UHASSELT

KNOWLEDGE IN ACTION

Faculteit Geneeskunde en Levenswetenschappen

master in systeem- en procesinnovatie in de gezondheidszorg

Masterthesis

A Data-Driven Approach to Develop a Multiple Sclerosis Disease Trajectory using Modelling Techniques for Synthetic Data

Nadia Rabah

Scriptie ingediend tot het behalen van de graad van master in systeem- en procesinnovatie in de gezondheidszorg

PROMOTOR :

Prof. dr. ir. Liesbet PEETERS

2024
2025

UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

UHASSELT

KNOWLEDGE IN ACTION

Faculteit Geneeskunde en Levenswetenschappen

master in systeem- en procesinnovatie in de gezondheidszorg

Masterthesis

A Data-Driven Approach to Develop a Multiple Sclerosis Disease Trajectory using Modelling Techniques for Synthetic Data

Nadia Rabah

Scriptie ingediend tot het behalen van de graad van master in systeem- en procesinnovatie in de gezondheidszorg

PROMOTOR :

Prof. dr. ir. Liesbet PEETERS

Table of contents

List of tables

List of figures

Preface

"Just keep swimming"

- *Dory, Finding Nemo*

A few years ago, I found myself uncertain of which path to follow, both academically and professionally. It was during this period of reflection that I came across this master and quickly realized how well it matched my passion and drive for innovation. For the first time in a long while, I felt a clear sense of purpose: the drive to improve the healthcare landscape, inspired by the perspective of someone who has experienced the system not just professionally, but also personally.

This master's thesis marks not only the culmination of my academic journey but also the result of a personal journey, marked by determination and resilience. The journey to completing this degree has been anything but easy as it has challenged me in many ways. Looking back, I'm extremely proud of how far I've come. This thesis is a small step in a much bigger journey, which will hopefully continue in the direction of a more innovative and inclusive healthcare system.

I would like to sincerely thank my promotor, Prof. Dr. ir. Liesbet Peeters, and my mentor, Marcel Parciak, for their guidance, feedback, and support throughout this research. A special thanks to Noorderhart Revalidation & MS Center for providing their real-world data, making it possible to realize my research.

Most deeply I would like to thank my husband, for believing in me, supporting me and most of all for his endless patience.

This thesis is the result of many contributions, and I'm deeply grateful to everyone who helped me keep swimming.

Nadia Rabah

Abstract

Background

Multiple sclerosis (MS) is a complex, chronic autoimmune disease which affects over 2.8 million people worldwide. The disease course is highly heterogeneous and complex, highlighting the need for personalized treatment and data-driven decision making by neurologists. Therefore, collecting health data is thus essential to better understand the disease progression and to optimize and improve patient care. As a possible solution, synthetic health data could pose a promising alternative to clinical trials and real-world data by simulating realistic but not real patient trajectories without compromising privacy. Currently, multiple models for different chronic conditions exist but no model has been built to this day for the disease MS. This thesis addresses that gap by developing an MS disease trajectory with Synthea, an open-source synthetic patient generator.

Methodology

To model the disease trajectory of MS, the open-source synthetic patient generator, Synthea, was used. A data-driven approach was implemented by using two key-inputs: (1) evidence-based literature and (2) anonymized patient data from Noorderhart Revalidation & MS Center (Belgium).

Results

The results are divided into 2 different outcomes. First is the focus on the data-driven approach, elaborating on the sources and rationale behind the data used to design the model. The second outlines the actual model, detailing the various states used and the disease specific pathways.

Discussion

The disease trajectory for MS that was developed was built using a strong evidence-based foundation and furthermore successfully generates synthetic patients. Key elements such as EDSS scores, disease modifying therapies and progression pathways were integrated to enhance the patient journey to reality. For future research, raising awareness on the quality and quantity of real-world data are crucial to expand the potential of the model. Lastly, validation and expansion or adjustment to the model are advisable in collaboration with MS experts.

Introduction

In 2020, an estimated 2.8 million people worldwide were living with multiple sclerosis (MS)—a 30% increase compared to 2013 (1). MS is a chronic autoimmune disease, characterized by inflammation that damages the protective myelin sheath around the nerves in the brain and spinal cord. As a result, nerve signal transmission can be slowed or blocked, leading to various neurological symptoms such as fatigue, optic neuritis, cognitive dysfunction, sensory and motor impairments which can possibly result in reduced quality of life and disability (2, 3).

The onset of the first neurological symptom typically occurs between the ages 20 to 40 years (3). The pathophysiology of MS is currently unclear, however some risk factors which include genetic factors but also low vitamin D levels, the Epstein-Barr virus infection and exposure to cigarette smoke have been identified (3). Multiple sclerosis presents itself in various types, and is defined in the following subtypes: clinically isolated syndrome(CIS), relapsing-remitting MS(RRMS), secondary progressive MS (SPMS), primary progressive MS (PPMS) and progressive-relapsing MS (PRMS). The latter was removed from the standardized classification as there was too much overlap with other subtypes (4).

The complexity and highly heterogeneous nature of this disease makes it extremely challenging to manage. While treatment falls into two categories, symptomatic relief and disease-modifying therapies, the unpredictable course of the disease demands for a personalized approach. These personalized approaches require detailed data in order to allow neurologists to make informed decisions related to the diagnosis or treatment. Therefore, access to data and data-driven decision making is crucial to achieve the goals of improving patient care (2, 5, 6).

This highlights a critical gap in MS research: the lack of accessible and cost-effective patient data for data-driven decision making. Clinical trials, in particular randomized clinical trials (RCT) have limited generalizability and high operational costs (7-10). Due to these constraints there has been a growing interest for alternative data sources, such as real-world data (RWD). According to the U.S. Food and Drug Administration(FDA), "Real-world data are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources" (11).

This data is particularly important for MS, to accelerate research and improve treatment outcomes (12-14). RWD has advantages as it contains a broader patient population in comparison to traditional clinical trials, but comes with challenges such as data quality concerns, privacy risks and legal or ethical barriers to access (7, 13, 15).

To address some of these challenges associated with clinical trials and RWD, synthetic data offers a valuable complementary resource by generating a purpose-built mathematical model that replicates the complexity of real health data (16). By training this model, synthetic data is generated that mirrors the key characteristics of the original dataset (17). Synthetic data can simulate realistic patient journeys that ultimately could accelerate data science initiatives and possibly bridging the gaps in data access in research (16, 18). One of the tools used for generating synthetic data is Synthea, an open-source patient generator that produces the lifespan of synthetic patients. Currently there are multiple synthetic models available for chronic conditions such as lung cancer or Alzheimer's disease (15). These disease-specific models allow researchers to explore data and stimulate the development of early ideas and solutions before gaining actual restricted datasets (19).

Despite the potential of synthetic data for chronic conditions, unfortunately there currently is no open-source model that captures the complexity and disease trajectory of multiple sclerosis within Synthea. Therefore, this thesis will aim to address that gap by developing a realistic disease trajectory for MS using Synthea.

The goal of this thesis is to answer the following research questions below:

- What key elements define the disease trajectory of MS in a way that reflects a realistic patient journey?
- How can the progression of MS be modelled and implemented in an open-source tool for generating synthetic health data?

Method

An open-source synthetic patient generator, called Synthea, was selected as a framework to model the medical history of synthetic patients and thus the disease trajectory of an MS patient. The aim of this tool is to provide synthetic data, meaning data that resembles realistic but is not actual real patient data. This results in data that is free of cost, privacy and security restrictions which enables research in different sectors such as academic research where otherwise data is often unavailable, due to legal or practical issues (20).

