General Information

1. General Information

1. Project Title

Application of a Carolina Open Health Data (CaOHD) Open Infrastructure for Exploring Real-World Outcomes for Drug Exposures

2. **Brief Summary**. Provide a **brief non-technical description** of the study, which will be used in IRB documentation as a description of the study. Typical summaries are 50-100 words. Please reply to each item below, retaining the subheading labels already in place, so that reviewers can readily identify the content. PLEASE NOTE: THIS SECTION MAY BE EDITED BY THE IRB FOR CLARITY OR LENGTH.

Purpose:

We propose Carolina Open Health Data (CaOH) as an Observational Medical Outcomes Partnership (OMOP)-based open infrastructure to support large-scale, real-world evidence of drug effectiveness and adverse events.

Electronic health records contain data on clinical phenotypes and drug exposures that represent "real-world" practice rather than the idealized setting simulated by clinical trials. Indeed, clinical trials are studied in relatively homogeneous and "healthy" patient populations. However, when drugs are used in real-world clinical practice, it is common for patients to have comorbidities and/or concomitant medication use and/or environmental exposures and/or be of an ethnicity, sex, or age that was not studied pre-clinically or selected for in clinical trials. These biases in patient recruitment for clinical trials often reduce the effectiveness of drugs in real-world clinical practice and may result in unexpected adverse events and toxicity, thus resulting in removal of drugs from the market after US FDA approval and incurring major financial losses in investments for clinical trials (Wysowski and Swartz, 2005; Nagaich and Sadhna, 2015; Onakpoya et al. 2019).

We hypothesize that through the application of statistical and machine learning algorithms, we will be able to evaluate drug effectiveness and safety in real-world populations, meaning heterogeneous patient populations with uncontrolled comorbidities; concurrent medications; and diverse environmental exposures, family history, racial/ethnic backgrounds, sexes/genders, and ages. Moreover, by adopting the Columbia Open Health Data (COHD) open-source model, we will minimize the effort required to establish a similar OMOP-based open infrastructure at UNC, and we will be able to leverage the extensive experience and resources available through the Observational Health Data Sciences and Informatics (OHDSI) community (Observational Health Data Sciences and Informatics,2021).

This effort will build on our several existing efforts. First, we will leverage our collaborative work with colleagues at Columbia University Irving Medical Center (CUIMC) and their COHD open infrastructure (Ta et al. 2018). COHD exposes data on counts and patient prevalence of conditions, procedures, drug exposures, and patient demographics, and the co-occurrence frequencies between them. Count and frequency data are derived from CUIMC's OHDSI database. COHD's 5-year dataset contains 29,964 single concepts (10,159 conditions, 10,264 drugs, and 8,270 procedures) and 15,927,195 concept pairs from 1,790,431 patients. Exclusion of rare concepts (count?=?10) and Poisson randomization enable open data sharing by eliminating risks to patient privacy. The software code supporting COHD is available in a public GitHub repository (https://github.com/WengLab-InformaticsResearch/cohd_api). *Importantly, the COHD team* will not *have access to identifiable UNC Health patient data; rather, they will advise us on the COHD software code and the OMOP common data model.*

The proposed effort will also build on the Integrated Clinical and Environmental Exposures Service (ICEES) that we developed to openly expose and explore semi-aggregated clinical data that have been integrated with a variety of public data sources on environmental exposures (Fecho et al. 2019). Like COHD, ICEES is also open source and supported by a custom data integration pipeline termed FHIR Patient data Integration Tool (FHIR PIT) (Xu et al. 2020). ICEES was developed initially to support open research on clinical and environmental determinants of asthma and related common pulmonary disorders; however, the framework itself is flexible, extensible, and disease-agnostic. Indeed, we have extended ICEES to support research on drug-induced liver injury, primary ciliary dyskinesia, and coronavirus infection. The software code supporting ICEES is available in a public GitHub repository (https://github.com/ExposuresProvider/icees-api).

The proposed CaOHD is a natural extension of both the COHD and ICEES efforts and will provide a complementary approach, focused on post-marketing drug effectiveness and safety research and surveillance. Importantly, the ICEES and COHD teams have worked closely together since the launch of the NCATS Biomedical Data Translator ('Translator') program and have multiple joint publications (e.g., Ahalt et al. 2019; Fecho et al. 2019; Fecho et al. 2023). Our collaborative work will help to ensure the success of the proposed work.

Our work to date on ICEES has been based on the i2b2, PCORnet, and FHIR common data models. However, with NC TraCS' recent move to OMOP as a preferred common data model, the time is right for ICEES to follow suit. Moreover, NC TraCS is currently preparing for a pilot of the "OMOP Repository of Data for Research, Deidentified" (ORDR-D), which will provide UNC researchers with access to a repository of OMOP-derived patient data, stripped of all PHI via HIPAA Safe Harbor. The proposed CaOHD will directly complement that work, but it will focus on co-occurrences of concept pairs (drugs, conditions, procedures) and drug–related adverse outcomes, adopting the COHD framework.

Participants:

We will *not* be enrolling any subjects as part of this work, and thus no individual patient will be contacted by members of the study team, nor will informed consent be obtained. Rather, we will be working with existing biomedical data only and requesting data only on phenotypes that are shared by ten or more patients.

Procedures (methods):

We propose a CaOHD open infrastructure to support large-scale, real-world research on drug-related health outcomes and adverse events.

We will not apply any inclusion or exclusion criteria; rather, we will request a five-year dataset (2018–2022) on all patients represented in the NC TraCS OMOP data repository who have had at least two visits to the UNC Health system within the five-year study period.

Initially, we will focus on the OMOP concepts of conditions, drugs, and procedures for both inpatient and outpatient visits. CaOHD will allow users to openly explore UNC Health data and ask questions such as: *Which procedures and patient demographics are most likely to be associated with postoperative opioid-related respiratory depression? Can we predict adverse drug reactions based on comorbidities? Which racial and socioeconomically disadvantaged groups have the highest rates of drug-related adverse outcomes, and for what drugs?*

After we have implemented and tested the CaOHD open infrastructure, and demonstrated its value in the context of driving use-case questions, we will adapt our FHIR PIT data integration pipeline to integrate the OMOP-derived patient data with environmental exposures data, as is done with ICEES (**Figure 1**). That work will enable us to extend CaOHD to ask environmental health–related questions such as: *Does exposure to high levels of airborne pollutants increase the risk of 1-year mortality after a cardiac event? Is attendance at a public school located within <500 meters from a major roadway or highway associated with nonresponse to treatment among school-age children with asthma? Does residential location within < 1 mile of a landfill increase the risk of neurologic disease among patients >65 years?*

Study Design: Retrospective data collection for a five-year dataset, January 1, 2018 through December 31, 2022.

Specific Data Elements Requested: DOB, US Census Bureau census block group for primary home residence, visit dates, and OMOP conditions, drugs, and procedures (i.e., condition_occurrence table, drug_exposure table, procedure_occurrence table). *Of importance, geolocations and dates will be used solely for linkage to environmental exposure estimates; all data elements will be deidentified per HIPAA Safe Harbor method after the data integration step and before openly exposing the data via CaOHD.* **Figure 1** provides a high-level overview of our FHIR PIT data flow and integration pipeline.

Specific Inclusion Criteria: None.

