

Medical Product Safety Surveillance: Data Quality in the Sentinel Initiative

Judith C. Maro, PhD

Harvard Medical School and Harvard Pilgrim Health Care Institute

December 3, 2019

Sentinel Program Overview

What is the Sentinel System?

One of the FDA's biggest jobs is to make sure drugs, vaccines, and medical devices are safe. FDA wants to know if patients get bad side effects from these products. To make it faster and easier to learn about problems, FDA created a special program called the Sentinel System.

How the Sentinel System Works



Sentinel System's 3 important parts

- Information: The system looks at billing claims and patient records.
- **Expert Team: Sentinel** works with scientists. doctors and computer experts.
- **Computer Programs:** They study large groups of patients who take the same medicine. or use the same device.



Personal privacy

- No one at FDA or the **Sentinel Operations** Center has access to your name, address, or any other information that identifies you.
- For more information. visit sentinelinitiative.org.



Sentinel asks questions like:

- How many patients take the same drug?
- How many patients are getting bad side effects (swelling, bleeding, etc.)?
- Are side effects more common after taking one drug than after another drug that treats the same problem?



How does FDA use the information?

- FDA can choose to collect more information.
- FDA can provide updated safety information for patients and providers.
- If you have concerns about your own medical products, please contact your doctor.

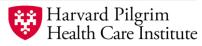
Sentinel Infrastructure: Available Data Elements

Collaborating Organizations

Lead: Harvard Pilgrim Health Care Institute







Data & Scientific Partners





































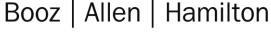




NYC-CDRN

New York City Clinical







IBM Watson Health



MEDICAL CENTER





Kaiser Permanente Washington Health Research Institute





(itin) PEDSnet















PaTH Network







Sentinel Data Philosophy

- 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

Administrative Data							
Enrollment	Demographic	Dispensing	Encounter	Diagnosis	Procedure		
Patient ID	Patient ID	Patient ID	Patient ID	Patient ID	Patient ID		
Enrollment Start &	Birth Date	Dispensing Date	Service Date(s)	Service Date(s)	Service Date(s)		
End Dates	Sex	National Drug Code	Encounter ID	Encounter ID	Encounter ID		
Drug Coverage	Zip Code	(NDC)	Encounter Type and	Encounter Type and	Encounter Type and		
Medical Coverage	Etc.	Days Supply	Provider	Provider	Provider		
Medical Record	Medical Record		Facility	Diagnosis Code &	Procedure Code &		
Availability			Etc.	Туре	Туре		
				Principal Discharge Diagnosis	Etc.		
				Diagnosis			

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

Innationt 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.

Mother-Infant Linkage Data

Single Patient Example Data in Model

	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	3	30		
PatID1	10/14/2005	00185094098	30	3	30		
PatID1	10/17/2005	00378015210	30	4	15		
PatID1	10/17/2005	54092039101	30	3	30		
PatID1	10/21/2005	00173073001	30	3	30		
PatID1	10/21/2005	49884074311	30	3	30		
PatID1	10/21/2005	58177026408	30	6	50		
PatID1	10/22/2005	00093720656	30	3	30		
PatID1	10/23/2005	00310027510	30	1	15		