The foundation of the disease trajectory model was developed iteratively, using two primary sources which are evidence-based and data-driven. First, source data from the extensive literature search was used to define the disease trajectory. Secondly, data from Noorderhart Revalidation Center was used to incorporate the probability pathways into the model.

With these two methods, crucial elements of the disease trajectory were identified. To define the trajectory and to ensure its reliability, the patient's diagnosis trajectory was included which are based on the 2017 McDonald criteria which contains a set of guidelines to help provide an accurate diagnosis (21).

As not all patients receive their MS diagnosis immediately, the diagnosis of clinically isolated syndrome (CIS) was included. This term describes the first onset of potential multiple sclerosis and typically applies to young adults (22). The common disease modifying therapies were also added to the module, focusing on the five most used therapies.

Additionally, the Expanded Disability Status Scale(EDSS) scores were integrated, as this is a way of measuring how much impact MS has on the patient. This score ranges from 0 to 10, with the higher the score the higher the effect of disability on the person. This score is very important as neurologists use it to monitor changes over time in the level of disability in a patient (23). Lastly, the relevant comorbidities were included that are associated with this disease.

Construction Disease Trajectory within Synthea

The following simplified example of childhood ear infections shows the flow of a generic module within Synthea from diagnosis at an encounter with their pediatrician, until the probabilistic pathway for treatment with either an antibiotic or a painkiller (24).

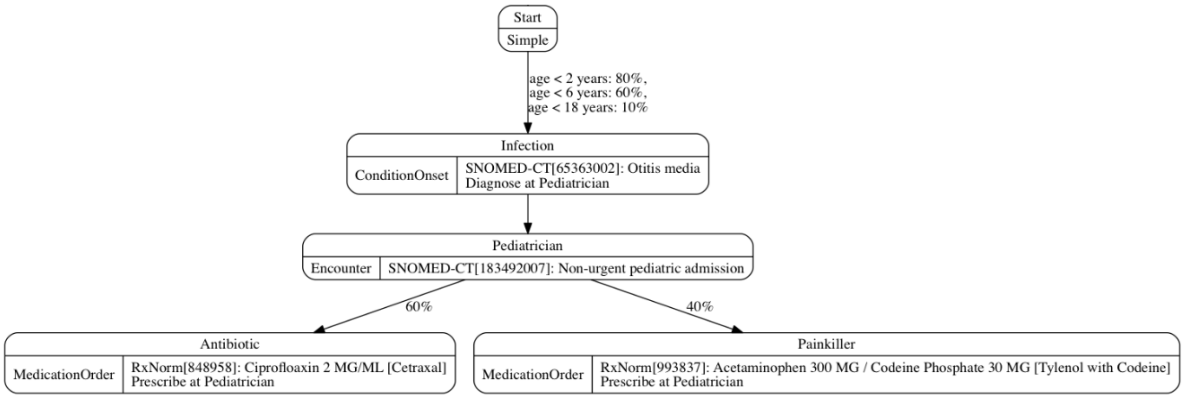

Figure 1: Example of a simplified Synthea disease trajectory (15)

While building the module, it was taken in account that a new module must minimally have a name and a set of states. Synthea currently supports two different categories of states. The first are the seven different control states such as the Initial or Terminal state. An initial state would be where the module starts and the Terminal state is where the module ends. The second category are the clinical states, of which Synthea supports 11 different types. These clinical states can contain encounters, which in the example above in Figure 1 is the encounter with the pediatrician. Another example of a clinical state is the onset of a condition, which in this example is the otitis media. Furthermore, they are incorporated with medical terminology codes such as SNOMED-CT or RxNorm codes (15).

In this thesis, the Synthea module builder, a web interface, will be used to model, implement and define the disease trajectory of MS, ultimately resulting in a simulation model.

To standardize the medical terms such as symptoms or diagnoses SNOMED-CT were used (25). SNOMED-CT is a multinational standard, and has been accepted as a common global language for health conditions. The use of this standard in the model was crucial as this guarantees a standardized vocabulary for clinical terms used in the model (26). The use of RxNorm was done to have a system where the names of all types of medications are normalized, including generic and branded drugs. This increases the interoperability in different health IT systems which is crucial when a patient moves from one hospital or organization to another (27). Lastly, the same standardized language was applied for laboratory tests and clinical observations, using the Logical Observation Identifiers Names and Codes (LOINC) (28). An overview of the standardized terminologies and their descriptions can be found in Appendix 1.

Logical Observation Identifiers Names and Codes (LOINC) is a universal standard for identifying laboratory tests and clinical observations. It facilitates a smooth information exchange between hospitals, locally and internationally.

Lastly, to define the parameters and probabilities, anonymized patient data provided by Noorderhart Revalidation & MS center, located in Pelt, Belgium was used. This dataset was used to extract key percentages to integrate the probability pathways and subsequently construct the synthetic data.

Model availability

The model file itself is publicly available for download from the GitHub repository at:

https://github.com/NadiaRabah/MS-Disease-Trajectory-Synthea

The developed MS disease trajectory can be uploaded and accessed through the Synthea module builder interface at:

https://syntheticahealth.github.io/module-builder

Results

The results are presented in two different sections. The 'Source Data' section focuses on the data-driven approach, elaborating on the sources and rationale behind the data used to design the model. The second section 'Synthea Model Prototype' outlines the actual model, detailing the various states used and the disease specific pathways.

Source Data

For this section, data from literature and data from Noorderhart was used to implement the probabilistic pathways.

Initial Symptoms

The data starts with the presentation of the distribution of the initial symptoms experienced before the MS diagnosis, using a pre-onset delay of 20 to 40 years (3).

In Table 1, the majority (53,96%) of patients presents with symptoms classified as 'Other Initial Symptoms', followed by spinal cord (18,71%) and optic pathway (17,27%) symptoms. Less frequent are the supratentorial and brainstem-cerebellum symptoms (5,04%).

Table 1: Overview of initial symptoms

DESCRIPTION	n	%
Supratentorial	7	5,04%
Optic Pathways	24	17,27%
Brainstem-Cerebellum	7	5,04%
Spinal cord	26	18,71%
Other Initial Symptoms	75	53,96%
Total	**139**	**100%**

Comorbidities

Regardless of their initial symptom, all patients follow a linear pathway of the model until they reach the official MS diagnosis. From there on, the comorbidities have been assigned with a probabilistic pathway which was calculated with the following data in Table 2. This tableTable 2 also shows the calculated percentages for each comorbidity.

Table 2: Number of comorbidities in MS patients

DESCRIPTION	n	%
Surgery	31	3,50%
Cystitis	29	3,28%
Depression	29	3,28%
Infection	26	2,94%
Urologic disorder	25	2,82%
Migraine	18	2,03%
Hypertension	17	1,92%
Allergy	17	1,92%
Hypercholesterolemia	16	1,81%
Orthopedic Disorder	15	1,69%
Diabetes	10	1,13%
Cardiovascular Disorder	10	1,13%
Nicotine Abuse	9	1,02%
Epilepsy	7	0,79%
Cancer	3	0,34%
Herpes Zoster	1	0,11%
Other	8	0,90%
No comorbidities	614	69,38%
Total	**885**	**100%**

MS Subtypes

The following source data outline the MS subtypes states, which branch off between CIS and PPMS initially. The Sankey diagram in Figure 2 below shows the progression of persons with MS (PwMS), as found in the Noorderhart data.

Initially, 320 patients (36.16%) were assigned an MS subtype, while 565 (63.84%) remained unassigned. Among the assigned patients, 303 patients (94.69%) had clinically isolated syndrome (CIS), of which 298 (98.35%) transitioned further to a different MS subtype, 296 (99,33%) to RRMS and 2 (0,67%) to SPMS.

