

PERSPECTIVE OPEN



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

Rishi J. Desai¹✉, Michael E. Matheny², Kevin Johnson², Keith Marsolo³, Lesley H. Curtis³, Jennifer C. Nelson⁴, Patrick J. Heagerty⁵, Judith Maro⁶, Jeffery Brown⁶, Sengwee Toh⁶, Michael Nguyen⁷, Robert Ball⁷, Gerald Dal Pan⁷, Shirley V. Wang⁸, Joshua J. Gagne^{1,8} and Sebastian Schneeweiss¹

The Sentinel System is a major component of the United States Food and Drug Administration's (FDA) approach to active medical product safety surveillance. While Sentinel has historically relied on large quantities of health insurance claims data, leveraging longitudinal electronic health records (EHRs) that contain more detailed clinical information, as structured and unstructured features, may address some of the current gaps in capabilities. We identify key challenges when using EHR data to investigate medical product safety in a scalable and accelerated way, outline potential solutions, and describe the Sentinel Innovation Center's initiatives to put solutions into practice by expanding and strengthening the existing system with a query-ready, large-scale data infrastructure of linked EHR and claims data. We describe our initiatives in four strategic priority areas: (1) data infrastructure, (2) feature engineering, (3) causal inference, and (4) detection analytics, with the goal of incorporating emerging data science innovations to maximize the utility of EHR data for medical product safety surveillance.

npj Digital Medicine (2021)4:170; <https://doi.org/10.1038/s41746-021-00542-0>

BACKGROUND

The United States Food and Drug Administration (FDA)'s Sentinel System (referred to as "Sentinel" hereafter) uses distributed analytic tools and curated real-world longitudinal health insurance claims data for more than 100 million people from participating healthcare systems to generate insights regarding the safety of medical products^{1,2}. It is an important resource that informs drug labeling, drug safety communications, FDA Advisory Committee meetings, and other regulatory decisions^{3,4}. Health insurance claims data currently form the backbone of Sentinel, owing to their complete capture of outpatient pharmacy dispensing records, medical encounters, and hospitalizations during well-defined periods of health plan enrollment. Reliable implementation of drug safety analyses is achieved in a timely manner within Sentinel through the Active Risk Identification and Analysis (ARIA) system, which consists of modular programs that apply sophisticated epidemiologic study designs and analyses to distributed health insurance claims data organized in a common data model¹.

In a recent analysis of ARIA capabilities, some of the most frequently cited reasons for the inability of using the current claims-based Sentinel for safety investigations included the lack of clinical details to accurately identify health outcomes, missing or inaccurate measures of important confounding variables, or unavailability of computable phenotyping algorithms to identify study populations with acceptable accuracy⁵. Linking electronic health records (EHRs), which contain granular information related to clinical parameters, with insurance claims will likely address some of the current gaps in Sentinel's capabilities. In 2019, Congress required the FDA to create a Medical Data Enterprise to

enhance the Sentinel infrastructure by incorporating EHR data from at least 10 million lives⁶.

With all the opportunities for improvement, the inclusion of additional EHR data also introduces several methodological challenges that need to be addressed to facilitate robust analyses using a large-scale data infrastructure of combined EHR and claims data. The FDA has outlined an approach for achieving these goals in a 5-year Sentinel System strategic plan⁷. In this report, we describe a roadmap of the newly launched FDA Sentinel Innovation Center to strengthen the Sentinel System by outlining key challenges and proposed initiatives to address them.

IDENTIFYING AND INCORPORATING FIT-FOR-PURPOSE DATA SOURCES

To expand Sentinel's access to EHR data, a first-order goal is to develop the organizational framework and establish the governance, harmonization, and quality assurance processes for ensuring high-fidelity, fit-for-purpose data to support queries of regulatory importance. To support causal conclusions, establishing a clear temporal sequence of events based on data from sources with near-complete longitudinal capture is imperative. As most EHR sources in the US lack the ability to capture data when individuals receive care outside of the contributing healthcare systems, linkage of EHRs to insurance claims, which captures longitudinal data regardless of the care settings, is necessary to understand the completeness of longitudinal data. In establishing a query-ready distributed data network containing EHRs, the Sentinel Innovation Center will address key regulatory needs including determining where to

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

²Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. ³Department of Population Health Sciences, Duke University, Durham, NC, USA.