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

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

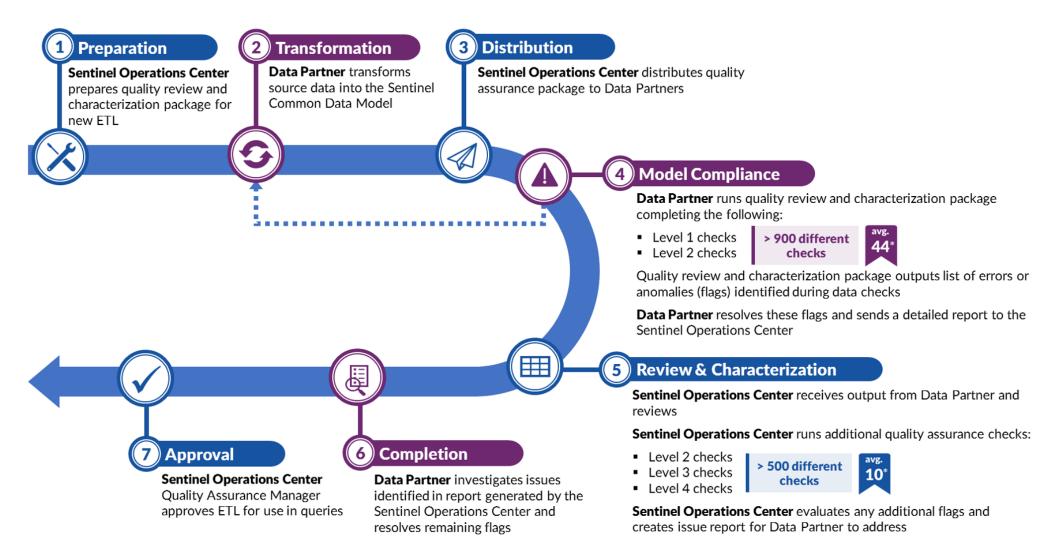
ENCOUNTER						
PATID	ENCOUNTERID	ADATE	DDATE	ENCTYPE		
PatID1	EncID1	10/18/2005	10/20/2005	IP		

	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



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

Data Quality Checks and Examples

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

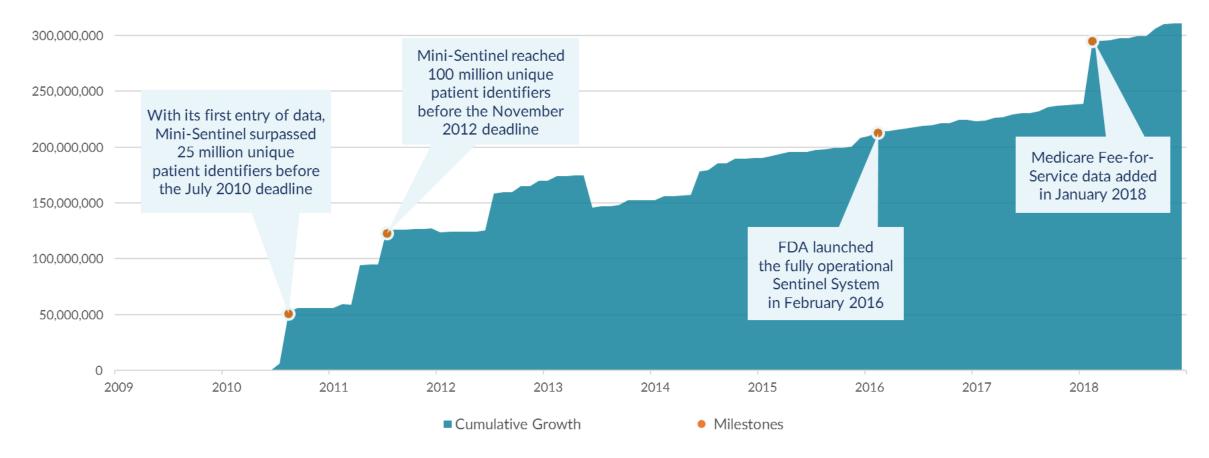
✓ There is no sizable percent change in the number of prostate cancer encounters by sex*

Cross-ETLs

*Under development

Growth of the Sentinel Distributed Database

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.

