**Harvard Data Science Review • Special Issue 1: COVID-19:
Unprecedented Challenges and Chances**

A Conversation on COVID-19 With Brief19 Editor-in-Chief and ER Physician Jeremy Faust

Jeremy Faust¹ Xiao-Li Meng^{2,3} Liberty Vittert^{4,5}

¹Department of Emergency Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States of America,

²Department of Statistics, Faculty of Arts and Sciences, Harvard University, Cambridge, Massachusetts, United States of America,

³Harvard Data Science Review, Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, United States of America,

⁴Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts, United States of America,

⁵John M. Olin Business School, Washington University in St. Louis, St. Louis, Missouri, United States of America

Published on: Jul 24, 2020

DOI: <https://doi.org/10.1162/99608f92.e9b6f961>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

On May 11, 2020, The *Harvard Data Science Review*'s Editor-in-Chief, Xiao-Li Meng, and Media Feature Editor, Liberty Vittert, conducted a virtual interview with Dr. Jeremy Faust, emergency physician at Boston's Brigham & Women's Hospital to discuss his take on the data behind COVID-19. The conversation's topics range from data quality when examining the number of cases and deaths, to the viable metrics behind when to re-open and the search for a possible vaccine. Using examples of pandemics from the recent past including HIV in the 80's and H1N1 in 2009, Dr. Faust explains how problematic it is to compare COVID-19 statistics with data from these pandemics due to the lack of similarities.

This interview includes both an audio recording and written transcript below. The transcript that appears below has been edited for purposes of grammar and clarity.

▶ 0:00 / 26:09

Audio recording of the interview

Xiao-Li Meng (XLM): Hello, everyone. I am Xiao-Li Meng, the founding Editor-in-Chief of *Harvard Data Science Review*, where we publish everything data science and data science for everyone. I'm joined by our Media Feature Editor, Liberty Vittert. Today, we're going to discuss a topic that is at the forefront of everyone's mind: the COVID-19 pandemic.

Liberty Vittert (LV): Man, it sure is the top of everybody's mind, Xiao-Li. And you know, what we're going to do today is take a look at the really tricky data that got us here, is keeping us here, and that which will get us out of it. As we record this in mid-May, there have been over six million diagnosed cases of COVID-19 and counting. And with a constant influx of data surrounding COVID-19, it can be really hard to make sense of exactly what's happening. We're going to examine the data surrounding the possibility of a vaccine, the real impact on hospitals and E.R. visits, and a whole heck of a lot more.

XLM: Well, in addition to the constant flux of data, there's also a constant flux of opinions. Thus it becomes more important than ever to listen directly to the experts. We're therefore particularly delighted today to hear directly from Dr. Jeremy Faust, emergency physician at the Brigham and Women's Hospital and a regular news contributor.

Jeremy, very nice meeting you, and thank you for taking time to do this with us. So, what we want to talk about today is really have a conversation about a lot of numbers running around, and a lot of things out there about coronavirus, the pandemic, the issue of treatment, social impact—everything. So, what do you take in terms of the data quality of the data we're seeing? How does it compare, for example, with the previous pandemic? Are we in a particularly bad situation compared to SARS or MERS, or are we in a better situation? Just sort of your general feeling about the whole issue of data quality before we go further.

Jeremy Faust (JF): First of all, thank you for having me, and it's nice to be here. I think that with the previous pandemics, the most recent one being the 2009 H1N1, the good news was that we did not have excess mortality that could be measured. Whatever happened in 2009 in the United States, it did not seem to cause a noticeable impact if you were agnostic to what was going on. If no one told you there was a pandemic happening in 2009, your eyes wouldn't catch it if you just watched the data and statistics day in and day out, week in and week out, year in and year out. If you do look at the data for all-cause mortality in the 1980s and 90s, you see 'Something's happening with young men in cities. That's clearly going on. What's happening?' And then you would ask the people there and they say, 'We have this problem called HIV and AIDS.' And that's what is explaining that rise. So that's where you have data and then you have a narrative to explain the data. That's the first step, to separate those two things. You come to 2009 and you hear about this global pandemic called H1N1 2009. And, again, the person who didn't know that wouldn't notice. They wouldn't be like, 'Oh, what happened there?' The biggest problem that we have in understanding coronavirus statistics is probably the fact that we actually have nothing really good to compare it to from the past in recent memory. So that's the biggest challenge at first, I would say.

