Causal Inference, Moral Intuition, and Modeling in a Pandemic

Forthcoming in *Philosophy of Medicine/The Examination Room*

Stephanie Harvard[1] and Eric Winsberg[2]
[1]The University of British Columbia
[2]The University of South Florida

Abstract: Throughout the COVID-19 pandemic, people have been eager to learn what factors, and especially what public health policies, cause infection rates to wax and wane. But figuring out conclusively what causes what is difficult in complex systems with non-linear dynamics, such as pandemics. We review some of the challenges that scientists have faced in answering quantitative causal questions during the COVID-19 pandemic and suggest that these challenges are a reason to augment the moral dimension of conversations about causal inference. We take a lesson from Martha Nussbaum-- who cautions us not to think we have just one question on our hands when we have at least two-- and apply it to modeling for causal inference in the context of cost-benefit analysis.

1. Born to be Wild

In August, 2020, a huge motorcycle rally held in Sturgis, South Dakota, led to a staggering 266,796 cases of COVID-19 across the United States, accounting for 19% of the 1.4 million new cases diagnosed nationwide in the month following the rally and resulting in public health costs of $12.2 billion.[1]

Or so researchers told the media. How *exactly* did they come to this conclusion? And should we trust it?

The Sturgis motorcycle rally was a real-world event, one that scientists had no control over. This means researchers couldn't use the same methods they use in scientific experiments, such as the randomized clinical trials (RCT) that have been used to test COVID-19 vaccines. In an RCT (as many readers of *Philosophy of Medicine* will know), study participants are randomly assigned to two groups: the 'treatment' group gets the experimental treatment (like a vaccine) and the 'control' group gets a placebo (something that has no real treatment value, like a salt water injection). When the randomization process in an RCT works as expected (combined with other methodological safeguards), there will be no differences between groups other than those due to chance. For example, the percentage of people who work from home will be roughly the same among participants in both groups— as will the percentage of

[1] https://www.nbcnews.com/news/us-news/sturgis-rally-may-have-caused-250-000-new-coronavirus-cases-n1239577

men, the percentage of healthcare workers, the percentage of people who live in large households, and so on (again, unless these differ simply due to chance). This is helpful for drawing accurate conclusions. When researchers analyze RCT results and conclude that vaccines are effective, they can be confident (though not certain) that *vaccines* cause reductions in COVID-19 cases, not something else.

If instead of testing vaccines using RCTs, researchers made vaccines publicly available and observed their effects among whoever took them, they would run into problems drawing conclusions. For example, it's possible that significantly more people who work from home would choose to take the vaccines than people who work outside the home— maybe because people who work from home are more likely to hear pro-vaccination arguments than others, and this causes them to get vaccinated in greater numbers, or maybe for some other reason that we don't understand. Because of this possibility, if someone skeptical asked "How do you know *vaccines* cause reductions in COVID-19 cases, not something else?", researchers wouldn't have a good answer for them. After all, it could be that *working from home* caused reductions in COVID-19 cases, not vaccines.

So how did researchers draw the above conclusions about the Sturgis motorcycle rally, a real-world event, not a scientific experiment? Well, they tried their best to use the *principles* of experimentation: they observed a real-world event and created treatment and control groups to compare. They identified American counties with many Sturgis rally-goers (the 'treatment' counties) and distinguished them from counties with few Sturgis rally-goers (the 'control' counties). To know which counties belonged in each group, the researchers used cell phone data: they identified out-of-towners whose phones were pinging in Sturgis during the rally, figured out their home counties, then determined which counties rally-goers were more likely to come from. Then, they assessed whether treatment counties experienced greater increases in COVID-19 cases after the Sturgis rally, relative to control counties (Dave et al. 2021).

