"Using computer simulation to help inform emergency preparedness"

Dr. Salah Uddin Khan
June 30th 2017

Hannah Golightly

This afternoon I am pleased to introduce today's speaker Dr. Salah Uddin Khan. Dr. Khan has a PhD in Public Health from the University of Florida and is currently working as a Postdoctoral Fellow in the Department of Population Medicine here at the OVC. Dr. Khan's research is focused on studying infectious disease transmission in populations and assessing the efficacy of health interventions. With that, I'll turn it over to Dr. Khan to take you through his presentation entitled: "Using Computer Simulation to Help Inform Emergency Preparedness."

Salah Uddin Khan

Thank you, Hannah, for the nice introduction and thank you OSRN for inviting me for the talk. So, here is a brief note on the topic I'd like to highlight during this presentation. We'll start with the meaning of the disease simulation model and why and when it can be useful, then move onto a couple of examples of disease simulation research that was done here in Ontario and, finally, finish with some of our group's ongoing and upcoming research on disease simulation in the swine population and how these findings could benefit the industry.

Since most of the audience have a variety of backgrounds, I'll try to speak in generalised terms so that it is easy going for most of the audience and I'll try to avoid many of the technical terms, and absolutely no equations.

All right, so sometimes a solution to a problem is so complex that our conventional tools or approaches do not always lead directly to a solution or to a right answer or, simply saying, we do not have the means to assess what-if scenarios. Here is an example. The real-life problems are often complex in nature and with the simulation models we tend to make simplified assumptions to get as close to the truth as possible and then refine the model until we're very close to the truth or the facts. The figure, here, shows a real-world problem and its solution and how it's mimicked through a model with simplified assumptions. It's just a hypothetical example.

Now, you might be wondering how to define a model. There are many definitions. Here is one that I find handy, describing the infectious disease model in a population, which is a useful, powerful and accessible tool. It's a tool for evaluating questions about the spread of infectious disease within and between many populations and our physical social space and time.

So, why do you use a model -- why and when -- to explain the infectious diseases that are currently not present in a population? For example, in Ontario, FAD is not present so during its absence we can model, given the population we have over here, how the disease would spread, how far would it reach and how many farms would it affect and how many animals do we need to cull or

would die from a hypothetical outbreak. So it gives us an understanding of the extent of the outbreak and helps us prepare for the upcoming events if it occurs. And some of those diseases are very devastating for the population, like classical swine fever, so rather than waiting to see what happens when a disease occurs, it's a proactive approach to see what happens through simulation and what are the interventions, what are our currently known interventions that could prevent these diseases and, of course, the what-if scenarios? So it's all about being proactive and understanding how our intervention would work.

As you've seen so far, I'm trying to avoid mathematical equations and too many technical terms during this talk, so this presentation could be easily understood by a broader audience. Here are the bare minimum data needs to build a disease simulation model in a population. These include locations. The population structure could be the location where the farms are, the size, how many animals are there, and then how they're connected.

For diseases to spread from one farm or one animal to another, they need to have some degree of connection. They need to connect somehow so it could be by the animal movements, it could be by movement of vehicles, it could be movement of people between the farms, or sharing farm machinery or equipment, or simply it could be because of the proximity because some of the diseases transmit through the air or because of close proximity or some of them might also transmit by vectors.
So, this also leads us to assess what pathogen are we trying to model, what type of disease are we trying to model and how each of the diseases has their own way of spreading. The mode and rate of transmission is different. Some could be airborne, some could be direct contact and some would stay viable in the environment a very long time and could be transmitted by corn mites. So, all of these need to be taken into consideration for a model.

So, before we go further, I'd like to talk about two examples of infectious disease simulation models down here in Ontario, then work through our ongoing and upcoming research. The first example is a population-level disease simulation model. So it was done for the province of Ontario. PRRS, was the disease, the model, and to do that, because the population-level data was not available, they randomly generated farm locations within the province. That was based on the 2011 census data and they also came up with the direct and indirect contact rates of the contact and the percentage of the - and then simulated what proportion of farms that could be infected and the size of the epidemic.

