Estimating infection fatality rates and herd immunity levels through demographics

Miguel Sánchez-Romero, Vanessa di Lego, Alexia Prskawetz, and Bernardo L. Queiroz

IIASA-RANEPA Symposium, June 29, 2020
Objective/Policy relevant:

- How can demography help to monitor the evolution of the COVID-19 epidemic?
  Fraction of infected people
  Fraction of detected people
Motivation

- Objective/Policy relevant:
  - How can demography help to monitor the evolution of the COVID-19 epidemic?
    - Fraction of infected people
    - Fraction of detected people

- Technical issues:
  - Uncertainty in the infected data: Many asymptomatic individuals
  - Uncertainty in the mortality data: Under-reported and over-reported
Motivation

- **Objective/Policy relevant:**
  - How can demography help to monitor the evolution of the COVID-19 epidemic?
    - Fraction of infected people
    - Fraction of detected people

- **Technical issues:**
  - Uncertainty in the infected data: Many asymptomatic individuals
  - Uncertainty in the mortality data: Under-reported and over-reported

- **Methodology:** Age-structured SEIR model with COVID-19 mortality (*asymptomatic*) and non COVID-19 mortality
Motivation

- **Objective/Policy relevant:**
  - How can demography help to monitor the evolution of the COVID-19 epidemic?
    - Fraction of infected people
    - Fraction of detected people

- **Technical issues:**
  - Uncertainty in the infected data: Many asymptomatic individuals
  - Uncertainty in the mortality data: Under-reported and over-reported

- **Methodology:** Age-structured SEIR model with COVID-19 mortality (asymptomatic) and non COVID-19 mortality

- **Calibration strategy:** Bayesian Melding (Poole and Raftery, 2000)
  - Takes information from: Model Inputs, Model Outputs, Data
  - Calibrate the model to evolution of the total number of deaths in the province of Hubei (China)
  - Run a sample of 3000 models for each of region/country analyzed generated with the IMIS algorithm
In-sample performance of the SEIR model

Figure 1: In-sample model performance of the SEIR models. Death toll in Hubei (China). Source: Black data points taken from ECDC COVID-19 situation update worldwide, as of 12 May 2020. Notes: Data (black dots). Simulation results (blue lines) based on 3000 random sets of parameter inputs drawn from the posterior distribution of $\pi^{[\Theta]}(\Theta)$. The gray area depicts the time interval used in the Bayesian melding.
Why is it so difficult to model COVID-19?

<table>
<thead>
<tr>
<th>Effective transmission rate</th>
<th>Removal rate</th>
<th>Inverse incubation period</th>
<th>Fatality adjustment factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corr: 0.935</td>
<td>Corr: 0.102</td>
<td>Corr: −0.723</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Correlation matrix of the SEIR model inputs based on the posterior distribution of $\pi^\Theta(\Theta)$. 
Infection fatality rates

Figure 3: Posterior distribution of the SEIR model inputs $\pi^{[\Theta]}(\Theta)$. 
Infection fatality rates across US States

Old states (IFR > 1.2%): Florida, Maine, North Dakota, Montana, West Virginia
Young states (IFR < 0.8%): Utah, Alaska, Georgia, Texas, District of Columbia

Figure 4: Expected infection fatality rates across US states (in %)
Demographics and fraction of people infected

- Infection fatality rate at time $t$:

$$\text{IFR} = \frac{\text{Total deaths}_t}{\text{Total infected}_t} = \frac{\text{Total deaths}_t}{\text{Population}_t \times \left(\text{Probability of being infected}_t\right)}$$ (1)

- Fraction of people infected until time $t$:

$$\text{Probability of being infected}_t = \frac{\text{Total deaths}_t}{\text{Population}_t \times \text{IFR}}.$$
Demographics and fraction of people infected

- Infection fatality rate at time $t$:

\[
\text{IFR} = \frac{\text{Total deaths}_t}{\text{Total infected}_t} = \frac{\text{Total deaths}_t}{\text{Population}_t \times \left(\text{Probability of being infected}_t\right)}
\]  

(1)

- Fraction of people infected until time $t$:

\[
\text{Probability of being infected}_t = \frac{\text{Total deaths}_t}{\text{Population} \times \text{IFR}}.
\]  

(2)
Figure 5: Observed vs. estimated fraction of people infected. Note: The observed data includes information from seroprevalence studies conducted in 67 regions: 19 regions in Spain (green dots), Connecticut, Missouri, New York City, and the state of New York (blue dots), and 44 regions in Brazil (red dots).
Table 1: Expected fraction of people infected in selected US States (in %) by June 26th, 2020: Ordered according to the fraction of people infected

