
Case Report

Case report: evaluation of an open-source synthetic data platform for simulation studies

Daniella Meeker¹, Crystal Kallem², Yan Heras³, Stephanie Garcia⁴, and Casey Thompson²

¹Keck School of Medicine, University of Southern California, Los Angeles, California, USA, ²Clinovations Government+Health, Washington, District of Columbia, USA, ³Optimum eHealth, LLC, Irvine, California, USA, and ⁴Office of the National Coordinator for Health Information Technology, Washington, District of Columbia, USA

Corresponding Author: Casey Thompson, MSN, RN-BC, Clinovations Government+Health, 1325 G Street, NW, Suite 500, Washington, DC 20005, USA; casey@govhealth.com

Received 15 November 2021; Revised 10 May 2022; Editorial Decision 4 June 2022; Accepted 28 July 2022

ABSTRACT

Simulation is a mainstay of comparative- and cost-effectiveness research when empirical data are not available. The Synthea platform, originally designed for generating realistically coded longitudinal health records for software testing, implements data generation models specified in publicly contributed modules representing patients' life cycle and disease and treatment progression. We test the hypothesis that Synthea can be used for simulation studies that draw parameters from observational studies and randomized trials. We benchmarked the results and assessed the effort required to create a Synthea module that replicates a recently published cost-effectiveness simulation comparing levofloxacin prophylaxis to usual care for leukemia. A module was iteratively developed using published parameters from the original study; we replicated the initial conditions and simulation endpoints of demographics, health events, costs, and mortality. We compare Synthea's Generic Module Framework to platforms designed for simulation and show that Synthea can be used, with modifications, for some types of simulation studies.

Key words: Electronic Health Records, computer simulation, privacy, software validation, leukemia

LAY SUMMARY

Clinical data are critical for research, which requires robust data infrastructures that support rigorous analyses and generate relevant information to strengthen the validity of findings. However, high-quality, real clinical health data (data from the health records of actual patients) can be difficult to access because of complex privacy and usage issues. Synthetic health data (the realistic, but not real, health data of fictitious patients) can provide a lower-risk data source to complement research use of real clinical health data. It can meet testing needs by providing realistic synthetic health records that reflect the characteristics of the general population, which can be used until real clinical health data are available. Simulations are used in healthcare research when desired empirical data are not available. This study evaluated Synthea™, a synthetic patient generator designed for software development and testing, for use in simulation studies. Synthea may also be useful for study planning, software testing, and development of analysis routines. While Synthea was not designed for simulation studies, it includes many features of commercial microsimulation software and Synthea modules can be configured, with reasonable effort, to support some simulation research.

INTRODUCTION

Simulations are used in health care when desired empirical data are not available, such as forecasting scenarios when changes in policies or practices may impact longitudinal population-level outcomes or cohort-level disease progression. This demonstration study focused on evaluating the use of Synthea V2.6.1 (Synthea) as a potential platform for simulation studies of patient healthcare utilization and outcomes. Such studies produce a simulated data set for statistical analysis comparing alternative treatment or policy scenarios or forecasting alternative future circumstances, including outcomes like disease progression, quality of life, and mortality.

Synthea¹ is an open-source, freely available software package designed to generate realistic, standards-based synthetic healthcare records for developing and testing technology operating against clinical data. Unlike other software that generates coded synthetic health data (eg, EMRBots),² Synthea enables user-specified models of healthcare and disease progression³ to generate longitudinal patient record sets with meaningful and technically realistic sequences of coded events. For this reason, researchers interested in health services research, economics, and policy have speculated that Synthea could be applied for simulation studies.⁴ Most existing Synthea modules use models derived from guideline pathways, rather than models that also include alternative or hypothetical pathways, including guideline-discordant pathways.⁵ Third-party evaluations of the validity of data generated with Synthea's existing modules revealed limitations to this guideline-based approach.³ However, Synthea's ability to support simulation-based research with appropriate models has not yet been evaluated.

