Epidemiological Investigation of COVID-19

Health Freedom Ohio Research Team

The objective of this paper is to review what is known about the epidemiology of COVID-19. The various epidemic models that have been used to drive policy will be described, with particular note regarding their strengths and weaknesses. Recent data from serologic surveys will be summarized, and the implications of these findings on meaningful epidemiologic parameters will be explained. Finally, we will elaborate on what parameters are needed in order to understand the true public health burden.

But first, we would like to start by saying that we do not downplay this virus. We do not claim it is a "hoax" or "conspiracy" or anything of the sort. Rather, we believe the proportions of this epidemic are being exaggerated. While it is tragic that people have been affected by this virus or have even succumbed to it, there is morbidity and mortality due to respiratory illnesses every year.

The Importance of Case Definitions

We have heard repeatedly that COVID-19 diagnosis kits are not widely available. This is surprising, given a recent report of 10,000 kits being sent to Uganda as a humanitarian effort [1]. This is part of the reason that the CDC has issued official guidance [2] that counts of COVID-19 disease as well as COVID-associated death include "presumed" cases that are not laboratory confirmed. Because of the similarity in presentation to other respiratory illnesses, this results in an over-report of cases, and therefore an overestimation of the public health burden. This CDC guidance has also resulted in an over-reporting of deaths due to COVID. It is important to re-iterate that there is a difference between determining deaths in individuals <u>with</u> disease versus deaths <u>of</u> the disease. Many reports have been presented on social media of doctors having death certificates corrected to say COVID-19 was the cause of death, even when it was only a co-occurring condition (for example, [3]). This inflation of death statistics has an adverse effect on reporting to the media and policy.

Epidemiologic Models

In Ohio, there have been two epidemiologic models mentioned by Dr. Amy Acton and Governor Mike DeWine as the basis for policy. A primary concern is that neither of these models has

been peer reviewed. This is important to mention because peer review is one of the hallmarks of science. In peer review, other experts in the field can evaluate the assumptions and procedures in a scientific study. Generally, broad changes to health policy are not made until a study is peer reviewed. Admittedly, there is a push to release data in the interest of saving lives, but the lack of peer review must be kept in mind. Second, neither of these analyses are easy to find for the public to evaluate them. Indeed, these have only been obtained through direct contact with the modelers or through the media scouring the internet.

Both of the Ohio models use differential equations to characterize the epidemic curve. This is generally standard procedure in mathematical modeling in infectious disease epidemiology, so in and of itself, this is not a problem. However, these types of models should utilize observed data as the basis for different parameters, and should also vary the values of those parameters as a sensitivity analysis.

The Ohio State model [4,5] was based on a survival density curve and claimed that it used the current case count data as its basis. This is the model that projected that at some point, there would be 10,000 cases of COVID-19 per day in Ohio, and social distancing would only reduce that to 7,000 at the peak per day. However, this model made a variety of assumptions that are not appropriate. First, in the online seminar presenting this model, the scientist developing the model noted wide variations when there were slight changes in the assumptions. He said that was “too early to tell” (exact quote) what the epidemic would actually look like. And yet, this model was taken as the basis for the “stay at home” order. The model was highly unstable. Second, because there was no data available on recovery from COVID-19, this parameter was simply “zeroed out” (again, exact quote). As we explain below regarding the IHME model from the University of Washington, that assumption is critical. This model also assumes homogeneity across individuals, which is an oversimplification, and it assumes “spread on contact”.

The preprint version of this model [5] assumed there were no false positives in the number of reported cases, some vague modeling of recovery, and incorporated illness “onset dates”, which are subject to recall bias. It is also concerning that the paper describing the model never actually showed a table of the parameter estimates and how they were varied. This is standard practice in mathematical modeling. The preprint paper also estimates hospital bed usage and explicitly states that these parameter estimates cannot be shown because hospital bed numbers are considered a “trade secret” according to the Ohio Revised Code. This seems strange, considering a simple internet search reveals a cleveland.com article with the number of beds in the state of Ohio [6].

Before we describe the model from the Cleveland Clinic, we wish to explain why some of these model limitations are truly significant. First, we know that the tests for COVID-19 are not terribly accurate. False positive tests would overestimate the morbidity of disease, and thus, over-estimate the “curve”. Second, cases that are recovered are no longer infectious. By not modeling this, the epidemic could “run away”; in other words, the number of infectious

individuals in the population could accumulate, and we know that after treatment, people are no longer infectious. Lastly, assuming homogeneity in the response to exposure assumes that everyone has the same risk profile. We know that isn't the case with COVID-19. Many studies have reported that elderly individuals and individuals with comorbidities are much more susceptible.

The model from the Cleveland Clinic [7,8] is similar mathematically to the Ohio State model. This mathematical model does not incorporate error terms, but the analyst did vary the parameter estimates, and thus did perform a sensitivity analysis. While the paper says that infection rates (R_0) are estimated from the data, the paper does not show how this was done. It is unclear if the actual reported case counts were even incorporated into the model at all. The resulting curve starts looking much like the existing data, then changes shape, and more concerning, it shifts, with a peak hospital usage in July. That isn't at all justified by the observed data that have been reported in Ohio. The model even incorporates social distancing and other variable restrictions, but that doesn't explain this sudden shift in the curve.

