

Dr. Alison Motsinger-Reif
Behind the Mask
December 10, 202

Barr: Good afternoon. Today is December 10, 2020, and I have the pleasure of speaking to Dr. Alison Motsinger-Reif. Dr. Motsinger-Reif is a principal investigator and the Chief of the Biostatistics and Computational Biology branch at the National Institute of Environmental Health Sciences in North Carolina. Thank you very much for speaking a little bit about your very interesting COVID-19 work which is more visual than a lot of the others. Dr. Motsinger-Reif, what was your inspiration for developing the COVID-19 pandemic vulnerability index dashboard in your county level scorecards?

Motsinger-Reif: I can't take any credit for the inspiration or the method. This has been an incredibly collaborative project. So, it's actually a partnership with Dr. David Reif at North Carolina State University and folks in his group along with folks at Texas A&M, including Ivan Rusyn and Weihsueh Chiu. They are data visualization informatics and environmental health and disaster response experts, and I do statistical analysis machine learning and computational work. They had, actually, been working and sort of thinking about visualizing and communicating risk data from a number of perspectives. This goes back to work that David had done in trying to visualize chemical exposure risk and they had been working on sort of translating what had been for chemical exposure to this sort of geospatial kind of visualization. They had been working on that in the context of flood risk, the hazard risk for floods, and they were working on that. Then COVID-19 hit, and we all shifted gears and pivoted and tried to apply this to COVID-19.

Barr: Have you worked with visualizations in the past for your project, or is this a very new experience for you?

Motsinger-Reif: I work in visualizations as a data scientist in general. Visualization is a huge part of communicating results, visualizing results, and making things clear and communicated. That being said, within that team I would say David Reif is the one that worked the most in data visualization and communication. He is really the driving force behind some of the graphics, where my group and I more immediately have worked with folks to decide which data needs to go in to this visualization, how do we weigh the different data types, and then how do we use that data to predict the number of cases and deaths? So, some of the visualization is new and really sort of driven in that collaborative partnership.

Barr: Yes. Did you draw from other tracking projects with your own visualization project?

Motsinger-Reif: Yeah. We, definitely, sort of pull data from all over the place including some of those other tracking projects. When we got started, the case counts and stuff came of the Hopkins tracker that

was there. We have now switched to USA Facts. As different things are open, we definitely, shamelessly, grab data from all over the place and try to incorporate it, but hopefully add value with the unique visualization and sort of clustering and other tools.

Barr: How did you decide on the design elements that are included in your visualizations, as well as, what you are going to even visualize, because that is also a very difficult part, there are a lot of things you could sort of look at with COVID-19?

Motsinger-Reif: The visualization, as I said, some of the tools were under development already in these other contexts so, some of it was, "This is what they were doing, and it was working". We were going to grab onto that, especially for some of the aesthetics of the map and that kind of thing. Some of it is a little obvious as well. We have gotten feedback as far as updating the look of it and what we were visualizing and some of those features from end users. Our most recent updates, for example, have come from direct feedback from folks at the CDC. They have mirrored our tracker on the CDC's main COVID-19 tracker and people that work in sort of communications and, with dashboards and other trackers. We have tried to build on what tools already existed and then take feedback on what people are responding to and what is useful.

Barr: How did you go with the radar chart given the number of different ways you can look at things?

Motsinger-Reif: Well, that, like I said, came from some history of projects of visualizing risk in other contexts and, really, one of the strengths of the radar chart is that you can jam a lot of information that is pretty intuitive, using that radar chart right. This might be a good time for me to share our screen, if you don't mind, so I can actually walk through the charts.

Barr: Sure. I had some other questions that maybe we could get to before then. How many different data sources are you pulling from, and what types of data are you getting and from where?

Motsinger-Reif: We pull data for a number of aspects. We pull demographic data at the county level from a number of sources, things like the number of cases off USA Facts, we pull social distancing real-time data from a tech company that distributes it, based on cell phone tracking information. We pull demographic and healthcare data from a number of sources. We picked data types that are both static, relatively static, and really dynamic—data that updates daily. So, information on residential density and the baseline demographics of the population in terms of how many people are in at-risk categories based on comorbidities, or age groups, or other factors is totally static.

