

Epidemiology and selection problems and further heterogeneities

by Tyler Cowen April 12, 2020 at 2:45 pm in **Data Source, Economics, Medicine, Science**

Richard Lowery emails me this:

I saw your post about epidemiologists today. I have a concern similar to point 4 about selection based what I have seen being used for policy in Austin. It looks to me like the models being used for projection calibrate R_0 off of the initial doubling rate of the outbreak in an area. But, if people who are most likely to spread to a large number of people are also more likely to get infected early in an outbreak, you end up with what looks kind of like a classic Heckman selection problem, right? In any observable group, there is going to be an unobserved distribution of contact frequency, and it would seem potentially first order to account for that.

As far as I can tell, if this criticism holds, the models are going to (1) be biased upward, predicting a far higher peak in the absence of policy intervention and (2) overstate the likely severity of an outcome without policy intervention, while potentially understating the value of aggressive containment measures. The epidemiology models I have seen look really pessimistic, and they seem like they can only justify any intervention by arguing that the health sector will be overwhelmed, which now appears unlikely in a lot of places. The Austin report did a trick of cutting off the time axis to hide that total infections do not seem to change that much under the different social distancing policies; everything just gets dragged out.

But, if the selection concern is right, the pessimism might be misplaced if the late epidemic R_0 is lower, potentially leading to a much lower effective spread rate and the possibility of killing the thing off at some point before it infects the number of people required to create the level of immunity the models are predicted require. This seems feasible based on South Korea and maybe China, at least for areas in the US that are not already out of control.

I do not know the answers to the questions raised here, but I do see the debate on Twitter becoming more partisan, more emotional, and less substantive. You cannot say that about this communication. From the MR comments this one — from Kronrad — struck me as significant:

One thing both economists and epidemiologists seem to be lacking is an awareness for the problems of aggregation. Most models in both fields see the population as one homogenous mass of individuals. But sometimes, individual variation makes a difference in the aggregate, even if the average is the same.

In the case of pandemics, it makes a big difference how that infection rate varies in the population. Most models assume that it is the same for everyone. But in reality, human interactions are not evenly distributed. Some people shake hands all day, while others spend their days mostly alone in front of a screen. This uneven distribution has an interesting effect: those who spread virus the most are also the most likely to get it. This means that the infection rate looks very higher in the beginning of a pandemic, but sinks once the super spreaders has the disease and got immunity. Also, it means herd immunity is reached much earlier: not after 70% of the population is immune, but after people who are involved in 70% of all human interactions are immune. At average, this is the same. But in practice, it can make a big difference.

I did a small simulation on this and came to the conclusion that with recursively applied Pareto-distribution where 1/3 of all people are responsible for 2/3 of all human interaction, herd immunity is already reached when 10% of the population had the virus. So individual variation in the infection rate can make an enormous difference that are be captured in aggregate models.

My quick and dirty simulation can be found here:
<https://github.com/meisserecon/corona>

See also Robin Hanson's earlier post on variation in R_0 . C'mon people, stop your yapping on Twitter and write some decent blog posts on these issues. I know you can do it.



23 Comments



print

Comments

anonymous April 12, 2020 at 2:52 pm [Hide Replies](#)

1

Yes! These nonlinear effects hobble both simple plans (open or closed) and simple models of compliance (those looking at "degree" of agreement).

Daksha April 12, 2020 at 3:15 pm [Hide Replies](#)

2

The first problem is the data problem. The incidents of occurrence are higher than is being reported. There are strategies designed to insure the data is understated. For example Trump has excused corporations from providing timely and accurate data. Deaths which are not tested as virus are not recorded as virus also understating the occurrence. Reliving on the federal data base in making decisions is to use bad data. There are better data bases not infected with the virus of manipulation. The Atlantic Constitution just repeated on the rank corruption the Governor has been utilizing in Georgia. The official base is understated partly because testing is nowhere it needs to be and various parties are working to insure it stays that way

anonymous April 12, 2020 at 3:22 pm [Hide Replies](#)

3

It's true that in normal times we don't worry too much about what other people die of, especially far away.

Suddenly it would be nice to be a full autopsy nation.

("The percentage of deaths for which an autopsy was performed declined more than 50 percent from 1972 through 2007, from 19.3 percent to 8.5 percent..")

[Respond](#)

[Respond](#)

Respond

3rdMoment April 12, 2020 at 2:58 pm **Hide Replies**

4

This thread describes a very interesting simulation from a mathematician, using connected graphs, that basically explores the same ideas about heterogeneity in R_0 .

https://twitter.com/gro_tsen/status/1241745979663155203

He also claims that we observe lower total infection rates in the real world than would be expected from simple SIR models.

anonymous April 12, 2020 at 3:05 pm **Hide Replies**

5

Also excellent.

