - Pragmatic clinical trials embedded in health systems
- Be humble about what the data can do – don't overpromise
- Have fun – Success is possible

pcornet

Distributed data network architecture: Lessons from PCORnet and FDA Sentinel

Thank You

Jeffrey Brown, PhD

Associate Professor

Harvard Pilgrim Health Care Institute/ Harvard Medical School

July 12, 2016