Sentinel

Signal Identification Methods in the Sentinel System

Judith C. Maro, PhD
Harvard Medical School
and Harvard Pilgrim Health Care Institute, Boston, MA

Agenda

- Regulatory Background
- Sentinel Data
- Sentinel Tools and Methods

Sentinel and the United States Food and Drug Administration's (FDA) Mandate

Sentinel

Section 905
Mandates creation of Sentinel

Section 901
New Food and Drug Administration Amendments Act (FDAAA) Postmarketing Requirements (PMR) authority

"The Secretary may not require the responsible person to conduct a study under this paragraph, unless the Secretary makes a determination that the reports under subsection (k)(1) and the **active postmarket risk identification and analysis system** as available under subsection (k)(3) will not be **sufficient** to meet the purposes set forth in subparagraph (B)."

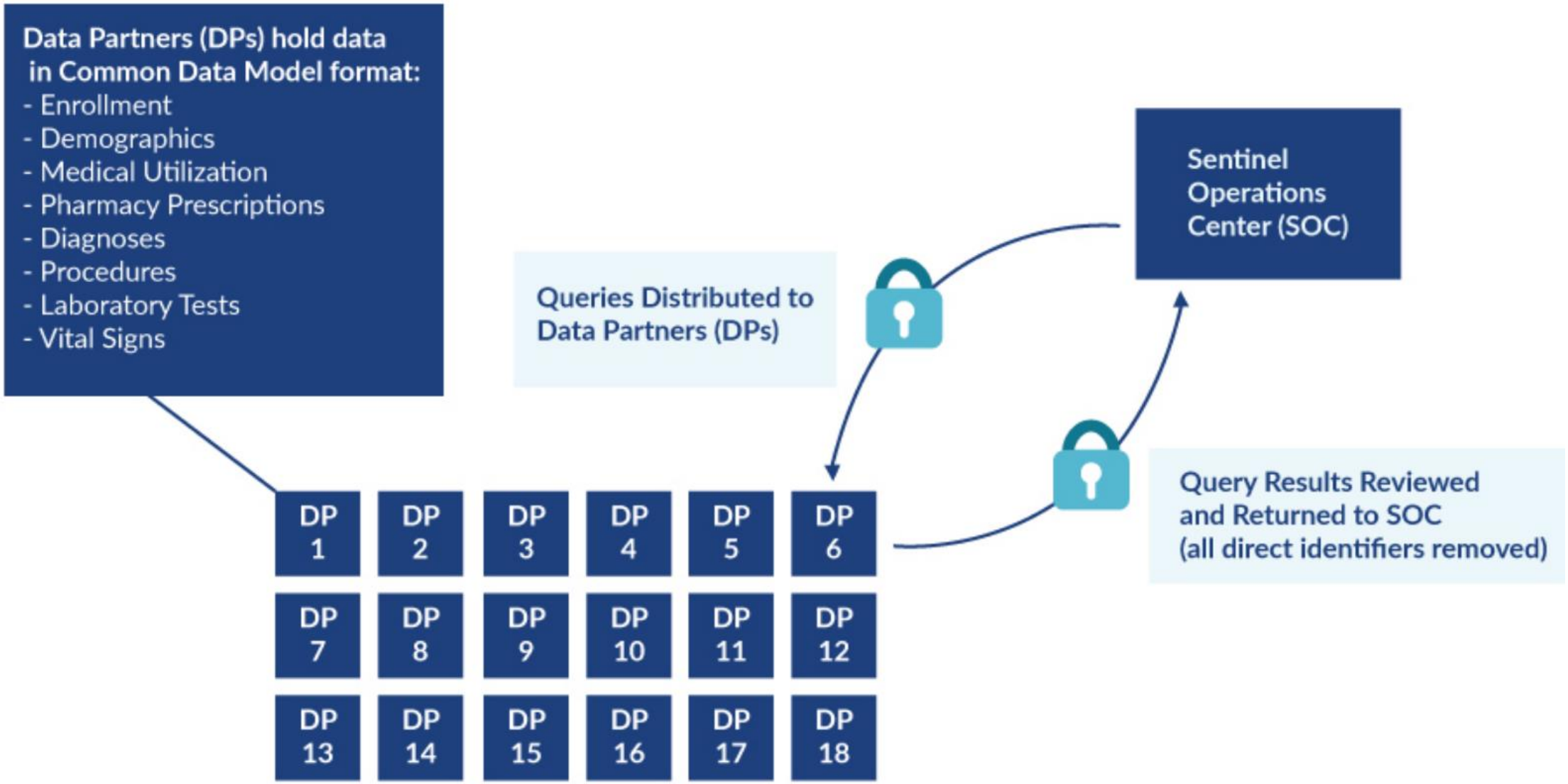
https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf

Sentinel Design Requirements

Sentinel

- Electronic health data for >100M persons
  - Include special populations (pregnant women, elderly)
  - Ability to link to external sources, e.g., National Death Index
  - Ability to access full text medical records
- Expertise in the way health care delivery and payment influence electronic healthcare data
- Rapid answers to many FDA safety questions
- Accuracy sufficient to support regulatory decision making
- Federal Information Security Management Act (FISMA)-compliant data security
- Ability to protect non-public information and to keep records on all data requests for public record-keeping

