Linking Our Lives - Champions
Series 3 Episode 4

SUMMARY KEYWORDS
ONS longitudinal study, Scottish longitudinal study, deprivation, health inequalities, employment, migration, data access, synthetic data, air quality, pregnancy exposure, health outcomes, census data, secure labs, research support, longitudinal studies.

SPEAKERS
Paul Norman, Christine Garrington, Tom Clemens

Christine Garrington 00:01
Welcome to Linking Our Lives a podcast about the ONS longitudinal Study. I'm Chris Garrington, and in this series we're telling you more about this unique data resource and its potential to forward our understanding of the changes to our society since 1971. I'm in conversation with researchers who've been and continue to be champions of this special Census based data that has so many applications and which they believe could and should be used more widely. Today, we're talking about not just the ONS Longitudinal Study, but its Scottish counterpart, the Scottish Longitudinal Study. I'm delighted to be joined by Paul Norman from the University of Leeds and Tom Clemens from the University of Edinburgh. Both are geographers who've used these studies to look at the links between deprivation, health, employment and the places where people live and work. I'd like to start by asking you both, if I may, why these studies provide such great data for research looking at these important matters. Tom, if I could start with you.

Tom Clemens 00:58
One of the things about the Longitudinal Studies that is so important is that they cover a lot of individuals. There's a lot of people included in these studies, the sample sizes are large. For a lot of the questions that I was interested in answering, some of the effects that we interested in are quite small. They're not that easy to detect, so you need larger sample sizes in order to study them. And the nature of the Longitudinal Studies in Scotland, we're a 5% sample in Scotland, which equates to roughly 350,000 people, which is a great resource, it's really large, and certainly compared to other potential data sources I might have considered, things like some survey data sources, much, much, much bigger in size, which allowed me to answer these questions with enough kind of statistical power and enough statistical certainty to be able to answer those questions.

Christine Garrington 01:38
Paul, your earliest experience of the LS was simply to use the area data in your research, but you were one of the first people really to see its bigger potential. So, what was that potential for you?

Paul Norman 01:48
Going back to the specific research question, which was about the relationship between subnational migration, health and deprivation, so people moving between different areas. The work that I've done with the area data, etc, simply used the year before the Census, so that's in answer to the question, you know, where were you living one year ago? Same address, different address, where was it, if different. So, it's another version of what Tom's just said that actually the time increment of one year ago in terms of people's migration or non-migration, was actually insufficient to capture the effect that it might have a on the individual and be on the areas people are moving to

and from. So it kind of needed that scaling up of the time frame of 10 and 20 years ago, that sort of decennial Census time increment. It needed that in order to detect the effect, you know if 10% of people move every year, which is approximately what it was, if not still is, you kind of need to scale that up by a longer time increment in order to detect the effect. That's another version of what Tom's been saying about you need sufficient numbers in the example I was working in, you needed sufficient time to find what had happened.

Christine Garrington 03:16
And presumably that's unique to the LS's compared with other studies?

Paul Norman 03:22
Not necessarily The Birth Cohort Studies, British Household Panel Survey, not necessarily called that anymore. You can do similar kinds of things, but actually the LS's across the UK are much bigger sample sizes, so whilst you can look at alternative sources, they're actually less likely in the applications we're dealing with to give you plausible results. In some of these cases, I've used some of those other sources and they are in themselves useful, but the sample size of the LS kind of trumps them.

Christine Garrington 03:57
Yeah. And understood. Tom, I'm wondering, looking back, was there someone who actively encouraged you to use the LS, someone who championed the data to you because they could see its potential to work well for you in your research, whether that was around your undergraduate studies or your PhD studies or or later?

Tom Clemens 04:18
It's gonna be a bit cliched, I guess, to say this, but my PhD supervisor was probably the biggest inspiration for this. As I said before, he was involved in the original setting up of the Scottish Longitudinal Study, and I remember as an undergraduate becoming very interested in health inequalities research, trying to understand the relationships between our social lives and our health. And he'd been fairly instrumental in publishing lots of work, particularly from a kind of geographical perspective, understanding spatial and geographical inequalities in health. So he'd written lots of stuff back in the day around this and that was what got me interested in this area as an undergraduate students. I'm very lucky to get to work with him later on. So yeah, Professor Paul Boyle, he's not necessarily working so much in this area anymore, but he was kind of instrumental throughout my PhD, encouraging me to use the data, helping me out with how to extract the most from the data that was available, and then also kind of inspiring me to continue that kind of work later on into research posts and later on in my academic career, as I moved to University of Edinburgh.

