

Synthea™

Synthetic Patient Generation

Dylan Hall – dehall@mitre.org

June 4, 2019



MITRE is a not-for-profit organization that operates federally funded research and development centers, sponsored by the federal government.

FFRDCs assist the United States government with:

- Scientific research and analysis
- Development and acquisition
- Systems engineering and integration

Department of Defense
Federal Aviation Administration
Internal Revenue Service
Department of Veterans Affairs
Centers for Medicare & Medicaid Services
Department of Homeland Security
Administrative Office of the U.S. Courts
National Institute of Standards and Technology

We also have an independent research program that explores new and expanded uses of technologies to solve our sponsors' problems.

THE U.S HEALTHCARE SYSTEM: WORST IN THE DEVELOPED WORLD

U.S. lags behind in healthcare innovation

True interoperability 'not even close'

High Demand for EHR Datasets

*Non-clinical or secondary uses including:
software development, testing, clinical training where realistic data is required*

Restrictions

Real patient records carry privacy, confidentiality, consent, policy, and legal restrictions

Privacy Risks

Deidentified and anonymized records have been successfully reidentified

Cost

EHR datasets are difficult to obtain.

Anonymized records are being bought and/or sold by federal and state health departments, hospitals, health insurers, pharmacists, government lobby groups, marketers, and data brokers



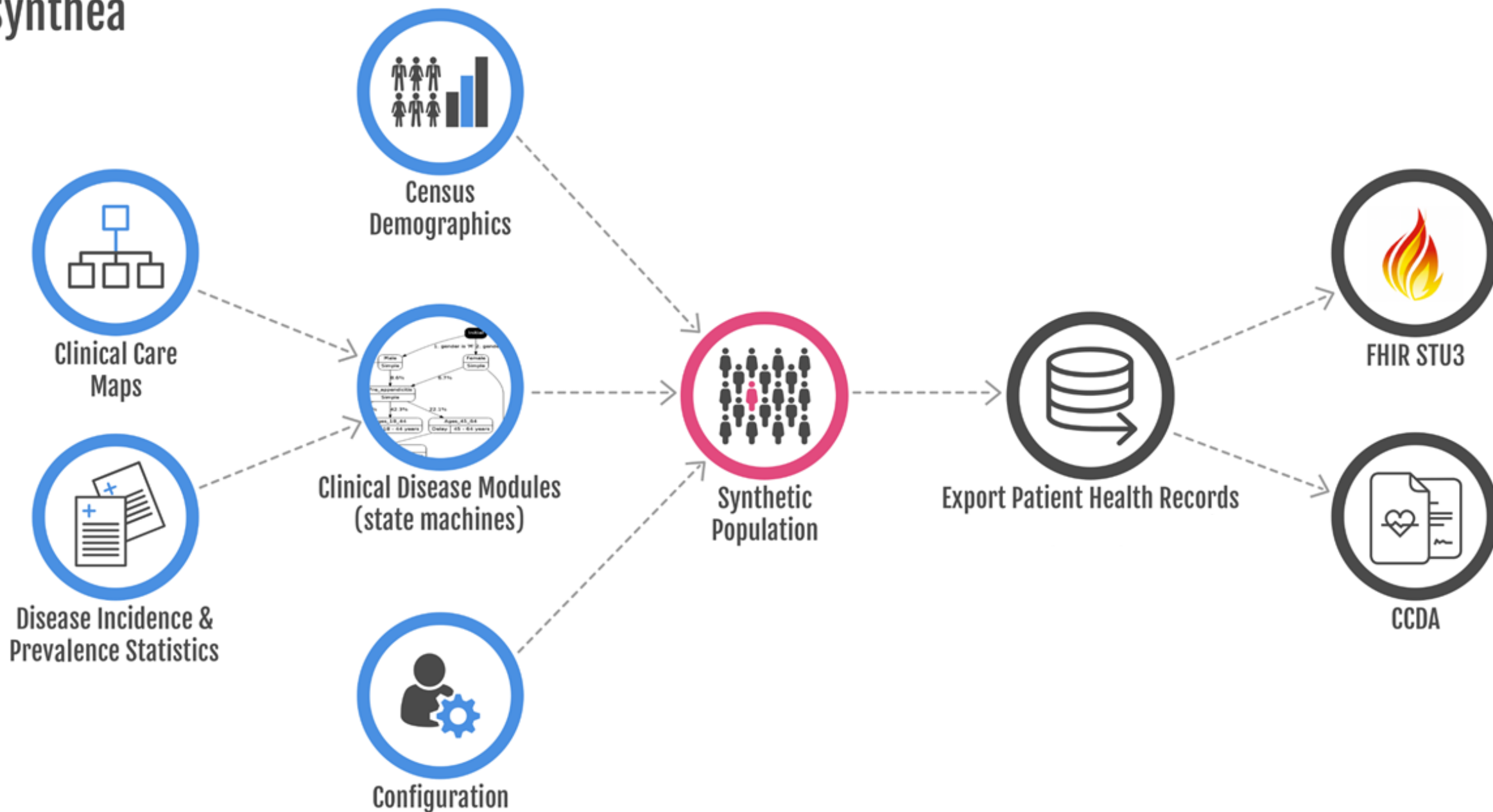
SYNTHEA

Synthetic Patient Generation

Realistic Health Data

No Cost, No Restrictions

Synthea



Reasons Patients Visit PCP		Causes of Premature Death (Years of Life Lost)
Routine Infant/Child Health Check	1	Ischemic Heart Disease
Essential Hypertension	2	Lung Cancer
Diabetes Mellitus	3	Alzheimer's Disease
Normal Pregnancy	4	COPD
Respiratory Infections (Pharyngitis, Bronchitis, Sinusitis)	5	Cerebrovascular Disease
General Adult Medical Examination	6	Road Injuries
Disorders of Lipoid Metabolism	7	Self-Harm
Ear Infections (Otitis Media)	8	Diabetes Mellitus
Asthma	9	Colorectal Cancer
Urinary Tract Infections	10	Drug Use Disorders

**As of August 2018: Veteran-Focused Modules,
In partnership with VHA**

“Top 5” Conditions Affecting Veterans

Hyperlipidemia

Hypertension

Osteoarthritis

Type 2 Diabetes

Major Depressive Disorder

PTSD

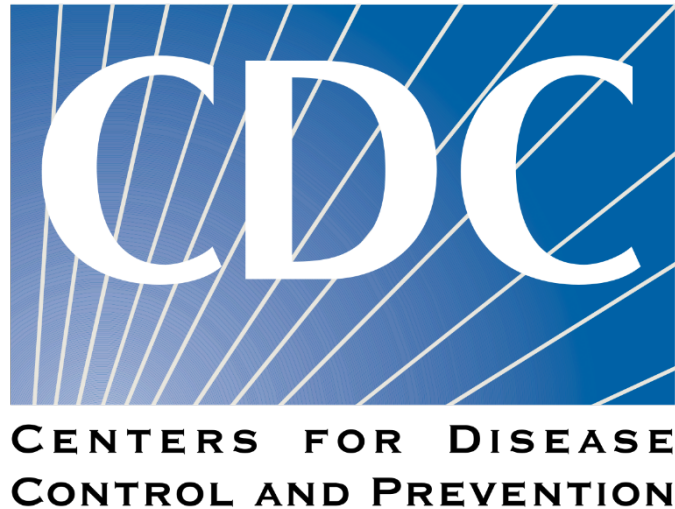
Cancers:

- Colorectal

- Lung

- Prostate

Publicly available data sources

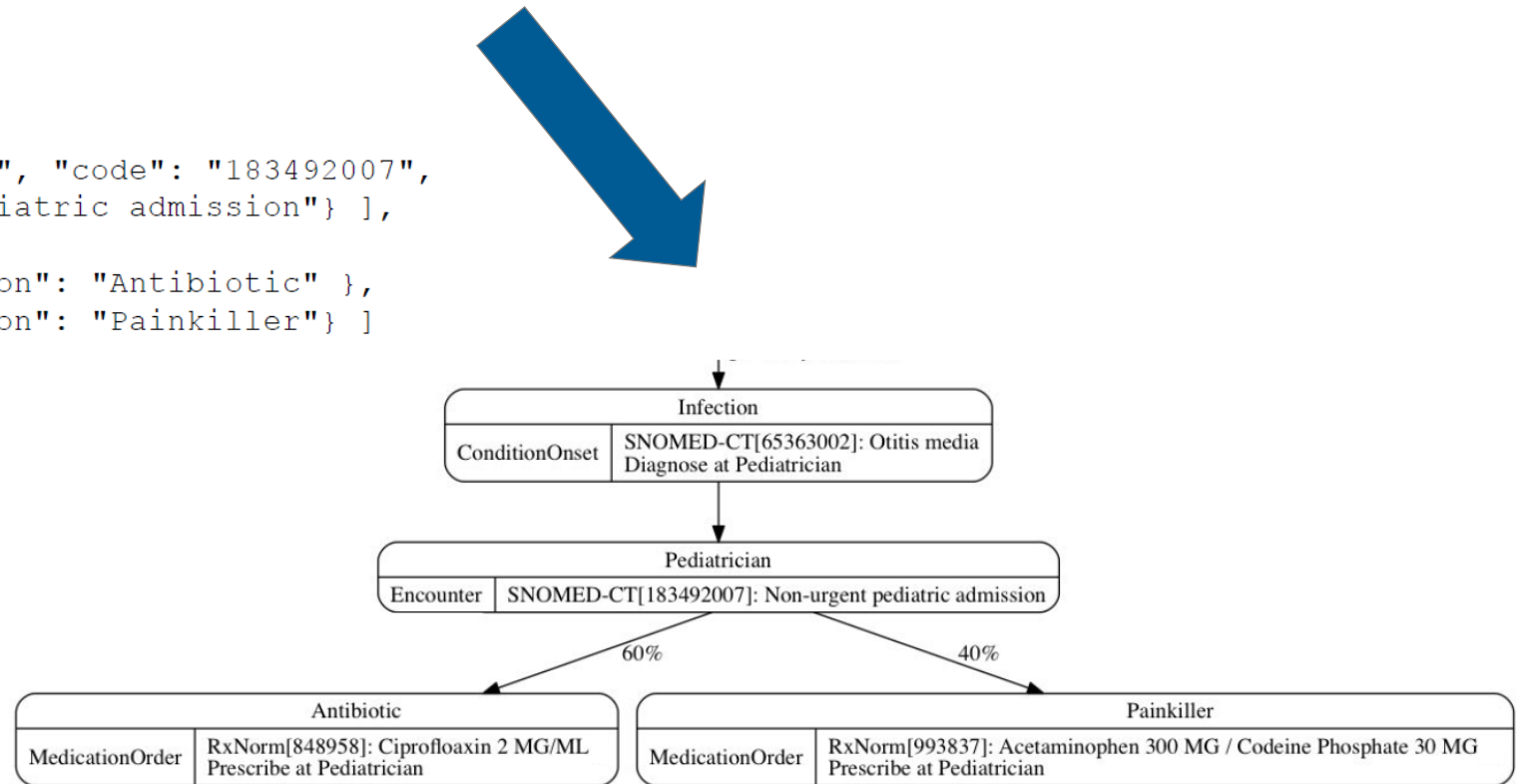


Clinical Pathways

*a.k.a.
care pathways,
critical pathways,
integrated care pathways,
care maps,
clinical practice algorithms*

Modules are written in JSON

```
"Infection": {
  "type": "ConditionOnset",
  "target_encounter": "Pediatrician",
  "codes": [ { "system": "SNOMED-CT", "code": "65363002", "display": "Otitis media" } ],
  "direct_transition": "Pediatrician"
},
"Pediatrician": {
  "type": "Encounter",
  "encounter_class": "ambulatory",
  "codes": [ { "system": "SNOMED-CT", "code": "183492007",
    "display": "Non-urgent pediatric admission" } ],
  "distributed_transition": [
    { "distribution": 0.6, "transition": "Antibiotic" },
    { "distribution": 0.4, "transition": "Painkiller" } ]
}
```



Synthea Module Builder UI

<https://synthetichealth.github.io/module-builder/>

Synthea Module Builder [New Module](#) [Load Module](#) [Download](#) [Add State](#) [Add Structure](#) [Help](#)

◀◀◀ **Exemplitis**

[Remove State](#) [Copy State](#)

Exemplitis
State Type: [ConditionOnset](#)

[Enter value](#)

Target Encounter: [Wellness Encounter](#)
[Add Assign to Attribute](#)

Codes

[remove](#)
System: SNOMED-CT
Code: [123](#)
Display: [Exemplitis](#)