<table>
<thead>
<tr>
<th>#</th>
<th>State</th>
<th>Population</th>
<th>Deaths</th>
<th>Fraction infected</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New Jersey</td>
<td>8,872,797</td>
<td>14,914</td>
<td>16.1</td>
<td>(9.2, 55.9)</td>
</tr>
<tr>
<td>2</td>
<td>New York</td>
<td>19,428,618</td>
<td>31,342</td>
<td>15.3</td>
<td>(8.7, 52.9)</td>
</tr>
<tr>
<td>3</td>
<td>Massachusetts</td>
<td>6,887,571</td>
<td>8,012</td>
<td>11.1</td>
<td>(6.3, 38.5)</td>
</tr>
<tr>
<td>4</td>
<td>Connecticut</td>
<td>3,558,382</td>
<td>4,307</td>
<td>10.9</td>
<td>(6.2, 37.8)</td>
</tr>
<tr>
<td>5</td>
<td>District of Columbia</td>
<td>701,570</td>
<td>546</td>
<td>9.7</td>
<td>(5.5, 33.6)</td>
</tr>
<tr>
<td>6</td>
<td>Rhode Island</td>
<td>1,054,930</td>
<td>927</td>
<td>8.0</td>
<td>(4.5, 27.6)</td>
</tr>
<tr>
<td>7</td>
<td>Michigan</td>
<td>9,983,147</td>
<td>6,134</td>
<td>5.8</td>
<td>(3.3, 20.2)</td>
</tr>
<tr>
<td>8</td>
<td>Illinois</td>
<td>12,644,110</td>
<td>6,847</td>
<td>5.5</td>
<td>(3.1, 18.9)</td>
</tr>
<tr>
<td>9</td>
<td>Maryland</td>
<td>6,010,630</td>
<td>3,142</td>
<td>5.3</td>
<td>(3.0, 18.4)</td>
</tr>
<tr>
<td>10</td>
<td>Delaware</td>
<td>970,344</td>
<td>507</td>
<td>4.7</td>
<td>(2.7, 16.2)</td>
</tr>
<tr>
<td>11</td>
<td>Pennsylvania</td>
<td>12,795,687</td>
<td>6,579</td>
<td>4.5</td>
<td>(2.6, 15.7)</td>
</tr>
<tr>
<td>12</td>
<td>Indiana</td>
<td>6,729,379</td>
<td>2,595</td>
<td>4.0</td>
<td>(2.3, 13.8)</td>
</tr>
<tr>
<td>13</td>
<td>Colorado</td>
<td>5,722,795</td>
<td>1,673</td>
<td>3.3</td>
<td>(1.9, 11.6)</td>
</tr>
<tr>
<td>14</td>
<td>Virginia</td>
<td>8,420,245</td>
<td>1,700</td>
<td>2.1</td>
<td>(1.2, 7.2)</td>
</tr>
<tr>
<td>15</td>
<td>Iowa</td>
<td>3,153,268</td>
<td>704</td>
<td>2.1</td>
<td>(1.2, 7.2)</td>
</tr>
<tr>
<td>16</td>
<td>Kentucky</td>
<td>4,450,886</td>
<td>553</td>
<td>1.3</td>
<td>(0.7, 4.3)</td>
</tr>
</tbody>
</table>

Source: Authors' calculations using data from https://coronavirus.jhu.edu/us-map, United States Mortality DataBase, and US Census Bureau.
Demographics and fraction of detected people

- Fatality rates (IFR and CFR) until time $t$:

$$\text{IFR} = \frac{\text{Total deaths}_t}{\text{Total infected}_t} = \frac{\text{Total deaths}_t}{\text{Population}_t \times \left(\text{Probability of being infected}_t\right)}$$  \hspace{1cm} (3)

$$\text{CFR} = \frac{\text{Total deaths}_t}{\text{Total infected and detected}_t} = \frac{\text{Total deaths}_t}{\text{Population}_t \times \left(\text{Probability of being infected}_t\right) \times \left(\text{Probability of being detected}_t\right)}$$  \hspace{1cm} (4)

$$\text{Fraction detected}_t = \frac{\text{Total infected and detected}_t}{\text{Total infected}_t} = \frac{\text{IFR}}{\text{CFR}}.$$  \hspace{1cm} (5)
Demographics and fraction of detected people

- Fatality rates (IFR and CFR) until time $t$:

  \[
  \text{IFR} = \frac{\text{Total deaths}_t}{\text{Total infected}_t} = \frac{\text{Total deaths}_t}{\text{Population}_t \times \left(\text{Probability of being infected}_t\right)}
  \]

  \[
  \text{CFR} = \frac{\text{Total deaths}_t}{\text{Total infected and detected}_t} = \frac{\text{Total deaths}_t}{\text{Population}_t \times \left(\text{Probability of being infected}_t\right) \times \left(\text{Probability of being detected}_t\right)}
  \]

- Fraction of people detected until time $t$:

  \[
  \text{Fraction detected}_t = \frac{\text{Total infected and detected}_t}{\text{Total infected}_t} = \frac{\text{IFR}}{\text{CFR}}.
  \]
**Table 2:** Expected fraction of infected individuals who are detected in the US (in %) by June 26th, 2020: Ordered according to the fraction of people detected.

