Modification and Use of Synthea to Account for Patient Vaccination Choice

Submission for the Synthetic Health Data Challenge

By:

Michael D. Teter
miketeter@yahoo.com

Christopher E. Marks
cemarks@alum.mit.edu

Challenge Category: I (Enhancement to Synthea)

Modification and Use of Synthea to Account for Patient Vaccination Choice

Abstract: Recent trends have shown an increase in the number of parents that request non-medical exemptions (NMEs) to routine pediatric vaccinations [1, 3]. The resulting decline in childhood vaccinations has led to increases in childhood disease outbreaks [6]. In its current form, Synthea [14] does not account for these variations in pediatric vaccinations [2, p. 15], which are the result of patient health care preferences that do not align with accepted optimal approaches to health care. For this effort, we modify the existing immunization workflow in Synthea so that it accounts for patient preferences and produces synthetic health records (or synthetic patients) with a more varied vaccination history, specifically with respect to the Measles, Mumps, and Rubella (MMR) vaccine. We demonstrate the utility of this modification by developing a prototype module for childhood measles that is responsive to whether the patient has received the MMR vaccination. To validate the results, we use linear regression to compare the synthetic data produced to existing public vaccination estimates [6]. Finally, we comment on how this development enhances the capability of Synthea and propose multiple avenues for future research efforts that, using this capability, could have significant impacts on public health and facilitate Patient Centered Outcomes Research (PCOR).

Github: `https://github.com/teterholdings/Synthea_NME`
YouTube: `https://youtu.be/BVlPj1qW75w`

1 Overview

One of the known limitations of Synthea is that it does not allow for variation in care trajectories [2, p. 15]. This limitation is significant in the context of the Office of the National Coordinator (ONC) objective of enabling Patient-Centered Outcomes Research (PCOR) [2, pp. 3,9], as it does not account for individual patient preferences or objectives in applying treatments. One area where patient preferences are rapidly changing, and that could have serious consequences on public health, is in the area of vaccinations [3, 10]. Additionally, many US states have permissive non-medical exemption policies, which make it relatively easy for parents to decline routine vaccinations for their children.

In this effort, we modify Synthea software to account for parent non-medical exemption decisions for their children. We create a prototype "Measles" module that takes advantage of this code modification, omitting the MMR vaccine from some patients' records and making these patients susceptible to measles infection. These updates make it possible to analyze where there might be high-risk regions for measles in the US, and investigate what-if scenarios based on different assumptions about NME rates in the future.

While our use-case is focused on pediatric vaccination, our work also has particular relevance

to COVID-19 vaccination in the United States. Because this vaccine is elective, and also as a result of political polarization, a large portion of the population has not received it [1]. The methods and results we present could be applied to COVID-19 and other vaccinations to analyze the effects of different vaccination rates in order to predict medical resource requirements, educate the public, and inform public policy decisions.

2   Methods

To prototype the non-medical exemption capability, we modified the immunizations module [12] workflow in Synthea to look for a patient NME attribute before adding the immunization to the patient's record. For example, before administering an MMR vaccine, the software first checks for an attribute called "`mmr NME`." If this patient attribute exists, the MMR workflow is skipped and the MMR record is omitted from the patient's records.

This small change in the Immunizations module enables a user to write drop-in JavaScript Object Notation (JSON) modules that set NME attributes for patients, resulting in an unvaccinated subset of the population. Like any other modules, these modules can be called at runtime or compiled into Synthea.

We created the `measles.json` module depicted in Figure 1 [13] to apply this new capability. The module uses the lookup table transition capability, developed recently to facilitate COVID-19 modeling, to apply NME rates that vary over geographic region and time. We set the default NME rate at or below the national average (in Figure 1 it is 1.86%, the national average), using the lookup tables to incorporate NME rates information in US states for periods where this data is publicly available [4].

If a patient has an NME, then the patient is susceptible to measles infection. Once again, we use a lookup table transition to call the measles infection submodule for a subset of these patients, using workflow based on the existing Synthea COVID-19 modules. As depicted in Figure 1, these patients loop through the "Wait Until Exposure" state in four-day increments until they either are exposed to measles or terminate the module. We designed the measles exposure lookup table for this simulation state transition to loosely track with measles infection rates in the United States over the past 10 years [8]. The default transition probabilities include a very small probability that the patient will be exposed to measles; otherwise the patient continues to cycle through the "Wait Until Exposure" state.

