May 20, 2020

R_0 for Covid-19 research: An early estimate and policy implications[1]

Avinash Dixit, Princeton University[2]

Introduction

If any pandemic spread faster than Covid-19, it is that of research about Covid-19: its epidemiology, the structure of the virus, development of vaccines, antibodies and treatments, its economic and political consequences, and so on. Covid-19 spread at the rapid rate of air travel, but research about it is spreading at the even faster speed of the internet. This alarming development calls for research into this new pandemic, and some thinking about how to cope with it. Here I make a start by using some data to calculate the R_0 for it. Then I offer some preliminary and highly speculative policy recommendations.

Data and results

Given the urgency of the problem, I chose to begin with some simple very rough calculations based on a small sample of data available at this point. In future work I hope to develop a fuller database and conduct proper statistical tests.

I searched Google Scholar for "Covid-19", and took a random sample of ten papers from the first three pages of results. These are shown in the Data Appendix. For each, I recorded the number N of citations it had attracted to the date (May 18), and the number M of references listed in the paper. I attributed the citations to all the references plus the paper itself, so the calculated rate of "infection" R equals N/(M+1). Averaging this over the data sample, the estimate of R_0 for the Covid-19-research pandemic is 34.

Of course I don't intend this to be interpreted as any precise estimate, and point out some sources of bias and error. But it seems safe to say that the R_0 for Covid-19-research greatly exceeds that for Covid-19 itself![3]

---

[1] This research was not funded by any makers of disinfectant beverages.

[2] I wrote this paper because at times like this we could all do with a laugh. I hope the paper will "go viral." For a more serious take, see (thanks to John Londregan for this reference) https://www.sciencemag.org/news/2020/05/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat#

[3] Estimates for the latter often range around 3, sometimes as high as 6; see https://www.livescience.com/new-coronavirus-compare-with-flu.html

Here are just some possible reasons for bias or error in my estimate. (1) The references listed in each paper include several that pertain to methodology or other non-Covid work. Therefore the true M is likely to be smaller, and the true R_0 bigger. (2) The N is as of today, and will grow over time. Therefore the eventual R_0 will be larger, perhaps substantially so, than my estimate. (3) The first few pages from Google Scholar may include the most influential papers, and the average paper may get fewer citations. That would make the true R_0 smaller than my estimate. (4) A citation may not indicate a primary source of infection. But it makes some contribution to the total viral load, and therefore helps measure the seriousness of the contamination.

These sources of bias and error call for more detailed and careful work in the future. But I don't think they will alter the conclusion that $R_0 >> 1$ for Covid-19 research. The implications for exponential growth are truly frightening.

Policy discussion

How can the huge R_0 for Covid research be brought down to a number less than 1? We can get some ideas from the "test, trace and isolate" procedures being developed and used for Covid-19 itself in many countries.

Test: Here we have an advantage. There are no asymptomatic carriers. Researchers know perfect well whether they are infected, and if they are, they want to display and publicize their research as fast and as widely as possible. So we need not do any testing and have no need to invade anyone's privacy.

Trace: Here again we have an advantage. Google Scholar and similar publicly available tools enable us to trace the contagion very easily.

Isolate: The editorial and peer review processes will be of great help in efforts to isolate and quarantine research. Desk rejection rates at the top journals are already approaching 80%. If similar scrutiny can be applied to what are currently essentially unsupervised self-publication platforms like the various arXiv's, that will prevent the public appearance of much research, and so slow down spread of the contagion. Peer review can also delay chain of infection. Economics has perfected this art; peer review in that field can stretch out over three or four years. Other disciplines have much to learn from this.

Treatment: The editorial process can be of immense help. It can delay the public's exposure to papers for a long time. It can also make the final result unreadable and therefore less likely to be disseminated widely. Requiring the authors of accepted papers to conceal their main message – wrapping it in all kinds of secondary extensions, tests, caveats, and so on – leads to substantial increase in the length of already-lengthy papers. Editors can thus ensure that each paper gets read by only a very small number of experts in the narrow topic, who are likely to be already infected in any case. Once again Economics leads the way in this, and others can follow.

Data Appendix

No.	Paper	Citations received (N)	References listed (M)	R=N/(M+1)
1	Mehta et al, Lancet 395, 1033-4	548	11	49.8
2	Bai et al, JAMA 323, 1406-7	671	6	111.8
3	Zheng et al, Nature Reviews Cardiology 17, 259-60	282	10	28.2
4	Gao et al, Bioscience Trends,	565	7	80.7
5	Rothan & Byrareddy, J Autoimmunity 109,	298	29	10.3
6	Fauci et al, NE J Med 382, 1268-9	185	12	15.4
7	Anderson et al, Lancet 395, 931-4	399	19	21.0
8	Dong et al, Drug Discoveries 14, 58-60	188	19	9.9
9	Hellewell et al, Lanet Global Health 8, e488-e496	283	35	8.1
10	Sohrabi et al, I J Surgery 76, 71-76	289	62	4.7
		Average =		34.0