NTDS_35

<u>Key:</u>

I: **Interviewer**
R: Respondent

I: Okay. So, I would like to start simply by... if you want to quickly introduce yourself and your broader role in PHE, and then move on to how you got involved in MEDMI.

R: So I'm Gordon Nichols; I'm an epidemiologist. So I spent the first 20 years of my life doing laboratory work, and then converted over to epidemiology half way through my career. I'm a scientist, a PhD trained scientist, not a medic. My background's been microbiology and epidemiology in infectious diseases in humans, so that's my main interest. I've worked on parasitology, bacteriology, virology and various aspects of that in my working life. Most of the activity has been related to diarrhoeal diseases, or gastrointestinal diseases in general. In 2013, I was seconded to the European Centre for Disease Control in Stockholm for a year, to work on climate change with Jan Semenza in the centre in Stockholm. As a response to that, and the people working in climate change here thought it would be good if I was involved in the MEDMI project. So while I was there, I was co-opted onto the project, which I wasn't originally a researcher on. That's the reason that I'm involved in MEDMI.

I: It's interesting that you had this gradual conversion from microbiology to epidemiology, into your career. How did that occasion...?

R: Right. I guess my PhD was a mixture of microbiology and epidemiology. I took a relatively new pathogen called cryptosporidium, which is a parasitic protozoa, and I worked on both the biology, the typing and the epidemiology of that pathogen. That was really the opportunity to switch, at that stage, rather than carry on in diagnostic work, to move over to the epidemiological side.

I: Okay. And was that an unusual combination?

R: It's not the mainstream route for scientists in epidemiology, but it's not unheard of. People do it. The career paths for scientists in epidemiology, in the microbiological side of epidemiology, are a bit diverse, a bit unplanned, so I think there's probably a need for better career routes for scientists within epidemiology.

I: Okay. Why are then unplanned? So there's no clear disciplinarity?

R: I guess it's different for the medics and the scientists. So I guess the medics have a clear progression. I think with the scientists it's much more diverse, and there's less infrastructure, infrastructural support for the scientists in moving forward. I think there are arrangements for training medics that aren't as well-developed for the scientists.

I: Okay, very interesting. So, going back to MEDMI, then. Basically you got involved some...?

1

R: It was probably in 2013, yeah. And so the project... I first started working on it in 2014, I think. I was when I came back from -

I: Stockholm.

R: - ECDC.

I: Okay. So what expectations and what goals did you get involved in MEDMI? How did the people who brought you in...? Who was it that brought you in and what was the expectation for you?

R: So it was Laura, Laura Fleming, brought me in, with help from (s.l. Cetiris 00:05:43) and probably there may have been one or two others who facilitated that because they know of me. So then I guess the project was a little bit light on people's microbiology experience, so because I've worked for 40 years on microbiology and epidemiology, and because I was working on climate change, they thought I was a good subject to work with. It also tied in with the stuff that I was doing in Stockholm.

I: Okay. And so I know you worked closely with Majid, obviously, on this project, who probably had been working quite a while before you arrived on the project.

R: So Majid was already... I think he was already appointed before I got involved in the project. I've worked with him particularly on the campylobacter side of the project. Campylobacter's a food poisoning organism, and so it was particularly on looking at the influence of weather on campylobacter. He's come from the finance industry, so his time series work with the finance industry, looking at time series models, fitted in with the analysis of time series in climate. So that's that.

I: Okay. So what was the situation with MEDMI when you started? What did you need to do in the beginning? Do you want to basically summarise the story of the project?

R: Okay, so going through from the start... I guess it has to be said that in addition to MEDMI, there's another project, which is the NIHR funded project on climate change. Or... it's environmental change and health. I'm involved in that, and so some of the activities... I've tried to keep them defined to the individual projects, but there's some crossover. So there's a campylobacter project within the NIHR work, and also a campylobacter project within the MEDMI one. The campylobacter made... the people who'd organised the project had flagged up as something they wanted to be worked on in the MEDMI project from the start, before I was involved. So I carried on with that. The way I split it up was, with the MEDMI project we were working with data from the UK, predominantly, and with the NIHR one, initially the project was on European data. There are other things being done in the NIHR project, so it wasn't just the campylobacter project, but I wanted to keep individual parts of the project separate so that I couldn't be accused of double-counting, which might be considered unethical.

