# Applications of RWD in FDA's Sentinel System

Sarah K. Dutcher, PhD, MS
Office of Surveillance & Epidemiology
Center for Drug Evaluation & Research
Food and Drug Administration

DIA Real-World Evidence Conference
October 24, 2024

# Disclaimer

The views expressed in this presentation are those of the presenter and not necessarily those of the US Food and Drug Administration.

The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to DIA, its directors, officers, employees, volunteers, members, chapters, councils, Communities or affiliates.

This presentation is incomplete without accompanying verbal commentary.

# Outline

1. Expanding Sentinel System data and analytic capabilities
   - Incorporating linked claims-EHR data
   - Leveraging artificial intelligence to evaluate medical products
2. Sentinel's PDUFA VII commitments
   - Pregnancy safety
   - Negative controls

# Recognizing the Need to Harness Alternative Data Sources


"Previously, our investments in post-market data have mostly focused on the development of systems to consolidate and analyze information derived from **healthcare payer claims**…Now we have the capacity to use clinical data derived from **electronic health records** to develop faster reporting on the performance of medical products in real world medical settings."

https://www.fda.gov/news-events/fda-voices/fda-budget-matters-cross-cutting-data-enterprise-real-world-evidence


"Strategic aim 3: Accelerate access to and broader use of real-world data"

"…FDA will focus its investment on innovations emerging from **new data science disciplines**, such as natural language processing and machine learning, and seek to **expand its access to and use of EHRs**"

https://www.fda.gov/media/120333/download

# Sentinel Innovation Center Vision

## Current Sentinel System Limitations

Inability to identify certain study populations of interest from insurance claims

Inability to identify certain outcomes of interest from insurance claims

Other limitations (inadequate duration of follow-up, the need for additional signal identification tools)

## Sentinel Innovation Center Initiatives

### Data Infrastructure


### Feature Engineering

Emerging methods including machine learning and scalable automated NLP approaches to enable computable phenotyping from unstructured EHR data

### Causal Inference

Methodologic research to address specific challenges when using EHRs such as approaches to handle missing data, calibration methods for enhanced confounding adjustment

### Detection Analytics

Development of signal detection approaches to account for and leverage differences in data content and structure of EHRs

