

# The Sentinel System: The Case for Reusable Tools and Analysis Ready Data in Distributed Data Networks

Richard Platt MD, MS  
Harvard Pilgrim Health Care Institute and  
Harvard Medical School  
for the Sentinel Investigators  
August 29, 2017

# Disclosures

- None related to this presentation

## Sentinel's charge

Assess the use, safety, and effectiveness of regulated medical products by using electronic healthcare data plus other resources

Create data, informatics, and methodologic capabilities to support these activities

**Speedily!**

# Sentinel partner organizations

## Lead – HPHC Institute

DEPARTMENT OF POPULATION MEDICINE



## Data and scientific partners



## Scientific partners



# An Ideal Distributed Network Should...

- Accommodate many data holders' data
- Incorporate new kinds of data as they become available
- Maximize local control of data and uses
- Minimize data exchange
- Include local experts in study design and interpretation
- Allow a study protocol to be implemented identically and efficiently across the network
- Support standardized, reusable components

# An Ideal Distributed Network Should...

- Accommodate many data holders' data
- Incorporate new kinds of data as they become available
- Maximize local control of data and uses
- Minimize data exchange
- Include local experts in study design and interpretation
- **Allow a study protocol to be implemented identically and efficiently across the network**
- **Support standardized, reusable components**

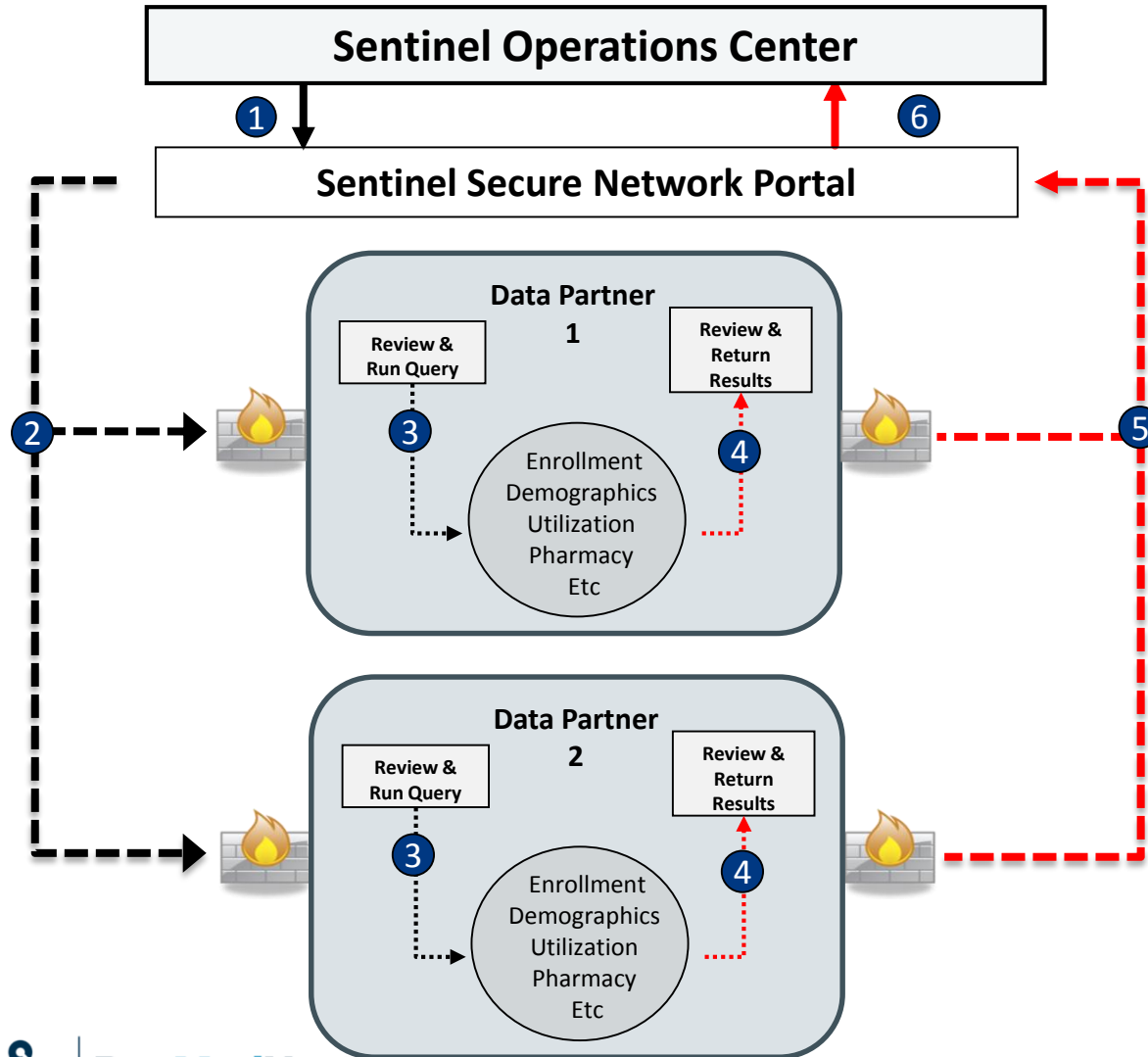
# Sentinel distributed database\*

Population with well-defined person-time for which most medically-attended events are known

- 425 million person-years of observation time
- 43 million people currently accruing new data
- 5.9 billion dispensings
- 7.2 billion unique encounters
- 42 million people with  $\geq 1$  laboratory test result

\* As of January 2017

# Sentinel distributed analysis



- 1- User creates and submits query
- 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 and returned



# Three ways to address questions

## Custom Programs



- Analysis as specified
- Custom inputs, custom output
- Longer execution

# 10 Selected protocol based assessments

- CDER
  - Dabigatran and several outcomes
  - Metabolic effects of 2<sup>nd</sup> generation antipsychotics in youth
  - Diabetes drugs and acute myocardial infarction
  - IV Iron and anaphylaxis
- CBER
  - IV Immune Globulin and thromboembolic events
  - Gardasil and venous thromboembolism
  - Influenza vaccines and pregnancy outcomes
  - Gardasil 9 and Pregnancy Outcomes
  - Prevnar 13 and Kawasaki disease
  - Blood components and Transfusion-Related Lung Injury (TRALI)

# Three ways to address questions

## Custom Programs



- Analysis as specified
- Custom inputs, custom output
- Longer execution

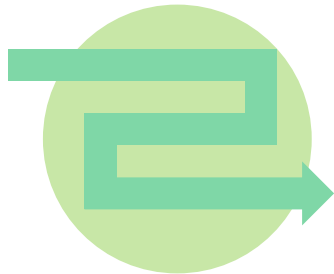
## Routine Analytic Framework (RAF)



- Off-the-shelf query “templates”
- Standard inputs, standard output
- Quick execution

# Three ways to address questions

## Custom Programs



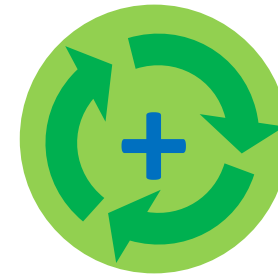
- Analysis as specified
- Custom inputs, custom output
- Longer execution