For a location, and taken from either 2016 or 2018, we pick data off the census and other sorts of resources like that, and then we take daily data on things like the number of cases. Then we can model

the number of transmissible cases and how many people are infectious at the time and things like the social distancing data that updates daily from tracking information. Everything is pulled in automatically. It updates daily.

Barr: What does it feel like to input this data? Also, I know cleaning up the data can be a huge challenge too.

Motsinger-Reif: Yeah, and it's active work when any of these other places change the format of the data or something. I mean saying "fix" isn't the right word, but we've got to fix the dashboard when something underlying it changes. There is a great programmer data scientist Skylar Marvel that works a lot on that. Folks in the team, Matt Wheeler is a statistician that does the prediction, so he has to deal with outliers and wonky data and stuff that is missing and try to make some rational choices.

Barr: Have you gotten a lot of wonky data or have most of the data sets that you've gotten been fairly straightforward?

Motsinger-Reif: We have experimented with a lot of data sets and the ones that were really wonky didn't make it in, and we tried to find similar data from some other source, for example, we tried multiple different data streams for social distancing. Different companies put data out there and some of the data that was out there wasn't necessarily applicable to urban settings, so we played with sort of different data streams and then we have changed sources as things get dynamic. As things have progressed. the data sources have gotten much more reliable. The amount of cleaning and the amount of wonky data that was coming out in March and April is very different than what is coming out now, so we've just sort of had to update for those sorts of things, things like the number of cases changing over the weekend, we've got to make sure we're smoothing through times and across holidays.

Barr: That is a lot to think about that people may not necessarily realize.

Motsinger-Reif: There is a lot of management and cleaning and a lot of really smart people's hard work.

Barr: How are you cleaning the data you have, because I imagine the amount of data you have is just very large, so are you using any particular tools to help you with the cleaning and inputting of some of this data?

Motsinger-Reif: So, for the statistical analysis, for example, the map that's really doing a lot of that modeling uses our software. I mean, there are lots of codes, specific packages to help do that. Skylar

Marvel works in a number of programming languages that can help streamline that and automatically pull in data, and then folks at NIEHS and the Office of Scientific Computing have helped us make sure things are stable. As many people as want to access the site and can hit it at once, it won't crash. A lot of those sort of pragmatic details to work out.

Barr: Definitely. What have been some of the challenges that you all have encountered so far and has there been anything that has been surprising?

Motsinger-Reif: There have been challenges in trying to make sure, and for me in particular, but making sure communication is clear. This is a dashboard that's meant to hit a lot of audiences so how do you describe some of this technical work in a way that is useful and meaningful across audiences, is something I know can be a challenge for me sometimes, and I am often working on. In some ways, the technical challenges are easier because we have more practice working through some of those than some of the communication and the kind of approaches, we have gotten input from the communications and web folks at NIEHS have helped us.

Barr: Does it mean that what you're looking at all over the course of the pandemic, like in terms of you wanted urban versus rural. but are there other cases like that where you have kind of included more nuance in your graphic representation or something else?

Motsinger-Reif: Yeah, we have definitely been adding data types in there as we learn more things. For example, there was a big sort of finding that air pollution was associated with COVID-19 risk so, based on that result we went and grabbed air pollution data and added it to the model. Then we do analysis to see if this data that we just grabbed is, actually, informative in trying to predict cases and deaths. How much should we weight that data type versus the other data streams that are coming in?

As we learn new things, we are trying to grab that data and update our models so that they reflect new information and then certainly things are changing, right? You saw the impact of social distancing when there were official shut-down orders, and quarantine orders, right, and that had much more of an influence on the trajectory than when once everywhere opened up pretty similarly. There have been both new insights and new data that we need to remodel and reconsider and then there have been underlying changes in how things are spreading and how people and policy have changed.