Sentinel Distributed Database

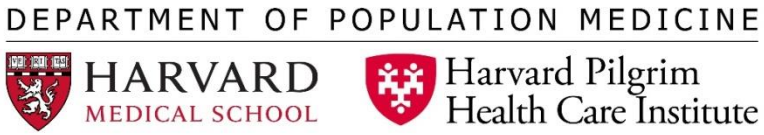


https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model

Collaborating Organizations

Sentinel

Lead – HPHC Institute



Data & Scientific Partners

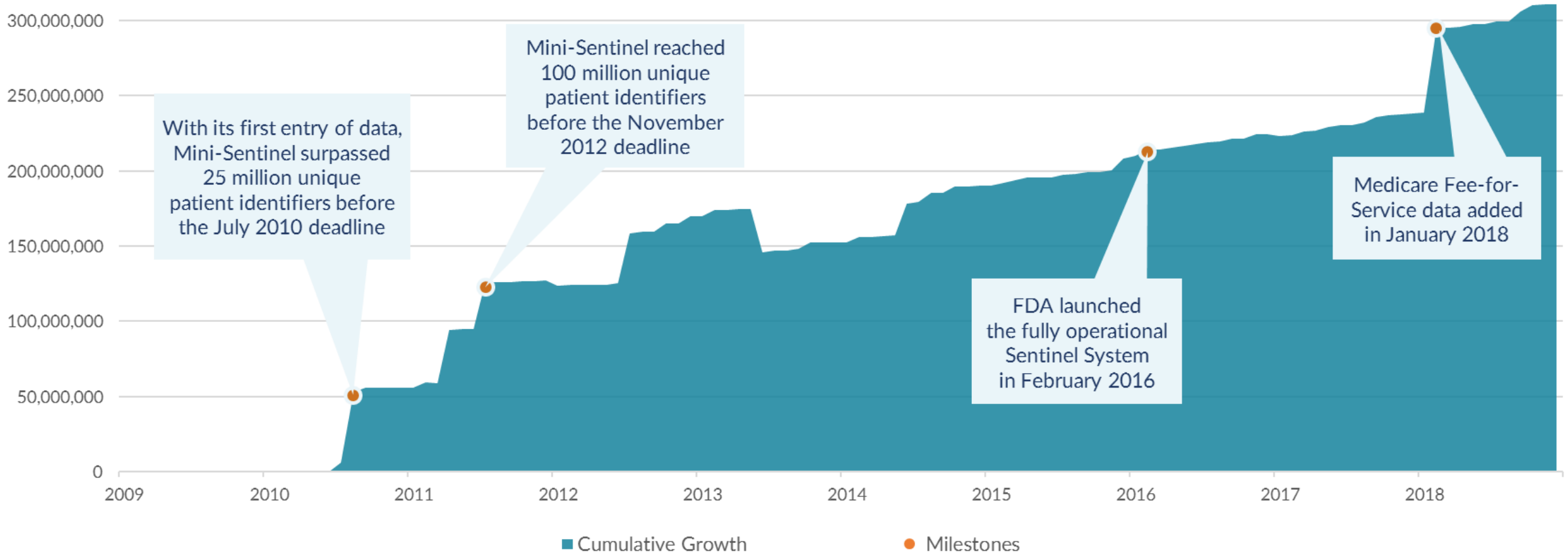
Scientific Partners

https://www.sentinelinitiative.org/collaborators

Growth of the Sentinel Distributed Database

Sentinel

- 70 million members currently accruing new data



The area above depicts the cumulative number of unique patient identifiers in the Sentinel Distributed Database from 2010 to present. If patients move health plans, they may have more than one patient identifier.

http://www.sentinelinitiative.org/sentinel/data/snapshot-database-statistics

Sentinel Common Data Model Guiding Principles

Sentinel

- Includes claims, electronic health record (EHR), and registry data and flexible enough to accommodate new data domains (e.g., free text).
  - Typically, we do not include empty tables – we expand as needed when fit for purpose.
- Data are stored at most **granular/raw level possible** with minimal mapping.
  - Distinct data types should be kept separate (e.g., prescriptions, dispensings)
  - Construction of medical concepts (e.g., outcome algorithms) from these elemental data is a **project-specific** design choice.
  - Sentinel stores these algorithms in a library for future use.
- Appropriate use and interpretation of local data requires the Data Partners' local knowledge and data expertise.
  - Not all tables are populated by all Data Partners➔site-specificity is allowed.
- Designed to meet FDA needs for analytic flexibility, transparency, and control.

Available Data Elements

Sentinel

Administrative Data

Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure
Patient ID	Patient ID	Patient ID	Patient ID	Patient ID	Patient 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 and Provider	Encounter Type and Provider	Encounter Type and Provider
Medical Record Availability	Etc.	Amount Dispensed	Facility	Diagnosis Code & Type	Procedure Code & Type
			Etc.	Principle Discharge Diagnosis	Etc.

Clinical Data

Lab Result	Vital Signs
Patient ID	Patient ID
Result & Specimen Collection Dates	Measurement Date & Time
Test Type, Immediacy & Location	Height & Weight
Logical Observation Identifiers Names and Codes (LOINC®)	Diastolic & Systolic BP
Etc.	Tobacco Use & Type
	Etc.

Registry Data

Death	Cause of Death	State Vaccine
Patient ID	Patient ID	Patient ID
Death Date	Cause of Death	Vaccination Date
Source	Source	Admission Date
Confidence	Confidence	Vaccine Code & Type
Etc.	Etc.	Provider
		Etc.

Inpatient Data

Inpatient Pharmacy	Inpatient Transfusion
Patient ID	Patient ID
Administration Date & Time	Administration Start & End Date & Time
Encounter ID	Encounter ID
National Drug Code (NDC)	Transfusion Administration ID
Route	Transfusion Product Code
Dose	Blood Type
Etc.	Etc.

Mother-Infant Linkage Data

Mother-Infant Linkage
Mother ID
Mother Birth Date
Encounter ID & Type
Admission & Discharge Date
Child ID
Child Birth Date
Mother-Infant Match Method
Etc.

https://www.sentinelinitiative.org/sentinel/data/distributed-database-common-data-model

Single Patient Example Data in Model

Sentinel

DEMOGRAPHIC

PATID	BIRTH_DATE	SEX	HISPANIC	RACE	zip
PatID1	2/2/1964	F	N	5	32818

DISPENSING

PATID	RXDATE	NDC	RXSUP	RXAMT
PatID1	10/14/2005	00006074031	30	30
PatID1	10/14/2005	00185094098	30	30
PatID1	10/17/2005	00378015210	30	45
PatID1	10/17/2005	54092039101	30	30
PatID1	10/21/2005	00173073001	30	30
PatID1	10/21/2005	49884074311	30	30
PatID1	10/21/2005	58177026408	30	60
PatID1	10/22/2005	00093720656	30	30
PatID1	10/23/2005	00310027510	30	15