## Routine Analytic Framework (RAF)



- Off-the-shelf query “templates”
- Standard inputs, standard output
- Quick execution

## RADaR: Rapid Analytic Development and Response: RAF + custom code

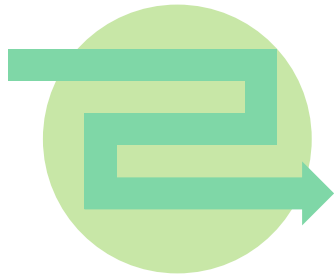


- Hybrid approach: custom code leveraging RAF
- Standard inputs, custom output

# Three ways to address questions

## Rapid Analyses

### Custom Programs



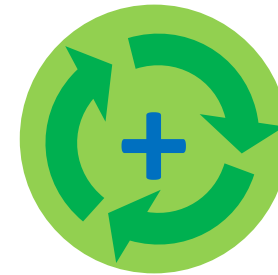
- Analysis as specified
- Custom inputs, custom output
- Longer execution

### Routine Analytic Framework (RAF)



- Off-the-shelf query “templates”
- Standard inputs, standard output
- Quick execution

### RADaR: Rapid Analytic Development and Response: RAF + custom code

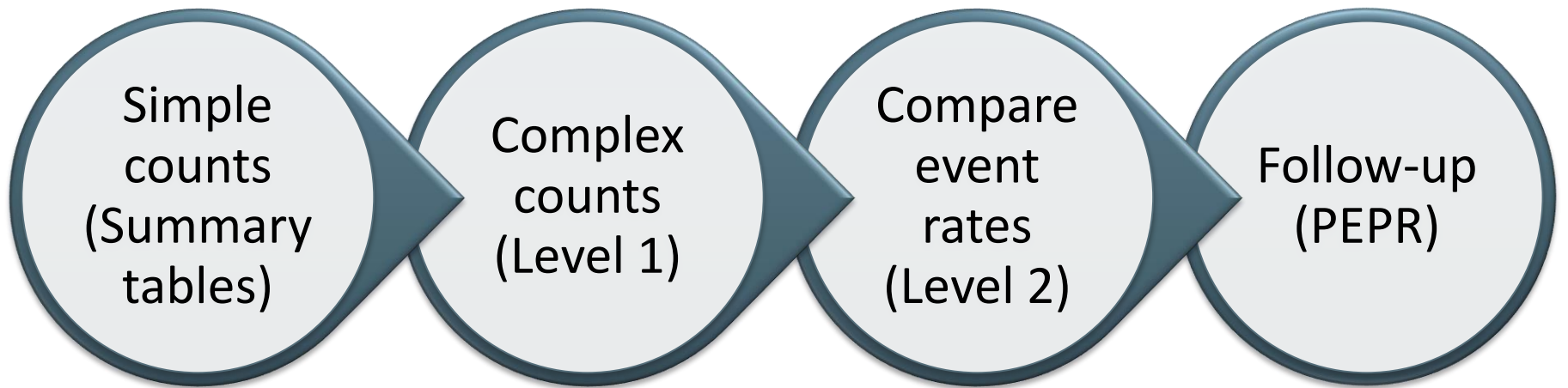


- Hybrid approach: custom code leveraging RAF
- Standard inputs, custom output

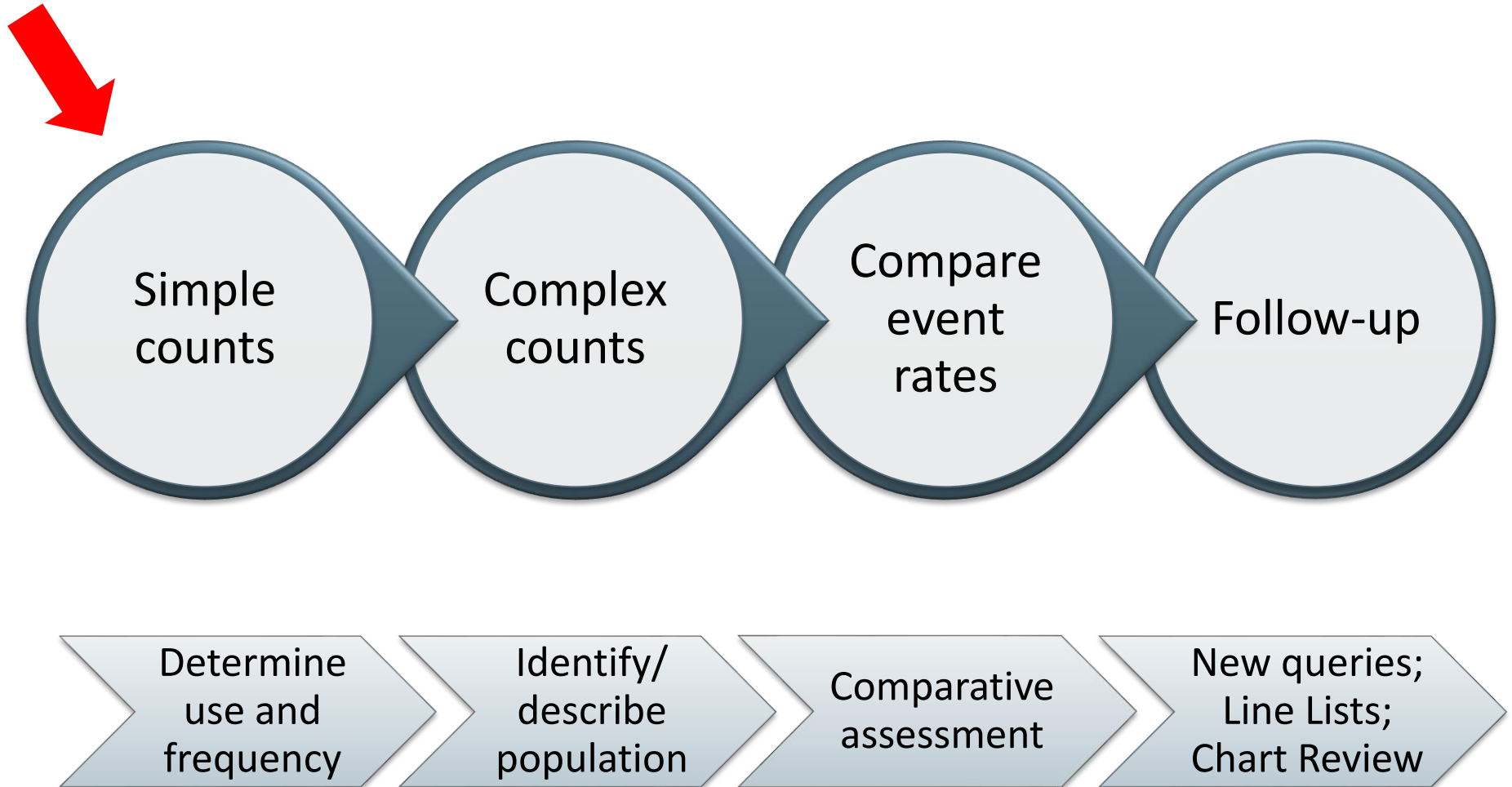
# Routine Analytic Framework tools

- Validated, flexible, and reusable analytic programs
  - Meet FDA's needs for responsiveness, data quality, reproducibility, and transparency
  - Run efficiently against the Sentinel Common Data Model and generate standardized output
  - Meets needs of Data Partners with diverse technical, data governance, security, and confidentiality requirements