From RRMS, 245 (82.77%) retained the RRMS classification as their final recorded diagnosis and 51 (17.23%) progressed to secondary progressive MS (SPMS) joining the two patients directly transitioning from CIS, bringing the total SPMS to 53 patients. All of these patients remained in SPMS as their final diagnosis.

Of the remaining assigned patients, 17 (5.31%) had primary progressive MS (PPMS), with 10 (58.82%) retaining this diagnosis. Lastly, 7 patients (2.19%) were diagnosed with progressive relapsing MS (PRMS), with all maintaining this subtype at their final recorded diagnosis.

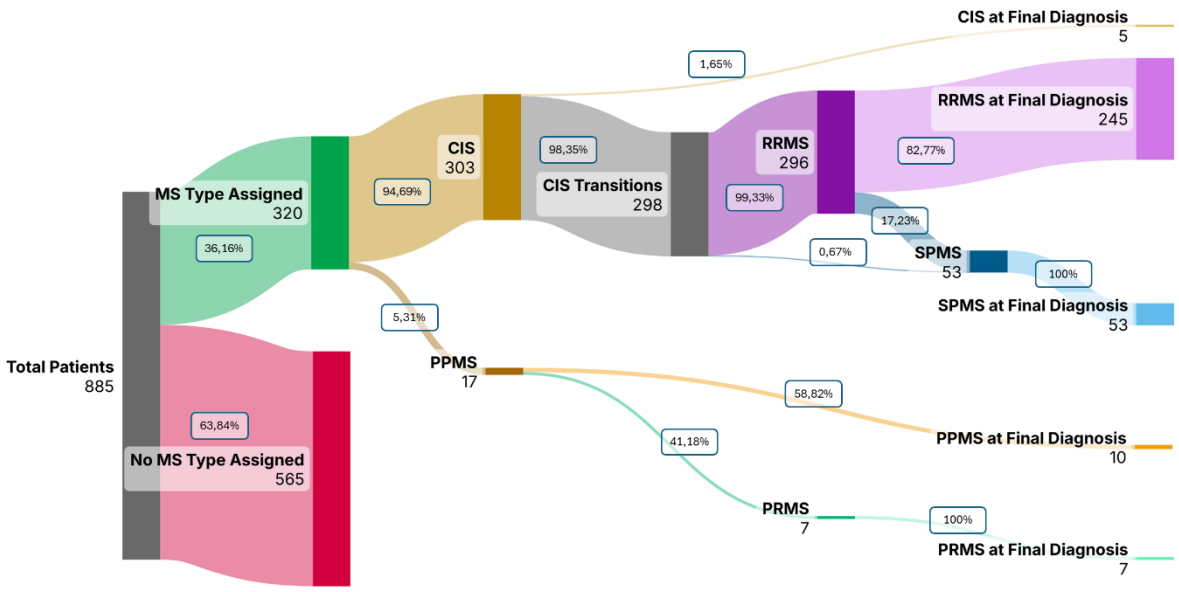

Figure 2: Sankey diagram overview of MS subtypes distribution and progression

Initial Expanded Disability Status Scale score per MS subtype

The baseline or initial EDSS scores were analyzed across the different multiple sclerosis subtypes. As shown in Table 3, the mean initial EDSS scores had a big variation between these subtypes. Patients diagnosed with SPMS exhibited the highest mean EDSS score (M =5.35, SD = 1.44), followed by those with PPMS (M = 4.69, SD = 2.05) and progressive-relapsing MS (M = 4.60, SD = 1.64). In contrast, patients with clinically isolated syndrome (CIS) had the lowest mean EDSS (M = 2.06, SD = 1.15), while those with relapsing-remitting MS (RRMS) showed a slightly higher mean (M = 2.70, SD = 1.53).

Approximately half of the population (n = 180) lacked an MS subtype classification and were thus grouped under the undefined group and were not used in the definition of the disease trajectory.

Table 3: EDSS scores per MS subtype

MS Subtype	n	Mean	Standard deviation
CIS	16	2.06	1.15
PPMS	8	4.69	2.05
PRMS	5	4.60	1.64
RRMS	110	2.70	1.53
SPMS	37	5.35	1.44
Undefined	180	3.71	2.19

Disease Modifying Therapies across MS Subtypes

Table 4 presents an overview of the distribution of disease modifying therapies (DMT) among patients with different MS subtypes. The absence of certain treatments (indicated with '—') does not imply that these DMT's cannot be used for the respective MS subtype. It reflects that these treatments were not among the five most frequently reported in the dataset and therefore were not included in the table.

Table 4: Distribution of the five most frequently reported DMT's per MS subtype

DMT	CIS	RRMS	SPMS	PPMS	PRMS
No DMT	290 (96,35%)	24 (16,90%)	—	6 (42,86%)	—
Aubagio	1 (0,33%)	—	—	—	—
Avonex	3 (1,00%)	28 (19,72%)	—	—	—
Betaferon	6 (1,99%)	31 (21,83%)	—	2 (14,29%)	—
Copaxone	—	29 (20,42%)	3 (7,32%)	—	—
Endoxan	—	—	18 (43,90%)	3 (21,43%)	3 (42,86%)
Gilenya	—	30 (21,13%)	2 (4,88%)	—	—
Lemtrada	—	—	4 (9,76%)	—	2 (28,57%)
Novantrone	—	—	14 (34,15%)	—	—
Ocrevus	—	—	—	2 (14,29%)	1 (14,29%)
Rebif 22mcg	1 (0,33%)	—	—	—	—
Rebif 44mcg	—	—	—	1 (7,14%)	1 (14,29%)
Total patients	301 (100%)	142 (100%)	41 (100%)	14 (100%)	7 (100%)

Among patients with CIS, the vast majority were not receiving any DMT (96,35%). A small proportion of patients were receiving disease-modifying treatments, including Betaferon (1.99%, n = 6), Avonex (1.00%, n = 3), Rebif 22mcg (0.33%, n = 1), and Aubagio (0.33%, n = 1).

Among the patients diagnosed with RRMS, the data showed a diverse use of DMT's. As shown in Table 4, the most frequently prescribed was Betaferon, administered to 21.83% of patients (n = 31), closely followed by Gilenya (21.13%, n = 30), Copaxone (20.42%, n = 29), and Avonex (19.72%, n = 28). A smaller proportion of RRMS patients (16.90%, n = 24) were not receiving any DMT at the time of data collection.

Among the 41 patients with SPMS the treatment data collection revealed that Endoxan was the most common (43.90%, n=18), followed by Novantrone (34.15%, n=14). Lemtrada (9.76%, n=4), Copaxone (7.32%, n=3), and Gilenya (4.88%, n=2) were less frequently prescribed.

The treatment distribution among 14 patients with PPMS showed that the majority (42.86%, n=6) received no DMT. Among treated patients, Endoxan was the most frequently prescribed (21.43%, n=3), followed by Ocrevus and Betaferon (14.29% each, n=2). Rebif (44 mcg) was the least common treatment (7.14%, n=1).

From the seven patients with PRMS, Endoxan was the most frequently prescribed treatment (42.86%, n=3), followed by Lemtrada (28.57%, n=2). Rebif with the dosage of 44mcg and Ocrevus were each used in one patient (14.29% each, n=1).

Synthea Model

The model initiates the patient's journey from the onset of initial symptoms to the formal diagnosis of MS, including the potential presence of comorbidities during this period. This part of the model was based on the 2017 McDonald criteria which is a set of guidelines to help provide an accurate diagnosis of MS (21).

