A Practical Overview of Real-Life Challenges Leveraging EHR Data for the FDA Sentinel System

Rishi J. Desai, MS, PhD
Associate Professor of Medicine
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine
Brigham and Women's Hospital, Harvard Medical School, Boston, MA

rdesai@bwh.harvard.edu @Rishidesai11

February 14, 2025

Disclosures

This work was supported by the US Food and Drug Administration Sentinel Innovation Center Contract under Master Agreement No. 75F40119D10037.

The contents are those of the authors and do not necessarily represent the official views of, nor an endorsement, by FDA/CDC/HHS, or the US Government.

What is Sentinel?

Public Law 110–85
110th Congress

An Act

To amend the Federal Food, Drug, and Cosmetic Act to revise and extend the user-fee programs for prescription drugs and for medical devices, to enhance the postmarket authorities of the Food and Drug Administration with respect to the safety of drugs, and for other purposes.

Sept. 27, 2007
[H.R. 3580]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

Food and Drug Administration Amendments Act of 2007.
21 USC 301 note.

This Act may be cited as the "Food and Drug Administration Amendments Act of 2007".

SEC. 905. ACTIVE POSTMARKET RISK IDENTIFICATION AND ANALYSIS.

(a) IN GENERAL.—Subsection (k) of section 505 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 355) is amended by adding at the end the following:

"(3) ACTIVE POSTMARKET RISK IDENTIFICATION.—

"(A) DEFINITION.—In this paragraph, the term 'data' refers to information with respect to a drug approved under this section or under section 351 of the Public Health Service Act, including claims data, patient survey data, standardized analytic files that allow for the pooling and analysis of data from disparate data environments, and any other data deemed appropriate by the Secretary.

"(B) DEVELOPMENT OF POSTMARKET RISK IDENTIFICATION AND ANALYSIS METHODS.—The Secretary shall, not later than 2 years after the date of the enactment of the Food and Drug Administration Amendments Act of 2007, in collaboration with public, academic, and private entities—

"(i) develop methods to obtain access to disparate data sources including the data sources specified in subparagraph (C);

"(ii) develop validated methods for the establishment of a postmarket risk identification and analysis system to link and analyze safety data from multiple sources, with the goals of including, in aggregate—

"(I) at least 25,000,000 patients by July 1, 2010; and

"(II) at least 100,000,000 patients by July 1, 2012; and

"(iii) convene a committee of experts, including individuals who are recognized in the field of protecting data privacy and security, to make recommendations to the Secretary on the development of tools and methods for the ethical and scientific uses for, and communication of, postmarketing data specified under subparagraph (C), including recommendations on the development of effective research methods for the study of drug safety questions.

"(C) ESTABLISHMENT OF THE POSTMARKET RISK IDENTIFICATION AND ANALYSIS SYSTEM.—

"(i) IN GENERAL.—The Secretary shall, not later than 1 year after the development of the risk identification and analysis methods under subparagraph (B), establish and maintain procedures—