# Rapid analysis querying sequence



# Querying sequence





# Simple counts (summary table queries)

- Counts of (new) users with exposure or condition

## Simple counts (summary table queries)

- Counts of (new) users with exposure or condition
- Example: Dispensing of evolocumab (PCSK9 inhibitor) without prior dispensing during preceding 180 days, by age, sex, and year

Age	2015		2016 (partial)	
	Male	Female	Male	Female
<44	5	2	61	28
45-64	85	61	569	335
65-74	42	35	231	222
75+	11	20	101	149
<b>TOTAL</b>	<b>261</b>		<b>1,696</b>	

## Simple counts (summary table queries)

- Counts of (new) users with exposure or condition
- Example: Dispensing of evolocumab (PCSK9 inhibitor) without prior dispensing during preceding 180 days, by age, sex, and year

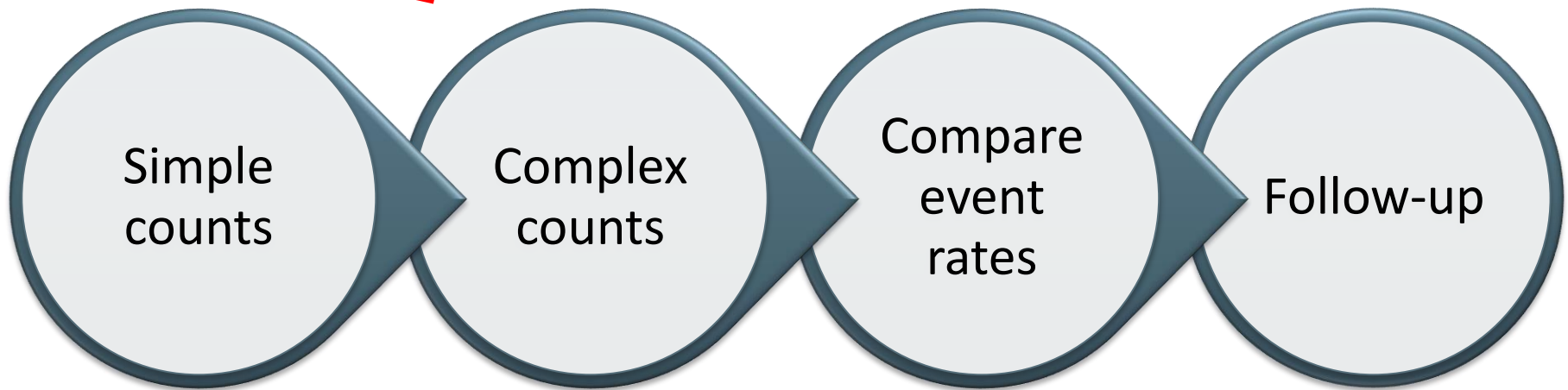
Age	2015		2016 (partial)	
	Male	Female	Male	Female
<44	5	2	61	28
45-64	85	61	569	335
65-74	42	35	231	222
75+	11	20	101	149
<b>TOTAL</b>	<b>261</b>		<b>1,696</b>	

- 49 such queries / 291 scenarios in 2016

## Summary tables at 33<sup>rd</sup> ICPE

- Early Post-Approval Surveillance of New Molecular Entity Uptake in the Sentinel Distributed Database
  - Today: Poster session B

# Querying sequence



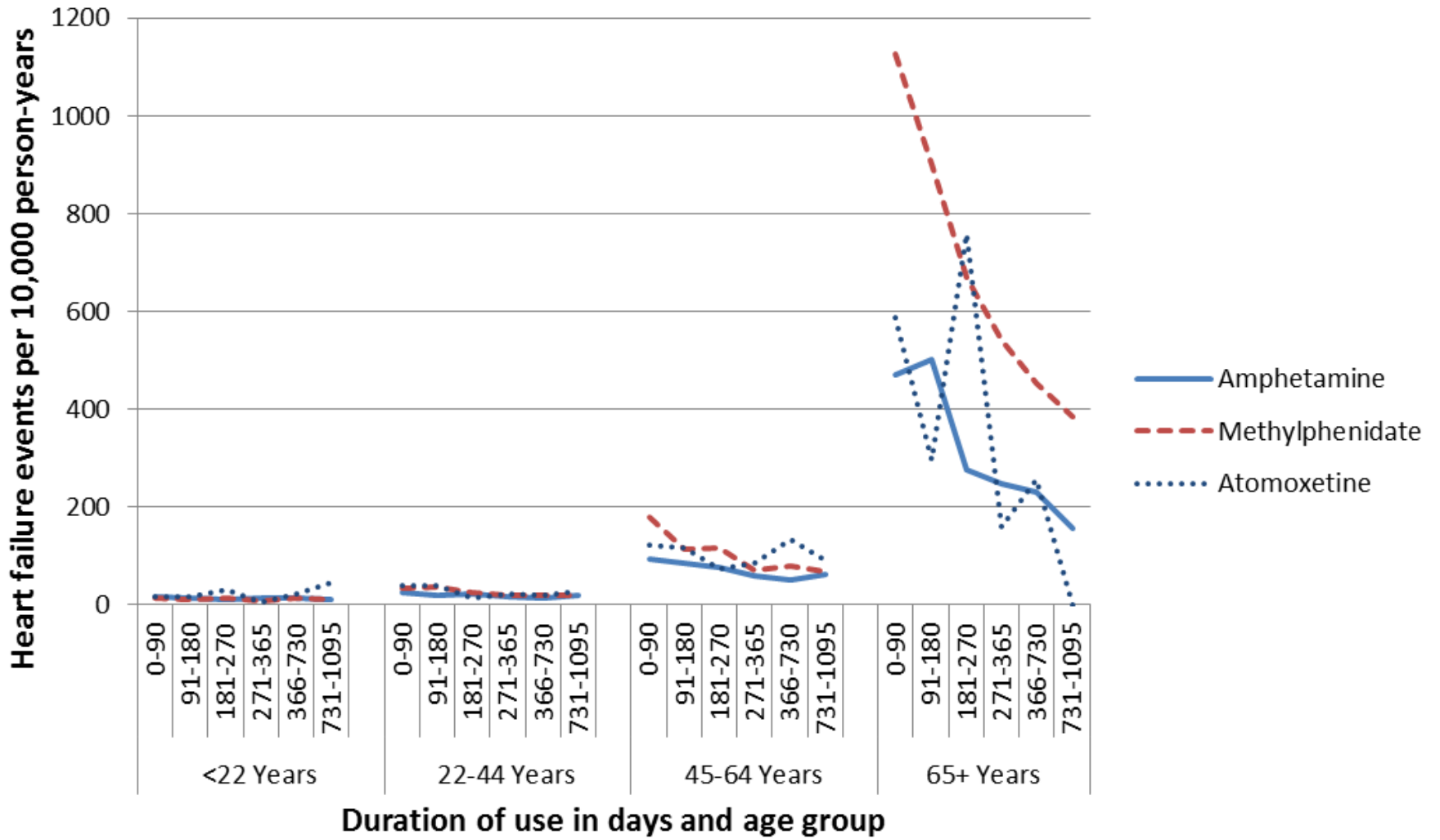
## Complex count queries (Level 1 / 1+)