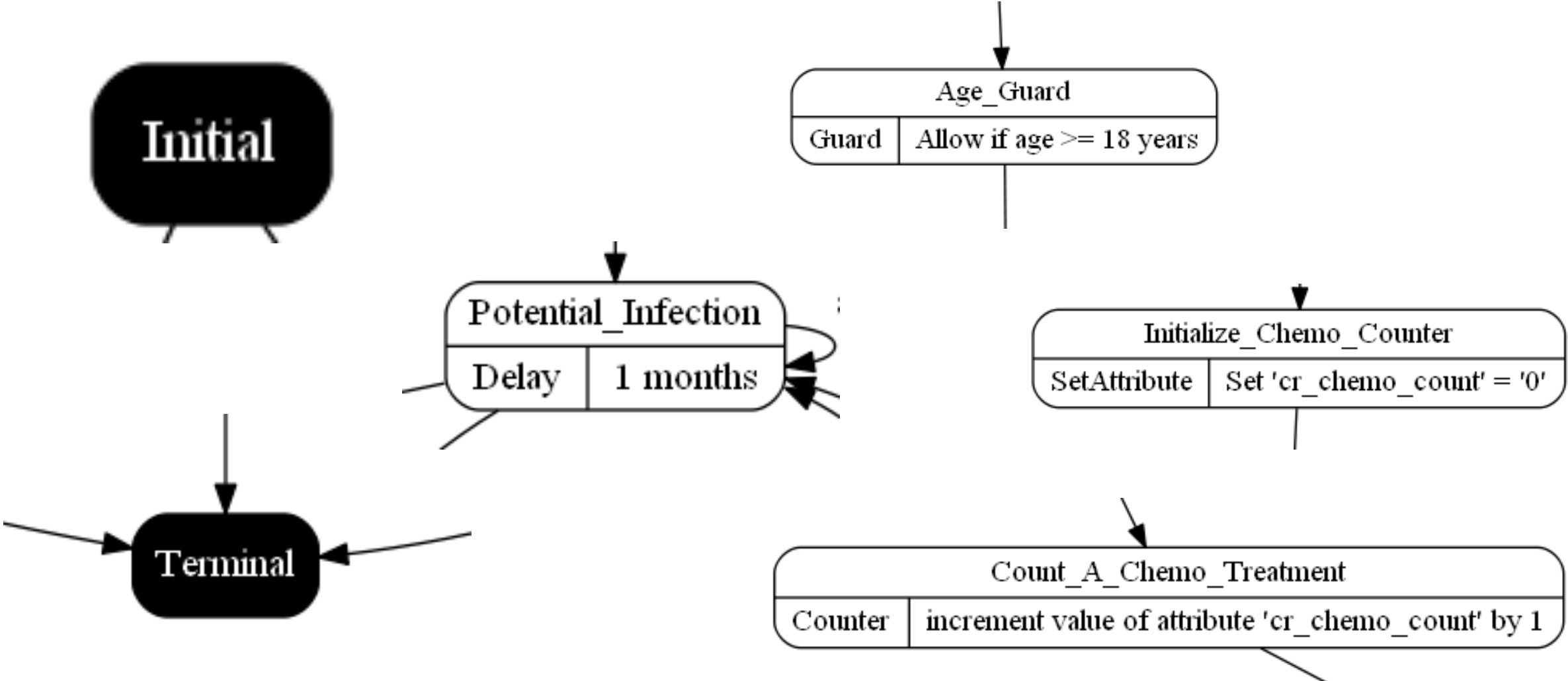
±

Transition Type: [Direct](#)

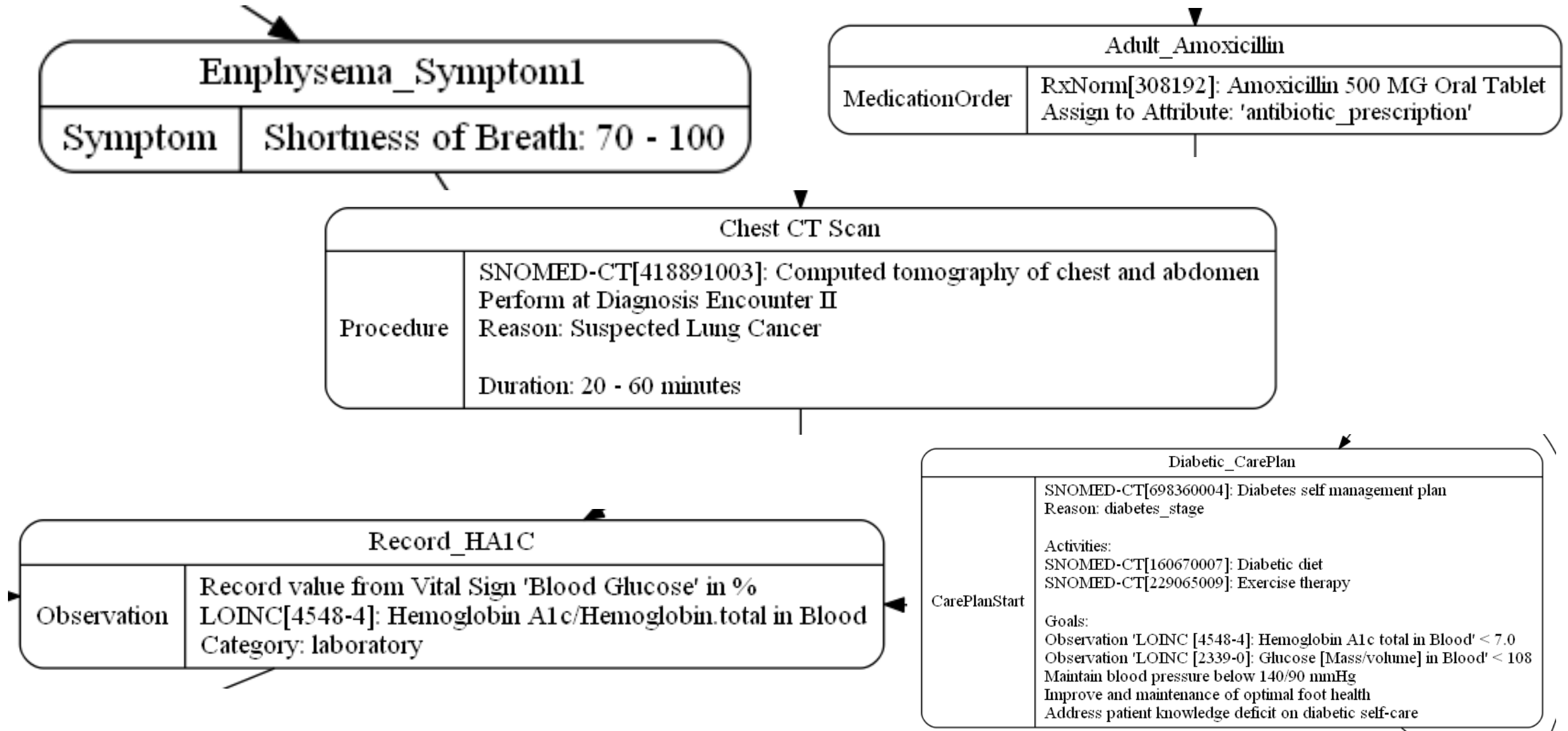
Transition To: [Wellness Encounter](#)

```
graph TD;
  Initial((Initial)) -- "1. gender is 'M'" --> Age_Guard;
  Age_Guard["Age_Guard  
Guard | Allow if age > 40 years"] -- "10%" --> Pre_Exemplitis;
  Pre_Exemplitis["Pre_Exemplitis  
Delay | 0 - 10 years"] --> Exemplitis;
  Exemplitis["Exemplitis  
ConditionOnset | SNOMED-CT[123]: Exemplitis  
Diagnose at Wellness_Encounter"] --> Wellness_Encounter;
  Wellness_Encounter["Wellness_Encounter  
Encounter | Wait for regularly scheduled wellness encounter"];
```

Control States control the flow through the state machine



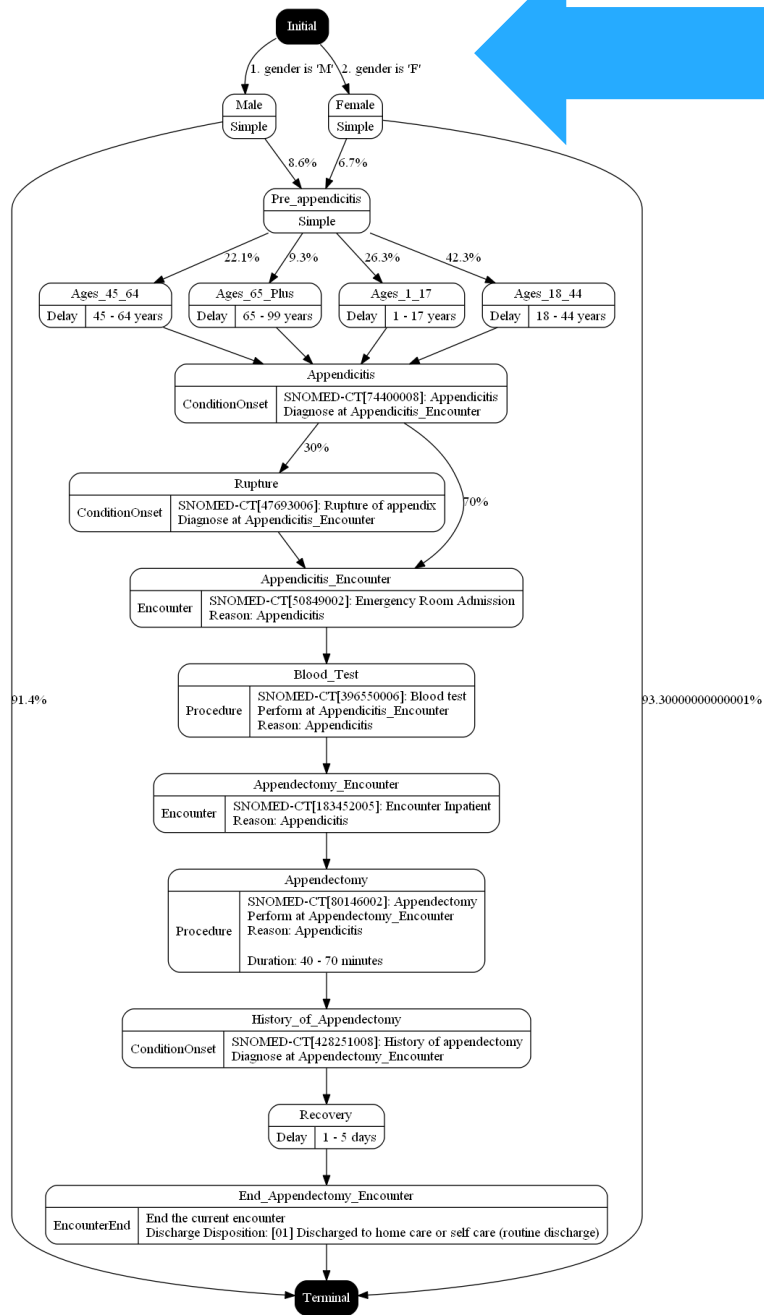
Clinical States drive the progression of disease and care



Model Validation

*Conditions and comorbidities:
Population - Incidence and prevalence
Individual - Frequencies and progression*

Example Module – Appendicitis



Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:

Smith292, John949

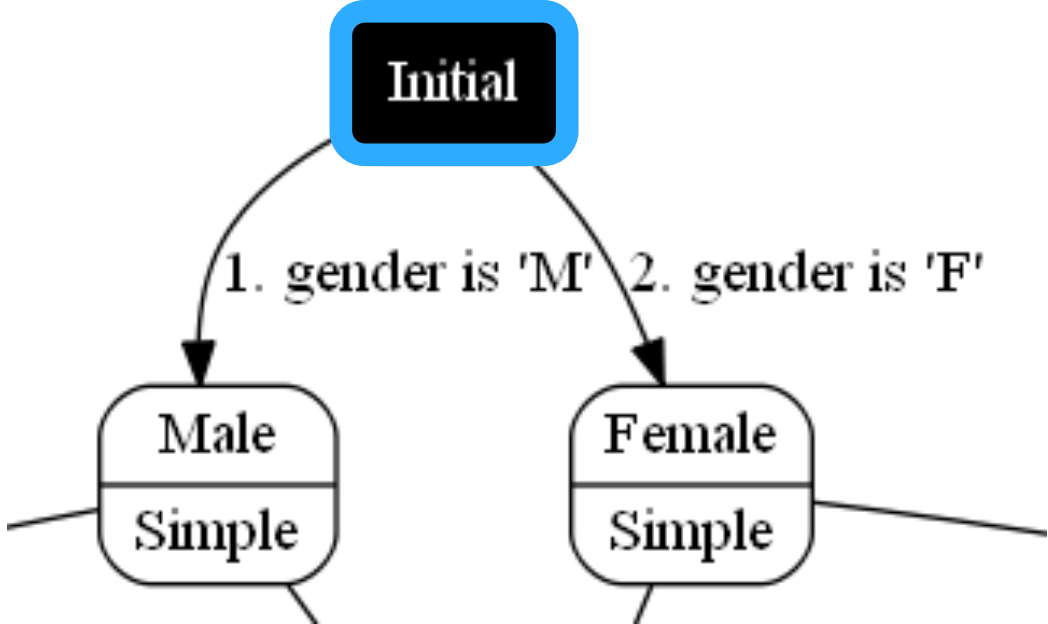
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