Now we contrast these models to the IHME model from the University of Washington [9]. The first noteworthy difference is that a preprint for this paper is available on medRxiv, a preprint server commonly used in biomedical science. Second, this model has adapted with time. As more data on the number of cases was recorded, the model changed, and this can be seen from the website [10]. Third, an advantage of this model is that it incorporates average length of stay as a proxy for recovery time, and put greater emphasis on deaths and hospitalizations than number of actual cases. Of course, this model also assumed accurate case reporting, thereby having the same limitation due to "false positives" as the two Ohio models. It also assumes that the cause of death was reported accurately, and we know that the CDC has issued guidance to report COVID-19 as the cause of death even if there were pre-existing conditions and/or other factors that ultimately led to death [2]. Another advantage of the IHME model is that it formally conducts age-adjustment, thereby adjusting the mortality rate for the age of the population. Since we know that COVID-19 disproportionately affects the elderly, this is important, so that mortality rates are not over-estimated. The model was also clear about the data it used as input. This model had peak hospital bed usage occurring at April 19, 2020. It is puzzling that, while the work that went into this model was so transparent, it was dismissed by the Governor and Dr. Acton.

It is true that modeling real time is a challenge. Very little was known about COVID-19, and scientists continue to learn. However, basic math would have revealed that the estimates from the Ohio State model were unreasonable, given those numbers were higher than those seen in more populated countries. Regardless, the models would have benefited from more sensitivity analysis, and additional parameters to allow for the unknown rather than dropping those parameters from the model entirely. Alternatively, rather than criticizing the IHME model, as the Ohio State paper did, they could have considered adopting aspects of it. Ultimately, if a model is to be used for making policy, it needs to use actual observed data, not only parameter values.

This is clear from the fact that the model that did use actual data (IHME) projected the burden of disease and on the health care system with values that were much closer to what actually happened.

Our Understanding of Prevalence

Two recent studies have been conducted in California, one by investigators at Stanford [11], and another at the University of Southern California (data only available from press releases and interviews with the media [12,13]). These two surveys shed light on the actual prevalence of COVID-19 in the population. Both studies sampled asymptomatic individuals. Both studies made an attempt to sample randomly, and when their samples did not accurately represent the demographics of the county of interest, they used statistical methods to adjust the prevalence values estimated from their study. The tests used had a sensitivity > 90% and specificity > 99%, which are quite accurate for detecting infection. Sensitivity and specificity are quantities that describe the accuracy of screening tests, and these terms are often misapplied in the media. Sensitivity is the ability of test to identify correctly all screened individuals who actually have the disease, and specificity is the ability of the test to identify only nondiseased individuals who actually do not have the disease. The Stanford study, which sampled in Santa Clara County, estimated an adjusted prevalence ranging between 2.4-4.1%. The USC study, which sampled in LA County, estimated a prevalence ~4%; this study released their data immediately to the public and additional analyses were ongoing.

Why are these values relevant? First, it shows that a fair proportion of the population is asymptomatic. Second, and more importantly, it gives a more accurate estimate of the prevalence of infection. Now, in addition to confirmed symptomatic cases, we know the more accurate number of cases. Cause-specific mortality is estimated as number of cases due to disease / number of individuals with the disease. All previous estimates of cause-specific mortality were overestimates because the denominator was too small. Now, we see that the cause-specific mortality may have been overestimated by as much as 50-85 times [11]. Said another way, for months, the reported cause-specific mortality rate for COVID-19 was 3-5%. With this new knowledge of the seroprevalence, the <u>mortality rate is 0.1-0.2%.</u> The Stanford study goes on to estimate seroprevalence for other populations, and report a prevalence of 10% in Italy. This is significant for two reasons. First, the mortality in Italy has been shocking to see, but the mortality rate may have been overestimated. Second, we know that the cause of death due to COVID-19 was exceedingly overreported in Italy, amplifying this problem.

Another recent study from Denmark [14] conducted a seroprevalence study of over 9000 healthy blood donors. The sensitivity of the antibody assay they used was slightly lower than in the USC study (82.5%) but the specificity was as good (99.5%). In this study, they estimated a seroprevalence of 1.7% after adjustment for sensitivity and specificity of the assay. They also

estimated the cause-specific fatality rate in individuals less than 70 years of age at 82 per 100,000. That's 0.082%. While this was a sample of healthy individuals that passed an eligibility screen in order to donate blood, in many ways, this represents a typical healthy population.

There was a COVID-19 outbreak on the Diamond Princess cruise ship [15]. This is an extreme situation of intense exposure, however among over 3700 passengers and crew, 712 were infected, 13 died, 645 recovered, and 54 are still recovering after 8 weeks. This demonstrates the wide variability of individual response to the same intense exposure. However, when one accounts for the age distribution of the passengers, the age-adjusted case fatality rate is 0.125% [16]. An article by a renowned epidemiologist goes on to extrapolate this case fatality rate to the US population demographics, estimated to be 0.05% [16]. According to the article, this is lower than seasonal influenza.