⁴Biostatistics Unit, Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. ⁵Department of Biostatistics, University of Washington, Seattle, WA, USA.

⁶Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA, USA. ⁷Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, FDA, Silver Spring, MD, USA. ⁸Present address: Johnson & Johnson, New Brunswick, NJ, USA. ✉email: rdesai@bwh.harvard.edu

source the EHR data with standing linkage to insurance claims and defining the minimum data elements necessary in order to address use cases that are currently difficult to address. Additionally, we will outline principles and strategies for determining how to organize both the structured, semi-structured and unstructured EHR data alongside insurance claims data in a common data model to facilitate standardized query implementation (Fig. 1).

A SYSTEM ROOTED IN A CAUSAL ANALYSIS FRAMEWORK

To comprehensively monitor the safety of marketed medical products, Sentinel investigations focus on both signal identification, which is intended to generate hypotheses regarding unsuspected adverse events, as well as signal refinement and evaluation, which is intended to test previously generated hypotheses and identify susceptible populations⁸. As regulatory decisions are predicated on inferring causal relations between a medical product exposure and adverse outcomes, a causal analysis framework fitting nonrandomized treatment allocation and secondary data is needed⁹. Even for signal identification based on data mining methods, attention to principles including clearly established temporality of confounder assessment preceding exposure followed by outcome surveillance, paired with analytic strategies that reduce confounding, is required to limit the number of spurious signals^{10–12}. For signal refinement and evaluation, in addition to these important principles, further considerations include explicitly specifying comparison groups, study populations, and specific outcome definitions ideally contemplated in a hypothetical ‘target’ trial which investigators then emulate in the secondary data of Sentinel^{13,14}. Analyses further should be accompanied by robustness evaluations to address the consistency of evidence with respect to alternative investigator decisions in study design, analysis, or variable measurement. To clarify challenges in developing and applying causal methods that leverage both claims and EHR data sources, the Sentinel Innovation Center will develop a causal analysis framework proposing a stepwise process that systematically considers key choices with respect to design and analysis that influence the validity of studies conducted with nonrandomized data. A standardized “industrial” process that will be outlined in this framework will serve as a valuable tool to inform the conduct and assessment of the quality of nonrandomized studies of drug-outcome evaluation.

EHR DATA-SPECIFIC INNOVATION NEEDS

While EHR data offer great promise for improving Sentinel’s capabilities, including improvements in computable phenotyping and confounding control, they also bring a range of measurement challenges that must be addressed. Expanded measurement procedures and analytic approaches are outlined in Fig. 2, identifying a range of Innovation Center projects. Briefly, we focus on distinct measurement challenges (1) to determine patient characteristics in the pre-exposure space (left side) where automated feature identification processes may feed directly into data-adaptive confounding adjustment procedures¹⁵ and (2) to identify and expeditiously validate health outcomes of interest in the postexposure space (right side) by augmenting claims-based algorithms with EHR-based data that may be structured, semi-structured, or unstructured. The Innovation Center has formed four main domains as strategic priority initiatives to deliver a query-ready, large-scale data infrastructure of combined EHR and claims data: data infrastructure, feature engineering, causal inference, detection analytics (Fig. 3).

DATA INFRASTRUCTURE DEVELOPMENT

EHR data are heterogeneous in their content, structure, completeness, and quality. To enable efficient and timely querying of EHRs to support surveillance activities, pre-processing to make them query-

ready is critical. To achieve this, the Sentinel Innovation Center is developing a principled approach to extend the Sentinel Common Data Model to include new data elements from structured and unstructured EHR data (Fig. 1). While common data models can help to impose a standard organization of the data across multiple sites, mapping of source values into the model can lead to omissions, data errors, and other data consistency issues¹⁶. Even though a data model standardizes the name of data elements, a lack of semantic interoperability across sites may remain. To address this challenge, we will investigate approaches to detect and mitigate data consistency issues and to develop and harmonize data across multiple EHR data sites. Further, we will focus on the development of a set of data quality metrics and approaches for the integration of structured and unstructured data elements from EHR into a common data model to facilitate reliable analyses of medical product outcomes using these data.