Specific Exclusion Criteria: None

3. Is this new study similar or related to an application already approved by a UNC-Chapel Hill IRB? Knowing this will help the IRB in reviewing your new study.

Yes

If yes, provide IRB study number and explain why this is relevant to the current study and why it would be useful for the IRB to know.

Yes, the proposed project is related to the following IRB-approved applications: study number: 16-2978; study number 20-0924; study number: 21-0099; study number: 21-0768; and study number: 21-2064. These relevant studies support the Integrated Clinical and Environmental Exposures Service (ICEES), which will be leveraged as part of the proposed work.

2. Project Personnel

1. Will this project be led by a STUDENT (undergraduate, graduate) or TRAINEE (resident, fellow, postdoc), working in fulfillment of requirements for a University course, program or fellowship?

No

2. List all project personnel beginning with principal investigator, followed by faculty advisor, co-investigators, study coordinators, and anyone else who has contact with subjects or identifiable data from subjects.

- List ONLY those personnel for whom this IRB will be responsible; do NOT include collaborators who will remain under the oversight of another IRB **for this study**.
- If this is Community Based Participatory Research (CBPR) or you are otherwise working with community partners (who are not functioning as researchers), you may not be required to list them here as project personnel; consult with your IRB.
- If your extended research team includes multiple individuals with limited roles, you may not be required to list them here as project personnel; consult with your IRB.

The table below will access campus directory information; if you do not find your name, your directory listing may need to be updated.

Liaison	Last Name	First Name	Department Name	Role	
University of North Carolina at Chapel Hill (UNC-CH)					
	Krishnamurthy	Ashok	Renaissance Computing Institute	Principal Investigator	view
	Ahalt	Stanley	Renaissance Computing Institute	Co-investigator	view
⭐	Fecho	Karamarie	Renaissance Computing Institute	Co-investigator	view
	Garcia	Juan	Computer Science	Other	view
	Reilly	Jason	Renaissance Computing Institute	Other	view
	Yi	Hong	Renaissance Computing Institute	Other	view

If your research includes personnel from a UNC Health Network Entity (NE), the UNC Health Office of Research Support and Compliance (ORSC) will review your IRB application and/or submitted UNC Health Collaboration Survey. You may be contacted by ORSC for additional information. IMPORTANT: In addition to obtaining IRB approval, you must also receive ORSC clearance for project personnel employed by the NE site(s). Project personnel MAY NOT proceed with research activities until you have obtained both approval from the IRB and clearance from the NE. Upon completed ORSC review, an ORSC NE Clearance Form will be provided and uploaded to the IRB application Study Documents section.

NOTE: The IRB database will link automatically to UNC Human Research Ethics Training database and the UNC Conflict of Interest (COI) database. Once the study is certified by the PI, all personnel listed (for whom we have email addresses) will receive separate instructions about COI disclosures. The IRB will communicate with the personnel listed above or the PI if further documentation is required.

3. If this research is based in a center, institute, or department (Administering Department) other than the one listed above for the PI, select here. Be aware that if you do not enter anything here, the PI's home department will be AUTOMATICALLY inserted when you save this page.

Department	Renaissance Computing Institute

3. Funding Sources

1. Is this project funded (or proposed to be funded) by a contract or grant from an organization EXTERNAL to UNC-Chapel Hill?

Yes

Is UNC-CH the **direct** recipient of any Federal funding for this study? You should answer 'yes' *only* if you are the grantee. You should answer 'no' if you are the recipient of a sub-award or contractor under the grant.

Yes

Funding Source(s) and/or Sponsor(s): Please list all entities that are providing monetary support or supplies (e.g., study drug, gifts, devices at no cost, or others that provide in-kind services).

Sponsor Name	UNC Ramses Number	Sponsor Type	Prime Sponsor Name	Prime Sponsor Type	Sponsor/Grant Number	Det
National Institutes of Health (NIH)	20-1992	Federal			1-OT2-TR003430-01|1-OT2-TR003430-01|3-OT2-TR003430-01S1|3-OT2-TR003430-01S2|3-OT2-TR003430-01S3|3-OT2-TR003430-01S4	view

2. Is this study funded by UNC-CH (e.g., department funds, internal pilot grants, trust accounts)?

No

3. Is this research classified (e.g. requires governmental security clearance)?

No

4. Is there a master protocol, grant application, or other proposal supporting this submission (check all that apply)?

- ✔ Grant Application
- ✘ Industry/Federal Sponsor Master Protocol
- ✘ Student Dissertation or Thesis Proposal
- ✘ Investigator Initiated Master Protocol
- ✘ Other Study Protocol

5. Is this a Clinical Study?
Check YES if this study involves research using human volunteers that is intended to add to medical knowledge. There are two main types of clinical studies: clinical trials and observational studies. Do NOT check yes merely because you are conducting research in a clinical setting or using clinical data.
Click here for additional definition of "Clinical Study"

No

6. Will this clinical study be listed in ClinicalTrials.gov, either by you or the sponsor?
Click here for ClinicalTrials.gov Guidance Information

No

4. Screening Questions

The following questions will help you determine if your project will require IRB review and approval.

The first question is whether this is RESEARCH (click for details)

1. Does your project involve a systematic investigation, including research development, testing and evaluation, which is designed to develop or contribute to generalizable knowledge? PLEASE NOTE: You should only answer yes if your activity meets all the above.

Yes

The next questions will determine if there are HUMAN SUBJECTS (click for details)

2. Will you be obtaining information or biospecimens through intervention or interaction with the individual, and use, study, or analysis of the information or biospecimens? This would include any communication or interpersonal contact between investigator and subject such as using in-person or online questionnaires/surveys, interviews, focus groups, observations, treatment interventions, etc. PLEASE NOTE: Merely obtaining information FROM an individual does not mean you should answer 'Yes,' unless the information is also ABOUT them.

No

3. Will you be obtaining, using, studying, analyzing, or generating identifiable private information or identifiable biospecimens collected through means other than direct interaction? This would include data, records or biological specimens that are currently existing or will be collected in the future for purposes other than this proposed research (e.g., medical records, ongoing collection of specimens for a tissue repository).
OR
Will you be using human specimens that are not individually identifiable for FDA-regulated in vitro diagnostic (IVD) device investigations?

Yes

The following questions will help build the remainder of your application.

4. Will subjects be studied in the Clinical and Translational Research Center (CTRC, previously known as the GCRC) or is the CTRC involved in any other way with the study? (If yes, this application will be reviewed by the CTRC and additional data will be collected.)

No

5. Does this study directly recruit participants through the UNC Health Care clinical settings for cancer patients **or** does this study have a focus on cancer or a focus on a risk factor for cancer (e.g. increased physical activity to reduce colon cancer incidence) **or** does this study receive funding from a cancer agency, foundation, or other cancer related group? (If yes, this application may require additional review by the Oncology Protocol Review Committee.)

No

6. Is the UNC Chapel Hill IRB taking or being asked to take responsibility for the oversight of research by individuals, groups or organizations outside of UNC Chapel Hill? Or you are asking the UNC Chapel Hill IRB to cede review to an External IRB. If so, a reliance agreement will need to be executed prior to conducting any research activities.

No

Location

1. Are UNC-affiliated researchers involved in research conducted at any locations outside of the United States?

No

Scientific Review

Scientific Review

All biomedical research conducted at the University of North Carolina at Chapel Hill involving procedures that pose greater than minimal risk must undergo scientific review. Scientific review is a process that evaluates the scientific merit of a protocol ensuring that only scientifically sound protocols are submitted for IRB review. Scientific review ensures that the research uses procedures consistent with sound research design and the research design is sound enough to reasonably expect the research to answer its proposed question.