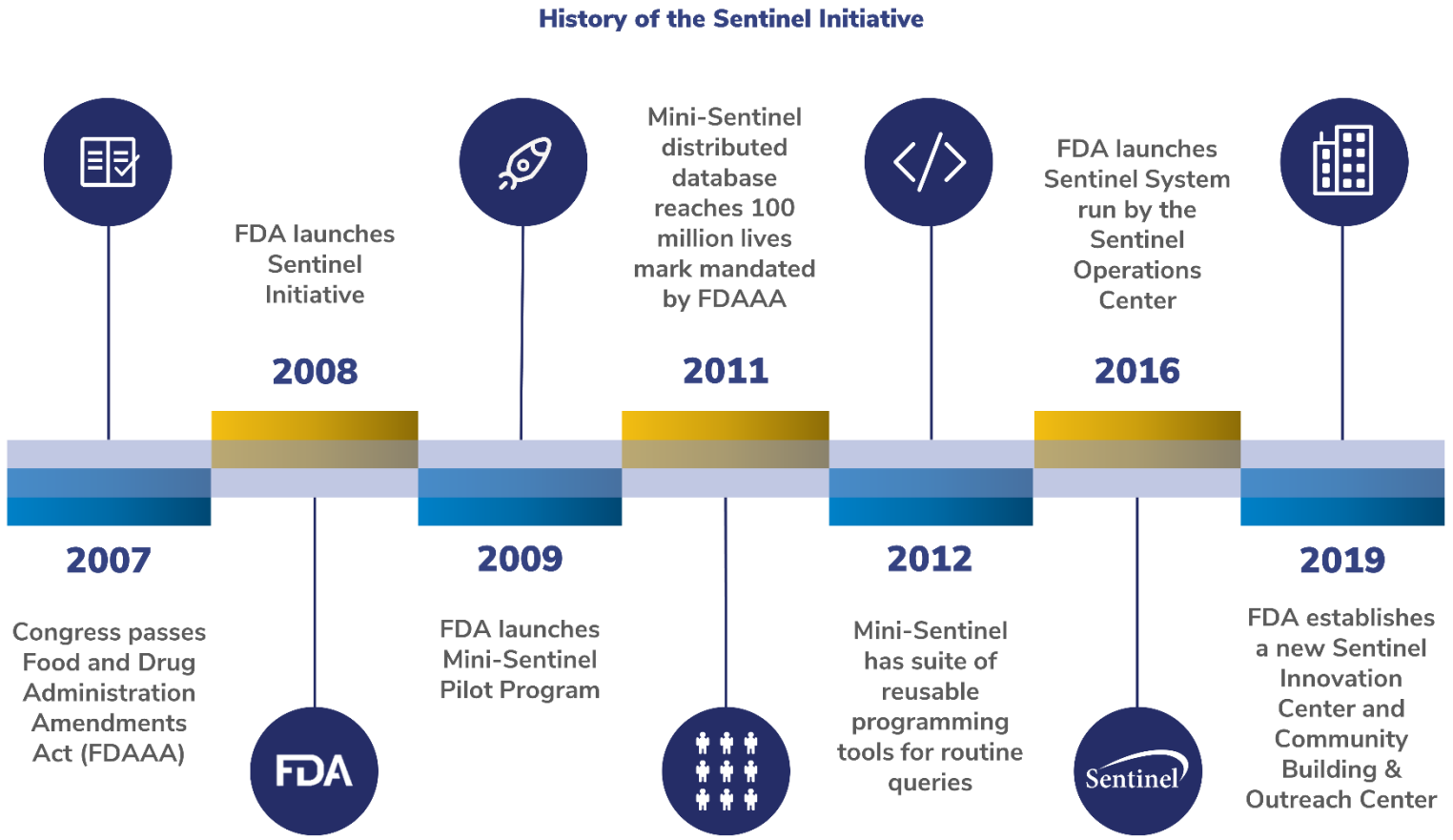
Establishment of a postmarket risk identification and analysis system

https://www.govinfo.gov/content/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf

FDA's Sentinel System

2007 FDA Amendments Act mandates FDA to establish ***active surveillance system*** for monitoring drugs using electronic healthcare data.

Through the Sentinel Initiative, FDA aims to assess the post-marketing safety of approved medical products.



Sentinel Distributed Database (SDD)

1. Aetna, a CVS Health company
2. Duke University School of Medicine: Department of Population Health Sciences (Medicare Fee-for-Service data)
3. Harvard Pilgrim Health Care Institute
4. HCA Healthcare
5. HealthCore, Inc. (Anthem, Inc. data)
6. HealthPartners Institute
7. Humana, Inc.
8. Kaiser Permanente Colorado Institute for Health Research
9. Kaiser Permanente Hawaii Center for Integrated Health Care Research
10. Kaiser Foundation Health Plan of the Mid-Atlantic States, Inc.
11. Kaiser Permanente Northwest Center for Health Research
12. Kaiser Permanente Washington Health Research Institute
13. Marshfield Clinic Research Institute
14. Optum (OptumInsight Life Sciences Inc. and Optum Labs®)
15. Vanderbilt University Medical Center, Department of Health Policy (Tennessee Medicaid data)

351.8 million unique patient identifiers (2000-2020)*

70.6 million members currently accruing new data

14.8 billion pharmacy dispensings

13.8 billion unique medical encounters

*Potential for double-counting if individuals moved between Data Partner health plans

Recognizing the need to harness alternative data sources and methods

FDA Budget Matters: A Cross-Cutting Data Enterprise for Real World Evidence

Share Tweet Linkedin Email Print

June 10, 2018

By: Scott Gottlieb, M.D.