Christine Garrington 05:11
Fantastic. And what about for you, Paul?

Paul Norman 05:13
Well, I'm afraid to say it was the same person. So, I was a mature student at the University of Leeds, doing a Masters in GIS. And Paul Boyle was at that stage at the University of Leeds, so he talked me into doing a PhD, and then he immediately left having done that, and went to St Andrews. But he did continue to be a supervisor, although I stayed in Leeds, also supervised by Phil Rees, but Paul was instrumental in Amy doing a PhD, and B me working on the LS, and then, as Tom said, he went on to be a leading light in setting up the Scottish version. And that was kind of some of that thinking was going on while I was doing my PhD, and the SLS followed soon after. So we owe the same person quite an academic and actually personal debt, because Paul was a very good friend and good socially at the same time as well. A small world.

Christine Garrington 06:13

It is that's wonderful to hear. Thank you for sharing that, and really nice to give Paul a proper and well-deserved name check by the sounds of things here. So, thank you very much for sharing that. I wonder if we might turn our attention now to actually working with the data, we've heard across earlier episodes about experiences of going into secure labs. But I wonder if there's something that you both might like to share with you know what it's been like, because it's changed a lot, right? And it's changing all of the time. So, let's go to you. Tom,

Tom Clemens 06:42
Yeah, I mean, it has changed a lot. It's changed for the better, it is still definitely challenging as a data set to access and work with, compared to particularly publicly available data sets, or, for example, data that you may have collected yourself. It definitely comes with a unique set of challenges. But I happen to think that those challenges are positive in many ways, certainly working in environments like Safe Data Havens, and working in environments where you are constricted in what you can do definitely makes you a better researcher. I can definitely credit having to be much more prepared for going into the safe haven, having a much clearer idea before I go in what sort of analysis I want to do, making sure I've got clear in my head what sorts of questions I'm going to be asking the data at that particular time. All of those things have been very good training that other researchers maybe don't need to do, because that, you know, their access to the data is much more straightforward. Certainly, things are changing. At the moment in Scotland, we've got a number of exciting things around data access, which in the future, might help them to make that process a bit easier. Not necessarily in the assets, but in other kind of administered data sources, like the Census data, there is the option for remote access, which might change things and make things a lot easier. But certainly at the moment, the day to day practicalities are going into, in Scotland, we have a very lovely building, we're based in the national records of Scotland. We have to go to a physical premises. It's very grand building, so can take some inspiration from the grand surroundings that you're in the Scottish Longitudinal Study, so that's great. In general, it's preparedness. It's being kind of much more diligent with your planning. I mean, these days, I don't just work exclusively with The Longitudinal Studies. I do have access to other data sets, and so it's nice to be able to work on those easy to access data sets. But when you work on the SLS, you just kind of get into that a bit of a groove where you put a lot of work in preparation before you arrive. And in some ways as well, it's nice to be in a room where you're completely separate from the internet. There's no emails going off, there's no phones going off. You're in a much kind of easier environment to concentrate. And so, I think it's much more easy to be productive in an environment like that. Yeah, I think we often associate working with this kind of data as being challenging and difficult, but that there are a huge number of policies that come with it.

Christine Garrington 08:39
Yeah, no, absolutely. I'll ask Paul the same question in a moment. But off the back of that, I wonder if you have any good tips. I mean, you've already said a few, really, about being organised and meticulous in your approach and preparedness. But I wonder if you have any other tips for somebody thinking about using the data for the first time. Particularly interested to know if you were supported by anybody as well, or if there's help out there that people can get to overcome some of those challenges, as well as sort of taking responsibility yourself, if you like. Paul, I'm sure there's a lot you'd agree with there. But let's ask you first about your experiences of working with the data over so many years. Is there things that stand out? Is there things that you'd like to share?

Tom Clemens 09:05
I think one of the things I wish I'd been told when I first started was don't be afraid to ask questions at every opportunity. That answer probably goes for lots of lots of questions, I think. But both the LS and the Scottish Longitudinal Studies have teams of staff who are willing to help, and certainly in the SMS, we've got some very friendly, knowledgeable staff who are ex researchers, in some cases, and are very familiar with some of the challenges. Researchers can be a bit reluctant sometimes to ask for help or kind of check certain things. And I would just urge particularly new researchers, don't be afraid to ask for help. I think the SS and ONS LS are the same. We're very understanding of the challenges, particularly for junior researchers that you're facing with doing

analysis in these kinds of environments. Alongside that, reiterating what I said before, really, which is being as well prepared and diligent in your planning before you go in and making best use of the time that you have available, knowing that if you can't finish by five o'clock, it will be another few days before you can get back in there. So planning your time. And being kind of realistic with what you can get done in the time there is probably the most important thing, I would say.

