

Linking Out Lives - Champions

Series 3 Episode 2

SUMMARY KEYWORDS

migrant health, ONS longitudinal study, health differences, data challenges, census linkage, socio-economic factors, mortality rates, cancer study, data support, safe pods, user-written code, aging migrants, intergenerational health, research opportunities, data accessibility

SPEAKERS

Christine Garrington, Joe Harrison, Matt Wallace

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, in many cases, 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, I'm really delighted to be joined by Matt Wallace and Joe Harrison, who've been using the LS to look at the health of immigrants. Indeed, both have used the data for their respective PhD research. Matt is a reader in social inequality at the University of Salford, and Joe is a research fellow at the University of St Andrews. Matt's interested in the health differences between international migrants, the children of migrants and non migrants, and how these differences affect wider population health. Whilst Joe's current PhD research aims to increase understanding of the different life courses experienced by the Pakistani community and their descendants in the United Kingdom and Norway. So Matt, if I may, I'm going to start with you. What first drew you to the ONS Longitudinal Study, and what was it that you wanted to use it to look at?

Matt Wallace 01:18

Migrant health is incredibly important. We have increasing kind of absolute and relative shares of migrants in many countries in the world, including in the UK, and it's become increasingly important to look at how their health differs from non migrants, or between different groups of migrants, or compared to the children of migrants. But the truth is, there aren't many data sets in the world that are well equipped to study the health of international migrants, and that's because they place a very heavy burden on the data. We need a health measure that we're interested in, we need to be able to identify migrants and non migrants using information like country of birth, nationality, citizenship, ethnicity. And we need other demographic and socio-economic characteristics that might affect the link between migration and health, such as education, labour, housing, relationships. Very few data sets, at least, in my opinion, are able to meet that heavy burden, but the LS meets that burden and then some, which is what drew me to using the LS for my PhD in the first place.

Christine Garrington 02:17

Oh, fantastic. We'll dig a bit more into that in a moment. But Joe, I wonder if I can ask you the same question about what drew you to the LS, and how you're using it perhaps in your PhD?

Joe Harrison 02:18

Yeah, of course. So similarly, I also was kind of driven by this is the only option for migrant mortality studies and migrant health studies in terms of sample size, it is really the only individual level data in the UK which would

allow that. So you could maybe do death rates with other data sources, but to get into life course elements, so socio-economic success, marital status and these sort of factors, the Census linkage is probably the only thing you can really use it to. For my research as well, I've been comparing to Norway, and in Norway, I've been very lucky to be able to access a complete population register, and this is the only thing that is comparable in the UK to that sort of statistical power and detail. It's not exactly the same, but it's in terms of driving my research, as comparing the two contacts. I need data that is reasonably comparable. And that's what drove me to the LS, really.

Christine Garrington 03:26

So you both touched on this a little bit, but I wonder if I can drill a bit deeper and ask why the LS is such a good data set for these types of questions about the health of people who've come to live in England and Wales from other countries. Matt, do you want to take that first?

Matt Wallace 03:41

So the LS is a veritable treasure trove for the kind of information needed to study migrant health. It started from 1971 which is when large scale migration had really taken off in the UK. So we're able to track people from 1971 for over 50 years now. We can track new arrivals between 1971 and the most recent year in the LS. There are various different pieces of health-related information in the ONS, longitudinal study on mortality, mortality from all causes, cause specific mortality from cancers, cardiovascular diseases, communicable non communicable diseases. People self report their own health at the Census, their own general health, whether they have a limiting, long term illness. And we can use all of this information to look at health differences between different people, particularly migrants and non migrants, because we also, in the LS have information on country of birth and ethnicity, so we can define both migrants and their children. So ultimately, we can contrast the health of different migrant groups. We can contrast the health of migrants and their children, combining ethnicity and country of birth. We can study changes in migrant health over time by looking along the LS, or we can study migrant health at snapshots in time just by using the Census data that's there. So there's so many things that we can do with it. It's so flexible and rich with all of this data.

Christine Garrington 05:04

So Joe, a veritable treasure trove. Are there any particular jewels in it that sparkle shine for you?

Joe Harrison 05:13

Yeah, I think the consistency and the accuracy it has with them, well, I don't, I can't test the accuracy, but the ICD codes and being able to identify cause specific that was been very useful for me in I did a cancer study, and so having this knowledge of the type of cancer, if the mortality, subsequently is from cancer, these sort of informations is, yeah, it's been invaluable to making a good PhD project, which obviously, with my intention. Mortality and health as a process is longitudinal, so having this longitudinal study is necessary really. It's sort of impossible to get from sort of a cross-sectional data set if anyone's ever tried to study long term health using something like Understanding Society, which is a great data source for certain things, but if you try to use it for mortality or aging help, it's really not got the numbers to be able to do anything of long term nature.

Christine Garrington 06:07

i'm conscious about this whole idea of champions. And I was, I was wondering, Joe, if there was anyone specifically who sort of actively encouraged you and therefore championed the data set to you for use in your PhD research. And I know you were both supervised at different times by Hill Kulu himself, an incredible supporter of the LS, but is it Hill? Is there anybody else, Joe?