If a patient is exposed, the patient transitions to the "Call Infection Submodule" state, which executes the measles infection workflow (not depicted). This workflow is based on the COVID-19 infection submodule, progressing through common measles symptoms and, in some cases, hospitalization and/or death.
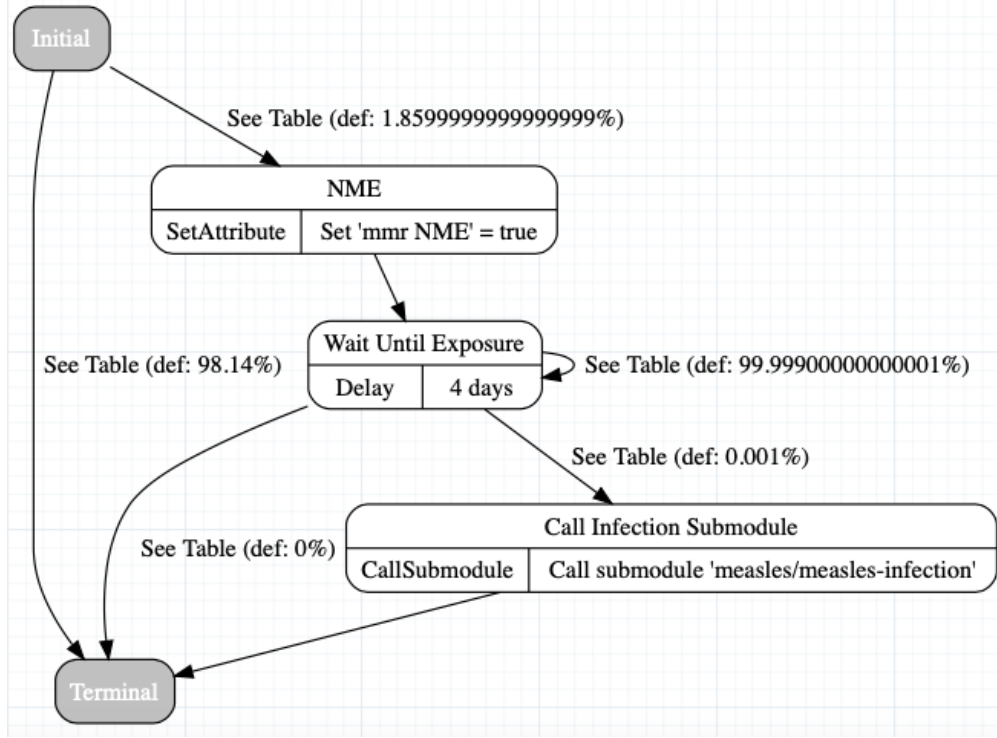
Figure 1: Measles NME module (`measles.json`).

3 Results and Validation

The Center for Disease Control (CDC) provides vaccine non-medical exemption rates by state for recent years [4], which we used to tailor the NME transition probabilities as discussed in the previous section. These same statistics can be used to validate the NME rates in the data in Synthea. For our implementation, we aim to ensure that synthetic NME rates and measles infection rates are approximately proportional to known data.

We generated approximately 5000 synthetic health records for each US state using our modified Synthea immunization and measles modules. In order to focus on the population that might have NMEs, we limited the record generation to patients ages 6-30. We then compared the rate of NMEs in kindergarten-aged children per state, per year against the known data points in the National Center for Immunization and Respiratory Diseases (NCIRD) [4] data. We found that the NME rates in the synthetic data correlated with the actual NME rates (correlation coefficient: 0.81, R^2: 0.65). This result validates that our implementation is producing more records with NMEs for locations and times where and when there are more people with NMEs in the United States. Figure 2 provides a scatter plot with the regression line and R^2 value for this analysis.