ENROLLMENT

PATID	ENR_START	ENR_END	MEDCOV	DRUGCOV
PatID1	7/1/2004	12/31/2004	Y	N
PatID1	1/1/2005	12/31/2005	Y	Y

DEATH

PATID	DEATHDT	DTIMPUTE	SOURCE	CONFIDENCE
PatID1	12/27/2005	N	S	E

ENCOUNTER

PATID	ENCOUNTERID	ADATE	DDATE	ENCTYPE
PatID1	EncID1	10/18/2005	10/20/2005	IP

DIAGNOSIS

PATID	ENCOUNTERID	ADATE	PROVIDER	ENCTYPE	DX	DX_CODETYPE	PDX
PatID1	EncID1	10/18/2005	Provider1	IP	296.2	9	P
PatID1	EncID1	10/18/2005	Provider1	IP	300.02	9	S
PatID1	EncID1	10/18/2005	Provider1	IP	305.6	9	S
PatID1	EncID1	10/18/2005	Provider1	IP	311	9	P
PatID1	EncID1	10/18/2005	Provider1	IP	401.9	9	S
PatID1	EncID1	10/18/2005	Provider1	IP	493.9	9	S
PatID1	EncID1	10/18/2005	Provider1	IP	715.9	9	S

PROCEDURE

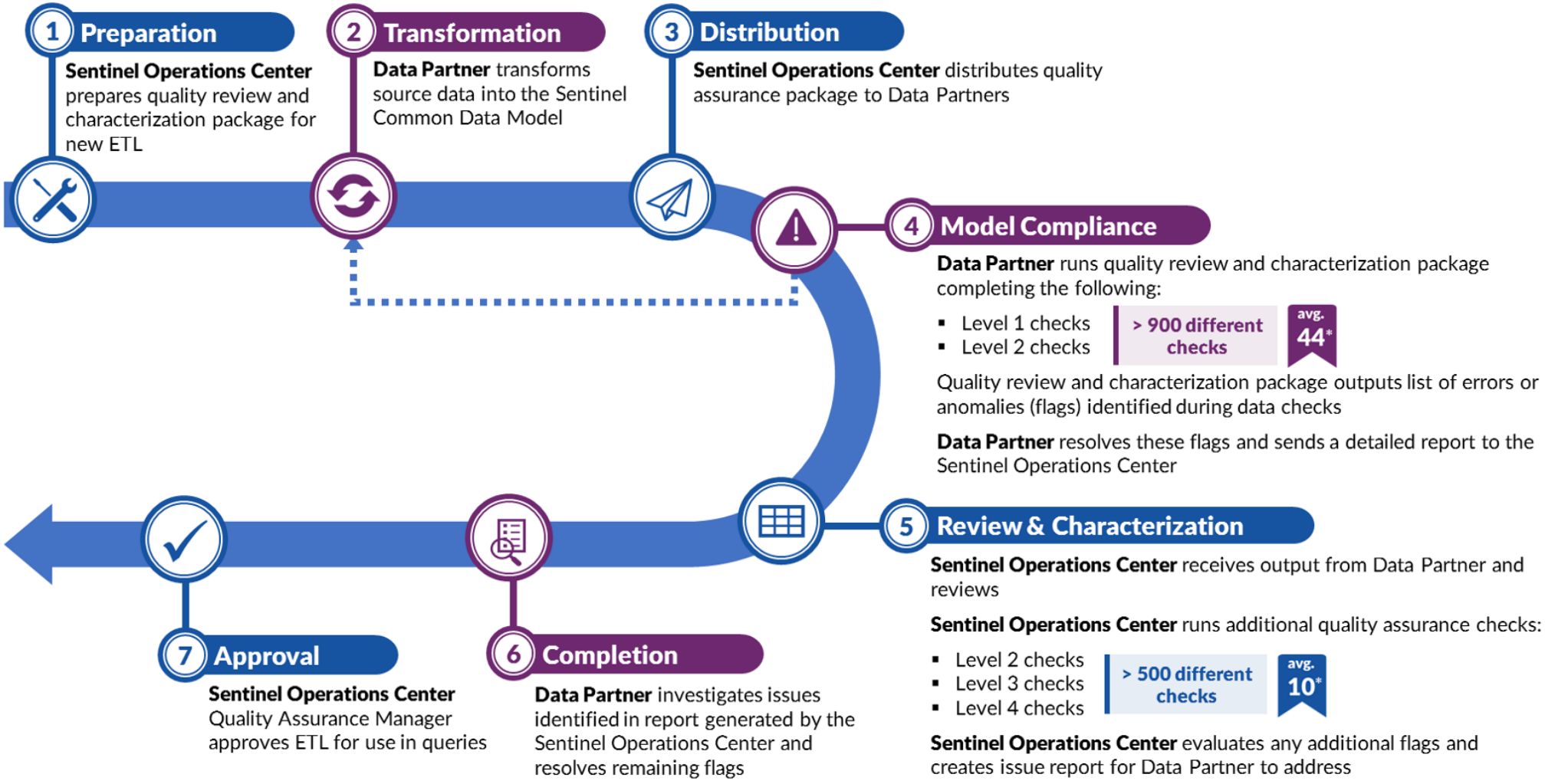
PATID	ENCOUNTERID	ADATE	PROVIDER	ENCTYPE	PX	PX_CODETYPE
PatID1	EncID1	10/18/2005	Provider1	IP	84443	C4
PatID1	EncID1	10/18/2005	Provider1	IP	99222	C4
PatID1	EncID1	10/18/2005	Provider1	IP	99238	C4
PatID1	EncID1	10/18/2005	Provider2	IP	27445	C4

CAUSE OF DEATH

PATID	COD	CODETYPE	CAUSETYPE	SOURCE	CONFIDENCE
PatID1	J18.0	10	U	S	E

Data Quality Review and Characterization Process

Sentinel



* On average, there are 44 flags identified by the program and 10 additional flags identified by the Sentinel Operations Center per ETL

https://www.sentinelinitiative.org/sentinel/data-quality-review-and-characterization

