

eIUS: Disease Control Experience Report 2

In the text: Some barriers in ‘{...}’ to provide more contextual information.

Interviewee profile

Senior researcher at an epidemiology department at a UK university

Time spent in research

“I’m involved in research probably about 75% of the time”, i.e. “personally doing research is probably about 25%” and “supervising post graduate students” accounts to “perhaps 50% of the total time”

Research area and Research question(s)

“Well, my research area is very broad in that my main **focus is on the epidemiology of diseases in animals**, so looking at population level, patterns of disease in animals and what affects who gets disease, why and when and that sort of things across a whole range of species, and they are often, farmed animals, cattle, commanding animals, dogs, cats, horses, but also poultry and wild animals so on as well, and a whole range of diseases, so **to some extent it’s a matter of applying a whole range of tools to address various different biological and veterinary questions**, and also I’m actually **conceptually interested in interdisciplinary research** because often those sort of questions bring in lots of different people, and recently I’ve actually become quite interested in starting to **work with people in the social sciences**, thinking about how animals diseases impact on human activity and wellbeing, but also how human activity impacts on the risk of animal diseases, so thinking about things from a range of angles, but it’s actually hard to pin down exactly what I do, I find.”

Research question (in detail) in the Avian influenza/poultry project example: “(..) the basic question we asked was, given that we think that poultry farms interact in this particular way with these probabilities of interacting with each other and that sort of thing, **if Avian Influenza occurred somewhere in Britain, first of all, where is it likely to go geographically**, that was the one question we wanted to ask, so are there particular parts of the country where it’s more likely to cause an epidemic than if it perhaps goes somewhere else, how many farms, how many individuals are going to be infected in an outbreak is clearly an important question for a number of reasons; one because it impacts on those people but from policy makers point of view, they need to know do I need to have ten people on standby or a hundred people on standby to try and control this outbreak, and then the geography comes into that as well because if it’s ten farms in one little corner of Britain it might be easier to control than if it’s ten farms scattered all over the place, so we’re trying to get an idea of numbers and how that might impact on control, those are the main questions, but then we’re breaking it down into which particular breed of bird, for example, what type of farm is more affected; whether you’re more affected if your part of the integrated company or if you’re just a little independent person etcetera, and whilst we have a lot of this

data for these farms, the data wasn't perfect so we're trying to help identify, well we think that these types of farms with these characteristics are going to be of more risk and you should be more concerned about those so that if you've got limited resources you can target these areas rather than those areas and that sort of thing, so there was some quite specific knowledge we were after which was about where is this likely to happen, where are the key areas, but some more general questions as well as to what sort of types of categories of farms are we most concerned about, so we're trying to summarise it across those different scales of information."

Research Lifecycle

Literature Review – Start of the research process

"Often we're **starting to think about areas that we have some knowledge of but not a detailed knowledge** of that particular area because it's a broad approach to different research places, we're not always tackling one particular biological issue which we might generate extensive knowledge for ourselves, we're often branching off into areas that we don't actually know a lot about; we might be working with people who are experts in that area, so **literature review is often the very first and really important thing to do**. To be honest, a lot of the research that we do is driven by, like everybody's perhaps, research calls; we're funded through a lot of government money and that sort of thing so a lot of our research projects are **responsive to the funding opportunity** rather than having an idea then we go out and seek particular funding for that particular project, so that often means again we need to **rapidly generate quite a lot of knowledge through literature and so on about a particular area.**"

Literature reviews and consulting with experts (especially in the project with **interviewee 1**, see exp report 1): "(...) particular project requirement for a commission by government to look at this issue and the literature review was a key component to the model building, and that was really important as well as consulting with particular experts in particular fields, so just drawing all that information together was the first step, but actually it continued throughout the fairly short term project; it was only six months, but literature review went on throughout the entire project pretty well up to the end."