Over time, as our experience with new medical products expands, our knowledge about how best to maximize their benefits and minimize any potential risks, sharpens with each data point we gather. Every clinical use of a product produces data that can help better inform us about its safety and efficacy.

The FDA is committed to developing new tools to help us access and use data collected from all sources. This includes ways to expand our methodological repertoire to build on our understanding of medical products throughout their lifecycle, in the post market. We don't limit our knowledge to pre-market information, traditional de novo post-market studies, and passive reporting. Newer methodologies enable us to collect data

FDA Commissioner Scott Gottlieb, MD

Focus on integrating more EHR data in Sentinel

npj | Digital Medicine

www.nature.com/npjdigitalmed

PERSPECTIVE OPEN

Check for updates

Broadening the reach of the FDA Sentinel system: A roadmap for integrating electronic health record data in a causal analysis framework

Rishi J. Desai[1✉], Michael E. Matheny[2], Kevin Johnson[2], Keith Marsolo[3], Lesley H. Curtis[3], Jennifer C. Nelson[4], Patrick J. Heagerty[5], Judith Maro[6], Jeffery Brown[6], Sengwee Toh[6], Michael Nguyen[7], Robert Ball[7], Gerald Dal Pan[7], Shirley V. Wang[1], Joshua J. Gagne[1,8] and Sebastian Schneeweiss[1]

OXFORD

JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

American Journal of Epidemiology, 2024, 00, 1–7

https://doi.org/10.1093/aje/kwae226

Advance access publication date July 16, 2024

Invited Commentary

A future of data-rich pharmacoepidemiology studies: transitioning to large-scale linked electronic health record + claims data

Sebastian Schneeweiss*[,1], Rishi J. Desai[1], Robert Ball[2]

The FDA Sentinel Real World Evidence Data Enterprise (RWE-DE)

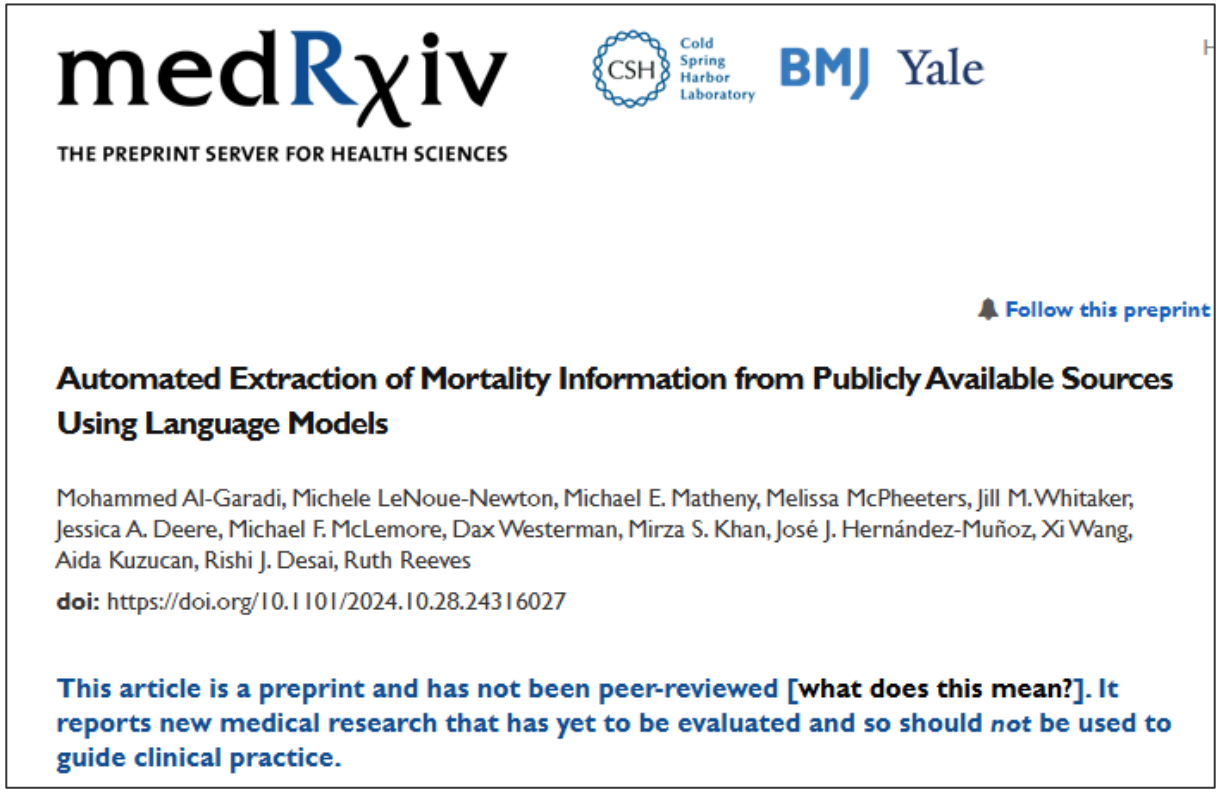
Rishi J. Desai[1] | Keith Marsolo[2] | Joshua Smith[3] | David Carrell[4] | Robert Penfold[4] | Haritha S. Pillai[1] | Joyce Lii[1] | Kerry Ngan[1] | Robert Winter[3] | Margaret Adgent[5] | Arvind Ramaprasan[4] | Meighan Rogers Driscoll[6] | Daniel Scarnecchia[6] | Daniel Kiernan[6] | Christine Draper[6] | Jennifer G. Lyons[6] | Anjum Khurshid[6] | Judith C. Maro[6] | Ruth Zimmerman[7] | Jeffrey Brown[8] | Patricia Bright[9] | José J. Hernández-Muñoz[9] | Michael E. Matheny[3,10] | Sebastian Schneeweiss[1]