XLM: So, you're saying that by now, if somebody, from outer space, without knowing the news, looking at that data compared to last year, the same period, they would say 'Ha! Something's going on.'

JF: That's correct, and that is highly unusual. They would notice the same thing in New Jersey, in New York, on the second week in September of 2001. They would say, 'Wait, whoa, what happened?' but then they say,

‘Oh, well, it went away. Obviously, this wasn't a sustained problem.’ Whereas today, in many states, we see a sustained problem that an outside observer with no knowledge would say, ‘This requires investigation.’

XLN: And also, this is not just about the United States. Everywhere, other countries, people look and say, ‘There’s definitely something going on globally,’ just by purely looking at the data.

LV: You said that there's nothing really good to compare it to, but you see in the news all the time, people compare it to the flu. Even the head of the CDC compares it to the flu. It's always: here's the mortality rate of COVID-19, here's the fatality rate of the flu, and that's how they compare it. So, is there a reason that you can't compare it to the flu?

JF: Yeah, there is a reason you can't compare it to the flu, and again, it's because it's sort of the garbage-in, garbage-out thing. We started with talking just about excess mortality, and that's the number of deaths that are happening over the expected baseline. This is a different question. This is about case fatality rate. This is the number of deaths divided by the number of cases. And the problem with that case fatality rate is it relies, of course, on an accurate determination of numerator deaths—so, counting properly—and denominator—the number of cases that you actually think are out there. I would propose that both are being incorrectly measured right now. At the moment, the deaths are closer to the cases. In other words, the death, counts are a little bit off, probably. In most places, like the United States, the case counts are way off. It's incredible. That's the 10,000-foot view, but the issue with seasonal influenza is that if you're going to compare anything to anything, anything to seasonal influenza, coronavirus or whatever else, you're kind of using the baseline assumption that the data that you're using as a comparison, the baseline statistics, is accurate. And the trouble is that—in my opinion, and I think I'm managing to convince people about this—that the data that we've been given about seasonal influenza is pretty inaccurate. It's pretty wrong. The methodology is terrible. And so, the number that gets floated around, I think you all heard, oh, yeah, the seasonal influenza, the case fatality rate is about zero-point-one percent. And this is accepted as dogma. But if you actually look into how that number is calculated, you immediately conclude that that number is grossly inaccurate. It is wildly, wildly wrong. It is insanely wrong.

XLN: In what direction?

JF: Oh, it way overestimates influenza. Influenza is not nearly as fatal as 0.1% percent. This would imply that 1 in a 1,000 people with influenza dies of influenza. That is just completely false. But what the CDC does acknowledge openly is that they feel that we are missing many influenza cases.

They feel that there are coefficients and modifying statistics that, for every 1,000 influenza cases we hear about, just active cases, we need to correct for that and say, ‘OK, well, we missed some because not everyone goes to a pediatrician. We missed some because not everyone gets admitted to the hospital. We missed some because, even if they do get admitted to the hospital, they get admitted with the ICD-10 code of asthma

exacerbation, or leukemia, or whatever it is.’ And they may not put on the coding influenza even though they had it. So, the CDC very actively corrects for what they think is undercounting of influenza. We are not doing that currently for COVID-19, but I have seen some very, very elegant ways of looking at how it is we may be undercounting coronavirus. It’s very substantial. The difference, I would say, is that, again, data and narrative. The data that I see, that people suggesting what the real numbers of cases are, just makes sense because I don’t truly believe that the case fatality rate of COVID-19 is above 1%, which is what we’re measuring right now. We know that the cases have to be higher. When I see estimates that that put the case fatality rates in a ballpark that I think is reasonable, then my narrative says, OK, we’re getting somewhere.