FEATURE ENGINEERING APPROACHES

Health insurance claims data, which have been the primary source of data within the Sentinel System, are highly structured with all information stored using standard terminologies such as International Classification of Diseases (ICD) codes for diagnoses and National Drug Codes (NDC) for filled prescriptions. Some of the EHR-based data such as administrative data, medication orders, and most laboratory testing results are recorded as structured or semi-structured fields, but a vast amount of potentially useful information is stored in visit notes (e.g., narrative descriptions of a patient’s signs and symptoms, family history, social history), radiology reports or images, and discharge summaries as unstructured data. Substantial engineering is needed to identify features from unstructured data that can be extracted and organized as structured data.

Natural language processing (NLP) and automated feature extraction are essential mechanisms to support scalable computable phenotyping from EHRs in signal detection and signal refinement activities. Development of automated feature extraction workflows that allow for time-contextualization is critical to enable the determination of temporality in confounder, exposure, and outcome assessment in Sentinel queries. However, there are numerous challenges in using these approaches at scale in national consortia. NLP tool performance between sites has been a challenge due to systematic data changes. Tool development further needs to be tailored to the eventual application of the derived features. For instance, identification of health outcomes of interest needs focused tools to ensure optimal performance characteristics including positive predicted value and specificity, while identification of potential confounders may utilize more flexible tools utilizing unsupervised approaches for high-dimensional feature vector generation. To address these challenges, the Sentinel Innovation Center has initiated a range of activities. The first set of projects aims to develop and validate algorithms for the identification of health outcomes of interest using focused NLP tools combined with machine learning approaches to identify complex concepts such as suicidal ideation. This work will expand on prior Sentinel activities that have demonstrated the proof of concept for use of algorithmic approaches in improving outcome identification^{17–19} and focus on developing a general framework for efficient and standardized processes. Another set of activities aim to build a semi-automated system of confounding adjustment that uses generalized NLP tools to enable large-scale extraction of features that appear sequentially and may serve as indicators of patients’ health trajectory. Finally, initiatives are also underway to improve the generalizability and transportability of NLP approaches across sites using statistical learning approaches²⁰.

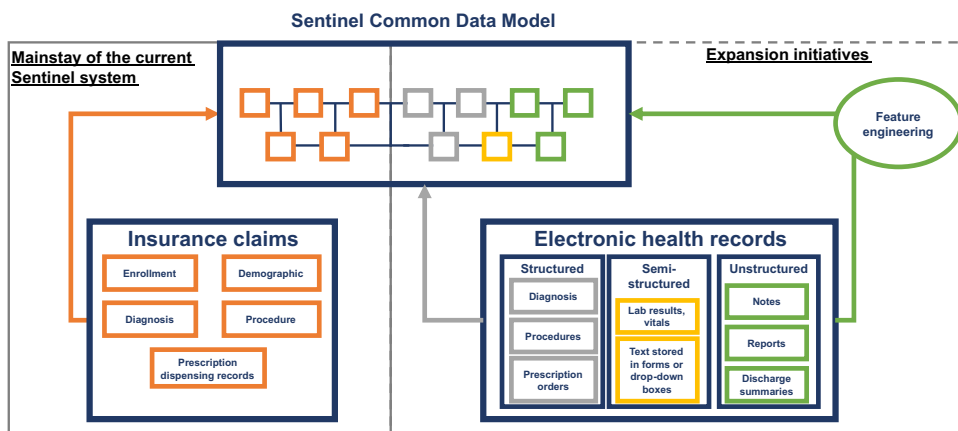


Fig. 1 Conceptual overview of the integration of claims data and electronic health records in Sentinel. Solid box on the left indicates data elements currently available in the Sentinel common data model, dotted box on the right indicates elements from electronic health records that will be considered for inclusion.

CAUSAL INFERENCE METHODOLOGY

Typical Sentinel investigations for drug safety conducted using insurance claims data sources have relied on restrictive study designs such as the active comparison of new user designs to achieve internal validity given the availability of large underlying populations; however, the use of EHR-claims linked sources may require alternate design choices, such as prevalent new user design²¹ to accommodate relatively smaller underlying populations. Tradeoffs when deviating from traditional design choices to accommodate available data assets need to be thoroughly investigated. Other unique challenges when using EHR for the outcome and confounder measurement include nonrandom missingness, or similarly the selective presence of data such as medical tests that may be ordered in light of the patients' prognosis, or incompleteness that occurs when outcomes recorded outside of the care systems are not available ("data leakage"). The Sentinel Innovation Center will focus on characterizing these challenges and developing strategies, methods, and tools to address them.

