7 April 2020

Call with Dobormir, James, Chloe, Eileen plus

David Swan, Daniel Reeves, Josh Schiffer.

Introductions: David described the work his group is doing, building an ODE model to study strains of covid propagating among cells, and looking to use our model at some point. Chloe described where we are with our model calibration, the roadblocks she’s hit and places where she wanted more clarity or where we needed to limit the amount of parameters we’re trying to estimate due to lack of sufficient data.

General idea: We have a complicated model so we will make lots of assumptions, based on data where we can, best guess where necessary. We discussed the following model components in some detail.

## Model parameterization notes

1. Geography. Should we include only King County for now or add Snohomish? Data may not be consistent across two sources, but it would increase the absolute number of cases for calibration.
   1. King County only for now.
   2. DD: What I heard is a strong preference to add Snohomish data to allow for more viral samples. We should plan to extend the analysis to both counties once we have our initial calibration for King County. Eileen, do you think you can work on integrated data set. I hope that Dan and Dave will provide sources.
2. Recovery rates/total time/infectivity. “Recovery” in our model will mean recovery from infectiousness, as opposed to the strict definition of doctors. Many studies have indicated a generation time of 5 days, with the expectation that patients shed a high viral load for 5 days, and continue for a couple of weeks at a lower load. Don’t expect those in hospital to have a high rate of infectivity due to strict measures taken (UW only has 1 documented case of a healthcare worker being infected from a patient).
   1. Reasonable scenario: 5 days
   2. Bad scenario: 20 days
   3. Infectivity from hospitalized: 10%
   4. DD: Chloe, based on this discussion I think we should change r1=0.2, r2=r3=0.15 and beta\_h=-0.1
   5. Josh 4/10: Not sure if you all have seen this: [https://covid.idmod.org/data/Social\_distancing\_mobility\_reductions\_reduced\_COVID\_Seattle.pdf](https://urldefense.proofpoint.com/v2/url?u=https-3A__covid.idmod.org_data_Social-5Fdistancing-5Fmobility-5Freductions-5Freduced-5FCOVID-5FSeattle.pdf&d=DwMFaQ&c=eRAMFD45gAfqt84VtBcfhQ&r=_bxm4-TOYbUwJOcCScT1-lpVji0pBDhe0mulnhpix3U&m=z0y5pDDyyz6YUeN5TSlKvJCfPg2_B_mFxNCUqFfkUvc&s=okVLNzQF3RMBlHGABp--qHo7qSdJDiGfGXzI4LZZjoc&e=) There is more than we need re: mobility but maybe the R0 estimates are helpful. I wish I knew what the y-axis was in Fig 2. Looks scaled for death but not even mentioned in the document.
   6. DD: I think IDM shows number of new cases daily in Figure 2.
   7. Josh 4/10: My 26000 foot view of the literature is that major groups are not agreeing on R0. I do not yet have any intuition why the London school is getting considerably lower R0 estimates. Maybe they are estimating a lower % of asymptomatics?
      1. [https://epiforecasts.io/covid/posts/global/](https://urldefense.proofpoint.com/v2/url?u=https-3A__epiforecasts.io_covid_posts_global_&d=DwMFaQ&c=eRAMFD45gAfqt84VtBcfhQ&r=_bxm4-TOYbUwJOcCScT1-lpVji0pBDhe0mulnhpix3U&m=z0y5pDDyyz6YUeN5TSlKvJCfPg2_B_mFxNCUqFfkUvc&s=uKPNlM5RLSUPjYFG4utQCLI_Op49UyAtwzzSYQY5XBQ&e=)
      2. [https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2020-03-30-COVID19-Report-13.pdf](https://urldefense.proofpoint.com/v2/url?u=https-3A__www.imperial.ac.uk_media_imperial-2Dcollege_medicine_mrc-2Dgida_2020-2D03-2D30-2DCOVID19-2DReport-2D13.pdf&d=DwMFaQ&c=eRAMFD45gAfqt84VtBcfhQ&r=_bxm4-TOYbUwJOcCScT1-lpVji0pBDhe0mulnhpix3U&m=z0y5pDDyyz6YUeN5TSlKvJCfPg2_B_mFxNCUqFfkUvc&s=v6g2tj-BlBDHoNWWZKSKIq_jqYboEdoK75DbtDtg0ug&e=)
      3. [https://covid.idmod.org/data/Social\_distancing\_mobility\_reductions\_reduced\_COVID\_Seattle.pdf](https://urldefense.proofpoint.com/v2/url?u=https-3A__covid.idmod.org_data_Social-5Fdistancing-5Fmobility-5Freductions-5Freduced-5FCOVID-5FSeattle.pdf&d=DwMFaQ&c=eRAMFD45gAfqt84VtBcfhQ&r=_bxm4-TOYbUwJOcCScT1-lpVji0pBDhe0mulnhpix3U&m=z0y5pDDyyz6YUeN5TSlKvJCfPg2_B_mFxNCUqFfkUvc&s=okVLNzQF3RMBlHGABp--qHo7qSdJDiGfGXzI4LZZjoc&e=)
      4. <https://urldefense.proofpoint.com/v2/url?u=https-3A__wwwnc.cdc.gov_eid_article_26_7_20-2D0282-5Farticle&d=DwICAg&c=eRAMFD45gAfqt84VtBcfhQ&r=d-DuB8n8nShKCn1831mgZa1LMBBcKBjqIEHLXY_WkCg&m=za59xC41ga8AvvImLlCyq0FacrWbTMzOxSlAwJwwHIA&s=hS1kmNet_vDjwcGSEPs6KFWIxq-FM2FRcpjVmBnalHM&e=>
   8. DD: Differences in R0 estimates were the main reason we decided against using them to back calculate transmission rates. It reminds me that we should calculate and plot the effective R0 we get over time with our model for comparison.