Despite the differences in intended application, Synthea has features in common with microsimulation software designed to simulate healthcare outcomes and utilization, including the Future Elderly Model and other purpose-specific simulation software.^{6,7} Like many microsimulation models, Synthea's model parameters are typically sourced from domain experts and published studies. Synthea initiates with a set of patients having predefined transition parameters and initial conditions, each evolving through discrete states over time according to probabilistic models or rules to produce an analytic population. By default, Synthea simulates the initial conditions in the Massachusetts population but can be modified to reflect custom demographics. However, Synthea is intended for "innovation, development, education, and other nonclinical secondary uses."⁸ It does not support the ingestion of record-level data for estimating model parameters for high-fidelity deidentification of existing data sets^{9,10} or include user-friendly ways to specify complex, multivariate models from such data.

We hypothesized that Synthea can be repurposed for some types of simulation studies without changes to source code. For example, our demonstration evaluated Synthea's capability to replicate a published simulation study that employed commercial software. Successful replication would indicate that Synthea can be applied in this way. This evaluation complements Synthea's originally intended use as a means to generate technically useful, realistic records from existing public data in a way that does not impact patient privacy.¹¹

MATERIALS AND METHODS

Reproducing an existing study with published parameters enabled us to benchmark results from a new Synthea module

against study results and report module development methods and effort (we note that parameter development is often the most time-intensive part of simulation study design and Synthea module building, but not informative in evaluating the Synthea platform for simulation studies, and thus not included in this evaluation).

A literature scan identified simulation studies with published endpoints and parameters for initial conditions, health states, transition probabilities, and service utilization. We selected a 2020 cost-effectiveness analysis comparing levofloxacin prophylaxis to usual care for pediatric patients undergoing an episode of chemotherapy for acute myeloid leukemia (AML) (McCormick).¹² To summarize the parameter development process used in this study (referenced in detail in McCormick,¹² S1), published parameters were drawn from 31 studies and an original retrospective cohort analysis of AML patients in the Pediatric Health Information System database. Probabilities of intensive care unit (ICU) admission and mortality were drawn from several published studies, and cost parameters were drawn from the Federal Supply Schedule.¹³ As described above, this approach to data-driven modeling from research cohorts and process models differs from the prevailing guideline-based approach in existing Synthea modules: guidelines do not incorporate pathways that include adverse outcomes and guideline discordant treatment patterns.

To evaluate Synthea usability for new adopters, our module builders were moderately experienced with Synthea's generic module framework (GMF). They gained experience by developing four new modules for Opioids, Pediatrics, and Complex Care.¹⁴⁻¹⁷ The primary module builder is a nurse informaticist specializing in data standards and healthcare interoperability. The secondary builder, consulted for troubleshooting and quality assurance after each iteration, is a bioinformatics researcher specializing in data standards and clinical model design. A researcher with experience in simulation studies resolved complex parameter alignment issues when McCormick results required conversion to Synthea's specification format.

To help inform potential users of Synthea's challenges and capabilities, we adopted an iterative approach to Synthea module building, testing the differences between Synthea's output and the output in McCormick's et al. in each version of the Synthea module as shown in Table 1. By reporting the alignment and technical solutions required in each iteration, we provide insight into the level of effort and technical details of steps required to repurpose Synthea for simulation studies. To verify reproducibility, we conducted an iterative development process, building an AML module in Synthea conforming to McCormick results and model specifications and generating a new patient data set for analysis. After each iteration, Synthea-generated patients' demographic attributes, treatment, and health outcomes were compared to McCormick results. Point estimates in McCormick tables and [Supplementary Material](#) were compared to Synthea data using chi-squared and *t* tests for binary and continuous variables, respectively. Pass-fail tests for each iteration were assessed and updates made until all comparisons "passed" with $P > .001$. Issues were tracked, and module builders identified and implemented solutions. We measured effort required for each development cycle, issues encountered, and resolutions. Module builders relied on Synthea's public documentation¹⁸ and GitHub issue tracker¹⁹ throughout the process. Data analysis was conducted using Stata15.²⁰