Publicly Available Formatted Data

Submit Comment

Medicare Claims Synthetic Public Use Files in Sentinel Common Data Model Format

Project Title	Medicare Claims Synthetic Public Use Files in Sentinel Common Data Model Format
Date Posted	Wednesday, March 27, 2019
Status	Complete
Deliverables	Sentinel's SynPUFs Software Toolkit
	SynPUFs Example Sentinel Modular Program Report
Related Links	Centers for Medicare and Medicaid Services Synthetic Public Use Files (SynPUFs)
Description	Sentinel has made available the CMS 2008-2010 Data Entrepreneurs' Synthetic Public Use Files (SynPUFs) in the Sentinel Common Data Model (SCDM) format. This transformation of data allows for the running of Sentinel's Routine Querying System tools, including the Cohort Identification and Descriptive Analysis (CIDA) tool, on the SynPUFs data. The CMS SynPUFs are available in the form of 20 mutually exclusive datasets, which together make up a 5% sample of the entire CMS database from 2008-2010. Each of the 20 datasets contains about 110,000 members. The intended use of these data in SCDM format is to generate familiarity with the CIDA tool and its capabilities and to allow for methodological expansion.

- 2.2M synthetic beneficiaries
- 20 mutually exclusive data samples

Mechanism to Transform Commercial Data

Submit Comment

SAS Code for Transforming the IBM MarketScan® Research Databases (MarketScan) into the Sentinel Common Data Model

Project Title	SAS Code for Transforming the IBM MarketScan® Research Databases (MarketScan) into the Sentinel Common Data Model
Date Posted	Tuesday, January 29, 2019
Status	Complete
Description	The Sentinel Operations Center and IBM Watson Health have partnered to make SAS® code available for transforming the IBM MarketScan® Commercial and Medicare Supplemental Databases into the Sentinel Common Data Model. If your organization currently licenses either of these databases and wishes to leverage the analytic infrastructure developed by Sentinel by transforming these data into the Sentinel Common Data Model, please click the 'Submit Comment' button on this page to request access. The Sentinel Operations Center will send you a MarketScan License Verification form. Contingent on license validation by IBM Watson Health, Sentinel will share the SAS code and documentation with your organization.

Sentinel Data Queries: Routine Querying Tools

Sentinel Infrastructure Supports Multiple Aims

Sentinel Infrastructure

Sentinel System

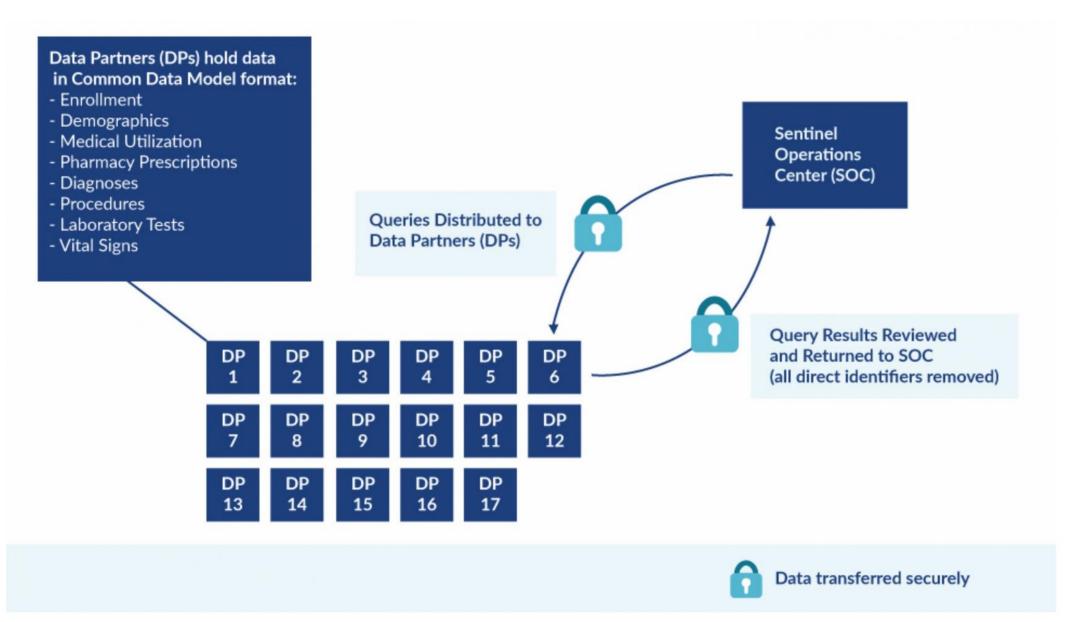
Routine queries and other activities that use pre-existing data

