

Center for Data-Driven Drug Development and Treatment Assessment (DATA) Whitepaper for Prospective Companies

Introduction

The University of Michigan's Institute for Data Science ([MIDAS](#)) has been awarded a planning grant by the National Science Foundation's [Industry-University Cooperative Research Centers \(NSF IUCRC\)](#) program to establish a research center for precompetitive research into patient phenotyping and pharmacovigilance using novel machine learning and artificial intelligence techniques. We are seeking industry partners whose research efforts align with the proposed Center's focus and who will be active collaborators in shaping the future paradigm of integrated and cost-effective patient care. The motivation for the Center, its research areas, and your potential role within it are detailed below.

Motivation

Despite the rigor of clinical trials and regulatory oversight for bringing new pharmaceuticals/biologics to market, many new therapeutics result in severe adverse events or have lower efficacy in the general population than pre-market data suggest. The ability to account for patient heterogeneity within clinical trials is limited by cost and clinical trial logistics such as patient recruitment. One potential solution is to leverage computational/in-silico approaches for differentiating amongst subgroups in the subject population. However, few machine learning techniques exist for teasing out the complexities of drug interactions, preexisting conditions, and unique patient phenotypes that result in disparate treatment effects and outcomes that are observed outside of the highly controlled setting of clinical trials. Moreover, the fragmented nature of healthcare and post-market surveillance data systems, their high latency, along with a primary emphasis on reporting only basic statistical information related to adverse events impede efforts to leverage this data for improving patient outcomes. The proposed Center's computational approach to health monitoring using novel techniques for identifying patient phenotypes related to adverse events, drug interactions, and treatment outcome in integrated public and proprietary datasets can address the healthcare and pharmaceutical data silos stymying comprehensive pharmacovigilance efforts while improving the state of patient care at the national level.

Center for Data-Driven Drug Development and Treatment Assessment (DATA) will focus on five main areas with the goal of significantly accelerating overall drug development and assessment of the outcomes associated with these treatments, while reducing national health care expenditures. These research areas represent the general capabilities that the Center will offer, but can be readily extended to meet the needs of participating industry partners:

a) *Development, testing, and validation of algebraic, machine learning, and artificial intelligence techniques for drug development, health monitoring and patient phenotyping with respect to drugs/treatments* – State of the art algebraic, machine learning, and artificial intelligence techniques together with advances in mobile/remote health monitoring will be brought to bear on discovering patient phenotypes related to treatment outcomes/efficacy and drug toxicity/adverse effects; using public, academic, and industry partner-provided patient and drug databases. These include advanced tensor algebraic techniques that were designed to systematically integrate auxiliary information currently ignored or under-utilized by extant methods. Additionally, novel machine learning-based methodologies for mobile/at-home patient monitoring will also be further developed.

b) *Provide an industry-wide and vendor-agnostic Secure Data Hub with third-party private search capabilities for the development, testing, validation and assessment of drugs and treatments* – By applying advanced algebraic and statistical methods, the Center will create large, encrypted, and integrated databases of patient and pharmaceutical data for which the participating companies need

only share full encryptions of their proprietary databases. Such databases enable third-party private searches, allowing partners to search securely and anonymously for data related to therapies or patient populations of interest.

c) Enable federated machine learning for drug design, health informatics and pharmacovigilance over encrypted databases – Center investigators have adapted several machine learning algorithms for use over encrypted data. The Center will further integrate these algorithms into a federated learning framework. Such technologies will enable machine learning over databases that cannot be integrated into the Secure Data Hub.

d) Enable data driven approaches to developing health care policies with respect to drugs and treatments – Utilizing the patient phenotypes, health analytics, and comprehensive adverse event/drug interactions data developed and collected by the Center, policy makers, professional organizations, and patient advocacy organizations will be able to better revise standards of care and patient outreach methods to improve patient outcomes in response to drugs and treatments while reducing costs related to unnecessary or ineffective therapies.

e) Provide unique lab resources for development and validation of patient phenotypes and new drug discovery – University of Michigan provides unique high throughput in-vitro and in-vivo capabilities, facilitated by multiple labs, centers and cores, which will be made available as resources for developing new drugs informed by discovered patient phenotypes and post-market surveillance data, closing the loop on the pharmaceutical development pipeline.

About the NSF IUCRC Program and Your Potential Collaboration

The proposed Center will leverage the existing community of data science and domain knowledge experts of MIDAS and the University of Michigan. MIDAS and its [affiliated faculty](#) will actively engage with industry partners to construct and implement a computational platform using novel algorithmic approaches to extract, validate, and integrate information from different drug/target datasets in a secure and vendor-agnostic manner.