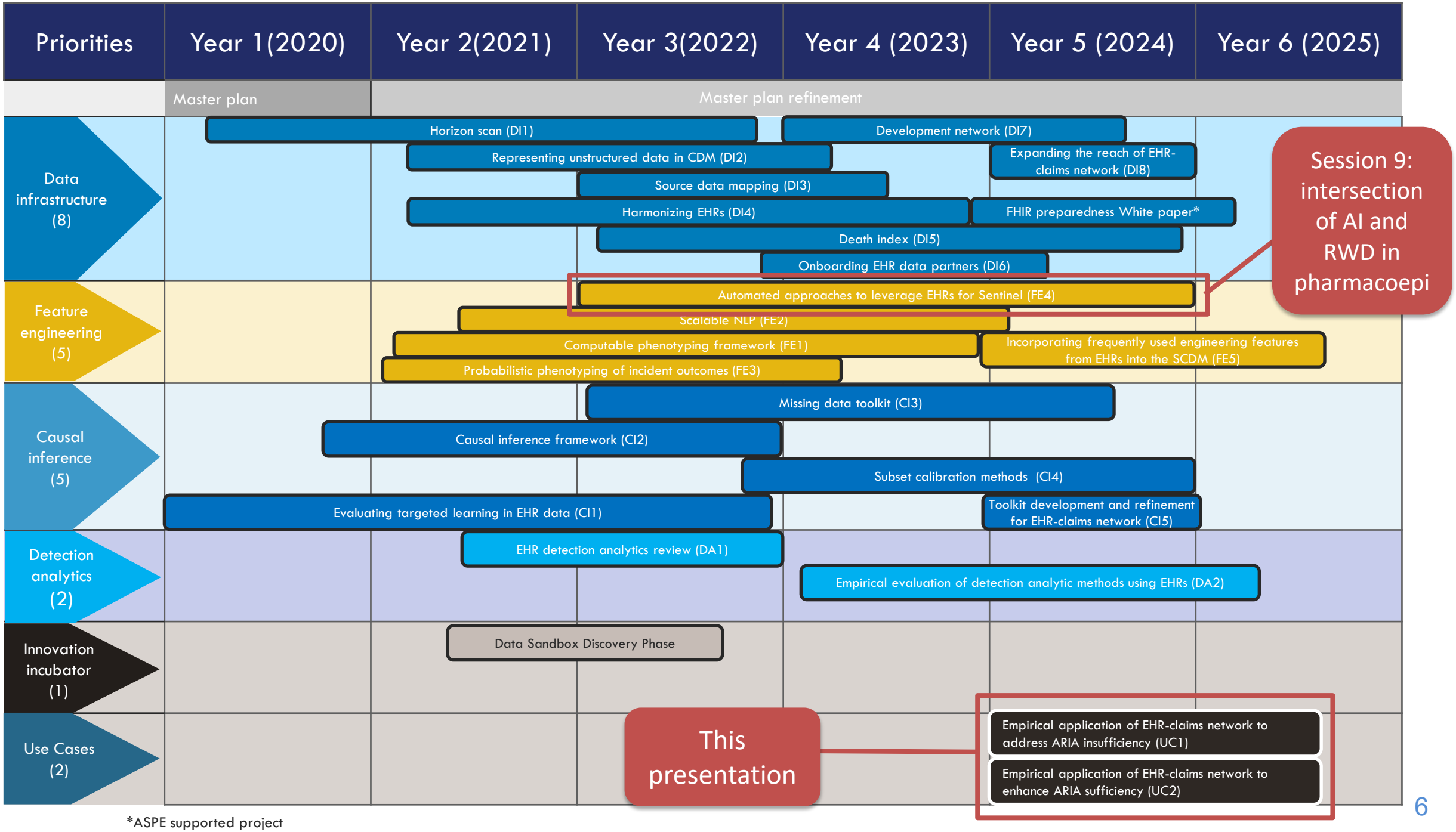
## Sentinel Innovation Center Vision

**RWE Data Enterprise**: A query-ready, quality-checked distributed data network containing EHR for at least 10 million lives with reusable analysis tools

**2020** → **2024**



Desai RJ, Matheny ME, Johnson K, et al. Broadening the reach of the FDA Sentinel system: A roadmap for integrating electronic health record data in a causal analysis framework. NPJ Digit Med. 2021;4(1):170.