The following part of the model focuses on the determination of the MS subtype (CIS, RRMS, PPMS, SPMS and PRMS), the initial Expanded Disability Status Scale (EDSS) scores, the different treatment pathways and the disease course – including cycles of relapses, maintenance, and recovery or progression in case of progressive disease forms. Treatments or visits to hospitals were implemented with a one day duration.

Initial to Diagnosis

The following section will explain each chosen state type the model was built on as depicted in Figure 3 below.

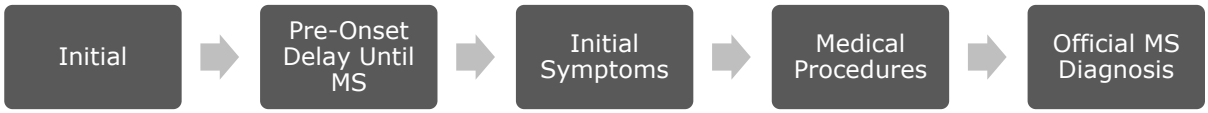

Figure 3: Patient journey from initial to official MS diagnosis

The model was constructed with an *Initial* state type and then proceeded to the *Pre-Onset Delay* state. This delay state type depicts a time interval between 20 to 40 years that reflects the potential age, which patients may experience their first MS-related symptom (3).

Following this, the model transitions to a simple state titled *Initial Symptoms* and branches off into the five potential initial symptoms a patient could experience before their MS diagnosis. These symptoms included the optic pathways, supratentorial regions, brainstem and cerebellum region, spinal cord and finally the category for other initial symptoms. Each of these symptoms were modeled using the symptom state type where each symptom was assigned with a predefined probability. The details of these probabilities were explained previously in the 'Source Data' section of this paper.

Regardless of the initial symptom pathway, all simulated patients in the model converge to the encounter state type *Neurology Consultation*, depicted with the appropriate SNOMED-CT terminology. Following this neurology consultation, a *MedicationOrder* state was incorporated to stimulate the model to initiate treatment. Within this model, a prescription for methylprednisolone was added to treat the acute MS symptoms and coded with a RxNorm identifier. To reflect clinical practice, this step was added to represent treatment with oral or intravenous methylprednisolone for 3-5 days, in the linear path, supporting faster recovery from relapses and initial symptoms (29).

Subsequently the encounter state prescribing Magnetic Resonance Imaging (MRI) was included into the model with the appropriate SNOMED-CT code, followed by a procedure state in regards to initiate a cerebrospinal fluid (CSF) collection, which is used for diagnostic purposes as documented into the McDonald criteria. After this procedure, an observation state evaluates the presence of oligoclonal bands in the CSF, using LOINC coding to reflect laboratory findings that may be relevant to the final

MS diagnosis (21). The encounter for all simulated patients was formally ended through an encounter end state of the consultation.

Once the simulated patient leaves the consultation, the treatment was started as an ambulatory treatment within an encounter state type (30). The delay has a time frame of 3-5 days, reflecting the duration of this treatment (29, 31).

After this delay, a procedure state was added into the model to simulate an MRI of either the spinal cord or the brain with a duration of 1 day and attributed SNOMED-CT code (32). This is followed by an observation state to assess the Gadolinium(Gd) enhanced lesions and non-enhancing lesions in the imaging to identify the diagnostic criteria (21).

From there on the model transitions into the delay time for diagnosis which was defined at a median time of 5.7 months after the initial consultation and in between the diagnostic observations and results in the confirmation of the diagnosis (33). This transitions from there on to the encounter state which symbolizes a new neurology consultation where the official MS diagnosis is reported.

In order to make the simulation as realistic as possible, a simple state type titled *Comorbidities* was designed into the model which branches off into the different comorbidities a patient can have. This was incorporated with multiple condition onset states, representing a different comorbidity for each state (e.g., cardiovascular disease, diabetes, epilepsy,…) and is assigned a probability each. These probabilities were discussed in the section 'Source Data'.

MS Subtype and Disease Modifying Therapy

Following the final diagnosis of multiple sclerosis (MS), the model transitions and branches off to determine the clinical subtype of the disease which is depicted in Figure 4.

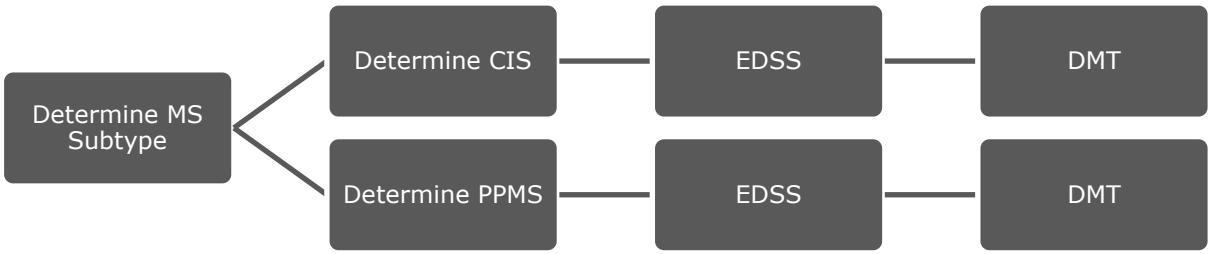

Figure 4: MS subtypes CIS and PPMS, and subsequential DMT

The two states that were initially incorporated into the model were the *ConditionOnset* state type for CIS and PPMS. Once the MS subtype has been determined, the initial EDSS for each subtype was assigned with a probability as well as the different DMT options.

Determine Clinically Isolated Syndrome (CIS)

For the simulated CIS patients, a *ConditionOnset* state is triggered with an assigned SNOMED-CT code that corresponds to it. The initial EDSS score was also incorporated into the model under an *Observation* state. Following this subtype assignment, the different treatment pathways were modeled via *MedicationOrder* states. These different treatment pathways include DMT's such as Betaferon, Avonex or no treatment at all. Finally, the consultation ended with an *EncounterEnd* state.

From the simple state titled *CIS* two different pathways were modeled, to either another *Neurology Consultation Encounter* state when the patient transitions to a different MS subtype or to the *Terminal* state when there is no transition to a different MS subtype.

Transition CIS-RRMS

At the end of determining CIS, an encounter state type related to a neurology consultation was modeled to allow transition from CIS to RRMS. When a simulated patient possibly transitions to RRMS within the model, the simulation will follow the top to bottom pathway where another initial EDSS score is assessed under the *Observation* state (Figure 5). From there on, the different treatment

pathways, as documented in the 'Source Data' section, were also modeled via *MedicationOrder* states. Finally, the consultation ended with an *EncounterEnd* state.

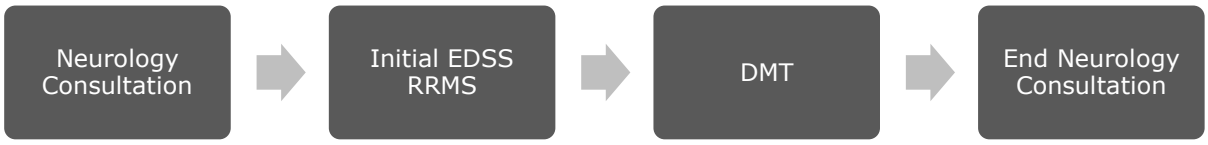

Figure 5: Neurology consultation course RRMS

Once the patient has been assigned to the simple RRMS state, there are 3 different pathways (Figure 6) a simulated patient could go through. The first would be another neurology consultation with an *Encounter* state, as the patient could transition to yet another MS subtype, this time SPMS. In the second pathway, the patient experiences a relapse with a delay of <1 month, goes through recovery with a delay of 1 year and then goes back to the simple RRMS state (34, 35).