- PRISM
- BloodSCAN
- ARIA

FDA-Catalyst

Routine queries + interventions and interactions with members and/or providers

Sentinel is a Distributed Data Network

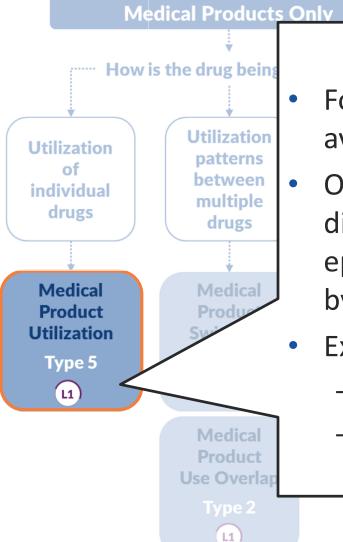


Active Risk Identification and Analysis (ARIA)



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

What are you investigating?



Outcomes Only

Medical Products & Outcomes

Medical Product Utilization (Type 5)

- Follow patient after "first valid" exposure episode for all available follow-up time in database.
- Output metrics include the number of patients, episodes, dispensings, and days supply; number of episodes by episode number, episode length; number of episode gaps by gap number, gap length.
- **Examples:**
 - Evaluate utilization patterns of obesity drugs
 - Exploratory study of biosimilar use in Sentinel

Risk Interval







Utiliz

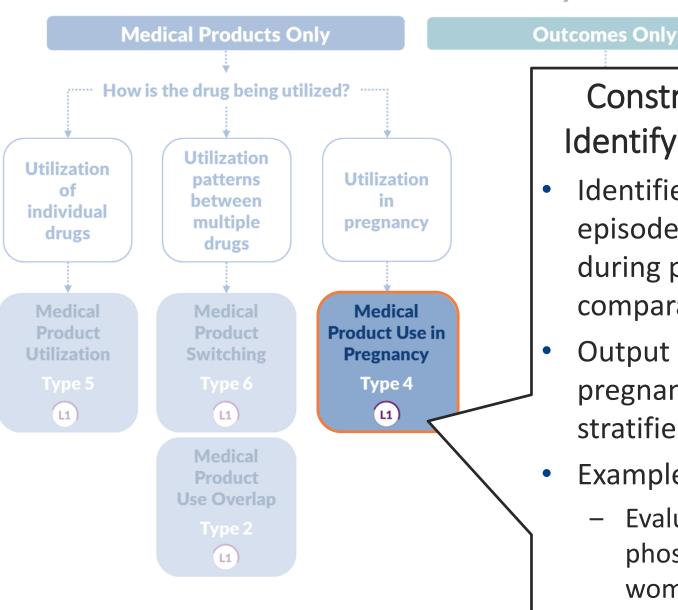
Utilization Patterns of Obesity Drugs

Project Title	Utilization Patterns of Obesity Drugs
Date Posted	Tuesday, March 19, 2019
Project ID	cder_mpl1r_wp129
Status	Complete
Deliverables	Sentinel Modular Program Report: Utilization Patterns of Obesity Drugs, Report 1 Sentinel Modular Program Report: Utilization Patterns of Obesity Drugs, Report 2
Description	This request examines utilization patterns of nine obesity drugs in the Sentinel Distributed Database (SDD) between January 1, 2008 and December 31, 2017. This request was distributed to 17 Data Partners on December 21, 2018.
Medical Product	benzphetamine bupropion/naltrexone diethylpropion liraglutide lorcaserin HCL orlistat phendimetrazine phentermine HCL phentermine/topiramate





What are you investigating?