Residual confounding due to selection into treatment groups driven by outcome risk factors is another salient challenge in nonrandomized studies. The issue is accentuated when using data that lack clinical granularity such as insurance claims. Availability of EHR sources that contain richer clinical information on factors not readily available in claims hold promise to improve confounding adjustment in nonrandomized studies. The Sentinel Innovation Center will evaluate the feasibility of improved confounding adjustment from EHR-based variables through a combination of automated feature generation algorithms and advanced statistical and machine learning approaches such as Super Learner and Targeted Maximum Likelihood Estimation (TMLE)^{22,23}. Super Learner, which is an ensemble algorithm for predictive modeling, can data-adaptively model confounder summary scores, such as the propensity score, and the outcome to address model misspecification in the setting of complex and high-dimensional data settings of EHRs. TMLE can incorporate data-driven methods for high-dimensional confounder selection to empirically identify confounder information not specified by investigators. Developing scalable tools to implement these innovative methods in real-time will enhance the ability of Sentinel to address the common threat of confounding. Additionally, the Innovation Center will also investigate methods such as negative control outcomes or exposures and enhance existing tools for quantitative bias analyses²⁴ to better understand the

robustness of findings, including the impact of residual confounding.

Availability of additional clinical information in EHR further opens new opportunities to (1) identify outcomes that are generally not identifiable with claims data alone or (2) to efficiently validate claims-based algorithms in a subset of patients with EHR data available. The use of advanced methods based on machine learning and NLP to expedite outcome identification or validation have the potential to increase the efficiency of traditional drug safety evaluations. For instance, data-adaptive validation techniques where human experts review batches of patient charts for endpoint validation based on claims or EHR data could be iteratively used to train algorithms to inform the selection of cases that have a higher likelihood of being a true positive in the next batch²⁵. Incorporating NLP-assisted technology that can sift through patient charts and present the most relevant chart based on pre-specified key terms to human expert reviewers can add further efficiency to the process. The Sentinel Innovation Center will consider methodologic research to improve outcome identification and validation, which is a vital requirement for reliable evaluation of medication safety in Sentinel.

DETECTION ANALYTICS

Data mining approaches such as TreeScan have been developed in insurance claims data for signal detection in the Sentinel infrastructure based on the grouping of ICD diagnosis codes into hierarchical levels¹⁰. EHRs offer a potentially promising complementary source of information for medication safety signal detection but may require tailored approaches to account for and leverage differences in data content and structure compared to insurance claims. Specifically, unstructured clinical narratives recorded in EHR may provide more complete capture of subtle adverse events that may not trigger formal coding or medical interventions, aspects that are observable in claims data. The addition of detailed information presents an opportunity to expand signal detection efforts that are currently used in Sentinel. While NLP-based identification of adverse events from unstructured clinical notes is feasible with currently available methods, relational identification of newly occurring adverse events in a temporal sequence to specific medication exposures is complex and the subject of active research²⁶. The Sentinel Innovation Center plans to develop a methodological framework and conduct empirical evaluations to identify and test the most promising approaches for EHR-based signal detection.