Example (Avian influenza/poultry project; also includes start of data gathering towards analysis): "(...) with the poultry project, we took our knowledge of the poultry industry as it exists, we did a lot of reading around about the industry and the disease that we're dealing with, Avian Influenza, consulted some experts, both veterinary experts but people also from the industry as well, and we actually sat down and drew what we thought were the **important compartments within this model**, the different aspects of poultry breeding and that sort of thing and how they interrelated to each other and then we started thinking about how we prioritised this thing, so we didn't necessarily read the literature and say okay, there's an important parameter, we need to do this in the model; we thought about what the model was and how; I mean basically the basic unit of the model was a poultry farm, and so we thought how do poultry farms relate to each other, so for example, they might belong to the same company, different farms are in the same company, there might be an independent farm, what do they do, are they producing eggs or are they producing chickens for meat; it's a very structured industry so the parent birds don't get used,

really, they just get used to produce birds that will then produce eggs for example and there's grandparents and so on, so it's a highly structured and organised industry, it's a very bizarre industry to some extent, so we actually sat down and tried to draw up a framework or a model of the industry and how these units were interested in the farms, how disease might transmit between them through various mechanisms and then sort information on how we parameterise that, be it through the literature or by talking to experts, or a bit of both, but for many it was actually very difficult to work out what was the most appropriate parameter, which I guess feeds into why the **computing was so important** because actually then doing sensitivity analysis where we can consider a whole range of parameters, the whole range for a particular parameter and do that in sort of a multidimensional sense where all parameters are varying a lot and then try and work out which parameters are most sensitive, would have the greatest affect in the model was actually quite an important step I think, because it was **such a data-sparse area** that we were working in."

Data collection process

In reference to the example under 'Literature Review – Start of the research process' (see last paragraph): "(...) there was an overlapping process" between data collection and data analysis.

Databases/tools/sites for literature/data collection: "(...) online databases were vital; PubMed's useful; I tend to use a range of them because they bring up different things; there is a specific veterinary one as well that online **search engine** that tends to report some of the journals that never make it into PubMed and so on and also has some conferences and that sort of thing in there as well so that's quite useful, so yeah, those **online databases** were fundamental." And: "**PubMed, Web of Knowledge** are probably the two main ones, but I try not to just do the search in one engine, I'll try and do it in a couple and then **cross reference** because I often find that they don't pick up the same." (PubMed: <http://www.ncbi.nlm.nih.gov/pubmed/> ; Web of Knowledge: <http://www.isiknowledge.com/> ; <http://wok.mimas.ac.uk/>)

Modelling/data collection (exemplifying the research practice and topic); also in reference to the project with interviewee 1 (see exp report 1; to contextualise this particular project within the general research of this interviewee): "Only a part of what I do is this type of modelling, this type of **mathematical simulation modelling is actually a fairly small part**, so I guess the data collection from this came from several sources; Defra [<http://www.defra.gov.uk/>] had collected some information through questionnaires and that sort of thing that was to help parameterise the model so we got access to that and we were able to analyse that. Defra also instigated what was known as the **National Poultry Register** [a database], which was where everybody who kept more than fifty poultry birds had to register that they did that and some information about how many birds, how they were kept, where they are and that sort of thing, so we actually had that providing information on the spatial scale, but also some background information about the types of birds and so on, which was vital, so that was posted by Defra but it was collected by Defra at the same time we were doing the study, so we were getting a lot of **very raw data with lots of input areas** and that sort of thing in it so both Defra and us were trawling through that, trying to work out what actually it all meant and what was real and what was error and that was quite time consuming. We did some more **qualitative interviews** with

people from the industry and that sort of thing, so I'll give you an **example from that study**; we were interested in the role that a feed lorry might have in transmitting the virus between the farms because there is some suggestion that, because the feed lorries often go from one farm to the next before they go back to base and that sort of thing, so we discussed that with people from the feed companies just to find out things like how far a truck might go on a particular route, how many farms it might visit, etcetera, etcetera, and through other things like, that it might go from one feed mill, visit three or four farms, end up at the other end of the country at a different feed mill, fill up again, and then go back to its original feed mill via three or four other farms; but then other days I worked things like how long the virus might survive in those sorts of environments, risk of transmission, we also needed information about where the trucks are likely to go (.); if you've got a company, do the trucks tend to stick within the company, how often are they washed, etcetera, etcetera, so a lot of that was collected by qualitative interviews, in terms of data about, for example the difference of the effect of Avian Influenza within different species of birds so there's some quite important bird differences, so ducks, for example, can be infected for a very long period of time and not show any clinical signs so you don't actually know they're infected, whereas chickens rapidly become sick and die, so you can reasonably rapidly detect they're infected, so there's these differences within breeds so that was mostly going through literature and talking to some experts, but the other sorts of things I do, I guess I'd break it down to possibly two main things, one is we actually do quite a lot of **fieldwork where we get out and we're actually sampling from populations in the field**, or **using stored data**, so again, we do quite a lot with **National Statistics** and various bits and pieces which tend to be **very large datasets**. For example, one of the main ones I work with is known as the **British Cattle Movement Data**, so perhaps surprisingly, every location in Britain that wants to keep a cow has to register that's its going to do that, and every cow has an ear tag that's a unique number, and every time a cow is born or it moves from farm A to farm B, or it dies; all of that is recorded, so you end up with a dataset which tells you where all the cows are at any one point in time and where they move and that sort of thing, and that's **millions of rows of information and we spent quite a bit of time crunching through that** as well."