Construct Pregnancy Episodes and Identify Medical Product Use (Type 4)

- Identifies live births to create pregnancy episodes and assesses medical product use during pregnancy episodes and in a comparator group of women.
- Output metrics include number of pregnancy episodes, medication use stratified by trimester.
- Example:
 - Evaluate utilization patterns of phosphodiesterase 5 inhibitors in pregnant women

Submit Comment

Phosphodiesterase Type 5 (PDE5) Inhibitor Utilization Among Women

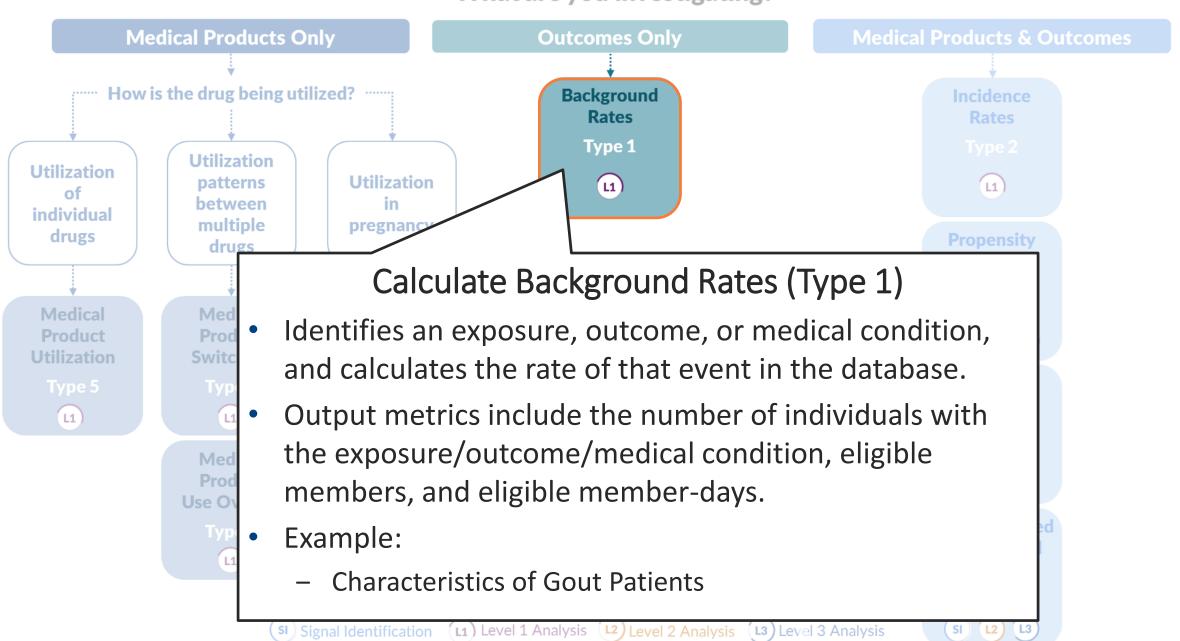
Utiliz indiv

> Me Pro Utili

Project Title	Phosphodiesterase Type 5 (PDE5) Inhibitor Utilization Among Women
Date Posted	Friday, October 12, 2018
Project ID	cder_mpl1r_wp111-112
Status	Complete
Deliverables	Sentinel Modular Program Report: Phosphodiesterase Type 5 (PDE5) Inhibitor Utilization Among Reproductive-Aged Women, Report 1
	Sentinel Modular Program Report: Phosphodiesterase Type 5 (PDE5) Inhibitor Utilization Among Pregnant Women, Report 2
Description	The goal of this query was to estimate phosphodiesterase type 5 (PDE5) inhibitor utilization among women in the Sentinel Distributed Database (SDD). Report 1 contains estimates of phosphodiesterase type 5 (PDE5) inhibitor use among reproductive-aged women. Report 2 contains estimates of PDE5 inhibitor use that occurred during a pregnancy ending in a live-born delivery or within 90 days prior to pregnancy start, among women. Data from January 1, 2001 to March 31, 2018 from 16 Data Partners contributing to the SDD were included in this report. This request was distributed to Data Partners on August 27, 2018.
Medical Product	phosphodiesterase type 5 (PDE5) inhibitor

(SI) Signal Identification (L1) Level 1 Analysis (L2) Level 2 Analysis (L3) Level 3 Analysis

What are you investigating?