Table 1. Synthea AML module iterations

AML Module Iteration	Challenges	Lessons Learned
Iteration 1	<ul style="list-style-type: none"> Initial version of module 	<ul style="list-style-type: none"> Developed expertise in states and transitions based on parameters in a simulation study
Iteration 2	<ul style="list-style-type: none"> Editing and enhancing cost table to match published costs for encounters, procedures, and medications 	<ul style="list-style-type: none"> Cost data are in specified in different places within Synthea, some cost data are contained within lookup tables, and the remainder are default costs in the Synthea demographics file
Iteration 3	<ul style="list-style-type: none"> Learning how to edit/enhance distribution table for table transition 	<ul style="list-style-type: none"> Table transitions are new to Synthea and not used by many modules. There is limited information related to this type of transition on the Synthea wiki
Iteration 4	<ul style="list-style-type: none"> Removed distribution table and used complex transitions 	<ul style="list-style-type: none"> Complex transitions for race/ethnicity caused a large proportion of patients to be eliminated from the module Costs were updated in Synthea cost lookup tables. When corresponding codes were not present in lookup tables, additional lines were added to accommodate new codes and affiliated costs
Iteration 5	<ul style="list-style-type: none"> Moved complex transitions to below the age guard/year guard in the module Run Synthea for a default city in a diverse area Create new city in Synthea with race and ethnicity which match McCormick 	<ul style="list-style-type: none"> Procedures and observations older than 10 years ago were not displaying in the output. A setting must be updated in the <code>synthea.properties</code> file to allow older observations and procedures to display in the output Updating the Synthea demographics file to create a new city requires updating the appropriate zip code and provider files and adding a latitude/longitude
Iteration 6 (V0.6a–0.6e)	<ul style="list-style-type: none"> Modified existing city in the Synthea default demographics file with parameters to match McCormick for base population along with table transition within the module to differentiate levofloxacin/nonlevofloxacin populations Added 22 additional age delays to the top of the module to accommodate for age 	<ul style="list-style-type: none"> Modifying an existing city with specific parameters, adding a distribution table within the module, and adding age delays in the module generated findings which matched McCormick

RESULTS

Developers designed initial Synthea module states and transitions based on the McCormick model diagram (Figure 1).

The AML module initialized with 25 states and 24 transitions and finalized with 47 states and 46 transitions. During development, 22 additional states were incorporated. The AML module reproduced McCormick model pathways, including assuming only febrile neutropenia patients develop bacteremia during the episode of care and routing only nonfebrile neutropenia patients to a terminal state. Transition probabilities and attribute distributions were based on McCormick Supplementary Table 1 parameters. Workarounds were developed to address Synthea's modeling interface limitations, resulting in the increased number of states and transitions. The final AML module is displayed in Figure 2.

Initial conditions

Synthea's default demographic data are based on city and state-level census data rather than specific patient cohorts.²¹ Replicating McCormick initial conditions for AML patients (eg, race, age) into Synthea's framework required configuring a city within the Synthea demographics file with parameters matching the McCormick population.

Transition probability specification

The AML module included state transitions for medication with levofloxacin, development of bacteremia, admission to an ICU, and mortality. The McCormick defining state transition for hypothesis testing was assignment to levofloxacin treatment or standard of care, and these groups were not demographically identical. Thus,

conditional transition probabilities were required. Synthea's GMF offers multiple mechanisms for conditional transitions, including *table-based transitions* enabling transition probabilities depending on patient demographics and health attributes. Synthea distribution tables generate a separate CSV file to be called upon during the simulation. Synthea documentation contained limited information on table transitions, so development required significant trial and error.

Continuous distributions of values

At the time of the study, Synthea supported only uniform probability distributions for transitions. Support for Gaussian distributions could have been useful when attempting to match certain McCormick distributions, including patients' age. Patient age at diagnosis (time elapsed between birthdate and date of diagnosis) was implemented as a Synthea *delay* state following birthdate. We approximated McCormick's Gaussian age distribution using 22 uniform distributions for delays between 0 and 21 years from birthdate to initial diagnosis. In response to user requests, Synthea developers recently added support for Gaussian distributions in the GMF, including distributions for laboratory values and delays. However, delays were not constrained to nonnegative values, resulting in simulated visits predating patient birth. Thus, the final version of the AML module relies on a series of uniform distributions rather than the new Gaussian features.