Barr: Yes. Do you know of any concrete examples of how your resource has influenced policymaking?

Motsinger-Reif: I have one really fun example of public health folks in Ohio working with companies. Apparently, utilities in Ohio, and I didn't know this at all, are competitive. In North Carolina, we've got only one gas company, one electric company, and that's the one I use, but in Ohio, they compete, and

they have people knock door-to-door with flyers and advertisements and they use the debate that in these areas the risk was too high and that they shouldn't have people going door-to-door. It was one kind of fun example, and it was somebody that reached out entirely because of the dashboard and we talked about it and she wanted to make sure she was interpreting things right. They sort of stopped that kind of door-to-door canvassing in a particularly high-risk area.

Barr: That's really good! I am very happy. What are your plans for this resource?

Motsinger-Reif: We are committed to continually updating the data types and that modeling. Our immediate goal is a commitment to making sure we're continuing to document and get the word out. We've got an initial curriculum drafted for some of the local schools. There are some schools in Wake county here that are working with high school students to learn about modeling and math and some of the computer science that goes into it as well as learning about some immediate plans to make sure our models stay up-to-date. We are thinking about new data types so that we are increasing the documentation and dissemination in a way that's useful.

Barr: You are building a lot of your visualizations on ArcGIS. Have you been happy with ArcGIS and how did you choose that particular platform versus other visualizing programs like Tableau?

Motsinger-Reif: This was entirely because it's what was already up and going—like building off of tools that were already developed. The easiest tool is the one you know regardless of sort of its actual advantages and disadvantages. Some of it was just that natural: the team knows how to work with this tool, it was under development, and so we're just going to keep going with what we know.

Barr: That's great. Well, I think now might be a really wonderful time to see some of y'all's work and to get a better sense of it. [Motsinger-Reif to share screen.]

Motsinger-Reif: Is that up for you?

Barr: Yes, it is.

Motsinger-Reif: Good. When I change, when I share a screen, it always changes the ratio, the visualization ratios here, so I'm actually just going to sort of go back. On VPN and across Teams this is not the site slowness. this is my internet slowness.

Barr: No problem. The internet has been slow. I think it's so many people being on at one time. I guess while we're waiting, we can talk about have you primarily been working at home or on campus or a combination and what has that experience been like?

Motsinger-Reif: I have been doing a combination. When I work on campus, I'm very low risk. I have got an office, and I can go close that door there. We have two young kids at home that are doing virtual school and one is seven and the other one just now two, so explaining to a two-year-old that you need the office, the home office, to yourself isn't always easy, so it's been nice to be able to be in different spaces though I am at home today. You figure it out like everybody is. It's been nice to go a little bit back and forth.

[Presentation on Dashboard shared on screen.] I will show you a quick example here: This is the dashboard. You see that it loads a map of the United States and it is colored by the overall PVI (pandemic vulnerability index). I'm going to walk there and show you an example: I am in Wake County. Let's start with how you interpret. I keep getting a bad network quality warning here, but here we go.

This is the radar chart that you mentioned, and it's a pie chart. There are many advantages to this, one of them being that we can display both the magnitude of vulnerability and, of course, each of the data streams. When you talk about cleaning it and how you process data, the PVI is calculated sort of all rank based, so we don't have to worry about distributions. And it naturally handles missing data and some of those issues but, basically, each data source and we have pulled in numbers. So, you will see here listed infection, measures of infection, and data about the population concentration and mobility at baseline, how many, what's the traffic and connectivity, and how many people commute, what's daytime and night time density, some of that kind of thing. Then, the daily updates on social distancing and testing rates locally and then those demographic data and then there is the pie chart.

Barr: How is the overall score deduced? By averaging all these different numbers next to the different data stream or by another way?