For example, research that involves experimental drugs or devices or invasive procedures requires scientific review. Additional examples can be found here.

At UNC, the Protocol Review Committee provides scientific review for oncology studies. For all other studies, scientific review can be conducted:

1. *Externally, by an independent organization that has no conflict of interest with the submitted research activity or;*
2. *Internally by the UNC Scientific Review Committee (SRC).*

Examples:

1. *An investigator-initiated study (with greater than minimal risk) regardless of the funding source must undergo scientific review by the UNC SRC unless it has been reviewed and approved by a protocol review committee empaneled for this purpose by Federal-funding agency (e.g., NIH, CDC, DOD) or a national foundation.*
2. *A multicenter industry/foundation sponsored protocol does not require scientific review by the UNC SRC.*

Note: Study section or FDA IND/IDE review is not adequate review to supplant review by the UNC SRC.

1. Does your research study require review by the Scientific Review Committee?
If you are unsure if your project requires SRC review, please contact the OCT at 919-843-2698.

Respond "no" if your research methods are limited to:

- the collection and/or use of existing data, documents or specimens and/ or,
- administering surveys; conducting interviews or focus groups, and/or
- prospective collection of biological specimens by non-invasive means (hair and nail clippings, sweat, dental plaque and calculus, saliva, sputum).
- Collection of blood from healthy adults by finger stick.

No

Select one or more of the following:

✔ This protocol does not involve interaction or intervention that poses greater than minimal risk to subjects. Minimal Risk-the probability and magnitude of harm or discomfort anticipated in the research are not greater than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.

✘ This protocol was reviewed by a protocol review committee empaneled for this purpose by Federal-funding agency (e.g., NIH, CDC, DOD) or a national foundation. Please provide evidence of this in the form of a letter from the project office if it is not evident from the protocol. Study section or FDA IND/IDE review is not adequate.

✘ The protocol was developed by an industry sponsor (e.g., pharmaceutical, device or diagnostics trials) and involves multiple research sites.

This research project does not require review by the UNC SRC. Please complete the remainder of the application.

Part A. Questions Common to All Studies

A.1. Background and Rationale

A.1.1. Provide a summary of the background and rationale for this study (i.e., why is the study needed?). If a complete background and literature review are in an accompanying grant application or other type of proposal, only provide a brief summary here. If there is no proposal, provide a more extensive background and literature review, including references.

We propose a Carolina Open Health Data (CaOHD) infrastructure as an Observational Medical Outcomes Partnership (OMOP)-based open infrastructure to support large-scale, real-world evidence of drug effectiveness and adverse events, including the role of social and environmental determinants.

Electronic health records contain data on clinical phenotypes and drug exposures that represent "real-world" practice rather than the idealized setting simulated by clinical trials. Indeed, clinical trials are studied in relatively homogeneous and "healthy" patient populations. However, when drugs are used in real-world clinical practice, it is common for patients to have comorbidities and/or concomitant medication use and/or environmental exposures and/or be of an ethnicity, sex, or age that was not studied pre-clinically or selected for in clinical trials. These biases in patient recruitment for clinical trials often reduce the effectiveness of drugs in real-world clinical practice and may result in unexpected adverse events and toxicity, thus resulting in removal of drugs from the market after US FDA approval and incurring major financial losses in investments for clinical trials (Wysowski and Swartz, 2005; Nagaich and Sadhna, 2015; Onakpoya et al. 2019).

We hypothesize that through the application of statistical and machine learning algorithms, the proposed CaOHD will allow us to evaluate drug effectiveness and safety in real-world populations, meaning heterogeneous patient populations with uncontrolled comorbidities; concurrent medications; and diverse environmental exposures, family history, racial/ethnic backgrounds, sexes/genders, and ages. Moreover, by adopting the Columbia Open Health Data (COHD) open-source model, we will minimize the effort required to establish a similar OMOP-based open infrastructure at UNC, and we will be able to leverage the extensive experience and resources available through the Observational Health Data Sciences and Informatics (OHDSI) community (Observational Health Data Sciences and Informatics,2021) and the Biomedical Data Translator Consortium.

This effort will build on several existing efforts. First, we will leverage our collaborative work as part of the Translator program (e.g., Ahalt et al. 2019) with colleagues at Columbia University Irving Medical Center (CUIMC) and their COHD open infrastructure (Ta et al. 2018). COHD exposes data on counts and patient prevalence of conditions, procedures, drug exposures, and patient demographics, and the co-occurrence frequencies between them. Count and frequency data are derived from CUIMC's OHDSI database. COHD's 5-year dataset contains 29,964 single concepts (10,159 conditions, 10,264 drugs, and 8,270 procedures) and 15,927,195 concept pairs from 1,790,431 patients. Exclusion of rare concepts (count?=?10) and Poisson randomization enable open data sharing by eliminating risks to patient privacy. The software code supporting COHD is available in a public GitHub repository https://github.com/WengLab-InformaticsResearch/cohd_api). *Importantly, the COHD team will not have access to identifiable UNC Health patient data; rather, they will advise us on the COHD software code and the OMOP common data model.*

The proposed effort will also build on the Integrated Clinical and Environmental Exposures Service (ICEES) that we developed to openly expose and explore semi-aggregated clinical data that have been integrated with a variety of public data sources on environmental exposures (Fecho et al. 2019). Like COHD, ICEES is also open source and supported by a custom data integration pipeline termed FHIR Patient data Integration Tool (FHIR PIT) (Xu et al. 2020). ICEES was developed initially to support open research on clinical and

environmental determinants of asthma and related common pulmonary disorders; however, the framework itself is flexible, extensible, and disease-agnostic. Indeed, we have extended ICEES to support research on drug-induced liver injury, primary ciliary dyskinesia, and coronavirus infection. The software code supporting ICEES is available in a public GitHub repository (https://github.com/ExposuresProvider/icees-api).

The proposed CaOHD is a natural extension of both the COHD and ICEES efforts and will provide a complementary approach, focused on post-marketing drug effectiveness and safety research and surveillance. Importantly, the ICEES and COHD teams have worked closely together since the launch of the NCATS Biomedical Data Translator ('Translator') program and have multiple joint publications (e.g., Ahalt et al. 2019; Fecho et al. 2019; Fecho et al. 2023). Our collaborative work will help to ensure the success of the proposed work.

Our work to date on ICEES has been based on the i2b2, PCORnet, and FHIR common data models. However, with NC TraCS' recent move to OMOP as a preferred common data model, the time is right for ICEES to follow suit. Moreover, NC TraCS is currently preparing for a pilot of the "OMOP Repository of Data for Research, Deidentified" (ORDR-D), which will provide UNC researchers with access to a repository of OMOP-derived patient data, stripped of all PHI via HIPAA Safe Harbor. The proposed CaOHD will directly complement that work, but it will focus on co-occurrences of concept pairs (drugs, conditions, procedures) and drug–related adverse outcomes, adopting the COHD framework.