3. Deaths. Data is available daily, and very likely includes cases that were not hospitalized. We may need to confirm with data source. We think the model is set to handle this by treating “hospitalized” as the same as a severe case. This is better than splitting them in the model because of the small number of total deaths (don’t want another parameter to calibrate). Daily historical data is not available to split by age and gender, though this is available currently.
   1. Case fatality rate by age (and gender): can be pulled intermittently to serve as a calibration point. (Alternatively can request complete historical data from King County.)
   2. DD: Right now we use CFR from Italy but mostly to inform the relative death rates by age and then rescale them with a calibration parameter alpha. Will be useful to get some estimates of CFR by age representative for WA state.
4. Asymptomatic. We don’t have data on this broadly, will need a serology test on a mass scale. Best data available so far is from the cruise ship where everyone was tested. Original reports were ⅓ of those infected were asymptomatic, this has drifted down to 20%.
   1. Best guess: 20% of infected are asymptomatic.
   2. DD: This is a very controversial point and experts disagree on this a lot. We should definitely explore at least 2 scenarios while waiting for serology data.
   3. Josh: Definitely need a sensitivity analysis.
5. Infectivity, diagnostic rates, social distancing, changes. There was a turning point in the rate of increase in diagnosed cases when we began doing more testing, and another when social distancing measures were put into place (this could be reflected in two waves, as big tech firms shut down offices earlier than Gov Inslee initiated stay home order). We reflect these shifts with changes in d (diagnostic rate) and sd (social distancing), which combine to reduce the rate of exposure of susceptible individuals (lambda). Contactedness would likely also change from social distancing and therefore would change the rate of exposure, but this has not been taken into account. We should be careful to avoid double counting.
   1. Timing of switch from d1 to d2: 3/2 minus 50 days. DD: Right now, the model assumes that the increase of testing coincide with the start of SD which is roughly 60 days after the start of the simulations
   2. Social distancing: implemented as a linear decrease to lambda for the 21 days after diagnostic rate switch.
   3. Contact matrix: from the UK, deemed our best guess
6. Testing efficacy and false negatives. Various factors will influence the number of tests done and the proportion of positives. This is especially true when looking at broader locations. If an area has relatively few positives, fewer tests are likely to be done (not a current concern for King County). If an area has an outbreak, more tests will be done, but the capacity may not meet the need, so likely only those with severe symptoms will be tested (captured through diagnostic rate parameters). When calibrating to data, false negatives will bias the positive case counts downward. There may be lower % of false negatives among hospitalized cases.
   1. Add function to model to capture number of tests performed as well as positives. Possibly split between hospitalized and not?
   2. DD: This should be part of the efficiency evaluation for a test-and-treat intervention
7. Hospitalizations. Data on hospitalizations may not be reliably available. It is on the Washington State dashboard but not split by county. The CDC has some of this data, but does not appear to have Washington State or King County.
   1. Josh may have access to UW data on hospitalizations, and that may be representative enough for King County.
      1. Josh 4/8: DOH is working on this.
      2. Josh 4/9: https://www.seattletimes.com/seattle-news/politics/coronavirus-hospitalizations-in-washington-state-sharply-higher-than-earlier-surveys-but-officials-say-curve-still-flattening/
   2. DD: This is a critical piece of data missing.
   3. DD 4/10: Eileen, Just to add one more to the mix is WSHA tracker at the <https://www.wsha.org/for-patients/coronavirus/coronavirus-tracker/>. Please check how it compares to the others.
   4. DD 4/10: Also, looking at the hospitalization data shows that it will be difficult to use it as a calibration target. Too many cases (about a third) a listed as suspected. I think we should use it as a validation target instead. Checking if our current hospitalizations are in the ballpark.
   5. Josh 4/10: Validation sounds great. For what it is worth, I hear that most cases listed as suspected are COVID.  Apparently it is clinically quite easy to distinguish from other things
   6. DD: Good to know. Another issue is that the data is for WA state and "roughly" half are KC. Not exact numbers are presented by county. Too many approximations to use in calibration.