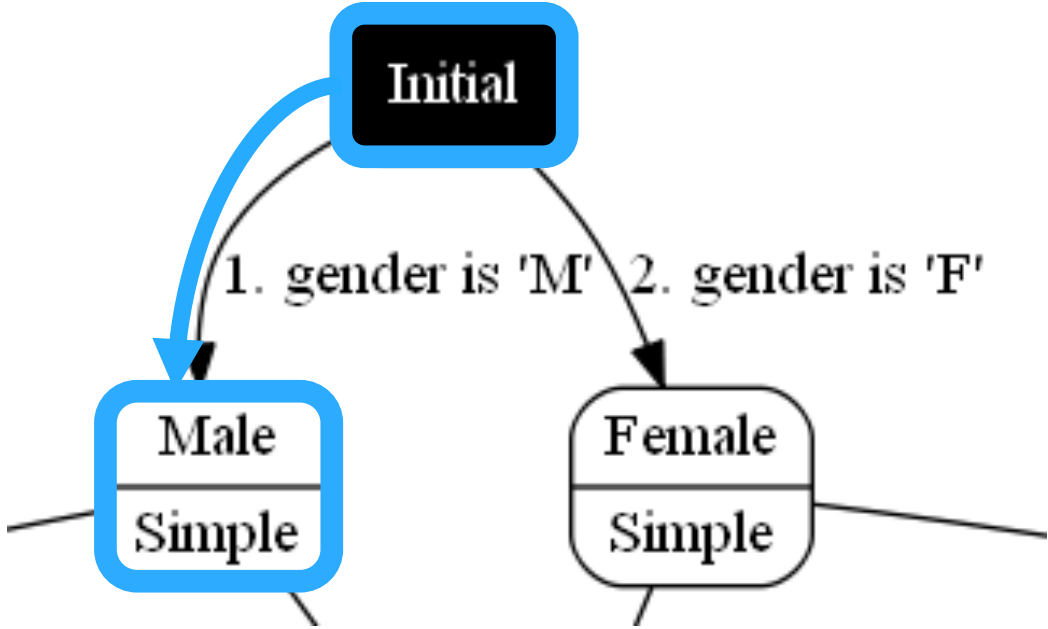
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

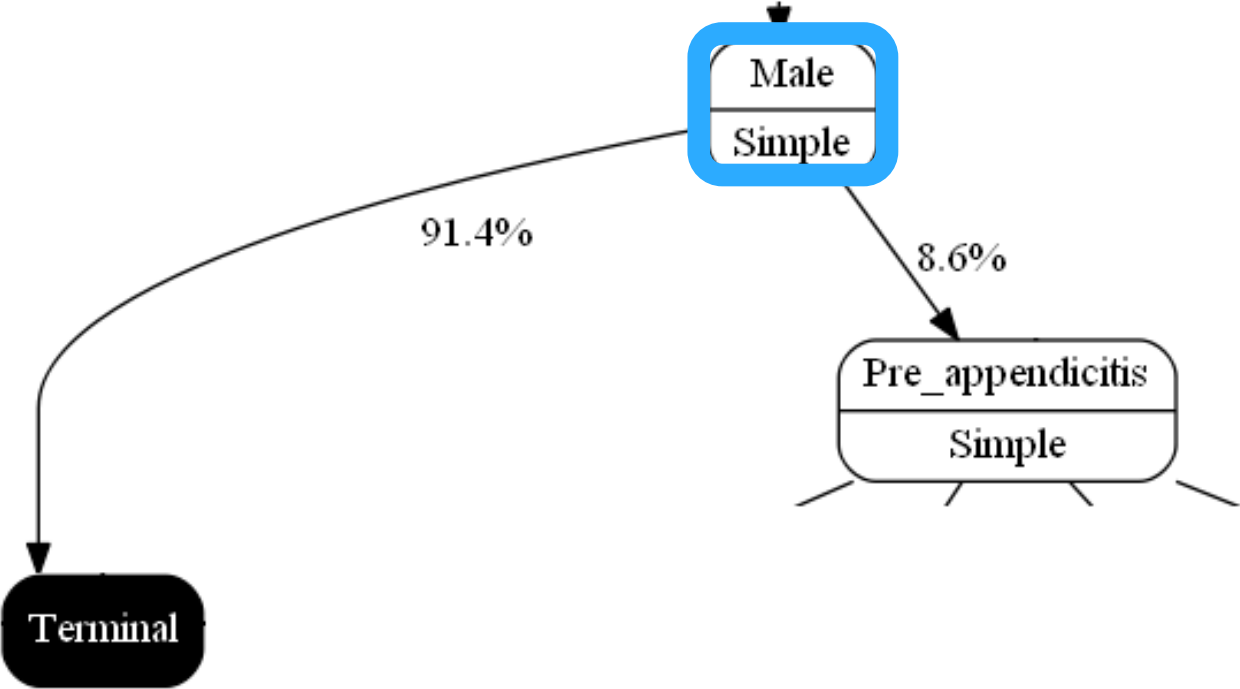
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

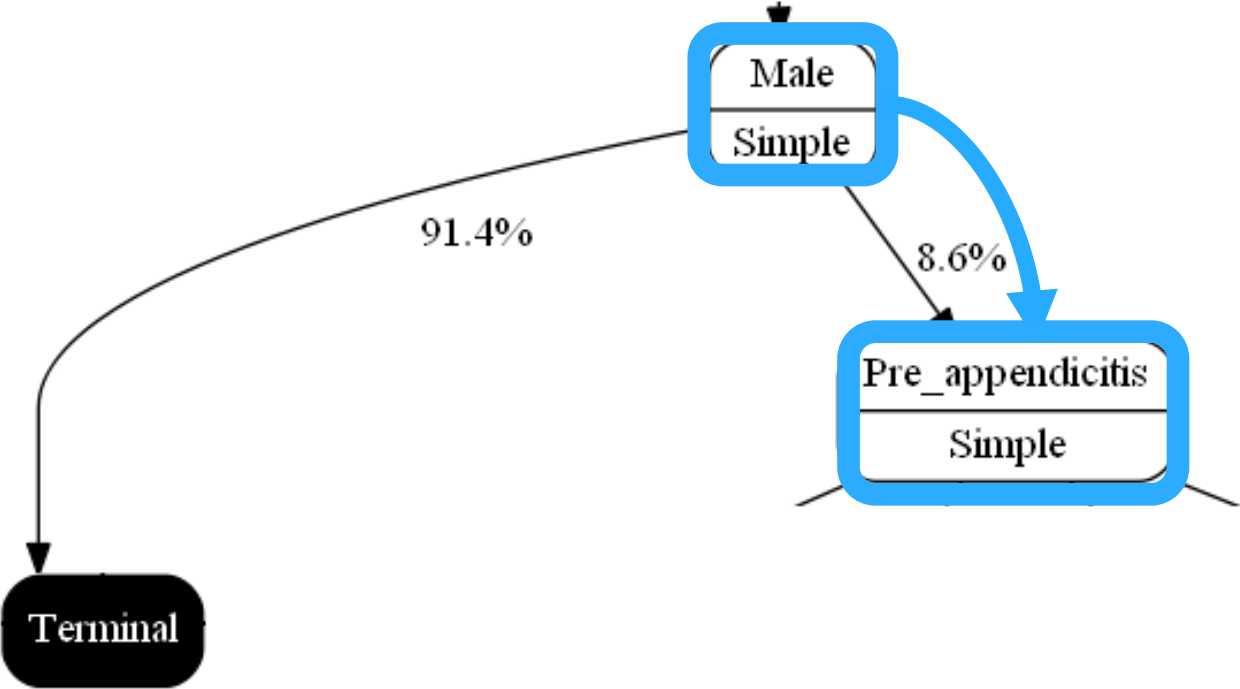
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

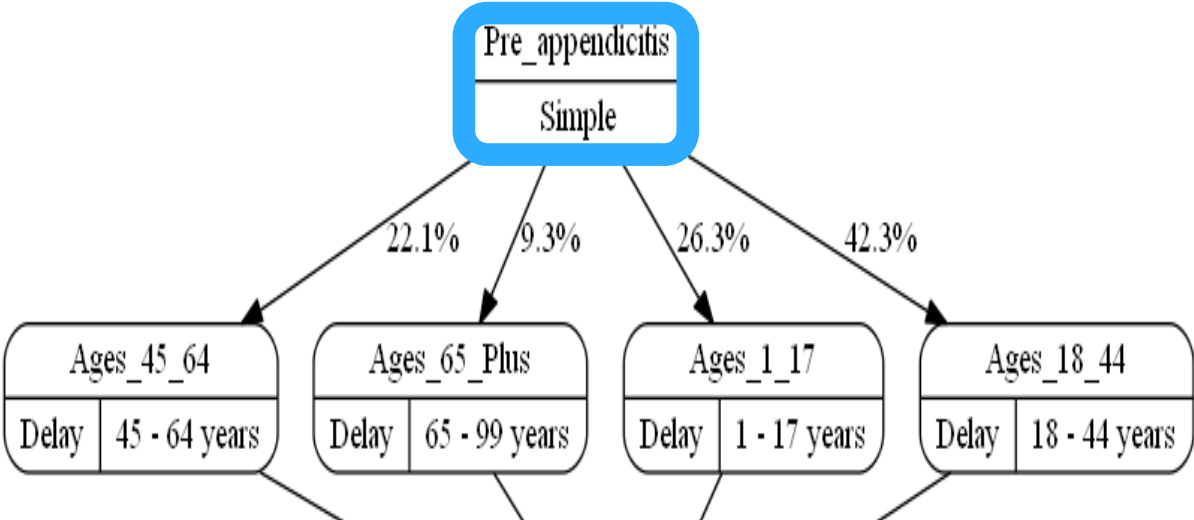
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

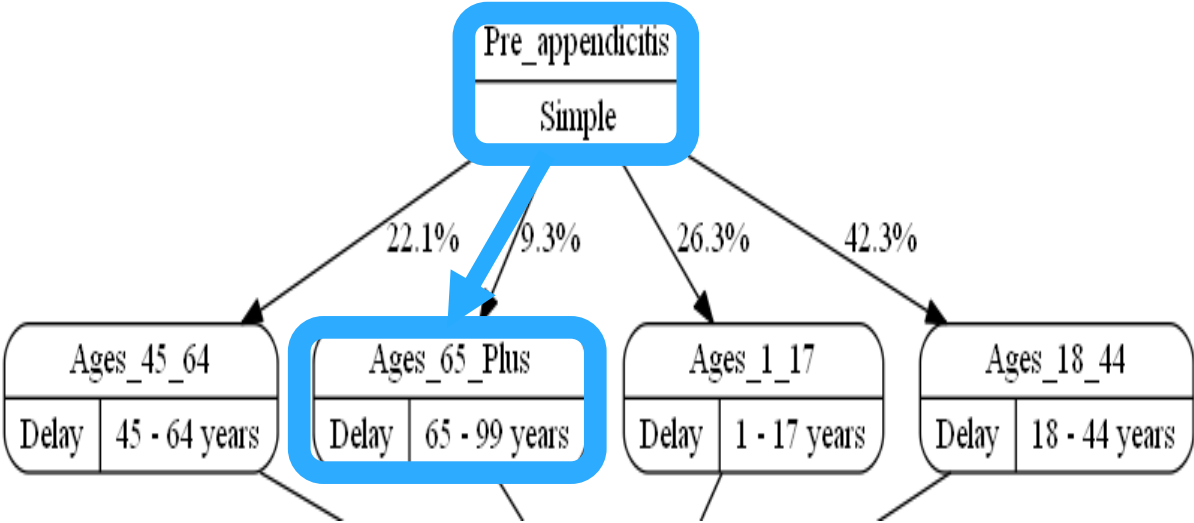
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

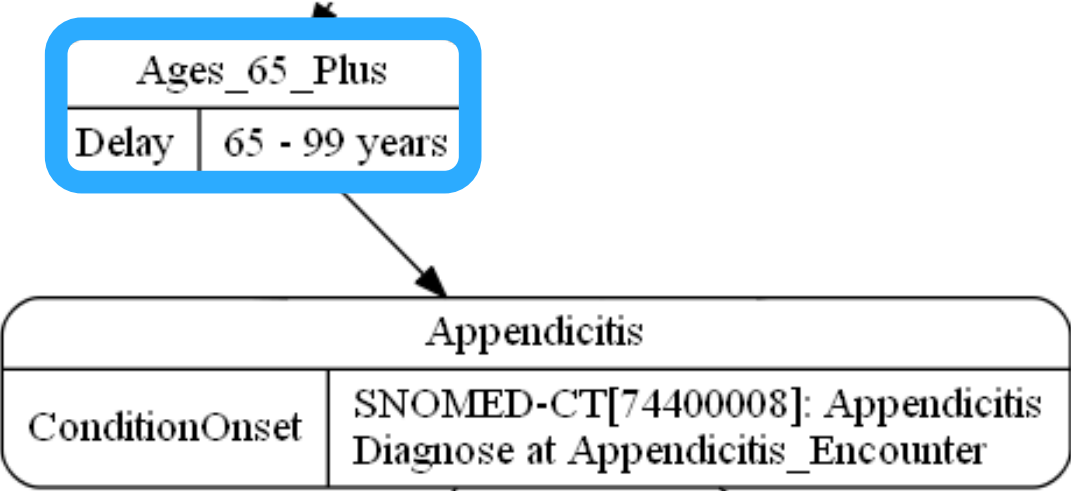
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
March 25, 1947