References:

Ahalt SC,* Chute CG, Fecho K, Glusman G, Hadlock J, Solbrig H, Overby-Taylor C, Pfaff E, Ta C, Tatonetti N, Weng C, and The NCATS Biomedical Data Translator Consortium. Clinical data: sources and types, regulatory constraints, applications. Clin Transl Sci 2019;12(4):329–333. doi: 10.1111/cts.12638. [K. Fecho, lead author] **All authors are listed in alphabetical order.*

Fecho K*, Bizon C, Issabekova T, Moxon S, Thessen AE, Abdollahi S, Baranzini SE, Belhu B, Chung L, Crouse A, Duby MP, Ferguson S, Friedman J, Forero L, Foksinska A, Gardner V, Glusman G, Hadlock J, Hanspers K, Hinderer E, Hobbs C, Hyde G, Huang S, Koslicki D, Mease P, Ramsey SA, Roach J, Rubin I, Shalev A, Schurman SH, Smith B, Soman K, Stemann S, Su AI, Ta C, Watkins PB, Williams MD, Wu C, Xu CH; and The Biomedical Data Translator Consortium. An approach for collaborative development of a federated biomedical knowledge graph–based question-answering system: question-of-the-month challenges. *J Clin Transl Sci 2023;7(1):E214. doi.org/10.1017/cts.2023.619. *Apart from the first five authors, all other authors are listed in alphabetical order.*

Fecho K,* Arunachalum S, Champion J, Chute CG, Gersing K, Glusman G, Hadlock J, Lee J, Pfaff E, Robinson M, Sid E, Ta C, Xu H, Zhu R, Zhu Q, Peden DB, and The Biomedical Data Translator Consortium. Sex, obesity, diabetes, and exposure to particulate matter: scientific insights revealed by analysis of open clinical data sources during a five-day hackathon. *J Biomed Inform* 2019;100:103325 [Special Communication]. doi: 10.1016/j.jbi.2019.103325. **Apart from first/lead and last/senior author, all other authors are listed in alphabetical order.*

Fecho K, Pfaff E, Xu H, Champion J, Cox S, Stillwell L, Bizon C, Peden D, Krishnamurthy A, Tropsha A, Ahalt SC. A novel approach for exposing and sharing clinical data: the Translator Integrated Clinical and Environmental Exposures Service. *J Am Med Inform Assoc* 2019;26(10):1064–1073. doi: 10.1093/jamia/ocz042.

Nagaich U, Sadhna D. Drug recall: An incubus for pharmaceutical companies and most serious drug recall of history. *Int J Pharm Investig.* 2015;5(1):13-19. doi:10.4103/2230-973X.147222

Observational Health Data Sciences and Informatics. The Book of OHDSI, January 11, 2021. https://ohdsi.github.io/TheBookOfOhdsi/.

Onakpoya IJ, Heneghan CJ, Aronson JK. Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature [published correction appears in BMC Med. 2019 Mar 2;17(1):56].*BMC Med*. 2016;14:10. doi:10.1186/s12916-016-0553-2

Ta CN, Dumontier M, Hripcsak G, Tatonetti NP, Weng C. Columbia Open Health Data, clinical concept prevalence and co-occurrence from electronic health records. *Sci Data.* 2018;5:180273. Published 2018 Nov 27. doi:10.1038/sdata.2018.273

Wysowski DK, Swartz L. Adverse drug event surveillance and drug withdrawals in the United States, 1969-2002: the importance of reporting suspected reactions. *Arch Intern Med*. 2005;165(12):1363-1369. doi:10.1001/archinte.165.12.1363

Xu H, Cox S, Stillwell L, Pfaff E, Champion J, Ahalt SC, Fecho K. FHIR PIT: an open software application for spatiotemporal integration of clinical data and environmental exposures data. *BMC Medical Informatics and Decision Making* 2020;20:article 53. doi: 10.21203/rs.2.19633/v1.

A.1.2. State the research question(s) (i.e., specific study aims and/or hypotheses).

Hypothesis:

We hypothesize that through the application of statistical and machine learning algorithms, the proposed CaOHD will allow us to evaluate drug effectiveness and safety, as well as the role of social and environmental determinants, in real-world populations, meaning heterogeneous patient populations with uncontrolled comorbidities; concurrent medications; and diverse environmental exposures, family history, racial/ethnic backgrounds, sexes/genders, and ages. Moreover, by integrating CaOHD with Translator, we will be able to suggest mechanistic explanations for real-world observations derived using CaOHD.

Study Aims:

Aim 1: Leverage our prior work on ICEES, the OMOP data repository at NC TraCS, and our collaboration with the COHD to implement the CaOHD.

Aim 2: Apply CaOHD to large-scale, real-world research on drug-related health outcomes and adverse events, including the role of social and environmental determinants.

Aim 3: Promote and integrate the CaOHD with the broader Translator ecosystem to generate mechanistic insights into CaOHD-derived real-world observations.

A.2. Subjects

A.2.1. Total number of subjects proposed across all sites by all investigators (provide exact number; if unlimited, enter 9999):

9999

A.2.2. Total number of subjects to be studied by investigators being provided oversight by the UNC IRB. (provide exact number; if unlimited, enter 9999):

9999

A.2.3. If the above numbers include multiple groups, cohorts, or ranges or are dependent on unknown factors, or need any explanation, describe here:

We propose a CaOHD open infrastructure to support large-scale, real-world research on drug-related health outcomes and adverse events, including the role of social and environmental determinants. Our inclusion criteria are all patients represented in the NC TraCS OMOP data repository who have had at least two visits to the UNC Health system within a five-year (2018–2022) study period. We do not have any exclusion criteria.

A.2.4. Do you plan to enroll subjects from these vulnerable or select populations:
If you will include children, prisoners or nonviable neonates or neonates of uncertain viability, please check the appropriate category below and complete the additional sections.

You should check "Pregnant women" if you specifically intend to recruit women who are pregnant or are not excluding pregnant women in biomedical research that is greater than minimal risk. Do not check if you are conducting a survey of the general public or conducting secondary data analysis or chart review not aimed at pregnant women.

Only check UNC-CH Student athletes, athletic teams, or coaches if you have specific plans to enroll these subjects. This is not applicable for intramural or club sports. For definitions and guidance see SOP 1201: Vulnerable subjects in research.

- ✔ Children (under the age of majority for their location)

Any minor subject who attains the age of majority during the course of the research study must provide consent as an adult, unless consent has been waived, which is requested in section D.3.1.

- ✖ Pregnant women
- ✖ Nonviable neonates or neonates of uncertain viability
- ✖ Prisoners, others involuntarily detained or incarcerated (this includes parolees held in treatment centers as a condition of their parole)

If an enrolled participant becomes incarcerated during the course of the research, they must be removed from the research project until such time as the IRB (and OHRP for NIH funded projects) approves the study to include prisoners, unless there is an immediate risk to the participant from ending treatments under the protocol.

- ✖ UNC-CH Student athletes, athletic teams, or coaches

A.2.5. Based on your recruitment plan and target sample population, are you likely to include any of the following as subjects? Select all that apply. This is not applicable to secondary data analysis or chart review.

Based on your responses, the consent form builder will insert the required text into your consent form template.

