NTDS_010

Key:

I: Interviewer
R: Respondent

I: To start with, thanks again for your time. I would like to basically start with what is your background, a brief description of your research interests and also your research projects before your collaborations to sale and in sale. So then we will start then talking more about the projects with sale.

R: Well, I'm an associate professor in health services research. I've had quite a mixed career. My research interest is fairly generic, obviously focused on health services research. My original research when I did my PhD was in the Common Cold Centre in Cardiff. That was really, I suppose, clinical trials mainly, again, very much patient-based research. That was centred around looking at the mechanism behind cough and how we could override that mechanism in order to treat cough. I then moved out of that area slightly but still within health services research and I was involved in a project with patients in the health service who had a condition called obstructed sleep apnoea. The project there was based around looking at the ways we could monitor and diagnose those patients. So it was looking at different techniques for that. All of my research really has been based around patients largely.

For a long time I had a role as a statistician, again largely based on working within health services research. The people I worked with have interest in gastroenterology mainly. I guess from the early projects I worked on, my main interests are in patient reported outcome measures and the use of routine data. Before I even engaged with sale, I was actually working on a project where we tried to replicate four funded research trials that had purposefully collected research data. What we wanted to see with those four projects was whether we could answer the questions posed by the research project just by using data that were collected as part of routine practice. So it was before the days of sale, really. We could access things like... one trial we looked at, for example, had a clinical information system that was routinely used in a particular hospital.

So for that project we could actually get a lot of routine data, whereas some of the other projects didn't collect anything, they had some maybe patient administrative data as part of their routine care, but the availability of data for some things was quite limited. What I did with one of those projects was I actually managed to develop a patient reported outcome score using routine data that I compared with an existing patient reported outcome tool. So there was some sense that if you had a good set of data it may be possible to get some kind of patient reported outcome measure using routine data. So I suppose that was my first introduction to routine data. I've then worked on many projects. I've got a big role now as part of the Clinical Trials Unit as Swansea. So a lot of the projects I've worked on are trial-based projects, so large multi-centre projects that look at multiple outcomes using multiple methods.

So, often there's a qualitative stream, there's a clinical outcome, and usually my role within those projects has been leading the patient reported outcome measure strand. I suppose, as part of my role, over the years I've developed links with a number of paediatricians. I've got an interest in paediatric research. I've been involved in some projects that have looked at whether children who presented to hospital with burns had worse outcomes up to about age five than those who didn't. So I suppose, again, we were utilising routine paediatric records to look at those outcomes, and I suppose, as part of that project in particular, one thing we seem to see in the data was that the children, although there weren't huge differences in terms of the outcome of those children, what we did see was the children who presented to hospital with a burn seemed to be moving more frequently than those children who weren't.

So I suppose that's where the idea... I suppose my own main project within sale came from that project. What we did in developing that project was to look at whether children who moved frequently within their early years had worse outcomes in terms of their educational outcome and things like admission to hospital for various conditions. So that's where the idea came to develop this project, utilising sale data. So that's the biggest experience with sale, but I have been involved in other projects that have used sale. One of them quite some time ago looked at whether you could identify potential patients to recruit to trials using routine data. So that was another project I was involved with. I've also had a little bit of involvement, which is probably off the back of the house moves project, with the Welsh Electronic Child cohort.

I've also been involved with a project with an orthopaedic surgeon who wanted to see whether the mode of admission to hospital had any effect on their long-term outcome, so basically the way they were triaged in hospital and we utilised sale data to look at things like subsequent admissions to hospital, subsequent GP admissions and whether there was any difference in those outcomes between the different modes of triage.

I: This is another project?

R: Yes. I wasn't the PI on that but I was involved in some of the analysis of the data which came from sale.

I: One thing that I was curious about was the children that move project. I've seen there are a few different publications for different aspects. It seemed to offer you the opportunity to do research that is not strictly medical as far as I could tell with my non-institutional judgment.

R: Yes, and I guess we bolted that onto the health outcomes. It was largely because of, I suppose, the research team that was involved in that research. I was working with Ronan Lyons and Sarah Rogers who have more of an epidemiological slant and they're looking at things like injuries. Also the original research I was talking about on the burns patients I was working with some paediatricians and they're also interested in the educational aspects and the developmental aspects. So it was an opportunity to look at a bigger broader picture than just their health outcomes.