Cost specifications

Synthea's GMF supports customization of lookup tables for costs and default values based on Medicare cost files with adjustments for each state. Synthea cost data are based on a simplified version of real-

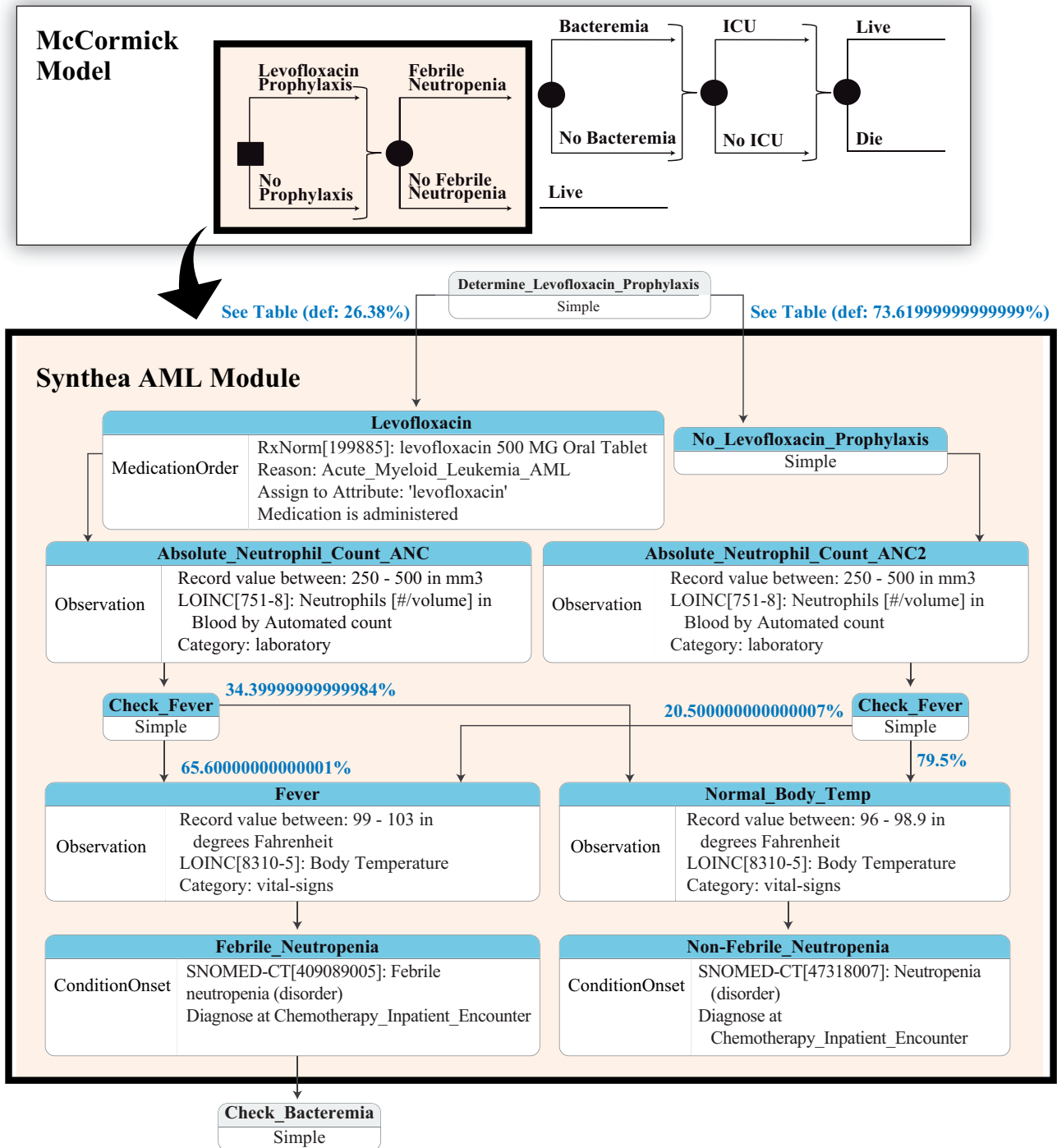


Figure 1. Comparison of McCormick model¹⁰ to Synthea AML module. Displays a comparison of a portion of the McCormick model to the Synthea AML module. AML: acute myeloid leukemia.

world costs—publicly available Medicare records are multiplied by state adjustment factors. Costs are provided for encounters, procedures, medications, and immunizations. Costs not provided in lookup tables are assigned a default cost in the Synthea properties file.²²

Levofloxacin was added to the medications cost lookup table with a McCormick-based specified cost. Bacteremia cost data were updated for the inpatient encounter in the encounters cost lookup table. ICU admission and mortality costs were updated in the procedures cost lookup table, and a code for death event was added.