Motsinger-Reif: So, it actually adds up over the ranks. For each one of those data streams, I can rank all 3000 plus counties from sort of lowest values to highest values and I can do that in context of each one of the data streams, and then it sums over those ranks. What the radar graph is showing is in all cases the bigger the piece of the pie, the slice of the pie indicates the higher risk. So, things like in green here the population density measures the denser population, the higher the risk of spread and basic infectious disease knowledge, the larger the piece of the pie. for Wake county, for example, it shows you your relative concentration compared to the rest of the country. The pie chart, the overall size of the slices shows the relative magnitude and then the coloring will show you the source of that. The PVI itself is the area of the overall size of the slices added up and then you can visualize what's the risk source.

Here is Wake county. You also can contextualize where you are in each one of those compared to the rest of the country. Here on the left what you see is this is the overall distribution of PVI in the country

right now. To the left are the lowest risk counties and to the right are the highest risk and this black bar shows you where Wake county is. Right now, we're about in the middle of the country. You can then compare for each one of data streams where we fall compared to other counties. So, we are to the lower end of the distribution for transmissible cases, we are clearly in the bottom half spread. You can see that we are a relatively dense county compared to the rest of the country in population mobility or towards the middle and in residential density. We are at the highest risk profile of social distancing. People are out and about; people are moving around here in Raleigh as much as anywhere else. In testing, we are about in the middle of the country in our counties testing rate, probably addition to graphics, so these try to show you where we are. We are relatively healthy compared to other counties, probably because we are a relatively young county when you look at our age distribution. We try to contextualize that, and then we have also got sort of reporting information. Here are three-day averages of cases and deaths, the number of days that there have been declining summaries on county numbers and then differently than other dashboards we show timelines of since mid-March our case counts deaths, then the overall PVI, and where we rank in the country. You could see that at one point it was quite a hot spot but then our relative vulnerability has gone down, and then we could [make] daily changes and then our machine learning models will actually predict, in medium terms, cases and deaths in the area locally.

There are also a ton of tools if you are on the data science side to cluster, to filter, to help visualize. You could, by default, we are showing the overall PVI in that blue color, but you can show just the top rating counties on overall vulnerability. You can click and sort and see "I want to look at top counties based on number of transmissible cases" or just based on age, or some of the comorbidities. You can go back like you can compare what did the vulnerability look like earlier in time, you can cluster counties.

One example case has been working with folks in HHS. They were interested in Orleans county Louisiana where New Orleans is, and apparently, they had really aggressively enforced some of the interventions. They had really proactively done testing and aggressively done some of the social distancing and quarantining and wanted to see what the effects of those interventions were compared to peer counties that didn't intervene. This was back in mid-summer to early summer and so we clustered, for example, on all the things that couldn't change, at least not easily. We looked at New Orleans, and we clustered to find the county that was as similar as possible on demographics, age demographics, comorbidities, and population density, all of that stuff that is just native to the area. But then we anti-clustered on, as dissimilar as possible, the interventions. And, interesting, the county that was as similar, most equivalent, to New Orleans on those static things was actually Fulton county Georgia where Atlanta is. Both looked very similar in population demographics and infrastructure and it was interesting to see. You could see in those timelines when New Orleans had intervened and they flattened their curve, if you look now, versus a peer county without those interventions. You saw the curve still spiking.

Barr: Very interesting.

Motsinger-Reif: Hopefully, that kind of thing can help decision makers and health policy people communicate. For some of those interventions to work, there's going to have to be public buy-in and it

needs to be rational and hopefully being able to say, "This is our peer county, this isn't somewhere that's totally different, this is our peer and here's what happened, we're not even modeling it, here's what happens when they made different interventions, here's what happened." There are a lot of data science tools to help you visualize or cluster and look at not just the overall vulnerability, but different sources of it and pick that apart and visualize it and process that big data that we tried to pull together.

Barr: Has the cluster feature been the most popular feature today, or are there others that have also been really used?

Motsinger-Reif: More of the visualization, I think. We've got a little bit more communication to do to make the clustering super accessible without some demonstrations that go into some of the documentation and use cases we're doing, but absolutely, for people that we've been able to orient, a little more clustering, filtering, and then the different visualizations, have been the ones that I have seen used, and had people talk to me about and sort of demonstrate.

