Modeling COVID-19

What is the current state of the art in modeling?

On March 26th 2020 the Health Service Utilization Forecasting Team from the Institute for Health Metrics and Evaluation (IHME) at the University of Washington published new models for the utilization of health resources. These models have been used extensively by the media and have been included in recent White House press briefings.

Fundamental ideas behind the model

The models are designed on the premise that deaths are the most reliable statistic for monitoring the transmission of COVID-19 as well as overall case-fatality rates. The authors therefore focus their efforts on developing a model of the death rates. The authors then assume that hospital service resource needs will be highly correlated with death rates and can be derived from the same models (see below).

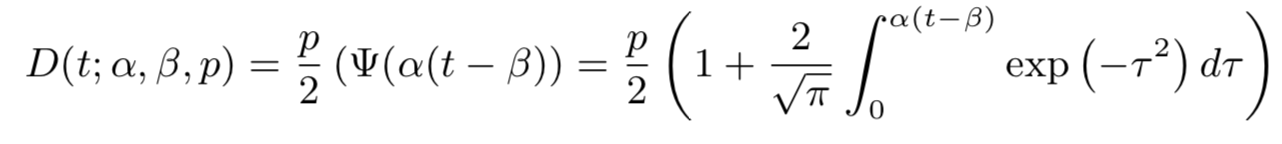
Data types and associated sources

There are four main data types used by the IHME team for constructing these models

* Deaths
  + WHO and government (both local and national) data sources to identify the number of COVID-19 deaths per day at the state/province level
* Social distancing policies
  + National and state governmental websites, executive orders, and newly initiated COVID-19 laws
* Hospital Beds, ICU capacity, and annual utilization by state
  + American Hospital Association
    - Have to estimate ICU utilization rates from limited data obtained in Italy, the United States, and China as well as parameters sourced from literature.
* Age-specific death rates
  + Data from China, Italy, Korea, and the United States

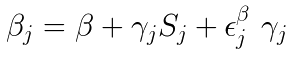
Functional form of the model

The IHME team chose to use a parameterized Gaussian error function to model the cumulative death rate for each location



where the function Ψ is the Gaussian error function (written explicitly above), p controls the maximum death rate at each location, t is the time since the death rate exceeded 1e-15, β is a location-specific inflection point (time at which rate of increase of the death rate is maximum), andαis a location-specific growth parameter (speed of infection).

Since data are available across multiple locations, the authors needed to develop ways to maintain meaningful relationships across the data sets. To do this they break the three parameters α, β and *p* into location independent and location specific components and use a so called “link function” to ensure consistency across locations. Examples of some of the link functions used in this project are as follows:



OSX_partition:Users:mpp:Desktop:Screen Shot 2020-04-05 at 9.35.56 AM.png

OSX_partition:Users:mpp:Desktop:Screen Shot 2020-04-05 at 9.36.00 AM.png

where j denotes the location, S is a social distancing value, γ, *u* are random effects*,* andε is a noise factor.

From Death Rates to Hospital Utilization

Once the death rate models are constructed the team estimated hospital service utilization using an individual-level microsimulation model. In short, they simulate deaths by age using average patterns from the US and other countries (there is not enough data for the United States yet for them to use it alone). Details on the specific estimates produced from this procedure are briefly summarized here:

* Estimate of hospital bed usage
  + From each simulated death a date of admission is estimated using the median length of stay for deaths from the global line list (10 days <75 years; 8 days 75+ years). The number of individuals requiring admission but discharged alive was then generated using the age-specific ratio of admissions to death.
* Daily ICU occupancy and ventilator use
  + Age-specific fraction of admissions requiring ICU care
    - Based on data from the US (122 total ICU admissions over 509 total admissions).
  + Fraction of ICU admissions requiring invasive ventilation
    - Estimated as 54% (total n = 104) - based on 2 studies from China.
  + Applied median lengths of stay of 12 days based on the analysis of available unit record data and 8 days for those admissions with ICU care.

In all data from this part of the analysis appeared to be difficult to obtain and the results are not nearly as robust as the estimates of death rates.

Technical details of constructing the model

The IHME team fit their data with the model described above using custom curve-fitting software. They are in the process of publishing the software on GitHub. On April 7th the first working example was merged with the master branch. For us to be able to leverage the code they have published in a meaningful way we need to spend time with the input scripts and see we can reproduce the fits with the initial conditions and constraints provided in the example and the data that they have made available to the public.

Summary and what this means for our work

The IHME model is constructed to accurately model death rates, which are then used to estimate hospital service utilization. The main data used for modeling the death rates are official death totals and social distancing policies. The extrapolation from death rates to hospital service utilization requires additional information about age related death rates and estimates of median lengths of hospital stay by age group.

Many experts consider this model to be the best available right now and the IHME team updates the data and fits constantly. The model has received some criticism for using not being able to accurately extrapolate the utilization of hospital resources from the death rates, and because of this, looking into the extrapolation procedure would be the best place to search for opportunities to improve the results. Other areas where we might be able to extend the model would be looking into increasing the granularity of the results by using county or city level data. This, however, would likely be an exceedingly difficult task and many of the parameters used in the procedure are currently being extrapolated from a national or even international level and are less likely to be reasonable estimates when dealing with smaller sample sizes.