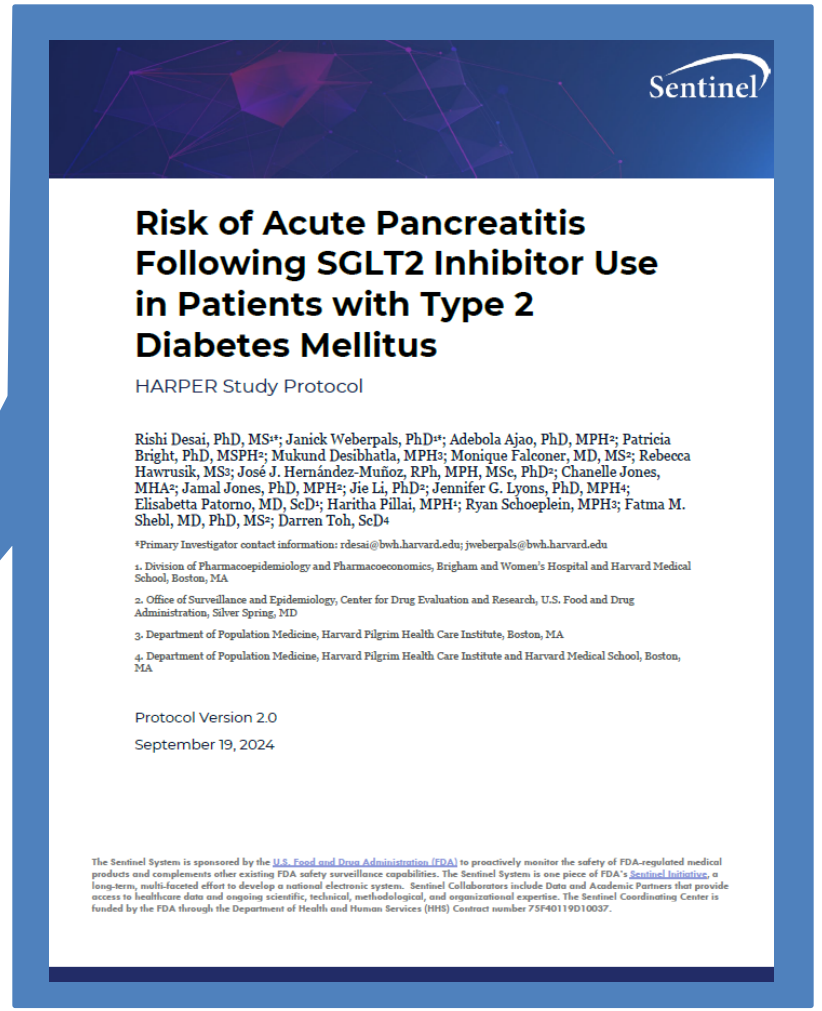

*ASPE supported project

# Use Case 1: Improving ARIA Sufficiency

- Due to limitations inherent to claims data, Sentinel's Active Risk Identification and Analysis (ARIA) system is sometimes deemed insufficient to address a regulatory question of interest
- Linking EHRs and claims data and incorporating advanced methods can overcome some of ARIA's current limitations

  **Aim 1**: For health outcomes of interest for which ARIA analyses were previously determined to be insufficient, conduct fitness-for-purpose analyses and assess the likelihood of successful development of computable phenotypes by incorporating rich EHR data and data-driven modeling methods

  **Aim 2**: Conduct a protocol-based pharmacoepidemiologic analysis to evaluate the complexities encountered and propose solutions for typical claims-based ARIA analyses that will be handled by linked EHR-claims data



Sentinel

**Risk of Acute Pancreatitis Following SGLT2 Inhibitor Use in Patients with Type 2 Diabetes Mellitus**

HARPER Study Protocol

Rishi Desai, PhD, MS[1*]; Janick Weberpals, PhD[1*]; Adebola Ajao, PhD, MPH[2]; Patricia Bright, PhD, MSPH[2]; Mukund Desibhatla, MPH[3]; Monique Falconer, MD, MS[2]; Rebecca Hawrusik, MS[3]; José J. Hernández-Muñoz, RPh, MPH, MSc, PhD[2]; Chanelle Jones, MHA[2]; Jamal Jones, PhD, MPH[2]; Jie Li, PhD[2]; Jennifer G. Lyons, PhD, MPH[4]; Elisabetta Patorno, MD, ScD[1]; Haritha Pillai, MPH[1]; Ryan Schoeplein, MPH[3]; Fatma M. Shebl, MD, PhD, MS[2]; Darren Toh, ScD[4]

[*]Primary Investigator contact information: rdesai@bwh.harvard.edu; jweberpals@bwh.harvard.edu

1. Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

2. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD

3. Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA

4. Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA

Protocol Version 2.0

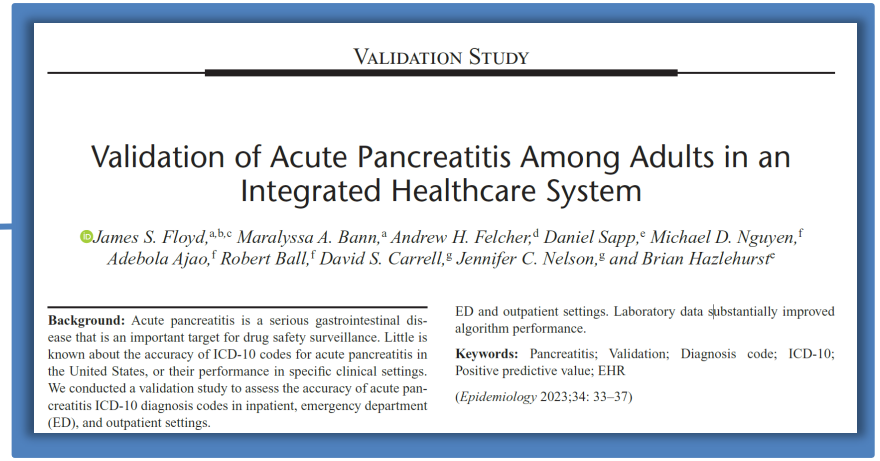
September 19, 2024

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number 75F40119D10037.

https://www.sentinelinitiative.org/methods-data-tools/methods/empirical-application-sentinel-electronic-health-record-ehr-and-claims

