# Empirical inference of Underlying Condition Probabilities Using Synthea-Generated Synthetic Health Data

Team TeMa

Dr. Michael D. Teter
miketeter@yahoo.com

Dr. Christopher E. Marks
cemarks@alum.mit.edu

Challenge Category: II (Novel Uses of Synthea Generated Data)

# Background

Problem Motivation

- Simulation is often used to investigate complicated phenomena for which analytic determination of outcome probabilities is intractable.
- Synthea is built in a way that makes it well-suited for this purpose.
  - It inputs conditional probabilities that can be validated.
  - Its outputs are the result of tailorable combinations of these input probabilities.

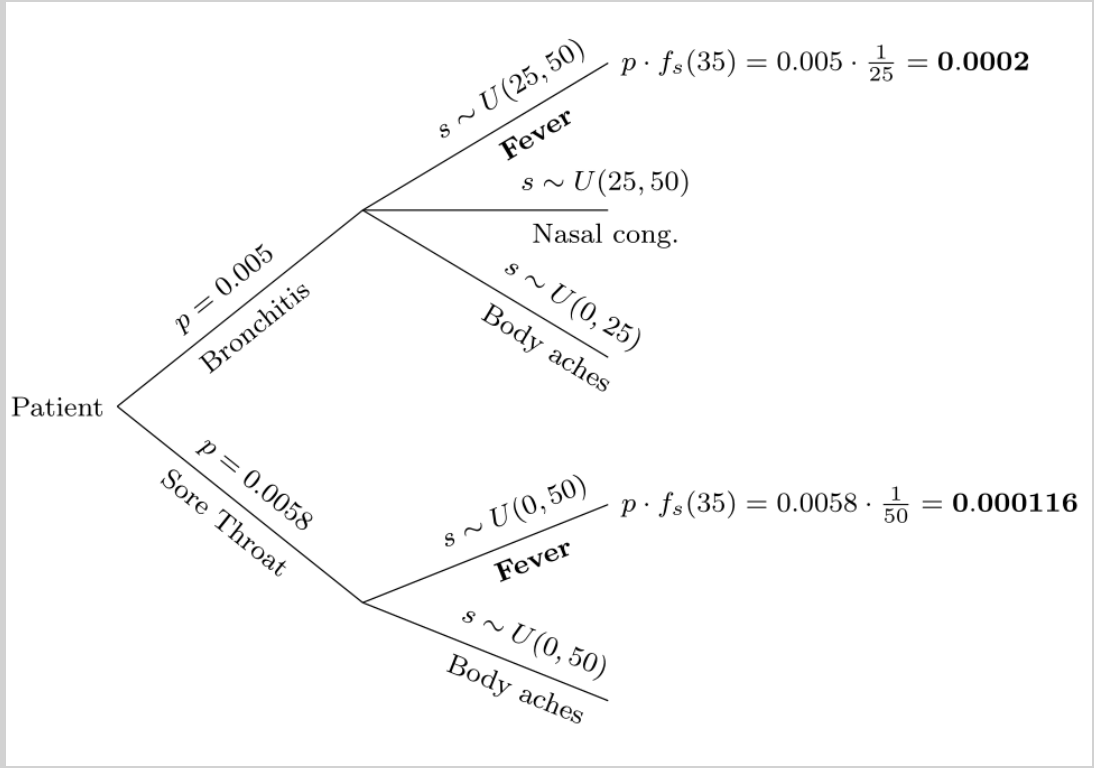
Our Task

- Use Synthea-generated data to investigate relationships between a patient's pathology and a given set of symptoms and severities.

