

Graham Casey Gibson, Dan Sheldon, Evan Ray, Nicholas Reich

March 22, 2020

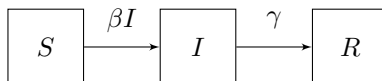
## 1 Introduction

COVID-19 has become a worldwide pandemic that has effected all continents, causing over 308,000 cases and 13,000 deaths. The viral agent (SARS-CoV-2) is extremely contagious with an estimate of  $R_0$  between 2.5 and 3.5. [?] With the arrival of the virus in the U.S., significant effort has been made to model the spread of the disease.

We see this effort as having two distinct endpoints. The first is to model the number of COVID-19 cases in the population and the second is to model the effects of COVID-19 on influenza surveillance networks such as ILInet. Since 2015, the CDC has been running an influenza-like-illness (defined by fever  $\geq 100$  and cough or sore throat, referred to as ILI) prediction initiative. The goal is to forecast the percent of patients who meet the symptomatic criteria, out of the patients who present to the network. Since the outbreak of COVID-19 in the U.S., efforts have shifted to harnessing influenza surveillance networks to monitor the pandemic. Because of the symptom overlap, it is expected that ILI will pick up COVID-19 cases without requiring a confirmed positive test. This is advantageous for two reasons, first, the network is already in place with a defined network of hospitals and providers reporting, and second, the lack of available testing means waiting for confirmed cases may delay an accurate estimate of the true number of cases.

In this work, we consider two approaches to model the effects of COVID-19 on ILInet.

## 2 Explicit SIR Model



Thanks to the efforts of JHU ([? ]) we have available case counts in a time series for all 50 states. In particular, we have a time series of confirmed cases we call  $C_{i,j}$  where  $i$  indexes calendar time and  $j$  indexes regions, in this case states.

We can model the confirmed cases using the SIR model as

$$C_{i,j} \sim N(I_{i,j}, \sigma^2) \tag{1}$$

$$\begin{bmatrix} S_{i,j} \\ I_{i,j} \\ R_{i,j} \end{bmatrix} = f \left( \begin{bmatrix} S_{i-1,j} \\ I_{i-1,j} \\ R_{i-1,j} \end{bmatrix} \middle| \beta_j, \gamma_j \right) \quad (2)$$

where  $f$  is the RK4 approximation to the differential equation parameterized by  $\beta, \gamma$ . We use,

$$\begin{aligned} I_0 &\sim \text{Unif}(0, 0.02) \\ \gamma &\sim \Gamma(k_\gamma, k_\gamma \hat{d}) \\ \beta &\sim \Gamma(k_\beta, k_\beta \hat{d} / \hat{R}) \\ \rho &\sim \text{Beta}(\kappa \hat{\rho}, \kappa(1 - \hat{\rho})) \end{aligned}$$

? These satisfy:  $\mathbb{E}[\gamma] = 1/\hat{d}$ ,  $\mathbb{E}[\beta] = \hat{R}/\hat{d}$ ,  $\mathbb{E}[\rho] = \hat{\rho}$ ? The user-selected parameters and current values are:

- $k_\gamma = k_\beta = 1$  shape parameters for  $\gamma, \beta$
- $\hat{d} = 10$  estimate of duration in infected compartment
- $\hat{R} = 2.2$  estimate of  $R_0$
- $\hat{\rho} = 0.3$  initial guess of detection rate
- $\kappa = 50$  concentration parameter for detection rate

## 2.1 Interventions

Currently in progress.

## 2.2 Incorporating ILI

In addition, we have the publicly available ILI data for the same calendar dates and regions, we call  $\text{ILI}_{i,j}$ . We consider this the sum of both non-COVID-19 case counts and COVID-19 case counts. That is

$$\text{ILI}_{i,j} = \frac{\text{ILI}_{i,j}^{\text{non-COVID-19 cases}} + C_{i,j}}{\text{ILI}_{i,j}^{\text{patients}}} \quad (3)$$

We have well developed process models for  $\frac{\text{ILI}_{i,j}^{\text{non-COVID-19 cases}}}{\text{ILI}_{i,j}^{\text{patients}}}$  that have been implemented in the FluSight challenge to predict ILI. However, we do not have well developed process models for  $\text{ILI}_{i,j}^{\text{non-COVID-19 cases}}$  and  $\text{ILI}_{i,j}^{\text{patients}}$  individually. In order to use our existing process models for influenza driven ILI we re-write (??) as

$$\text{ILI}_{i,j} = \frac{\text{ILI}_{i,j}^{\text{non-COVID-19 cases}}}{\text{ILI}_{i,j}^{\text{patients}}} + \frac{C_{i,j}}{\text{ILI}_{i,j}^{\text{patients}}} \quad (4)$$

Equation (??) allows us to use ILI process models to predict the first term. However, we still need to model  $ILI_{i,j}^{\text{patients}}$  for the second term. This model defines the likelihood over observed  $ILI_{i,j}$  and confirmed COVID-19 cases  $C_{i,j}$  for a given time-point and region  $j$  as:

$$\mathcal{L}(ILI_{i,j}, C_{i,j} | \beta, \gamma, ILI_{1:(i-1),*}, C_{1:(i-1),*}) = \mathcal{L}(ILI_{i,j} | ILI_{1:(i-1),*}) * \mathcal{L}(C_{i,j} | \beta, \gamma, C_{1:(i-1),*}) \quad (5)$$

which is the product of the likelihood over the observed  $ILI$  and the observed COVID-19 cases.

### 2.3 Confounding of ILI and COVID-19

As of March 2020, we expect  $ILI$  to be confounded between non-COVID-19 and COVID-19. That is, beyond this cutoff we do not observe  $ILI^{\text{non-COVID-19 cases}}$  but rather,  $ILI^{\text{non-COVID-19 cases}} + ILI^{\text{COVID-19 cases}}$ . In order to directly model  $ILI^{\text{non-COVID-19 cases}}$  we restrict the model to only use historical  $ILI$  data before March of 2020. We therefore use a kernel density estimate of  $ILI$  from historical seasons.

## 3 Latent SIR Model

A second option that avoid the complexity of modeling the expected number of patients who will present to the influenza surveillance network, is to treat the SIR curve as a latent variable without directly tying it to confirmed COVID-19 cases.