Other parameters were adjusted to avoid impacting McCormick costs during generation. Tables were customized in the developer’s local version of Synthea.

Execution of simulation

To execute the AML module,²³ builders implemented the model and custom lookup tables for costs and transition probabilities. Details are in the AML module companion guide.²⁴

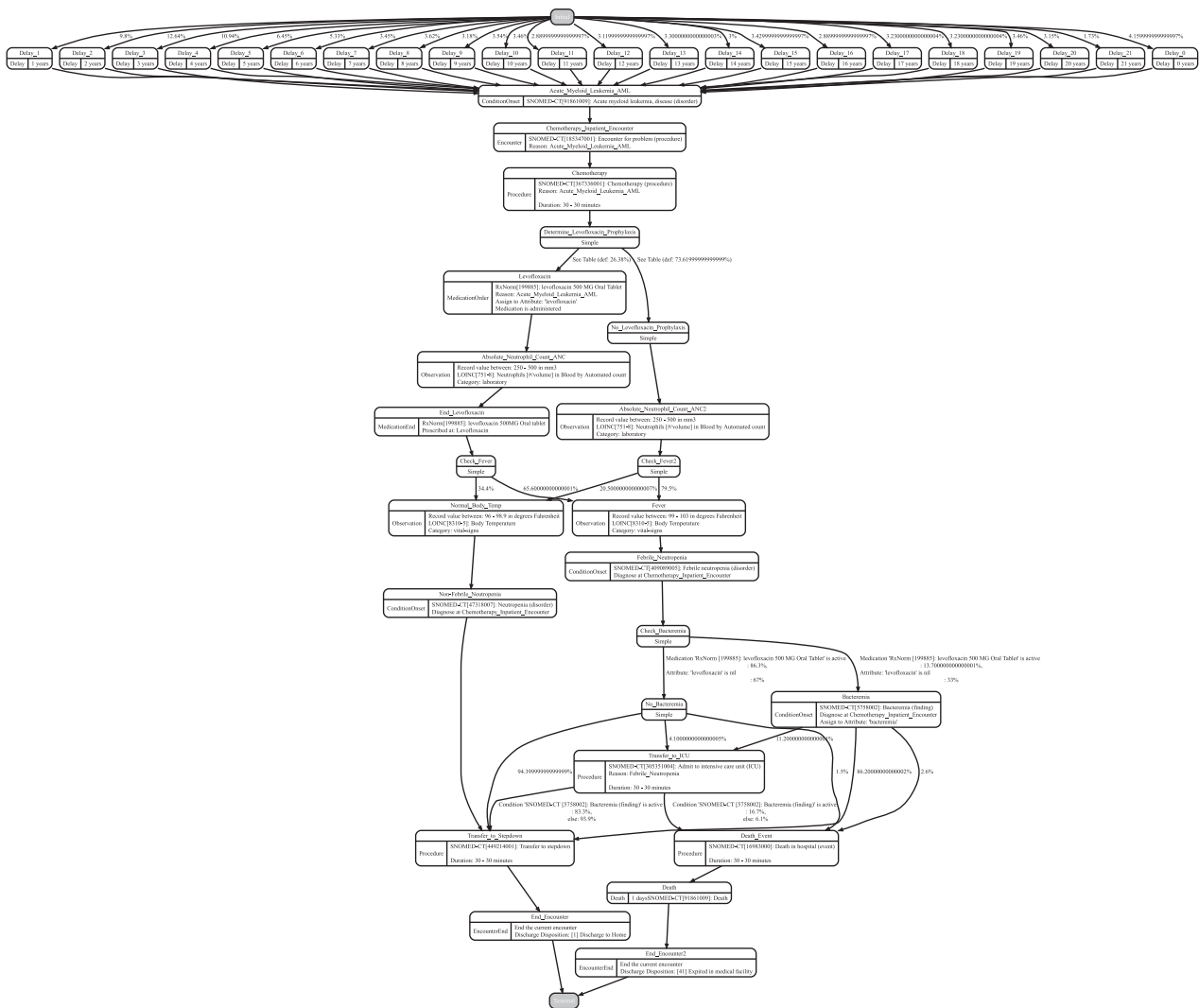


Figure 2. Final Synthea AML module (Iteration 0.6e).²¹ Displays the final version of the Synthea AML module as displayed in the module builder tool. AML: acute myeloid leukemia.

Statistical analysis

To fully reproduce the McCormick results, six AML module versions (with multiple sub-iterations of V0.6) were iteratively developed and tested. Module refinement continued until Synthea distributions for all patient attributes in both the treatment and standard of care populations were comparable to the point estimates reported in the McCormick results. Table 2 shows the results of comparisons between Synthea and McCormick for each sub-iteration of AML module V0.6. As described above, matching the Gaussian distribution for age was the most challenging aspect of module development, failing to replicate reference values in multiple attempts over the course of development.

The level of effort to develop the AML module was approximately 20% higher than other modules developed by the same team, largely due to the distinct objective of matching results of prior research rather than the more straightforward objective of matching guideline pathway diagrams without quantitative data validation. Matching McCormick parameters and a Gaussian distribution required creating numerous delay states and complex table transitions. Achieving a precise level of accuracy when comparing Synthea out-

put to McCormick results required developing and testing numerous iterations of the AML module.

DISCUSSION

Microsimulation models have been used for real-world applications in forecasting policies^{6,7} and assessing disease progression trajectories and intervention scenarios.²⁵ This demonstration shows that Synthea can be repurposed for simulation studies comparable to the reference study without advanced programming. Although work-arounds were required, replication was feasible. We found that Synthea could be readily modified to initialize simulations with a cohort matching the characteristics of the AML population. Synthea also has features allowing for more complex and conditional state transitions required to reproduce differential treatment allocation across our initial AML cohort. While reproducing these capabilities of simulation software required some nonintuitive uses of the Synthea GMF, the same approach could be used in future simulation studies using Synthea in other contexts.