# Pharmacoepidemiology Study in Linked EHR-claims Data

- **Context**: In 2017, ARIA was determined to be insufficient to assess the risk of acute pancreatitis with use of SGLT-2 inhibitors
  - No longer an active safety concern, but data challenges (e.g., outcome identification) are still relevant
- Diagnosis codes are known to have limited ability to identify acute pancreatitis (PPV: 55-66%), which raises concerns regarding the validity of prior studies due to outcome misclassification
- **Data Source**: Sentinel RWE Data Enterprise commercial network
- **Approach**: A cohort study using propensity-score fine stratification for confounding adjustment
  - Outcome: acute pancreatitis, defined using a probabilistic phenotyping algorithm
  - Applying multiple imputation methods to analytically address missingness in key confounding variables (e.g., HbA1c and BMI)
- **Status**: Results anticipated by the end of the year



VALIDATION STUDY

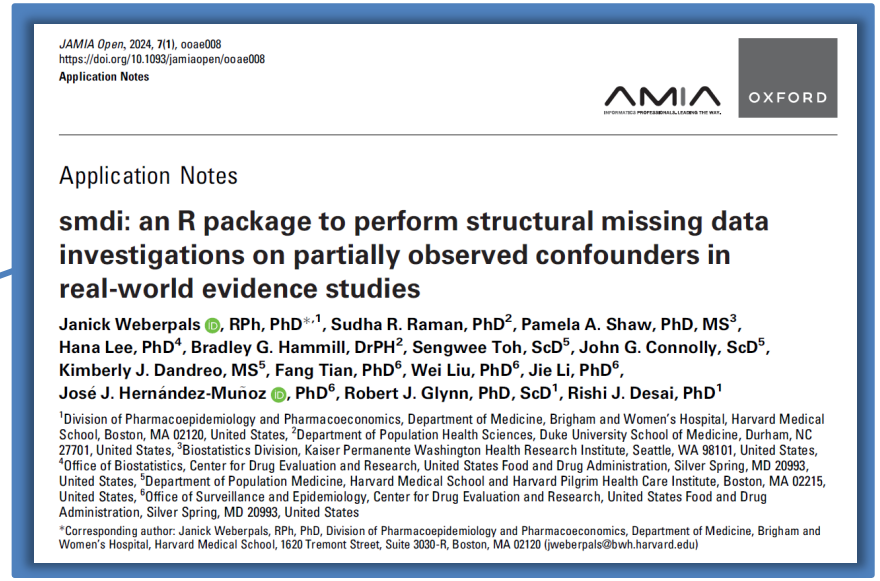
Validation of Acute Pancreatitis Among Adults in an Integrated Healthcare System

James S. Floyd,[a,b,c] Maralyssa A. Bann,[a] Andrew H. Felcher,[d] Daniel Sapp,[e] Michael D. Nguyen,[f] Adebola Ajao,[f] Robert Ball,[f] David S. Carrell,[g] Jennifer C. Nelson,[g] and Brian Hazlehurst[e]

**Background:** Acute pancreatitis is a serious gastrointestinal disease that is an important target for drug safety surveillance. Little is known about the accuracy of ICD-10 codes for acute pancreatitis in the United States, or their performance in specific clinical settings. We conducted a validation study to assess the accuracy of acute pancreatitis ICD-10 diagnosis codes in inpatient, emergency department (ED), and outpatient settings. Laboratory data substantially improved algorithm performance.

**Keywords:** Pancreatitis; Validation; Diagnosis code; ICD-10; Positive predictive value; EHR

(*Epidemiology* 2023;34: 33–37)



*JAMIA Open*, 2024, 7(1), ooae008
https://doi.org/10.1093/jamiaopen/ooae008
**Application Notes**

AMIA OXFORD

Application Notes

**smdi: an R package to perform structural missing data investigations on partially observed confounders in real-world evidence studies**

Janick Weberpals, RPh, PhD*,[1], Sudha R. Raman, PhD[2], Pamela A. Shaw, PhD, MS[3], Hana Lee, PhD[4], Bradley G. Hammill, DrPH[2], Sengwee Toh, ScD[5], John G. Connolly, ScD[5], Kimberly J. Dandreo, MS[5], Fang Tian, PhD[6], Wei Liu, PhD[6], Jie Li, PhD[6], José J. Hernández-Muñoz, PhD[6], Robert J. Glynn, PhD, ScD[1], Rishi J. Desai, PhD[1]