Access to data (secure/confidential data): "Those data are available from Defra under contract with a confidentiality agreement and that sort of thing, you can apply to have access to it, but it is all obviously confidential information so that raises other issues, so it's not strictly available online, no." Which means it is sent on request on a CD.

Confidentiality and security are important, e.g. for British Cattle Movement Data, which is not available for the general public, but to which the interviewees **institute has subscribed to**, i.e. members of the project/research group have **access to it on the institute's server**: "Yeah, we have it on server and we just use an **SQL frontend** thing and interrogate it that way. (.) it's all **password protected** and that sort of thing."

Data analysis

Analysis process (in general and example): "Again, it depends on the data and the question. With the stuff that we did with (.) [**interviewee 1**] that was not the analysis

but the simulations were done using **Condor** and then from that we extracted summary data that were analysed on PCs using various appropriate programs and that tends to be the way; most of the analysis we do will be done **on local desktop PCs.**”

More on analysis with interviewee 1’s group (example plus collaboration; also: solving barriers): “Yeah, I mean it was actually a really good group to work with because (.) [interviewee 1’s] background was not in biology but he’s very able to adapt and to talk; one of the issues we crossed disciplinary and interdisciplinary so it was just actually communicating with people, but both (..) [interviewee 1 and a Professor] who is a mathematician here and was (.) [interviewee 1’s] principal supervisor [in the past], I think this group had a strong ability to communicate across the [domains], so [at points] they couldn’t understand what we were saying and we couldn’t understand what they were saying; so that was really useful [communicating across the domains], (..) [interviewee 1] did do quite a lot of the interpretation and really we did it as an intuitive process so we’d **start off with a fairly simple breakdown of the data that was coming out of the simulations and stuff and then drill down a bit more into it to see what we could pull out of it**, but with that number of simulations and so on, actually summarising the data was difficult to get all the subtlety and nuance that the data held. I thought it was quite difficult to summarise that appropriately.”

{ {“Well if you use the example of this cattle movement database; the biggest problem with that is that it’s such a detailed database that you can ask so many questions of it, and so the biggest trick we’ve found is to really formulate what the question is that you’re asking and why you’re asking it and that sort of thing, and then go and look at the data, whereas if you just start looking at the data; you can learn so much and there’s a lot that’s interesting there but it actually becomes a very disjointed process (..)”.} }

Impact of Research question on analysis (plus example and collaboration): “(..) we’ve found it’s really important to determine a question, and that’s usually a policy based question (..) and often we’re trying to see whether a particular policy change has had an impact on the cattle population and the way people farm, so **we’ll break that policy down into different aspects and try and extract the data**, and again we work at different levels here; we, within the house **do quite a lot of analysis ourselves**, so for **example**, for that we’re doing a lot of **time series analysis**, but we also **work with people from the stats department in Lancaster**, and they take a more technically challenging approach to the data and probably a more rigorous approach and so we’re able to **partly compare the results** that we get from these **different approaches** as well (..)”.

Collaboration – see also under ‘Data analysis’ (marked) and Dissemination (tools)

About challenges of communication leading to example of how it can be solved: “(..) communication is all about giving and taking, so listening to people and also expressing your own thoughts, and the challenge really comes when you’ve got people from different backgrounds, because the language they use and the jargon that’s used and that sort of thing often makes that communication really difficult, and so I think initially it is a bit time consuming and can be intimidating, particularly talking to people from other disciplines where you don’t really understand what they’re trying to

say; you think you understand perhaps, but you're often wrong and one thing I mentioned earlier, **we've been doing some work with some social scientists, various people from the different social sciences so sociology and psychology and that sort of thing and we actually ended up having an away couple of days; we all just sat in a room and we just talked about things** and actually you find that often you can talk about the same things using a completely different language and **once you start to learn each other's language the whole thing becomes a lot easier**, so I think there's a language barrier to interdisciplinary research (..)".

Communication in general and use of tools: "(..) in terms of modes of communication I guess, generally, I find the best thing is to sit down with a **wipe board** and **talk** [= **face-to-face is the most important mode**] and nut things out but we certainly did a lot **via the internet, email, telephone**, that sort of thing as well."