Respond

Respond

Mogden April 12, 2020 at 2:59 pm **Hide Replies**

6

Now that Tyler has outed himself as an anti-elite bad person, we can ignore what he has to say.

Respond

Massimo April 12, 2020 at 3:10 pm **Hide Replies**

7

This alleged overestimation makes a lot sense to me, now that I see explained. Of course, I did not think to that myself. But I am not an epidemiologist, let alone a smart one.

However, it seems to me that this theory, if it holds, it might be over my intelligence, but it is still basic enough to be considered by at least one member of the legions of epidemiologists that existed since the inception of epidemiology, whenever that was.

What can explain that the models seem not to consider this apparently very important consideration? Might it be that overestimating is actually a feature, not a bug? It makes sense, considering the incentives of the epidemiologists. The more serious the issue, the more they become important in the eyes of society, and the more the budget of their departments increase. Again, might it be just a Public Choice issue?

Catinthehat April 12, 2020 at 3:47 pm [Hide Replies](#)

8

Overestimation is also a feature of the climate models. There's a tendency to make the situation look worse. It justifies your paycheck some. Why do we need so many climate scientists if the situation is under control

anonymous April 12, 2020 at 4:05 pm [Hide Replies](#)

9

Has that ever made sense? Total revenues for the oil and gas drilling sector were \$3.3 trillion in 2019, but somehow the financial incentives were *still* with environmentalists?

No, it sounds like the answer is somehow making it out in spite of financial incentives.

Very obvious incentives running the other way.

Catinthehat April 12, 2020 at 4:10 pm [Hide Replies](#)

10

The environmentalists have their own incentives. Their careers, their reputations, their convictions. Regardless the model predictions run hot vs the actual warming

That was good April 12, 2020 at 4:16 pm [Hide Replies](#)

11

Now explain why the Arctic climate models have been so flawed, as the actual reported data varies considerably from more comforting consensus predictions involving a longer time scale for what has been observed over the last decade.

Climate science modelling can be deeply flawed, and the Arctic is a prime example of just how unreliable it is.

Respond

Respond

Respond

Respond

Respond

International Pants Apparatus April 12, 2020 at 3:34 pm **Hide Replies** 12

>I do see the debate on Twitter becoming more partisan, more emotional, and less substantive.

This coming from a solidly-partisan Tyler Cowen?

I can only conclude that you've realized you are losing.

Respond

Simple answers to such questions April 12, 2020 at 3:35 pm **Hide Replies** 13

"I do not know the answers to the questions raised here"

Collect data and test people. It really is that simple in places like South Korea or Germany, and these days, Italy.

And ignore twitter completely, of course. Along with people whose models are not based on any actual data, not that an absence of data appears to be a problem for economists.

And as a final tip, anyone using the term herd immunity without reference to a vaccine is easily dismissed.

Respond

Raghuveer Parthasarathy April 12, 2020 at 3:49 pm **Hide Replies** 14

On Twitter: Over the past week or two I've looked at Twitter more, to see if there's anything useful I'm missing. I've reaffirmed that Twitter is awful -- just noise and bias--confirmation. I can understand how it's addictive, with a steady stream of novelty with a low bar to accessing it, but the lack of substance is really obvious, and the net result is simply to increase anxiety or cocoon people into bubbles. The only good things on it are (1) pointers to articles and blog posts; (2) tweets that are basically just images, for example of interesting foods; (3) glimpses into the psyches of groups that one would never intersect with, for example odd political fringes of different countries. I'm not actually sure #3 is actually insightful, rather than oddly entertaining. There are a handful of twitter threads that have something vaguely approaching the structure and depth of a real argument, but these just highlight how poor twitter is for this sort of layout, and I really can't understand why the tweet isn't instead: "here's a link to a real post or article."

George April 12, 2020 at 4:00 pm **Hide Replies** 15

I get a little disgusted when someone posts a fifteen-tweet thread to make their argument, using the 1/15, 2/15, 3/15 numbering system and all.

Good things blogs are dead and newspapers are gone.

Respond

mkt42 April 12, 2020 at 4:07 pm **Hide Replies**

16

Yes, I never go to twitter, with but two exceptions.

If someone such as Tyler links to it, then thanks to his selection/curation (as well as my estimation of whether I should click or not, I do so maybe 1/3 of the time) there's at least a 50% chance I'll find some interesting tweets. Without that, the chance of finding something worthwhile is less than 1%.

Once or twice a day twitter generates an email to me with tweets that it's algorithms think that I might find interesting. After perhaps a decade it's finally become semi-decent at that. I'll click on perhaps 10% of their proffered tweets, and be satisfied perhaps half the time.