Table 2. Summary of results for AML module V0.6

AML module iteration	V0.6a		V0.6b		V0.6c		V0.6d		V0.6e (w/age restriction)		V0.6e (w/delays)	
	Tx	SoC	Tx	SoC	Tx	SoC	Tx	SoC	Tx	SoC	Tx	SoC
Tx versus SoC	Tx	SoC	Tx	SoC	Tx	SoC	Tx	SoC	Tx	SoC	Tx	SoC
Population size	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Age (SD)	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Pass	Pass
Race: white	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Race: black	Fail	Fail	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Race: other/missing	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Hispanic ethnicity	Fail	Fail	Fail	Fail	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Bacteremia	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Fail	Pass	Pass	Pass
ICU	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass
Mortality	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass

AML: acute myeloid leukemia; ICU: intensive care unit; SoC: standard of care; Tx: treatment.

The McCormick simulation did not involve multivariate models for conditional state transitions or interactions between individuals. However, it is representative of a broad class of microsimulation studies. Future releases of Synthea may include user-friendly support for these features, extending the complexity of possible studies and models. Synthea currently lacks the capability to support complex multivariate distributions, but newly introduced support for nonuniform distributions (including univariate Gaussian distribution related to this study) may serve as a foundation for this capability. Enhancements to the documentation of Synthea's existing features may facilitate future simulation research.

For health economics research, Synthea can represent costs as they might appear in billing data for encounters, procedures, medications, and immunizations using billing codes provided in lookup tables. McCormick assigned a cost to each outcome and medication generically, thus cost comparisons can be calculated with these frequencies. Future analyses and module iteration might add this granularity and complexity and assess or extend Synthea's default values for costs. Although Synthea allows for more extensive modeling, the current strategy is sufficient for study designs like McCormick, where simple frequencies of generic simulated events are used to generate cost comparisons.