Data Quality Checks and Examples

Sentinel

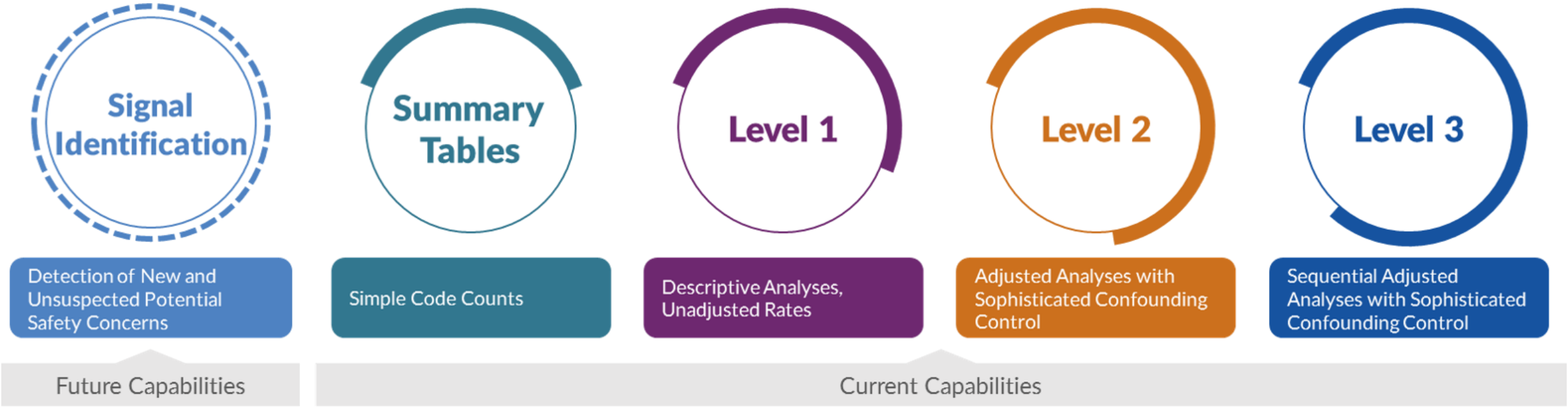
Level 1 Checks	**Completeness** ✓ Admission date is not missing value **Validity** ✓ Admission date is in date format	Sentinel Common Data Model Compliance
Level 2 Checks	**Accuracy** ✓ Admission date occurs before the patient's discharge date **Integrity** ✓ Admission date occurs within the patient's active enrollment period	Cross-Variable and Cross-Tabular
Level 3 Checks	**Consistency of Trends** ✓ There is no sizable percent change in admission date record counts by month-year	Cross-ETLs
Level 4 Checks	**Plausibility** ✓ There is no sizable percent change in the number of prostate cancer encounters by sex*	Cross-ETLs

*Under development

https://www.sentinelinitiative.org/sentinel/data-quality-review-and-characterization

Active Risk Identification and Analysis (ARIA)

Sentinel



- Template computer programs with standardized questions
- Parameterized at program execution
- Pre-tested and quality-checked
- Standard output

https://www.sentinelinitiative.org/active-risk-identification-and-analysis-aria

Cohort Identification and Descriptive Analysis (CIDA)

Sentinel

OVERVIEW

The purpose of this repository is to document version 7.3.0 of the Sentinel Routine Querying System. Functional documentation sections describe the capabilities of the tools in the system. Technical documentation sections specify the tools' inputs and outputs and provide the information required to build analytic packages to address research questions of interest.

SENTINEL ROUTINE QUERYING SYSTEM TOOLS

Sentinel's Routine Querying System includes three tools:

The **COHORT IDENTIFICATION AND DESCRIPTIVE ANALYSIS (CIDA) TOOL** identifies and extracts cohorts of interest from the Sentinel Distributed Database based on requester-defined options (e.g., exposures, outcomes, continuous enrollment requirements, incidence criteria, inclusion/exclusion criteria, relevant age groups, demographics).

The CIDA tool calculates descriptive statistics for the cohort(s) of interest and outputs datasets that may be useful for additional analyses. The CIDA tool may be used alone or in conjunction with the Propensity Score Analysis Tool or the Multiple Factor Matching Tool.

There are six cohort identification strategies available:

- Type 1: **Extract information to calculate background rates**
- Type 2: **Extract information on exposures and follow-up time**
- Type 3: **Extract information for a self-controlled risk interval design**
- Type 4: **Extract information for medical product use during pregnancy**
- Type 5: **Extract information for medical product utilization**
- Type 6: **Extract information on manufacturer-level product utilization and switching patterns**

https://dev.sentinelsystem.org/projects/SENTINEL/repos/sentinel-routine-querying-tool-documentation/browse

Signal Identification Methods and Future Tools

Signal Identification in the Sentinel System

The Food and Drug Administration Amendments Act (FDAAA) of 2007 mandated that FDA "create a robust system to identify adverse events and potential drug safety signals." Federal Food, Drug, and Cosmetic Act Section 505(k)(3)(C)(i)(3)(cc) (21 U.S.C. 355(k)(3)(C)(i)(III)(cc)). FDA defines **signal identification as a process of systematically evaluating potential adverse events related to the use of medical products without prespecifying an outcome of interest**. Several statistical approaches exist in Sentinel that can be applied to the electronic healthcare data to detect new and unsuspected potential safety concerns. These analytic tools are not intended to establish causal associations between medical products and potential adverse events. These approaches provide information about unexpected elevated frequencies of a health outcome after product exposure and should always be followed by clinical review and/or safety studies specifically designed to quantify the magnitude of effect with confounding control targeted at the specific outcome of interest.

http://www.sentinelinitiative.org/sentinel/surveillance-tools/signal-identification-sentinel-system