# Our Methods (1 of 2)

## Empirical Bayes

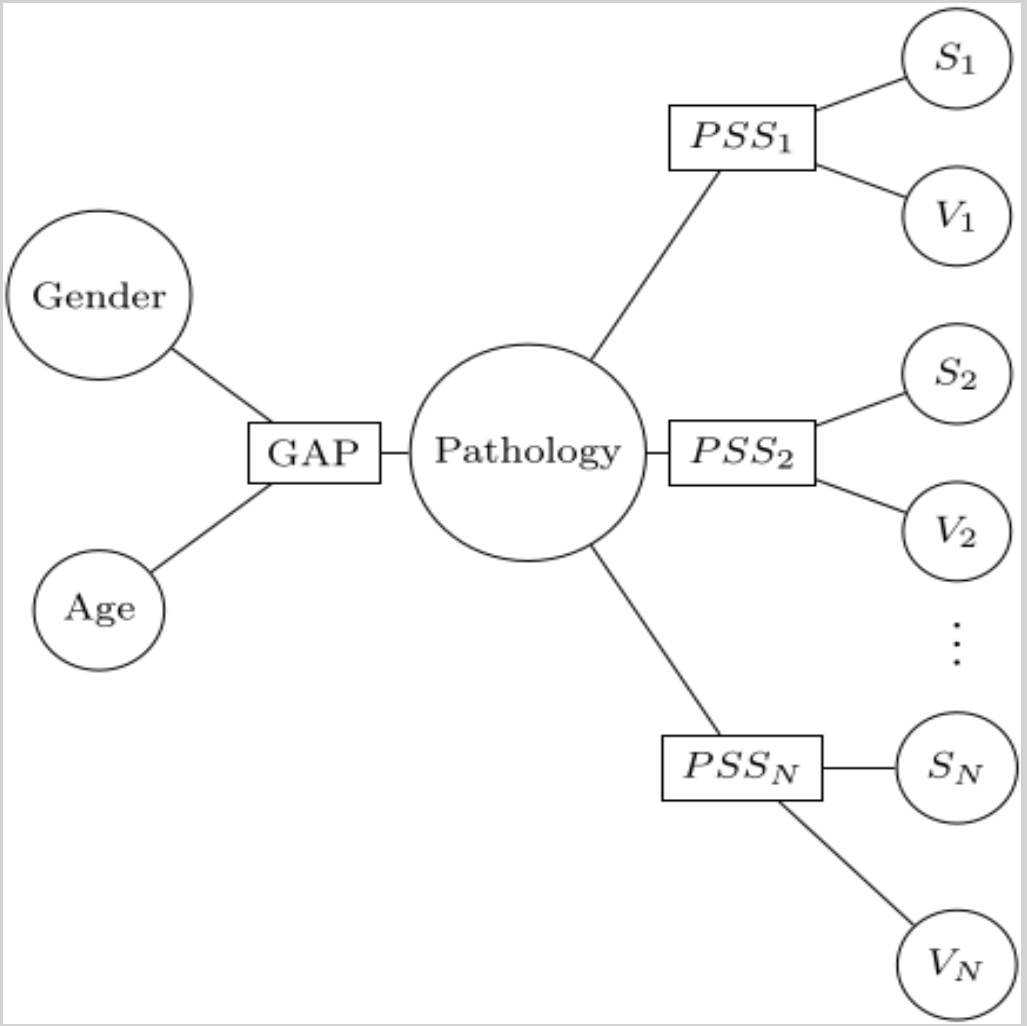
- Canonical Bayesian analysis
- Using empirical distributions in Synthea data, no need to enumerate a complicated tree.



# Our Methods (2 of 2)

## Bayesian Network

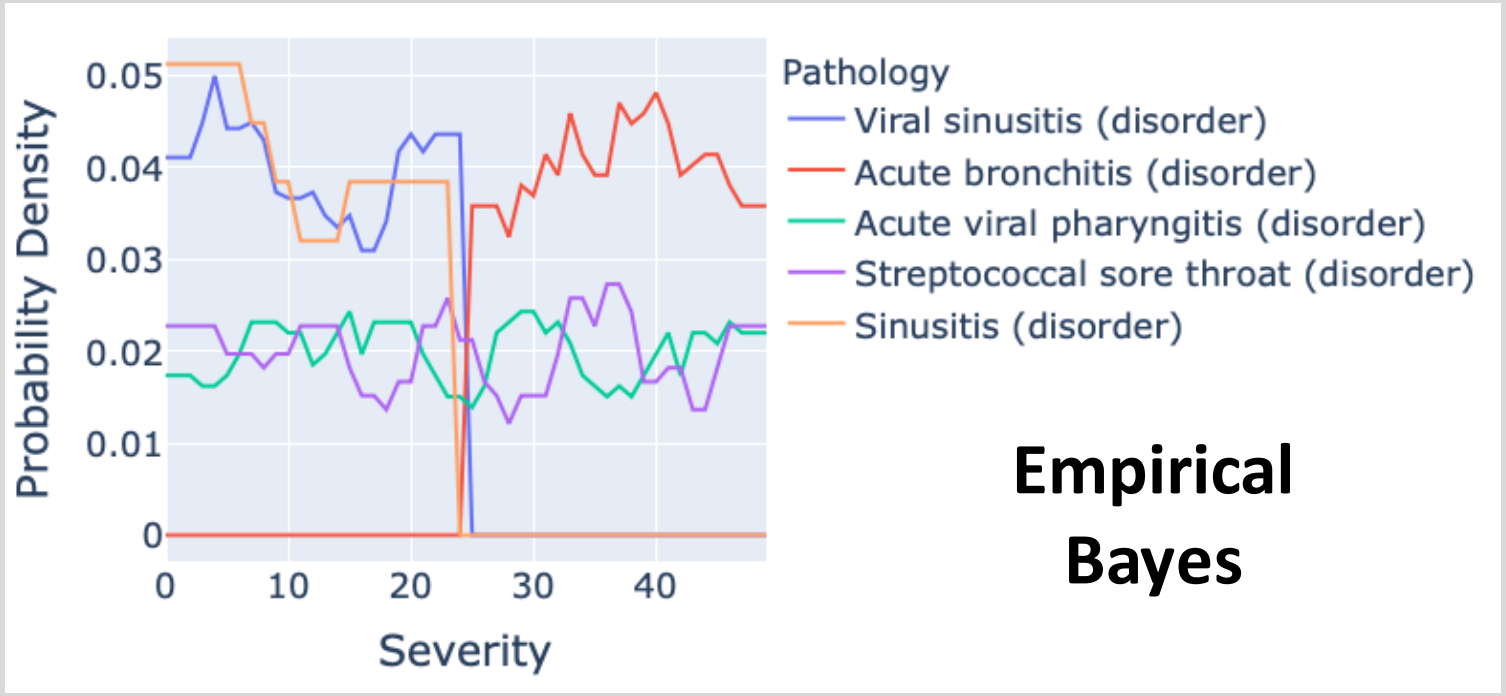
- Graph-based machine learning method.
- We decide which variables are related.
- More versatile than the strictly empirical model.