In some circumstances, we would have a reason to trust the conclusions reached in the above sort of analysis (called a 'difference-in-difference' analysis). If we had a reason to think that American counties with many Sturgis rally-goers are very similar to other American counties in all respects other than motorcycle rally enthusiasm—in other words, *very similar in all other respects relevant to SARS-CoV-2 transmission*—then we would have a reason to trust the conclusions. Or, if we knew there were some differences between American counties, but those differences were guaranteed *to stay fixed over time*, allowing researchers to adjust for them, then we would have a reason to trust the conclusions. Under those circumstances, we could be confident that the only important sudden difference between the treatment and control groups was their exposure to people who went to the Sturgis rally. We could safely assume that had the Sturgis rally never taken place (in an imaginary world with the opposite facts to our own, what scientists call the 'counterfactual' world), the designated 'treatment' and 'control' counties would have experienced the same trajectories in SARS-CoV-2 transmission. We could make what researchers call the "common trends assumption" (Dave et al., 2021, p. 780).

But we *don't* have reason to think that American counties with many and few Sturgis rally-goers, respectively, are all that similar. On the contrary, these counties are known to be different in countless respects, including geographic, demographic, economic and social characteristics, some of which might be expected to change over even short periods of time. So, it's just not safe to make the common trends assumption: we can easily think of all sorts of plausible reasons that different American counties might have experienced different COVID-19 trajectories, even if the Sturgis rally never happened (Dowd 2020).

Since the Sturgis rally study makes the common trends assumption under unsafe conditions, we have *at least one* reason not to trust its conclusions. In fact, many researchers have given numerous other reasons[2] not to trust this well-publicized study.[3] But when it comes to reasons not to trust something, often one is enough.

2. What Causes What?

Throughout the COVID-19 pandemic, people have been eager to learn what factors, and especially what public health policies, *cause* infection rates to wax and wane. And researchers have not been shy about providing answers. Indeed, the COVID-19 pandemic has been a period of what Jacob Stegenga has called "fast science" (Stegenga 2020; Schliesser & Winsberg 2020): we have been bombarded with studies telling us *what causes what*, and these studies have very quickly changed people minds about the best public health policies. But figuring out conclusively what causes what is difficult at the best of times, and even more difficult in complex systems with non-linear dynamics— systems like economies, the Earth's climate or *infectious disease spread*. Systems like these frequently produce interesting patterns that inspire scientific investigation: we want to explain why inflation rose and then fell in a particular period in Western Europe, why earthquakes happen when they do, why COVID-19 mortality rates in one country are higher than another at a particular time, and so on. But when we investigate such patterns we face a challenge, two sides of one perplexing coin. First, in complex, non-linear systems, there is an enormous number of possible causes to consider. Second, complex, non-linear systems are capable of making interesting patterns more or less out of nothing: the effects of these systems— from earthquakes, to snowflakes, to the Great Red Spot on Jupiter— sometimes have no *macroscopic* explanation, no explanation that humans will find satisfying or useful. They are just the characteristic patterns of non-linearity.

Figuring out *what causes what* (what scientists call 'causal inference') ultimately requires identifying all the possible causes of something and then ruling them out one by one, until we're left with a strong argument about the one that's left standing. This is why scientists tend to like RCTs: although RCTs are not infallible, they're a pretty strong method for ruling out possible causes (Howick 2011), for establishing that a certain cause-and-effect relationship exists under specific conditions *at least.* But many questions that have come up during COVID-19 have been difficult to address with RCTs— policy-relevant questions like, *do mask mandates cause reductions in COVID-19 cases?*— which has left researchers to use observational data and modeling methods not unlike the ones used in the Sturgis rally study (Mitze et al., 2020). This is

---

[2] https://twitter.com/RexDouglass/status/1303379252742479872?s=20;
https://twitter.com/ashishkjha/status/1303536487259148291?s=20;
https://twitter.com/AssumeNormality/status/1303427792693059587?s=20

[3] In addition to NBC news, this was reported in Newsweek, USA Today, CBS news, the San Francisco Chronical, and by Rueters.

not necessarily the end of the world: if we can use observational methods to establish an inverse correlation between mask mandates and COVID-19 case numbers, then rule out every other possible cause (every 'confounding factor') anyone can think of, then we might be able to conclude that mask mandates *cause* reductions in COVID-19 cases. Incidentally, this is what we've done with smoking: we've never conducted an RCT on the topic, but we know that smokers get cancer more often than non-smokers, because we've ruled out almost every possible confounding factor anyone has ever thought of (Doll 2002). But, importantly, we didn't do this in six months or a year. That smoking causes cancer is not a conclusion of *fast science.* On the contrary, it was the conclusion of many decades of research, a long process of triangulating evidence and weeding out studies we had one or more reasons not to trust.