Using these same 5000 synthetic health records, we examined the measles infection rates and again compared against known measles infections data from the Center for Disease Control [8]. Figure 3 shows bar plots of measles infection rates in the United States from (a)
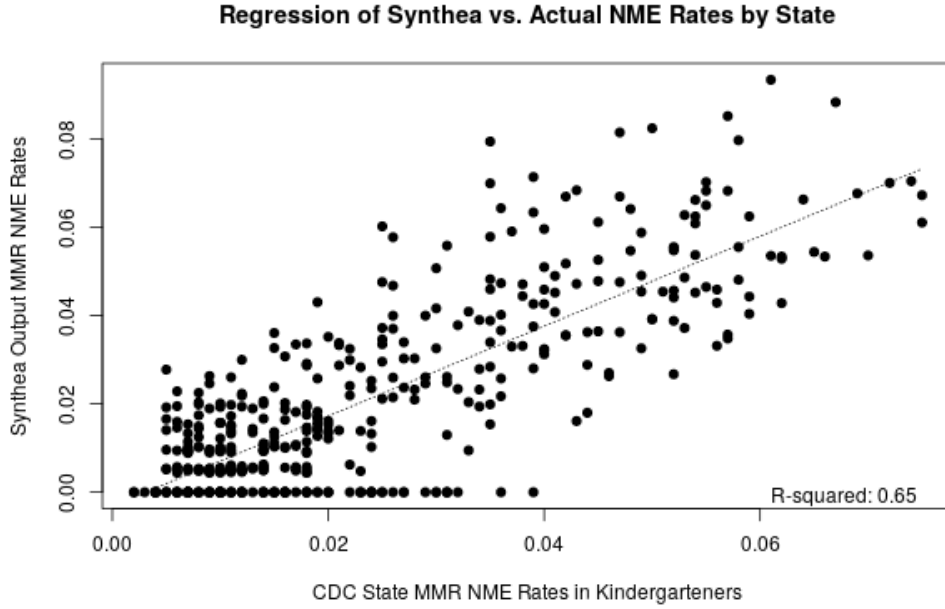
Figure 2: Regression of synthetic data NME rates against known NME rates in the US.

the synthetic (Synthea-generated) data, and (b) CDC monitoring. The synthetic measles infection rates track proportionately with observed infection rates.

(a) Synthetic measles rates.
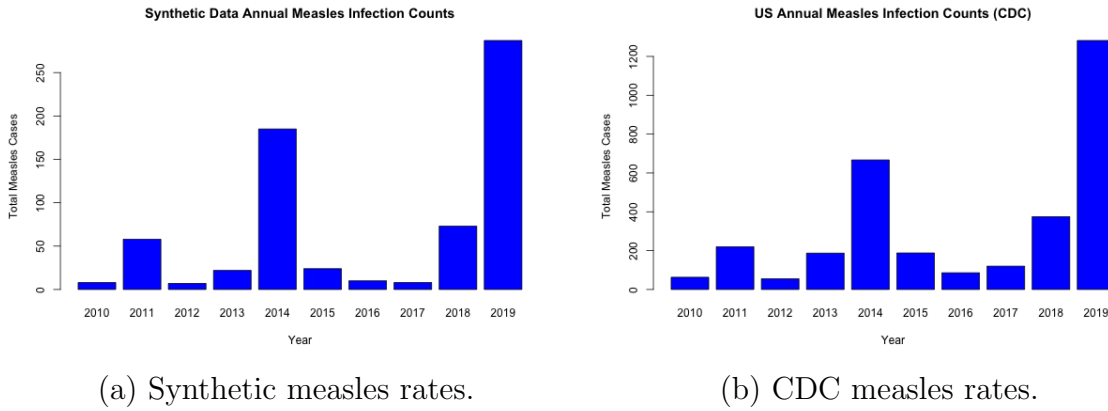

(b) CDC measles rates.

Figure 3: Comparison of measles infection rates in synthetic and CDC data.

The synthetic output can also be spatially compared to known measles incidence. Figure 4 provides a density map of the 2019 US measles outbreak [7], along with a heat map of all of the Synthea-generated measles cases. A qualitative comparison of the two maps shows some similarities and some differences. Consistencies between the two plots result from the infections occurring in population centers where NME rates are high, most notably in the northeast and northwest. Discrepancies are inevitable; other than preferring populated areas for measles infections, we did not make an effort to geographically match known outbreaks. What is important is the qualitative similarity between the two maps: each as a few places

