Request for Proposal

Sentinel Innovation Center:

Enhancing the validity of pharmacoepidemiology studies through the inclusion of semi-structured and unstructured electronic health record (EHR) data in confounding adjustment and outcome ascertainment

Department of Population Medicine
Harvard Medical School / Harvard Pilgrim Health Care Institute

Landmark Center
401 Park Drive
Suite 401
Boston, MA 02215

March 2022

Enhancing the validity of pharmacoepidemiology studies through the inclusion of semi-structured and unstructured EHR data in confounding adjustment and outcome ascertainment

Table of Contents

History of Modifications

Version	Date	Modification	Author
1.0	03/11/2022	Original version	Sentinel Operations Center
2.0	03/23/2022	Added disclaimer language and watermark	Sentinel Operations Center

1 Background

1.1 Sentinel Innovation Center

The Sentinel Initiative is a multi-faceted effort by the US Food and Drug Administration (FDA) to develop a national electronic system that complements previously existing methods of safety surveillance. The Sentinel System is comprised of three centers: the Sentinel Operations Center (SOC), the Sentinel Innovation Center (IC), and the Community Building and Outreach Center (CBOC). The centers collaborate to advance regulatory science using the Sentinel System.

The IC is led by the four lead hubs across the country (listed below), in collaboration with the Harvard Pilgrim Health Care Institute, in which the Sentinel Operations Center resides.

- Brigham and Women's Hospital and Harvard Medical School, Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics
- Duke University School of Medicine, Duke Clinical Research Institute
- Vanderbilt University Medical Center, Department of Biomedical Informatics
- Kaiser Permanente Washington Health Research Institute & University of Washington School of Public Health

The IC was created in part to respond to FDA's Medical Data Enterprise Initiative to build a new system containing electronic health records (EHRs) representing minimally 10 million lives. The IC is a test bed to identify, develop, and evaluate innovative methods to study medical product safety and effectiveness using real-world data (RWD) and enhance real-world evidence (RWE) capabilities through advanced analytics. The IC has developed a scientific blueprint for the next five years, called the IC Master Plan, which has four key strategic priority areas: data infrastructure; feature engineering; causal inference; detection analytics. Achieving the vision to enhance RWE capability requires:

- Expanding data element availability (i.e. feature extraction) in EHRs by finding efficient ways to extract, standardize and quality check unstructured and semi-structured data
- Developing new analysis tools that utilize these extracted features including state-of-the-art approaches to identify clinical phenotypes, and to adjust for confounding in medical product safety and effectiveness studies

Administrative claims and selected clinical data elements have been the primary source of data within the Sentinel System. These data are highly structured, whereas much of the potentially useful information contained within EHRs is in the form of unstructured data, including visit

notes (e.g., narrative descriptions of a patient's signs and symptoms, family history, social history), radiology reports or images, and discharge summaries. This unstructured information requires substantial engineering to identify features that can be extracted and organized in the form of structured or analyzable data.

Natural language processing (NLP), a computational linguistics technology, can be used to read, interpret and organize health data that may be contained within unstructured data fields within EHRs. NLP and automated feature extraction accelerate reading of complex information and can help researchers and clinicians to understand complex clinical narratives. They are essential mechanisms to support scalable computable phenotyping (identification of health outcomes of interest) and confounding control, both of which are important for supporting queries in Sentinel.

2 Project Description

This request for proposal (RFP) activity solicits proposals that intend to build a semi-automated system that uses advanced methods and tools, including NLP, to enable large-scale outcome identification, confounder feature extraction, and longitudinal contextualization of EHR data. The project will demonstrate how unstructured EHR data could help improve the validity of population-based pharmacoepidemiologic studies within linked health care claims and EHR data environments. Specifically, the project will involve an "end-to-end" implementation of a full pharmacoepidemiologic study using one or more product-outcome pairs to demonstrate how NLP and other advanced analytic methods or tools can be used to accurately identify study outcomes and covariates from unstructured EHR data.

The project goal and requirements are as follows:

2.1 Project Goal: Conduct a demonstration of a population-based pharmacoepidemiologic study, from study design through implementation, of one or more exposure-outcome pairs using a linked health care claims and electronic health record data source.