I: How did this opportunity to work with sale come about?

R: I think it came off the back of the project. I suppose I hadn't really considered using sale when I discussed it with the other members of the team like Ronan and Sarah. It was at the time when the Welsh Electronic Child Cohort project was just starting to develop. In discussions with Ronan, he highlighted that sale had the potential to look at outcomes like education and health outcomes. So it was really in the course of developing the project and the idea and working with other members of staff within the college that the idea came about. It was fortunate that it was at a time when the WECC research was going on as well. We managed to piggy-back on some of... in terms of funding it was quite a small pot of funding. I think had we not piggy-backed on the back of the WECC project, it would have been quite difficult to undertake.

What we managed to do was bolt it on to the WECC project which had a big team of statisticians and a lot of collaborators and we managed to get them to do a bit of work, fortunately. I suppose that's one of the things with the sale data, in theory it all sounds great and there's a wealth of data and a mass of data there, but I suppose the process of getting that data out, for somebody like me who doesn't really have any experience, requires quite a lot of resource and quite a lot of time.

I: So is it because it's quite technical and complex, or also because there's long procedures?

R: I suppose a bit of both. I think in terms of the process for me on that project, again I think we were quite fortunate because a lot of the governance processes had already been pushed through for the WECC project and we just kind of bolted what we wanted to do onto that project. So I didn't really experience any difficulties as far as the governance processes go. That went through fairly quickly. I guess it's fortunate that the statistical team that were working on the WECC project were doing all the data cleaning, were doing all the coding. If I'd have done that it would have been a huge learning curve for me because I don't query data in the way that they do. I don't know how to use R. I don't know how to create queries and all of those kinds of things. So if I'm involved in projects that use sale, it's usually where the sale team have created an analysis-ready dataset which I guess from their side requires a lot of resource.

But I think had we not had all those resources in place it would have been quite a lengthy process in terms of the learning curve for me. I've got a PhD student who's utilising sale. She's had to learn data linkage processes, developing queries in R. So it required quite a lot of understanding. I think if you're not querying and developing those queries on a day-to-day basis, it's quite a hard process. I think they're not insurmountable difficulties. I think it's more about ensuring there are resources and support in place in order for those things to happen. So for me if I was developing a research proposal that I wanted to use sale for, I would make sure that I costed in time for a sale technician and all of those things. So what I get out of it would be the analysis-ready dataset rather than me having to do all those steps together.

I: Is this very different from other work you've done with routine data?

R: Yes, because the project I was talking about was quite some time ago. So the routine data available were very limited. We're only talking about... we might have been replicating a trial where there were 250 patients. So the scale of the data was much smaller whereas for our house moves project we're talking about a quarter of a million records. So we've got a quarter of a million children over five years. So the scale of the data is enormous.

I: It also involves technical knowledge that is different.

R: Yes. With my clinical trials hat on, what we've done, I suppose, historically is we've had our outcomes, we've used a patient reported outcome measure or we've looked at side effects of treatment or surgical outcomes and all of those kinds of things. As a trials unit, particularly in Swansea, because we've got this resource available, what we're trying to do and what we're trying to develop on a long-term goal is to employ routine data outcomes into our trials. So a couple of trials I've been involved with, we've had a component of our main trial where we may recruit a couple of hundred or couple of thousand patients. What we've also bolted-on to that is a routine data aspect where we can look at a much larger cohort of anonymised data. So there's one project, for example, that I'm involved with called Prismatic, which is looking at a computerised system that is used in general practice that gives every patient a risk core.

The risk score is their risk of being admitted to hospital in the following year as an emergency. So what the trial is evaluating is whether the implementation of this system and how they manage patients using this system has any effect on their rate of emergency admission. So the main trial we did patient reported outcome measures, we did some qualitative research with clinicians and patients about how they felt they were being treated and so on, but what we're also doing is collecting all the data for all the patients in every practice that we've recruited and looking at the long-term outcomes, so the number of emergency admissions to hospital, the number of GP visits and so on. So there are, I guess, two separate cohorts of data but we're utilising the sale data to look at a much bigger sample over a longitudinal period.