## Interventions

1. DD 4/10: Josh, I will work on specific intervention scenarios to simulate when we are happy with the calibration. Will send out for discussion today or tomorrow.

## Calibration

#### General plan (Chloe 4/8)

1. Chloe 4/8: A quick update on the data I'm using. It's unfortunately no longer possible to download data from the [king county dashboard](https://kingcounty.gov/depts/health/communicable-diseases/disease-control/novel-coronavirus/data-dashboard.aspx) in a non-image format and deaths have gone missing too. Cases and deaths are available by county on the [washington state dashboard](https://www.doh.wa.gov/Emergencies/Coronavirus), but not in downloadable form. So I think we need to contact someone to see if there is some protected link we can get access to to download data or if they can send us a spreadsheet weekly or something. Here's the contact info for the king county data set: <https://data.kingcounty.gov/Health-Wellness/King-County-Covid-19-Dashboard-link-/hwrx-ns5c>. I'm happy to make the request but maybe you'd prefer to do it Dobromir since you're the project lead? Could be a good time to ask about hospitalization data too? In the meantime I'm using the county-level data from the [nytimes](https://github.com/nytimes/covid-19-data) that Eileen found, but that is problematic since it doesn't match the data on the king county dashboard, so it won't agree with the age-specific snapshots in time.
   1. Josh 4/8: I think that having data through at least 3/31 will be enough to get started: social distancing was in place by then. When considering the 2nd wave scenarios, we will certainly hope to implement test and treat +/- PEP quite early (before or at the same time as social distancing)
   2. Dobromir 4/8: This is a valid concern but I agree with Josh that we should continue with what we have to date. The calibration scheme should include fitting 3 pieces of data:
      1. Cumulative number of cases over time up to March 31
      2. Proportion of cases by age based on the point estimate on March 31
      3. Number of deaths by age based on the point estimate on March 31

Ideally, we will supplement these with data on the hospitalizations by age and with the timing on the peak of daily cases/deaths.