These simulation models, the researcher assesses the rate of direct and indirect contact and indirect means of contact, by the vehicles and personnel, and then had both of them put together in a model and say, let's say we have two different scenarios. In one scenario, the model, we only assess what would be the rate or extent of the outbreak and the size of the epidemic due to direct contact. And the second one would say, what would be the extent of the epidemic when we have both direct and indirect contact present for the disease to disperse?

Initially they made several assumptions of the rate and type of contacts between different production type, production classes, the structure of population as well. And they ran this model for over a thousand simulations and then presented their average of the outcomes. So, here, I'm going to spell out the key findings and their implications. They ran the model for a variety of scenarios. Here are the results comparing the epidemic size for direct contact, that is on the left, and the combination effect of both direct and indirect contact on the right.

The direct contact scenario resulted in a lower epidemic size, the median percentage of infected farms varied between 32% to 37%, while the combined -- that is both direct and indirect contact mode -- the percentages of infected farms ranged between 42% and 49% so when they had both direct and indirect contact, the size of the epidemic were larger. The combined scenario of direct and indirect contact also simulated larger epidemic size compared to the direct contact only.

The larger proportions of indirect contact were attributed to the vehicles moving between farms. This means if adequate biosecurity measures were taken in place to limit PRRS transmission by indirect contact, there is a chance that the farmer's would have minimised the size and the extent of the outbreak.

The second disease modeling research I'm going to highlight is also on PRRS but it's somewhat different in terms that, here, the population is not for the entire province, it's only a farm. And then they've broken the farm down to different units. One is the gilt isolation unit and then the breeding and gestation unit and there's farrowing units. And then, each of these, they modeled the disease for each of the animals within these units so the animals were agents for this simulation.

This paper approaches with a what-if view. Here they say, what if PRRS is introduced in a farrow-to-wean farm consisting of about 1000 sows, which is typical of North America, and how would the outcome look like if the immune status of the herd varies? They wanted to compare the efficacy of two type of interventions. The first one is the live activated vaccination and the live-virus inoculation, the second one, to minimize the size of the outbreak in the herd. And they measured the outcome as infection, not disease. So when we say infection it means the virus is present and replicating within the host, and for disease, in general, it says they will have some clinical outcome.
So the figure shows the size of the epidemic after an introduction of the PRRS infected sow in a herd, after 60 days of isolation and run over 100 simulations. It's a very busy slide so I'll go over the figures one by one. The X axis shows the day after virus introduction and the Y axis shows the total number of infected sows.

So, on the top three figures, that is A, B and C, shows the baseline simulated, and so there was no intervention there and simulated the number of infected sows after introducing one, five and ten infected sows. So for simulation A, they introduced one infected sow in the farm and for B, they introduced five infected

sows in the farm and for C, they introduced ten and looked to see the size of the epidemic and how many days it took to develop the outbreak, the epidemic. So this is 100 simulations so that's why it looks busy and it's somewhat different. And so, because this is the stochastic model, every time you run the model, it's somewhat different than previous.

The figure on the modified live-virus vaccination consisted vaccine efficacy they'd consider what if the vaccine efficacy is about 95%? That is for G, and for H it was 80%, and for I it was 70%. The figure for live-virus inoculation first was the varying degree of infectiousness from the distribution, second was the simulation after 56 days and the third figure after 30 days of isolation. So when the breeding sows were brought in they were in the gilt isolation room so they varied the duration of isolation in the breeding room.

So, in all cases, the intervention could not completely prevent piglets from being infected, so they also had concerns over the negative impact of maintaining PRRS positive population as was done through live-virus inoculation. So they concluded the management changes to reduce exposure bacteria to imminent losses a macro-level approach to reduce the infection within the breeding herd, to reduce mortality among infected nursery pigs.

So, they compared three different scenarios of three different immune statuses. One was the baseline and the modified live-virus vaccination and third one was the live-virus inoculation and then they introduced infected sows and see the size of the epidemic and, resulting the nursery pigs, what happens to them?

In all cases they saw that the virus passed through the offspring and then they said, okay, all three of them were somewhat, well, the live-virus inoculation was somewhat reducing the viruses within the breeding sows but the farrowing piglets, the piglets were still getting infected so they recommended for a different approach.

