The Office of the National Coordinator for Health Information Technology

Synthetic Health Data Generation to Accelerate Patient-Centered Outcomes Research

Project Goal

Support patient-centered outcomes research (PCOR) by expanding the capabilities of Synthea™ to produce high-quality synthetic health data that increases the number and variety of synthetic health records available for researchers, health IT developers, and informaticians.

Synthetic health data can provide a lower risk data source to complement research and support testing needs until real clinical health data are available.

Objectives

- Convene a multidisciplinary panel of experts to provide guidance for the selection of use cases and module development.
- Develop Synthea modules in three focus areas – patients with complex care needs, opioid use, and pediatric populations.
- Engage a broad community to validate the realism and demonstrate potential uses of Synthea-generated synthetic health data.
- Disseminate project outputs that future Synthea users can refer to when developing Synthea modules and/or using Synthea-generated synthetic health data.

Major Project Components

- Prioritized Focus Areas & Use Case Selection
- Synthea Module Development
- Synthea Enhancement & Data Validation
- Project Outputs – Dissemination

Prepared for the Office of the National Coordinator for Health Information Technology under contract number HHSP233201500099I/75P00119R00375.

Prioritized Focus Areas

Opioid Use | Pediatric Populations | Complex Care Needs

Use-Case Selection Criteria

The following evaluation criteria helped guide use case identification and prioritization:

Importance

- Clinical significance of use case aligns with prioritized focus areas
- Increases number and variety of synthetic health records

Feasibility

- Availability of clinical care maps
- Availability of incidence and prevalence statistical data
- Use case scenarios, flows, and clinical concepts can be supported by the Synthea Generic Module Framework

Reliability

- Scientifically reliable incidence and prevalence rates

Use

- Allows for validation of realism and/or potential use of generated synthetic health records
- Is not inhibited by the limitations of Synthea

Existing Modules

- Does not duplicate or overlap with existing modules/submodules

Selected Use Cases

Use Case	Prioritized Focus Area	Importance
Sepsis	Complex Care Needs	Leading cause of death in critically ill patients in the United States.
Prescribing Opioids for Chronic Pain & Treatment of Opioid Use Disorder	Opioid Use	Opioids prescribed for ~ 20% of patients with non-cancer pain symptoms or pain-related diagnoses.
Cerebral Palsy (sialorrhea)	Pediatric Populations	Uncontrolled hypersalivation occurs in ~ 40% of pediatric cerebral palsy patients.
Spina Bifida	Pediatric Populations	Most common, permanently disabling birth defect associated with life.
Acute Myeloid Leukemia	Pediatric Populations	Replicates study parameters comparing levofloxacin prophylaxis to usual care for leukemia patients undergoing chemotherapy.

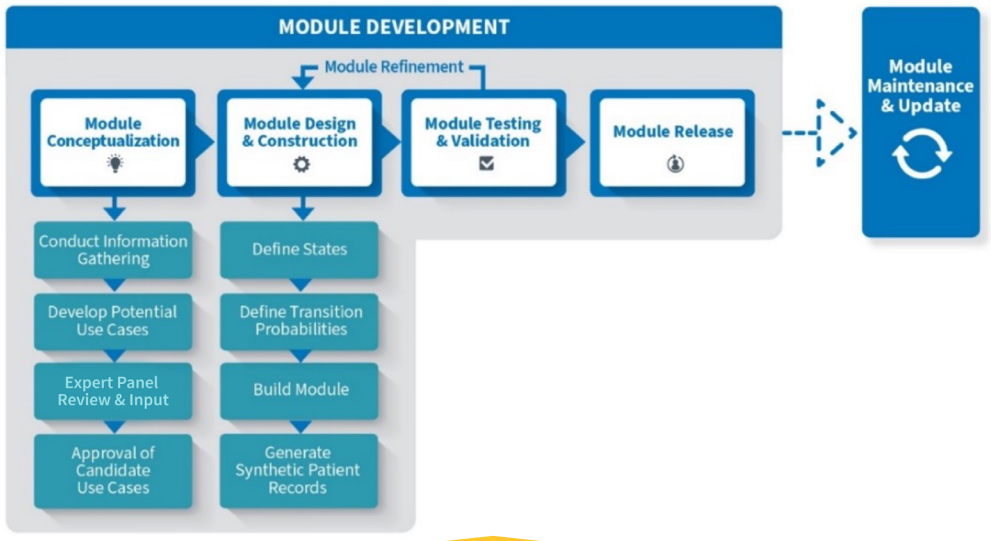
Synthea Module Development

Synthea modules are a foundational element driving the generation of synthetic health data.

Enhancing Module Quality

- Standardizing the module development process supported iterative module development and provides a methodology for future module builders.
- Developing module Companion Guides provided essential technical information for new modules and serves as a model for future module developers and implementers.

Module Development Methodology



Five new Synthea modules and Companion Guides were developed, validated, and published.

Synthea Enhancement & Data Validation

Synthetic Health Data Challenge

The Synthetic Health Data Challenge launched on January 19, 2021 and invited proposals for enhancing Synthea or demonstrating novel uses of Synthea-generated synthetic health data. Selected proposals moved on to the development phase and competed for $100,000 in total prizes.

Results

The Synthetic Health Data Challenge drew participants from across the United States who designed innovative solutions to enhance Synthea capabilities and validate synthetic data output.

Winning Solutions

First Place - $40,000

- **CodeRx** - Medication Diversification Tool

Second Place - $15,000

- **The Generalistas** - Virtual Generalist: Modeling Co-morbidities in Synthea
- **Team LMI** - On Improving Realism of Disease Modules in Synthea: Social Determinant-Based Enhancements to Conditional Transition Logic

Third Place - $10,000

- **Particle Health** - The Necessity of Realistic Synthetic Health Data Development Environments
- **Team TeMa** - Empirical Inference of Underlying Condition Probabilities Using Synthea-Generated Synthetic Health Data
- **UI Health** - Spatiotemporal Big Data Analysis of Opioid Epidemic in Illinois

Project Outputs - Dissemination

Click Below to Access

Project Background | **Final Report** | **Modules and Companion Guides** | **Synthetic Health Data Challenge Winning Solutions Webinar**