Distributed work in the project with interviewee 1: "(..) we were based in two sites [in the same city] during that project (..)", two main researchers at each; in such a case f2f meetings take place "at least weekly and in the crunch weeks, more often than that".

In general: "(..) we really do try and collaborate across institutions (..)"

Conferencing tools: "So some sort of conferencing software or something? Absolutely yeah." E.g. with "a PhD student based in Ethiopia who is supervised by two people here and somebody else up in Edinburgh, so we'll often do a **Skype conference call** for that and that works quite well. I recently did a PhD via **teleconference** which was fantastic; it worked brilliantly well, that was with somebody in Canada." The interviewee uses Skype very frequently, mainly for calls (without video).

On other **conferencing/eLearning/research event environments/applications:** "I think that sort of **technology actually is fantastic**, {{and I know this University has invested in some software for doing that sort of thing; I've not actually used it, but it's a Microsoft-y thing [based on Sharepoint], but}} I'm sitting in on an eCourse at the moment which I just log in to and you can watch people's slides and listen in and that sort of thing and join in (..)". {{"I can actually see a massive application it's the jump; it's that little bit of investment in time and effort that I need to become familiar with them and make sure that the other people I'm trying to communicate with are familiar with them as well that I haven't quite got through yet, but I suspect within the next year or two I will. I hope.}}}

Dissemination

Collaborating for dissemination/use of tools: "Yeah, I mean to be honest, for most of what we tend to do in **word**, just with **email**, where various people contribute different parts of it until you get something that looks like a whole and then different people will edit it and **track changes** and all that sort of thing. I mean I do, for some projects, use things like **Confluence** [a corporate wiki system: <http://www.atlassian.com/software/confluence/>] which you might know, and the University now has something (..) [called **VOCAL**, a collaboration platform: <http://www.liv.ac.uk/csd/vocal/index.htm>], so these are online collaboration tools like you say in terms of like a **Wiki** or something like that, and we have used them, for

one project we're working on we have used that, effectively, to try and write the **project report**, so there's various people working on the project and within this online environment we've set up effectively, the table of contents, what we think the report's going to look like and as people work on various bits they just slide it in there so you've always got the most up to date version on the link, and it works, to some extent, as long as people keep it up to date, so we do that sort of thing, but I think in that particular case we have several people each equally integral to writing the report, whereas often it's largely falling to one person to do there bulk of it with input from other people and in that case it's usually easier for them to keep it locally on their PC. I was going to **try and use Google Documents at one point** for this sort of thing but actually the limited size was far too small, so yeah, that's what we tend to do, and then we'll meet and discuss."

Important journals/publications:

Nature is the most important publication – but a publication together with interviewee 1 (on the Avian influenza/poultry project) “didn't get through; for some of these things it is a little bit challenging because although that project wasn't; the research question itself was quite specific, it's about the risks of Avian Influenza in Britain”. So in the end “finding a journal that necessarily wants it is not always easy, so that particular paper went to some people; the **proceedings of the Royal Society B**, which is the biological science”. “(.) [Interviewee 1] has presented it at some more **mathematical biology conferences** and we present some of the work at the more **veterinary conferences**, and it's been presented to people like **Defra** as well.” Also “we were asked to write a piece for some **poultry industry magazine**, which we did, just describing it”.

On publishing interdisciplinary: “There are some journals now that are much better at that, there's another **Royal Society journal** called **Interface** which does deal with that sort of interface between different disciplines and so on, so there are some good journals out there”

Other important elements about/in the research:

In general on using and adopting technology in research: “(.) I guess in terms of the technology side of things I'll come back to the issue with my research is that it tends to be very diverse and it's very methods diverse as well so while I had my skills, often projects are requiring other people with other skills and certainly work in that sort of area, so part of what I try to do is work out what methods are out there, what could I do, and that's where things like Condor is useful, I don't necessarily need to know how to use it but I know it exists and I know people can use it and often I see that's actually my role in a lot of the research that we do, is actually trying to identify those sorts of technologies and those sorts of things or methods or whatever, and then bring the people together who have those skills who have perhaps never met before and put a very applied question together that, between us, we can address, so I guess what I'm trying to say is that actually, a lot of these things, I have no idea how they work but that's probably not a big issue because there are people out there who do and it's knowing about it; what I'm trying to say is actually a little bit of knowledge in those circumstances can be very, very useful, you don't need all the technical detail.”