Smith292, John949

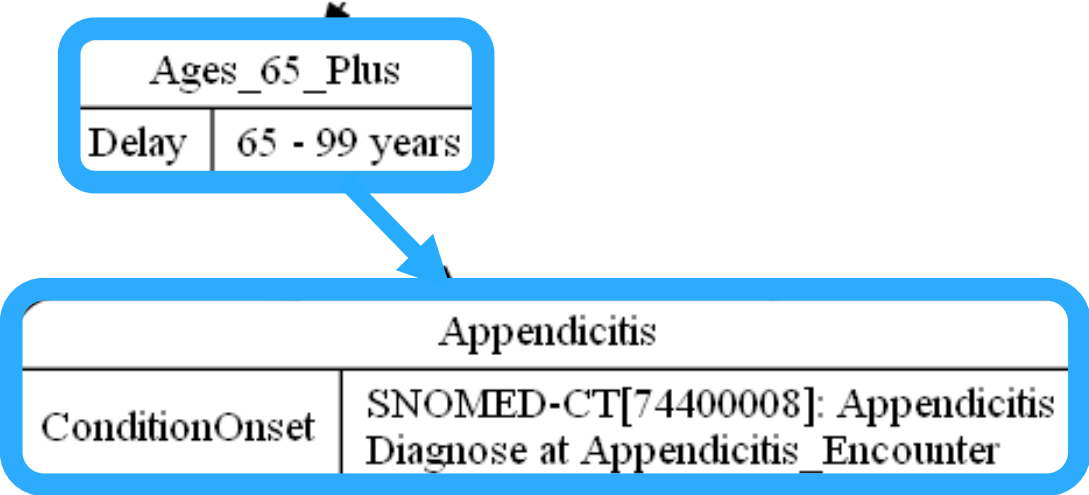
Born: March 25, 1947

Gender: Male

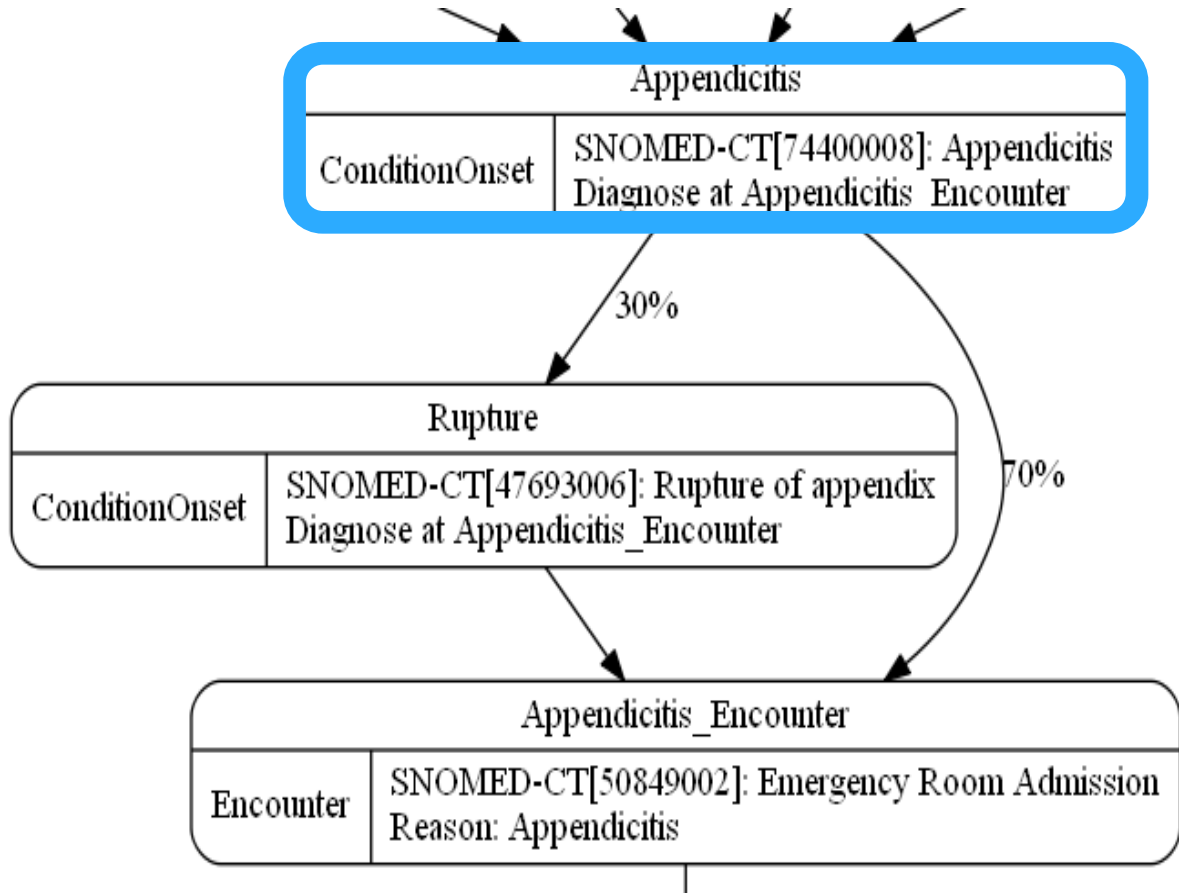
Encounters:

Conditions:

Procedures:



**Current Date of Simulation:
July 23, 2013**



Current Date of Simulation:
July 23, 2013

Smith292, John949

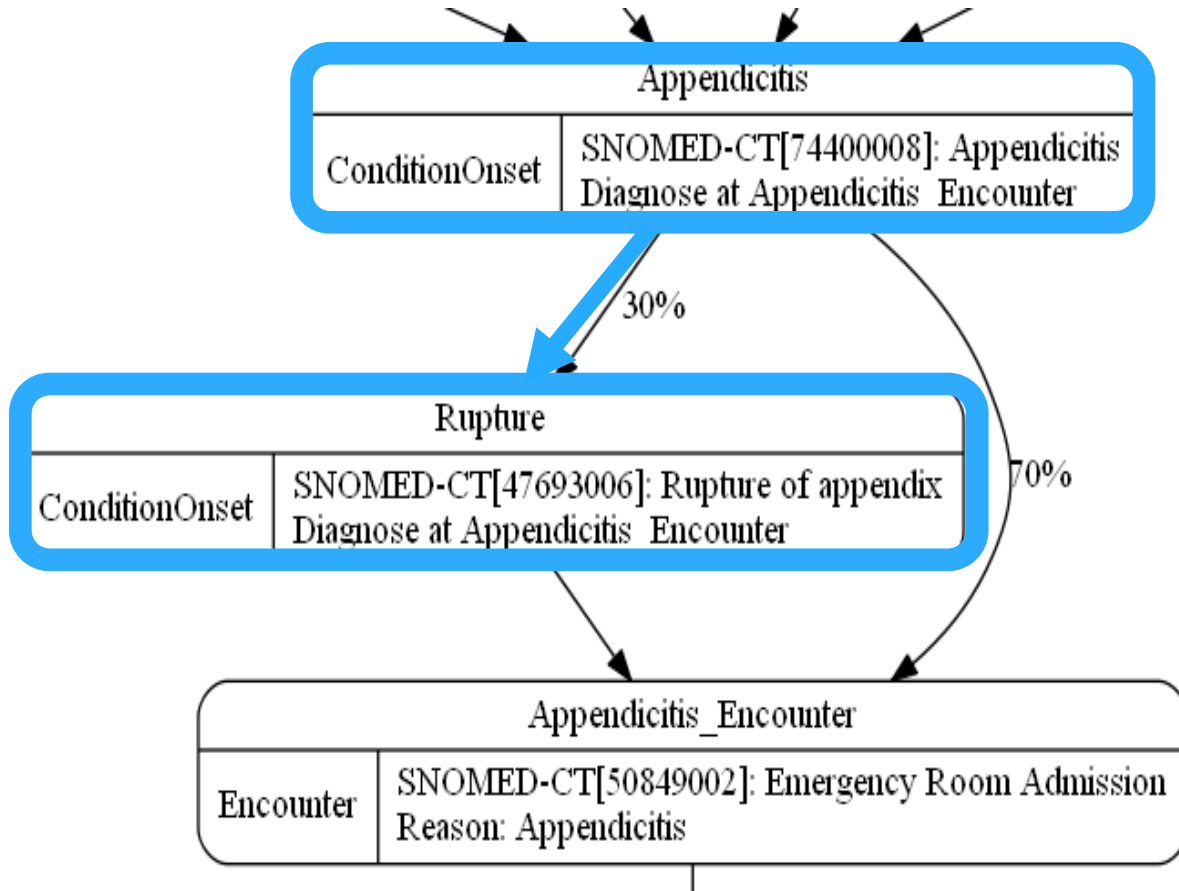
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
July 23, 2013

Smith292, John949

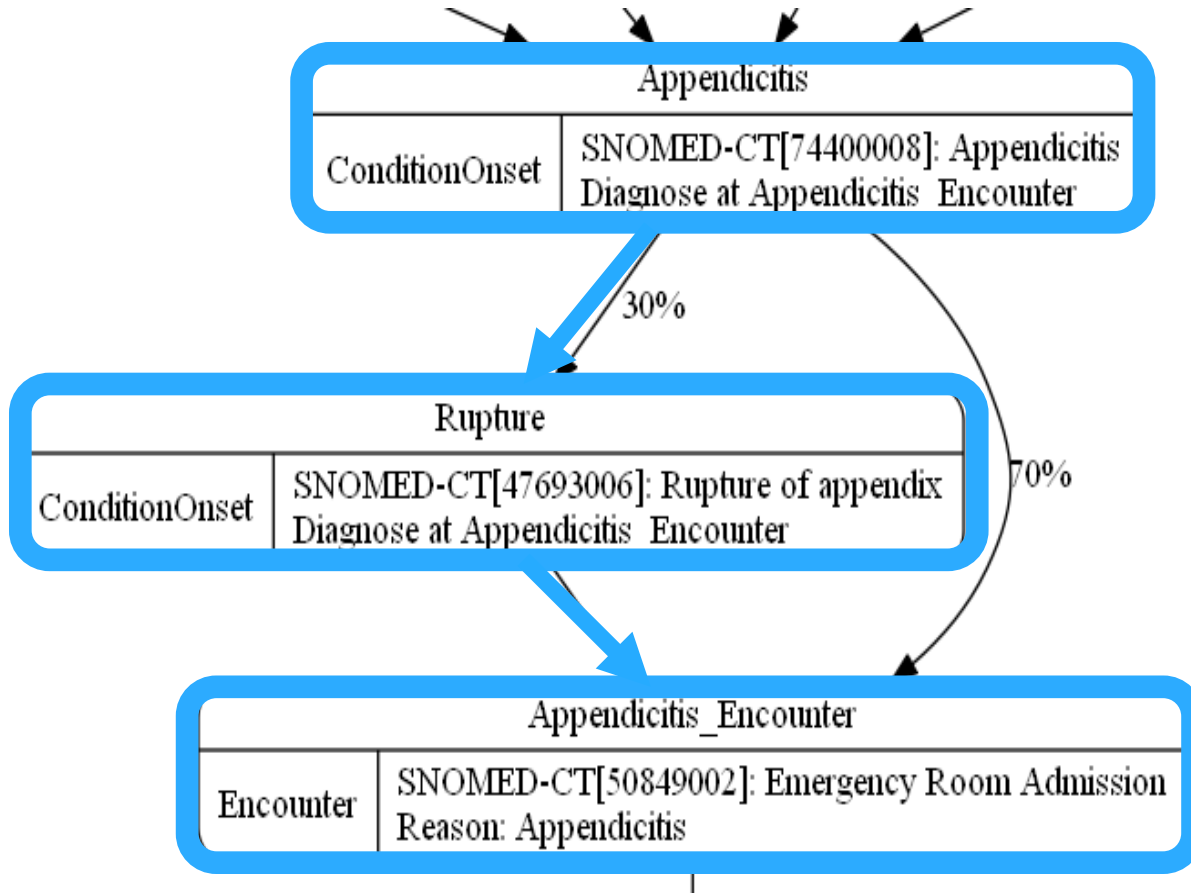
Born: March 25, 1947

Gender: Male

Encounters:

Conditions:

Procedures:



Current Date of Simulation:
July 23, 2013

Smith292, John949

Born: March 25, 1947

Gender: Male

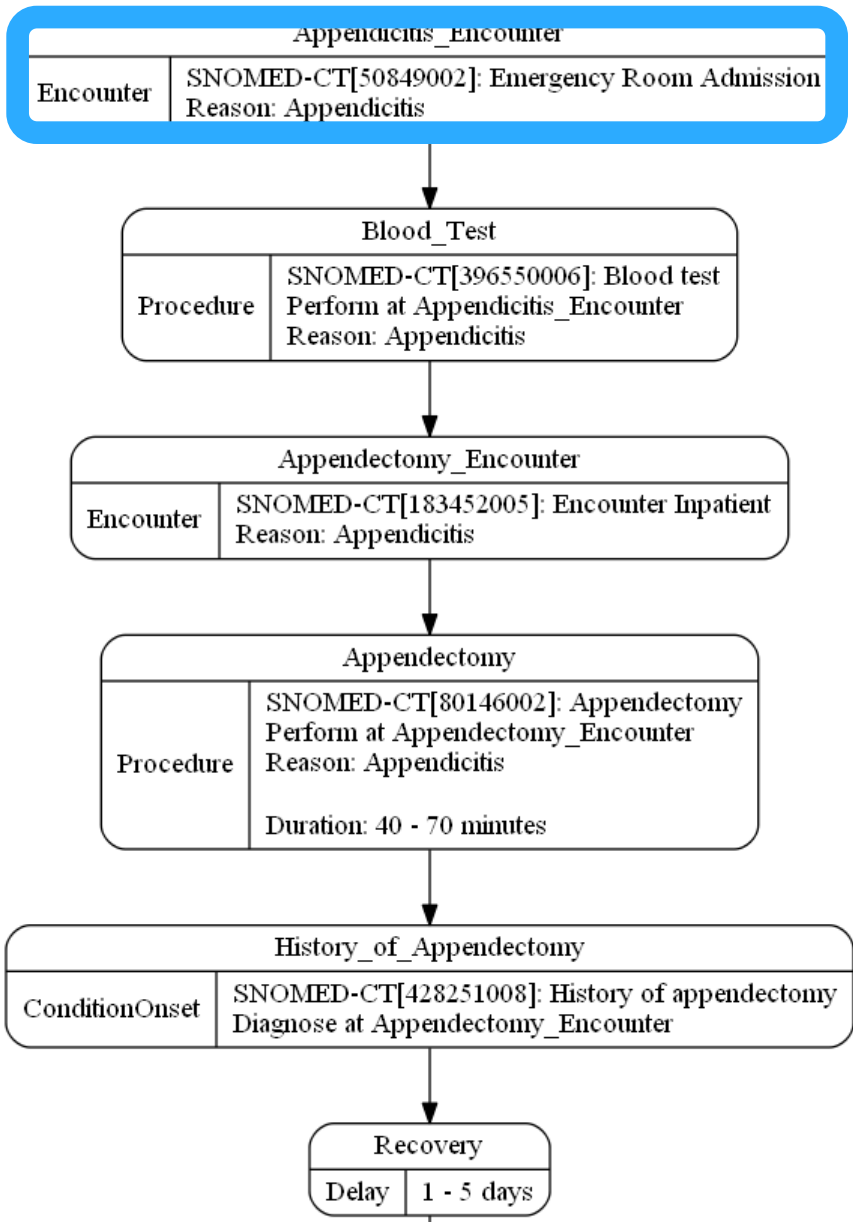
Encounters:

Emergency Room Admission,
July 23, 2013 →

Conditions:

Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:



Current Date of Simulation:
July 23, 2013

Smith292, John949

Born: March 25, 1947

Gender: Male

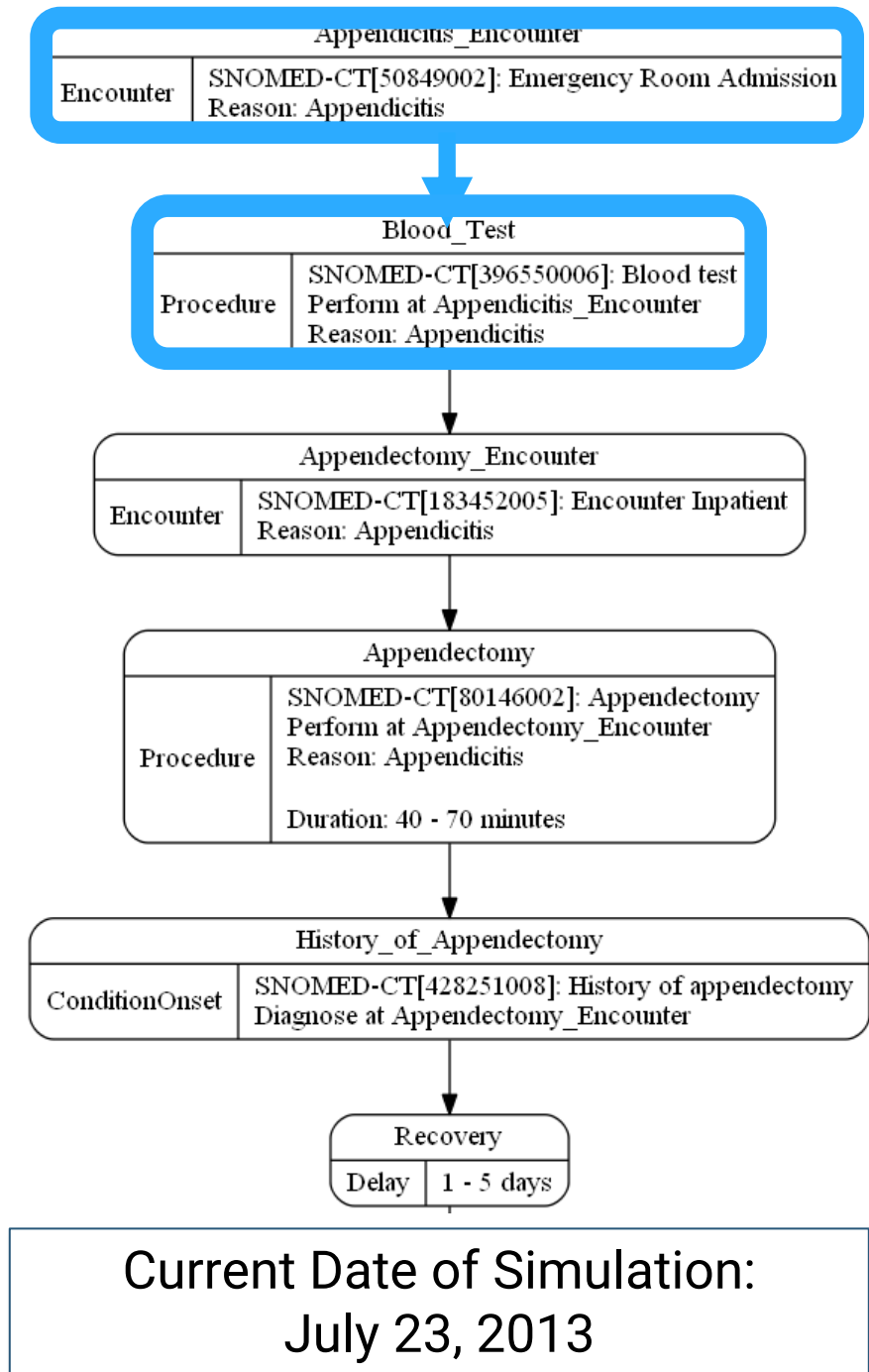
Encounters:

Emergency Room Admission,
July 23, 2013 →

Conditions:

Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:



Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

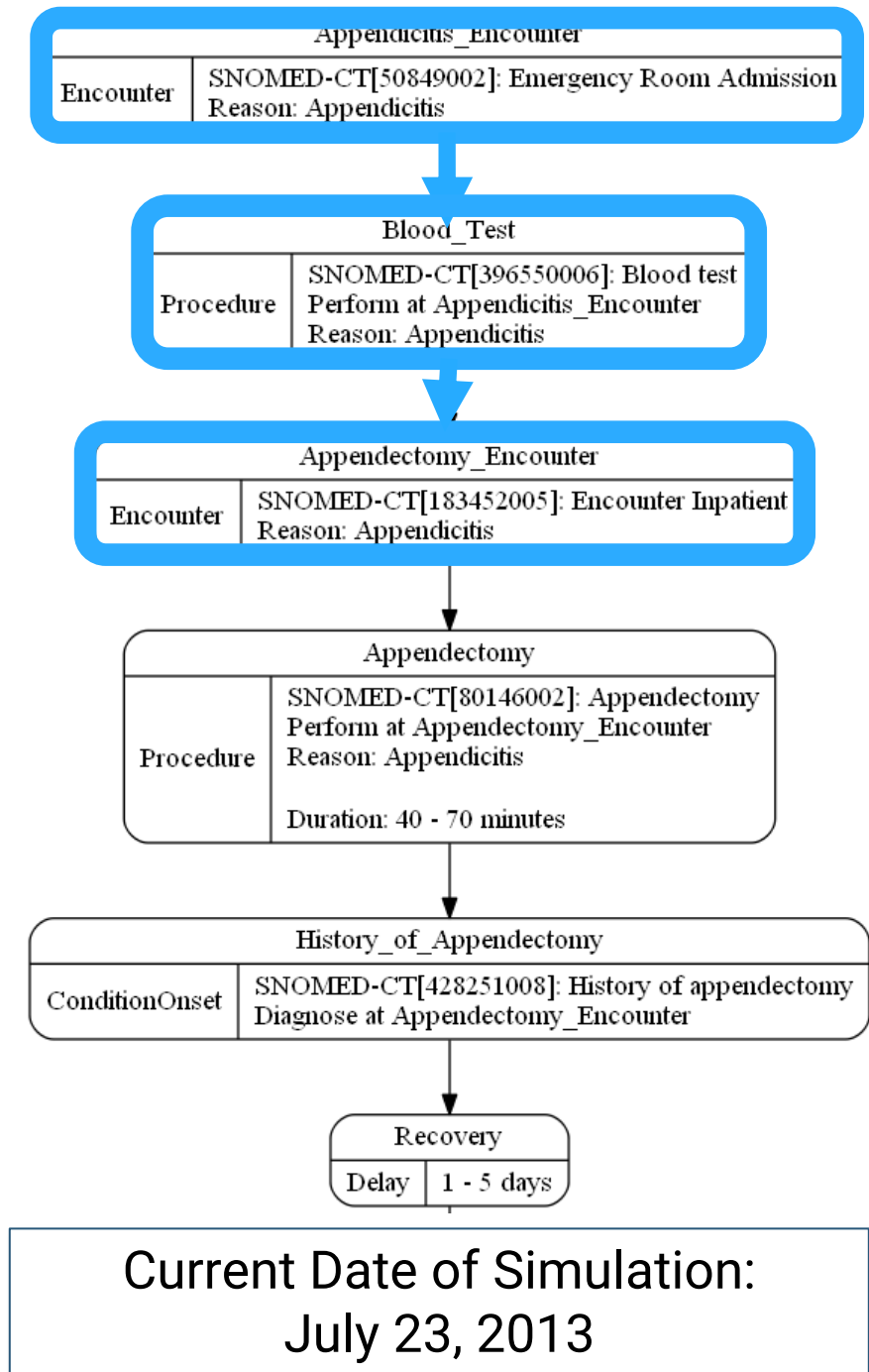
Emergency Room Admission,
July 23, 2013 →

Conditions:

Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013

Procedures:

Blood Test, Performed July 23, 2013



Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

Emergency Room Admission,
July 23, 2013 → July 23, 2013

Inpatient Encounter,
July 23, 2013 →

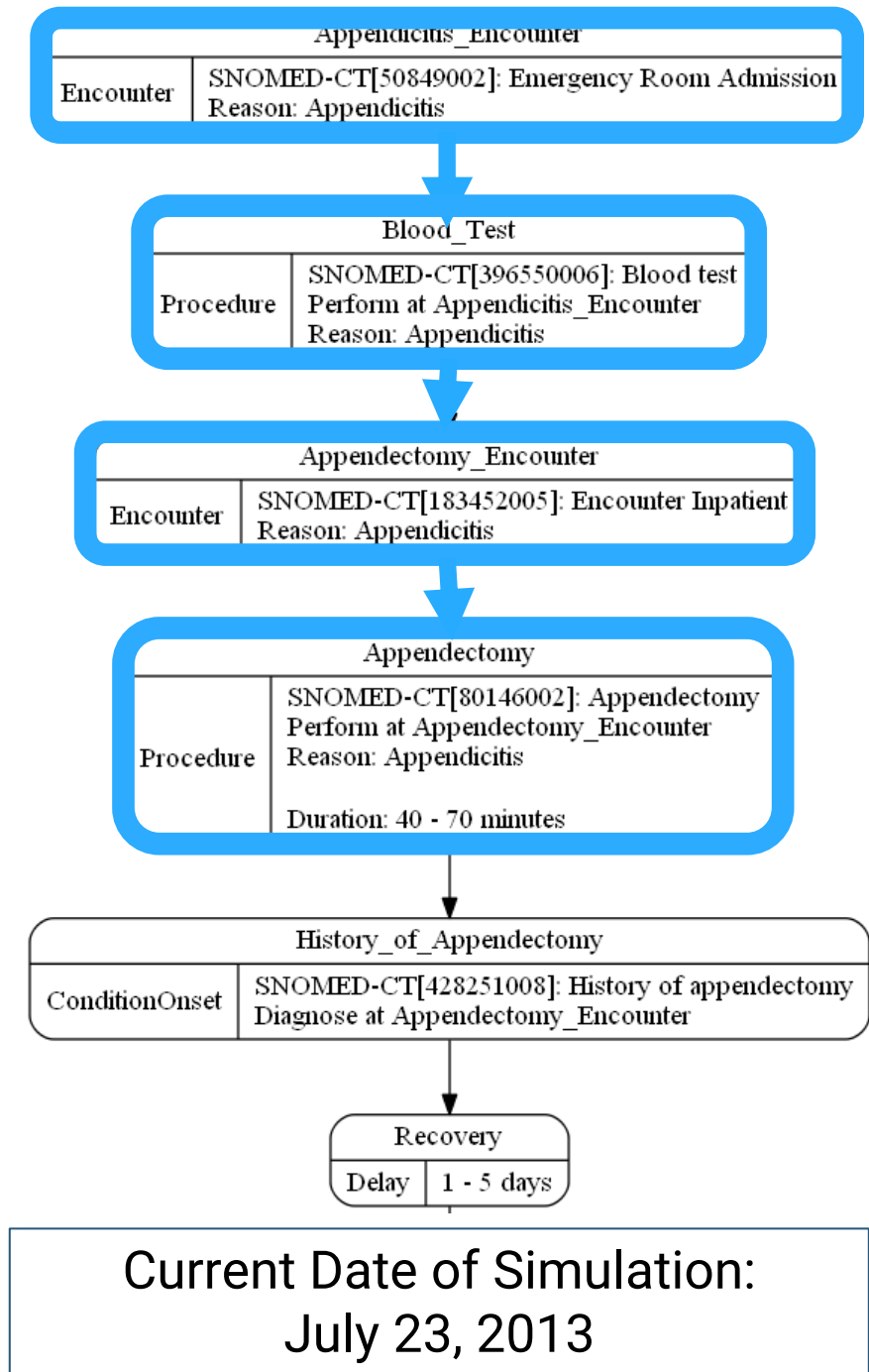
Conditions:

Appendicitis, Diagnosed July 23, 2013

Rupture of appendix, Diagnosed July 23, 2013

Procedures:

Blood Test, Performed July 23, 2013



Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

Emergency Room Admission,
July 23, 2013 → July 23, 2013

Inpatient Encounter,
July 23, 2013 →

Conditions:

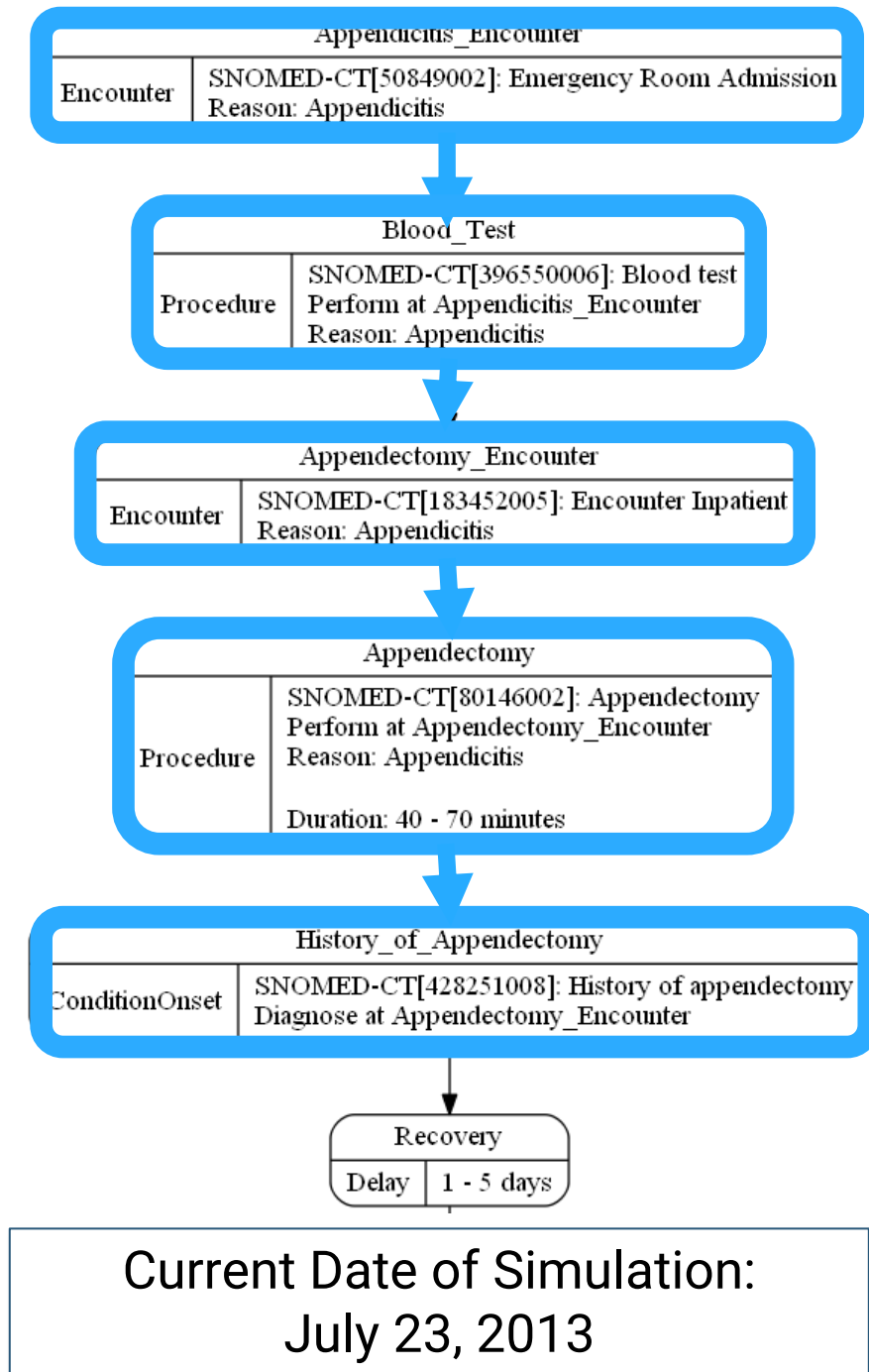
Appendicitis, Diagnosed July 23, 2013

Rupture of appendix, Diagnosed July 23, 2013

Procedures:

Blood Test, Performed July 23, 2013

Appendectomy, Performed July 23, 2013



Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

Emergency Room Admission,
July 23, 2013 → July 23, 2013

Inpatient Encounter,
July 23, 2013 →

Conditions:

Appendicitis, Diagnosed July 23, 2013

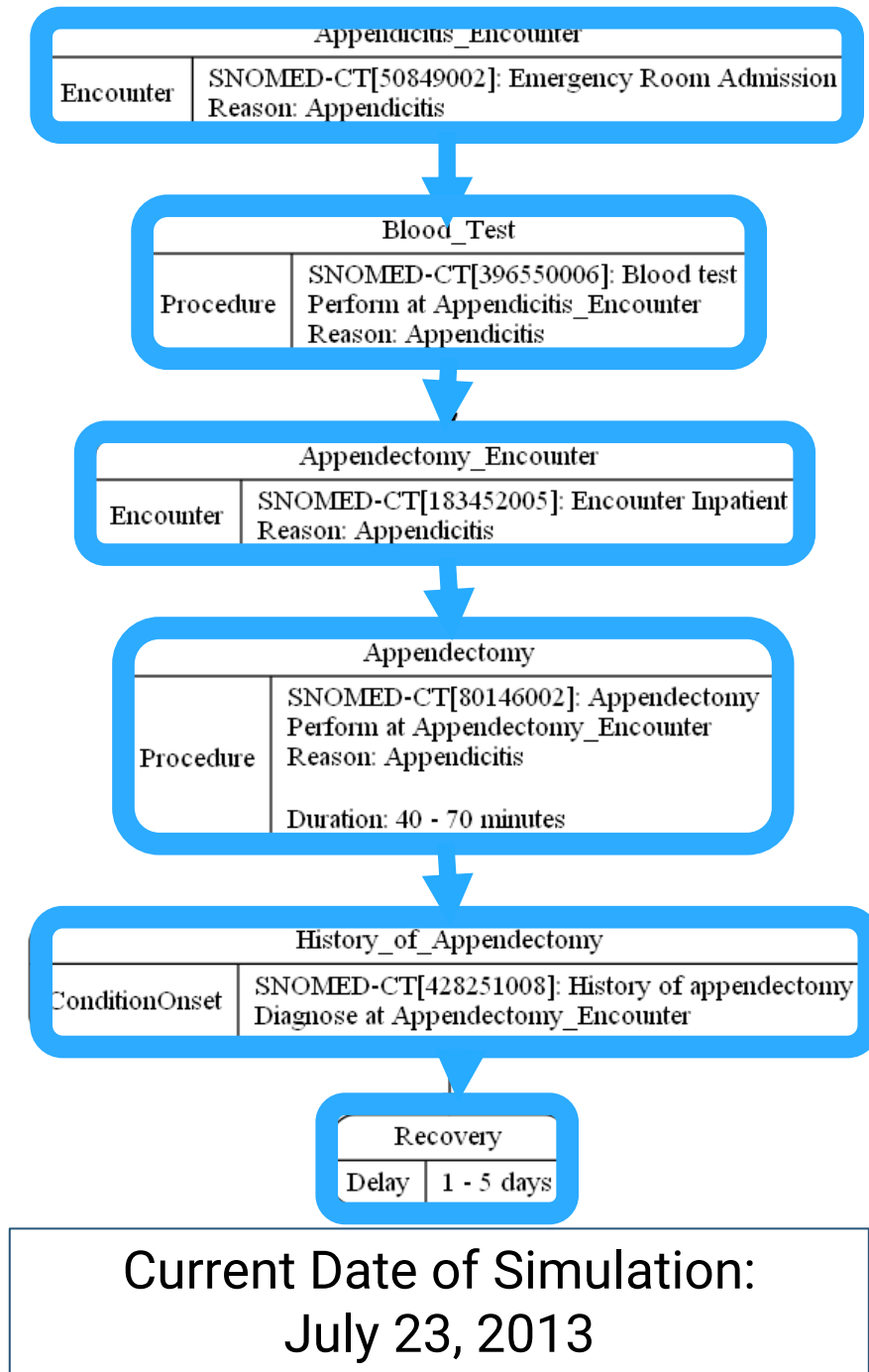
Rupture of appendix, Diagnosed July 23, 2013

History of appendectomy, Diagnosed July 23 2013

Procedures:

Blood Test, Performed July 23, 2013

Appendectomy, Performed July 23, 2013



Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

Emergency Room Admission,
July 23, 2013 → July 23, 2013

Inpatient Encounter,
July 23, 2013 →

Conditions:

Appendicitis, Diagnosed July 23, 2013

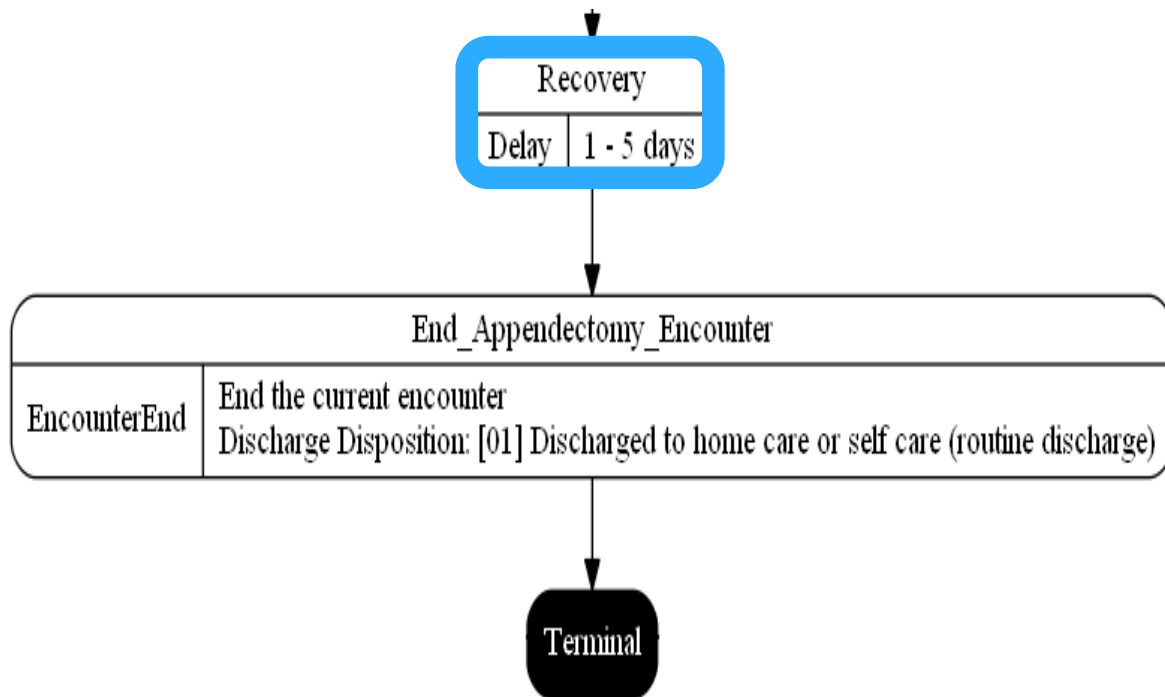
Rupture of appendix, Diagnosed July 23, 2013