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA | [2]Department of Population Health Sciences, Duke University, Durham, North Carolina, USA | [3]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA | [4]Kaiser Permanente Washington Health Research Institute, Seattle, Washington State, USA | [5]Department of Health Policy, Vanderbilt University Medical Center, Nashville, Tennessee, USA | [6]Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, Massachusetts, USA | [7]HealthVerity, Philadelphia, Pennsylvania, USA | [8]TriNetX, LLC, Cambridge, Massachusetts, USA | [9]Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, FDA, Silver Spring, Maryland, USA | [10]Geriatrics Research Education and Clinical Care Center, Tennessee Valley Healthcare System VA, Nashville, Tennessee, USA

Desai et al. npj Digital Medicine 2021
Schneeweiss et al. AJE 2024
Desai et al. PDS 2024

Challenges in Data Capture with EHRs: A Salient Example of Recording of Mortality

References available as preprints



medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory BMJ Yale

Follow this preprint

Automated Extraction of Mortality Information from Publicly Available Sources Using Language Models

Mohammed Al-Garadi, Michele LeNoue-Newton, Michael E. Matheny, Melissa McPheeters, Jill M. Whitaker, Jessica A. Deere, Michael F. McLemore, Dax Westerman, Mirza S. Khan, José J. Hernández-Muñoz, Xi Wang, Aida Kuzucan, Rishi J. Desai, Ruth Reeves

doi: https://doi.org/10.1101/2024.10.28.24316027

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Cold Spring Harbor Laboratory BMJ Yale

Follow this preprint

Augmenting Fact and Date of Death in Electronic Health Records using Internet Media Sources: A Validation Study from Two Large Healthcare Systems

Michele LeNoue-Newton, Mohammed Ali Al-Garadi, Kerry Ngan, Haritha S. Pillai, Ruth M Reeves, Daniel Park, Dax M Westerman, Jose J. Hernandez-Munoz, Xi Wang, Aida Kuzucan, Shirley Wang, Kueiyu Joshua Lin, Candace Fuller, Melissa McPheeters, Michael E. Matheny, Rishi J. Desai

doi: https://doi.org/10.1101/2025.01.24.25321042

This article is a preprint and has not been certified by peer review [what does this mean?]. It reports new medical research that has yet to be evaluated and so should *not* be used to guide clinical practice.

Motivation

- In safety and effectiveness evaluations of medications, mortality is often an important outcome of interest to researchers, clinicians, and patients.

- However, information regarding deaths that occur outside of a clinical setting are frequently missing in real-world data sources such as electronic health records (EHRs) or insurance claims.

- Using data from two Sentinel partner sites: Mass General Brigham (MGB) and Vanderbilt University Medical Center (VUMC), we aimed to quantify under-recording of mortality and attempted to augment this missing information using publicly available information regarding deaths in the US.