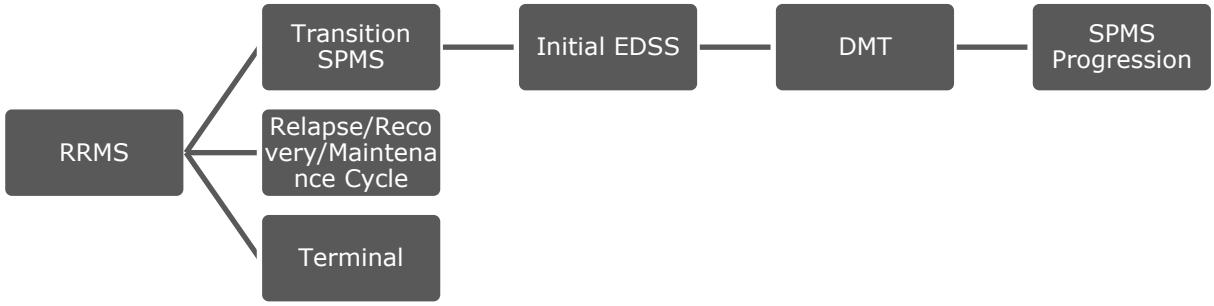

Figure 6: RRMS pathways

When the simulated patient experiences no relapse, the model was designed as such that the pathway of maintenance was incorporated. In this cycle, the patient would go from maintenance with a delay of one year back to the simple state of RRMS.

The third and final pathway from the RRMS simple state is the *Terminal* state.

Transition CIS/RRMS to SPMS

The model was designed as such that both pathways from CIS and RRMS can transition to SPMS. This was designed with a *ConditionOnset* state titled Transition SPMS and was defined with a SNOMED-CT code. This pathway was also designed from top to bottom, where the initial EDSS score was assessed under the *Observation* state. Subsequently, the different treatment pathways were modeled through the *MedicationOrder* states. These different pathways include DMT's as Endoxan, Novantrone, Lemtrada, Copaxone and Gilenya as documented in Table 4 in the section 'Source Data'. As a final pathway the end of the neurology consultation was modeled through an *EncounterEnd* state.

In order to reflect the progression of SPMS a cycle was added that goes from SPMS progression with a delay of one year, back to the simple state of having SPMS. Once the simulated patient has gone through all these different pathways, the terminal state has been reached.

Determine PPMS

The pathway for PPMS was modeled with first the *ConditionOnset* state and linked with the appropriate SNOMED-CT code. Followingly, the initial EDSS score was implemented in the model under an *Observation* state. From there on, the different treatment pathways for PPMS were modeled through the *MedicationOrder* state. These different pathways include treatments such as Ocrevus, Betaferon, Refib, Endoxan and no treatment, all linked to a RxNorm code. As a final pathway the end of the neurology consultation was modeled through an *EncounterEnd* state (Figure 7).

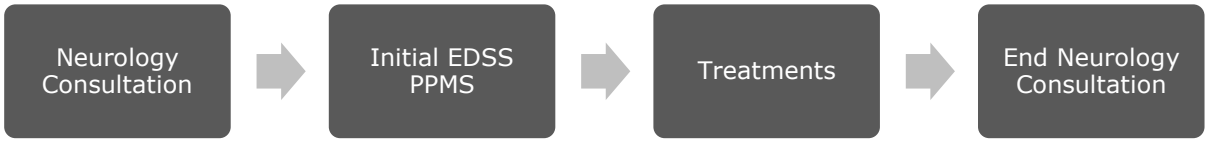

Figure 7: Neurology consultation course PPMS

Transition from PPMS to PRMS

From the simple state PPMS, the simulated patients can diverge into three different pathways (Figure 8).

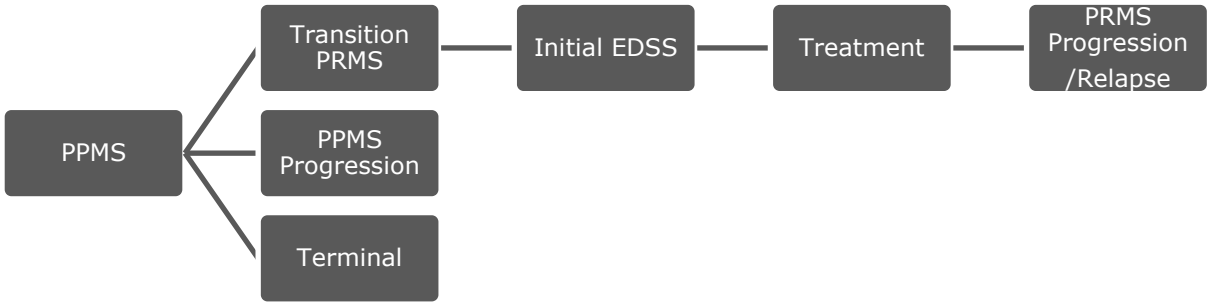

Figure 8: PPMS pathways

The first is where the model branches off to the terminal state, the second depicts the cycle of progression of PPMS. Finally, the model branches off to an *Encounter* state for a neurology consultation to discuss the transition from PPMS to PRMS. Logically, a new *ConditionOnset* state was designed for PRMS with the appropriate SNOMED-CT code.

Followingly, the initial EDSS score was implemented in the model under an *Observation* state. From there on, the different treatment pathways for PRMS were modeled through the *MedicationOrder* state. These different pathways include treatments such as Endoxan, Lemtrada, Rebif and Ocrevus, all linked to a RxNorm code. As a final pathway the end of the neurology consultation was modeled through an *EncounterEnd* state.

The patients then are simulated to a simple state for PRMS, followed by a progression/relapse cycle and finally transitioning to the terminal state.

Discussion

This thesis presents a model of the MS disease trajectory, developed using the Synthea open-source tool.

To define a realistic patient journey, key elements were drawn from scientific literature, patient data from Noorderhart and diagnostic criteria such as the McDonald criteria (21). Key elements such as the preclinical disease onset, diagnostic process, EDSS scores, DMT's, and transitions between MS subtypes were used. These elements address the first research question:

- What key elements define the disease trajectory of MS in a way that reflects a realistic patient journey?

To answer the second research question,

- How can the progression of MS be modelled and implemented in an open-source tool for generating synthetic health data?

the model was built in Synthea using disease specific states, probabilistic pathways and standardized medical coding systems (SNOMED-CT, RxNorm). Additionally, a data-driven approach was used to define disease progression within the model, ensuring that the transitions reflected real-world patterns observed in literature and the patient data provided by Noorderhart.

Strengths

A major strength of this thesis is its methodology, the model was developed using evidence-based data to emphasize its realism, by using not only existing literature but also real-world data provided by Noorderhart. Key factors such as the delays, diagnostic criteria and treatment options were incorporated based on literature and data. The diagnostic criteria were based on the 2017 McDonald guidelines for the diagnosis of MS (21). For the probabilistic pathways and the different treatment options, the EDSS scores but also the disease progression, the real-world data from Noorderhart was used. This was done to ensure that the model reflects the actual clinical outcomes such as comorbidities and the initial symptoms but also the treatment strategies. A key strength of the model is its flexibility, allowing researchers to adjust these parameters to fit specific needs.

One of the model's main strengths is that it follows the Synthea structure and criteria, enabling the generation of synthetic patient data which can be found in Appendix 2. Furthermore, a solid foundation has been made with this model, which could be adapted in the future as the individual components such as the states could be updated or expanded. With this adaptability, the model could be changed to the specific purposes of different research areas or other purposes.