I: Okay, so this other project doesn't involve the same people as MEDMI, I imagine?

2

R: So there are some people who are across both projects.

I: **Okay. And this NIHR, was that antecedent to the MEDMI or parallel? Did it start later?**

R: I think the MEDMI project was prior to the start of the NIHR. So one of the difficulties with the MEDMI project was I think it was a bit slow in starting up, as I understand it.

I: **Yeah. This has come across in so many ways.**

R: Initially the campylobacter work was attractive for that, in that we already had the data. So it allowed us to do some of the things we wanted to do in terms of method development with existing datasets. With all this climate stuff there is a problem in actually getting the data that you want to get in order to do the analytical work. So that worked out quite well, because we could start off with an existing dataset.

I: **Okay. So the data that you had already, was that health data? What kind of data was that?**

R: So it's data on campylobacter cases reported through national surveillance, and then linked to weather data. About five years ago, with people at the Met Office, I arranged a project linking individual cases per day to the weather conditions in the previous 30 days for temperature, rainfall. So it was local data linkage. It was done through the laboratory postcode, and that was because first of all we don't have the domestic postcode of all patients.

I: **So of some patients but not all of them?**

R: You would miss data. And secondly because there's an issue of confidentiality in terms of using the patient postcode. It's argued that patients could be identifiable through patient postcode. I think in the way that we use the data, it doesn't make any practical difference from that perspective, but it's one of those things that it's important to make sure isn't compromised using human health data. So we'd previously done this data linkage, and we got the datasets, and therefore we could apply a range of different approaches to analysing the data, to see if we can improve on available techniques. So it really is a method development approach to try and improve those methods which have been used in the past.

I: **So you mean the MEDMI project is a method development approach?**

R: The MEDMI, yeah.

I: **Why again? So you had developed this work before in the other project, five years ago?**

R: So the other project... we hadn't developed the methods, we'd just done the data linkage. The results we'd published from it were a very simple analysis. So it was really taking that dataset, reusing it...

I: **Demonstrating that this kind of work is...**

R: Can give you more information, yeah.

I: **Okay, I see. So basically, part of your involvement was also bringing this work into MEDMI and making it available on the platform?**

R: Yeah.

I: **So in that sense of the various dimensions of data that MEDMI is about, environmental, weather... this kind of data linkage project had already all the datasets that it needed, or did it need to be?**

R: It had the datasets... we agreed at the start that we would use the data we'd god, not get additional data, but that we'd look at methods for extracting data in the same way, or in a similar way, that we could apply to a much larger dataset. And obviously the infectious disease data we hold over the last 25 years is about 13 million records, I think, so the intention is to try and build up a way of systematically applying that data to a larger range of pathogens so that we can then do some analysis on a range of different pathogens, and not just the one we're working on. But using the campylobacter as a good place to start, because there are a lot of cases, so we can take a large volume of data and then develop the methods using that large volume of data, that can then be applied, hopefully, to smaller numbers of cases.

I: **Okay, yeah.**

R: So the other intention was to work up the methods in order to apply the data over a larger time series. So the dataset we had was for five years. We have 25 years of data, but in the past we've only been able to link five years of data. So the intention was to develop a wider range of weather parameters, and then to do it over a longer time period, and then come up with some answers to the question of what weather influences have on campylobacter action. So that was the overall approach, and as part of that, Majid was also involved in extracting human disease data and then, with the Met Office, linking cases to... So he's not involved in the linkage, he was involved in extracting the data, getting the ethical approval, etc, etc, and then... It was Christophe Saran at the Met Office who was working on the methods for extracting the data. Doing the data linkage is actually quite a process-intensive job. I think it's been a job to develop the code for doing it, and also it takes quite a long time to process, even though they've got one of the biggest computers in the country. In addition to that work, there's also been some work, some preliminary work, on the seasonality on the range of different pathogens. So obviously if you're looking at the effect of weather, weather changes through the season, and so we've done some preliminary work on how the different pathogens are affected by the season. We did a small sub-project through Mark Cherry. So Mark has done a paper which we still need to finish off, looking at seasonality across the whole range of pathogens, to try and give us a better focus on what organisms might be affected by climate change.