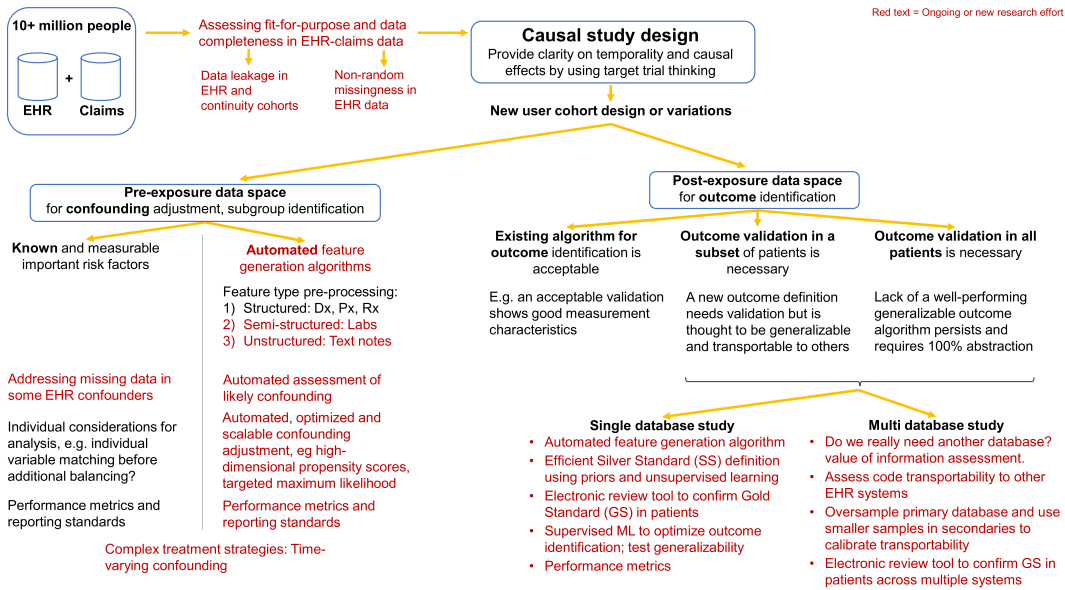


Fig. 2 Methodological research needs to support FDA safety decision-making with linked EHR-claims data. A summary of key infrastructure, design, and measurement challenges is described. Red text indicates ongoing or future research activities.

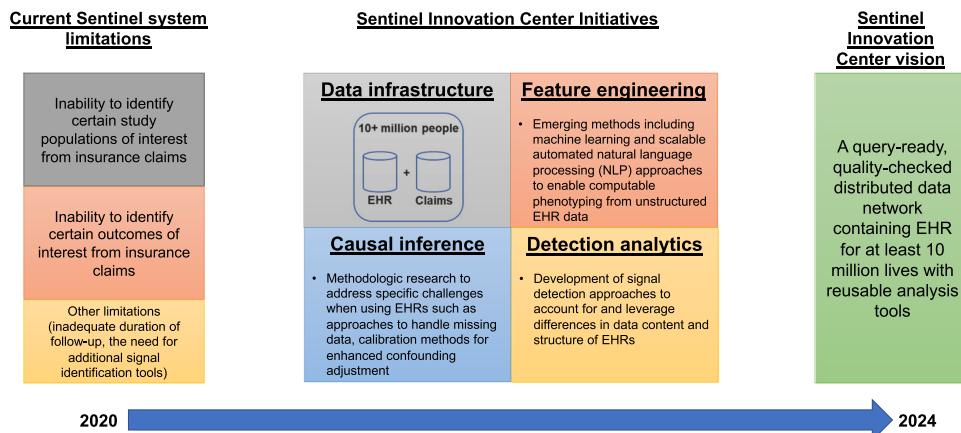


Fig. 3 The Sentinel Innovation Center initiatives and vision. Arrow at the bottom indicates timeline for the proposed activities.

DEMONSTRATION PROJECTS TO CALIBRATE THE VALIDITY OF RWE STUDIES

The Sentinel Innovation Center will launch several demonstration projects focusing on use cases typical for Sentinel; associations of medications for chronic conditions with infrequently observed outcomes where EHR data linkage has the potential to offer enhancements with respect to population identification, confounding adjustment, or outcome measurement. In these demonstration projects, we will emulate the corresponding target trial using principles and methodology developed across several Sentinel Innovation Center initiatives described above. When available, we will benchmark results against those observed in the corresponding pivotal clinical trial findings¹⁴.

TRANSPARENCY AND REPRODUCIBILITY OF PROCESS AND IMPLEMENTATION

Transparent communication of study specifications in a pre-registered protocol, akin to a prospective trial protocol, is a feature that helps increase confidence in results of hypothesis-confirming

studies from nonrandomized studies using secondary EHR-claims data. A focus of the Sentinel Innovation Center will be to facilitate EHR adaptation of existing tools such as the structured template and reporting tool for real-world evidence (STaRT-RWE) planning, implementation and communication template²⁷, to prepare and be able to register study protocols for future Sentinel queries. Additionally, adherence to principles for transparent reporting of analytic assumptions and diagnostics in line with good reporting practices²⁸ as well as transparent research practices including sharing of the analytic codes will be emphasized.