The NSF IUCRC program has been enabling long-term research partnerships between industry, academia, and government for over 40 years. The IUCRC program facilitates the establishment of a thematic center focused on pre-competitive research projects, in which participating members have the ability to solicit and select proposals for development. The NSF funds a single center at the national level for any one research area and has indicated its support for our proposal to be **the** NSF-funded center fulfilling the national need for improved health monitoring and post-market surveillance.

The creation of a new center is a multi-phase process. We are currently in the planning phase, during which we will determine the Center's organizational structure, operating procedures, intellectual property (IP) policies, and initial experimental plan.

During this planning phase, financial commitment letters will be required for all companies who wish to become members of the proposed Center. Membership entails royalty-free use of intellectual properties developed through the Center's research efforts, along with voting rights on the Industry Advisory Board (IAB), with each member receiving a pro-rated number of votes based on their membership status. We anticipate the following annual membership fee structure: Full Membership at \$80,000 for one vote, \$120,000 for one and one-half votes, or \$160,000 for two votes; and Associate

Membership at \$40,000 for one-half vote. Members exercise their voting rights through the Center's IAB, choosing which pre-competitive research projects to fund. The NSF provides a minimum of \$150,000 annually in the first 5 years of the Center's operations, with additional matching funds depending upon industry participation. The NSF IUCRC program stipulates that 90% of all membership and NSF monies directly fund research projects.

Meetings were held with prospective Center members on August 17-19, during which attendees had an opportunity to learn more about the Center and review multiple pre-competitive research project proposals for potential inclusion in the initial set of proposals to be evaluated by the Industry Advisory Board (meeting recordings are available on request). Prospective members and MIDAS will now work together to finalize the proposal through mid-November, at which time financial commitment letters are due. Upon completion of the planning phase, the full proposal will be sent to the NSF for consideration by December 8, 2021. We anticipate the NSF proposal review process to take six to nine months, with a potential Center start date of September 2022.

Benefits of Participating in DATA

1. Define problems/projects, have a team of U-M/industry members work on the project, and have access to all resulting solutions at precompetitive level
 - a. Note that the IDC rate for the projects defined within IUCRC is 10%, as opposed to 56% for direct projects with U-M.
 - b. The process of agreeing/signing contracts with U-M (or any other university) is often very long and time-consuming, while the projects defined within IUCRC are processed and funded very quickly.
2. Access to highly skilled talent (students, postdocs, etc.) for future hiring
 - a. Ability to observe potential future employees throughout the course of one or more projects and assess their capabilities.
 - b. We will have regular workshops through which students/postdocs can have direct discussions regarding their career plans and options.
3. Access to additional funding directly from NSF
 - a. NSF provides grants to industry partners through programs such as [INTERN](#) and [REU](#).
 - b. Students/interns learn the job before getting hired by a company while getting paid by the NSF.
4. Ability to interact with other industry partners in a precompetitive setting and learn from each other's interests, needs, and capabilities
 - a. At least some parts of the ideas/solutions to data/algorithmic challenges may come from other industry partners.
 - b. Larger companies have the opportunity to learn about the capabilities of smaller companies participating in the Center.
 - c. Smaller companies have the opportunity to show their capabilities to larger entities.

Access to Resulting Solutions and IP

NSF emphasizes that centers must target common problems and pains by focusing on precompetitive partnerships. All solutions (data, algorithms, etc.) can be used by all participants but only within the framework of the Center. All companies can use the resulting solutions for the applications defined/funded by the Center but not as a product that they market. The Center focuses mainly on deployment of IP via a public, shared resource mechanism such as open source, open access (Creative

Commons), or public domain. However, it is *very* possible that the work conducted under the Center will result in new IP. The ownership of the technology will be governed by the Center's bylaws. U-M will work with the companies for technology transfer.

Resources Provided by University of Michigan

1. Expertise in data science and algorithms for health informatics. Samples of the algorithms and computational resources provided by the Center include:
 - Computational drug toxicity screening and drug interaction modeling using fully homomorphic encryption and coupled tensor methods (CMMC/CTMC)
 - Efficient and effective models for mobile/outpatient monitoring and prediction of adverse events using physiological signals
 - Tensor based machine learning methods for fusion of multimodal patient data for diagnostic and prognostic modeling.
 - Patient phenotyping using transparent artificial intelligence (AI), which enable interpretation of the recommendations/predictions created by computational methods
2. Data and lab resources related to health informatics and pharmacovigilance
U-M has already established successful research cores, labs, and resources that will be instrumental to the success of DATA, such as:
 - [Clinical Pharmacogenomics Laboratory](#)
 - [Institute for Healthcare Policy and Innovation](#)
 - [Michigan Drug Discovery](#)

To learn more about the Center for Data-Driven Drug Development and Treatment Assessment (DATA), please contact Dr. Kayvan Najarian (kayvan@med.umich.edu) or Dr. Jonathan Gryak (gryakj@med.umich.edu).