INFORMATICS PROFESSIONALS. LEADING THE WAY.

Drug Safety Surveillance with EHR Data: A Promise Not Yet Fulfilled

S16: Oral Presentations - Pharmacoinformatics

Joshua C. Smith, PhD
Vanderbilt University Medical Center
#AMIA2023

Disclosure

I and my spouse/partner have no relevant relationships with commercial interests to disclose.

Disclaimer

- This project was supported by Task Order 7540119F19002 under Master Agreement 75F40119D10037 from the U.S. Food and Drug Administration (FDA).

- The views expressed in this presentation represent those of the presenter and do not necessarily represent the official views of the U.S. FDA.

Learning Objectives

After participating in this session, attendees will:

- Become familiar with the current state of EHR-based medication safety signal identification
- Understand limitations of existing research
- Appreciate how adoption of best practices would benefit future work in this area

Introduction

- Premarket clinical trials characterize adverse reactions to medications, but controlled studies do not reflect the variability of the general population
- Pharmacovigilance programs such as FAERS and Sentinel protect patient health and safety by identifying adverse event **signals** through post-marketing surveillance of spontaneous reports and insurance claims data
    - **Spontaneous reporting systems** are limited to voluntary reports of suspected ADEs, subject to reporting bias, and provide limited information on patients
    - **Claims data**, while well-suited for longitudinally following patients after exposures, often lack clinical granularity and under-capture subtle events that do not trigger formal coding

Introduction

- In contrast, **Electronic Health Records** (EHRs) contain comprehensive structured and rich, unstructured clinical narratives
- Leveraging EHR data may improve safety signal identification and is one of the primary goals of the FDA Sentinel Innovation Center
  - EHR data may help address shortcomings of current approaches
  - Current signal identification methods used for claims and spontaneous reports may not fully leverage the breadth and depth of EHR data
  - EHR-based methods should ideally incorporate myriad data types, control for the impact of confounding, and address variability in terms of data collection processes, gaps, and quality across health systems

Objective

- To support the expansion of EHR-based pharmacovigilance, we conducted a literature review of current practices in the use of routinely collected EHR data for medication safety signal identification
  - Studies were **eligible** if they (1) analyzed patient-level EHR data collected through routine clinical care; and (2) implemented, evaluated, or proposed analytic methods for identifying or discovering adverse event signals.
  - This **included** studies that performed hypothesis-free discovery, as well as those that attempted to identify associations between specific drug-event pairs or among sets of related exposures and events
  - We focused on original research and **excluded** commentaries, prior literature reviews, and studies focused on recognizing known ADEs or using non-routine data such as pharmacogenomics

Search Criteria

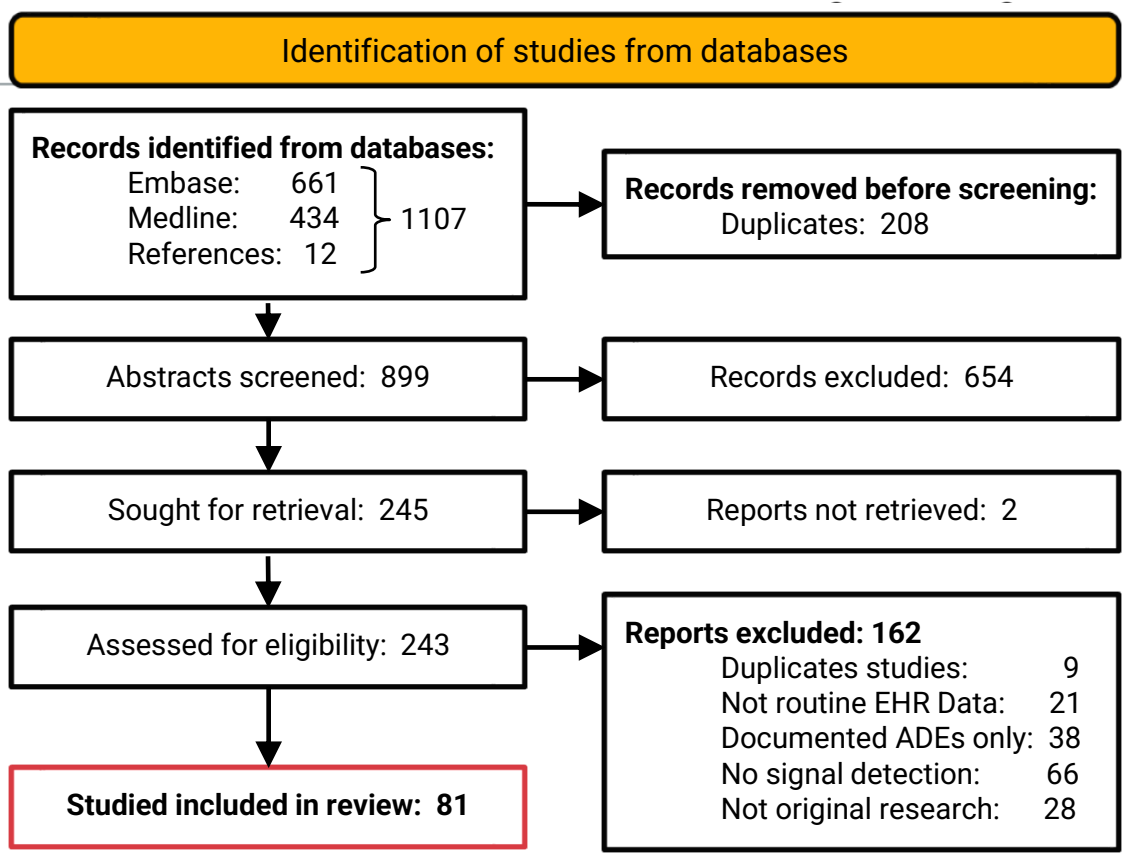
Biomedical domain		Analytic methods		Data context