- ✖ Decisionally impaired individuals
  (e.g., Mini mental state examination (MMSE), Montreal cognitive assessment (MOCA))
- ✖ Children who are wards of the State (Foster children)
- ✖ Non-English-speaking individuals
- ✖ UNC-CH Students
- ✖ UNC-CH Employees
- ✖ People, including children, who are likely to be involved in abusive relationships, either as perpetrator or victim.
  This would include studies that might uncover or expose child, elder or domestic abuse/neglect. (See SOP Appendix A)

A.2.6. If any of the above populations are checked (excluding 'Decisionally impaired individuals' and 'Children who are wards of the State (Foster children)'), please describe your plans to provide additional protections for these subjects.

N/A

A.2.7. Age range of subjects:

Minimum age of subject enrolled	0
	years
Maximum age of subject enrolled	99
» If no maximum age limit, indicate 99	
	years

A.2.A. Children

Research involving children (45 CFR 46 Subpart D or 21 CFR 50 Subpart D)

A.2.A.1. Why is it necessary to involve children as subjects for this research? If the study addresses a condition that particularly affects children, please explain.

This study aims to support a rich integrated analysis of real-world outcomes of drug exposures and social and environmental determinants of health and disease. Children are exquisitely sensitive to drug exposures and environmental factors and thus are important to include in the study.

A.2.A.2. Describe potential for direct benefit to children participating in this study OR if no prospect of direct benefit to children participating in this study, explain how research is likely to yield generalizable information about the condition. If applicable, please explain how benefit would differ for children randomized to active (i.e. treatment or intervention) versus placebo (i.e inactive or control) groups.

The children whose data we will be analyzing will receive no direct benefit from this study. However, the study will support a rich integrated analysis of real-world outcomes of drug exposures and social and environmental determinants of health and disease. Children are exquisitely sensitive to drug exposures and environmental factors and thus are important to include in the study. Our aim is to study real-world outcomes, which is not possible with traditional clinical trial design (i.e., randomization to treatment or intervention versus placebo).

A.2.A.3. Describe the unique risks associated with children AND discuss your plans to minimize the risks and provide additional protections.

This study will not involve any direct contact with subjects. Rather, we will be using existing biomedical data (i.e., electronic health records) to assess real-world outcomes of drug exposures and other environmental exposures using existing biomedical data (i.e., electronic health records). As such, the children whose data we will be analyzing will not face any unique risks other than a possible breach of confidentiality. However, we will minimize this risk by: limiting the number of PHI elements to only those that are needed to link clinical data with environmental exposures data (patient geocodes, visit dates, and birth dates); storing the data within RENCI's secure enclave; and restricting data access to only those team members who specifically need access. We will adapt all security measures that have been embedded within the design of our current ICEES instances (see General Information, 3. Related IRBs).

A.3. Inclusion/exclusion criteria

A.3.1. List required characteristics of potential subjects (i.e., inclusion and exclusion criteria). If not covered, list also characteristics that would preclude their involvement.

We will *not* be enrolling any subjects as part of this work, and thus no individual patient will be contacted by members of the study team, nor will informed consent be obtained. Rather, we will be working with existing biomedical data only and requesting data only on phenotypes that are shared by ten or more patients.

Specific Inclusion Criteria: Patients represented in the NC TraCS OMOP data repository who have had at least two visits to the UNC Health system within the five-year study period.

Specific Exclusion Criteria: None

A.3.2. Justify any exclusion based on race, gender or ethnicity

N/A

A.3.3. Will pregnant women or women who become pregnant be excluded?

No

A.4. Study design, methods and procedures

Your response to the next question will help determine what further questions you will be asked in the following sections.

A.4.1. Will you be using any **methods or procedures commonly used in biomedical or clinical research** (this would include but not be limited to drawing blood, performing lab tests or biological monitoring, conducting physical exams, administering drugs, or conducting a clinical trial)?

No

A.4.2. Describe the study design. List and describe study procedures, including a sequential description of the study activities. When relevant, include in your description what subjects will be asked to do.

We propose a CaOHD open infrastructure to support large-scale, real-world research on drug-related health outcomes and adverse events, including the role of social and environmental determinants.

Our inclusion criteria are all patients represented in the NC TraCS OMOP data repository who have had at least two visits to the UNC Health system within a five-year (2018–2022) study period. We do not have any exclusion criteria.

Initially, we will focus on the OMOP concepts of conditions, drugs, and procedures for both inpatient and outpatient visits. CaOHD will allow users to openly explore UNC Health data and ask questions such as: *Which procedures and patient demographics are most likely to be associated with postoperative opioid-related respiratory depression? Can we predict adverse drug reactions based on comorbidities? Which racial and socioeconomically disadvantaged groups have the highest rates of drug-related adverse outcomes, and for what drugs?*

After we have implemented and tested the CaOHD open infrastructure, and demonstrated its value in the context of driving use-case questions, we will adapt our FHIR PIT data integration pipeline to integrate the OMOP-derived patient data with environmental exposures data, as is done with ICEES. That work will enable us to extend CaOHD to ask environmental health–related questions such as: *Does exposure to high levels of airborne pollutants increase the risk of 1-year mortality after a cardiac event? Is attendance at a public school located within <500 meters from a major roadway or highway associated with nonresponse to treatment among school-age children with asthma? Does residential location within < 1 mile of a landfill increase the risk of neurologic disease among patients >65 years?*

Study Design: Retrospective data collection for a five-year dataset, January 1, 2019 through December 31, 2023.

Specific Data Elements Requested: DOB, US Census Bureau census block group for primary home residence, visit dates, and OMOP conditions, drugs, and procedures (i.e., condition_occurrence table, drug_exposure table, procedure_occurrence table). *Of importance, geolocations and dates will be used solely for linkage to environmental exposure estimates; all data elements will be deidentified per HIPAA Safe Harbor method after the data integration step and before openly exposing the data via CaOHD.*

Specific Inclusion Criteria: None.

Specific Exclusion Criteria: None.

Study Aims:

Aim 1: Leverage our prior work on ICEES, the OMOP data repository at NC TraCS, and our collaboration with the COHD team to implement the CaOHD.

Aim 2: Apply CaOHD to large-scale, real-world research on drug-related health outcomes and adverse events, including the role of social and environmental determinants.

Aim 3: Promote and integrate the CaOHD with the broader Translator ecosystem to generate mechanistic insights into CaOHD-derived real-world observations.

A.4.3. If subjects are assigned or randomized to study "arms" or groups, describe how they are assigned.

- *Describe the methods of computing the randomization schedule (if any) and maintaining blinding (if any).*
- *Who will perform these computations?*
- *How will you verify each subject's eligibility prior to randomization?*

N/A

A.4.4. Describe any follow up procedures.

N/A

A.4.5. Once this study has been approved by the IRB, for how many months or years will this study be active (you are collecting data or have access to identifiers)?

We anticipate that the study will remain active for one year after IRB approval. Access to PHI elements is required for integration of the OMOP clinical data elements with environmental exposures data. After the integration step, all identifiers will be stripped of PHI per HIPAA Safe Harbor.

A.4.6. Will this study use any of the following methods?

✗	Audio Recording
✗	Video Recording
✗	Behavioral observation - (e.g., Participant, naturalistic, experimental, and other observational methods typically used in social science research)
✗	Pencil and paper questionnaires or surveys
✗	Electronic questionnaires or surveys
✗	Telephone questionnaires or surveys
✗	Interview questionnaires or surveys
✗	Other questionnaires or surveys
✗	Focus groups
✗	Diaries or journals
✗	Photovoice
✗	Still photography
✗	Unencrypted Messaging with Participants (e.g., text messages, unencrypted emails)

A.4.7. If there are procedures or methods that require specialized training, describe who (role/qualifications) will be involved and how they will be trained.