Another project, a gastroenterology project I've been involved with again a trial, the main outcome measures were things like surgery, patient reported outcome measures and all those kind of things, qualitative research. What we're also hoping to do is longer-term linking the data we have on those patients into sale to look at their long-term outcomes. So as a trials unit I guess we're fortunate that we've got the sale data bank and we're trying to build up our portfolio of trials that are engaging with the e-health and the sale aspect.

I: So you're trying to converge as you've got these partnerships?

R: Yes. So just trying to make use of this resource in order to strengthen, in terms of big development and outcomes we can demonstrate that we can do more than just this cohort, we can look far beyond that.

I: What does it mean to try to converge and scale your... does it mean you need to hire different kinds of expertise, more people from an organisational point of view for you to involve routine outcomes?

R: I suppose historically it's all been... I suppose five years ago it was kind of done in-kind. So Ronan would say, "We can do this and we can do that," and that's when probably not a lot of research was going on. Given the scale of projects now and the scale of data, if we were to consider using sale data, we would need to make sure that we resource that into a bid. So we would put a proportion of a sale technician into our proposals. We'd make sure that we had co-applicant expertise from the sale team on our bids. So, for example, Prismatic, the project I was talking about in general practice, Martin Heaven is our sale liaison link. Rohan is working with our statistician to build up the data and link the data and so on. Traditionally, I don't think those resources were appropriately costed in. I was exactly the same, probably. You think about these ideas but you don't think about what goes on behind the scenes and all the underlying work that's necessary to build and get all that data in place.

So we're trying to engage much more with sale team, with David and with Ronan to make sure that we're appropriately costing those personnel in. As a trials unit... and I suppose the trials unit as it is now is currently being directed by Helen Snooks. She's only currently acting as the interim director because we're waiting for a new director to come in-post since the retirement of our old director. Helen also leads the Swansea arm of Prime Centre Wales which is primary and emergency care. Both of those units have recently been refunded by Health and Care Research Wales. As part of that core funding, we've got some core funding for a bit of a qualitative person, a statistician, an IT specialist and so on, to make sure everything is in-place for our core infrastructure. As part of that proposal we've actually costed in a day's worth of a sale person within that bid.

So we realised... there's potential there to use sale and it's more than just... that sale person is not on a specific project but is more about making sure we engage with the sale team, that we can appropriately cost up future bids and there's a link between us as a trials unit and the sale team. It's the same with Helen's team. They've actually costed in... I think they've managed to get a full-time person to be able to develop proposals that will use the sale data.

I: If you can open these a bit more, so I understand that you're talking about the idea of developing a more continuous relationship with the sale. So what you just said about not having collaboration on a specific project. It's not really just about the costing in but also a different way of costing, if I can analyse it in this way. So it's not just bounded by this project but it seems like in order to develop ways of working and also a legacy of learned...

R: Yes, and it's building up that experience. So rather than starting from scratch at every single project stage, thinking about, "What do we need? Who do we need to engage with?" We'll develop that sale expertise within our infrastructure.

I: So what is this sale... from your point of view, the sale expertise about? So what are you trying to accelerate?

R: In terms of our infrastructure it's about ensuring that we know all the appropriate processes. They'll build up experience in terms of their knowledge of the kind of data that exists within sale. So I suppose it's more of a generic overarching role rather than a technical data query kind of role. So it's about making sure we know who we go to, to think about getting the governance processes in place, somebody who can link between David's team and our team, and bring that sale general experience. So they can sit on the sale committees. They can represent the trials unit... sale have a monthly sale analyst meeting or forum. So they can gain general experience about what's currently going on, whether there are new datasets coming in, whether there are specific issues with data.

So they've got quite a good working knowledge of what is possible, I suppose, within sale, and what is feasible for us as a trials unit to add-on to bids in order to strengthen them. So it is very much quite a general role that sits between the trials unit and sale but has quite a broad knowledge of what sale can do.

I: Going back to working with the sale data, you said about how you work with Rohan and the others that work with the dataset to work out actually the dataset. So for you when you work with the data what kind of data is given to you? Is it an image of a file? Is it a dynamic querying system?

R: For me it's more like an analysis-ready dataset. So what I would see would be an SPSS dataset with specific outcomes that we'd specified as part of the project. So the simplest example was the project I worked on with the orthopaedic surgeon which was on open tibia fractures. So, basically, what we asked the sale team, or what the surgeon asked the sale team, was, "We need to identify patients who are triaged by this method and triaged by this method and what we want to do is look at the number of GP visits they had within this period and the number of A&E visits they had in this period and the number of in-patient visits." So what we saw then was specific outcomes for each of the, say, 100 patients that sale identified within that dataset.