#### Details (Chloe 4/9)

1. Source data: We’ll proceed without the hospitalization data for now, and we can add that in later. Unfortunately I don't have the age breakdown for Mar-31, I have it for Mar-29 when it was downloaded, and I typed it in by hand for Apr-7. I decided to use the nytimes county data so that we could have deaths and cases, which should be much better for calibration than cases alone. Because that data doesn't exactly align with the king county dashboard data, I used the age data as proportions.
   1. DD: Eileen, can you explore what is the difference between KC dashboard and nytimes data and where it may come from. If they are qualitatively the same we should expect reasonably close optimal solutions, yes? A but nervous to ignore the official KC data in favor of other sources.
   2. DD: Chloe, do we have the complete list of parameter values and ranges somewhere in one place where we can review and comment?
   3. Chloe: You can see the parameter values (params\_fix) and upper/lower bounds (params\_kc\_lower and params\_kc\_upper) for the calibrated ones in R-proj/globalparameters.R If the UW wants to run the R code, they can look at R-proj/simulation.R for examples.
2. Calibration targets and parameters: I have a simple calibration, of d2, bstar, beta\_d, delta1, h\_3, h\_4,  sd, and alpha, based on 4 criteria (cases, deaths, age dist of cases on 2 dates, age dist of deaths on 2 dates).
   1. DD: Reading your post, my main question is what simulation start time is assumed now? I guess that including it into the calibration will improve the fit and most likely change the optimal parameter sets.
   2. Chloe: Right now I'm adding 45 days before the first detected case, so I think that makes the start date Jan 14.
   3. DD: I would like to hear from the rest of the team on the weighting (especially Josh). Both, cases and death data seem equally reliable to me. Not sure if we want to put different weights on data on cumulative vs. age proportions. Also, do we need to match proportions at 2 different data points? I presume they remain relatively stable in the model outcome so differences in the real data may be difficult to resolve and hitting both targets could be tough.
   4. Chloe: Running the model with the 'default' parameters, the diagnosed proportions change from [0.04 0.25 0.43 0.29] at t=60 to already quite different at t=80 [0.10 0.42 0.32 0.15] The deaths don't change as much but they might be a lagging indicator.
   5. Josh 4/9: I never know the right answer to the weighting question so I tend to favor equal if possible. I think if you fit cumulative cases & deaths, then the day to day shifts in # of tests and reporting will have less of an effect on the data fitting. I agree one datapoint is fine.
   6. James 4/10:
      1. I believe we should try to fit deaths (broken down by age) with a much higher priority than cases.
      2. I like the idea of fitting to Iceland, with its very high testing rate. But IF we are going to fit to another location, it should be Diamond Princess mortality data in my opinion. In that case, we know that there was 100% testing done post quarantine. This means that we can have a known number of infected people and what proportion of them die. It is only necessary to fit to the cumulative number of deaths to this point. Fitting to Iceland would require another dynamic model, which is fine but obviously more work.
      3. What data from Iceland suggests is that the number of cases in most areas is dramatically under-reported. Therefore I would not trust any empirical estimates of m\_i reported so far, least of all from Italy. Again, we can inform this parameter by EITHER fitting to Diamond Princess OR just using case fatality in Iceland. I was hopeful that we could use an estimate from another modeling group, like Imperial, but after reading their methodology it is apparent that they assumed perfect diagnosis in China among 50-59 year olds.
      4. As I believe that cases are under-reported in King County, we should allow d2 to be much smaller. My guess is that there could easily have been 5x as many cases as have actually been reported. Granted many of these could have been prior to the increase in testing, but it seems probable that d2 could be truly around .02, so the range of possible values should at least include that.
      5. Where does the estimated range of 0.5-0.8 for recovery come from. Is this supposed to be .05-.08?
      6. As I noted on the github, h = (1-m)\*r
3. Calibration methods: I use mean squared error on the time series, so later values will count more since they are larger numbers. We could log the data if we'd prefer a more equal weighting. Right now it's just using least squares. We could put it into a maximum likelihood framework, but would then have to decide on an error distribution and the covariance matrix.
4. Calibration time: It takes about 15 min to run on my laptop.
5. Sample output in email chain.
   1. DD: James, please take a look on the overall calibration scheme. Would like to hear your take if we need adjustments and what "goodness of fit" should be expected.
   2. DD: I have engaged one CS student from UW to investigate the differences between MATLAB and R outputs.

Other notes

* Expect that the signal for social distancing will be very delayed, even more so for deaths.
* CDC starting serology, but don’t expect results until summer.
* IHME is not a mechanistic model, rather they fit a bell curve to China data on deaths, some very strong assumptions.
* Other data from Children’s - Josh will send a link