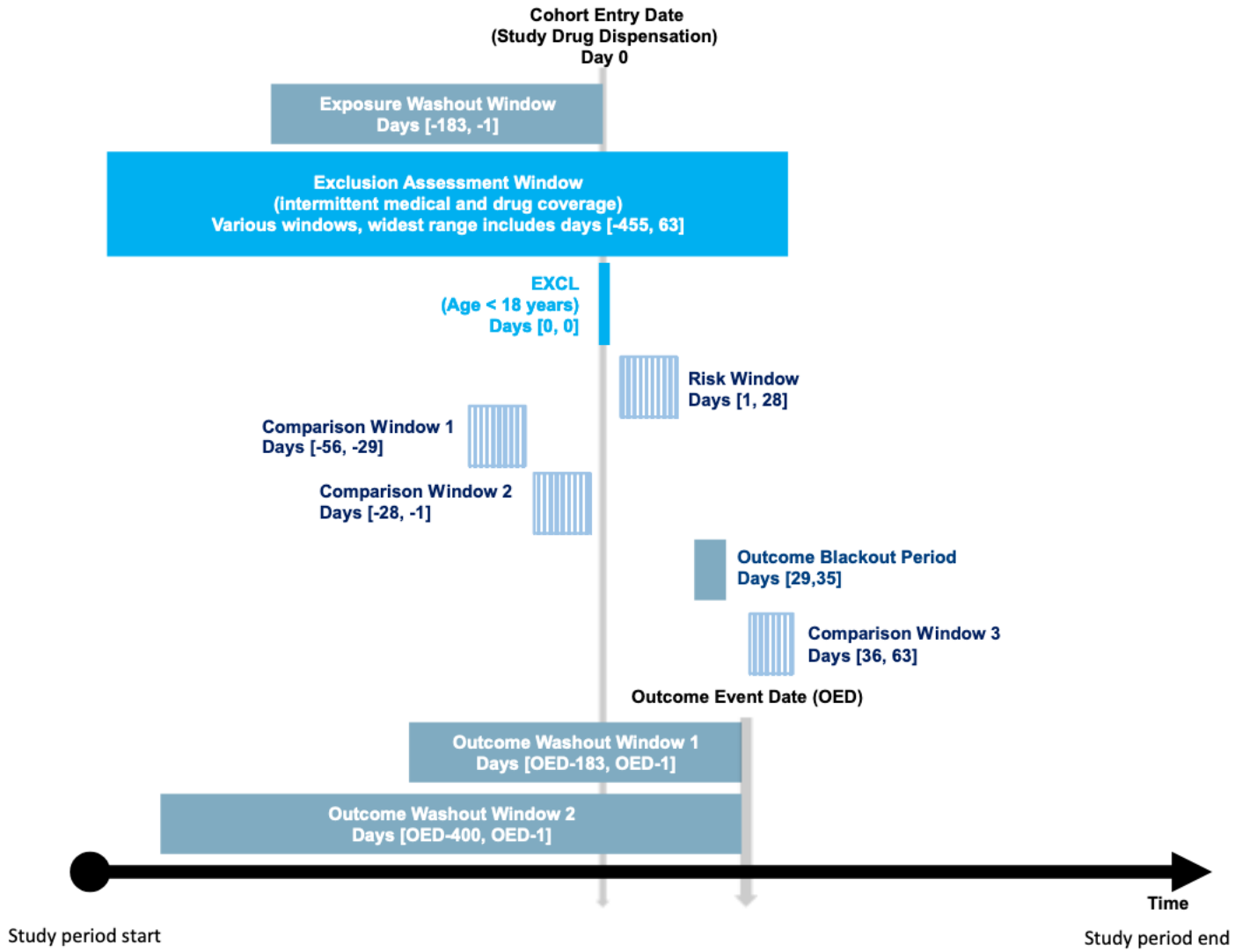
Overview of Signal Identification Techniques Utilized by the Sentinel System

Method	Study Design or Contrast	Test Statistic	Control for Multiple Testing	Adjustment for Trends in Healthcare Utilization
Information Component Temporal Pattern Discovery	Compares the rate of events in multiple prespecified control and risk windows relative to the timing of a first dispensing using a self-controlled design, while adjusting for general dispensing patterns across the database	Ranks alerts based on the delta in Information Component between the risk and control windows	No, however, uses a shrinkage estimator to reduce false positives due to random variability or rare events	Yes
Propensity Score Based TreeScan	Compares the rate of events in a prespecified risk window between persons newly exposed to a drug of interest who are matched by propensity score to a cohort of new users of a comparator drug	Ranks alerts based on the log-likelihood ratio, a measure of observed vs. expected counts, using a Bernoulli probability model	Yes, via Monte Carlo hypothesis testing	No
Self-Controlled TreeScan	Compares the rate of events in prespecified control and risk windows within the same person			Optional
Sequence Symmetry Analysis	Compares whether an event occurs more frequently after exposure to a medication than before medication exposure using a self-controlled design	Ranks alerts based on magnitude of absolute difference in sequence orders and presented unadjusted p-values from chi-square tests	No	No
Tree-temporal TreeScan	Compares the rate of events across multiple risk and control windows within the same person that do not require explicit pre-specification of the windows. Effectively combines the benefits of TreeScan with a temporal scan of many possible risk windows	Ranks alerts based on the log-likelihood ratio, a measure of observed vs. expected counts	Yes, via Monte Carlo hypothesis testing	Optional

http://www.sentinelinitiative.org/sentinel/surveillance-tools/signal-identification-sentinel-system

Self-Controlled Designs

Figure 1. Design Diagram



https://www.sentinelinitiative.org/sites/default/files/Methods/Evaluation_of_Three_Self-Controlled_Methods_for_Signal_Detection_Protocol-v2.pdf