Quantification of unrecorded mortality information in EHRs

Reference standard data 2015-2020 (State vital records MA, VT, CT)

Deterministic linkage

Patient list from EHRs

Mass General Brigham

Total deaths in reference standard	Total deaths recorded in EHRs
65,139	**21,595 (33%)**

Reference standard data 2015-2020 (National Death Index)

Deterministic linkage

Patient list from EHRs

VANDERBILT UNIVERSITY MEDICAL CENTER

Total deaths in reference standard	Total deaths recorded in EHRs
50,952	**29,223 (57%)**

Augmentation approach: Sources

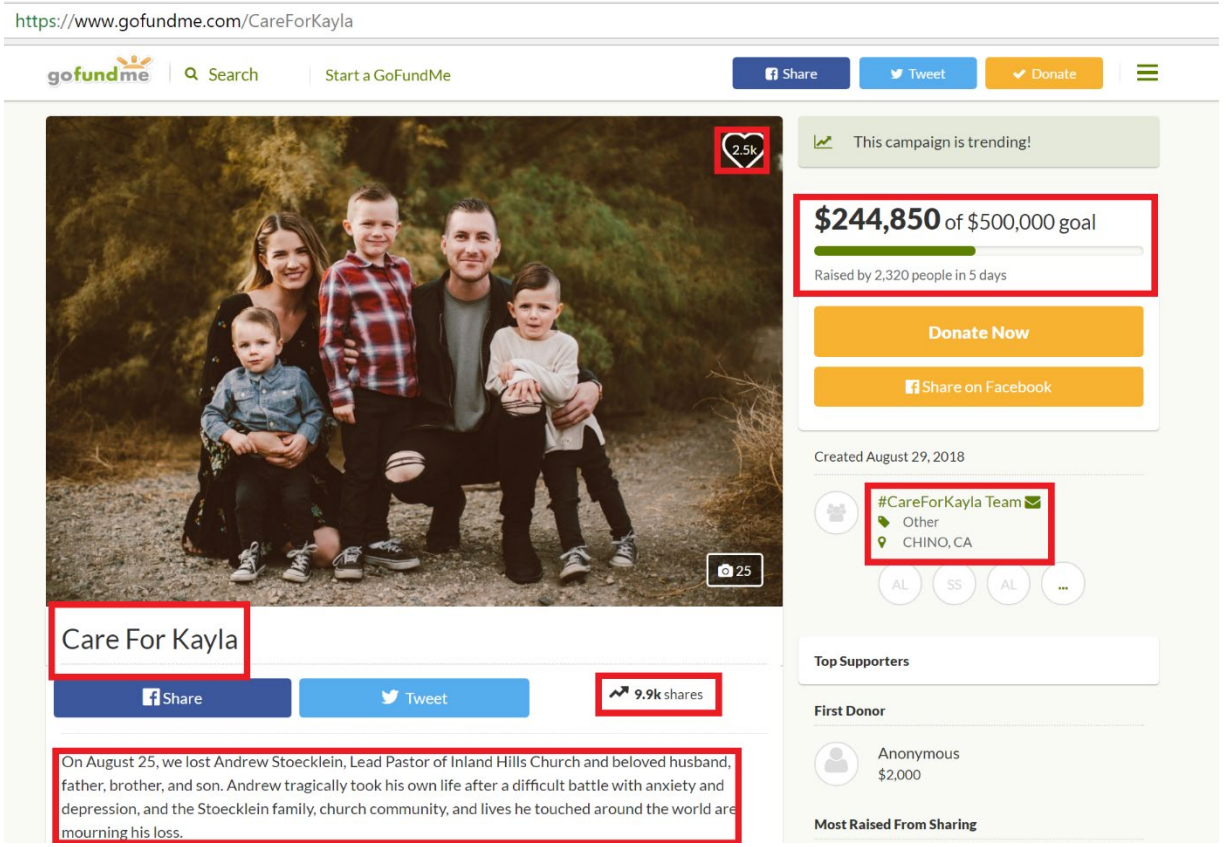
Publicly available death data

- Obituaries
- Memorial and crowdfunding websites (e.g., GoFundMe)
- Social media (Twitter)