I: **Okay. So this work, for example, and the other strands of work, how interdependent are they?**

R: I think they all fit into the overall aim, which is to gain a better understanding of how climate and other environmental variables influence infectious disease, but they're not dependent on each other. I guess the initial campylobacter project... probably subsequent work to it is quite dependent on

the initial project, but that work will go ahead anyway. And given that the next MEDMI project may not have infectious disease within it, it may be that it's taken forward within the NIHR project, because that has a slightly longer time period to go.

I: **Going back to bringing in the data into MEDMI and linking it... so these two phases were separate, that you told me about? Part of it was about bringing in linkage that you had worked on five years ago and making that available, but then there was also work of generating new links?**

R: Yeah.

I: **Is that correct, with Christophe? So the linkages that you had already available were, is it, five years? So that was basically...**

R: It was Christophe who did that originally. But he's also been working on developing new datasets.

I: **Okay, the work five years ago of linking... So basically you had this five years' work of linkages that you could bring into MEDMI and you could already start work on from a research point of view. And then, in parallel, you could also extend this method to produce new linkages over the 25 years. So this is going to be also linkages that you will exploit in future work?**

R: Yeah. The future work may not be in MEDMI; it may be in the NIHR.

I: **Could you say more about how the data linkage is an intensive process? Beyond that it obviously involves great numbers of records, and then the computing needs to be long.**

R: I guess what's involved, what's originally involved, is you have a geographic location and a temporal location, and you want to marry those two together and give the equivalent of weather before that period. So the individual record, geographic record location was the specimen data of the patient and laboratory. And so we used the laboratory postcode because it was very complete. Every record has which laboratory it was tested at. And so we had the postcode of all the laboratories, and we thought for the purposes of weather that the local laboratory would be a good enough surrogate for the individual person's postcode. So it was equivalent, within a few miles, so the weather probably wouldn't have changed all that much. We don't really know to what extent the weather parameters would be affected by those different distances. One of the small sub-projects has been to try and test that, to test the difference between data derived from the postcode of the patient and the postcode of the laboratory. We can use that with the subset of cases where we have the domestic postcode and the laboratory postcode. We can do that algorithm for linking to the weather before that. We've had quite a lot of difficulty in getting the data for this. It's taken a while, and I think Majid has that data now, but I still haven't seen an analysis. I'm hoping that it's going to show that there's very little difference.

I: **There is no difference, yeah.**

R: There's very little difference. But we need to be able to prove that, or to determine how much difference there is.

I:　　**Yeah. So this is all still ongoing work.**

R:　　Still ongoing, yeah.

I:　　**It's too early to tell. So why does it take a lot of time?**

R:　　I think it's to do with the intensive nature of the processing, and writing the right computer script to do it. It's as simple as that.

I:　　**Okay. So there had to be many attempts, and refining...**

R:　　Yes. So it's taken much longer than we thought it would to get that data out, and I think that's down to the Met Office taking a long time to produce the data in the right form.

I:　　**To produce these links.**

R:　　And still now we don't have all the data that we want. So there's still some pressure needed to try and generate the datasets that we want.