Propensity-Score Matched Designs

Sentinel

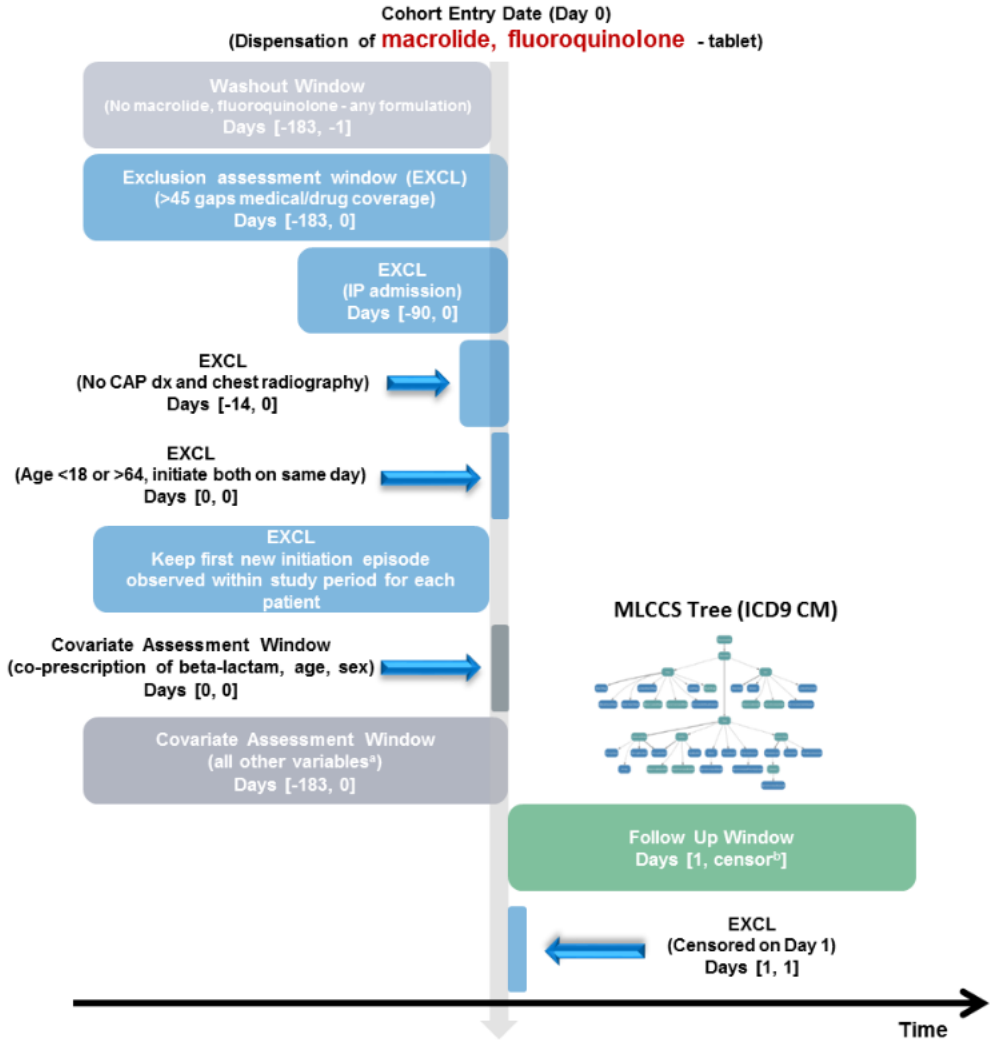
Example 1

[a]All other variables considered in candidate global propensity scores

- Age (continuous)
- Gender
- Metastatic cancer
- Tumor
- Arrhythmia
- Congestive heart failure
- Dementia
- Renal failure
- Weight loss
- Hemiplegia
- Alcohol abuse
- Pulmonary disease
- Coagulopathy
- Complicated diabetes
- Anemia
- Fluid and electrolyte disorder
- Liver disease
- Peripheral vascular disorder
- Psychosis
- Pulmonary circulation disorders
- HIV/AIDS
- Hypertension
- Degenerative disease of central nervous system
- Durable medical equipment
- Vaccine administration
- Screening examinations and disease management training
- Pap smear
- HPV DNA test
- Mammogram
- Fecal occult blood test
- Colonoscopy
- PSA test
- Number of inpatient hospitalizations
- Number of outpatient visits
- Number of emergency department visits
- Number of unique generics
- Prior prescription of penicillins
- Prior prescription of cephalosporins
- Prior prescription of sulfonamides
- Prior prescription of tetracyclines
- Prior prescription of aminoglycosides
- Co-prescription of beta-lactam
- Pregnancy at time of initiation
- Empirically selected

[b]Censoring

- 183 days
- Sep 30, 2015
- Discharged dead
- Disenroll medical or drug (45 day gaps allowed)



http://www.sentinelinitiative.org/sites/default/files/Methods/Development_and_Evaluation_Global_Propensity_Score_Data_Mining_with_Tree-Based_Scan_Statistics_Protocol.pdf

Tree-Based Scan Statistics are Enabled by:

Sentinel

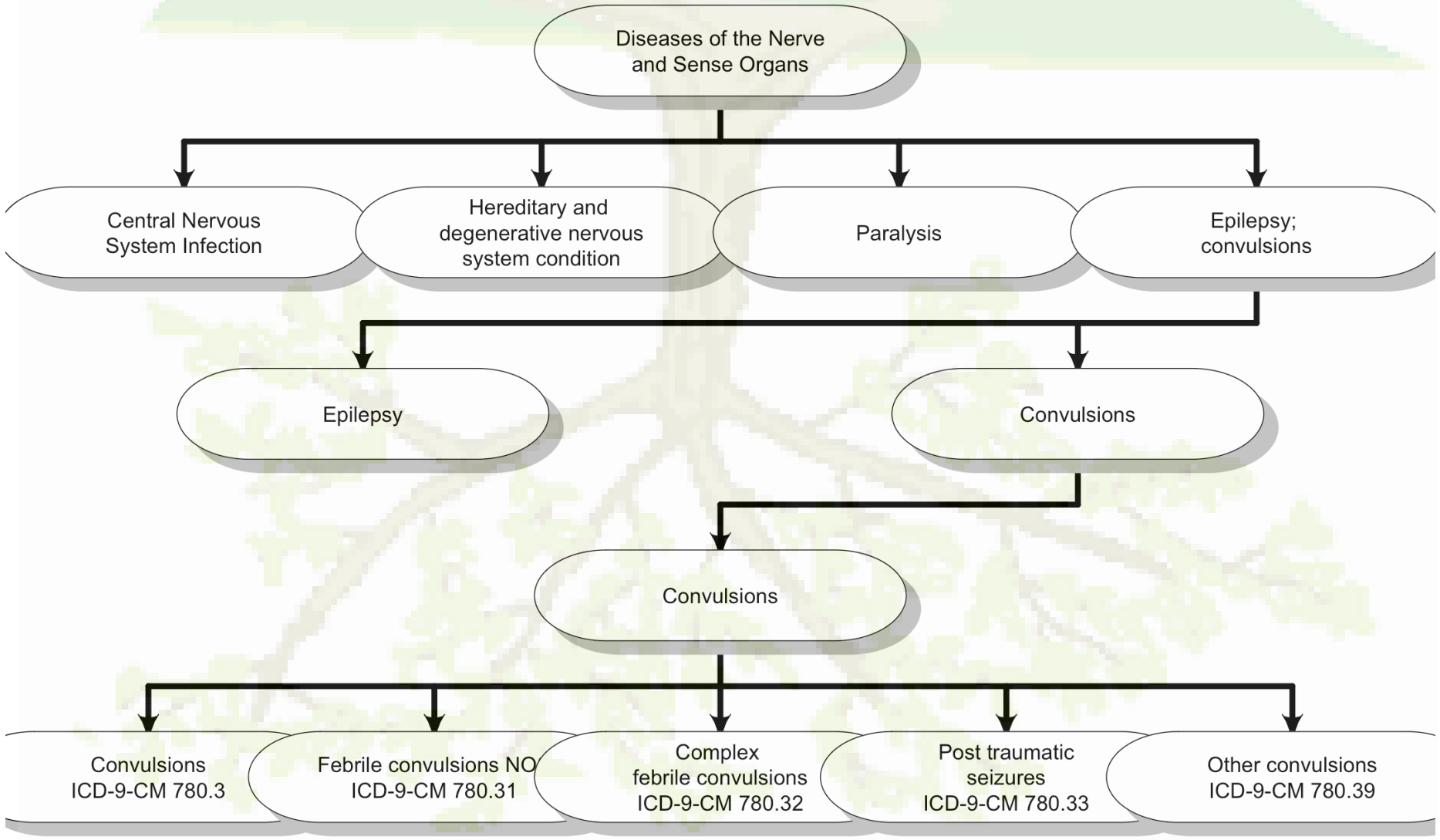
- A signal detection / data-mining method
- Automatically adjusts for multiple hypothesis testing
- Scans electronic health data that are grouped into hierarchical tree structures

TreeScan

http://www.treescan.org

Data Arranged in a Tree Structure

Sentinel



System Organ Class
Gastrointestinal disorders

High Level Group Term
Gastrointestinal signs and symptoms

High Level Term
Nausea and vomiting symptoms

Preferred Term
Nausea

Lowest Level Term
Feeling queasy

Study Designs Compatible with TreeScan Analytics

Sentinel

		TreeScan Analytics					
		Poisson Model		Bernoulli Model		Tree-Temporal Model	
		Unconditional	Conditional	Unconditional	Conditional	Unconditional	Conditional
Study Designs	Self-Controlled Design			X	X	X	X
	Propensity Score or other Fixed Ratio Match Design			X			
	Stratified Cohort Design	X	X				

Unconditional means the null hypothesis relies on an external input about the expected outcomes.
Conditional means the null hypothesis is determined by the characteristics of the incoming data set.

How has TreeScan been evaluated thus far?

Sentinel

Simulated Datasets

- **Advantages**
  - Artificially inject “excess risk” of variable specific sizes
  - Allows quantitative assessment of method under “experimental conditions” where “truth is known”
- **Limitations**
  - Simulated data has a range of realistic representations. Early simulations are quite artificial

Empiric Assessments

- **Advantages**
  - Empiric testing with real data
  - Allows assessment of method under real life conditions
  - Can be effective method to assess performance if test case is well characterized
- **Limitations**
  - Can be challenging to interpret unexpected results
  - Need additional information to investigate unexpected results

Self-Controlled Designs

Sentinel

Sentinel

PDS Pharmacoepidemiology & Drug Safety

Official Journal of the International Society for Pharmacoepidemiology

ISPE

PC

ORIGINAL RI

Mening
Vaccine
method

Prep
Kullc

Rongxia Li
Stanley Xu,

First publish

Assessment
Papillomavi
Controlled T
Signal–Dete
System

W Katherine Yih, J
Carolyn Balsbaugh, D
Martin Kulldorff

American Journal of E
1269–1276, https://d
Published: 23 Febru

Submit Comment

Evaluation of Three Self-Controlled Methods for Signal Detection: TreeScan, Sequence Symmetry Analysis, and Information Component Temporal Pattern Discovery

Project Title	**Evaluation of Three Self-Controlled Methods for Signal Detection: TreeScan, Sequence Symmetry Analysis, and Information Component Temporal Pattern Discovery**
Date Posted	*Wednesday, April 24, 2019*
Status	In progress
Deliverables	Evaluation of Three Self-Controlled Methods for Signal Detection: TreeScan, Sequence Symmetry Analysis, and Information Component Temporal Pattern Discovery Protocol
Description	The aim of this methods project is to compare the relative performance of three analytic methods, TreeScan, Sequence Symmetry Analysis (SSA), and Information Component Temporal Pattern Discovery (ICTPD) in signal detection capability (both type I and type II error) using a simulated dataset as well as concordance in alerting when using an empiric dataset. The Workgroup will use the same dataset(s) to examine health outcomes of interest using all three methods. The Workgroup will select at least one drug evaluation example with a well-known safety profile for evaluation of TreeScan, SSA, and ICTPD.