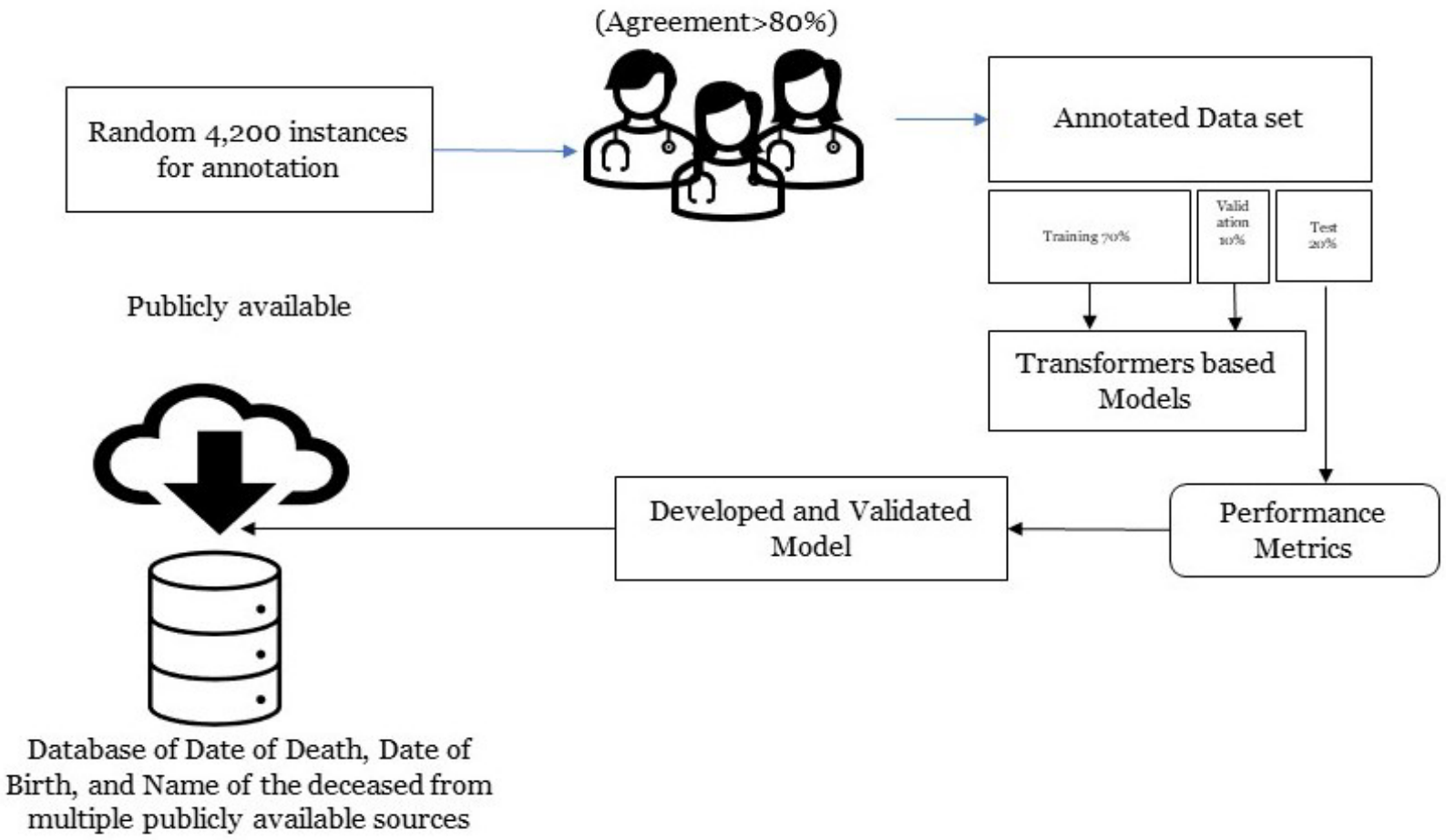
Augmentation approach: Challenge

Key challenge: A lot of information stored as free-text, challenging to extract and use at scale