with large numbers of cases, a more substantial number of places with low infection counts, and a large portion of the country (north & central states) with very few or zero infections.
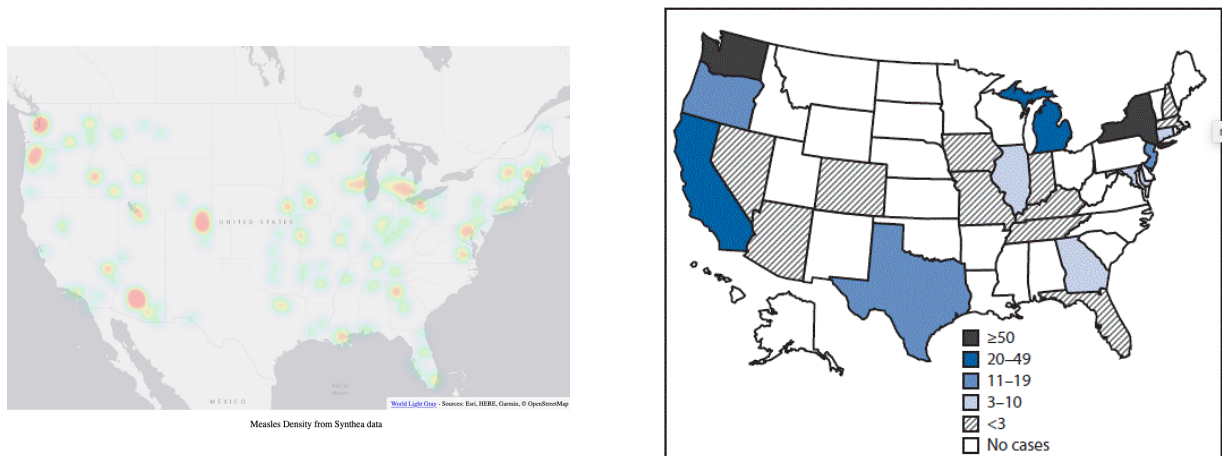
(a) Synthea-generated measles density across US.

(b) CDC map of 2019 US measles density

Figure 4: Comparison of 2019 US measles outbreaks with Synthea output.

Varying the lookup tables naturally produced different results. Lower probabilities, including a lower default NME probability in the measles module (Figure 1), combined with larger sets of generated records tended to produce results that performed better in our qualitative and quantitative validation checks. This result is not surprising; NMEs and measles are low-density events. To study them in that context one needs a large statistical population. In most of our test cases, we generated 1000-5000 records per state. In order to obtain useful rates of NMEs and measles infections from these relatively small sets of records, we set probabilities that were higher than observed rates in the US. While these probabilities had the intended effect of boosting the number of measles cases, they had the unintended effect of introducing "herd immunity," which ended up inhibiting our ability to model subsequent infections.

4   Implementation Details

We cloned the Synthea source code from github [11], made the changes described in the preceding sections, and recompiled into a "jar" with dependencies to execute. We found that, although we could drop in the measles module and measles-infection submodule, we could not get Synthea to load lookup tables from a local path at runtime. Therefore, we put our prototype modules and lookup tables in the source code so that the `synthea-with-dependencies.jar` contained all of our work. The code, along with documentation of the steps we followed is on github at `https://github.com/teterholdings/Synthea_NME`.

5 Utility

The very concept of accounting for the patient's preferences and health objectives is fundamental to PCOR [9]. In its current state, Synthea does not account for these preferences, which can result in patient records that are not representative of the population. By capturing one such patient preference, childhood MMR vaccine NMEs, we have provided a method for incorporation of more diversity in synthetic data generation that will enable researchers to more directly investigate and optimize patient-centered outcomes.

The synthetic patient records generated by our prototype include non-medical MMR exemptions for some patients and, consistent with recent findings (see, e.g., [5]), reflect a corresponding increase in childhood measles in the US. These modifications to Synthea can be useful in analyzing the impacts of non-medical exemptions to vaccines. Our current, prototype implementation includes workflows in which patients are hospitalized, or die from measles. This prototype can be used immediately to compare the costs, in terms of hospitalizations and deaths, of different NME scenarios in the US.

The outputs of the simulation under different scenarios would be valuable in informing public policy decisions for NMEs, as well as for educating the public. It would be very useful, for example, to understand and to be able to compare the costs of NMEs for different vaccines, in order to make informed decisions on which vaccines might be "negotiable," and which are not. More detail in the infection submodule, e.g., inclusion of additional long term effects of childhood measles, would increase the utility of these NME cost analyses. Finally, this synthetic data would be very useful in projecting medical resource and care requirements, especially with respect to pediatrics and complex care, in the most likely future scenarios.