Characteristics of Gout Patients and Use of Urate-Lowering Therapies

Project Title	Characteristics of Gout Patients and Use of Urate-Lowering Therapies
Date Posted	Friday, March 22, 2019
Project ID	cder_mpl1r_wp123, cder_mpl1r_wp126
Status	Complete
Deliverables	Sentinel Modular Program Report: Characteristics of Gout Patients and Use of Urate-Lowering Therapies, Report 1
	Sentinel Modular Program Report: Characteristics of Gout Patients and Use of Urate-Lowering Therapies, Report 2
	Sentinel Modular Program Report: Characteristics of Gout Patients and Use of Urate-Lowering Therapies, Report 3
Description	The goal of this request was to assess characteristics of gout patients and use of urate lowering therapies (ULT) among individuals in the Sentinel Distributed Database (SDD). This request contains three reports:
	 Report 1 examines counts of individuals with gout diagnoses, and cardiovascular morbidities and gout severity among those individuals. Report 2 contains counts of individuals using the ULTs febuxostat and allopurinol, and captures switching between ULT drug products and doses.
	 Report 3 contains cumulative exposure duration of febuxostat and allopurinol prior to dose or drug switching.

Utilization of individual drugs

Medical Product Jtilization

L1

What are you investigating?

Medical Products Only

Outcomes Only

Incidence

Rates

Type 2

(L1)

Develop Unadjusted Incidence Rates (Type 2)

- Identifies an exposure of interest and looks for the occurrence of health outcomes of interest (HOIs) during exposed time.
- Output metrics include number of exposure episodes and number of patients, number of health outcomes of interest, and days at-risk.
- Example:
 - SGLT-2 Inhibitor Use and Incidence of Diabetic Ketoacidosis

Propensity Score **Analysis**



Multiple Factor Matching



Self-Controlled **Risk Interval** Design





(L1)







SGLT-2 Inhibitor Use and Incidence of Diabetic Ketoacidosis in Patients with Diabetes Mellitus

Project Title	SGLT-2 Inhibitor Use and Incidence of Diabetic Ketoacidosis in Patients with Diabetes Mellitus
Date Posted	Tuesday, March 19, 2019
Project ID	cder_mpl1p_wp026
Status	Complete
Deliverables	Sentinel Modular Program Report: SGLT-2 Inhibitor Use and Incidence of Diabetic Ketoacidosis in Patients with Diabetes Mellitus
Description	The goal of this request was to estimate rates of diabetic ketoacidosis (DKA) among new users of sodium-glucose cotransporter-2 (SGLT-2) inhibitors canagliflozin, dapagliflozin, empagliflozin, or sitagliptin in the Sentinel Distributed Database (SDD). Data from March 1, 2013 through June 30, 2018 from 17 Data Partners contributing to the SDD were included in this report. This request was distributed to Data Partners on November 28, 2018.
Medical Product	canagliflozin dapagliflozin empagliflozin sitagliptin sodium-glucose cotransporter-2 (SGLT-2) inhibitor
Health Outcome	diabetic ketoacidosis

Utili:

Pro Jtili

What are you investigating?