Other research applications

Our results also indicate Synthea can be used for study planning, software testing, and development of analysis routines for research studies. For complex experimental designs, like stratified or cluster randomized trials, simulations are often used in power analysis to prepare for clinical trials. Synthea can generate data for power analysis and sample size requirements. For example, should a follow-up to the McCormick study include a randomized trial of levofloxacin prophylaxis, investigators might apply our Synthea AML module to help plan sample size requirements and prepare data infrastructure. Because an arbitrary number of simulated patient records can be produced, Synthea has become the "gold standard" for performance benchmarking in health data management strategies.^{26,27} Synthea can also produce realistic patterns in healthcare records, comparable to data generated in Electronic Health Records (EHRs) during pragmatic trials. Informaticians can generate and use data to apply and test software that might, in a forthcoming trial or registry project, use EHR data. Once a database is designed with realistic data, anal-

ysis routines for data safety and other monitoring reports can be developed and tested before beginning data collection, reducing the time between implementing research infrastructure and initiating enrollment.

CONCLUSION

Synthea was not designed for simulation studies but includes many features of commercial microsimulation software. Synthea offers flexibility for estimating coded cost data in a way that allows even greater variation than the McCormick reference study. The user interface is not optimized for complex specifications of initial conditions and transition probabilities, but Synthea modules can be configured, with reasonable effort, to support some simulation research. Although Synthea is an open-source tool, it is actively supported by a team of MITRE developers who address new feature requests, consider the clinical validity of contributed modules, and manage other issues identified by the user community.

ACKNOWLEDGMENTS

The authors would like to thank Maureen Tan, for copy editing support and for reference formatting assistance. We would also like to thank Carmen Smiley from the Office of the National Coordinator for Health Information Technology and Anita Samarth from Clinovations Government+Health, for leadership and contributions on the overarching project that examined the scope and utility of synthetic data for patient-centered outcomes research. The work described here is part of the *Synthetic Health Data Generation to Accelerate PCOR* project that was supported by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund. Finally, the authors acknowledge the MITRE Synthea team for their assistance in responding to issues encountered over the course of this work.

FUNDING

This work was funded through U.S. Department of Health and Human Services Office of the Secretary Patient-Centered Outcomes Research Trust Fund under interagency agreement number 750119PEN0002, contract number HHSP233201500099I, task

order number: HHSP75P00119F37004 with Clinovations Government+Health.

AUTHOR CONTRIBUTIONS

All authors acknowledge that they: (1) made substantial contributions to the conception or design of the work, or the acquisition, analysis, or interpretation of the data for the work; (2) contributed substantially to the drafting and final approval of the version to be published; and (3) agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

SUPPLEMENTARY MATERIAL

Supplementary material is available at *JAMIA Open* online.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

Data are available from a GitHub Repository: <https://github.com/casey7083/Demonstration-Study-Output>.