For readers who are nervous about where we're going: our point here is certainly not that every claim so far about what caused what in the COVID-19 pandemic is wrong. At a high level, our basic understanding of how SARS-CoV-2 is transmitted clearly encourages us to make certain causal inferences in the face of certain observations. For example, the frequent observations of COVID-19 outbreaks in specific types of settings, like long-term care facilities, and of disproportionate COVID-19 case numbers among people working public-facing jobs and living in crowded housing encourage us to infer that certain conditions cause SARS-CoV-2 transmission— and *certainly* they encourage us to care about who is most affected by these conditions, which are linked to poverty and structural racism (Cevik & Baral, 2021). And, to be clear, we (the authors of this paper) do infer that the Sturgis rally, during which many people spent long periods of time together in close proximity in indoor spaces, like bars, caused many COVID-19 cases to occur. Our point is *not* that people should be skeptics (or denialists) at every level on every COVID-related question until scientists converge on a verdict or collectively throw up their hands. Rather, our point here is this:

There are *specific types* of causal inference questions that are enormously difficult to answer in a pandemic. Just one example is '*how many* COVID-19 infections did the Sturgis rally cause?'[4], which is very different from the easier question '*did* the Sturgis rally cause COVID-19 infections?'. A whole set of other relevant examples are questions related to the effectiveness of public health interventions, which have been implemented, often concurrently, at different levels in different places in the complex real world to combat COVID-19.[5] When analyzing the effects of these interventions, scientists have a large menu of confounding factors to consider and they're generally restricted to using observational data and modeling methods with important

---

[4] Or, indeed more importantly, 'how many *more* COVID-19 infections did the Sturgis rally cause than the alternative actions of would-be Sturgis rally-goers had the rally been cancelled?'. We will get to this type of quantitative counterfactual question shortly.

[5] Vinay Prasad (2021) has detailed eight methodological challenges that arise when assessing the effectiveness of public health interventions, focusing on the larger-scale ones sometimes called 'lockdowns'. Prasad, too, expresses the opinion that we are a long way off from knowing the effects of such interventions and, for some, we may never know. We don't repeat his detailed descriptions of the methodological challenges here, but our big-picture summary of those challenges is consistent with his message. We note that of course a decrease in the effective contact rate (in a compartment model) or a decrease in equivalent parameters in an agent-based model will result in fewer infections. It has to. But this is different than showing that lockdowns qua real political interventions are guaranteed to work. In any case answering a quantitative counterfactual question like 'how much infection does lockdown prevent compared to X' is different than answering the mechanistic question 'would lockdowns interrupt chains of infection transmission?'

limitations. This includes even the most cutting-edge machine learning methods, which have strict methodological requirements that are difficult to meet, particularly in a pandemic.[6] Furthermore, there is a possibility that certain patterns just won't have a big-picture cause— this makes causal inference especially hard, for obvious reasons. The sheer difficulty that we face in knowing *what causes what* under these conditions is a reason to augment discussions about modeling for causal inference. Above all, it's a reason to augment the moral dimension of those discussions.