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02120, United States, [2]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC 27701, United States, [3]Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA 98101, United States, [4]Office of Biostatistics, Center for Drug Evaluation and Research, United States Food and Drug Administration, Silver Spring, MD 20993, United States, [5]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA 02215, United States, [6]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, United States Food and Drug Administration, Silver Spring, MD 20993, United States

*Corresponding author: Janick Weberpals, RPh, PhD, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont Street, Suite 3030-R, Boston, MA 02120 (jweberpals@bwh.harvard.edu)

Floyd JS, Bann MA, Felcher AH, et al. Validation of Acute Pancreatitis Among Adults in an Integrated Healthcare System. *Epidemiology*. 2023;34(1):33-37.
Bann MA, Carrell DS, Gruber S, et al. A comparison of manual and automated approaches to developing computable algorithms for identifying acute pancreatitis. Under review.
Weberpals J, Raman SR, Shaw PA, et al. smdi: an R package to perform structural missing data investigations on partially observed confounders in real-world evidence studies. *JAMIA Open*. Apr 2024;7(1):ooae008.

# Use Case 2: Strengthening ARIA Sufficient Analyses

- Although ARIA analyses provide vital information to the FDA to aid in regulatory decision making, often uncertainties remain due to lack of data availability in insurance claims for critical variables pertaining to the research question (e.g., residual confounding, lack of validated outcome algorithms)

  **Aim 1**: Rapid confounder balance evaluation of factors unmeasured in Sentinel claims data

  **Aim 2**: Correcting claims analyses for unmeasured confounding using subset calibration tools

  **Aim 3**: Real-time validation of code-based algorithms

  **Aim 4**: Identifying use of cannabis-derived products (CDP) from free-text notes

  **Aim 5**: Expand on a principled quantitative bias analysis (QBA) at the design stage that could allow for better understanding of the uncertainties associated with potential unmeasured confounding



https://www.sentinelinitiative.org/methods-data-tools/methods/empirical-application-sentinel-ehr-and-claims-data-partner-network

# Identifying Use of Cannabis-Derived Products (CDP) From Free-Text Notes

- **Context**: There is increasing interest in the potential utility of cannabis for a variety of medical conditions, as well as research on the potential adverse health effects from use of cannabis
  - FDA has not approved cannabis for the treatment of any disease/condition, but has approved Epidiolex®, a cannabis-derived product (CDP) in the form of a cannabidiol (CBD)
  - Despite increased patient usage of unapproved cannabis-derived products in recent years, CDP is not currently captured in structured claims data
  - Usage of cannabis-derived products, however, may be recorded in unstructured patient-reported data EHRs
- **Data Source**: Sentinel RWE Data Enterprise development network (Vanderbilt University Medical Center)
- **Approach**: Develop a method for capturing patient usage of CDP from linked EHR-claims data and perform exploratory analyses to characterize patients using CDP
  - Identify individuals with suspected CDP exposure in structured EHR data
  - Use NLP tools and algorithms to find CDP exposure in text from clinical notes (iterative process)
  - Analyze patient cohorts identified based on exposure to Epidiolex, CBD, and other CDP to understand demographics, clinical characteristics, and comedications
- **Status**: results anticipated in Spring 2025

U.S. Food and Drug Administration. "FDA and Cannabis Research and Drug Approval Process." https://www.fda.gov/news-events/public-health-focus/fda-and-cannabis-research-and-drug-approval-process
Carrell DS, Cronkite DJ, Shea M, et al. Clinical documentation of patient-reported medical cannabis use in primary care: Towards scalable extraction using natural language processing methods. Subst Abus. 2022;43(1): 917-924.

