Synthetic Health Data Challenge Frequently Asked Questions

BASIC FACTS ABOUT PCOR AND SYNTHETIC HEALTH DATA

What is PCOR?

PCOR (Patient-Centered Outcomes Research) is a research field focused on producing scientific evidence comparing the effectiveness of various medical prevention and treatment options while also considering patients' health care preferences, values, and the questions they face when making health-care decisions. Robust data infrastructures that support rigorous analyses and generate relevant information strengthen the validity of PCOR findings.

What are synthetic health data?

Synthetic health data (sometimes called synthetic health records) is realistic (but not real) patient data and associated health records. This realistic data for fictional patients, which models patients from birth until death, is free of protected health information (PHI) and personally identifiable information (PII) constraints. Synthetic health data can be generated to meet the specific interests of PCOR researchers and developers for testing theories, data models, algorithms, and prototype innovations.

Why is synthetic health data important?

Researchers and developers depend on clinical data for testing research algorithms and/or technology while awaiting access to real clinical data. Unfortunately, cost, patient-privacy concerns, and other legal restrictions can make high quality, health- and health-care related data difficult to access. Anonymized data (data from the health records of actual patients with personal information stripped away) is often used. However, the risk of re-identification of anonymized data is high and, especially for rare conditions, impossible to completely eliminate.

Further, because of a variety of interoperability issues, it can be difficult to bring data together from different resources for the purpose of robustly testing analysis models, algorithms, or assisting in the development of software applications. After securing data, there are several processes that must be done before beginning to apply or use the data. For example, a researcher or heath IT developer will typically need to aggregate, de-identify, and analyze data before testing the effectiveness of algorithms and modeling approaches used in matching and disease modeling techniques. Interoperability issues also make it difficult to compile large amounts of data from different sources for the purposes of robustly testing analysis models or assisting with the development of software applications. Synthetic health data also offers the kind of built-in interoperability and integration of clinical and claims data that rarely exists in the real world.

BACKGROUND ON PCORTF, ONC, AND FEDERAL SUPPORT FOR CHALLENGES

What is PCORTF?

The Office of the Secretary (OS) of Health and Human Services, through the Patient-Centered Outcomes Research Trust Fund (PCORTF), is charged with coordinating relevant federal health programs to build data capacity for comparative clinical effectiveness research. Information about the OS-PCORTF Strategic Framework for PCOR data and its use can be found at https://aspe.hhs.gov/patient-centered-outcomes-research-trust-fund-faqs.

Who is ONC?

ONC (Office of the National Coordinator for Health Information Technology) is a federal entity that, as part of its broad public-service mission to coordinate nationwide efforts to implement and use the most advanced health information technology and the electronic exchange of health information, leads and collaborates on projects that inform policy, standards, and services specific to the adoption and implementation of a data infrastructure for PCOR. One of ONC's current projects, funded by PCORTF, is the Synthetic Health Data Generation Engine to Accelerate PCOR. The Synthetic Health Data Challenge is an important component of that project. More information about ONC can be found at https://www.healthit.gov/.

Why does the Federal government support Challenges?

Challenges enable the Federal government to tap into the expertise and creativity of the public. Under a directive calling for innovative ways to generate ideas and collaboration, Challenges are policy tools that can foster participation in government activities through the process of co-creation. Challenges may offer a variety of prizes, including cash, recognition, or the deployment of a winning solution. For more information about Challenges, visit https://www.healthit.gov/topic/innovation/health-it-prizes-and-challenges-faqs.

OVERVIEW OF THIS CHALLENGE

Is the Challenge part of a larger project?

The Synthetic Health Data Challenge is an important part of the Synthetic Health Data Generation Engine to Accelerate PCOR project. The goal of the project is to enhance Synthea™, an open-source synthetic health-data generator, and to support PCOR research needs by increasing the number and diversity of available synthetic patient records. The project targets the areas of opioids, pediatrics, and complex care, because of the unique characteristics of these data needs. Increased availability of synthetic data for these priority areas will help expedite testing of research algorithms and technology. Part of the project evaluated existing Synthea data modules to assess opportunities for development and enhancement. To learn more about the Synthetic Health Data Generation Engine to Accelerate PCOR project, visit https://www.healthit.gov/topic/research-evaluation/synthetic-health-data-generation-accelerate-patient-centered-outcomes. To learn more about Synthea, visit: https://github.com/synthetichealth/synthea/wiki

What is the Synthetic Health Data Challenge?

The Synthetic Health Data Challenge (Challenge) is a prize competition that invites innovators, researchers, and technology developers to create and test innovative and novel solutions aimed at further cultivating the capabilities of Synthea and the synthetic health data it generates. The Challenge was implemented to

demonstrate novel uses and validate the realism of Synthea-generated synthetic health records. Information about the Challenge is available at: https://www.challenge.gov/challenge/synthetic-health-data-challenge

What are the categories for Challenge entries?

The Challenge has two entry categories:

Category I – Enhancements to Synthea. Solutions in this category include, but are not limited to, development or enhancement of Synthea modules and development of solutions that enhance or address limitations of Synthea.

Category II – Novel Uses of Synthea-Generated Synthetic Data. Solutions in this category include, but are not limited to, novel uses of Synthea generated data for research and technology development.

When does the Challenge take place?

The entire Challenge takes place between January 19, 2021 and July 13, 2021.

What are the Challenge phases?

The Challenge is broken into two (2) phases:

Phase I – Proposal for Innovative Models: Participants will submit a written proposal describing their proposed solution, including methodology and intended outcomes. Selected Phase I proposals will proceed to Phase II.

- Proposals are invited from teams or individuals.
- Any number of proposals can be submitted.
- Judges use scoring criteria to determine the Phase I proposals that move on to Phase II.

Phase II – Prototype/Solution Development: Phase I proposals that are selected to proceed to Phase II will develop their prototype/solution at this stage.

- Only Phase I winning proposals compete during Phase II.
- Judges use scoring criteria to determine the winning solutions from among the Phase II submissions.
- Winning solutions are eligible for cash awards.

What awards are Challenge participants eligible for?

The Challenge offers up to $100,000 in total prizes:

- Up to two (2) first place winners will receive $25,000.
- Up to two (2) second place winners will receive $15,000.
- Up to two (2) third place winners will receive $10,000.
- Honorable Mentions may be awarded but will not receive a monetary prize.

As an additional incentive, each winning entry and honorable mention will be invited to present during a public webinar in October 2021.

How will entries be judged?

Technical reviewers with expertise relevant to the Challenge will evaluate Phase I and Phase II submissions based on a list of criteria detailed on the Challenge webpage. The submissions and evaluation statements will then be reviewed by federal employees serving as judges. Judges have, at a minimum, basic knowledge and understanding of synthetic health data, PCOR research, and health IT standards and tools.

For Phase I, the judges will determine which submissions will move on to Phase II. For Phase II, the judges will select up to six (6) Challenge winners and may also name honorable mentions. Judging is subject to a final decision by the Award Approving Official: Donald Rucker, MD, National Coordinator for Health Information Technology.

Where is there more detailed information about the Challenge?

For more detailed information about all aspects of the Challenge, including how to enter, eligibility rules, key dates, and submission requirements, go to https://www.challenge.gov/challenge/synthetic-health-data-challenge.

Who should be contacted with questions about the Challenge?

Send feedback or questions about the Challenge to SyntheticDataChallenge@govhealth.com.