2.2 Project Requirements:

2.2.1 The proposal must meet the following criteria:

a. Use a linked health care claims and EHR data source. It is assumed that the exposure (e.g., an outpatient medication) can be reliably measured using structured claims data (e.g., outpatient pharmacy records).

b. Select a pharmacoepidemiologic use case or case study example that would benefit from the use of semi-structured or unstructured EHR data in i.) confounding adjustment and ii.) outcome identification. Use cases will not be compelling if confounding control and outcome ascertainment are already adequately managed via structured data alone.
c. Demonstrate how structured, semi-structured and unstructured EHR data can be leveraged in the pharmacoepidemiologic use case or case study example(s).
d. Demonstrate how NLP methods can be used to identify and contextualize pre-exposure confounding variables and incorporate unstructured EHR data into confounding adjustment. Proposals will be more competitive if bidders demonstrate the transportability of their methods to other similar data sources. Proposals will also be more competitive if bidders demonstrate the scalability of their method within a single data source (e.g., to very large sample sizes).
e. Demonstrate how NLP methods can be used to identify outcomes in unstructured EHR data and describe a robust approach to validate these methods. Rapid chart validation procedures would be one validation method, though other validation approaches are acceptable. Similarly, transportability and scalability of these methods to identify outcomes will make proposals more competitive.
f. Develop an evaluation framework that clearly describes when and how extraction of unstructured EHR data adds value to a linked claims-EHR study above and beyond what could be learned/studied in structured data alone. Demonstration of measurable and objective improvements over the status quo at reasonable cost and effort will make proposals more competitive.
g. Report on the complete pharmacoepidemiologic study demonstration, including the evaluative component, by making all source code freely available for replication in additional data sources.

2.2.2 With respect to semi-structured and unstructured EHR data, proposals will be more highly ranked if the methods have the following characteristics:

a. Semi-structured and unstructured text should be able to be leveraged with minimal investigator interaction, allowing the features extracted to be used as input to subsequent models (e.g., propensity scores).
b. Semi-structured and unstructured text that is utilized to develop features must demonstrate accurate temporal relationships (i.e., be accurately date-time

stamped) and be able to distinguish between "history of condition" and present onset of condition.

c. Semi-structured and unstructured text should be leveraged in a human-readable way. For example, proposals that use unstructured text information to predict health outcomes of interest or to allow clinicians to rapidly validate health outcome information must annotate medical charts in a clinician-friendly manner such that the information can be rapidly accessed and understood (i.e., long strings of alphanumeric identification codes that require lookup tables and additional work for clinician reviewers will not be considered adequate).

d. Semi-structured and unstructured text should easily be able to populate chart abstraction or curation forms to facilitate chart validation, preferably in a semi-automated way with minimal investigator interaction.

e. Semi-structured and unstructured text that is extracted should demonstrate the validity and accuracy of the extraction effort such that investigators have confidence in its representation of the raw data.

2.3 Use Case Examples

For illustrative purposes, the following examples show the need for having both structured and unstructured data to assess a product-outcome association. Bidders are permitted to use one of these examples or identify a separate exposure-outcome pair of their choosing for this investigation. The chosen case examples that use semi-structured and unstructured text to ascertain outcomes and/or covariates must meet criteria listed in Section 2.2.

1. The effect of sacubitril/valsartan (compared to valsartan alone) on renal function in patients with and without diabetes and heart failure with preserved ejection fraction.
    - The outcome is estimated glomerular filtration rate (eGFR), which can be obtained from semi-structured or unstructured EHRs.
    - Ejection fraction is needed to define the cohort and potentially used as a baseline confounder.
    - HbA1c may also be used to define the cohort and potentially used as a baseline confounder.
    - A secondary analysis of a randomized controlled trial on this topic is available for reference here: https://pubmed.ncbi.nlm.nih.gov/35119183/
2. The impact of therapeutics on liver function, particularly those for which adverse events of drug induced liver injury have been reported.

- Potential therapeutics to be investigated could include: daclizumab, a monoclonal antibody used to treat multiple sclerosis; tolvaptan, used to treat low blood sodium or slow kidney function decline.
- The outcome of interest is drug induced liver injury.
- There is currently insufficient supplemental clinical data available in claims data alone to assess the product-outcome pair relationship.

3 Proposal Guidelines

3.1 Proposal Content Requirements:

Each proposal should contain the following information:

1. A detailed description of your proposal, including how it meets all aims/criteria set forth in the project description section (Section 2.2), a project plan detailing how the activities will be operationalized. Proposals will be limited to eight pages, not including CVs/Biosketches, references or budget documents.
    a. Within the proposal, describe how access to and inclusion of one or more linked claims-EHR databases, containing unstructured and semi-structured data, will be secured for the duration of the project.
    b. Within the proposal, describe the dissemination plan for all developed tools and programming code. Anything developed in the course of this project is expected to be made freely publicly available.
2. NIH-formatted Biosketches or CVs demonstrating bidder's expertise in: pharmacoepidemiology and causal inference; use of EHR data in pharmacoepidemiologic studies, tools or software development; NLP or other artificial intelligence methods; observational database studies.
3. Inclusion of a project timeline that supports full project implementation including proposed deliverables and milestones, and access to any needed data or tools. Inclusion of anticipated resources you will assign to this project (number, role, title, prior experience). Inclusion of any detail on project management methodologies to be used.
4. Inclusion of a budget containing hourly rate by role. Please use the table in **Appendix A** to indicate budget and time estimates.