REFERENCES

1. GitHub Synthetichealth/Synthea. Home. <https://github.com/synthetichealth/synthea/wiki>. Accessed August 8, 2021.
2. Kartoun U. A methodology to generate virtual patient repositories. *arXiv*. 2016;1608.00570.
3. Chen J, Chun D, Patel M, *et al*. The validity of synthetic clinical data: a validation study of a leading synthetic data generator (Synthea) using clinical quality measures. *BMC Med Inform Decis Mak* 2019; 19 (1): 44. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6416981/>. Accessed April 15, 2021.
4. Chido J. Synthetic patient records help deliver real health outcomes. *MITRE Project Stories*. February 2019. <https://www.mitre.org/publications/project-stories/synthetic-patient-records-help-deliver-real-health-outcomes>. Accessed May 13, 2021.
5. GitHub Synthetichealth/Synthea. Getting started. <https://github.com/synthetichealth/synthea/wiki/Getting-Started>. Accessed May 13, 2021.
6. Goldman DP, Lakdawalla D, Michaud P-C, *et al*. *The Future Elderly Model: Technical Documentation*. Roybal Center for Health Policy Simulation, University of Southern California; 2015. <https://healthpolicy.usc.edu/research-program/health-policy-simulation/>. Accessed June 14, 2021.
7. Goldman DP, Leaf DE, Tysinger B. *The Future Americans Model: Technical Documentation*. Roybal Center for Health Policy Simulation, University of Southern California; 2016. https://cehd.uchicago.edu/wp-content/uploads/2019/12/fam_techdoc.pdf. Accessed June 14, 2021.
8. Walonoski J, Kramer M, Nichols J, *et al*. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *JAMIA* 2018; 25 (7): 921.
9. Tucker A, Wang Z, Rotalinti Y, Myles P. Generating high-fidelity synthetic patient data for assessing machine learning healthcare software. *NPJ Digit Med* 2020; 3 (1): 147. <https://doi.org/10.1038/s41746-020-00353-9>. Accessed August 9, 2021.
10. Foraker RE, Yu SC, Gupta A, *et al*. Spot the difference: comparing results of analyses from real patient data and synthetic derivatives. *JAMIA Open* 2020; 3 (4): 557–66.
11. Walonoski J, Klaus S, Granger E, *et al*. Synthea™ novel coronavirus (COVID-19) model and synthetic data set. *Intell Based Med* 2020; 1: 100007.
12. McCormick M, Friehling E, Kalpatthi R, *et al*. Cost-effectiveness of levofloxacin prophylaxis against bacterial infection in pediatric patients with acute myeloid leukemia. *Pediatr Blood Cancer* 2020; 67 (10): e28469.
13. VA Federal Supply Schedule Service. <https://www.fss.va.gov/>. Accessed August 4, 2019.
14. Synthea Generic Module Builder: Cerebral Palsy. Treatment of sialorrhea in cerebral palsy. https://synthetichealth.github.io/module-builder/#cerebral_palsy. Accessed April 30, 2021.
15. Synthea Generic Module Builder: Prescribing Opioids for Chronic Pain and Treatment of OUD. Prescribing opioids for chronic pain and treatment of opioid use disorder. https://synthetichealth.github.io/module-builder/#prescribing_opioids_for_chronic_pain_and_treatment_of_oud. Accessed April 30, 2021.
16. Synthea Generic Module Builder: Spina Bifida. Spina bifida. https://synthetichealth.github.io/module-builder/#spina_bifida. Accessed April 30, 2021.
17. Synthea Generic Module Builder: Sepsis. Sepsis. <https://synthetichealth.github.io/module-builder/#sepsis>. Accessed April 30, 2021.
18. GitHub Synthetichealth/Synthea. Home. Welcome to the Synthea™ Wiki! <https://github.com/synthetichealth/synthea/wiki>. Accessed June 7, 2021.
19. GitHub Synthetichealth/Synthea. Issues. <https://github.com/synthetichealth/synthea/issues>. Accessed June, 7, 2021.
20. StataCorp. New in Stata 15. <https://www.stata.com/stata15/>. Accessed June 7, 2021.
21. GitHub Synthetichealth/Synthea. Default Demographic Data. <https://github.com/synthetichealth/synthea/wiki/Default-Demographic-Data>. Accessed June 22, 2021.
22. GitHub Synthetichealth/Synthea. Cost data. <https://github.com/synthetichealth/synthea/wiki/Cost-Data>. Accessed May 13, 2021.
23. Synthea Generic Module Builder: Acute Myeloid Leukemia. <https://synthetichealth.github.io/module-builder/>. Accessed September 22, 2021.
24. GitHub Synthetichealth/Synthea. Module companion guides. <https://github.com/synthetichealth/synthea/wiki/Module-Companion-Guides>. Accessed October 1, 2021.
25. Kausch SL, Lobo JM, Spaeder MC, Sullivan B, Keim-Malpass J. Dynamic transitions of pediatric sepsis: a Markov chain analysis. *Front Pediatr* 2021; 9: 743544.
26. Shoumik FS, Talukder MIMM, Jami AI, Protik NW, Hoque MM. Scalable micro-service based approach to FHIR server with golang and NoSQL. In: 2017 20th International Conference of Computer and Information Technology (ICIT). Piscataway, NJ: IEEE; December 22–24, 2017; Dhaka, Bangladesh.
27. smileCDR/Performance Testing. Benchmarking smile CDR at scale in AWS. <https://www.smilecdr.com/benchmarking-smile-cdr>. Accessed June 15, 2022.