# Sentinel's PDUFA VII Commitments

**M. ENHANCEMENT AND MODERNIZATION OF THE FDA DRUG SAFETY SYSTEM**

FDA will contin
scientific approa
prevention, and
regulatory activ
system will imp
access to needed

User fees will p
2) optimization
capabilities and
analytic capabil
the understandi

**2. Optimization of the Sentinel Initiative**

The us
continu
Effecti
quality
metho
compr
capabi
to addr
eviden

i. Pregnancy Safety

The goal of pregnancy safety
studies is to inform labeling o
or evaluate safety signals in a

(1) FDA will develop a fr
types of post-market p
used, incorporating kn
market studies have be
identifying gaps in kn
demonstration projects
such as, but not limite
studies, anticipated ex
(FRP) and pregnant w
proposed risk mitigati
type of risk to be detec
address the use of preg
data sources including
efficient means of obta

ii. Use of Real-World Evidence – Negative Controls

FDA is building Sentinel/BEST methodology to improve understanding of robustness evaluations used to address the consistency of RWE with respect to study design, analysis, or variable measurement. FDA will develop new methods to support causal inference in Sentinel/BEST that could address product safety questions and advance our understanding of how RWE may be used for studying effectiveness.

(1) By September 30, 2023, FDA will hold a public workshop on use of negative controls for assessing the validity of non-interventional studies of treatment and the proposed Sentinel Initiative projects.

(2) FDA will initiate two methods development projects by September 30, 2024 to 1) develop an empirical method to automate the negative control identification process in Sentinel and integrate it into the Sentinel System tools; and 2) develop a method to use a double negative control adjustment to reduce unmeasured confounding in studying effectiveness of vaccines.

(3) By September 30, 2027, FDA will publish a report on the results of the development projects.

# Sentinel's PDUFA VII Commitments: Milestones

| Milestone | Due Date |
|---|---|
| FDA will hold a public workshop on post-market safety studies in pregnant women | September 30, 2023 |
| FDA will:<br>• Publish a workshop report describing the proposed framework<br>• Initiate demonstration projects | September 30, 2024 |
| FDA will, based on the results of the demonstration projects:<br>• Update the proposed **framework**<br>• Develop a **guidance or MAPP/SOPP** as appropriate to implement a standardized process for determining necessity and type of pregnancy postmarketing studies including PMRs | September 30, 2027 |
| FDA will hold a public workshop on use of negative controls for assessing the validity of non-interventional studies of treatment | September 30, 2023 |
| FDA will initiate two **methods development projects** | September 30, 2024 |
| FDA will publish a **report** on the results of the development projects | September 30, 2027 |

# Sentinel PDUFA VII Commitments: Pregnancy Safety Demonstration Projects

| Commitments | |
|---|---|
| A | **CDER:** Assess the performance of **pregnancy registries** versus electronic healthcare **database studies** to **detect a signal** when the exposure to medication in pregnancy is **relatively common**. |
| B | **CDER**: Assess the performance of **single arm safety studies** versus signal identification methods using electronic healthcare data **to detect a signal** when the **exposure to medication** in pregnancy is anticipated to be **low.** |
| C | **CDER:** Assess the performance of **pregnancy registries** versus electronic h… to **evaluate a signal** when the exposure to medication in pregnancy is **rela…** |
| D | **CDER:** Assess the performance o… **rmations (MC…** signal detection and eval… but not al… |
| E | **CBER**: Assess the **perf**… lth record (EHR) and claims-linked healthcare data for a … utcomes (e.g., spontaneous abortion, stillbirth, congenital m… **ant women**. The parameters of the pregnancy-outcome algor… al usability with therapeutic products. |

Using a product with low exposure to imitate a descriptive pregnancy safety study (DPSS)

Similar approaches for "A" and "C" since both involve registries and a common exposure

Differs in regulatory goal: signal detection vs. evaluation

# Sentinel PDUFA VII Commitments: Use of RWE - Negative Controls

| Commitments | |
|---|---|
| 1 | **CDER:** Develop an empirical method to **automate the negative control identification** process in Sentinel and **integrate** it into the Sentinel System tools |
| 2 | **CBER**:Develop a method to use a **double negative control adjustment** to reduce unmeasured confounding in studying effectiveness of vaccines |

An automated approach to find disconnected negative controls: Data-driven Automated Negative Control Estimation (**DANCE**) algorithm



Kummerfeld E, Lim J, Shi X. Data-driven Automated Negative Control Estimation (DANCE): Search for, Validation of, and Causal Inference with Negative Controls. Journal of Machine Learning Research. 2024;24:1-35.

# QUESTIONS?

FDASentinelSystem@fda.hhs.gov


FDA
U.S. FOOD & DRUG
ADMINISTRATION