History of appendectomy, Diagnosed July 23 2013

Procedures:

Blood Test, Performed July 23, 2013

Appendectomy, Performed July 23, 2013



Current Date of Simulation:
July 23, 2013

Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

Emergency Room Admission,
July 23, 2013 → July 23, 2013

Inpatient Encounter,
July 23, 2013 →

Conditions:

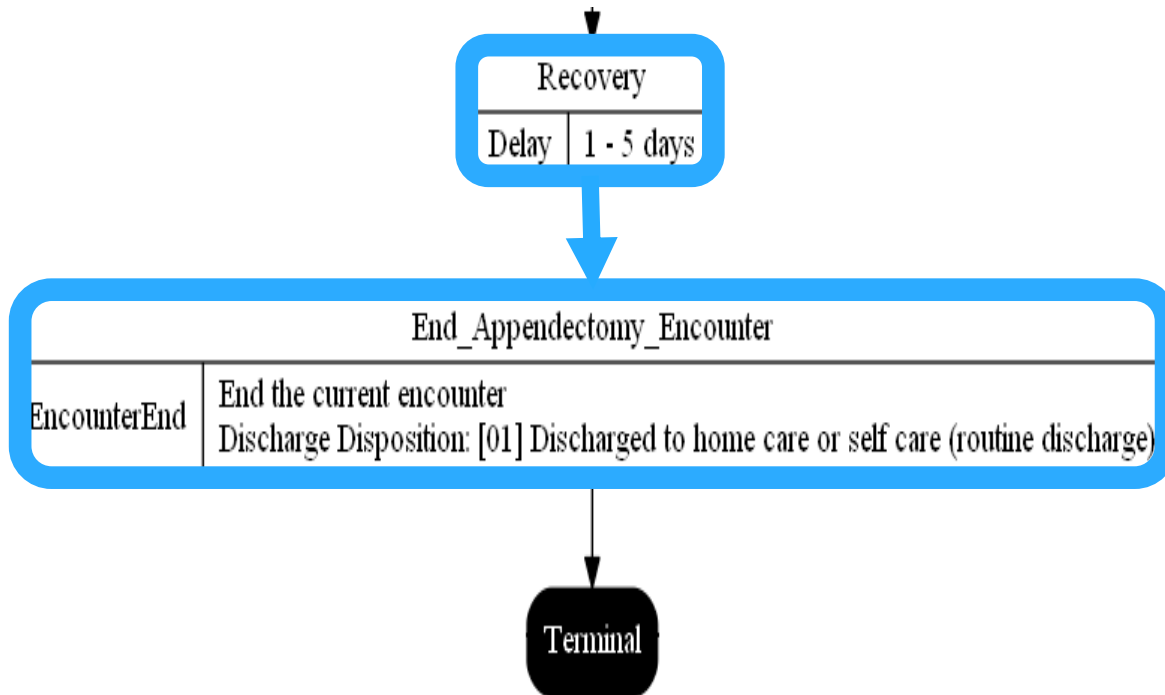
Appendicitis, Diagnosed July 23, 2013

Rupture of appendix, Diagnosed July 23, 2013

History of appendectomy, Diagnosed July 23 2013

Procedures:

Appendectomy, Performed July 23, 2013



Current Date of Simulation:
July 27, 2013

Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

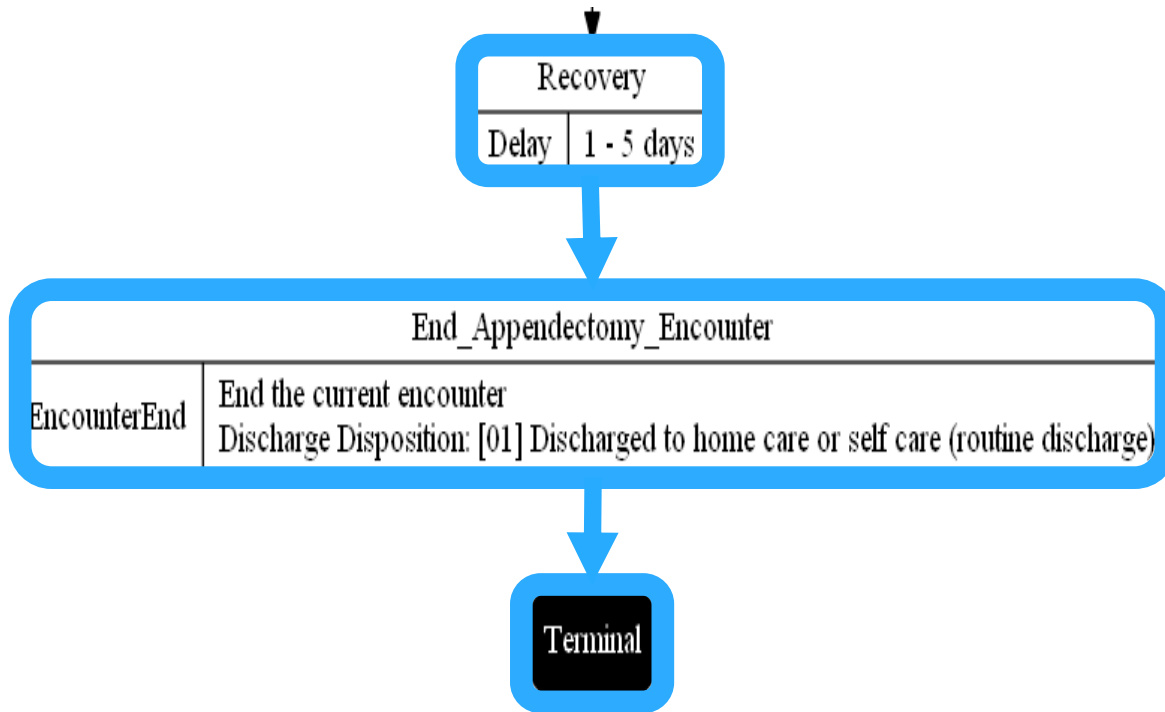
Emergency Room Admission,
 July 23, 2013 → July 23, 2013
 Inpatient Encounter,
 July 23, 2013 → **July 27, 2013**

Conditions:

Appendicitis, Diagnosed July 23, 2013
 Rupture of appendix, Diagnosed July 23, 2013
 History of appendectomy, Diagnosed July 23 2013

Procedures:

Appendectomy, Performed July 23, 2013



Current Date of Simulation:
July 27, 2013

Smith292, John949

Born: March 25, 1947

Gender: Male

Encounters:

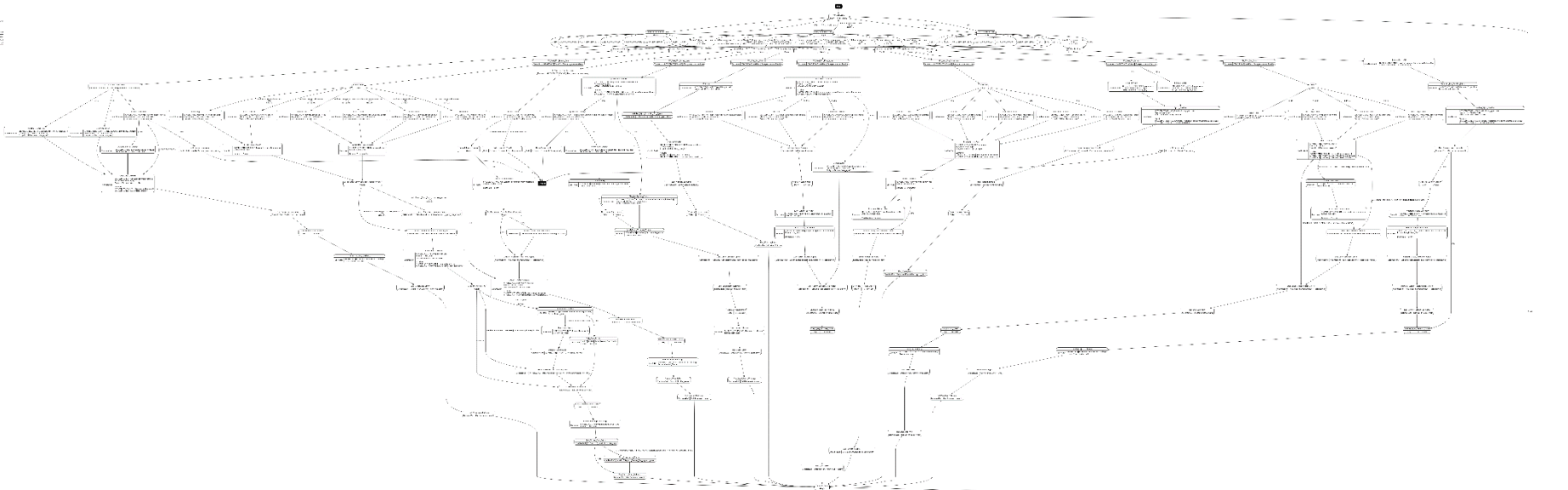
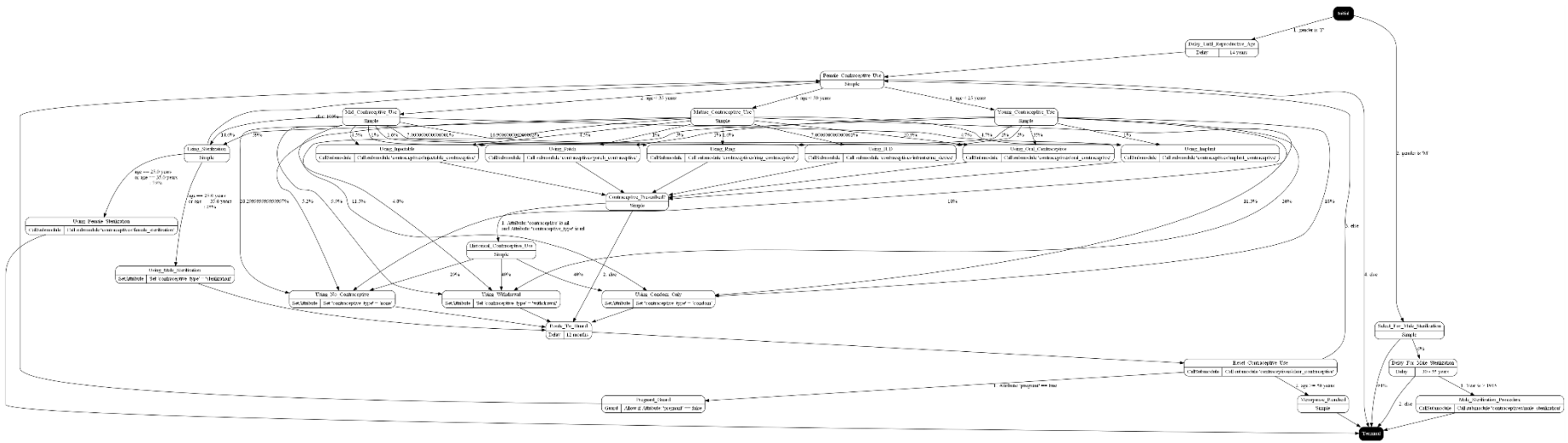
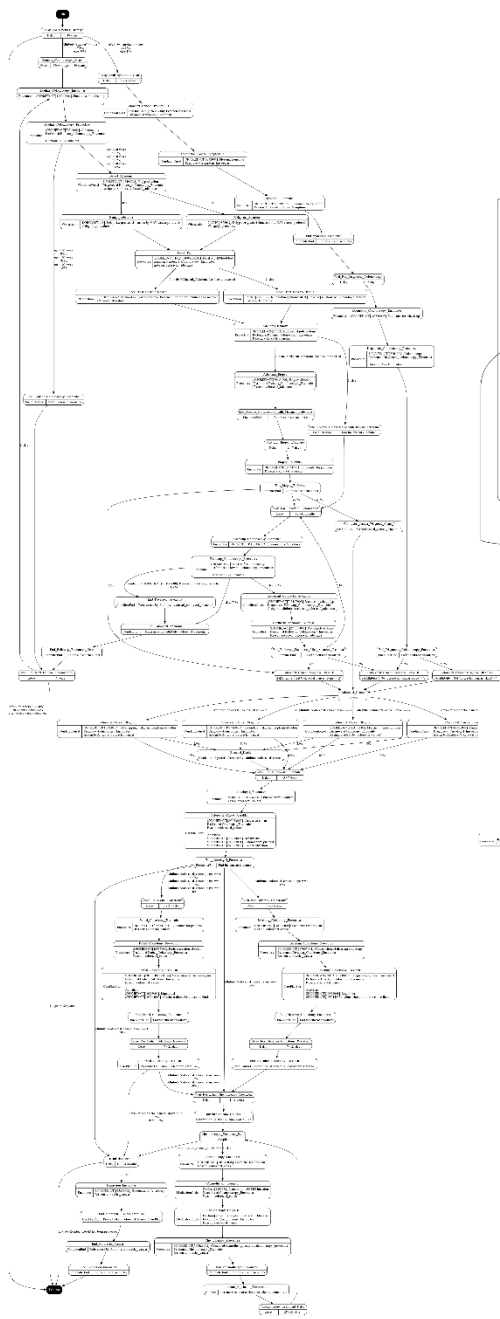
Emergency Room Admission,
July 23, 2013 → July 23, 2013
Inpatient Encounter,
July 23, 2013 → July 27, 2013

Conditions:

Appendicitis, Diagnosed July 23, 2013
Rupture of appendix, Diagnosed July 23, 2013
History of appendectomy, Diagnosed July 23 2013

Procedures:

Appendectomy, Performed July 23, 2013



Export Formats



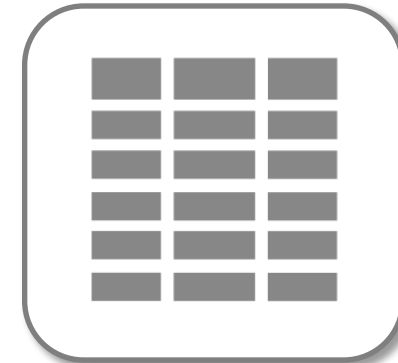
HL7® C-CDA®



HL7® FHIR®

A screenshot of a web-based data table with a blue header and yellow body. The table has multiple columns and rows of data, including headers like 'Patient ID', 'Patient Name', and 'Patient Address'. The data is presented in a clear, readable format.