MeSH
• Product Surveillance, Postmarketing
• Pharmacovigilance
• Drug-Related Side Effects and Adverse Reactions
• Drug Interactions
OR
*Keywords**
• Pharmacovigilance
• Drug/medication +
  • Surveillance
  • Safety
  • Interaction(s)
  • Reaction(s)
  • Adverse reaction(s)
  • Side effect(s)
  • Toxicity/toxicities
	AND	*MeSH*
		• Data mining
		• Artificial intelligence
		• Machine learning
		• Algorithms
		• Natural language processing
		• Pattern Recognition, Automated
		• Models, Statistical
		OR
		Keywords
		• Text/data mining
		• Natural language processing/NLP
		• Machine learning
		• Artificial intelligence
		• Deep learning
		• Data-driven
		• Signal detection
		• Signal identification
			AND	*MeSH*
				• Electronic health records
				• Medical records
				• Medical Records Systems, Computerized
				OR
				Keywords
				• Electronic health record(s)/EHR(s)
				• Electronic medical record(s)/EMR(s)
				• Clinical narrative(s)/note(s)/text
				• Medical record(s)
				• Observational clinical data

* Keyword searches in titles and abstracts

Results

- Search returned 1095 publications + 12 added from references for a total of **1107 publications**
- Abstracts were screened by two reviewers
- Full articles screened by primary reviewers
- 81 studies included in final review



Characteristics of reviewed studies*

	N	%
Analysis frame		
Retrospective	75	93%
Prospective	6	9%
Study design		
Cohort	54	67%
Case-control	22	27%
Self-controlled	12	15%
Signal Identification Target		
Specified associations	27	33%
Across many exposures	44	54%
Across many outcomes	34	42%
Methods		
Disproportionality	48	59%
Regression	25	31%
Machine learning and data mining	30	37%
Sequential analyses	3	4%

	N	%
EHR components accessed		
Demographics	46	57%
Medication orders	75	93%
Diagnostic/procedural codes	60	74%
Laboratory results	28	35%
Vital signs	3	4%
Clinical text	21	26%
Control for confounding		
None reported	32	40%
Adjustment for demographics	35	43%
Adjustment for other drug exposures	14	17%
Adjustment for comorbidities	22	27%
Statistical considerations		
Explicit consideration of temporal constraints	74	91%
Discussion of missing data	6	7%
Adjustment for multiple testing	20	24%

**Categories not mutually exclusive*

Discussion

The current state of research involving EHR-based signal identification is promising but would benefit from a more systematic approach

Methods and data models should also take advantage of the full breadth and depth of EHR data

Important to both define success metrics and estimate performance to better understand the types of signals that would be captured poorly in EHR data.

Development of a common data model for tailoring longitudinal EHR datasets for pharmacovigilance studies would simplify application and evaluation

Development of a common resource of known adverse event signals and control drug-event associations would support comparative evaluations

Discussion

Increased use of propensity score methods could leverage EHR data to address high-dimensional confounding

Distributed analyses across health systems could leverage larger datasets while protecting privacy; prior applications in claims-based data provide roadmap for wider implementation with EHRs

Simple approaches could provide initial screening and hypothesis generation to motivate more detailed, complex investigations

Recent advances (e.g., DDI-WAS) highlight potential for hypothesis-free discovery applications

Opportunities to leverage hierarchical relationships among clinical concepts through approaches such as tree-based scan statistic with TreeScan®

Next Steps

- A goal of the **Sentinel Innovation Center (IC)** is to expand signal identification methodology and capacity to carry out surveillance activities with EHR data
- **Sentinel IC Development Network**, linked claims and EHR data from IC lead hubs at Mass General Brigham, Kaiser Permanent Washington, Duke, and VUMC
- **Commercial Data Partner Linked EHR+Claims Network** (TriNetX, HealthVerity)
- *Development and Evaluation of EHR Information Extraction Pipeline and Tree-Based Scan Statistic Methods for EHR-Based Signal Detection*, an IC detection analytics project with Shirley Wang at MGB
- *Comparison of Adverse Drug Event Signal Detection Methods in EHR Data*, a Master's thesis project with Julie V. Kim at VUMC

Acknowledgements

The FDA Sentinel Initiative

The FDA Sentinel Innovation Center

Innovation Center Detection Analytics Workgroup

My coauthors:
Sharon E. Davis · Luke Zabotka · Rishi J. Desai · Shirley V. Wang
Judith C. Maro · Kevin Coughlin · José J. Hernández-Muñoz
Danijela Stojanovic · Nigam H. Shah

Thank you!

joshua.smith@vumc.org

Davis, et al. Use of Electronic Health Record Data for Drug Safety Signal Identification: A Scoping Review. Drug Saf. 2023 Aug;46(8):725-742.