<table>
<thead>
<tr>
<th>#</th>
<th>State</th>
<th>IFR</th>
<th>CFR</th>
<th>Fraction detected</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Iowa</td>
<td>1.07</td>
<td>2.54</td>
<td>42.20</td>
<td>(12.19, 74.19)</td>
</tr>
<tr>
<td>2</td>
<td>Virginia</td>
<td>0.97</td>
<td>2.81</td>
<td>34.53</td>
<td>(9.97, 60.71)</td>
</tr>
<tr>
<td>3</td>
<td>Kentucky</td>
<td>0.99</td>
<td>3.72</td>
<td>26.61</td>
<td>(7.69, 46.79)</td>
</tr>
<tr>
<td>4</td>
<td>Delaware</td>
<td>1.12</td>
<td>4.60</td>
<td>24.29</td>
<td>(7.01, 42.70)</td>
</tr>
<tr>
<td>5</td>
<td>Maryland</td>
<td>0.98</td>
<td>4.75</td>
<td>20.65</td>
<td>(5.96, 36.30)</td>
</tr>
<tr>
<td>6</td>
<td>Illinois</td>
<td>0.99</td>
<td>4.88</td>
<td>20.35</td>
<td>(5.88, 35.79)</td>
</tr>
<tr>
<td>7</td>
<td>Rhode Island</td>
<td>1.10</td>
<td>5.56</td>
<td>19.82</td>
<td>(5.72, 34.84)</td>
</tr>
<tr>
<td>8</td>
<td>Colorado</td>
<td>0.87</td>
<td>5.26</td>
<td>16.61</td>
<td>(4.80, 29.21)</td>
</tr>
<tr>
<td>9</td>
<td>Indiana</td>
<td>0.97</td>
<td>5.88</td>
<td>16.47</td>
<td>(4.76, 28.95)</td>
</tr>
<tr>
<td>10</td>
<td>Pennsylvania</td>
<td>1.14</td>
<td>7.40</td>
<td>15.33</td>
<td>(4.43, 26.96)</td>
</tr>
<tr>
<td>11</td>
<td>District of Columbia</td>
<td>0.80</td>
<td>5.36</td>
<td>14.95</td>
<td>(4.32, 26.29)</td>
</tr>
<tr>
<td>12</td>
<td>Massachusetts</td>
<td>1.05</td>
<td>7.41</td>
<td>14.10</td>
<td>(4.07, 24.79)</td>
</tr>
<tr>
<td>13</td>
<td>New York</td>
<td>1.06</td>
<td>8.01</td>
<td>13.18</td>
<td>(3.81, 23.17)</td>
</tr>
<tr>
<td>14</td>
<td>New Jersey</td>
<td>1.04</td>
<td>8.74</td>
<td>11.91</td>
<td>(3.44, 20.94)</td>
</tr>
<tr>
<td>15</td>
<td>Michigan</td>
<td>1.05</td>
<td>8.85</td>
<td>11.88</td>
<td>(3.43, 20.88)</td>
</tr>
<tr>
<td>16</td>
<td>Connecticut</td>
<td>1.11</td>
<td>9.35</td>
<td>11.86</td>
<td>(3.43, 20.86)</td>
</tr>
</tbody>
</table>

Higher percentage of people detected in rural areas

\[ \hat{y} = 0.111 + 0.489 \times x, \ R^2 = 0.37 \]

Figure 6: Relationship between percentage detected and percentage living in rural areas. US states by June 26th, 2020
Conclusions

- There are too many models capable of replicating the observed pattern of deaths → Best strategy Bayesian Melding

- The IFR in the US varies from 0.7% to 1.2% depending on the age structure of the population

- The fraction of people ever infected of COVID-19 can be indirectly estimated → In the US the most hit states have a fraction of infected people close to 15%, which is far from herd immunity levels

- The fraction of people detected depends on how densely populated the region. However, there is region that has detected more than 50% of the total infected people.
Thank you for your attention

The model details is based on a previous version of this paper. “How many lives can be saved? A global view on the impact of testing, herd immunity and demographics on COVID-19 fatality rates”. Miguel Sanchez-Romero, Vanessa di Lego, Alexia Prskawetz, Bernardo L Queiroz. medRxiv 2020.04.29.20084400; doi: https://doi.org/10.1101/2020.04.29.20084400