Propensity Score Matched Designs

Sentinel

Epidemiology. 29(6):895–903, NOV 2018
DOI: 10.1097/EDE.0000000000000907, PMID: 30074538
Issn Print: 1044-3983
Publication Date: 2018/11/01

Print

Data Mi
Propens
Statistic

Shirley V. Wang; Ju
Gagne; Elisabetta
Sebastian Schnee

+ Author Informat

Submit Comment

Development and Evaluation of a Global Propensity Score for Data Mining with Tree-Based Scan Statistics

Project Title	**Development and Evaluation of a Global Propensity Score for Data Mining with Tree-Based Scan Statistics**
Date Posted	*Friday, August 10, 2018*
Status	In progress
Deliverables	Development and Evaluation of a Global Propensity Score for Data Mining with Tree-Based Scan Statistics: Protocol

Stratified Cohort Designs with Referent Cohort

Sentinel

PDS Pharmacoepidemiology & Drug Safety
Official Journal of the International Society for Pharmacoepidemiology
ISPE

Original Repor

Drug safe

Martin Kulldor
Richard Platt,

First published

Read the full t

Pharmaceutics **2013**, *5*(1), 179-200; https://doi.org/10.3390/pharmaceutics5010179

Open Access Article

Drug Advers
the Gamma
Tree-based S

Jeffrey S. Brown [1,2,*]
Inna Dashevsky [1], M
K. Arnold Chan [5,6],
Lisa Herrinton [2,10],
David Smith [2,13] an

eGEMs AcademyHealth The Journal for Electronic Health Data and Methods

Start Submi

Reading: Statistical Power for Postlicensure Medical Product Safety Data Mining

Share:

Empirical research

Statistical Power for Postlicensure Medical Product Safety Data Mining

Authors: Judith C. Maro, Michael D. Nguyen, Inna Dashevsky, Meghan A. Baker, Martin Kulldorff

Strengths of TreeScan

Sentinel

1. Takes advantage of hierarchical nature of clinical concepts in the form of a tree structure.
2. Investigator does not need to understand how particular outcomes are coded (i.e., can be indifferent to the granularity of the outcome data)
3. Formal control for multiple hypothesis testing (Overall Type 1 error)

Limitations of TreeScan

Sentinel

1. All outcomes are treated identically across the tree (8000+) regardless of their time of onset, severity, etc.
2. Complex outcomes (algorithms such as 2 codes within X days of each other) are not tested with TreeScan.
3. Individual study designs have limitations depending on the design chosen.

Selected Findings from Pilot Work

Sentinel

- Most important decision is ultimately based on study design.
- Self-Controlled Methods
  - Best when applied to stable patients (eg, contraceptives, vaccines)
  - Moderate performance for statins; Better performance possible with more careful exclusion criteria for recently hospitalized / unstable patients
  - Poor performance for acutely ill, unstable patients
- Propensity-Score Adjustment Methods
  - Best when obvious referent product to compare.
  - Even partial degrees of adjustment provide large improvements in performance as compared to no adjustment.

Expansion of TreeScan to a Sequential Framework

Sentinel

Submit Comment

Sequential TreeScan Signal Identification Methods Development

Project Title	**Sequential TreeScan Signal Identification Methods Development**
Date Posted	*Tuesday, December 11, 2018*
Status	In progress
Description	The aim of this methods project is to enable and pilot test sequential TreeScan analyses over time. This project will develop adjustments to tree-based scan statistics (Unconditional Bernoulli) that will enable sequential versions of TreeScan for the fixed-window self-controlled and propensity score matched approaches. Sequential TreeScan will also be performed on an agreed-upon example problem (i.e., a test case) in a non-distributed but routinely updated data source (Optum Clinformatics).

http://www.sentinelinitiative.org/sentinel/methods/sequential-treescan-signal-identification-methods-development

General Resources

Sentinel

Trainings and Public Meetings

- Public Sentinel Training at FDA - Day 2 of the Tenth Annual Sentinel Initiative Public Workshop
- Implementation of Signal Detection Capabilities in the Sentinel System, Duke Margolis Public Meeting
- 2018 ICPE Presentation: Data Mining for Adverse Drug Events with a Propensity Score Matched Tree-Based Scan Statistic
- 2018 ICPE Presentation: Signal Detection using TreeScan with Drug Classes: Pilot Projects in Sentinel
- 2017 ICPE Workshop: TreeScan™: A Novel Data-Mining Tool for Medical Product Safety Surveillance
- 2017 ICPE Presentation: Promises and Challenges of Screening for Adverse Events in Sentinel
- Guidance for Industry: Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment, March 2005

Projects and Publications

- Evaluation of Three Self-Controlled Methods for Signal Detection: TreeScan, Sequence Symmetry Analysis, and Information Component Temporal Pattern Discovery
- Data Mining for Adverse Drug Events with a Propensity Score Matched Tree-Based Scan Statistic
- The U.S. Food and Drugs Administration's Sentinel Initiative: Expanding the Horizons of Medical Product Safety
- Statistical Power for Postlicensure Medical Product Safety Data-Mining
- Infrastructure for Evaluation of Statistical Alerts Arising from Vaccine Safety Data Mining Activities in Mini-Sentinel
- Drug Adverse Event Detection in Health Plan Data Using the Gamma Poisson Shrinker and Comparison to the Tree-based Scan Statistic

http://www.sentinelinitiative.org/sentinel/signal-identification-sentinel-system/resources

Programming Resources

Submit Comment

TreeExtraction Documentation

Project Title	**TreeExtraction Documentation**
Date Posted	*Friday, June 29, 2018*
Status	In progress
Deliverables	Sentinel Reusable Programs: TreeExtraction Program v1.2 SAS Package Toolkit: TreeExtraction v1.2 Macros and Input Files Sentinel Reusable Programs: TreeExtraction Program v1.3 SAS Package Toolkit: TreeExtraction v1.3 Macros and Input Files Sentinel Reusable Programs: TreeExtraction Program v1.4 SAS Package Toolkit: TreeExtraction v1.4 Macros and Input Files CDER Supporting Tree and Mapping Files CBER Supporting Tree and Mapping Files

http://www.sentinelinitiative.org/sentinel/surveillance-tools/software-toolkits/treeextraction-documentation

Discussion

Sentinel

Acknowledgements

- Thanks to my many colleagues within the greater Sentinel Initiative including our many collaborating institutions

- Questions: info@sentinelsystem.org