- Counts and rates of events within user specified periods, among populations identified using complex “and/or/not” relationships.
- No assessment of causality

## Complex count queries (Level 1 / 1+)

- Counts and rates of events within user specified periods, among populations identified using complex “and/or/not” relationships.
  - Example: Rates of first diagnosis of heart failure or cardiomyopathy among new users of different drugs used to treat ADHD, by age and duration of exposure

**Figure 1. Rate of heart failure events (per 10,000 person years) by age group, medication, and duration of use**





## Complex count queries (Level 1 / 1+)

- Counts and rates of events within user specified periods, among populations identified using complex “and/or/not” relationships.
  - Example: Rates of first diagnosis of heart failure or cardiomyopathy among new users of different drugs used to treat ADHD, by age and duration of exposure
- 53 queries, 800+ scenarios in 2016

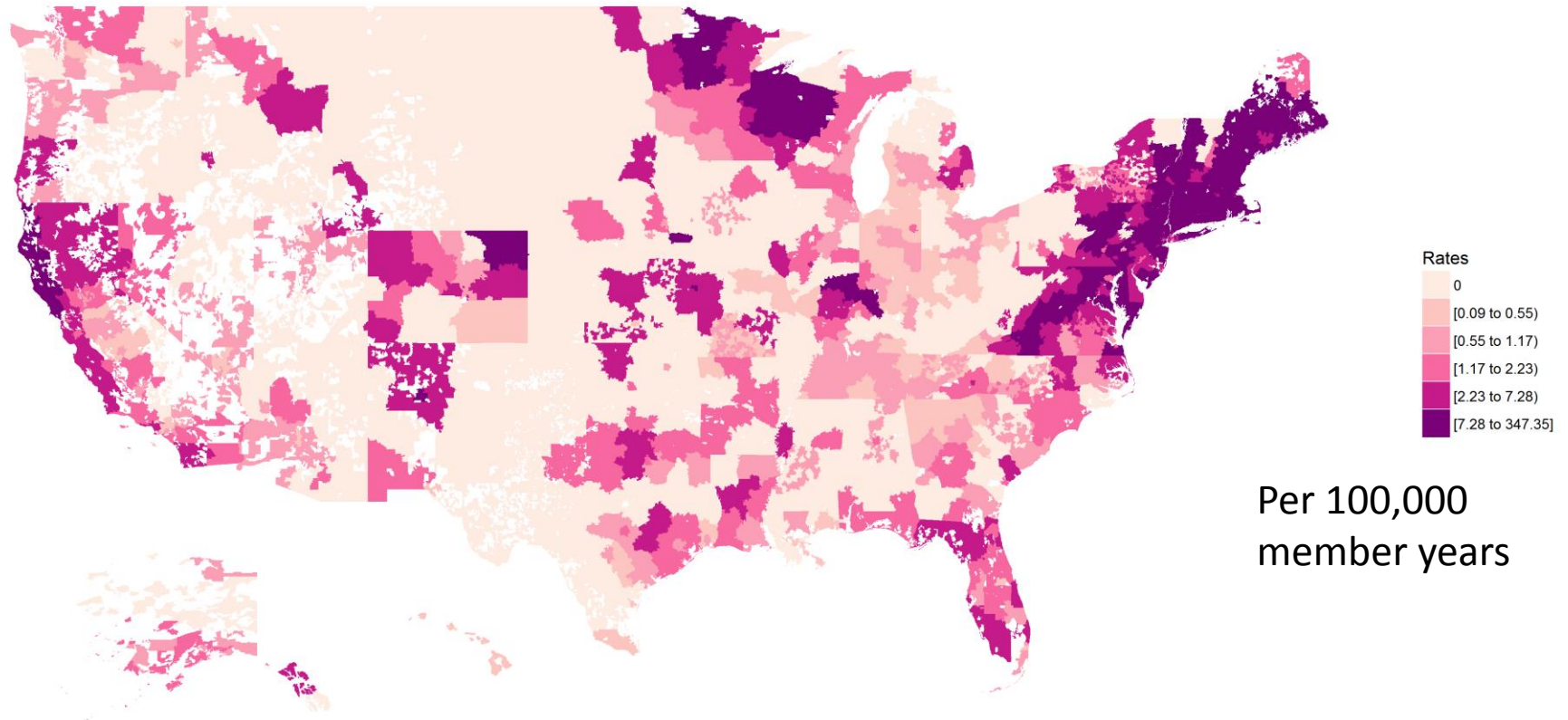
# Complex count queries (Level 1 / 1+)

- Counts and rates of events within user specified times, among populations identified using complex “and/or/not” relationships.
  - Example: Rates of first diagnosis of heart failure or cardiomyopathy among new users of different drugs used to treat ADHD, by age and duration of exposure
- 53 queries, 800+ scenarios in 2016
- *New uses*
  - *Medications errors (name confusion, dosing errors)*
  - *Geographic location stratification*