3.2 Evaluation Criteria

Proposals will be evaluated on the following criteria:

- Overall proposal suitability: Each proposal will be evaluated based on how clearly it addresses the scope and needs described herein. Quality is a higher priority than cost.

- Organizational experience: Bidders will be evaluated on their experience and expertise as it pertains to the scope of this project.
- Case studies and examples: Bidders will be evaluated on the quality of the provided example projects, applications, and case studies.
- Cost: Bidders will be evaluated on the projected cost of their proposed solution. Cost will factor heavily into selection of award recipient particularly if proposals are evenly ranked re: quality and suitability.
- Dissemination opportunities: Bidders will be evaluated on the ability of their methods to be scaled up within a single system for large-scale studies and to be applied in other similar settings (i.e., transportability).

3.3 Submission Instructions

- This RFP represents the requirements for an open and competitive bidding process. Proposals will be accepted until **Friday April 29, 2022** at 11:59pm ET. Please submit your proposals to InnovationCenterRFP@sentinelsystem.org. Any proposals received after this date and time will not be considered. All proposals must be signed by an official agent or representative of the company submitting the proposal.
- If the organization submitting a proposal must outsource or contract any work to meet the requirements contained herein, this must be clearly stated, and budgeted for, in the proposal. Any proposals which call for outsourcing or contracting work must include a name and description of the organizations being contracted. Additionally, proposals must be all-inclusive to financially cover any outsourced or contracted work.
- Contract terms and conditions will be negotiated upon selection of the winning bidder for this RFP. All contractual terms and conditions will be subject to review by Sentinel Operations Center and will include scope, budget, deliverable schedule, and other necessary items pertaining to the project. Please note that Sentinel Operations Center reserves the right to decline to select a winning bidder if submitted proposal(s) are determined to not meet project specifications as outlined in this document, or if otherwise warranted.

4 Request for Proposal Timeline

- All proposals in response to this RFP are due no later than **Friday April 29, 2022** at 11:59pm ET.
- Please submit your proposals to InnovationCenterRFP@sentinelsystem.org

- Evaluation of proposals will be conducted in May-early June 2022. If additional information or discussions are needed with any bidders during this window, the bidder(s) will be notified.
- A selection decision for the winning bidder will be made no later than June 30, 2022.
- Notifications to bidders who were not selected will be completed by June 30, 2022.
- Upon notification, contract negotiation with the winning bidder will begin immediately.
- As noted in the Proposal Guidelines, Sentinel Operations Center reserves the right to decline to select a winning bidder if submitted proposal(s) are determined to not meet project specifications as outlined in this document, or if otherwise warranted.

5 Award Information

- This RFP has been issued and will be managed by Harvard Pilgrim Health Care (HPHC). HPHC independently issued this RFP as a subcontracting opportunity and is not serving as an agent of the FDA or any other federal agency.
- HPHC funding for this activity and its associated subcontracts will be awarded under FDA Sentinel Contract # 75F40119D10037.
- This award will be made available via subcontract from Harvard Pilgrim Health Care, which houses the Sentinel Operations Center, and is a prime awardee of the FDA Sentinel System contract.
- This award will be capped at $1.6 million in total cost.
- Project timeline should be commensurate with the proposed activities but should not exceed a 24-month period.

6 Ownership

- All products developed, including source code, must be made freely available to the general public.

7 Contact Information

Prospective bidders may contact InnovationCenterRFP@sentinelsystem.org with any written questions about the proposal. Questions will be responded to within 5 business days and corresponding responses will be made publicly available via the Sentinel Initiative website.

8 Appendix A: Budget Template

IMPLEMENTATION				
LABOR:*				
Name	Role	Rate	Hours	Cost
		$0.00	0	$0.00
		$0.00	0	$0.00
		$0.00	0	$0.00
		$0.00	0	$0.00
		$0.00	0	$0.00
		$0.00	0	$0.00
TOTAL LABOR			0	$0.00
Other Costs:				
				$0.00
				$0.00
				$0.00
TOTAL Other Costs				$0.00
TOTAL DURATION (Weeks)				
TOTAL COSTS				$0.00

*Please utilize the 2022 NIH Guidance on salary limitations for grants and cooperative agreements to determine rate allowances. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-076.html