I:　　**Okay. Why is that difficult to produce these things on the patient level postcode?**

R:　　It's to do with the format the data is held in at the Met Office. The Met Office holds weather parameters that vary from changes per minute to changes per year, or decade, so it has just a vast amount of data, and extracting it in exactly the right format is more difficult than it seems. Some of the datasets are held in the format that's relatively easy to extract, and some there's no published data on the time scale and resolution that we want. The way that the weather data is held... You have ground stations, and the ground stations record the weather, and so there's gaps between the individual ground stations, and the different parameters measured in different ways. So they have to do interpolations to estimate the area between the ground stations and then produce the data in the format that they want it. So some things which I would have thought would have been easily collectible, like sunlight, they only produce on a weekly, monthly basis - it's monthly, I think - so we haven't been able to get daily or weekly sunshine recordings. It's the same for a whole lot of other parameters. It's trying to get the exact parameters we want, that's more difficult. I think it's to do with the ease, or lack of, being able to extract it.

I:　　**Okay. So how was the process of deciding what parameters you want, in the sense of interfacing with the other colleagues, and with the colleagues in the other operating institutions, in the Met Office?**

R:　　Well, I think... We've supplied lists of what we want, and I think it's...

I:　　**How did you come to an agreement, since it is not an easy answer? So you had to figure out and discuss what it is that you need.**

R:　　Well, I think it's predominantly... We asked what they'd got; they told us what they'd got, and then we said, "That's not quite what we want," and then, "This is what we want," supplied details of what we want. So we've got some of what we want, but we haven't got it all yet.

I: **Right. And were there consequences of this pace, and issues in the pace of the research or in the research questions you wanted to investigate?**

R: Yeah. Obviously if you can't get the data you want, it's more difficult to... But it didn't stop us from doing what we said we would do, which is the analysis of the data we'd already got. So that's been the main thing, to do that data, but also to work on the methods. So we've done a lot on developing the methods, and also on planning what we're going to do in the future.

I: **Okay. But would you have also wanted to work on the new links if you had done... earlier?**

R: Oh, yeah.

I: **So it was good to have this other option to...**

R: Yeah.

I: **And in respect, instead, to health data, which is typically more... it's got confidentiality issues which need to be sorted out. In your project, what was the situation? You had all the data you needed, or did you need to access it?**

R: For the MEDMI project, a lot of the data confidentiality side was sorted out by Majid. So he went through the data governance side of it, and that was done with Laura, I think. So I didn't get too involved in that side of it. The data we actually used, I've already used previously, and I'm confident it's extracted in a way which is non-identifiable, and so there's no issues of confidentiality associated with it.

I: **Yeah. Okay. And so what do you...? What do you think of all the potential of the MEDMI to make these linkages to other researchers? Because you have a very good understanding of the datasets and the linkages because you have produced them. I would be curious to know... How is it important to know about the kind of challenges and the kind of questions that you were thinking about when you were producing the linkages in order to use them well?**

R: Okay. So I guess for infectious disease work... The issues with infectious disease are... there's a whole lot of stuff, a whole lot of different parameters that can influence infectious disease apart from climate. So climate is an important contributor, but it's only one of a range of different contributors. So that has to be born in mind, and I think in the infectious disease world, that probably explains some of the lack of enthusiasm for working on climate change, because so many other things affect cases that working on climate change might be thought to be a bit trivial. However, we believe that it's important to understand the influence of the environmental and weather parameters in order to get a better understanding of the epidemiology, in that it's the epidemiology that's going to influence what happens in the future. Some of the things which are quite obviously affected by other parameters, or maybe are vaccine preventable, it's still useful to understand the influence of the other parameters on those.

I: **Yeah. If you were to give recommendations for MEDMI 2, what particular kinds of areas would you focus on to realise more of the potential of this?**

R: Well, I still feel there's a whole lot more that the infectious disease side needs in terms of understanding future climate change, and in terms of use of data. So what we've done first part, in the first MEDMI project. It has been really valuable. First of all in looking at the data linkage mechanisms that might be useful for analysing the effect of climate, but also in working on some of the methods for analysing that data. The fact that you've got cases where you've got a mixture of geography and time means that when you apply something like temperature to that, you've got three different parameters, and you can look much more closely at how those parameters interact. The seasonal distribution of cases, for example, has an element which is probably related to the weather parameters, but you can separate it out geographically, because the temperature differs across the country. So if you apply it in that way, you can learn more than if you just used the averaged data. So I think we've demonstrated that local data linkage has a strong potential for answering some of the questions which weren't very easily answered by using just averaged data. It's providing a greater resolution, and ability to tease out the weather parameters from the seasonal and the geographic.