# Babesiosis Rates 1/1/2008 to 10/30/2015

## 3-digit ZIP code, Location Certain

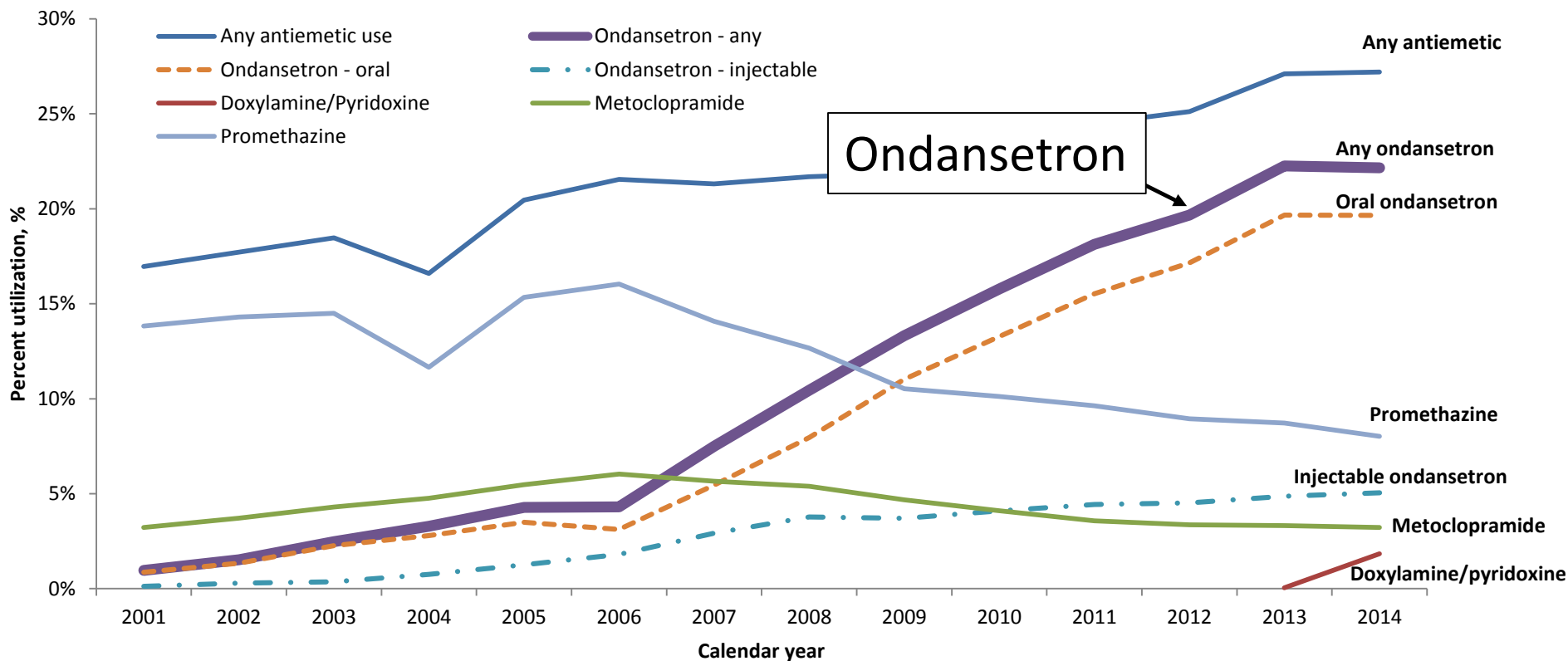
### Babesiosis Rates



## 8 Sentinel Complex Counts (L1) at 33<sup>rd</sup> ICPE

- Opportunities for Rapid Monitoring of New Cancer Treatments – Tyrosine Kinase Inhibitors
- Impact of ICD-10-CM on Selected CV-Related Events
- Identification of Name Confusion Medication Errors
- RCT Using FDA's Sentinel Infrastructure
- Dispensings of Influenza Antiviral Medications as a Source of Data for Influenza Surveillance
- Types 1 & 2 Diabetes Mellitus ICD-9-CM Codes Among New Users of Drugs Labeled for Type 2
- Use of TNF-alpha Inhibitors During Pregnancy
- Trends of Tdap Vaccination during Pregnancy

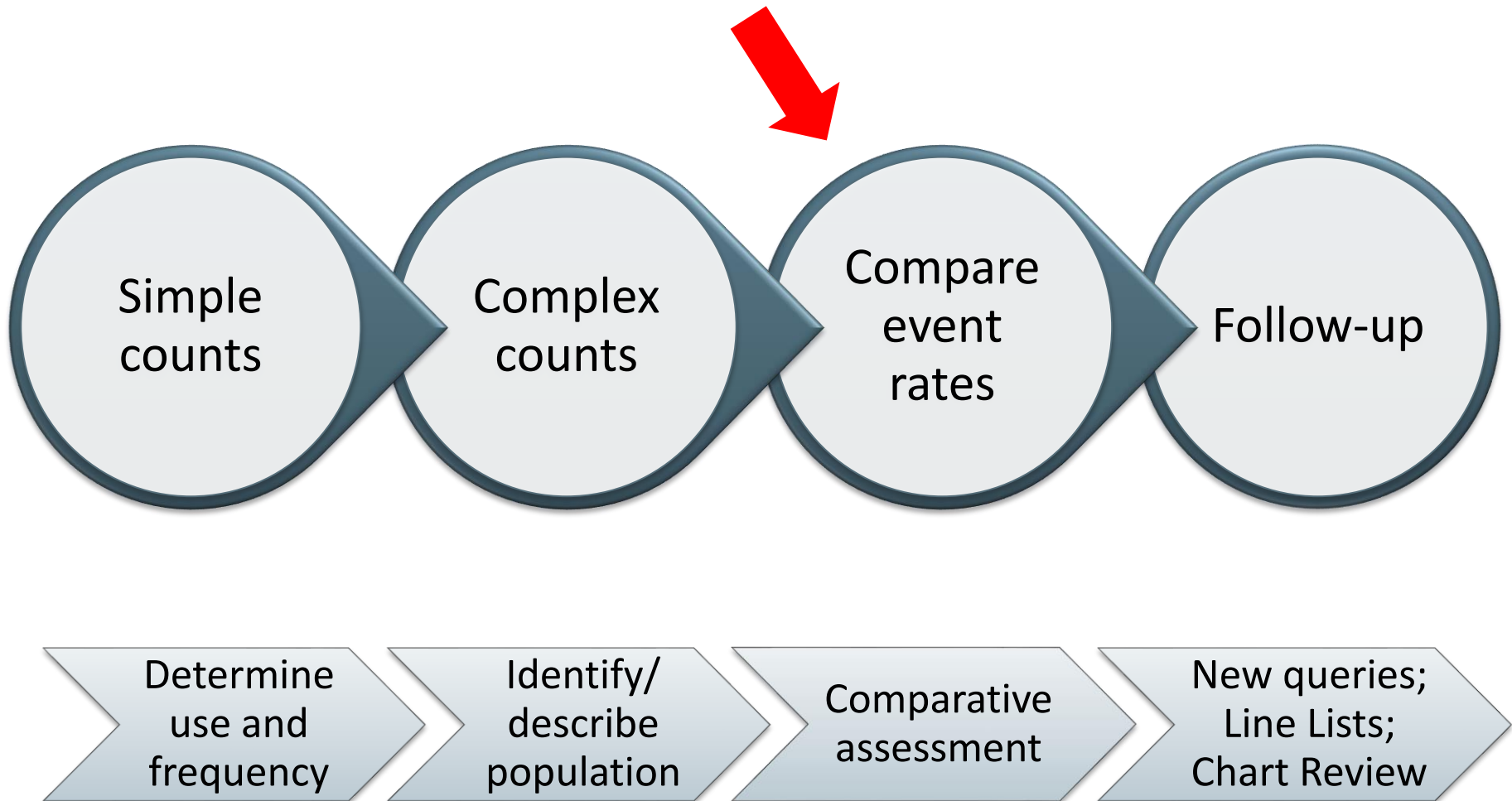
# Use of antiemetic drugs among live birth pregnancies in the Sentinel Distributed Database, 2001-2014<sup>a,b</sup>



<sup>a</sup> Dashed lines for oral and injection ondansetron form represent a portion of all total ondansetron use as shown by the solid purple line. Summation of oral and injection utilization sums to greater than total ondansetron use since some women received both products.