Paul Norman 09:53
Yes, one thing that might rather surprise people listening to this who have either used the LS more recently or are thinking of using it, is that I actually worked on a data download, so I was actually supplied with the data set. So, this was around the year 2000 I guess, emailed a data set, as it were, which was an extract from the England and Wales LS, and it was disclosure controlled for confidentiality checks before I was sent it. But there were no unique attribute records. So, there was a count of people who had similar attributes or the same attributes in terms of the categorical variables to other people. So, all my early work around the start to the 21st century, and therefore up to 1991 in the LS timeframe, was actually a data set that I could use on my own computer. What that enabled me to do was the exact opposite of what Tom's been talking about, which was to bumble along and try things and fumble, and so it was much less focused. Now, there are pros and cons to that, all the advantages and so on that Tom's described, I totally go with. Yes, by being time restricted and location constrained, in some ways, does make you more on task when you were there. However, if you're learning new techniques, which you probably will be doing, maybe somebody doing a PhD, for example, then the longitudinal element means that there's a steep ish learning curve. So, something that is has changed in the meantime, which I think is very useful, is the creation of synthetic data sets, which emulate the variables you have in your file. And that is something that you can have more ready access to develop your code syntax, whatever, what you are going to do. So, if you have a synthetic version of the same data, you can then make sure that what you take with you or send in in advance for your assistance to put in your workspace. Then that does help you make your time actually in the safe pod much more productive, because you know the syntax, etc will run.

Christine Garrington 12:37
And what about the support that's available? Obviously, our podcast is produced by the Celsius team. Lots of expertise and support available there. I think it's really important, and I'm sure you'd agree that academics who were looking to use it for the first time. I mean, they have to come through that sort of process anyway, but to know that there is a lot of expertise there is key, right,

Tom Clemens 12:56
Yeah, and to have a good relationship with your support person is essential. So again, because I've been doing this for a while now, it was the Institute of Education who were the academic support, and some truly excellent people, legends in their time, were there. So that was brilliant when I was starting off. And the same kinds of things were happening then has happened now with day meetings and people presenting their work, and support officers saying what they can do, and that kind of thing. One thing that is really, really nice is when you see the Support Officer having contributed quite a lot to somebody's work, is they end up being named as a co-author on the paper, or something like that. And I think that is illustrative of how much help somebody can be at Celsius, etc, to aid the research process.

Christine Garrington 13:49
Tom, I wonder if you would say, you know, whether working with the SLS has in some way sort of shaped the direction of your research and your career. And if so, how would it have been different in some way without the SLS?

Tom Clemens 14:20

This is a really interesting question for me, because I think my PhD started with accessing this COVID Longitudinal Study. And I'll be honest, I've never really looked back, and I've never really looked at any other forms of data so much. There's a few small pieces of work where I've looked at some survey data and other sources, but really, my entire career has been using data from the Scottish Longitudinal Study, but also other kind of sources of initiative data that we can link to the SLS. I think I can quite honestly say I probably wouldn't be an academic if it wasn't for the Scottish Longitudinal Study, which I always find interesting. So, I think it's testament to really how much can be done with the SLS. I mean, I think my research is only a small part of all of the projects that use the SLS, but it's been enough to easily sustain my career up to this point. You know, my career has moved away now from unemployment and health related research into kind of more environmental health type research. So I've become much more interested in the last few years in various ways in which the environment, particularly things like air quality, poor air quality, can impact health in various ways. As my interests have changed, the data and methods that I use has stayed the same, largely because the SLS is such a flexible resource that can be, and the ONS LS is such a flexible resource that can be used to answer a whole wide range of questions, really.

Christine Garrington 15:41
And Paul, what about for you?