Medical Products Only

Outcomes Only

Self-Controlled Risk Interval Design (Type 3)

- Identifies an exposure of interest, identifies an observation window relative to the exposure date, and examines the occurrence of outcomes during that window.
- Output metrics include number of exposure episodes, exposed individuals, individuals with an HOI in the risk and/or control windows, and censored individuals
- Example:
 - Seizure Risk following Ranolazine

Incidence

(L1)

Propensity Score **Analysis**



Multiple **Factor** Matching



Self-Controlled Risk Interval Design

Type 3





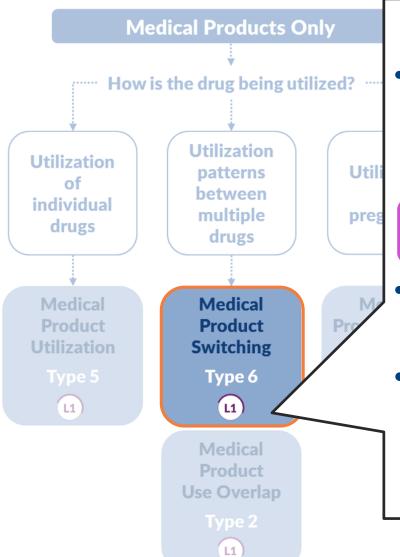


Submit Comment

Seizure following Ranolazine Use

Project Title	Seizure following Ranolazine Use
Date Posted	Thursday, January 3, 2019
Status	Complete
Deliverables	Sentinel Modular Program Report: Seizure following Ranolazine Use, Report 1
	Sentinel Modular Program Report: Seizure following Ranolazine Use: a Self-Controlled Risk Interval Analysis, Report 2
	Sentinel Modular Program Report: Seizure following Ranolazine Use: a Self-Controlled Risk Interval Analysis (an update to cder_mpl2p_wp002), Report 3
	Sentinel Analytic Packages: Seizure following Ranolazine Use: a Self-Controlled Risk Interval Analysis
Related Links	Prevalent and Incident Dispensings of Ranolazine
	2017 ICPE Symposium: Integrating Sentinel into Routine Regulatory Drug Review: A Snapshot of the First Year
	Seizure Algorithm Defined in "Seizure following Ranolazine Use: a Self-Controlled Risk Interval Analysis"
	Use of FDA's Sentinel System to Quantify Seizure Risk Immediately Following New Ranolazine Exposure

What are you investigating?



Switching Patterns (Type 6)

Captures utilization and switching patterns for userspecified groups that are based on any collection of National Drug Codes, Procedure Codes, etc.

Brand

Generic A

Generic B

Generic C

- Output Metrics include treatment episodes, switching patterns (e.g., $A \rightarrow B$, $A \rightarrow B \rightarrow A$), utilization metrics
- Examples
 - Metoprolol Extended Release
 - Lamotrigine Extended Release

Risk Interval Design

Submit Comment

Evaluation of Switching Patterns in FDA's Sentinel System: A New Tool to Assess Generic Drugs

Project Title	Evaluation of Switching Patterns in FDA's Sentinel System: A New Tool to Assess Generic Drugs
Date	Friday, August 17, 2018
Location	Drug Saf. 2018 Aug 17. doi: 10.1007/s40264-018-0709-4
Description	The aim of this study was to develop and implement a tool for analyzing manufacturer-level drug utilization and switching patterns within the U.S. Food and Drug Administration's Sentinel System. A descriptive tool was designed to analyze data in the Sentinel Common Data Model and was tested with two case studies, metoprolol extended release (ER) and lamotrigine ER, using claims data from four Sentinel Data Partners. This developed tool was able to elucidate novel utilization and switching patterns in two case studies. Such information can be used to support surveillance of generic drugs and biosimilars.









Sentinel's Public Documentation and SAS Program Depot (Public GIT) dev.sentinelsystem.org

Data Quality Review and Characterization Programs

Quality Assurance (QA) Package

Overview

This document describes the program package used to perform quality assurance (QA) review and characterization of data in the Sentinel Common Data Model (SCDM) format. This program package helps to ensure the data meets the necessary standards for data transformation consistency and quality.