VISION FOR STRENGTHENING THE CURRENT SYSTEM THROUGH FOSTERING INNOVATION

Enhancements in the current Sentinel system proposed by the Sentinel Innovation Center that are summarized in this report will be achieved through strategic investment in focused research projects working in harmony. The new system containing granular clinical information through the integration of EHRs will broaden the scope of the current system by expanding on the types of

medication safety and effectiveness questions that are insufficiently addressed by insurance claims only. This may include questions where insurance claims data do not have sufficient information to identify the underlying population of interest, the outcome of interest, or a critical confounder to support a valid comparative analysis. Additionally, this system will also strengthen investigations that are possible to implement with insurance claims only through activities such as enabling extensive clinical characterization in a subset of patients selected for treatment with different medications and facilitating outcome algorithm development and validation for implementation in broader insurance claims-based network.

CONCLUSION

The FDA Sentinel Innovation Center has outlined a set of initiatives and a project portfolio²⁹ to deliver a query-ready distributed data network containing EHR linked to claims and reusable analysis tools to enhance the capabilities of the current system. These initiatives will incorporate data science innovations, such as scalable natural language processing and machine learning, to maximize the potential of EHR for medical product safety surveillance.

Received: 23 September 2021; Accepted: 28 November 2021;
Published online: 20 December 2021

REFERENCES

- Ball, R., Robb, M., Anderson, S. & Dal Pan, G. The FDA's sentinel initiative—a comprehensive approach to medical product surveillance. *Clin. Pharmacol. Therapeutics*. **99**, 265–268 (2016).
- Platt, R. et al. The FDA Sentinel Initiative—an evolving national resource. *N. Engl. J. Med.* **379**, 2091–2093 (2018).
- Sentinel Initiative [Internet]. FDA Advisory Committee Meetings; 2021. <https://www.sentinelinitiative.org/communications/fda-advisory-committee-meetings>. Accessed March 1, 2021.
- Sentinel Initiative [Internet]. FDA Safety Communications; 2021. <https://www.sentinelinitiative.org/communications/fda-safety-communications>. Accessed March 1, 2021.
- Brown, J. S., Maro, J. C., Nguyen, M. & Ball, R. Using and improving distributed data networks to generate actionable evidence: the case of real-world outcomes in the Food and Drug Administration's Sentinel system. *J. Am. Med. Assoc.* **27**, 793–797 (2020).
- Gottlieb, S. FDA budget matters: a cross-cutting data enterprise for real world evidence. *Food and Drug Administration*. <https://www.fda.gov/news-events/fda-voices/fda-budget-matters-cross-cutting-data-enterprise-real-world-evidence> (2018).
- Sentinel System Five-Year Strategy 2019–2023. Food and Drug Administration, 2019. Available at <https://www.fda.gov/media/120333/download>. Accessed February 2, 2021.
- Pray, L. & Robinson, S. *Challenges for the FDA: The Future of Drug Safety: Workshop Summary* (National Academies, 2007).
- Schneeweiss, S. & Patorno, E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocr. Rev.* **42**, 658–690 (2021).
- Wang, S. V. et al. Data mining for adverse drug events with a propensity score matched tree-based scan statistic. *Epidemiol.* **29**, 895 (2018).
- Wang, S. V. et al. A general propensity score for signal identification using tree-based scan statistics. *Am. J. Epidemiol.* **190**, 1424–1433 (2021).
- Nelson, J. C. et al. Integrating database knowledge and epidemiological design to improve the implementation of data mining methods that evaluate vaccine safety in large healthcare databases. *Stat. Anal. Data Min.* **7**, 337–351 (2014).
- Hernan, M. A. & Robins, J. M. Using big data to emulate a target trial when a randomized trial is not available. *Am. J. Epidemiol.* **183**, 758–764 (2016).
- Franklin, J. M. et al. Emulating randomized clinical trials with nonrandomized real-world evidence studies: first results from the RCT DUPLICATE Initiative. *Circulation*. **143**, 1002–1013 (2020).
- Schneeweiss, S. Automated data-adaptive analytics for electronic healthcare data to study causal treatment effects. *Clin. Epidemiol.* **10**, 771 (2018).
- Schneeweiss, S., Brown, J. S., Bate, A., Trifiro, G. & Bartels, D. B. Choosing among common data models for real-world data analyses fit for making decisions about the effectiveness of medical products. *Clin. Pharmacol. Therapeutics*. **107**, 827–833 (2020).
- Ball, R. et al. Evaluating automated approaches to anaphylaxis case classification using unstructured data from the FDA Sentinel System. *Pharmacoepidemiology Drug Saf.* **27**, 1077–1084 (2018).
- Bann, M. A. et al. Identification and validation of anaphylaxis using electronic health data in a population-based setting. *Epidemiology* **32**, 439–443 (2021).
- Gibson, T. B. et al. Electronic phenotyping of health outcomes of interest using a linked claims-electronic health record database: findings from a machine learning pilot project. *J. Am. Med. Inform. Assoc.* **28**, 1507–1517 (2021).
- Shi, X., Li, X. & Cai, T. Spherical regression under mismatch corruption with application to automated knowledge translation. *J. Am. Stat. Assoc.* 1–12 (2020).
- Suissa, S., Moodie, E. E. & Dell'Aniello, S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiol Drug Saf.* **26**, 459–468 (2017).
- Lendle, S. D., Fireman, B. & van der Laan, M. J. Targeted maximum likelihood estimation in safety analysis. *J. Clin. Epidemiol.* **66**, S91–S98 (2013).
- Wyss, R. et al. Using super learner prediction modeling to improve high-dimensional propensity score estimation. *Epidemiology* **29**, 96–106 (2018).
- Lash, T. L., Fox, M. P., Cooney, D., Lu, Y. & Forshee, R. A. Quantitative bias analysis in regulatory settings. *Am. J. Public Health* **106**, 1227–1230 (2016).
- Collin, L. J. et al. Adaptive validation design: a Bayesian approach to validation substudy design with prospective data collection. *Epidemiol.* **31**, 509 (2020).
- Liu, F., Jagannatha, A. & Yu, H. Towards drug safety surveillance and pharmacovigilance: current progress in detecting medication and adverse drug events from electronic health records. *Drug Saf.* **42**, 95–97 (2019).
- Wang, S. V. et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ* **372**, m4856 (2021).
- Langan, S. M. et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ* **363**, k3532 (2018).
- Innovation Center (IC) Master Plan. Available at <https://www.sentinelinitiative.org/news-events/publications-presentations/innovation-center-ic-master-plan>. Accessed February 11, 2021.