XLM: I want to follow up on the competing risk issues. I was fascinated by this idea of how do you count the deaths due to the coronavirus, particularly with people who have these underlying conditions, right? Now, as you said, I think I can see that, by looking at the population, you can see whether there’s excessive deaths, so you can get to what that says. But in specific cases, if you want to determine, ‘what if somebody would have died because of a disease anyway, but the coronavirus made the person die two days earlier?’ From the medical profession perspective, how do you count that death? Does it count as a coronavirus death or count as the other thing? How do you deal with those issues?

JF: So, we don’t deal with that very well, but let me give you the ultimate example. Right now, if the entire world were to have coronavirus, every single person on Earth had it, seven billion people had it—and then in 100 years, follow up, you ask how many of them died? The answer is 100% of us died. We will all be dead someday. So, the question is, when do you stop counting? And so, you’re saying, ‘Oh, look, this person died and but two days later, they would have died,’ and the question is—OK, well, not two days, two days is too soon. Obviously, we can’t tell. A month out, was the one-month risk of this person’s death? Well, hard to say. What was the one-year risk, what was a two-year risk? You have to zoom out farther and farther and farther, and then, of course, eventually you zoom out too far and we’re all dead. You have to basically get granular but not too granular. I think one year is good. I think six months is good. But also, again, narrative. You go to a nursing home and you say a question, like—I actually ask nursing homes this question—‘How many deaths have you had this year, all causes, between January and today?’ And they said—and this is a small place—they said, “We had 15 deaths, it was crazy.” I say, ‘Well, you’re a nursing home, like, how many do you usually have?’ And they said, ‘Usually maybe three or four.’ OK. That’s a piece of information. I can’t go by that, but that just gets at the point of what you’re saying, is how many of them would have died anyway? And they’re saying, ‘Well, you know, about a third,’ in a four-month timeframe.

LV: What about if you came in from an alien planet, you’d see that there was something that happened in these past two months with the excess mortality rates. But what does that let us do? Like, OK, we all recognize that there’s an issue. Does that excess mortality rate—is that a metric that you think can be used to tell us to do something or to not do something?

JF: I think it can be. I think we're studying that very carefully, and, fortunately, this is the first time we've ever had to confront a situation where we had to ask ourselves: can we use this information or is it a sort of like a marker of something? I don't think there's any historical precedent for using it to guide policy, which I think is what makes it exciting. So, for example, if there were a sudden spike in highway deaths—whatever the number is per day, it's 40,000 per year, just do the math—but that's our baseline accepted deaths from car accidents. If suddenly that number were to spike, we would not accept that. We accept the 40,000 for reasons that, you know, are really hard to comprehend when you think about it. But it's just society. We just do. But we wouldn't accept 45,000 or 50,000, especially if it was determined that was a modifiable thing. Say the reason is that every single tire that was produced in the last six months was faulty, and there's always deaths happening, let's fix the tires. Or, there was terrorism and people are putting grenades on the highways, we better stop that. When it comes to everything in our life, we have, it seems to be, some acceptable loss parameter associated with living. So, for highways, it's 40,000 deaths per year. And for gun violence, it's 40,000 per year. Opioids, we're trying to get it lower, because we're not OK with the fact that it's 40,000 per year. But right now, infectious disease we are—this is when you get away from all-cause mortality, but it makes the point: we're not OK with this huge spike. So, people say, 'Oh, we don't shut down for flu.' I say, 'That's true because, A, the numbers are tiny, and, B, for many, many years, we've been operating at sort of a level of saying, OK, this is OK. This is our society's acknowledgement, but this is as good as we can do with flu and other things like it.' Whereas in this moment we're not OK with that. You have to start to take data and then apply your policy brain to it.

LV: OK, if we say we're OK with 40,000 deaths from traffic accidents every year, how do we determine that number for coronavirus? Because there's going to be some number because we have to open up sometime. How do we get to that number that's acceptable?