Analytic programs that are executed against data that is not in SCDM format will likely yield errors. Successful execution of the QA package indicates that the source data adheres to SCDM rules. Note that data must be in the form of SAS® datasets in order to use these analytic programs.

Folder Structure

- docs: is where specifications are saved; specifications provide details about the request parameters and functionality of the QA package
- dplocal: is where datasets with patient identifiers are saved. For more information about Sentinel's privacy standards, please refer to The Sentinel System Principles and Policies.
- inputfiles: is the subfolder containing all input files and lookup tables needed to execute a request. Input files contain information on what tables should be output and the type of analyses conducted on the variables in each table
- msoc: is where aggregated program results are saved
- sasprograms: contains the file(s) to be executed

Requirements

- UNIX/Linux or Windows environment
- SAS version 9.3 or higher
- SCDM formatted data (Medicare Claims Synthetic Public Use Files are available in the Sentinel Common Data Model Format here)

Cohort Identification and Descriptive Analysis (CIDA)

SENTINEL ROUTINE QUERYING SYSTEM OVERVIEW

The purpose of this repository is to document version 8.0.3 of the Sentinel Routine Querying System, also known as the Query Request Package (QRP). This system is comprised of cohort identification and analytic modules.

This documentation describes QRP capabilities and provides the information required to build guery packages (i.e., input and output specifications) to address guestions of interest.

COHORT IDENTIFICATION AND DESCRIPTIVE ANALYSIS (CIDA) MODULE

QRP's Cohort Identification and Descriptive Analysis Module (CIDA) 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).

CIDA calculates descriptive statistics for the cohort(s) of interest and outputs datasets that may be useful for additional analyses.

CIDA Cohort Identification Strategies

- 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

Downloading Sentinel Analytic Packages **Sentinel Analytic Packages**

Overview

A Sentinel analytic package is a standard folder structure containing detailed user-defined specifications, input files, SAS® macros, and SAS programs used to conduct Sentinel's routine querying analyses. A package allows the user to select the cohort(s) of interest in order to examine their health profile and outcomes.

Sentinel's analytic request packages are intended to run on data formatted in accordance with the Sentinel Common Data Model (SCDM). Note that data must be in SAS datasets to use these analytic programs.

Analytic Request Packages Available for Download

Request ID	Summary
cder_mpl2p_wp011	Osteoporotic Fractures following Lupron Depot-PED Use: A Multiple Factor Matched Analysis
cder_mpl2p_wp016	Non-Melanoma Skin Cancer following Hydrochlorothiazide Use: A Propensity Score Matched Analysis
cder_mpl2p_wp007	Severe Uterine Bleed following Novel Oral Anticoagulants Use: A Propensity Score Matched Analysis
cder_mpl2r_wp008	Acute Myocardial Infarction and Hospitalized Heart Failure following Saxagliptin or Sitagliptin Use: A Propensity Score Matched Analysis
cder_mpl2p_wp009	Stroke, Gastrointestinal Bleeding, and Intracranial Hemorrhage following Apixaban or Warfarin Use in Patients with Non-Valvular Atrial Fibrillation: A Propensity Score Matched Analysis
cder_mpl2p_wp006	Seizure following Ranolazine Use: A Self-Controlled Risk Interval Analysis (an update to cder_mpl2p_wp002)
cder_mpl2p_wp005	Stroke following Atypical Antipsychotic or Z-Hypnotic Use in Patients with Prior Use of Selective Serotonin Reuptake Inhibitors (SSRIs): A Propensity Score Matched Analysis
cder_mpl2p_wp001	Venous Thromboembolism following Continuous or Extended Cycle Contraceptive Use: A Propensity Score Matched Analysis
cder_mpl2p_wp004	Stroke following Typical or Atypical Antipsychotic Use in non-Elderly Patients: A Propensity Score Matched Analysis
cder_mpl2p_wp002	Seizure following Ranolazine Use: A Self-Controlled Risk Interval Analysis

Questions?

info@sentinelsystem.org