So, now, let's talk about our ongoing and upcoming research. So, as you've seen previously, because of the privacy issues, getting the exact population structure, the right population structure for the pig farms in Ontario is difficult. Famers do not want to reveal their location because of the privacy reasons but for the simulation model it is very important that we get the right population structure to run the simulation. And once we have the simulation we wanted to look at what happens if some foreign animal diseases, for example, classical swine fever, or FAD, gets introduced in Ontario, the commercial swine population, how far would it spread and assess the known interventions. How efficient would that be?

So let's start with our disease population model, how we came up with the synthetic swine population in Ontario. So, why is it needed? Because it's a crucial first step to get the right population structure that we could introduce into the model so the outcome doesn't go very far from the reality and because of privacy issues, we did not have access to individual-level farm data, their locations, their farm size and production classes, so we came up with the

approach to create a synthetic population that mimics the actual population, or very close to the actual population within Ontario.

The first road bump we hit was the agricultural census. The latest one was from 2011 and that had the complete data. Although there is 2016 coming up, we still did not get the detailed information when running this model. So the 2011 census was our way to go.

So, when we looked at the census, it had incomplete data as you can see from this table. The total number of pigs, that is the number of missing information. Almost half of them, in many cases, were missing. So, with so much missing information we are struggling to get a grasp on the population structure in Ontario.

To add to that, the census doesn't give you the location of the farms so we figured a way around. We said, okay, given that the information is not enough to find locations of the farms as well as the production classes, the different production classes and the sizes, the number of animals within each farm, we will try to look at the factors that influence the farm structure. What are the factors support a farm within the province of Ontario?

So there are a couple of them. It could be zoning and local regulations, the farms need to be away from the population structure, the population centers, the arable places and they should be within the agricultural zone and the availability of the land, it should not be on some of the government's or recreational zones. The farms need to have access to transportation and there could be geographical barriers like rivers. You cannot place a farm on a river.

So we pulled up all this information. How? So, what we did, we broke down Ontario into one square kilometer grids so there were about a million grids cells like this one over here, and then on top of that we put on different layers of attributes that would support or influence a pig farm structure in Ontario. So in this figure here you see a river. On top of that we added the road networks and then once we had this information we say, given this information, how far a cell is from a river or a road or a population center and then calculated the distance.

So once we had these numbers in place we multiplied them and divided them by a factor of 100 to get a combined suitability surface and that score ranged between zero and one -- here you can see that -- and then plotted these scores over Ontario so Ontario looks like this. As you can see, the places that are darker are more likely or more suitable to support a pig farm structure in Ontario.

So, since we got the score we said, okay, we had different attributes that would generate a model. How to validate the model? So, to validate the model we went OMAFRA and talked to the previous registry section and they helped us out with some anonymous data that gave us the actual location of the previous registry data, to say here are the actual locations and here is your model, and now you can compare and see which model works better.

So the model we preferred had a very high area under the curve as the accuracy metrics, that consisted of proximity to road networks, population center -- that is away from the population center and residential zones -- away from the camp and recreational sports, away from the government and institutional of land uses and then away from the other crown lands and within agriculture ecumene. So this combination worked better. So, when this combination worked better we said, we need to define a cut-off within the scores, we had score zero to one. Zero means that it's unlikely that this place would support a pig farm and one means it's absolutely likely that you could support a pig farm. So we need to draw a line where we could say, above this point, we would populate and generate pig farms.

So we ran a bunch of our statistics and two of them were true skill statistics which is sensitivity and specificity minus one, as well as Kappa, and then looked at where it had the highest specificity and then defined a cut-off that, for this model, it was 0.82. So for anything over 0.82 and above, we took them and separated them. So for those surfaces we overlaid the census data. The census data, the smallest unit was the CCS, that is census consolidated subdivision. For each of the CCS, they will have a number of farms.

Let's say for CCS A, that's an example, they had 100 farms. So we say, okay, within CCS A we have these many surfaces that scored 0.82 and above and within this surface we randomly generated 100 farms, so that was proportionate to the number that was given at the 2011 census. So once we have those random points generated for the farms -- here you can see red dots -- these are random points based on the likelihood score and proportionate to the 2011 census.