Paul Norman 15:42
Well, what one aspect is my most cited paper, according to Google Scholar, is one using the LS, so one has to say, therefore, that it's an important paper and data source, if only from that perspective. I guess two main strands to my work are how places have changed over time, along with people moving between those places. So one strand has driven the other in that an awareness of the LS and how you are capturing people's moves between different areas or staying in the same area, and in attaching deprivation scores quintiles to the LS member record, so you can see whether people are moving between more or less deprived areas, and their health outcome as a result, has kind of driven me to create comparable deprivation measures over time. And that's very much an area-based piece of work. And I'm therefore thinking about how one develops one in order to inform the other. Whilst there is a lot of my work which is area based, that work is then of the output to then attach to the LS, to use directly with the LS. But it's also the case that there's other work that I do whereby people who are interested in, for example, cancer outcomes, and they have people's diagnosis and their registration, their survival, and again, they want to attach those records to area deprivation or other area characteristics, measurements, urban, rural, etc. You know, as I say, the things are complementary and working with one style of data and time increments and so on, makes you aware of what may or may not be possible with another style of data. So, you're always thinking through the implications of how you've done things and why you've done it like that, and what are the pros and cons, because what you find when you do an analysis will be affected by some of those operational decisions. I wouldn't not have a career without the LS, but it certainly wouldn't be so fruitful and so interesting. It's probably a quarter of what I do has been with the LS. Put it that way.

Christine Garrington 15:52
That's a substantial amount of work. So then Paul, are there one or two key findings from the work that you've outlined there this amazing array of work using the ONS LS that you think in some way have made a difference, or are simply something new that we just didn't know before. What would those things be?

Paul Norman 18:28
So there was an early ish. The thoughts of health selective migration come from the 19th century, and the work of the registrar generals then so on. So, it's not new, but what you could find out has definitely increased and improved through the use of the LSS. So, there is a general movement of healthy people from more deprived areas to less deprived areas. And as that happens, health inequalities either develop or increase, or at least maintained. So, we find in aggregate that less deprived areas have better health, however measured, than more

deprived areas. But part of that is that people move between areas and healthier people tend to move from more to less deprived areas. There is a movement the other way that less healthy people tend to end up in more deprived areas, but the numbers are fewer, and there is quite an element of non-migration. People are less likely to move if they're in more deprived areas and in poor health. So, the sort of health inequalities work is better informed by that. But then there's always extra layers, so that that's a fair summary of the situation, but, but you find people at different ages are doing different things, that people are moving in the sort of preschool era differently to how they move during school ages. Then there are student type moves. Yes, and the description that I gave a minute ago is actually more telling for people in midlife. So, when careers develop or don't develop, as the case may be, and then people around retirement and older move in different directions for different reasons, to do with their health as good health enabled retirement migration, there's perhaps poor health required for people moving to formally and informal care a little bit later in life. So, the LS is the best data source in England and Wales to be able to tell what is going on with that.

Christine Garrington 20:37
And Tom, you alluded to some of your more recent work that you've been doing around the environment, and I know you've been doing some work looking at the links between where people can buy tobacco and crime, some really, really interesting work. But I wonder again, are there one or two key findings from the work that you've done that you think in some way have made a difference, or have told us something new that we didn't know before.

Tom Clemens 21:01
I think some of some of the more recent work around air quality exposure during pregnancy are probably the best here to showcase what the SLS and ONS LS studies in general can do. So, this was a study a few years ago now, but we're interested in trying to understand a bit better what impacts living in kind of, or being exposed to high levels of background poor air quality might have during pregnancy. And at the time, much of what we knew about this was based on studies in areas where air quality was notoriously bad. There's been quite a few studies in places like China and Beijing, in particular studies in India that showed there were quite dramatic impacts of poor air quality, but at kind of levels of poor air quality that were really, really high. And so one of the things that we're interested in as well, we know that that exposure levels at that sort of magnitude may be really harmful for things like the risk of having a low birth weight baby, or the excess risks of having a pre term baby. But what we didn't really know so much was, well, are those effects still kind of observable, or can we still detect them in areas where you know, or in context where air quality perhaps isn't quite as bad. In other words, we're really interested in this question of whether there's a kind of lower threshold, you know, a safe, almost like a safe level of exposure. One of the challenges of this is, but when we're looking at these kinds of questions, sometimes the statistical effect size of interest in are quite small, and so we what we needed was a big data set with lots of people in it. We needed to be able to link those people to their maternity records, their hospital records, the records that are generated when they give birth. And then, of course, we needed to be able to connect those records to some measures of air quality. And the only really data source that was allowed us to do that was in this case, it's got this longitudinal study. And of course, in Scotland, we're fortunate in that air quality is, on average, much better than it is in some of the other countries where this these kinds of studies have been done. So, we were able to kind of construct this really neat data set where we had information about where people were living at the time that they gave birth, as well as some modelled air pollution data allow us to kind of estimate their exposure to poor air quality. We didn't, indeed find some kind of interesting results of a particularly particulate matter pollution, which is a very an area of kind of great concern at the minute, but particularly particle size is less than two and a half microns. That kind of get your really small, ultra fine particles that can pass through the respiratory system into the blood, even at the relatively low exposures that you see in Scotland. We did detect excess risk of a preterm delivery, as well as associated risks with fetal development and low birth weight. So that was really important study, just to kind of really emphasise that there really isn't a safe exposure level for poor air quality when it comes to pregnancy. And so, I think on that sense, it was really important study. The other thing