N/A

A.4.8. Are there cultural issues, concerns or implications for the methods to be used with this study population?

No

A.5. Benefits to subjects and/or society

A.5.1. Describe how this study will contribute to generalizable knowledge that will benefit society.

The proposed CaOHD open infrastructure will support research will allow us to evaluate drug effectiveness and safety, as well as the role of social and environmental determinants, in real-world populations. Moreover, the proposed CaOHD will be integrated with the broader Translator ecosystem, thereby allowing us to harness the power of the many tools and algorithms that are available as part of the Translator program to generate mechanistic insights into disease pathology and suggest potential drug targets for CaOHD-derived real-world observations.

A.5.2. Does this study have the potential for direct benefit to individual subjects in this study?

No

Consider the nature, magnitude, and likelihood of any direct benefit to subjects. If there is no direct benefit to the individual subject, say so here and in the consent form, if there is a consent form. Do not cite monetary payment or other compensation as a benefit.

Explain

This study provides no direct benefits to individual subjects.

A.5.3. Are there plans to communicate the results of the research OR results of any clinical tests administered for the research back to the subjects?

No

A.6. Risks and measures to minimize risks

A.6.1. All research poses a potential risk for a breach of confidentiality. Please describe your plan for minimizing this risk.

All research projects involving clinical data carry a certain level of risk of breach of confidentiality. However, we will minimize this risk by: limiting the number of PHI elements to only those that are needed to link clinical data with environmental exposures data (patient geocodes, visit dates, and birth dates); store the data on a server within RENCI's secure enclave; and restrict data access to only those team members who specifically need access. We will adapt all security measures that have been embedded within the design of our

current ICEES instances (see General Information - Related IRBs).

For each of the following categories of risk you will be asked to describe any items checked and what will be done to minimize the risks.

A.6.2. Psychological

- ✖ Emotional distress
- ✖ Other

A.6.3. Describe any potential psychological risks checked above and what will be done to minimize these risks

No Answer Provided

A.6.4. Social

- ✖ Loss of reputation or standing within the community
- ✖ Harms to a larger group or community beyond the subjects of the study (e.g., stigmatization)
- ✖ Other

A.6.5. Describe any potential social risks checked above and what will be done to minimize these risks

No Answer Provided

A.6.6. Economic

- ✖ Loss of income
- ✖ Loss of employment or insurability
- ✖ Loss of professional standing or reputation
- ✖ Loss of standing within the community
- ✖ Other

A.6.7. Describe any potential economic risks checked above and what will be done to minimize these risks.

No Answer Provided

A.6.8. Legal

- ✖ Disclosure of illegal activity
- ✖ Disclosure of negligence
- ✖ Other

A.6.9. Describe any potential legal risks checked above and what will be done to minimize these risks

No Answer Provided

A.6.10. Physical

- ✖ Medication side effects
- ✖ Pain
- ✖ Discomfort
- ✖ Injury
- ✖ To a nursing child or a fetus (either through mother or father)

A.6.11. Describe any potential physical risks checked above, including the category of likelihood and severity, and what will be done to minimize these risks. Where possible, describe the likelihood of the risks occurring, using the following terms:

- Very Common (approximate incidence > 50%)
- Common (approximate incidence > 25 - 50%)
- Likely (approximate incidence of > 10 - 25%)
- Infrequent (approximate incidence of > 1 - 10%)
- Rare (approximate incidence < 1%)

Describe severity of risks using the following grading scale:

- Mild- No disruption to the subject's ability to perform daily activities; may include non-prescription intervention only
- Moderate- Temporary interference with daily activities; may include prescription intervention
- Severe- Interference with daily activities; medically significant but not life threatening
- Life threatening

Examples:

Rare (< 1%) and Severe: blindness
Rare (< 1%) and Mild: dry skin, dry mouth, transient headache

If you are using these terms differently than described above, please provide your study-specific definitions.

Phase 1 trials: Due to limited experience, incidence may be better described as the number of events that have occurred in the total number of animals/humans studied.

No Answer Provided

A.6.12. Unless already addressed above, describe procedures for referring subjects who are found, during the course of this study, to be in need of medical follow-up or psychological counseling

No Answer Provided

A.6.13. Are there plans to withdraw or follow subjects (or partners of subjects) who become pregnant while enrolled in this study?

No

A.7. Data and safety monitoring

A.7.1. When appropriate, describe the plan for monitoring the data to ensure the safety of participants. These plans could range from the investigator monitoring subject data for any safety concerns to a sponsor-based data and safety monitoring board or committee (DSMB, DSMC, DMC), depending on the study. For studies that do not raise obvious safety concerns, you may still describe your plans for monitoring the study as it progresses.

Not applicable.

A.7.2. If not already addressed above, describe the plans for aggregate review of unanticipated problems (including but not limited to adverse events) across all sites, in order to monitor subject safety.

Not applicable.

A.7.3. What are the criteria that will be used to withdraw an INDIVIDUAL SUBJECT from this study or halt the research intervention (e.g., abnormal lab tests, allergic reactions, failure or inability to comply with study procedures, etc.)?

Not applicable.

A.7.4. Are there criteria that will be used to stop the ENTIRE STUDY prematurely (e.g., safety, efficacy, unexpected adverse events, inability to recruit sufficient number of subjects, etc.)?

No

A.7.5. Will this study involve a data and safety monitoring board or committee?

No

A.8. Data analysis

A.8.1. Summarize the statistical analysis strategy for each specific aim.

Our primary goal is to deploy a CaOHD open infrastructure to support research drug-related health outcomes and adverse events, including the role of social and environmental determinants. We will be applying both traditional statistical approaches (e.g., Chi Square analysis) and exploratory analyses (e.g., random forest models). Moreover, as part of Translator, we will be conducting a sophisticated graph analysis intended to generate mechanistic hypotheses to real-world observations derived from CaOHD.

A.8.2. If this is a pilot study, please describe the future study and say how its study design, aims, sample size, and methods differ from the pilot study you are proposing.

Not applicable.

A.8.3. Provide a compelling justification for the proposed sample size in terms of the likelihood of achieving each aim.

As noted above, our primary goals is to deploy a CaOHD open infrastructure to support research drug-related health outcomes and adverse events, including the role of social and environmental determinants We will not be conducting hypothesis-drive research and will not be performing a power calculation. The sample size is listed as "9999" because we are requesting a five-year dataset, January 1, 2019 through December 31, 2023 on all patients available in the CDWH.

A.8.4. Summarize the plans for data management.

We are requesting PHI (i.e., geocodes and dates) in order to integrate the OMOP data set at the patient level using a custom software pipeline that our team has developed. We will store the data on a secure RENCI server and restrict data access to only those team members who specifically need access. We will adapt all data management and security measures that have been embedded within the design of our current ICEES instances.

A.9. Identifiers

A.9.1. Check which of the following identifiers you already have or will be receiving, or select "None of the above."