3. Causes, Costs, and Benefits

There are good reasons to distinguish between policy-oriented causal inference questions like '*did* the Sturgis rally cause COVID-19 infections?' (a qualitative question) and '*how many* COVID-19 infections did the Sturgis rally cause?' (a quantitative question). For one, in our current context, it seems like the qualitative question is easier to answer. Armed with just a basic understanding of how SARS-CoV-2 is transmitted and of what behaviors occurred at the Sturgis rally, we could probably convince most people that this event caused COVID-19 infections to happen— but we would need to do quite a bit more work to develop a specific estimate of *how many* infections it caused and convince people that our estimate was anywhere close to right. That said, many causal inference questions that invite a simple yes/no answer are also tremendously difficult to answer, so that's not our main reason for distinguishing between these types of questions. Rather, our main reason is that the quantitative question *specifically* yields the ingredients for a cost-benefit analysis, and cost-benefit analysis is sometimes the *only good use* for quantitative answers to causal questions.

Cost-benefit analysis is a method that's popular among policy-makers and other people who need to make decisions. This type of analysis starts by thinking of two or more options that we might pursue: for example, permitting the Sturgis rally (Option 1) or prohibiting it (Option 2). Cost-benefit analysis is used in many different fields, so the details of what happens next depend on who you ask. But, generally speaking, analysts think of some way to assign different 'weights' to each option, such that they end up with an aggregate number attached to each one which represents how desirable that option is (Nussbaum 2000, p.1028). A common way of doing this is to make the numbers correspond to 'willingness to pay' values (since our willingness to pay for something links to how desirable we think it is), but this isn't necessary: as the philosopher Martha Nussbaum (2000) points out, the numbers can correspond to anything.

The important point here is that cost-benefit analysis requires generating *specific* information about what is expected to happen under each option so we can attach desirability weights to those outcomes. For example, how many COVID-19 infections would happen if we were to permit the Sturgis rally versus prohibit it? What are the economic costs associated with each of those options? And so on. It is easy to see that questions like 'did the Sturgis rally cause COVID-19 infections?' doesn't generate the sort of data that can be used in a cost-benefit analysis. If the idea is to use observational data to inform future decisions like 'should the

---

[6] An example of a machine learning method that has been used with observational data to assess the effectiveness of public health interventions in the pandemic (e.g., the effectiveness of mask mandates, Mitze et al. 2020) but which has strict methodological requirements is the 'synthetic control method' (Abadie 2021). It's outside the scope of this paper to examine whether these methodological requirements are likely to have been met in a pandemic, but the authors have discussed this issue elsewhere: https://www.youtube.com/watch?v=NmJ89ujITLo&t=3s; https://www.youtube.com/watch?v=nhhgFHE82Iw&t=8s

Sturgis rally be permitted in the next pandemic, if it's like COVID-19?', we need specific estimates like '*how many* COVID-19 infections did the Sturgis rally cause in 2020?'.

Importantly, the sort of causal reasoning that goes on in a cost-benefit analysis is closely related to a specific sense of the word 'cause', which we should distinguish from another sense. Here are the two senses:

1) Something (e.g., the Sturgis rally) was mechanically implicated in something else (e.g., SARS-CoV-2 transmission)

2) Something (e.g., the Sturgis rally) is such that *had it never occurred* there would not have been an effect (e.g., a specific rate of SARS-CoV-2 transmission)

The causal reasoning that goes on in a cost-benefit analysis relates to the second sense of the word 'cause' far more closely than the first. In a cost-benefit analysis, the relevant question is not whether the Sturgis rally will cause infections, *period*, it's whether permitting the Sturgis rally will cause *more* infections to happen than prohibiting it would (and, specifically, *how many* more).

It may seem, at first blush, patently obvious that permitting the Sturgis rally would cause many, many more COVID-19 infections than would otherwise happen, and patently obvious that the rally should be prohibited for that reason. But we might also consider certain possibilities: perhaps, if we were to prohibit the Sturgis rally (through whatever structural mechanisms exist, like forced closures of venues scheduled to hold rally events, etc.), would-be Sturgis rally-goers would find other ways to meet with each other, with comparable frequency and under conditions similarly favorable to SARS-CoV-2 transmission (e.g., in private homes). If this were true, there might be no appreciable difference in the expected number of COVID-19 infections under Option 1 versus Option 2, respectively. Or, perhaps, there *would* be many more COVID-19 infections under Option 1, but our estimates of *other* undesirable outcomes under Option 2 would lead us back to choosing Option 1 anyway. In a cost-benefit analysis, we're allowed to count whatever we want: whatever we decide to count just reflects our values. If, for example, we chose to count economic costs and overall quality of life (using a measure that registered not only COVID-19 infections but many other aspects of health and well-being), there is some possibility that we would decide to permit the Sturgis rally. If we find that we're not open to even considering that possibility, that's a good sign that cost-benefit analysis is not for us, at least not in this context. We should use a different method to inform our decision-making, such as trusting our moral intuition that permitting a motorcycle rally mid-pandemic is not for the best.