that that study did that was kind of unique to what the SLS can bring was that we were able to look at exposures, not just at home, but also where people were working. Because obviously The Longitudinal Studies have access to workplace location, so we were able to bring together those two sorts of exposure in ways that hadn't been done before and in ways that were unique to The Longitudinal Studies. Thanks for sharing those really important findings from the LS research and the SLS research that you've done to date. I wonder if either or both of you have got plans. I think Paul, you've already alluded to the fact that you have got plans to use the soon to be released Census, 2021, data, when that becomes part of the studies. Paul, any plans? And what can you tell us?

Paul Norman 24:08
It's what people might say, turning the handle, doing the same thing again. So, what's actually going on at the moment is a set of projects which are known as LS beta tests. I presume the Alpha tests are carried out behind very closed doors at ONS to check the data. So, the beta tests are by some academic projects. People were asked to apply to do this. And in fact, I'm doing kind of the same thing again, that I did before. There are various reasons to do that. And one is that if I attach area deprivation to the LS members and also some coordinates to allow distance to move to be measured, and rerun the type of thing that I did before, I would expect very similar results to emerge. So, whilst that may not sound like a great piece of research in terms of its findings, one of the purposes of these beta test projects is to check that the LS is well configured and fit for the purpose that we want it to be. So, this is all part of the effectively, the support via both Celsius and ONS, because everybody wants to make sure that data are fit for the purpose that we would want it to be. So, I wouldn't expect any startlingly different results to the type of thing that I've done before, which would reassure in terms of data quality, as it were. However, fingers crossed, it still will generate at least one academic paper to know how health inequalities are being maintained, exaggerated, or perhaps, you know, fingers crossed, being alleviated somewhat, is still a useful exercise in itself.

Christine Garrington 25:52
And Tom, what about for you?

Tom Clemens 25:54
There's lots of exciting things going on in Scotland at the moment in relation to the Scottish Longitudinal Study. So, we have, I think probably the most exciting thing is the 2022 Census data. We don't currently have access to that, but this is obviously the next Census data we're looking to incorporate into our data set. So that would give us Census records from 1991, 2001, 2011, plus 2022, another 10 year's worth of information. So that's pretty exciting. There's lots of project's kind of in the pipeline, looking to kind of extend previous work, look at how things have changed since the last Census and so on. And I think one of the really exciting things that the SS has actually had, it for a while now, but I think there's, there's lots of exciting projects to be done with this kind of data, and that is the fact that we can now observe where people have moved in between Censuses. So that's a really potentially important, exciting development, because it allows us to look at changes in people's environments, quite small timescales and in between Census records. That's particularly interesting for someone like me is interested in various environmental exposures, because it allows us to look at over an individual's now kind of probably 20 year lives, 20-year period in their lives. We can look at changes in things like air quality exposure. We can look at changes in another area of interest of mine, which is availability of certain unhealthy products like alcohol and tobacco products, and how accessibility to retailing premises that sell those products can influence people's health behaviour choices. And so with this new kind of residential movement data, we can start to see, well, what happens if someone moves from a high exposure area and whatever exposure interest into a low exposure area. And we can start to look at changes in health that result from this change, which is a really powerful kind of analytical design, a pretty big step forward, I think, in what sorts of things we can do with SLS data that we have. Paul Norman and Tom Clemens, thank you so much for being champions of the ONS Longitudinal Study and for your contribution to research in so many areas, actually. And for our listeners, you can find an accompanying Linking Our Lives blog for this episode on the Celsius website. This podcast is produced by

the team at Celsius at UCL, with support from Research Podcasts. You can follow us on X at @CelsiusNews, and you can also find us on LinkedIn. Thank you very much for listening.