Barr: Is there a feature that you think that is not being used that you think is really, really, really, useful and should be used more?

Motsinger-Reif: I think, one of the things, a more recent feature we have added, is the ability to overlay different time points. To really be able to overlay, you can pick for example, there are a lot of things to toggle on and off. I have got information on the cases and deaths and then I can pick to look at the PVI at a different date. And those are relatively new features that I don't think we have shared, but I can look at, basically, the lag time in that reporting of what was the vulnerability, and now I see the consequence of that two weeks later or three weeks later or one month later.

Barr: That's very interesting. You said you do a lot with machine learning and then you had the prediction part with this project. What was it like to do that for this resource?

Motsinger-Reif: We have worked with it. It is team science at the most team-based. We have done some machine learning to help decide which data streams need to go in and their relative weights. Then, we've had a group of people, Fred Wright and Yihui Zhou at North Carolina State, for example, that have done sort of classical epidemiological modeling, really rigorous traditional biostatistics, to help support that. The prediction is actually done with a Bayesian two-step nested model, that Matt Wheeler who's a really fantastic Bayesian statistician, has developed, and it's like I said, the epidemiological modeling has helped give us estimates what you can interpret in a classical framework. Bayesian models are to predict as accurately as possible. Bayesian statistics is a wonderful area that can do amazing things but is computationally really expensive. And again, on happenstance, he had been doing some brilliant work on speeding up Bayesian computation that literally he's come up with some tricks that take what would

take tens of thousands of CPU (central processing unit) years and now can do it in six hours. He built off of some really recent computational methods that let you do really sophisticated machine learning that we can update every night on the data on his machine instead of a giant cluster. It was pulling in people that were already working on some interesting different pieces of this. We were able to pull together and unify.

Barr: What is your role? You said, you have done a lot of the statistics with this project, but can you talk about it more in detail?

Motsinger-Reif: My personal lab, and I am trying to shout out to everybody that I possibly can, John House has done a lot of work with evaluating the data streams of that machine learning: which data needs to come in, which data streams need to come in, and which ones aren't as informative. We have done a lot of that. Some of my personal role is very much a functionary, working with people to get the website actually on the web, and make sure it's ready so that people can ping it, and coordinating work with the communications team and some of the write up and pulling in Fred Wright to do the modeling and pulling in Matt Wheeler to do the prediction and then we work with him. For example, John House does all the evaluations for him. He works on building the models, improving them, we help evaluate and say how well is it doing, make sure we connect all those pieces, and get it on the web.

Barr: That's really great. Definitely more of a process than maybe by happenstance people would realize.

Motsinger-Reif: Thanks for letting me share.

Barr: Who do you think are the people who are most using this resource?

Motsinger-Reif: I can tell. We can track sort of who. We don't know what they are using it for, but we've got, like I said, a lot of interactions and questions from folks at CDC. We are getting an emerging number of local public health people like that example in Ohio. There are folks in North Carolina that we have contacted, their work with other folks at HHS that work and coordinate with FEMA in making some of that sort of resource, prioritization, and planning [decisions]. That's who we are hoping will use it and, you know, some more of our first hits and interactions, when you look at who is logging on: college campus officials, people that are making decisions at that level.

Bar: It's very nice you make it available to everybody in the United States and in the world, I guess, but you know not all of us have that understanding of what is being shown and I was just wondering if that affected how you went about your design? What is the audience you had in mind?

Motsinger-Reif: Yeah. That's definitely a challenge. So, on the part of the dashboard I showed, on the dashboard itself we do have a quick start guide, sort of, like, if you have no idea what we are talking about. I will share again because it's easier to show than describe. So, when we talk about documentation, right there is a start guide that hopefully walks you through, like at a really lay level, what kind of data have we pulled in how do you read that? What are you looking at how to interpret the scorecard at a lay level? At least, it is our intention.

Barr: That's really helpful.