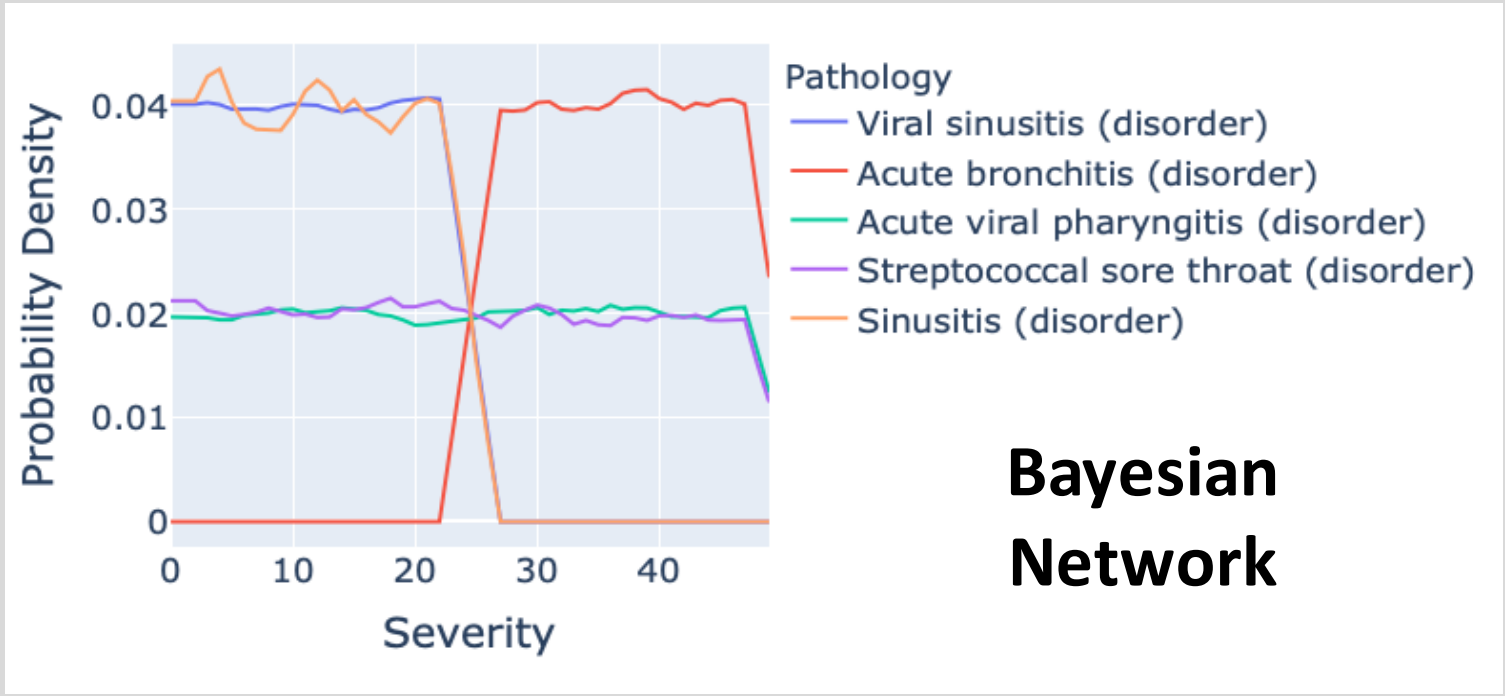
# Example Outputs

Patient

- 5 years old
- Female
- Fever




**Empirical Bayes**




**Bayesian Network**

# Validation

- Internal Validation through code testing
- External Validation through comparison with existing tools.
- A useful method for validating Synthea!

External validation: comparison with WebMD

| WebMD | Synthea Bayes |
|---|---|
| Bacterial Pneumonia | Viral sinusitis (disorder) |
| Middle Ear Infection | Acute bronchitis (disorder) |
| Viral Pneumonia | Acute viral pharyngitis (disorder) |
| Influenza (Flu) Child | Streptococcal sore throat (disorder) |
| Strep Throat | Sinusitis (disorder) |

# Summary & Next Steps

- Our approach can be extended to account for demographics, encounter types, patient location, patient history, etc.
- Compare results to known distributions as a way of **validating Synthea.**
- Identify areas where Synthea can be improved.
  - Standardization!
- Look for real-world applications.
  - Rare pathologies?