Additionally, with the use of the different standardized medical coding systems such as SNOMED this model enhances the interoperability with different systems. As these vocabularies are internationally recognized, the synthetic data could become interoperable, in the sense that every researcher or clinician would be able to understand the medical terms used in a standardized manner. An additional advantage could be that researchers in institutions could use the model and the synthetic data without having to put time into adjustments in this medical terminology, allowing for efficient collaborative research (36).

Although the model is not intended for use in the clinical world, it could serve as a valuable resource for machine learning. In this way, research teams have access to realistic data and it enables exploration and testing before gaining access to sensitive or perhaps restricted patient datasets. This approach could lower the barriers that exist in the acquisition of real-world clinical data and could support health innovations (19).

Lastly, the model stimulates the generation of a diverse synthetic population by the incorporation of multiple different variables such as the initial symptoms, comorbidities, EDSS scores, disease progression and the different treatment pathways. This not only enhances the representation of various patient profiles but also the external validity of the synthetic data by capturing the variability and complexity of the real-world disease trajectory for MS patients. With such variability this also reinforces the need for an MS disease trajectory and thus the model that was developed.

Limitations

The primary limitation of this thesis was the absence of quantity and quality in the real-world patient data. For instance, the dataset lacked 662 patients who had no symptoms but sought out a neurologist for a possible diagnosis. This posed a challenge as it restricted the capability of the model to expand and required the use of external research papers for certain parameters. The same was applicable for the MS types where 2/3 of the data was also missing, totaling for 562 patients. As a result, these missing data variables were excluded from the calculation of the probabilistic pathway.

An additional technical limitation was the fact that the SNOMED-CT, RxNorm codes etc. needed to be completed before actual synthetic data could be generated. This was not clearly stated by Synthea and thus delayed the timelines in general, trying to search for the cause of this problem.

Although synthetic data can be a promising solution to address challenges caused by data gaps and privacy risks, it is noteworthy that synthetic data still has privacy risks, as the data implemented can be vulnerable for privacy attacks (16, 37).

Lastly, the model was also not entirely completed due to the lack of data for the relapse and progression states. These states were added to the model but the actual clinical data to incorporate the probabilities is missing. Future research should therefore prioritize the model completion by incorporating missing clinical data if possible and perhaps expanding the model working together with an MS expert. The model presented in this thesis can serve a basis, so the validation could be done in the future to ensure its accuracy.

Recommendations

Further research is necessary to validate the generated synthetic data from this model. To validate the model, the synthetic data should be compared to the data of Noorderhart. Metrics for this validation could be the MS subtype distribution, the DMT and disease progression pathways. A second step would be to validate it against data from public health registries.

By using expert's opinions, such as neurologists that are specialized in the treatment of MS, the model could be adjusted or expanded where needed to improve its completeness. Lastly, an essential element would be to improve the data quality in clinical settings by emphasizing the importance of real world data to clinicians. This includes the systematic and accurate documentation of patient data, as well as raising awareness about the importance and tremendous value for applications such as model developments such as these.

Conclusion

This research presents a model of the disease trajectory for MS, developed using the Synthea open-source tool. It captures the preclinical disease onset phase all the way through the diagnosis, disease progression and associated DMT paths. By integrating disease-specific states, probabilistic pathways, and standardized medical coding, the model was designed to realistically reflect the clinical complexity of MS and provide a solid foundation for the generation of synthetic health data. Due to the lack of accessible RWD and the possible privacy problems associated with it, synthetic data has the potential to accelerate future data-driven health innovations (16, 18).

Furthermore, one of the key strengths of the model lies in the evidence-based and data-driven construction. The model was mapped by either using existing literature but also the existing real-world patient data from Noorderhart. This was done to ensure that the model resembles the real world's journey of a patient as much as possible.

In summary, this model provides a solid foundation for the specific disease simulation of MS patients that can be expanded and adjusted for a range of different non-clinical applications, including algorithm testing and synthetic data research. This thesis also fills a critical gap by developing the first known Synthea model for multiple sclerosis. This lays a foundation for future development of MS models but also the generation of synthetic patient datasets for complex chronic conditions.

References

1. Walton C, King R, Rechtman L, Kaye W, Leray E, Marrie RA, et al. Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. Mult Scler. 2020;26(14):1816-21.

2. Marcus R. What Is Multiple Sclerosis? Jama. 2022;328(20):2078.

3. Usta NC, Boz C, Terzi M. Early onset multiple sclerosis and the effect of disease onset age on neurological disability in multiple sclerosis. Clin Neurol Neurosurg. 2023;224:107528.

4. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. Neurology. 2014;83(3):278-86.

5. Peeters LM, van Munster CE, Van Wijmeersch B, Bruyndonckx R, Lamers I, Hellings N, et al. Multidisciplinary data infrastructures in multiple sclerosis: Why they are needed and can be done! Mult Scler. 2019;25(4):500-9.

6. Giuffrè M, Shung DL. Harnessing the power of synthetic data in healthcare: innovation, application, and privacy. NPJ Digit Med. 2023;6(1):186.

7. Rogers JR, Lee J, Zhou Z, Cheung YK, Hripcsak G, Weng C. Contemporary use of real-world data for clinical trial conduct in the United States: a scoping review. J Am Med Inform Assoc. 2021;28(1):144-54.

8. Jenkins V, Farewell V, Farewell D, Darmanin J, Wagstaff J, Langridge C, et al. Drivers and barriers to patient participation in RCTs. Br J Cancer. 2013;108(7):1402-7.

9. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review. Contemp Clin Trials Commun. 2018;11:156-64.

10. McGinley MP, Manouchehrinia A. The landscape of multiple sclerosis registries: Strengths and limitations. Mult Scler. 2024;30(3):281-2.

11. Administration(FDA) USFaD. Real-World Evidence 2024 [Available from: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

12. Peeters LM, Parciak T, Kalra D, Moreau Y, Kasilingam E, van Galen P, et al. Multiple Sclerosis Data Alliance - A global multi-stakeholder collaboration to scale-up real world data research. Mult Scler Relat Disord. 2021;47:102634.

13. Schlegel DR, Ficheur G. Secondary Use of Patient Data: Review of the Literature Published in 2016. Yearb Med Inform. 2017;26(1):68-71.

14. Trinh QD. Understanding the impact and challenges of secondary data analysis. Urol Oncol. 2018;36(4):163-4.

15. Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, et al. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J Am Med Inform Assoc. 2018;25(3):230-8.

16. Adrian JJeaW. Synthetic Data - what, why and how? ; 2022.

17. Riemann R. Synthetic Data European Data Protection Supervisor [Available from: https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en#:~:text=Synthetic%20data%20is%20artificial%20data,undergoing%20the%20same%20statistical%20analysis.

18. Gonzales A, Guruswamy G, Smith SR. Synthetic data in health care: A narrative review. PLOS Digit Health. 2023;2(1):e0000082.

19. Walonoski J, Klaus S, Granger E, Hall D, Gregorowicz A, Neyarapally G, et al. Synthea™ Novel coronavirus (COVID-19) model and synthetic data set. Intell Based Med. 2020;1:100007.

20. Corporation TM. Synthetic Patient Generation 2025 [Available from: https://syntheticheath.github.io/synthea/.

21. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. 2018;17(2):162-73.

22. Efendi H. Clinically Isolated Syndromes: Clinical Characteristics, Differential Diagnosis, and Management. Noro Psikiyatr Ars. 2015;52(Suppl 1):S1-s11.

23. UK) MSSMS. Expanded Disability Status Scale (EDSS) [Available from: https://www.mssociety.org.uk/living-with-ms/treatments-and-therapies/getting-treatment-for-ms/expanded-disability-status-scale#:~:text=The%20Expanded%20Disability%20Status%20Scale%20(EDSS)%20is%20a%20way%20of,3.5%2C%204%20and%20so%20on.