Motsinger-Reif: And then, we've got details. If you want to know more about the data that we've pulled and where it comes from, there is a quick description of what do we mean by transmissible cases, and where do we get that data from, here on details page. I would say another one of the use cases of the transmission we also have, has been other data scientists, that are accessing, clean. It takes a lot of work to pull it together and you see it; one of the other, actually, not the visualization linking to the data that we have pulled together and we update it nightly, and it's all structured and formatted consistently. It will load in a second, but we make all that data available.

Barr: That's great.

Motsinger-Reif: On GitHub, another use has been just making sure we're sharing the data and that some of those efforts of pulling it together and cleaning and formatting it consistently can help other people do modeling.

Barr: I'm sure. Have you had anybody say that they have used your data?

Motsinger-Reif: We've gotten a couple of citations we can track. We have a paper on medRxiv. The research paper is under review at the journal now and you can track who has cited your source. We are starting to see that other papers have cited that paper so hopefully there's some uptake.

Barr: Definitely. Oh, this is really lovely. I think we are going now to transition from your role as a scientist to you as a person living through the pandemic. So, you've mentioned a little bit about some of your challenges and opportunities but can you mention, maybe more in detail, what have been some other challenges for you as well as personal and professional opportunities?

Motsinger-Reif: Certainly. I mentioned we've got two little kids at home and schools closing and daycares closing has presented us the challenges that it has presented every other parent in those times with successes and failures. I often joked: the little one was 18 months old when everything shut down, it was like really the age that they will kill themselves if you don't watch. I joked, the kid could conjure blades! If you took your eye off for a second, and we only have three pairs of scissors in the house, but he could always find [them]. He would get scissors. All of a sudden, he is tall enough to grab the knives, and so there were a few months of that. Like, oh, my gosh! How many hazards do we have here and how can he always find them? And then, getting the six-year-old, now seven-year-old, to try to do first grade online and use a computer and log in. Nothing unusual, but just those challenges.

And then professionally, there have been opportunities to work with people and think about things that I haven't done, but it's challenging to not be with people, to not be able to have those interactive conversations for you.

There's nothing particularly unique about the challenges, but the kids at home has been my personal challenge and then also opportunity. Right? I wouldn't have spent as much time with them, the boys wouldn't have gotten to know each other as well. Their age gap is enough that they wouldn't have been in school ever at the same time so they've gotten to bond and hang out differently than they ever would have. So try to take the upside with the downside.

Barr: Are you guys doing anything fun to help with the pandemic, any particular hobbies to help you cope with the pandemic?

Motsinder-Reif: Yeah. My husband and the boys especially love [being] outside. They love hiking. So, they've leaned into that kind of thing very much. I like to cook and so I have leaned into that kind of thing more, and just some of the usual things; we got an outdoor movie screen, got a projector and—

Barr: That's fun.

Motsinger-Reif: —on the back deck.

Barr: It's really fun.

Motsinger-Reif: Some of those things and then, Zoom calls with grandparents.

Barr: This is a fun question. What has been your favorite place to go online to kind of relax or unwind during the pandemic?

Motsinger-Reif: Like browsing on a site?

Barr: Browsing on a site or social media.

Motsinger-Reif: I do no social media. So, I will actually say, one of my sanest things is not talking on them. I don't do Facebook, I don't do Twitter, Instagram. I have no presence there. So, that's nice. I guess I get blogs; they are interesting, and I guess Netflix. It's Netflix, it's binging whatever it is, I caught up on. The Crown was the last one I finished.

Barr: Is there anything else that you would want to share as an NIH scientist but also as a person who is living through this pandemic like everybody else in America?

Motsinger-Reif: I am relatively new to NIEHS. Today is actually my two-year anniversary.

Barr: Congratulations.

Motsinger-Reif: I was an extramural scientist. I worked at NC State as a professor and it's been really a privilege to be at NIH and get to watch firsthand science just pivot. It's been amazing to watch this giant organization. A crisis happens and watching people learn new things and start doing research on a completely new problem. It's been really neat and a privilege to see how hard people are working and just sort of the united mission. I guess has been really fun to see that is different at NIH than it was in academia.