I: What kind of specificities or peculiarities do the sale data have for you in respect to other routine data?

R: I suppose there's always the issue of non-matching. So if we're linking questionnaire data with sale data there's sometimes a proportion of patients we can't match. There are potentially different issues with different projects, I think. For the house moves projects, for example, we used a change of (ph 0.28.36) *Ralf* as an indication of a house move. Now, that relies on a parent actually registering that change of address with a healthcare provider. It's unlikely. Particularly we were looking at the first year of life. Potentially there's an issue of under-reporting. So parents who repeatedly move, if they move even to the house next door, they're unlikely to register every time they move. So there are issues with under-reporting. I suppose, with things like GP visits, some of the studies we've been involved with you can't distinguish between a telephone call and a prescription and whether that's actually a visit to the GP or a visit to the nurse. So some of the fine detail I think may not be there.

I think, again going back to the house moves project, it's sometimes about... there needs to be more definition of what your outcomes are. So we had to define what kind of conditions we're looking at, what kind of codes that we want to search for. So if we're looking at respiratory conditions we had to think about... there's a huge raft of respiratory conditions. I think it's imposing the restrictions on the data. It was things like do we look at searching for a particular respiratory condition across all twelve diagnostic codes, or really do we need to focus on the first and second positions, for example. So I guess there was probably more clarity needed in terms of what we wanted out of the data. Whereas, I suppose when you do design a trial and you purposefully collect data, you know before...

I: What you want to generate.

R: Yes.

I: So because of the issues like under-reporting, the change of Ralf, and stuff, I guess those are issues more like that anyway are also present in other kinds of routine...

R: Yes.

I: It's more general to the methodology, whereas this one is maybe more related to just the (unclear 0.31.12) of sale, or the perceived ill definition of what is possible just because it's very big.

R: Yes. It is. And I suppose what you don't get with sale data is a qualitative understanding of the reasons. So I'm in the middle of addressing reviewers' comments. We're writing a paper on the health outcomes related with frequent moves. Although we can say that, "This proportion of people moved and that might be a bad thing," what we don't know is why these individuals moved. Was it a move to a better location? We can get some understanding of whether it was to a better or worse location but we don't know whether people moved because of a job, because of a split in the family, whereas I suppose when you design a trial, certainly with the trials we do, we employ often a qualitative component.

So we can get quantitative outcomes. We can say, "This happened. This many people had surgery. This many people had side effects. This many people stopped the treatment." What the qualitative data will do is give us some understanding of why they stopped taking the treatment, whereas you don't get that qualitative detail with routine data.

I: How would you describe the process where you were saying it was a bit more difficult to define what we were looking for because... so how do you describe the process? Who are the people that solved the problem and what kind of communication...?

R: I think we were quite fortunate. We had a really good diverse research team. So as part of the research team we had people like Ronan and Sarah who were the sale senior researchers. We also had more technical people who were used to querying the data and generating the data. We had statisticians. What we also had, which I think was really important for the house moves project, was we had three paediatricians. So they had an in-depth understanding of what would be important and what the usual scenarios were in clinical practice. So when we were considering, for example, whether we would search all the diagnostic codes to see whether somebody had a respiratory condition, they would say, "Well, as a doctor in a clinic, I would only ever... if it was important to write it in the first or second position."

So we made decisions based on engaging with the clinical people and I think that's quite important. To me I think there's a huge amount of data in sale with incredible power but it's not losing sight of the clinical aspect and the clinical importance of the data and the on-the-ground input as to what the data should look like and what is important in the data and I suppose all those kind of elements. So we had quite a lot of engagement with them and I think they played a role in deciding. At the end of the day it was just defining what we wanted to do, but it was just getting some clinical justification for our decisions, really, that was based on real practice and the way children, patients are seen.

I: Another thing that I was curious about, there is now all this stuff of linking various datasets, or inferring. It's grounded on the development of algorithms to sort and sift through the data, and then the validation of these algorithms. So what is your insight into this? Is that different from other kinds of research in these kinds of topics?

