Carnegie Mellon University
Center for Innovation in Health

Data Access Opportunity for Research on Improving Patient Safety

The Center for Innovation in Health (CIH) is seeking CMU researchers interested in gaining early access to a large medical record dataset to pursue cutting-edge AI and ML approaches to improve patient safety.

In partnership with the University of Pittsburgh and the Jewish Healthcare Foundation, the CIH has launched a secure computing environment with access to a dataset (see description below) with the aim of fostering research on patient safety at CMU, and in particular on medication errors. This effort is part of CIH's Initiative for Patient Safety Research which has a goal to provide data relevant to developing at CMU autonomous and AI technologies for improving patient safety.

The CIH is seeking short proposals from researchers at CMU for early access to this data. Selected researchers will receive access to the data via our secure computing environment for use in patient-safety-related research. In turn, we ask that the selected researchers provide the CIH and the data team feedback on the data set, the computing environment, and other issues they encounter, as well as their understanding that this is an early "beta" effort that may experience technical or administrative hiccups.

This data access effort is supported by separate grants from the Jewish Health Foundation (JHF) to CMU and the University of Pittsburgh. CMU's Initiative for Patient Safety Research is part of JHF's RAPS initiative.

To apply for data access, please submit an application following the instructions below.

Description of the Data

The University of Pittsburgh, through CIH, is granting access to their Medication Error Avoidance at Regional Scale (MEARS) dataset. The MEARS Database contains data between 2015 and 2018, organized using the Observational Health Data Science and Informatics Common Data Model (OHDSI CDM), for approximately 90,000 patients with PACE insurance, and approximately 560,000 patients without PACE claims. PACE is a prescription assistance program for adults age 65 and older. The spontaneous reporting data combines legacy and current FDA Adverse Event Reporting System (FAERS) data into a harmonized database. Pharmacovigilance signal calculations are available for all drugs X adverse reactions in FAERS from 2004 – 2021. Additional data included are 1) NLP-extracted data on adverse events reported in all drug product labeling, 2) drug indications from the RepoDB database, 3) and a comprehensive terminology system that has more than 3 dozen clinical terminologies including SNOMED, RxNorm, LOINC, CPT, and ICD.

This dataset enables rapid exploration, insight, and innovation, catalyzing research to identify positive or negative deviations related to patient safety and medication errors. We are excited to offer access to this database to researchers with expertise in statistics, operations, AI/ML, computational biology, computer science, human and organizational behavior, system and UI design, psychology, information theory and other fields to conduct analyses and develop methods to detect, diagnose, and communicate events related to medication errors and patient safety. Ultimately we seek methodology to reliably allow providers to detect anomalous situations, recommend actions, and predict changes in patients' conditions.

Relevant Research Areas

Proposals should address a challenge in patient safety, with preference given to those that tackle aspects of medication errors. Prior research on patient safety is not required, but proposals must present a credible route to advancing patient safety.

Secure Computing Environment

The data is hosted in an Azure dedicated server environment run by The Department of Biomedical Informatics (DBMI) at the University of Pittsburgh. Access to the data requires logging into this system using a specific secure VPN. Computation must be conducted on this system and none of the primary data may leave the system. For peer reviewed academic publications, tables and figures with aggregate counts based on a dataset of less than 20 and the underlying raw Data may not be published without University of Pittsburgh's written consent. Anyone who is granted access must complete training specific to this data set on the requirements for data security and privacy and sign a data use agreement that includes specific restrictions.

The environment provides access to a Jupyter Hub allowing users to create multiple instantiations of Jupyter notebooks with access to Python and R.

How to Apply

Please submit a proposal that includes:

- A title for the proposed project.
- A description of how the data will be used to develop technologies for improving patient safety.
- The primary deliverables and outcomes that are anticipated, including (a) research publications and (b) problem solutions.
- The anticipated timeline to reach milestones in the project.

The above items are limited to 1 page in total.

On the 2nd page of the proposal, please provide:

- A list of the faculty supervisors of the project. Identify the lead supervisor. The lead faculty member must have their primary appointment at CMU. CMU-Pitt collaborations are encouraged but absolutely not required. Access for individuals outside of CMU or

Univ of Pittsburgh is not possible at this time. The lead faculty member will be responsible for reporting and compliance requirements.
- A list of at most 3 researchers (who may be Ph.D. students and that may overlap with the faculty supervisors), who will have direct, login access to the data and the secure computing environment.
- The statement, at the end, that says (verbatim): “We understand and agree that, if selected, the data will be provided on a best effort basis and that this is a beta testing program designed to work out a process for providing access to this data. We further agree that: (1) the data will be used only for the purpose stated in this application; (2) we will acknowledge the CIH and its grants (using language provided by the CIH) in any publication that uses the data; (3) we will provide feedback and suggestions regarding any issues that we encounter with the data and its access to the CIH team; (4) we will participate in relevant and appropriate CIH efforts to report results to the JHF and to seek other funding to support this effort; and (5) other restrictions and requirements apply.”

A bibliography of relevant past work or citations in the proposal may also be included on page 3 and beyond (and not counted in the page limits).

The proposal must be a single file in PDF format and sent to centerinnovhlth@andrew.cmu.edu with the subject “IPSR Data Access Proposal”.

Proposals must be received by **March 20, 2024** to be considered for this first round of access. Notification of decisions is expected by April 5, 2024. In this first phase, we expect to select between 2-4 proposals for access. Proposals not selected in this round will be retained and considered for future rounds if they occur.

Review Criteria

Proposals will be reviewed by CIH-affiliated faculty for their relevance to patient safety, the track record of the team, and the likelihood of the project resulting in impactful research.

What Selected Researchers Will Get

- Access to the secure computing environment and the data for use in the research described in their proposal.

Requirements of Selected Researchers

- Complete the relevant security training hosted by the CIH and the team at the University of Pittsburgh.
- Agree to the CIH-provided data use and security requirements outlined in the Data Use Agreement with Pitt..
- Agree to provide feedback and reports on results to the CIH.
- Agree to acknowledge CIH and Pitt/MEARS in publications that use the data.

- Agree to provide short, non-confidential research summaries that will be communicated to our funders or used to describe the outcomes of the effort.
- Agree to participate in at least one meeting with researchers selected by CIH from CMU and Univ Pittsburgh to explore potential collaborations.
- Agree to consider participation in, as appropriate, future collaborative research or funding efforts related to the data or patient safety.

Important Dates

- Proposals Due: Midnight (EDT) on Wednesday, March 20, 2024.
- Decisions Expected By: Friday, April 5, 2024.