From this mini-explainer of cost-benefit analysis, we hope three things are clear. A cost-benefit analysis works well if we can answer 'yes' to three questions:

Q1) Can we generate good estimates of expected outcomes under alternative options?

Q2) Do we know what desirability weights we want to assign to different outcomes under alternative options?

Q3) Have we identified a 'decision-relevant' threshold to help us choose between our options, using the results of our analysis? (For example, do we know much we're willing to pay to avoid a COVID-19 infection or how many aggregate 'quality of life' points we're willing to lose to avoid a COVID-19 death?)

And from our earlier discussion about causal inference, we hope two more things are clear. First, we *can't* always answer 'yes' to the first question: if we are using observational data to study a complex, non-linear system over a short time period, we shouldn't think we can generate good estimates of expected outcomes under alternative options. Second, Q2 and Q3 do not refer to causal questions. Rather, they refer to moral questions— dogged ones that seldom leave quantitative causal questions alone in the policy sphere.

4. The Questions on our Hands

Philosophy of Medicine readers will have heard the following platitude: the research questions that scientists pursue reflect social values. If scientists try to answer questions about cancer, it reflects an interest in cancer and a judgment that cancer research is a moral priority, a potentially lucrative endeavor, or both, and so on. Philosophers of science famously find this insight boring (Elliott & McKaughan 2009). However, this boring insight is the seed of a rich and important conversation about the moral dimensions of modeling for causal inference. This conversation gets particularly rich when we introduce certain considerations, three of which we've alluded to above. One, sometimes research questions are subtly different from one another, and it takes some effort to see the difference and to establish that different methods would be needed to answer each one. Two, some research questions are difficult or maybe even impossible to answer, particularly in short order. Three, some research questions are only good to answer if we plan to answer other questions too. An additional consideration, which should further augment our morally-charged conversation, is that science is an enormously important structure in society. We should want science to produce good answers to important questions, to not ignore important questions that it could give good answers to, and to refrain from giving bad answers to important questions or any answers at all to questions that just don't matter. Above all, we should want science to *avoid* giving people one or more reasons not to trust it, whenever possible and even, sometimes, at a cost. So, choosing which causal inference questions to address with modeling has important—and interesting— moral dimensions. We can't explore them all deeply here, but we can encourage future conversation about a couple of them.

To start, we should consider a lesson taught by Martha Nussbaum (2000), which places cost-benefit analysis in context. The lesson builds on two facts:

1) Cost-benefit analysis is just *one of two possible methods* to answer the question of 'what to do?' in a choice situation— what Nussbaum calls the Obvious Question. The competing method to answer the Obvious Question is something like 'duty analysis': instead of tallying up all the expected costs and benefits of each of our options, we zero in on a single perceived duty and act to honor it without doing any calculations.[7]

2) Regardless of the method we choose to answer it, the Obvious Question isn't the only question we should care about. A separate question is what Nussbaum calls the Tragic Question: the question of *whether any of our options is morally acceptable.* Imagine that we reach a point where we come to believe that a virus is so dangerous that it is impermissible for the government to allow migrant workers in our country to travel home; perhaps we think this will cause an unacceptable level of viral spread.[8] However, we also know that it is

[7] (John, 2020) also discusses some of the difficulties of using cost benefit analysis in the Covid-19 pandemic and discusses some alternatives.
[8] Something like this effectively happened in Bavaria in July, 2020, though many might dispute, in this particular case, whether allowing the migrant workers to go home was in fact morally impermissible.

impermissible to effectively imprison migrant workers. The fact that no course of action looks morally permissible, Nussbaum would remind us, does not let us off the moral hook from asking whether this state of affairs tells us something about the immorality of the underlying situation.