R: Yes. To me, I wasn't really involved in that particular aspect of it, but I suppose it's all about validation of that data and making sense and justifying the outcomes. I don't think it's largely different from a lot of what goes on anyway, so things like the project I mentioned, Prismatic, is a predictive risk tool. So that uses data to generate a risk score, which is not dissimilar to what sale can do, but I think it's the testing and the validation and that has to be founded on something. I think these risk scores are often not adopted and utilised until they have been tested. A lot of the data I was using wouldn't be based on development of algorithms. It would have just been counts and frequencies. I guess it's just having valid reasons for the assumptions. I guess as long as you can define and have a justified reason and rationale for why you've done something.

I: I was thinking about the study where you were exploring the potential for sale to select a cohort of people, the pragmatic control trial enrolment problem. The way I was reading it, I perceived the problem of the system is viable because it's very secure and very anonymised. Then actually then it becomes... is there any solution, any discussion about this?

R: I'm not aware. It is a huge issue. That is a huge potential issue that you can't identify. It has the potential to identify individuals that can be recruited, but then you can't actually go back and approach those people because you can't identify them. I know some people have gained consent to go back and identify individuals. So I suppose there's potential there as long as there are appropriate governance processes in-place and there's appropriate consent from the patients and so on. I think it's testing that process and ensuring that it is viable, it is robust, it doesn't breach security issues. It's potentially developing those processes whereby there is the potential to identify but it has been done appropriately. So whether that can be done within a study that has been fully consented and has been designed to test that process I don't know, but I know there are huge potential issues with that. So I probably don't know enough about that to...

I: In respect to other... can I ask you a last question?

R: Yes.

I: There were comments on this in the papers about anxiety in children and young people. It would be interesting to understand in respect with other routine data that you have been working with, sale data obviously links together lots of data that has been recorded or generated for various administrative purposes and various different kinds of people with different kinds of training. I copied this quote, you were saying, 'These results may reflect the genuine increase in anxiety in the young

population or increased awareness on the part of GPs.' It was the anxiety in children and young people, the incidence of anxiety in treatment prescription.

R: This is the prescription study. Okay.

I: I wanted to ask you about your reflections and is there a limit that can be... that you've been considering, or that you have an idea how it could be overcome by training people or something?

R: The mental health thing Anne led on. So she's probably more... she would be a better person to ask that question of. There are training issues. To be honest I haven't really thought any more on that one.

I: I can imagine in the long-run about this system, it will keep being developed and becoming more and more sophisticated as in general we hope for a closing in of the loop between research and the clinical. My assumption is that... the rationale behind my question was there will be a point where there needs also to be standards that can start to be shaping GP practice, GP data collection practices.

R: There is still a lot of inconsistency in the way data is collected. It would make things simpler if everybody used the same reporting/recording tools. I suppose it's the training in terms of how people record the data. If you've got one researcher working on a research project, there is limited variability in the data, whereas if you've got thousands of people recording there's potentially a lot of variability.

I: Are there initiatives to the study related to sale but more generally?

R: I think there are. There are moves towards standardised data collection, certainly within some areas they're looking at single patient records and standardising what is collected for patients and so on. I don't know enough detail about what the initiatives are but certain fields like gastroenterology have initiatives to try and collect things like patient reported outcome measures routinely. I think there are plenty of initiatives there but I think it's crossing all the disciplines and getting minimum standards across all of those.

I: The data that you work on from the sale. Have you considered sharing it and is it possible in terms of licensing and things like that?

R: Not as far as I know. From the studies I've been involved with, there are very stringent restrictions in terms of what you can do with the data and what you can report, if there are small numbers when we're doing some of the analysis they have to be checked by the analysis team prior to releasing the outputs.

I: So the results of your work in terms of coding data and interpreting data, so the dataset doesn't stay with you?

R: No. That's retained within sale. There are restrictions in terms of how long you can access that for. You can obviously extend that, but certainly from my use of the data, which was more in the final stages in terms of the analysis and the outputs, they would check that there was no possibility of identification, if we had less than five cases we weren't able to report the data for risk of potential identification. So, yes, I suppose if you knew more about the data there could be the potential to access share, but no, I don't have enough knowledge about using the sale data, but as far as the work I've had with sale I haven't done that and wouldn't do that. I think there are measures...

I: It's not necessary in particular also. Okay. Thanks a lot.

(End of recording)