References

[1] Anita Chabria. Anti-vaccine and alt-right groups team up to stoke fears of COVID-19 vaccine. *Los Angeles Times*, Dec 18, 2020. URL `https://www.latimes.com/california/story/2020-12-18/anti-vaxxers-team-up-alt-right-against-covid-19-vaccine`.

[2] Stephanie Garcia, Crystal Kallem, and Casey Thompson. Synthetic health data challenge phase I informational webinar. Webinar slides, February 2021. URL `https://www.challenge.gov/assets/netlify-uploads/20210202_challenge-phase-i-informational-webinar-materials.pdf`.

[3] Azhar Hussain, Syed Ali, Madiha Ahmed, and Sheharyar Hussain. The anti-vaccination movement: A regression in modern medicine. *Cureus*, 10(7), Jul 3, 2018. doi: 10.7759/cureus.2919. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6122668/`.

[4] National Center for Immunization and Respiratory Diseases (NCIRD). Vaccination coverage and exemptions among kindergartners (data), 2021. URL `https://data.cdc.gov/Vaccinations/Vaccination-Coverage-and-Exemptions-among-Kinderga/ijqb-a7ye`.

[5] Avril Nimblett-Clarke. The role of vaccine exemptions in the resurgence of measles. *JAAPA*, 34(2):36–40, February 2021.

[6] Jacqueline K. Olive, Peter J. Hotez, Ashish Damania, and Melissa S. Nolan. The state of the antivaccine movement in the United States: A focused examination of nonmedical exemptions in states and counties. *PLOS Medicine*, 15(6):1–10, 06 2018. doi: 10.1371/journal.pmed.1002578. URL `https://doi.org/10.1371/journal.pmed.1002578`.

[7] Manisha Patel, Adria D. Lee, Susan B. Redd, Nakia S. Clemmons, Rebecca J. McNall, Amanda C. Cohn, and Paul A. Gastañaduy. Increase in measles cases — united states, january 1–april 26, 2019. *MMMR Morbity Weekly Report*, 68:402–404, 2019.

[8] Manisha Patel, Adria D. Lee, Susan B. Redd, Nakia S. Clemmons, Rebecca J. McNall, Amanda C. Cohn, and Paul A. Gastañaduy. Increase in measles cases — United States, January 1–April 26, 2019. *Morbididty and Mortality Weekly Report (MMWR)*, 68(17): 402–404, 2019. doi: 10.15585/mmwr.mm6817e1. URL `https://www.cdc.gov/mmwr/volumes/68/wr/mm6817e1.htm`.

[9] Patient-Centered Outcomes Research Institute. Patient-centered outcomes research. URL `https://www.pcori.org/research-results/about-our-research/patient-centered-outcomes-research`. Accessed February 2021.

[10] Ranee Seither, Caitlin Loretan, Kendra Driver, Jenelle L. Mellerson, Cynthia L. Knighton, and Carla L. Black. Vaccination coverage with selected vaccines and exemption rates among children in kindergarten — United States, 2018–19 school year. *Morbididty and Mortality Weekly Report (MMWR)*, 68(41):905–912, 2019. doi: 10.15585/mmwr.mm684e1. URL `https://www.cdc.gov/mmwr/volumes/68/wr/mm6841e1.htm`.

[11] Synthea contributors. Synthea source code repository. URL `https://github.com/synthetichealth/synthea`. Accessed February 2021.

[12] Synthea Developers. Synthea immunizations java module, 2021. URL `https://github.com/synthetichealth/synthea/src/main/java/org/mitre/synthea/modules/Immunizations.java`. Accessed July 2021.

[13] Synthea Developers. Synthea module builder, 2021. URL `https://synthetichealth.github.io/module-builder/#measles`. Accessed July 2021.

[14] Jason Walonoski, Mark Kramer, Joseph Nichols, Andre Quina, Chris Moesel, Dylan Hall, Carlton Duffett, Kudakwashe Dube, Thomas Gallagher, and Scott McLachlan. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association*, 25(3):230–238, 08 2017. ISSN 1527-974X. doi: 10.1093/jamia/ocx079. URL `https://doi.org/10.1093/jamia/ocx079`.