Obituary Database Record		
Name	**Residence**	
Lillian M. Follett	Whitinsville, MA	
Age	**Date of Birth**	**Date of Death**
93	12/20/1927	12/01/2021
Obituary / Death Notice		
N. Smithfield, RI – Lillian M. Follett (Kollett), age 93, of North Smithfield, passed away on Wednesday, December 1st at Overlook Nursing Home in Pascoag, RI after a period of declining health. She was the beloved wife of the late Douglas L. Follett for over 50 years until his death in 2014. Lillian was born in Uxbridge on December 20, 1927 to the late Frederick W. and Mildred P. (Aldrich) Kollett. She enjoyed traveling with her husband and their children, as they were growing up. She had an affinity for gardening and canning vegetables (especially pickles). She made her own root beer. She was a very involved mother and was extremely devoted to her children. In later years, Lillian loved outdoor concerts and one of her favorite memories was on her 50th wedding anniversary with Douglas, when the entire family enjoyed a cruise around Newport. Lillian is survived by her loving children: David Follett and his wife Susan of Cumberland, RI, Bruce Follett and his wife Paula of Johnston, RI, Diane Lockwood and her husband John of Millville, Christine Mercadante and her husband Rob of Millville; siblings: Gordon Kollett and his wife Jacqueline of Uxbridge and Barbara Heldenbergh and her husband Richard of Millville; sister in law: Linda Kollett of Norton; grandchildren: Michael Follett and his wife Cynthia, Ryan Follett, Amye Follett, Madison Follett, John Lockwood and his wife Ashley, Erin Chinappi and her husband Andrew, Courtney Mercadante and her husband Nicholas Marcotte and Dayna Mercadante; and 7 great-grandchildren. Lillian was predeceased by her siblings: Evelyn Bouvier, Raymond Kollett, Jane Therien, Francis Kollett and Elaine Kollett. Calling hours for Lillian will be held on Thursday, December 9th from 4-7 P.M. at Buma Funeral Home, 101 N. Main Street, Uxbridge. Interment private. In lieu of flowers, donations in Lillian's name may be made to: The Nature Conservancy		



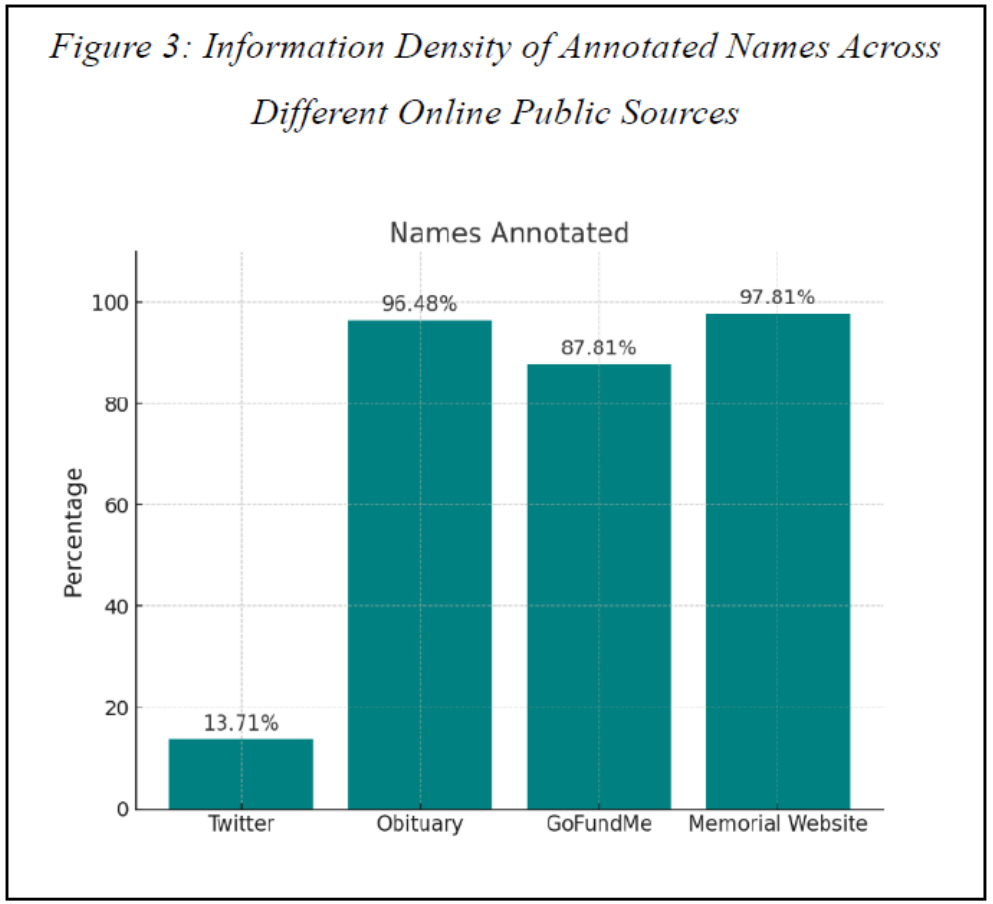
Augmentation approach: Training language models