I: **So what would be the requirements in terms of the composition of the teams, and skill sets of the team for MEDMI 2 project, or a repeat of MEDMI 1. Because I think there's been some traces of this, obviously...**

R: Okay, so the difficult...

I: **It's very interdisciplinary.**

R: It's interdisciplinary. So it's good having experts on the particular mathematical approaches. So Majid's been very good on time series, on different methods in time series, but he's got limited infectious disease background, and so that's... It's not a big issue, but it's easier if someone's been through infectious disease training, to relate some of the complexities of the biological work.

I: **Yeah, of course. Somebody with your background, would you have had enough mathematical expertise? Does this new kind of work exceed the training that health scientists or biological scientists have got in mathematics? Because there's an interesting...**

R: I guess epidemiology is inherently a mathematically intensive discipline, but it also requires the infectious disease understanding as well. I mean, I've done courses at... both European and university-based courses on mathematical epidemiology, so I'm quite comfortable with methods, although I don't necessarily use them or do them directly, I understand the concepts of them. So I don't think I have a problem with understanding maths. I have maybe some details with the details, but people like Majid are good at going through the details. So that's not been a problem.

I: **So it is important to have very advanced mathematical experts on these teams?**

R: It is. But it's also important to have good epidemiologists. So understanding... One of the projects outside of MEDMI that we're doing is looking at methods that have been used in climate change work, so mostly mathematical methods. Some of the comments from my colleague (s.l. Jani 00:41:03) have indicated that in some situations it looks like the method is already in place, and someone just applies a dataset to it and produces a paper without necessarily understanding the epidemiology. So yeah, I think you need a mixture of both the epidemiology and the maths.

I: **Right. Is it important to understand the underlining epidemiology to understand the phenomenon that is in discussion in the paper, or is it more also to draw the consequences?**

R: Well, both really. I guess with the... Is the noise level getting...?

I: **I think it should still be...**

R: It should still be okay for you? Okay. No, I think the... Is it going to be a problem? Say it again.

I: **I was just asking you why having the epidemiological knowledge is still important, and I was, for example, asking is it... Because I can imagine you can have an epidemiologically very consistent analysis, but is it more important to also draw, then, implications and consequences of the...?**

R: Yes. Understanding what the results mean is very important, yeah, and also understanding... One of the issues of climate change is you've got predictions through a model of what's going to happen in the future, but some of the difficulties with those predictions are that there may be other things which affect, rather than the model, which affect what happens in real life. So an example would be that salmonella is influenced by temperature. Salmonella grows up in food at warmer temperature which is contaminated, and then you get food poisoning. So when it's warmer, in theory you should have more food poisoning. In practice in Europe, we have an intervention campaign to reduce salmonella through vaccination of chicken flocks, and so salmonella cases are going down, even though it's getting warmer. So that's the sort of thing you have to understand. It's where the disease fits in the overall setting, rather than just a simple relationship between one real parameter and the cases.

I: **Okay. I think we have exhausted the questions that I wanted to ask you.**

R: Good.

I: **If there's anything else you would like to add, or some kind of interesting comment.**

R: So management of MEDMI, I think, has been broadly... I've been broadly happy with. I don't know what's going to happen with the next stage, whether there will be a MEDMI 2, and if so, whether it'll include any of the infectious disease site. We've still got work to publish yet, but we've got a couple of papers already. One's been submitted, and the other one we're still working on. So there will probably be some additional work out of it. I would have said, all in all, the infectious disease part of it I've been quite happy with. That's not

a criticism of the non-infectious-disease part, it's just that I haven't had too much to do with it.

I: Yeah, of course. Okay. Thanks a lot.

(End of recording)