Then, from the secondary data, we defined the class. How many of them would farrow to finish, how many of them would be nursery piglets? So this types of production classes, as well as the number of animals. So, our next step would be, once we had this population structure, we would look into several foreign animal diseases that could affect the Ontario swine industry and see what are the known intervention and their efficacy. For example, it could be vaccination, it could be zoning and culling and what are the combinations of intervention that works better than others.

So, in summary, synthetic population that we generated, that captured the key characteristics of the observed population and also preserved the privacy of the individuals so these points don't lead back to the individual producer. Rather, it gives us an estimate of how the swine farming population is structured over here in Ontario.

For the FAD, foreign animal disease modelling, we are trying to be correct and assess the interventions, vaccinations, zoning, culling. So, many of you might know that the OIE has come up with some changes to their regulations regarding zoning so, right now, the latest we have is from June 2016, that says that zoning would be done through an agreement between the planning partners so the sentence is somewhat loosely defined. So as well as your

partner for Ontario here, we would say United States would be our biggest partner for trading for our pigs and for our pig products.

So we need to come up, whenever a foreign animal diseases hits the country or the province, we need to come up with a rationale and say, hey, listen, we had an outbreak, we had a zone that is of three, four, five or ten kilometers and we did this amount of culling and this amount of surveillance, and the rest of the province is free of disease. So instead of calling out the entire country as infected, we can set zones and with this simulation we can come up with a rationale and say, within this zone, given our model, we see we can contain the outbreak in this given space and we can still continue our trade so, in that way, we can minimize the loss of trade.

Also important that these interventions could give us a heads-up of how much resources do we need whenever this CSF or FAD hits the provinces and what are the interventions that would work? So, given that our advanced knowledge on these interventions, we would be better prepared to tackle these kinds of devastating effects on the industry.

So, with that, I would like to thank our team members. Dr. Greer, she's leading our research group, as well as Terri O'Sullivan, Zvonimir Poljak as well as the people from OMAFRA, Janet, Tim and Alexander, that helped us with the data, as well as review our recent work. So, thank you and happy Canada Day. And if you have any questions, please.

Hannah Golightly

So if you have any questions please feel free to type them in using the Adobe Connect chat function and in the meantime I'll ask Dr. Khan speak to some of the questions that he has prepared, some of his frequently asked one's, on his research.

Salah Uddin Khan

So as we're talking, why population structure is important for infectious disease modeling? So, as you have seen from this example that we've provided, infectious disease needs hosts. The size of the host population within each units sphere, for agriculture let's say the size of the farms, how many animals are within that farm, how they're connected, are they in one pen or are they in one enclosure, are there a couple of farms that they're in very close proximity or are there farms where the animals move regularly in-between? So if one farm is infected, the chances are the farm that is connected to it would be infected after some time. So this population structure and the connection between them is important to mimic the infectious disease spread models.

So the second question would be can we develop models to see disease transmission within farms? So that was one of the questions I was asked before so I, kind of, brought up the second example where Drs. Greer and Poljak, they showed that, yes, you can also define a model to look at the disease transmission within a farm. Here, instead of having the farms as unit you have the animals as a unit, as an agent, and then run the model.

Another question that pops up often is that whenever you run the model, the stochastic model, the outcomes are often a little different from the first one. Why is that? It's the disease is a biological process and it's also because, to capture this biological process, we define the model to be somewhat different. Let's, for example, the second paper we discussed, we showed here that the authors wanted to capture the uncertainty of the contacts, the rate of contacts, the amount of contacts an animal would have every day. It could be they could have been contacted by two infected animals or one infected animal or five uninfected animals, so there is a bit of uncertainty. It's a balance of the process. To capture this balance of the process, we enter this uncertainty within our parameters so every time we run a model, it's a bit different. So how do we present these outcomes?

So, since the model is somewhat different, not so much in most cases, so we run it for several times; a hundred or thousand times and then average out, say, what's the average number of animals being infected after running this model? So we could say the average would be we would have ten infected animals and it could vary from five to 12. So that is how we present our findings and, because it's a biological process, it would always vary somewhat.