Human Readable
HTML or
Plain Text



CSV



You tell us!

Easy to Set Up – Under an Hour

No Cost, No Restrictions

SyntheticMass

1,000,000 synthetic patients
available for download today
free of cost, free of restrictions

<https://syntheticmass.mitre.org>

Source of Data:

Geographic region:

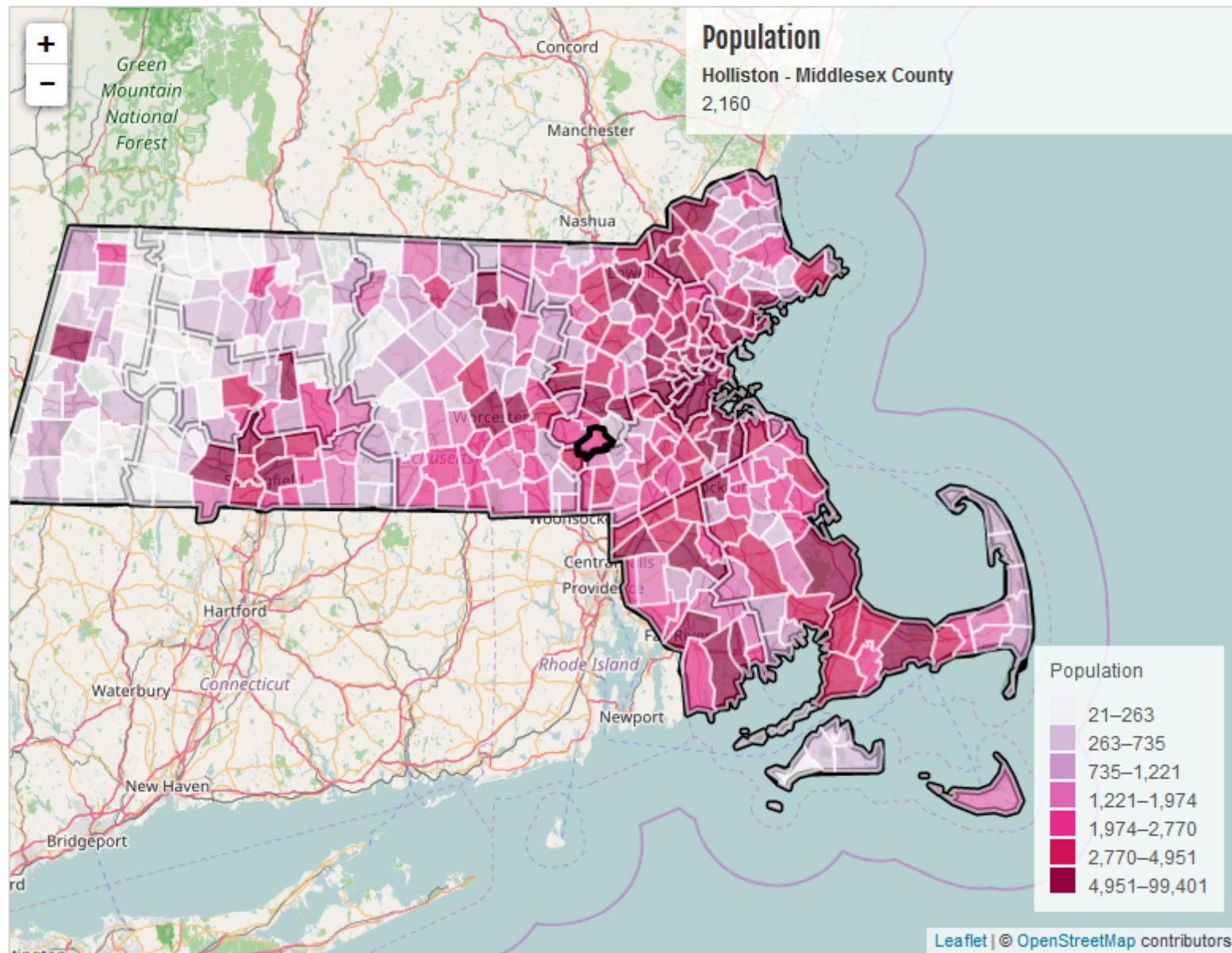
Data Value:

[Zoom map to all](#)

Population

Number of Residents

Region Type	Cities and Towns
Data Set	Synthetic data generated from Synthea
Total Population	1,009,150
Mean	2,875
Max	Boston Cities and Towns: 99,401
Min	Gosnold Cities and Towns: 21



[Sort by Population](#) | [Sort by name](#)

Source of Data: Synthea

Geographic region: Cities and Towns

Data Value: Population

[Zoom map to all](#)

Bedford

× Close

County	Middlesex
Population	2,106
Population Density	154.2 (per mi ²)
Area	14 sq. mi. Zoom

Demographics

Female Population	50.6%	(151 of 351)
Male Population	49.4%	(201 of 351)
Diabetes Prevalence	6.2%	(313 of 351)
Opioid Addiction Prevalence	0.9%	(124 of 351)
Heart Disease Prevalence	4.9%	(20 of 351)

Name	Gender	DOB
Terry123, Bessie159	female	03.Jun.1918
Hudson573, Colleen618	female	19.Dec.1921
Herman43, Hettie503	female	25.Apr.1939
O'Reilly259, Aidan944	female	21.Dec.1944
Aufderhar874, Carlee265	female	26.Sep.1951
Skiles93, Julianne73	male	07.May.1952
Kualis486, Darwin648	male	18.Mar.1957
Gorczyan106, Kali960	female	16.Feb.1958

Patient Record

× Close



Family name: Kuvalis486
 Given name: Darwin648
 Address: 6057 Stoltenberg Throughway
 City, State: Bedford, MA
 Postal Code: 01730

[Download Patient Data \(FHIR JSON\)](#) | [Download Patient Data \(CCDA XML\)](#)
[Send Data via Direct Message](#)

Height	187.09 cm	DOB	18.Mar.1957
Weight	128.39 kg	Age	60
Blood Type	n/a	Gender	male
Vision	n/a	Race	White
		Ethnicity	Nonhispanic
		Spoken language	n/a

[Observations](#) [Conditions](#) [Medications](#) [Allergies](#) [Care Plans](#) [More](#)

Conditions	Date of Onset	Date Resolved
Prediabetes	03.Jan.1991	n/a
Diabetes	04.Nov.1993	n/a
Diabetic retinopathy associated with type II diabetes mellitus (disorder)	03.Dec.1999	n/a
Nonproliferative diabetic retinopathy due to type 2 diabetes mellitus (disorder)	03.Dec.1999	n/a

How are people using Synthea?

Software Development & Integration

HL7 FHIR Connectathons since May 2017

PULSE@MassChallenge

IHIC Datathon, Athens October 2017

FHIR DevDays, Boston June 2018

Education

Risk-free data for students and researchers in clinical informatics

Data Analysis

Model validation – performing analytics to compare synthetic data to real data
Building tooling & models before moving to real data

Performance Testing

Use Synthea to generate hundreds, thousands, or even millions of synthetic patient records, to test system performance at scale

Citations of Synthea JAMIA Paper

<https://doi.org/10.1093/jamia/ocx079>

PhUSE 2017

Paper RW04

Use of Fast Healthcare Interoperability Resources (FHIR) in the Generation of Real World Evidence (RWE)

Regina Zopf, FDA, Bethesda, MD, USA
Jeff Abolafia, Rho, Chapel Hill, NC, USA
Bhargava Reddy, UCB, Raleigh, NC USA

ABSTRACT

There is a lot of interest in the clinical trial community to understand what information (EHRs) to support clinical trials. The use of FHIR has been endorsed by Information Technology (ONC) and is widely being used by EHR vendors.

A pilot was conducted to assess whether data that are typically of interest to patients can be extracted from medical records through FHIR resources to support clinical trials. Specifically, we assessed data in the [SyntheticMass](#) [Synthea repository](#) that includes diabetes diagnosis, medical history, concomitant medications, vital signs and

Deliverables from this pilot include a sample annotated Case Report Form (CRF) Harmonization (CDASH), Study Data Tabulation Model (SDTM) and FHIR R4 summary of our process followed and pilot experience, including challenges

Influence of simulation on electronic health record use patterns among pediatric residents

[Evan W Orenstein](#) ✉, [Irit R Rasooly](#), [Mark V Mai](#), [Adam C Dziorny](#), [Wanczyk Phillips](#), [Levon Utidjian](#), [Anthony Luberti](#), [Jill Posner](#), [Rebecca Tenney-Soeiro](#), [Chris P Bonafide](#)

Journal of the American Medical Informatics Association, ocy105,




<https://doi.org/10.1093/jamia/ocy105>

Published: 21 August 2018 [Article history](#) ▼

Ethical Issues in Secondary Use of Personal Health Information.

May 2018

Project: [Electronic Health Information Security and Privacy](#)

 Thomas Gallagher ·  Kudakwashe Dube ·  Scott Mclachlan

Characterizing Allegheny County Opioid Overdoses with an Interactive Data Explorer and Synthetic Prediction Tool

Theresa Gebert
Department of Statistics and Data Science
Carnegie Mellon University
theresa@stat.cmu.edu

Shuli Jiang
Department of Computer Science
Carnegie Mellon University
shulij@andrew.cmu.edu







Jiaxian Sheng
Department of Computer Science
Carnegie Mellon University
jiaxians@andrew.cmu.edu

Evaluating Text Analytic Frameworks for Mental Health Surveillance

Benjamin Mayer, Josh Arnold, Edmon Begoli, Everett Rush, Michael Drewry
Oak Ridge National Laboratory (ORNL)
Oak Ridge, Tennessee, USA
{mayerbw,arnoldjr,begolie,rusheniii,drewrymf}@ornl.gov




Kris Brown, Eduardo Ponce, Sudarshan Srinivasan
Electrical Engineering and Computer Science (EECS)
University of Tennessee, Knoxville
Circle Park Dr., Knoxville, TN 37996
{kbrown42,eponcemo,ssriniv3}@utk.edu

Teaching data science fundamentals through realistic synthetic clinical cardiovascular data

 Ted Laderas,  Nicole Vasilevsky,  Bjorn Pederson,  Melissa Haendel,  Shannon McWeeney,  David Dorr

doi: <https://doi.org/10.1101/232611>

Fast and simple comparison of semi-structured data, with emphasis on electronic health records

 Max Robinson, Jennifer Hadlock, Jiyang Yu, Alireza Khatamian, Aleksandr Y Aravkin,  Eric W Deutsch, Nathan D Price, Sui Huang,  Gustavo Glusman

doi: <https://doi.org/10.1101/293183>

OSEHRA Synthetic Patient Data Project Group



"This group will develop an open source toolset for generating clinically valid synthetic patient data and loading it into VistA (and potentially other healthcare IT products). Existing products and services such as Synthea and MiHIN will be surveyed, and the group will collaborate on requirements and specifications."

- Chair: Bo Dagnall, Perspecta

<https://www.osehra.org/groups/synthetic-patient-data-project-group>

Recent Additions

- **Anemia Module from OSEHRA Working Group**
- **Patient Provider Selection Behavior**
 - Nearest, Quality, Random, Network
- **Weight Loss Module**
- **Split Records by Provider**
- **Run Modules in Isolation**
- **Vital Signs Generator**
- **Individual Clinicians**
- **FHIR R4 Support**
- **Numerous Bug Fixes**

What's Next?

Join our open source community!

Submit a Pull Request

Report an Issue

Request a Feature

Let us know how you use Synthea!

Contact

- **Dylan Hall, Software Engineer**
- **Jason Walonoski, Project Lead**

dehall@mitre.org

jwalonoski@mitre.org

<https://synthetichealth.github.io/synthea/>



MITRE

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our federally funded R&D centers and public-private partnerships, we work across government to tackle challenges to the safety, stability, and well-being of our nation.

Learn more www.mitre.org