By posing the tragic question, we have a chance to figure out if anything structural has led us to face our dilemma and if, by changing it, we might avoid it again in the future. Maybe we can fix the structural arrangements that made the migrant workers' situation tragic. If the answer to this question is 'yes', then we should make the structural change. But if it is 'no', we are still not done: if we're political actors we still need to consider *making reparations* to the victims of our forced moral wrongdoing— to do whatever it takes to help to remedy the injustice that the migrant workers face.

Nussbaum's specific lesson is not to let the Obvious Question prevent us from posing the Tragic Question. However, a more general lesson can be learned from Nussbaum too. This is the lesson not to let ourselves "*believe that we have only one question on our hands, when in fact we have at least two."* (Nussbaum 2000, 1008, italics ours).

We should reflect carefully on how Nussbaum's general lesson applies to us when we're answering quantitative causal questions. As we made clear, quantitative causal questions are very much a part of cost-benefit analysis: but they are not the only part. Rather, cost-benefit analysis is one part quantitative (Q1) and two parts moral (Q2, Q3). Following Nussbaum, *we should not let the quantitative question distract us from the moral questions that are an intrinsic part of cost-benefit analysis.* We should always remember that quantitative causal inference does nothing to answer these questions, which can be resolved only through moral debate. Furthermore, if we don't have a plan to answer Q2 or Q3, we have a reason to scrutinize our motives for asking the quantitative causal question in the first place. Let us explain what we mean by that.

Imagine for the moment that we're capable of generating a perfect quantitative answer to the question: '*how many* COVID-19 infections did the Sturgis rally cause?'. At the same time, imagine also that we have no answers to Q2 or Q3 (and no plan to obtain them), so the purpose of that quantitative answer is *not* to inform a cost-benefit analysis. So what *is* the purpose of having that quantitative answer? Perhaps we think it is to inform policy-makers who want to make a decision on the basis of a single perceived duty, such as preventing COVID-19 infections. Setting aside the obvious point that this would be a controversial view of how such policy-making should proceed, it seems like policy-makers would only need the *quantitative* answer if they had identified a decision-relevant threshold, i.e., how many COVID-19 infections they have a duty to prevent. Otherwise, policy-makers could make a decision on the basis of a qualitative answer, like 'the Sturgis rally *will* cause COVID-19 infections'. So, it seems, even if we're capable of generating a perfect answer to the quantitative causal question, we don't have a clear reason to do it unless we also plan to answer Q3 (like we said, this is a dogged question that seldom leaves quantitative causal questions alone in the policy sphere). This raises at least one other moral question, which is the one about whether it's good or bad for science to be giving answers to questions that just don't matter.

5. Sometimes a Fantasy

---

https://www.euronews.com/2020/07/28/coronavirus-germany-and-austria-battle-local-outbreaks-as-italy-sets-1-000-face-mask-fine

There is another potential purpose for the quantitative answer which we should consider, keeping in mind that purposes come from people and scientists are people too. This is a *rhetorical* purpose. For example, in some circumstances, it's possible that certain scientists might have a preferred method for decision-making (e.g., honoring a single perceived duty, *not* doing a cost-benefit analysis) and corresponding option (e.g., prohibit motorcycle rallies during pandemics) and they want to persuade policy-makers to side with them. If that's the case, then it would be useful for those scientists if the quantitative answer effectively distracted policy-makers from other moral questions, specifically those that cost-benefit analysis encourages us to ask and answer. The *mere possibility* that answers to quantitative causal questions might serve a uniquely rhetorical purpose warrants further discussion. We can think of a few reasons for scientific models not to be used as a rhetorical devices, starting with the worry that this could affect people's trust in science.