In Ohio, we simply do not have the resources to do broad testing to better estimate the cause-specific mortality. The authors of the USC study even caution against broad testing, claiming that not all tests have the same accuracy.

Other Important Metrics

In the press conference on April 28, Dr. Acton mentioned that additional metrics would be followed soon. Here, we would like to make some points about what these metrics should be, and why they are important.

Physicians (whom we choose to keep anonymous) have noted that the true burden on the healthcare system is noted by deaths and hospitalizations. The entire purpose of social distancing was to ultimately reduce the burden on the hospital system, not prevent the entire state from getting sick. Thus, to evaluate the impact of social distancing, we need to estimate the change in deaths and change in hospitalization rate. As the governor pointed out on April 28, indeed, there are new hospitalizations and deaths every day. That is how disease outbreaks work – events accumulate. What is of interest is the rate of change in those metrics. Is the number of new deaths per day (or per week) declining? Is the number of new hospitalizations per day (per week) declining? Then you know you are on the other side of the peak.

This is why transparency is important. The inclusion of presumptive cases and deaths into the numbers presented to the media is misleading. Currently, only the cumulative numbers in deaths and hospitalizations are being presented. Length of stay is also important. An overnight hospitalization is not as severe as several days in the hospital. These have been denied to the public claiming privacy and HIPAA. This is completely contradictory. The number cases per county is being updated daily, often pinpointing where outbreaks occur, for example in nursing

homes or prisons. For lower populated counties, an increasing case count is nearly identifiable. There have been instances where individuals have been double counted, and it has been plastered on social media. Hospitalizations, discharges, and length of stay are no less private than these numbers, and are much more meaningful.

The Big Picture

It is important to keep things in perspective. As of April 28, 2020, COVID-19 has affected 16,128 Ohioans (confirmed numbers – "probable" cases are inappropriate to count). This is an incidence of 0.137%, far below Dr. Acton's infamous "guesstimate". There have been 757 confirmed deaths, which as stated above, is likely inflated because the CDC has ordered that cause of death always be reported as COVID-19 even if there was another cause of death. The mortality rate cannot be estimated without an accurate estimate of the total seroprevalence in the state. Supposing it is 1%, on the lower end of the Stanford confidence interval, the potential mortality rate is 0.075%, yes, less than 1%.

In 2018, there were 464 deaths due to influenza and 1961 due to pneumonia. The data for 2019 and 2020 are not yet available. Together, that is 3 times the number of COVID-19 deaths. So both in terms of absolute counts and rates, COVID-19 mortality is not worse than that for seasonal respiratory illnesses.

This also sets a terrible precedent. Much of the policy has been driven by this idea that asymptomatic individuals can spread COVID-19 infection. The literature suggests this may be true of other viral infections as well, such as influenza and other respiratory illnesses. Does this mean that we will be mandated to wear masks every flu season for the rest of our lives? Will businesses always be forced to close? We must ask ourselves what makes COVID-19 unique, and it certainly isn't the epidemiology.

Cited references

1. https://www.cleveland19.com/2020/04/28/case-western-reserve-university-sends-coronavirus-test-kits-african-country-uganda/
2. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/faq-surveillance.html
3. https://www.youtube.com/watch?v=wxDALi8encs&fbclid=IwAR2WYKJliqE92ch872O1b66Wnnkomg_f7lP1umdjrAJJ0fMg5UgkSqXN2Ww
4. https://video.mbi.ohio-state.edu/video/player/?id=4888&title=Mathematical%20Models%20of%20Epidemics%3A%20Tracking%20Coronavirus%20using%20Dynamic%20Survival%20Analysis
5. https://idi.osu.edu/assets/pdfs/covid_response_white_paper.pdf?fbclid=IwAR3QvE2H3hMUnSh7GJwocXgwaq9HTlXgkfah-lGYDJh5hJqbNaKxi679Gwc
6. https://www.cleveland.com/datacentral/2020/03/how-many-hospital-beds-are-near-you-details-by-ohio-county.html
7. https://github.com/sassoftware/covid-19-sas
8. https://github.com/sassoftware/covid-19-sas/blob/master/CCF/docs/seir-modeling/SAS-ETS-COVID-19-SEIR-model.pdf
9. https://www.medrxiv.org/content/10.1101/2020.03.27.20043752v1
10. https://covid19.healthdata.org/united-states-of-america/ohio
11. https://www.medrxiv.org/content/10.1101/2020.04.14.20062463v2
12. https://pressroom.usc.edu/preliminary-results-of-usc-la-county-covid-19-study-released/
13. https://www.youtube.com/watch?v=C_jXKcp4Zyg&feature=youtu.be
14. https://www.medrxiv.org/content/10.1101/2020.04.24.20075291v1.full.pdf
15. https://pubmed.ncbi.nlm.nih.gov/32109273/
16. https://www.statnews.com/2020/03/17/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/