24. Walonoski J. Generic Module Framework Aug 1, 2022 [Available from: https://github.com/syntheticheath/synthea/wiki/Generic-Module-Framework.

25. Walonoski J. Generic Module Framework: Basics May 28, 2020 [Available from: https://github.com/syntheticheath/synthea/wiki/Generic-Module-Framework%3A-Basics.

26. Medicine NLo. Overview of SNOMED CT October 14, 2016 [Available from: https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html.

27. Health I. RxNORM 101: What you need to know for clinical documentation 2023 [Available from: https://www.imohealth.com/resources/rxnorm-101-what-you-need-to-know-for-clinical-documentation/.

28. Yeh CY, Peng SJ, Yang HC, Islam M, Poly TN, Hsu CY, et al. Logical Observation Identifiers Names and Codes (LOINC(®)) Applied to Microbiology: A National Laboratory Mapping Experience in Taiwan. Diagnostics (Basel). 2021;11(9).

29. Myhr KM, Mellgren SI. Corticosteroids in the treatment of multiple sclerosis. Acta Neurol Scand Suppl. 2009(189):73-80.

30. Chataway J, Porter B, Riazi A, Heaney D, Watt H, Hobart J, et al. Home versus outpatient administration of intravenous steroids for multiple-sclerosis relapses: a randomised controlled trial. Lancet Neurol. 2006;5(7):565-71.

31. Morrow SA, Fraser JA, Day C, Bowman D, Rosehart H, Kremenchutzky M, et al. Effect of Treating Acute Optic Neuritis With Bioequivalent Oral vs Intravenous Corticosteroids: A Randomized Clinical Trial. JAMA Neurol. 2018;75(6):690-6.

32. Roberts K. MRI and MS: 7 things you need to know: MS Society; 2024 [

33. Fernández O, Fernández V, Arbizu T, Izquierdo G, Bosca I, Arroyo R, et al. Characteristics of multiple sclerosis at onset and delay of diagnosis and treatment in Spain (the Novo Study). J Neurol. 2010;257(9):1500-7.

34. Koch MW, Moral E, Brieva L, Mostert J, Strijbis EM, Comtois J, et al. Relapse recovery in relapsing-remitting multiple sclerosis: An analysis of the CombiRx dataset. Mult Scler. 2023;29(14):1776-85.

35. Nazareth TA, Rava AR, Polyakov JL, Banfe EN, Waltrip Ii RW, Zerkowski KB, et al. Relapse prevalence, symptoms, and health care engagement: patient insights from the Multiple Sclerosis in America 2017 survey. Mult Scler Relat Disord. 2018;26:219-34.

36. OHDSI. Standardized Data: The OMOP Common Data Model [Available from: https://www.ohdsi.org/data-standardization/.

37. Pezoulas VC, Zaridis DI, Mylona E, Androutsos C, Apostolidis K, Tachos NS, et al. Synthetic data generation methods in healthcare: A review on open-source tools and methods. Comput Struct Biotechnol J. 2024;23:2892-910.

Appendices

Appendix 1: Overview of clinical codes terminology used

Vocab	Code	Name	Domain	Class
SNOMED	56397003	Neurologist	Provider	Social context
RxNorm	6902	Methylprednisolone	Drug	Ingredient
SNOMED	816077007	MRI of brain	Procedure	Procedure
SNOMED	748661000000100	MRI of spinal cord	Procedure	Procedure
SNOMED	413017004	Cerebrospinal fluid oligoclonal band screening test	Measurement	Procedure
LOINC	32358-4	Oligoclonal bands [Presence] in Cerebral spinal fluid	Measurement	Lab test
SNOMED	788751009	Corticosteroid and/or corticosteroid derivative therapy	Procedure	Procedure
LOINC	LA24655-5	Abnormal brain MRI	Meas Value	Answer
SNOMED	24700007	Multiple sclerosis	Condition	Disorder
SNOMED	257556004	Surgery	Observation	Qualifier Value
SNOMED	38822007	Cystitis	Condition	Disorder
SNOMED	257551009	Infection	Condition	Disorder
SNOMED	41368006	Disorder of urinary tract	Condition	Disorder
SNOMED	37796009	Migraine	Condition	Disorder
SNOMED	59621000	Essential hypertension	Condition	Disorder
SNOMED	1300212001	Allergy Status	Observation	Observable Entity
SNOMED	890601000000107	Hypercholesterolemia	Condition	Disorder
SNOMED	712823008	Acute Depression	Condition	Disorder
SNOMED	928000	Disorder of musculoskeletal system	Condition	Disorder
SNOMED	73211009	Diabetes mellitus	Condition	Disorder
SNOMED	128292002	Chronic disease of cardiovascular system	Condition	Disorder
SNOMED	56294008	Nicotine dependence	Condition	Disorder
SNOMED	84757009	Epilepsy	Condition	Disorder

SNOMED	395099008	Cancer Confirmed	Condition	Context-dependent
SNOMED	4740000	Herpes zoster	Condition	Disorder
SNOMED	445967004	Clinically isolated syndrome	Condition	Disorder
SNOMED	428700003	Primary progressive multiple sclerosis	Condition	Disorder
SNOMED	426373005	Relapsing remitting multiple sclerosis	Condition	Disorder
SNOMED	343601000000105	Secondary progressive multiple sclerosis	Condition	Disorder
SNOMED	724778008	Progressive relapsing multiple sclerosis	Condition	Disorder
LOINC	LP241977-0	EDSS	Observation	LOINC Method
RxNorm	152605	Betaferon	Drug	Brand Name
RxNorm	153326	Avonex	Drug	Brand Name
RxNorm	1310526	Aubagio	Drug	Brand Name
RxNorm	135779	Copaxone	Drug	Brand Name
RxNorm	1012896	Gilenya	Drug	Brand Name
RxNorm	202590	Endoxan	Drug	Brand Name
RxNorm	82050	Novantrone	Drug	Brand Name
RxNorm	1594659	Lemtrada	Drug	Brand Name
RxNorm	1876381	Ocrevus	Drug	Brand Name
RxNorm	228271	Rebif	Drug	Brand Name