There is a layer of complexity waiting for us, which is added by the fact that we're often *not* capable of generating accurate answers to quantitative causal questions. If generating even *accurate* quantitative answers to causal questions might serve questionable purposes and have undesirable downstream effects, we have a pretty clear reason to worry even more about generating *inaccurate* answers.

We take it to be obvious that scientists engaged in policy-relevant modeling have a duty not to engage in practices that will obviously lead to harm. Producing *inaccurate answers* seems to come in two flavors: 1) the answer is off by an amount that frustrates people's purposes; 2) the answer is just fantasy: it is foreseeable that there's no legitimate way of getting at the real answer or ever testing if the answer given is close to right. At the very least, it seems like there are conditions under which scientists have a duty to avoid producing inaccurate answers of the second kind. This may be just our (the authors') moral intuition, but a recent qualitative study among health economists suggests that at least some policy-oriented modelers share it. Describing a moral decision that had come up in their practice, one modeler said:

"we made the decision not to simulate anything in that area because the lack of evidence would just produce results that were just fantasy" (Harvard, Werker & Silva, 2020, p.7).

We think the attitude being expressed here reflects moral considerations that scientific researchers ought to keep in mind when engaging with causal questions. While the pressure on scientists to produce answers may rise rapidly in contexts like pandemics, there are good reasons not to forget what epistemic powers are available to us at any given time.

References

Abadie A. Using synthetic controls: feasibility, data requirements, and methodological aspects. Journal of Economic Literature 2021, 59(2), 391–425 https://doi.org/10.1257/jel.2019145

Cevik M, Baral SD. Networks of SARS-CoV-2 transmission. Science. 2021 Jul 9;373(6551):162-163. doi: 10.1126/science.abg0842.

Dave D, McNichols D, Sabia JJ. The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19. South Econ J. 2020 Dec 2:10.1002/soej.12475. doi: 10.1002/soej.12475. Epub ahead of print. PMID: 33362303; PMCID: PMC7753804

Doll R. Proof of Causality: Deduction from Epidemiological Observation. *Perspectives in biology and medicine* 2002;45(4):499-515.

Dowd JB. The Sturgis Biker Rally Did Not Cause 266,796 Cases of COVID-19. Slate Sep 10 2020. https://slate.com/technology/2020/09/sturgis-rally-covid19-explosion-paper.html

Elliott K, McKaughan D. How values in scientific discovery and pursuit alter theory appraisal. Philosophy of Science Vol. 76, No. 5. https://doi.org/10.1086/605807

Harvard S, Werker GR, Silva DS. Social, ethical, and other value judgments in health economics modeling. Soc Sci Med. 2020 May;253:112975. doi: 10.1016/j.socscimed.2020.112975.

Howick J. The Philosophy of Evidence-Based Medicine. 2011; Oxford: Wiley-Blackwell.

Nussbaum M. The costs of tragedy: some moral limits of cost-benefit analysis. Journal of Legal Studies 2000, Vol. 29, No. S2. https://doi.org/10.1086/468103.

Prasad V. Do 'Lockdowns' work? Medpage Today Feb 4 2021. https://www.medpagetoday.com/opinion/vinay-prasad/91054

Schliesser E, Winsberg E. Climate and coronavirus: the science is not the same. The New Statesman 2020 Mar 23. https://www.newstatesman.com/politics/economy/2020/03/climate-coronavirus-science-experts-data-sceptics

Stegenga J. Fast science and philosophy of science. The BJPS Blog. Accessed July 25, 2021. http://www.thebsps.org/auxhyp/fast-science-stegenga/

Mitze T, Kosfeld R, Rode J, Wälde K. Face masks considerably reduce COVID-19 cases in Germany. Proc Natl Acad Sci U S A. 2020 Dec 22;117(51):32293-32301. doi: 10.1073/pnas.2015954117.

John, S. (2020) The Ethics of Lockdown: Communiction, Consequnces, and the Separateness of Persons” Kennedy Institute of Ethics Journal