- ✗ Names (this would include names/signatures on consent forms)
- ✗ Telephone numbers
- ✓ Any elements of dates (other than year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death. For ages over 89: all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 and older
- ✓ Any geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code and their equivalent geocodes (e.g. GPS coordinates), except for the initial three digits of a zip code
- ✗ Fax numbers
- ✗ Electronic mail addresses
- ✗ Social Security numbers
- ✓ Medical record numbers
- ✗ Health plan beneficiary numbers
- ✗ Account numbers
- ✗ Certificate/license numbers
- ✗ Vehicle identifiers and serial numbers (VIN), including license plate numbers
- ✗ Device identifiers and serial numbers (e.g., implanted medical device)
- ✗ Web universal resource locators (URLs)
- ✗ Internet protocol (IP) address numbers
- ✗ Biometric identifiers, including finger and voice prints
- ✗ Full face photographic images and any comparable images
- ✗ Any other unique identifying number, code, or characteristic, other than dummy identifiers that are not derived from actual identifiers and for which the re-identification key is maintained by the health care provider and not disclosed to the researcher
- ✗ None of the above

A.9.2. For any identifiers checked, how will these identifiers be stored in relationship to the research data?

- ✓ with the research data (i.e., in the same data set and/or physical location)
- ✗ separate from the research data (i.e., coded with a linkage file stored in a different physical location)

Provide details about the option you selected above:
We are requesting PHI (i.e., geocodes and dates) in order to integrate the various data sources at the patient level using a custom software pipeline that our team has developed. We will store the data on a secure RENCI server and restrict data access to only those team members who specifically need access. We will adapt all data management and security measures that have been embedded within the design of our current ICEES instances.

A.9.3. Are you collecting Social Security Numbers to be used as a unique identifier for study tracking purposes for national registry or database? (Do not check yes if collecting SSN *only* for payment purposes; this will be addressed later.)

No

A.10. Confidentiality of the data

A.10.1. Describe procedures for maintaining confidentiality of the data you will collect or will receive (e.g., coding, anonymous responses, use of pseudonyms, etc.).

All identified data will be stored on (and not moved or copied from) a secure server supplied and managed by RENCI, which has both an established infrastructure for the storage of PHI data for research purposes and an established history of compliance with HIPAA and institutional regulations. Any data that will be moved from the RENCI server will be derived deidentified versions of the raw dataset, with none of the requested PHI included.

A.10.2. Describe how data will be transmitted among research team (i.e., personnel listed on this application).

Patient data will not be transmitted from its location on RENCI's secure server, as all study team members will be able to access the server. The only data that might leave that location (be "transmitted," either via email or other file transfer mechanism) will be derived deidentified versions of the raw dataset, with none of the requested PHI included.

A.10.3. Are you collecting sensitive information such as sexual behavior, HIV status, recreational drug use, illegal behaviors, child/physical abuse, immigration status, etc?

No

A.10.4. Do you plan to obtain a federal Certificate of Confidentiality for this study? If your Sponsor will obtain or has already obtained a Certificate of Confidentiality, please select "Yes". You should also select "Yes" if your study is NIH funded and has been issued a CoC under the updated NIH policy.

NOTE: Investigators utilizing ANY federal funding to conduct this research should review the COC website to determine if their funding agency issues COCs automatically (as the NIH does) or if they might need to apply for the COC via the online COC system. Unfunded and non-federally funded investigators may also apply for a COC via the online COC system.

Information needed for CoC request:

- IRB approval documentation
- Name of Institutional Official: Carley Emerson, as designated by Penny Gordon-Larsen
- Email Address of Institutional Official: carley_emerson@unc.edu
- Phone: 919-966-6893

Yes

A.10.5. If this study is limited to data collection by survey or interview, discuss the potential for deductive disclosure (i.e., directly identifying subjects from a combination of indirect IDs).

Not applicable.

A.10.6. Will any of the groupings or subgroupings used in analysis be small enough to allow individuals to be identified?

No

A.11. Data sharing and transmission

A.11.1. Check all of the following who will receive **identifiable data** (contains any of the 18 identifiers listed above) outside the immediate research team (i.e., not listed as personnel on this application)? *

- ✔ No one
- ✘ Coordinating Center
- ✘ Statisticians
- ✘ Consultants
- ✘ Other researchers
- ✘ Registries
- ✘ Sponsor and/or its designee(s)
- ✘ External labs for additional testing
- ✘ Journals
- ✘ Publicly available dataset
- ✘ Other

A.11.2. For any recipients checked above, explain the confidentiality measures to be taken

No Answer Provided

A.12. Post-study disposition of identifiable data or human biological materials

A.12.1. Describe your plans for disposition of data or human biological specimens that are identifiable in any way (directly or via indirect codes) once the study has ended. If you plan to destroy linkage codes or identifiers, describe how and when this will be done.

After the study has ended, RENCI's standard data archive process will be implemented. This process removes all access to the secure server that stored the identified patient data and archives the data on an encrypted server accessible by no one other than RENCI IT staff.

Part C. Existing Data, Records, Specimens

C.1. Data Sources

C.1.1. What existing records, data or human biological specimens will you be using? (Indicate all that apply or select 'None of the above'):

- ✔ Medical records in any format.

  ALERT: You must check both boxes: 1) Medical records in any format and 2) Electronic medical record using Epic, or you/your study team will not be granted access to Epic for research purposes.

- ✘ Electronic medical records using Epic, WebCIS or other electronic system
- ✔ Carolina Data Warehouse for Health (CDW-H) (for UNC and its affiliates only)
- ✘ Carolinas Collaborative Data Request and Review Committee (DRRC)
- ✘ Paper medical records

If you access the medical records of fewer than 50 patients under a full or limited waiver of HIPAA, submit a copy of your IRB approval letter and a completed Research Disclosure Form to Health Information Management (HIM). Do not submit this information to the IRB. For additional information about this process, you should contact HIM directly at : 919-595-5591 or 919-966-1225 or 919-595-5580.

- ✘ Data already collected from another research study

Were the investigators for the current application involved in the original collection?	--

✗ Patient specimens (tissues, blood, serum, surgical discards, etc.)

Has the clinical purpose for which they were collected been met before removal of any excess?	--

✗ Data already collected for administrative purposes

✗ Student records (You will need to satisfy FERPA requirements: see SOP 3101, section 3.1 for guidance)

✗ UNC Dental Records

✗ Data coming directly from a health plan, health care clearinghouse, or health care provider?

✗ Publicly available data

✗ Other

✗ None of the above

For EACH data source checked above, provide a description of the data, proposed use, how data were collected (including consent procedures), and where data currently reside.

We will be requesting a five-year dataset (2018–2022) on all patients represented in the OMOP data repository within NC TraCS who have had at least two visits to the UNC Health system within the five-year study period. We will be focusing on the OMOP concepts of conditions, drugs, and procedures for both inpatient and outpatient visits. We will be linking the OMOP dataset to other datasets (e.g., environmental exposures), as described elsewhere in this application. As such, we will be requesting the minimal PHI elements required for the linkage between the OMOP dataset and the environmental datasets (i.e., geocodes, study dates, birth dates). The data currently reside within the secure enclave at NC TraCS UNC SOM. We will be requesting a waiver of consent.

C.1.2. Describe your plans for obtaining permission from the custodians of the data, records or specimens (e.g., pathology dept, tissue bank, original researcher):

We will submit a 'CaOHD Request' form, per NC TraCS protocol.

C.1.3. Do the custodians of the data, records or specimens require a data use agreement?

No

C.2. Coding and Data Use Agreements

C.2.1. When you receive these data, records or human biological specimens will they be coded? Coded means identifying information that would enable the research team to readily ascertain the individual's identity has been replaced with a number, letter, symbol, or combination thereof (i.e., a code). If you will not be using existing materials, check "No."