JF: My baseline assumption is that it's not when the number of deaths per day goes away, it's when the number of deaths per day is such that we no longer see a huge increase in excess mortality. Right now, we're getting 1,000 deaths per day. If that goes down to 150 or 125 deaths per day, then at that point you get to the proximate cause issue. It's just really damaging society in a way that's outsized. The closer to regular death rates you get, the more you have to say, well, wait a minute, why are we shutting down for this? Because it's not really impacting our society in a way that deviates from our expected norms. Currently, we're nowhere near that, so I'm concerned that we're opening too soon. But, as with excess death that goes lower and lower and lower—if that occurs—then it's like, yeah, you get the problem of, why are we shutting down just because the person with cancer happened to have coronavirus on the death certificate?

XML: Liberty's question reminds me, a lot of these norms—as you said, 40,000 deaths from traffic accidents of those things—most people don't know these numbers. These numbers were somehow historically just formed that way. They've become accepted norm. But now, because coronavirus is on everybody's mind, you will have people argue that every death is too many, we understand that kind of argument. Trying to actually

set a number like, what would be accepted? Because we know that basically there'll be a new norm, unfortunately, because there are new reasons for people to die. I think from next year or this year, you will see that the normally-expected number will go up. Then trying to set the number not by its own force, by some kind of policy, I can see this generates lots of controversy, particularly if that number kind of discriminates against a particular group versus others. That's the kind of a real issue I want from your perspective, how do you think about these issues? What are the guiding principles, from your perspective, thinking about accepting this new reality?

JF: One of the things that we need to take into account is the fact that individual actions and choices are not self-limited. So, for example, this is the argument that some people would say: 'Well, if you are addicted to drugs and you just do it in the privacy of your own home and you die of an overdose, that didn't hurt the society other than the fact that the government apparently has an interest in protecting its citizens' lives. But it's not contagious in the sense that you're sitting at home and overdosing from Fentanyl is sad, but it doesn't actually put me at risk.' This particular problem is different because it's a communicable infectious disease, and so your liberty could infringe upon mine because I don't want to get the virus. I don't want to give it to my parents who are in their 70s. Therefore, when we make these calculations, it's not simply that we're saying what's an OK number of deaths for other people to risk for their own communities. But this thing can cross state lines. It can cross all kinds of issues. So, it is much more difficult, but also require a lot more aggressive thinking on this, because your right to go about your business only goes so far as to where it potentially poses public safety that we don't accept on a regular basis. You know, that's why we have police officers patrolling the freeways to keep us all safe. There's no way to do that with coronavirus. So now you go out with your COVID-19, and I can't tell, you give it to me. My parents then die in a month or two because I didn't know I got it from you. There's no patrol for that. That's why it's really, really amazing to me that the individual liberty people are out there talking about liberate me, liberate that. What about the rest of society? You know, your choices matter. Your right to carry a bomb on an airplane—it doesn't go very far, does it? 'Oh, trust me, I got it. I won't light it or anything. I won't set it off. I just want to bring it from my friend to my other friend. I'm just the conduit.' Yeah. I don't think so, pal. You don't get to put us all at risk. So, when we make these policies, inherent in them—but again, I don't think people verbalize this—is a subconscious, but it should be conscious, assessment of what these actions mean for people outside of the immediate circle upon which they're apparently impacting.

LV: You know, when we're talking about data and data transparency around this pandemic, one of the big issues is all of the drug companies that are now working on therapies and vaccines. I saw a tweet earlier that you put out about how the FDA gave Gilead Sciences an emergency use authorization, but they wouldn't release the data that they had the emergency use authorization around this new therapy. Where is this data transparency coming in with all of the different therapies and vaccines that are currently being studied?