<sup>b</sup> Not all Mini-Sentinel data partners contributed data for the entire study period

# Querying sequence



## Comparison of rates (Level 2 / 2+)

- Propensity score matched relative rates or hazard ratios comparing outcomes among two cohorts identified by complex count program  
or
- Adjusted self-controlled risk interval analysis

## 4 Sentinel Comparisons (L2) at 33<sup>rd</sup> ICPE

- **Propensity score matched**
  - Venous thromboembolism after cyclic vs non-cyclic combined oral contraceptives
  - Stroke after antipsychotic use in the non-elderly
- **Self-controlled risk interval**
  - Seizure after gadolinium-enhanced imaging
  - Seizure after ranolazine

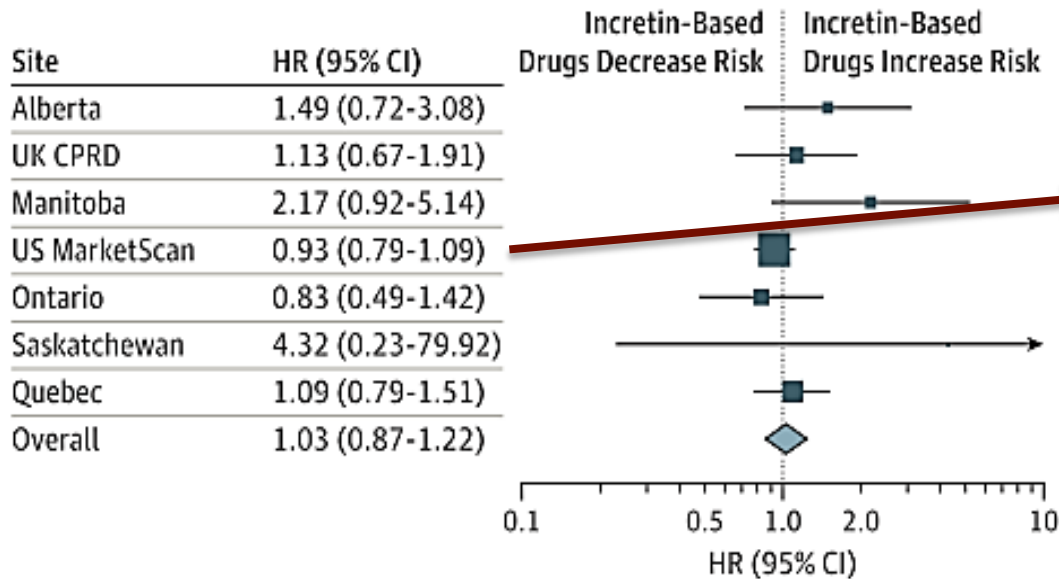


## Comparison of rates (Level 2 / 2+)

- Propensity score matched relative rates or hazard ratios comparing outcomes among two cohorts identified by complex count program  
or
- Adjusted self-controlled risk interval analysis
- 11 queries / 100+ scenarios in 2016

# Incretins and Pancreatitis - Results

Figure 2. Association Between the Use of Incretin-Based Drugs and the Risk of Acute Pancreatitis Among Patients With Type 2 Diabetes



CNODES MarketScan  
0.93 (0.79 - 1.09)

Sentinel Truven\*  
0.95 (0.81 - 1.12)

\* Truven Health MarketScan®  
Commercial Claims and Medicare  
Encounters Database

The reference category was current use of 2 or more oral antidiabetic drugs. The size of the boxes is proportional to the weight of a given participating site in the random-effects meta-analysis. The  $I^2$  (percentage of the total variance due to between-study heterogeneity) was 13.6% ( $P = .33$  for heterogeneity). CPRD indicates Clinical Practice Research Datalink; HR, hazard ratio.

Azoulay et al.  
JAMA Intern Med.  
2016;176:1464

## ABSTRACT

**Background:** Distributed health data networks use distributed databases for efficient, privacy-protecting, and effective public health research and surveillance activities. Distributed regression analysis (DRA) is a novel analytic method that does not require transferring of patient-level data in multi-database studies but produces results statistically equivalent to those from pooled patient-level data analysis. The execution of DRA has been largely manual and labor-intensive. We describe a new approach to conduct automated DRA in the FDA's Sentinel system, a distributed network using multiple electronic health data sources for medical product safety monitoring.

**Objective:** Implement a method within the existing PopMedNet™ (PMN) open-source platform used in Sentinel to allow automated, iterative, privacy-protecting, and scientifically accurate DRA in a real-world setting.

**Methods:** The project had 2 work streams: (1) develop DRA analytic code in SAS for multivariable-adjusted regression models and (2) enhance PMN to process DRA automated communication cycles within the distributed data network. We developed a new capability in PMN to enable the analysis center to (1) automatically aggregate site-specific intermediate statistics to compute or update the parameter estimates, which are returned to the data partners for subsequent iterations, and (2) to allow this iterative process to continuously refine the statistics until the model converges. The main outcome of interest was confirmation of analytic code accuracy and execution of DRA in a real-world setting. The DRA analytic code was validated against test data using results from pooled patient-level data analysis as a benchmark. PMN automation was tested internally and with external data partners.

**Results:** PMN software development was an iterative process where the implementation ensured that the functionality developed within the PMN code base would not impact existing Sentinel workflows or system functions. We developed and validated PMN's ability to perform regression analysis using only summary-level intermediate statistics and produce statistically equivalent regression parameters as pooled individual-level data analysis.

**Conclusion:** This work can be leveraged in the future for DRA in Sentinel and other networks. The functionality is agnostic to statistical software and can be extended to R and other software.

**Funding:** Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I.

## OBJECTIVE

- Implement a system to conduct distributed regression analysis (DRA) within the existing PopMedNet™ (PMN) open-source platform
- DRA key development features:
  - Automated
  - Iterative
  - Privacy-protecting
  - Scientifically accurate
  - Demonstrated use in a real-world setting

## BACKGROUND

- Distributed databases enable efficient, privacy-protecting, and effective public health research and surveillance activities
- DRA is a novel analytic method that **does not require transferring of patient-level data**
- DRA produces results statistically equivalent to pooled patient-level data analysis
- To date, DRA has been largely manual and labor-intensive