No

Part D. The Consent Process

D.3. Full or partial waiver of consent

The default is for subjects to give informed consent. A waiver might be requested for research involving only existing data or human biological specimens. More rarely, it might be requested when the research design requires withholding some study details at the outset (e.g., behavioral research involving deception). In limited circumstances, parental permission may be waived. This section should also be completed for a waiver of HIPAA authorization if research involves Protected Health Information (PHI) subject to HIPAA regulation, such as patient records.

D.3.1. Are you requesting any of the following:

✓ a waiver of informed consent in its entirety
✗ a waiver or alteration of some of the elements of informed consent
✓ a waiver of HIPAA authorization (If you are accessing patient records for this research, you must also request a waiver of HIPAA authorization)

Will you access the records of 50 or more patients under this waiver?

Yes

If you access the records of fewer than 50 patients under this waiver, submit a copy of your IRB approval letter and a completed Research Disclosure Form to Health Information Management (HIM). Do not submit this information to the IRB. For additional information about this process, you should contact HIM directly at : 984-974-3226.

To justify a waiver of the requirement for informed consent, you must affirm, by checking each of the following items that apply to this study. Provide an explanation.

Explain how the research involves no greater than minimal risk to subjects or to their privacy

There will be no patient contact for this study, and the requested PHI is the minimum necessary for integration of clinical and social/environmental exposures data at the patients and visit level. In addition, we are following all recommendations provided by an external cybersecurity consultant following a comprehensive risk assessment of ICEES.

Explain how the waiver will not adversely affect the rights and welfare of subjects (Consider the right of privacy and possible risk of breach of confidentiality in light of the information you wish to gather.)

The risk of breach of privacy/confidentiality is minimal. We are following our current practices with the ICEES instances, which invoke the NC TraCS established procedures for the provision and storage of PHI data. Moreover, we are following all recommendations provided by an external cybersecurity consultant following a comprehensive risk assessment of ICEES.

Please explain why it would not be possible to conduct the study with only de-identified data (i.e. without any identifiers listed in A.9.)

This project would be impossible to conduct without the ability to obtain the PHI noted in section A.9, namely, geocodes and dates, which will be used to integrate data sources. After integration, the data will be stripped of all PHI per §164.514(b) of HIPAA, 'Safe Harbor' method for patient de-identification of medical records.

Explain how the requirement to obtain consent would make the research impracticable, e.g., most of the subjects are lost to follow-up or are deceased.

The proposed work will use existing biomedical data on subjects within the CDWH. There will be no patient contact for this study, and the requested PHI is the minimum necessary for integration of clinical and social/environmental exposures data at the patient level. The expected number of subjects is quite large, given that we are requesting a dataset of all patients within the NC TraCS OMOP data repository who have had at least two visits over a five-year study period. As such, the research would be impracticable without a waiver of informed consent and HIPAA authorization.

Explain (or indicate if not applicable) how, when appropriate, there are plans to provide subjects with pertinent information after their participation is over. (e.g., Will you provide details withheld during consent, or tell subjects if you found information with direct clinical relevance? This may be an uncommon scenario.)

Not applicable.

Explain how the risk to privacy is reasonable in relation to the importance of the knowledge to be gained

We propose a CaOHD open infrastructure to support large-scale, real-world research on drug-related health outcomes and adverse events, including the role of social and environmental determinants. We will then integrate CaOHD with the broader Translator infrastructure, which will allow us to suggest biological mechanisms to explain the real-observations derived from CaOHD. Given the importance of work in this area, as well as the steps we will take to minimize risks of breaches to confidentiality, we believe that the benefits of the proposed work far outweigh the privacy risks.

In addition, please check the following and provide a brief explanation to justify a waiver of HIPAA Authorization (see SOP 1801, 2.3).

Explain how not recording or using Protected Health Information (PHI) would make the research impracticable.

This project would be impossible to conduct without the ability to obtain the PHI noted in section A.9, namely, geocodes and dates, which will be used to integrate data sources. After integration, the data will be stripped of all PHI per §164.514(b) of HIPAA, 'Safe Harbor' method for patient de-identification of medical records. The proposed work will use existing biomedical data on subjects within the UNC PCD Registry who have already consented to release their survey data and EHR data. There is no need to obtain consent. There will be no patient contact for this study, and the requested PHI is the minimum necessary for integration of clinical and social/environmental exposures data at the patient level.

D.3.2. If your request for a waiver applies to some but not all of your subject groups and/or consent forms, please describe and justify

Not applicable.

D.3.3. Does this request for waiver support a study design that involves deception or withholding of information?

No

Attachments

This submission requires the following attachments

Document Type
Grant Application

This submission includes the following attachments

File Name	Document Type
Ahalt_Translator_Y4ProgressReport_Y5FundingRequest FINAL OCT2023.pdf	Grant Application

view attachments

Addenda

Data Security Requirements

view addenda

If Principal Investigator of this study is a Student or Trainee Investigator, the Faculty Advisor certifies the following:

I accept ultimate responsibility for ensuring that this study complies with all the obligations listed above for the PI.

By certifying below, the Principal Investigator affirms the following:

I will personally conduct or supervise this research study. I will ensure that this study is performed in compliance with all applicable laws, regulations and University policies regarding human subjects research. I will obtain IRB approval before making any changes or additions to the project. I will notify the IRB of any other changes in the information provided in this application. I will provide progress reports to the IRB at least annually, or as requested. I will report promptly to the IRB all unanticipated problems or serious adverse events involving risk to human subjects. I will follow the IRB approved consent process for all subjects. I will ensure that all collaborators, students and employees assisting in this research study are informed about these obligations. All information given in this form is accurate and complete.

This study proposes research that has been determined to include Security Level 3 data security requirements. I agree to accept responsibility for managing these risks appropriately in consultation with departmental and/or campus security personnel. The Data Security Requirements addendum can be reviewed here.

Certifying Signatures:

Signature: Electronic Signature Received **Date:** 1/30/2024 06:11:43 AM
Ashok Krishnamurthy

The expectation is that this approval is being given on behalf of the head of the Department, Division, or Center. If the chair or director is an investigator on this project or otherwise conflicted in approving it, the Vice-Chair or Chair's designee should review it. By approving, you are certifying the following on behalf of your department, division or center:

- This research is appropriate for this Investigator and our department
- The investigator(s) are qualified to conduct the research
- There are adequate resources (including financial, support and facilities) available
- For units that have a local review committee for pre-IRB review, this requirement has been satisfied
- I support this application, and hereby submit it for further review

This study proposes research that has been determined to include Security Level 3 data security requirements. I agree to accept responsibility for managing these risks appropriately in consultation with departmental and/or campus security personnel. The Data Security Requirements addendum can be reviewed here.

- If you are approving for other purposes (e.g., CTRC, DSMB, IBC, PRC, RSC, or other review committees), you affirm the following: The proposed submission is approved and may be forwarded for IRB review.

Department Approval Signatures:

By signing in the appropriate space, the Department Chairperson(s) is indicating only that he/she has seen and reviewed this submission

Department: Renaissance Computing Institute
Signature: Electronic Signature Received **Date:** 1/30/2024 12:28:54 PM
Name & Title: Asia Mieczkowska, Director, Strategic Research Operations