What continues to drive me nuts is when someone, especially the people who are allegedly good tweeters, posts a tweet that says "read this, it's great" and then a link -- with no explanation of why it's great. I've learned not to bother clicking on those, unless I have nothing better to do with my time. Even if it's a person who generally posts quality tweets, that link is usually disappointing. When they provide some description of what it's about, I can better decide if I want to click on it or not.

Respond

Respond

Makiavelli April 12, 2020 at 3:52 pm **Hide Replies**

17

I did a small simulation on this and came to the conclusion that with recursively applied Pareto-distribution where 1/3 of all people are responsible for 2/3 of all human interaction, herd immunity is already reached when 10% of the population had the virus.

interesting. Maybe he can explain this in a conference of "experts" who all share their similarly quick-and-dirty Excel macros.

It should be obvious that it would be moronic to make policy decisions based on these hypotheses, but the experts are so in love with their models that I fear someone might go ahead and push for "let it rip" herd immunity based on this.

Respond

mkt42 April 12, 2020 at 3:52 pm **Hide Replies**

18

"One thing both economists and epidemiologists seem to be lacking is an awareness for the problems of aggregation"

Kronrod's (not Kronrad's) post was fine, although economists would probably call it a problem of heterogeneity rather than of aggregation but that's a minor semantic quibble.

The problem with this call, not just by Kronrod but by innumerable commenters and pundits, for corrections for heterogeneity, confounding factors, covariates, etc. is that in the early stages of the epidemic we lack the data to accomplish this sort of analysis.

We can try to make all sorts of nice upgrades of our models, but we quickly find that our data sets have insufficient observations or missing variables or both.

We can make some crude and obvious observations and corrections for age: older people are more vulnerable. Maybe gender: males seem to be more vulnerable, but is that due to gender or to other variables?

But only now are we starting to get data that allows us to go beyond that with much confidence. County level data in the US is probably still not useful except for finding that the virus started in big cities and spread elsewhere. Even country level data is questionable -- what was different in Italy? Can China's data be trusted?

One small advantage that economists have: the importance of distinguishing between endogenous vs exogenous variables is built into their thinking. Most analysts recognize that R_0 is an endogenous variable and not a constant, but economists are more likely to build models that attempt to analyze its determinants (note again however the limitations of the data) and to recognize the limitations of models that utilize R_0 but that lack a good way of determining what the value of R_0 will be.

Respond

Where are the kids in these models?

April 12, 2020 at 4:02 pm

Hide Replies

19

Which is an extremely simple question that the heterogeneity modellers should have no problem answering. Schools were in session until the lockdown, and children certainly spread the virus, so it should be no problem for such models to already provide answers to how much schools are involved in spread.

And if such models are unable to provide any insight into elementary or high school spread, then maybe it might make sense to gather more data on a previously unknown virus instead of wasting time playing with models.

Respond

Randall Parker April 12, 2020 at 4:04 pm **Hide Replies**

20

We need to know a lot more about the demography and lifestyles of the infected versus the non-infected: occupations, hobbies, family composition, commuting methods, work space density, meeting frequency, store shopping frequency, and the like. Collecting this ought to be part of the flow of getting tested.

Respond

Catinthehat April 12, 2020 at 4:05 pm **Hide Replies**

21

Ro does have a large variance which most models ignore for simplicity. There are the super spreaders. The observation that if you're more likely to be infect others, you're more likely to be infected yourself is a good one. There is probably a lower limit to the size of an infection cluster below which it stops and doesn't propagate. There are hermits who infect no one. In the middle are a bunch of average people who still have to go to school, to work, to church, to a gym class, go out to lunch and might see a movie, take a trip and the like. Many clusters seem to start from an event and I am not sure they can be modeled from a super spreader, The Gangelt Carnival near Heinsberg, the choir in Washington state, the Kirkland lifecare nursing home (started from a birthday party there), the soccer match in Lombardy, the Church gathering in Korea. I think a single infected man traveling on a plane infected 37 people on that same plane with SARS in 2002 and SARS may be less infectious than Covid-19 Such events will occur for the average person especially before the economy is in shutdown

Respond

Matthew W April 12, 2020 at 4:11 pm **Hide Replies**

22

"late epidemic R_0 "

Isn't R_0 defined as R at $t=0$? How does late epidemic R_0 make sense? Do these models really keep R constant?

Respond

So individual variation in the infection rate can make an enormous difference that are be captured in aggregate models.

Not at equilibrium. The hospital entry and exit rates have to be stable. So hospitals expect the occasional burst in R_0 and allocate reserve bed space for the expected occurrence. The hospitals act like a hedge fund.

Respond

Add Comment

Email *

Author *

Website

Comment

☐ Notify me of follow-up comments by email.

☐ Notify me of new posts by email.

Submit

Cancel