## METHODS

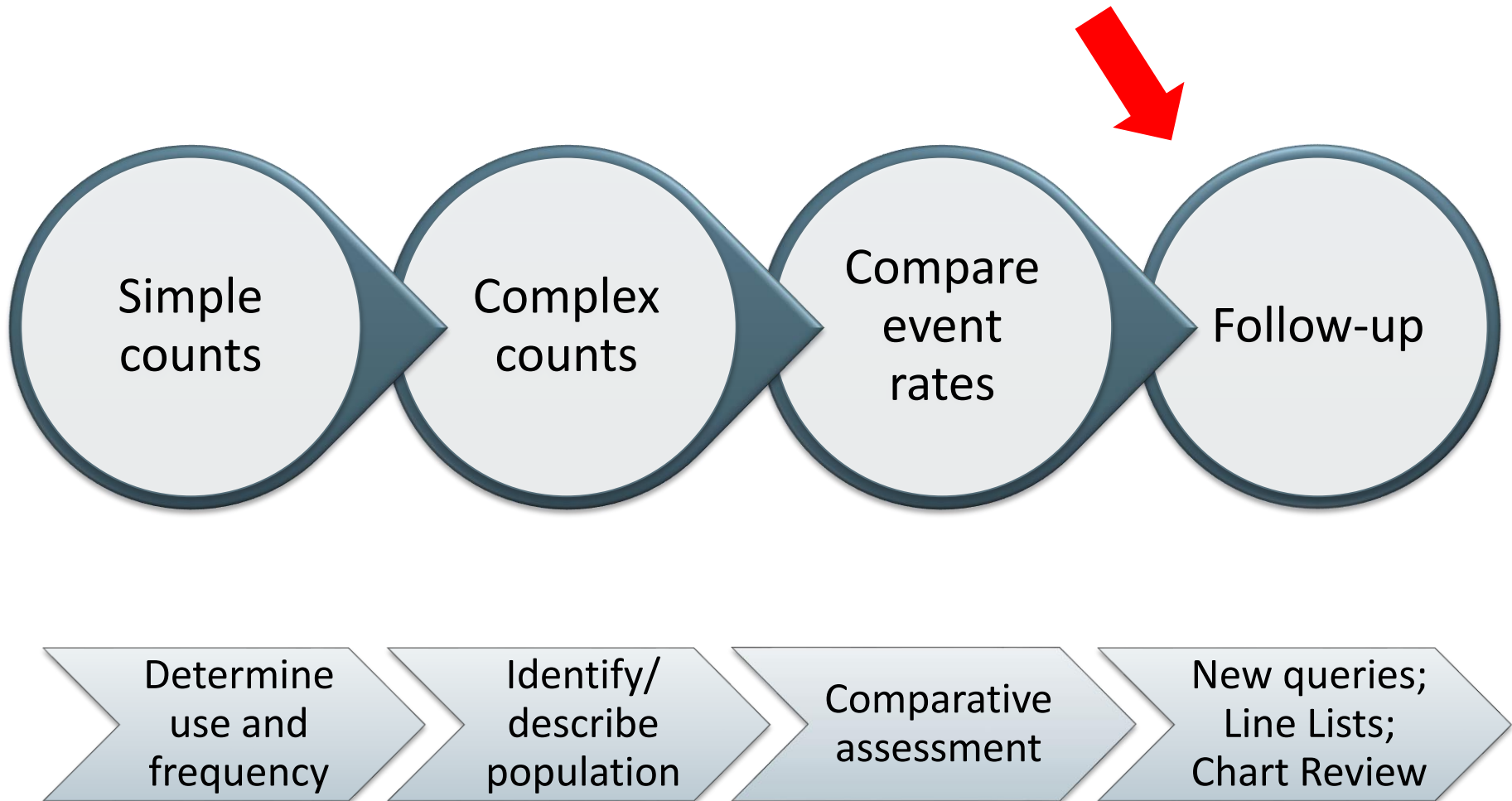
The project had two work streams:

- **Work stream 1:** Develop DRA analytic code in SAS for multivariable-adjusted regression analysis
- **Work stream 2:** Enhance PMN to process DRA automated communication cycles within the distributed network

Extend PMN to enable the Analysis Center to:

- Automatically aggregate site-specific intermediate statistics to compute regression parameter estimates, which are returned to the data partners for subsequent iterations
- Allow this iterative process to continuously refine the regression statistics until the model converges

# Querying sequence



# Patient Episode Profile Retrieval (PEPR)

Episode Detail						^ Incidence: F = first observed; I = incident; blank = prevalent # Primary Dx: P = primary; S = secondary; X = N/A ~ Med enroll segment containing the admission date of the encounter or the drug enroll segment containing the dispensing date									
Days from expos	Enc type	L O S	Clinical code			Code description	Incidence^	P Dx#	Node (Y/N)	Main expos (Y/N)	Any vacc (Y/N)	Rx days supp	Rx amt	Cov start~	Cov end~
			Cat	Type	Code										
0	AV		DX	09	V0382	Need Proph Vacc Agnst Strep Pne					1			-386	1260
0	AV		DX	09	V068	Need Proph Vacc Against Oth Comb Dz	F				1			-386	1260
0	AV		DX	09	V202	Routine Infant/Child Health Check								-386	1260
0	AV		PX	C4	90471	Immunization Admin	F				1			-386	1260
0	AV		PX	C4	90472	Immunization Admin Each Add	F				1			-386	1260
0	AV		PX	C4	90669	PCV7 Vaccine Im					1			-386	1260
0	AV		PX	C4	90710	MMRV Vaccine Sc	F			1	1			-386	1260
0	AV		PX	C4	99392	Prev Visit Est Age 1-4	F							-386	1260
4	AV		DX	09	0090	Inf Colitis Enterit & Gastroenterit	F							-386	1260
4	AV		PX	C4	99213	Office/Outpatient Visit Est	F							-386	1260
7	IP	1	DX	09	27651	Dehydration	I	P						-386	1260
7	IP	1	DX	09	53550	Uns Gastrit & Gastroduodit No Hemorr	I	X						-386	1260
7	IP	1	DX	09	7862	Cough	I	X						-386	1260
7	IP	1	DX	09	78703	Vomiting Alone	I	S	1					-386	1260
7	IP	1	PX	C4	71020	Chest X-Ray 2Vw Frontal & Latl	F							-386	1260
7	IP	1	PX	C4	74000	X-Ray Exam Of Abdomen	F							-386	1260

Day 0, office visit  
Routine health check  
Immunization

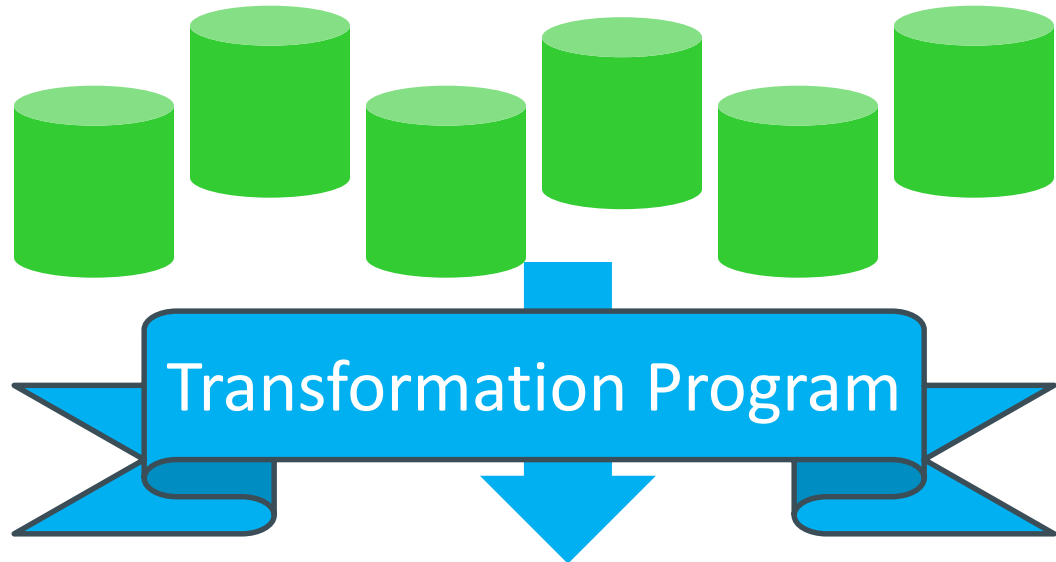
Day 4, office visit  
Gastroenteritis

Day 7, hospitalized  
Vomiting / cough  
Dehydration  
Gastroenteritis

**Rapid Response Requires  
Robust Data Quality Assurance –  
In Advance of Its Use**

# Every Data Partner transforms its data into the Sentinel Common Data Model

Data Partners' Source Database Structure



Transformed Database in Sentinel CDM Format

Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure
<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>
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		Amount dispensed	Facility	Diagnosis code & type	Procedure code & type
				Principal discharge diagnosis	