ACKNOWLEDGEMENTS

This project was supported by Master Agreement 75F40119D10037 from the US Food and Drug Administration (FDA). The FDA approved the study protocol including the statistical analysis plan, and reviewed and approved this manuscript. Coauthors from the FDA participated in the results interpretation and in the preparation and decision to submit the manuscript for publication. The FDA had no role in data collection, management, or analysis.

AUTHOR CONTRIBUTIONS

Concept: R.J.D., M.E.M., K.J., K.M., L.H.C., J.C.N., P.J.H., J.M., J.B., S.T., M.N., R.B., G.D.P., S.V.W., J.J.G., and S.S. Writing: R.J.D. and S.S.—Initial draft; M.E.M., K.J., K.M., L.H.C., J.C.N., P.J.H., J.M., J.B., D.T., M.N., R.B., G.D.P., S.V.W., and J.J.G.—critical revisions. Funding: R.J.D. and S.S.

COMPETING INTERESTS

Dr. Desai reports serving as Principal Investigator on research grants from Bayer and Novartis to the Brigham and Women's Hospital on studies outside of the submitted work. Dr. Schneeweiss is a consultant to Aetion, a software manufacturer of which he owns equity. His interests were declared, reviewed, and approved by the Brigham and Women's Hospital and Partners HealthCare System in accordance with their institutional compliance policies. Dr. Ball is an author on US Patent 9,075,796, "Text mining for large medical text datasets and corresponding medical text classification using informative feature selection." At present this patent is not licensed and does not generate royalties. Dr. Gagne is currently an employee of Johnson & Johnson. Dr. Nelson reports research funding from Moderna for service on their safety monitoring committee. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Rishi J. Desai.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021, corrected publication 2021