JF: I'm extremely concerned about this particular way of doing things. What happened was, there's a bunch of studies going on about Remdesivir, which is an investigational drug that might or might not work for anything. It's been around for a while. This is a drug that's been looking for a disease to treat, and they tried with coronavirus. A bunch of anecdotes came out saying it seems to work. If you look at the quality, those were bad anecdotes. They weren't even good anecdotes. It's just unbelievable. And yet that made front page news. Then you had a series of studies getting done that were more serious, and Anthony Fauci announced *some* results from the one that was being done by the National Institute of Allergy Infectious Diseases. He just announced *some* of the results, but they hadn't actually released the paper, and this very much goes against the scientific community's general experience for how this goes. You don't do science by press release. Anthony Fauci has enormous credibility, but temporarily, to borrow a phrase from popular culture, he sort of 'jumped the shark' in the sense of just putting stuff out there without any ability for people like me to look at it and say, 'Yeah, wait a second,' because let me tell you, time and time again, when the drug companies release press releases around some new drug or some new study, they have the press release, surprisingly—I'm sure you'll be shocked to learn—is not a hundred percent accurate in terms of what it means. For people like me and others in the medical community, it's our job to go through and to tear that apart and help people understand what it means, so when he did this, it was just like, OK, well, it's Tony Fauci, so we're going to give him a little bit a little bit of leeway. But at the same time, it's even more concerning because it's from him, and I really worry that other groups will follow suit and do the same thing.

Now we're going to get an emergency use authorization because of something that Tony Fauci just told them. It makes no sense. There's no downside in showing us the data. If anything, it would seem to suggest they have something to hide. Now, that may not be the case. It may just mean that they aren't ready or whatever. But the first reaction is not, oh, of course, that's going to be fine—especially in a situation in which we know that the study that was from which the data was released had very recently been altered in order to change what they were studying. So, when Tony Fauci gets up there and says, "We reached our primary objective in this study," it is about the most disingenuous thing I've ever heard come across his mouth, because two weeks earlier, the primary objective had been vastly different. I mean, can you imagine if you had a drug that was having a major study that was being looked at for mortality? And then at the last minute, you decided, 'You know what, let's not look at that. Let's just see if the people who got better go home sooner.' Does that pass the whiff test for you? It's not mine.

LF: I like this whiff test. I think that's a good one.

XLM: The question I have is actually one I was reminded of by another article coming out of the *Harvard Data Science Review's* special issue. This is a professor from MIT and his team. They basically tried to study if the FDA can change their criteria for these kinds of vaccines during outbreak time? For example, don't insist on the type-I error being so tiny because you have to worry about the other side. So, they did a bunch of these Bayesian calculations, essentially taking into account the utility functions—, because death is involved in all

this stuff—and basically came up with a conclusion which seems that the type-I error could be larger. I think the number is probably, you could tolerate, say, 16 percent. The question I have is—there are detailed calculations and lots of assumptions made—but I just want to ask you, as a physician, as a doctor, how would you take that kind of suggestion, from your perspective that now these drugs are approved with a different standard taking account a more nuanced utility, from the medical perspective, do you have any concerns with that? Do you feel like that's the right thing to do?

JF: Well, you have to make sure that whatever it is you learn doesn't make you dumber. What I mean by that is, that you started from a place of ignorance and you say, 'I don't know,' I don't know works. And then you conclude something works, but you're wrong. You now know less because you're farther away from the truth. So, when it comes to a vaccine, I'm going to punt a little bit on your question because I'm not going to chime in on what they ought to allow or they ought not allow, because we have, on one hand, no time. On the other hand, whatever we come up with is it. That's it. That's all we're going to have. We're not going to have Coke, Pepsi, RC Cola. You're going to have a vaccine that's out there and it's going to be supported, and if we get it wrong, we are phenomenally screwed, beyond belief. The number of deaths that you would fail to prevent if you didn't correctly develop the vaccine, the delta there is just staggering. I would just hate to think about it.

XLM: When you think about these errors, are you thinking about the vaccine failed to prevent the deaths as it promised? Or are you thinking about the vaccine itself causing all kinds of side effects? Which one is more crucial to you?

JF: I am far more concerned about the lives saved. I think that the safety record on vaccinations is fantastic. There's been some recent writing about some of these vaccines having their side effect profiles under-reported. When you look at those, it is very clear that, in my opinion, those are not accurate reports. And I don't say it just because it's my opinion. I say that because when you report side effects or adverse events, you have to have a very, very rigid gold standard upon which to measure that.