Lab Result	Vital Signs	Inpatient Pharmacy	Inpatient Transfusion	Death	Cause of Death
<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>	<b>Person ID</b>
Result and specimen collection dates	Measurement date and time	Administration date and time	Blood product code and type	Death date	Cause of death
Test type, immediacy & location	Height and weight	Encounter ID	Encounter ID	Source	Source
Logical Observation Identifiers Names and Codes (LOINC ®)	Diastolic & systolic BP	National Drug Code (NDC)	Blood type	Confidence	Confidence
Test result & unit	Tobacco use & type	Route	Administration start and end dates and times		
		Dose			

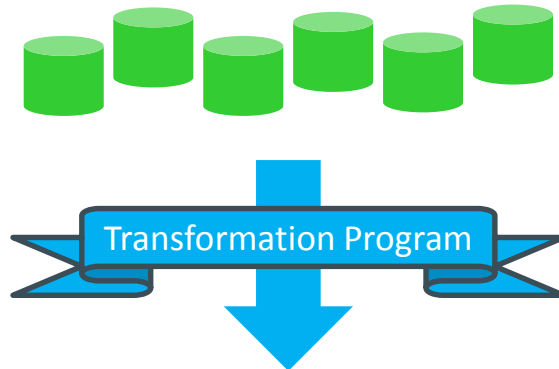
# The database is dynamic – updates overwrite the preceding data!

Data Partner Source Database Structure

Transformed database in Sentinel CDM Format

Timeframe of Data in Database

Data Delivery 1

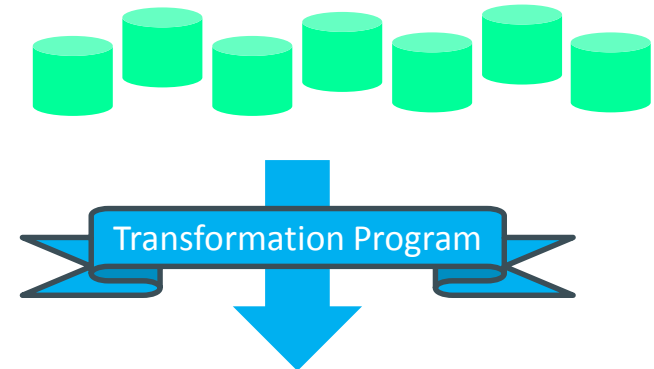


Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure
Person ID Enrollment start & end dates Drug coverage Medical coverage Medical record availability	Person ID Birth date Sex ZIP code	Person ID Dispensing date National Drug Code (NDC) Days supply Amount dispensed	Person ID Service (SNOMED)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

Lab Result	Vital Signs	Inpatient Pharmacy	Inpatient Transfusion	Death	Cause of Death
Person ID Result and specimen collection dates Test type, secondary & tertiary Logical Observation Identifier Names and Codes (LOINC ID) Test result & unit	Person ID Measurement date and time Height and weight Diastolic & systolic BP Tobacco use & type	Person ID Administration date and time Encounter ID National Drug Code (NDC) Route Dose	Person ID Blood product code and type Encounter ID Blood type Administration start and end dates and times	Person ID Death date Source Confidence	Person ID Cause of death Source Confidence

1/1/2000 1/1/2016

Data Delivery 2



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

Lab Result	Vital Signs	Inpatient Pharmacy	Inpatient Transfusion	Death	Cause of Death
Person ID Result and specimen collection dates Test type, secondary & tertiary Logical Observation Identifier Names and Codes (LOINC ID) Test result & unit	Person ID Measurement date and time Height and weight Diastolic & systolic BP Tobacco use & type	Person ID Administration date and time Encounter ID National Drug Code (NDC) Route Dose	Person ID Blood product code and type Encounter ID Blood type Administration start and end dates and times	Person ID Death date Source Confidence	Person ID Cause of death Source Confidence

1/1/2000 4/1/2016



# The quality assurance process

Send a standard QA program to check DP's data in waiting



## Data Partner

Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure
Person ID Enrollment date & end date Drug coverage Medical coverage Medical record availability	Person ID Birth date Sex ZIP code	Person ID Dispensing date National Drug Code (NDC) Date supply Amount dispensed	Person ID Service setting Encounter ID Encounter type & provider Facility	Person ID Service setting Encounter ID Encounter type & provider Diagnosis code & type Principal discharge diagnosis	Person ID Service setting Encounter type & provider Procedure code & type

Lab Result	Vital Signs	Inpatient Pharmacy	Inpatient Transfusion	Death	Cause of Death
Person ID Order and specimen collection dates Test type, reference & results Logical relationships between tests and other data Test result & unit	Person ID Measurement date and time Height and weight Blood pressure Respiratory rate & SpO2 Temperature & type	Person ID Administration date and time Encounter ID National Drug Code (NDC) Trade Dose	Person ID Blood product code and type Encounter ID Blood type Administration date and time and other	Person ID Death date Source Confidence	Person ID Cause of death Source Confidence

## Compliance Checks

**Level 1:** Completeness, validity, accuracy

**Level 2:** Cross-variable and cross-table integrity

## Judgment Call Checks

**Level 3:** Trends: consistency

**Level 4:** Logical: plausibility, convergence

# Sentinel Quality Assurance Statistics

- The QA team (six people) reviews ~50 data updates per year from 17 Data Partners
- Since 1/1/2016, the dataset has needed to be re-refreshed and QA package re-run 16 times to fix an issue
- In the latest data deliveries from the 5 largest DPs, 25 checks required DP follow-up
  - 22 of the 25 were Level 3 checks

## In closing

The Sentinel System supports timely regulatory decision making through

- Increasingly sophisticated, reusable tools that
  - Protect privacy and
  - Execute efficiently in diverse computing environments
- Well curated, analysis ready data

# Sentinel is a National Medical Product Monitoring System

LEARN MORE



## ABOUT

- Background
- Coordinating Center
- Privacy and Security
- The Sentinel System Story



## SAFETY ASSESSMENTS

- Active Risk Identification and Analysis System
- Assessments of Drugs
- Assessments of Vaccines, Blood, & Biologics



## DATA & SURVEILLANCE TOOLS

- Distributed Database and Common Data Model
- Complementary Data Sources
- Routine Querying Tools
- Validations and Literature Reviews



## COMMUNICATIONS

- FDA Safety Communications
- Publications and Presentations
- Sentinel Initiative Events
- Report Finder

## Latest Postings

### SPOTLIGHT

- Sentinel Initiative Public Workshop - Ninth Annual  
*Tue, 11/08/2016*

### STUDY PROTOCOLS & SURVEILLANCE PLANS

- Influenza Vaccines and Birth Outcomes Protocol (PRISM)  
*Fri, 01/20/2017*
- Identify and Evaluate Manufacturer-Level Drug Utilization and Switching Patterns in Sentinel  
*Mon, 12/12/2016*

### MODULAR PROGRAMS

- Querying Tools: Overview of Functionality and Technical Documentation  
*Tue, 12/27/2016*
- Influenza antiviral drug use 2010-2015  
*Mon, 10/31/2016*

**Thank you!**