Step 1

Define diagnosis entities and perform data extraction

A Construct an inventory of diagnoses to study

HXX.XXX
HYY.YYY
HZZ.ZZZ
} diagnosis_entity_name

B Extract monthly counts of patients observed with each diagnosis entity

Step 2

Estimate counterfactual expectations of monthly utilization levels for each diagnosis entity

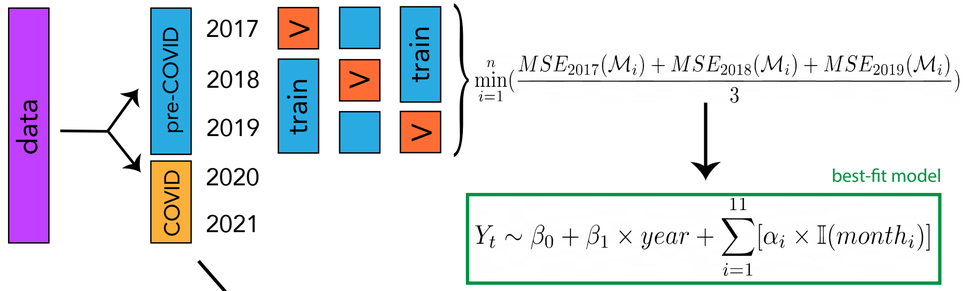
A Declare candidate counterfactual model specifications $\mathcal{M}_1 \ldots \mathcal{M}_n$

$$Y_t \sim \beta_0 + \beta_1 \times year + \sum_{i=1}^{11}[\alpha_i \times \mathbb{I}(month_i)]$$

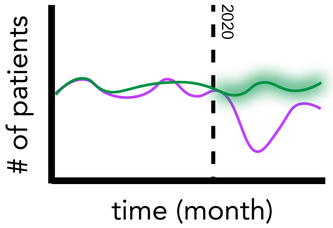
$$\vdots$$

$$Y_t \sim \beta_0 + \beta_1 \times year + \sum_{i=1}^{11}[\alpha_i \times \mathbb{I}(month_i)] + \beta_2 \sin\frac{2\pi t}{12} + \beta_3 \cos\frac{2\pi t}{12} + \beta_4 \sin\frac{2\pi t}{3} + \beta_5 \cos\frac{2\pi t}{3}$$

B Select the best-fitting model on pre-pandemic data using leave-out-one-year cross validation



C Generate counterfactual predictions for the pandemic period (2020-2021) using the best-fit model and a distribution that accounts for the dispersion of the training data to estimate 95% prediction intervals



Step 3

Calculate deviations from expectation

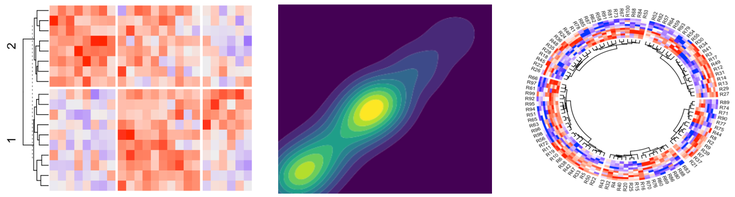
A Exclude diagnosis entities with poor out-of-sample counterfactual model performance (RMSPE ≥ 12.5%)

B Calculate monthly deviations from expectation δ for all remaining diagnosis entities

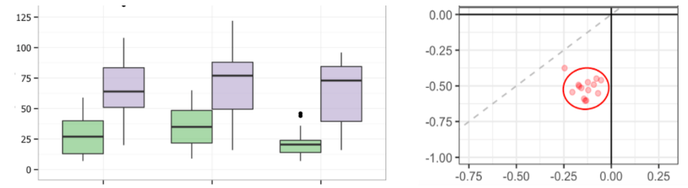
$$\delta = \frac{observed - expected}{expected}$$

Step 4

Identify factors associated with care utilization patterns



Pattern detection using data visualization and cluster analysis



Hypothesis testing using external indicies of disease severity