LV: One of the big problems is the unreported deaths, but I don't mean that in the normal sense. I mean the person who chooses not to go to the hospital today to get checked for a little skin issue that's melanoma they die from in three years, or the mental health issues that are coming about with this lockdown or losing their job and not being able to have any money, and more child abuse because kids are at home and are not being looked at at school. You made a really interesting tweet about your death number—how many people die to save how many jobs. But where does that real comparison come in, that cure-worse-than-the-disease, from your medical perspective?

JF: We definitely know that, because of the shutdown, because of shelter-in-place orders, there's a decrease in the number of people coming to emergency rooms to seek care. That sounds very scary, doesn't it? As an E.R. doctor, let me just tell you that of a grand majority of those visits are not helpful to the patients. They are helpful in the sense of if you have shortness of breath, how are you to know whether it's anxiety, whether it's a

little asthma flare from allergies, whether it's a heart attack, whether it's some kind of blood clot, which means you have cancer—maybe, in some cases—how are you to know? You don't. So, you have to come to the E.R. to get checked. But, by and large, the number of people that I save because of those visits is extremely low. The more you look at this as an E.R. doctor, the more humbling it is, because you realize there are so few times when I've saved lives. I know those saves. I cherish those saves. I have thank-you letters from people saying 'thank you for saving my ass,' you know, and I love those letters. But when I think about the way we usually do business in hospitals in this country, this crisis actually unveils something about that. And that is, for every single person that I admit to the hospital for chest pain—and I'm not talking about the subgroup of people who actually have already been diagnosed by me as having a heart attack; that's a smaller subset—but just the actual everyone I mean, admitting to the hospital because I'm worried about their chest pain being cardiac in some way—the number of patients I have to admit to the hospital to save a single life is astronomically higher than anybody wants to admit. The number of people that we're saving every day, I have to confess, is a lot lower than you were led to believe when we all watched *E.R.*, or whatever show was on TV.

But as time goes on and, if we're doing this for years, then all of a sudden that melanoma on your neck or whatever, that maybe we could have gotten a year or two more life out for you, that begins to add up. We're not there yet, after six weeks, and with child abuse, suicide—there's no evidence that suicides are up. There's no evidence to support that. We know there's evidence the heart attacks are down, but the numbers are so low that—I did the back-of-the-envelope calculation—that would have probably account for maybe a couple hundred deaths in New York during this period. Even if you assume that all chest pain is the big, bad heart attack, I actually went with the most conservative, parsimonious view when I was doing it on the back of a—not literally napkin. It was a piece of paper. But just to make the point. So, yes, we would eventually get there. And that's a major concern of starvation. I mean, we can imagine that in this country, people do have food security problems. We will get there eventually if we do not open up at some point. But my assessment is we're not there yet by a long shot.

LV: Jeremy, thank you so much for taking the time to be with us. And thank you for all the work that you're doing in hospitals.

XML: This is very educational to me, and so thank you very much, I really appreciate it.

JF: I'm happy to do it any time.

LV: Xiao-Li, the issue that I'm left with, sort of the takeaway for me, is just this balance of when you figure out to reopen and what counties and states and governments can really use to decide when to reopen, and the balance of excess deaths and people's lives with their economic livelihood.

XML: That is the question. We all have to think very hard about it. I certainly hope that data and data science would bring enough useful information for us to make the very best possible decisions. Regardless of how long

this pandemic is going to last—and we certainly all hope it will end soon—Liberty and I wish you the very best as we all go through this unprecedented crisis. Remember, we are all in this together, so let's take care of ourselves as well as each other. Thank you for listening.

Disclosure Statement

Jeremy Faust, Xiao-Li Meng, and Liberty Vittert have no financial or non-financial disclosures to share for this article.

©2020 Jeremy Faust, Liberty Vittert, and Xiao-Li Meng. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.