Appendix 2: Generated synthetic patient data

Table 1: List of generated patients

ID	BIRTHDATE	PREFIX	FIRST	MIDDLE	LAST	MARITAL	RACE	ETHNICITY	GENDER	BIRTHPLACE	ADDRESS	CITY	STATE	COUNTY	HEALTHCARE EXPENSES	HEALTHCARE COVERAGE	INCOME
81859ec4-bd56-265b-54f2-973b55054014	2/07/2023		Scot349	Colin861	Roob72		white	Non hispanic	M	Berlin Massachusetts US	468 Hackett Tunnel	Methuen	Massachusetts	Essex County	288256	133984	727133
9bf05617-382e-1715-274d-5c3334331d60	15/05/2011		Kiersten731		Dickinson688		white	Non hispanic	F	Waltham Massachusetts US	968 Rau Skyway	Worcester	Massachusetts	Worcester County	140000	828676	17324
d34fae2e-aee5-93e4-cde5-3212ec72ef92	28/08/1982	Mrs.	Jacqualine965		Hodkiewicz467	M	white	Non hispanic	F	Nizhny Novgorod Nizhny Novgorod Oblast RU	629 Kreiger Ranch	Quincy	Massachusetts	Norfolk County	1397078	0	86727
2f0862d5-3404-00f2-cd36-8f713c68a24f	16/11/1992	Mrs.	Tisa11	Lawanda300	Kub800	M	white	Non hispanic	F	Revere Massachusetts US	198 Cruickshank Overpass Apt 72	Boylston	Massachusetts	Worcester County	1316182	0	95433
7ee64453-195d-aecc-f494-d36a3873c970	18/09/1999	Ms.	Eva64	Lorena247	Concepción765		white	hispanic	F	Portsmouth Saint John Parish DM	559 Kilback Trail Apt 51	Bedford	Massachusetts	Middlesex County	1159716	0	52224
e059c8ab-8424-660c-72bf-279b3393bf40	28/06/1967	Mrs.	Maryjane289	Serina556	Romaguera67	D	asian	Non hispanic	F	Ashland Massachusetts US	939 Daugherty Crossing	Lowell	Massachusetts	Middlesex County	1971354	0	35606
e3453340-d387-3f4c-a13a-19e2fdb812af	24/01/1984	Mrs.	Laureen100	Jay242	Runte676	M	white	Non hispanic	F	Plymouth Massachusetts US	249 Lemke Frontage road	Brockton	Massachusetts	Plymouth County	1496285	52029	98383
5852562a-8aa8-017b-89e1-10226adef051	12/03/1935	Mrs.	Alayna598	Lilliam592	Emard19	W	white	Non hispanic	F	Boston Massachusetts US	887 Wilkinson Esplanade	Seekonk	Massachusetts	Bristol County	1994984	196016	196452
ab44f067-2309-d3e4-9a35-ab246c4c3568	9/03/2022		Wayne846	Tristan353	Weber641		white	Non hispanic	M	Boston Massachusetts US	831 Crist Trail Apt 10	Bedford	Massachusetts	Middlesex County	370096	133984	126805
0392f0ab-5dbe-9c9d-9046-4af3e34be4f5	15/04/2022		Marilu588	Emilia403	Valdivia496		white	hispanic	F	Guatemala City Guatemala GT	420 Welch Dam Suite 60	Newton	Massachusetts	Middlesex County	370096	133984	41635

Table 2: List of generated conditions

START	STOP	PATIENT	ENCOUNTER	SYSTEM	CODE	DESCRIPTION
8/10/2020		d34fae2e-aee5-93e4-cde5-3212ec72ef92	932bb9d3-2ac7-152a-fbef-c2c6e857c2b4	SNOMED-CT	24700007	Multiple Sclerosis
8/10/2020		d34fae2e-aee5-93e4-cde5-3212ec72ef92	932bb9d3-2ac7-152a-fbef-c2c6e857c2b4	SNOMED-CT	445967004	Clinically isolated syndrome
8/10/2020		d34fae2e-aee5-93e4-cde5-3212ec72ef92	d857f545-5aa1-8135-b957-53b52ef8330a	SNOMED-CT	426373005	Relapsing remitting multiple sclerosis
26/04/2020		2f0862d5-3404-00f2-cd36-8f713c68a24f	a41cb788-18fa-b293-1123-ad78f4a8b016	SNOMED-CT	24700007	Multiple Sclerosis
26/04/2020		2f0862d5-3404-00f2-cd36-8f713c68a24f	a41cb788-18fa-b293-1123-ad78f4a8b016	SNOMED-CT	445967004	Clinically isolated syndrome
15/03/1994		e059c8ab-8424-660c-72bf-279b3393bf40	dfa7d0a0-3a18-52e2-dd5f-768f9e902b3c	SNOMED-CT	24700007	Multiple Sclerosis
15/03/1994		e059c8ab-8424-660c-72bf-279b3393bf40	dfa7d0a0-3a18-52e2-dd5f-768f9e902b3c	SNOMED-CT	445967004	Clinically isolated syndrome
15/03/1994		e059c8ab-8424-660c-72bf-279b3393bf40	d512c452-9568-4f84-40dc-17cf05506801	SNOMED-CT	426373005	Relapsing remitting multiple sclerosis
9/01/2013		e3453340-d387-3f4c-a13a-19e2fdb812af	32492962-5986-3b89-8ebc-e5a6968fe3bb	SNOMED-CT	24700007	Multiple Sclerosis
9/01/2013		e3453340-d387-3f4c-a13a-19e2fdb812af	32492962-5986-3b89-8ebc-e5a6968fe3bb	SNOMED-CT	56294008	Nicotine dependence
9/01/2013		e3453340-d387-3f4c-a13a-19e2fdb812af	32492962-5986-3b89-8ebc-e5a6968fe3bb	SNOMED-CT	445967004	Clinically isolated syndrome
9/01/2013		e3453340-d387-3f4c-a13a-19e2fdb812af	531bf720-a43c-334a-fa99-73ce69ed4ae1	SNOMED-CT	426373005	Relapsing remitting multiple sclerosis
13/05/1963		5852562a-8aa8-017b-89e1-10226adef051	633f91a9-0439-dfe5-704b-0eee4e132963	SNOMED-CT	24700007	Multiple Sclerosis
13/05/1963		5852562a-8aa8-017b-89e1-10226adef051	633f91a9-0439-dfe5-704b-0eee4e132963	SNOMED-CT	445967004	Clinically isolated syndrome
13/05/1963		5852562a-8aa8-017b-89e1-10226adef051	1aec9af3-4d01-b82e-4a37-c78fae3f11a6	SNOMED-CT	426373005	Relapsing remitting multiple sclerosis

Table 3: List of generated medication orders

START	STOP	PATIENT	PAYER	ENCOUNTER	CODE	DESCRIPTION	BASE COST	PAYER COVERAGE	DISPENSES	TOTAL COST	REASON CODE	REASON DESCRIPTION
6/05/2020 2:12		d34fae2e-aee5-93e4-cde5-3212ec72ef92	b046940f-1664-3047-bca7-dfa76be352a4	ef850f94-fce6-6fd2-aade-b1669dcca61a	6902	Methylprednisolone	12994	0	61	792634		
8/10/2020 2:12		d34fae2e-aee5-93e4-cde5-3212ec72ef92	b046940f-1664-3047-bca7-dfa76be352a4	d857f545-5aa1-8135-b957-53b52ef8330a	152605	Betaferon	12994	0	56	727664		
22/11/2019 0:55		2f0862d5-3404-00f2-cd36-8f713c68a24f	b046940f-1664-3047-bca7-dfa76be352a4	98cdf539-f5d0-d63b-893d-634ea6144949	6902	Methylprednisolone	12994	0	67	870598		
10/10/1993 6:58		e059c8ab-8424-660c-72bf-279b3393bf40	e03e23c9-4df1-3eb6-a62d-f70f02301496	f7eb6391-d18b-4283-a723-5d91fb9e935c	6902	Methylprednisolone	12994	0	385	5002690		
15/03/1994 5:58		e059c8ab-8424-660c-72bf-279b3393bf40	e03e23c9-4df1-3eb6-a62d-f70f02301496	d512c452-9568-4f84-40dc-17cf05506801	135779	Copaxone	12994	0	379	4924726		
7/08/2012 10:09		e3453340-d387-3f4c-a13a-19e2fdb812af	d31fccc3-1767-390d-966a-22a5156f4219	bfbfefef-a6c9-fdb2-77e3-5019466be420	6902	Methylprednisolone	12994	0	155	2014070		
9/01/2013 9:09		e3453340-d387-3f4c-a13a-19e2fdb812af	d31fccc3-1767-390d-966a-22a5156f4219	531bf720-a43c-334a-fa99-73ce69ed4ae1	153326	Avonex	12994	10395	150	1949100		
8/12/1962 6:09		5852562a-8aa8-017b-89e1-10226adef051	d31fccc3-1767-390d-966a-22a5156f4219	938840e4-4434-3024-aa33-a63d8d733808	6902	Methylprednisolone	12994	0	760	9875440		