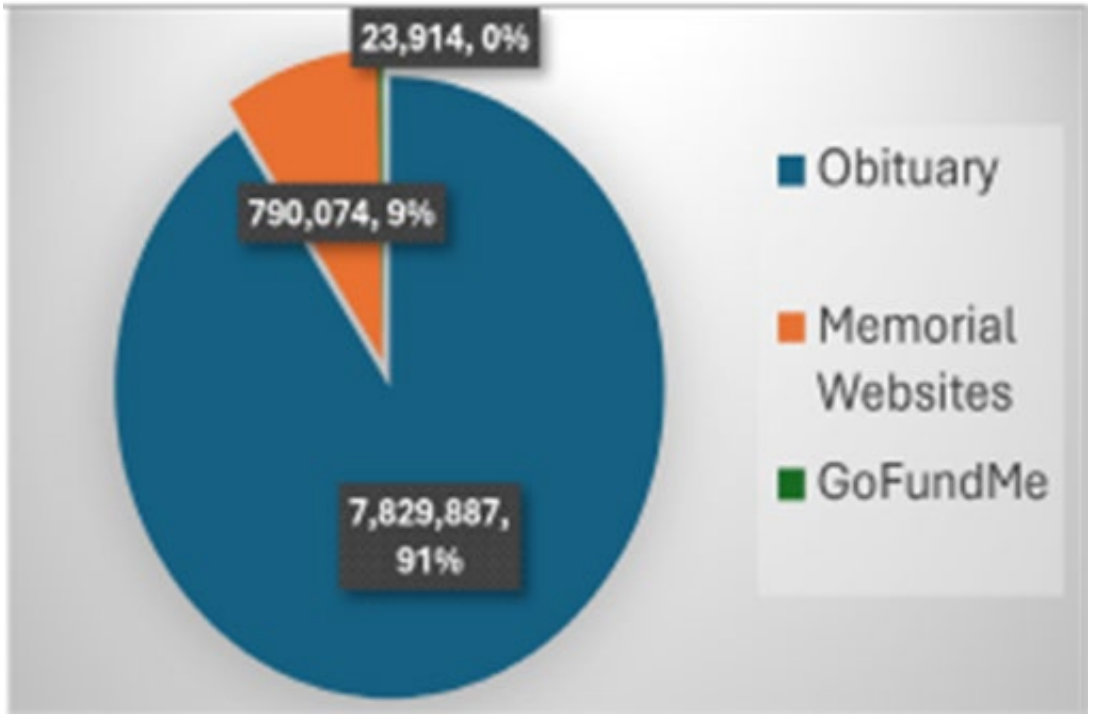
Model performance against manual labels

	RoBERTa		
	Precision (PPV)	*Recall (Sensitivity)*	*F1-score (95% CI)*
Decedent Name	0.86	0.84	0.85 (0.84-0.86)
Date of Death	0.87	0.91	0.89 (0.88-0.90)
Date of Birth	0.95	0.93	0.94 (0.92-0.94)
Micro Avg	0.88	0.88	0.88 (0.86-0.90)

Results: Extraction of deaths from publicly available information



Final dataset of 8.1 million extracted unique death records



Results: positive predicted values (PPV) of extracted deaths vs reference standard

When linking 8.1 million death records from public data to patient lists from the two healthcare systems using exact matching on first name, last name and date of birth

- PPV of **98.2%** at MGB and **98.9%** for VUMC with accurate date of death within +/7 days of reference standard

- PPV of **94.7%** at MGB and **96.3%** for VUMC with accurate date of death same as reference standard

Results: Increase in sensitivity of death capture with extracted deaths

Mass General Brigham

Total deaths in reference standard	Total deaths recorded in EHRs	Total deaths extracted from public information that are not recorded in EHRs
65,139	**21,595 (33%)**	**15,661 (24%)**

VANDERBILT UNIVERSITY MEDICAL CENTER

Total deaths in reference standard	Total deaths recorded in EHRs	Total deaths extracted from public information that are not recorded in EHRs
50,952	**29,223 (57%)**	**9,191 (18%)**

Summary

- Death capture in EHRs is sub-optimal, ranging from 33% to 57% completeness between our two systems
- Deaths extracted from public sources had high PPVs when validated against NDI and state vital files
- Augmenting death capture using publicly available information increased completeness meaningfully compared to relying on EHRs alone