Joe Harrison 06:29

Well, I think mine works in tandem from Hill. I feel like Matt actually gave a lot of knowledge to Hill from studying it. And then that's come back to me. So a lot of my frameworks and statistical methods are built off Matt's earlier

papers using the LS. So very grateful to be here with Matt today. So they would be like, they're probably the central papers that I've built from. So yeah, I would say Matt and Hill and the work that they did five or six years before I started, yeah, has been part of it. And a big shout out to the Celsius team as well, they've done a lot to help come up with ideas and make sure ideas are feasible.

Christine Garrington 07:04

What about for you? Matt, Tell us a bit more about your experiences then working with the data and what you've learned, if you like, that might benefit others who could consider using it, because that's partially what we want to achieve here. We want to, you know, to engage with more people around the data and make sure that they're aware of it and what it can do, what it could do for them and their research. So tell us a bit about that Matt.

Matt Wallace 07:05

Yeah, I mean, I mean, the answer is pretty straightforward, right? The answer, the answer is Hill Cooler, he alerted me to the data. He kind of provided me with that support and attention to put together a high-quality proposal to get my, my kind of PhD funded, to get me to where I am today, but I have very fond memories and the Celsius support team goes through different iterations, right? But the support team when I did my PhD was Chris Marshall, Rachel Stuchbury and Wei Xun, and they were always kind of patient, enthusiastic, engaged, understanding, particularly when your kind of learning the ropes of a new data set, but also new software to be able to kind of lean, organise and analyse that data. So kind of eternally grateful to that team as well. My experience was, was overall highly positive, right from the people promoting, to the people supporting the data. I used the LS a long time ago now, right? I think, or longish anyway, around 2012 to 2015. And thankfully, processes have evolved a little bit since I used the data. So when I wanted to do anything with the data, I either had to travel down to Pimlico to one of the ONS offices, or I had to send syntax files to one of the Celsius support team. And that's why I kind of valued their patience so much when you make errors or when things go wrong. But now, increasingly, there are safe pods around the UK where people can travel to to use the LS. You know, you don't have to go down to London anymore, or send files via email, or whatever. There may even be new avenues, because it's been a while since I use the LS. I mean in general; I mean research always seems like a race, but I think you know, what have I learned using the LS? Take your time, and kind of get to know the data. Get to understand its quirks and intricacies, because like any data set, it has its quirks and intricacies that you need to understand to use it properly, and to use it properly and to use it accurately and to produce reliable findings. Make use of the incredible amount of resources that are available to help you understand the data. There's an incredible amount of metadata. There are code books, data dictionaries, user guides, there's training, there's workshops. Celsius give you every possible platform to be successful with using the LS. So take those opportunities when they come, because they are there.

Christine Garrington 09:48

And what about for you, Joe, as someone perhaps, has been using it more recently. Is there any more up to date information that you could share to to top up what Matt has said there?

Joe Harrison 09:57

No. I would say similarly. So I am more fortunate than Matt, I didn't have to go to London every time I was able to use the safe pod. So also, thank you to the University of Dundee Safe Pod team, because they hosted me a lot. Take advantage of the information in there, and also prepare your own information beforehand as best you can, especially if you're going to these safe settings where your restricted access, so you can't use the internet, for example. So I prepared preliminary code, even though I wasn't actually sure what I was doing fully, but so that you can just save time, especially in a safe one, you might have a limited amount of time, you might just have the morning, so be aware of that. There is a lot of user written code, so I actually used, because country of birth has been asked at every Census, how to use that, I think it was written by Matt, actually. So how to shape the data in a way so you can find the modal country of birth that people have selected over time. There's also lots of geographic ones, so I don't use geography particularly, but there's a lot of code around how to get various levels

of geography. So if people are looking to use something more spatial, and they should look into the code that has already been written and is available in the safe settings and the SRS to use. And the other advice I would maybe give is when selecting your variables, some variables seem to overlap, but I would just select both, because that is a way of making sure that you have the information you need, rather than going back and asking for a variable, because that can cost a lot of time.

Christine Garrington 11:24

Yeah. Brilliant. Some great tips there, Joe, thank you. Now, as we speak, you've got a couple of soon to be published papers using the LS, looking, as you mentioned, at cancer and immigrants, and comparing the mortality of immigrants in the UK and Norway. As you said, what key things have you learned there? So you know, what are we finding out from this data?

Joe Harrison 11:45

So in terms of cancer, so a paper is going to come out in the next couple of weeks, being accepted on all cancer incidence and mortality of Pakistanis and Bangladeshis in England and Wales. Here, we're finding what we've known for a long time is that low cancer mortality exists for these groups, but I've started to, well, I've hoped to identify that this is driven by actually not getting cancer. There's low cancer incidence after incidents. Mortality is around the same as the white British majority population. And also starting to see that for descendants, there may be different mechanisms in the cancers they are getting. It's hard to pinpoint exactly, because it is quite a small group. They're not getting that many cancers, but the cancers may be more aggressive, and early life mortality may be higher. I'm saying may, I'm being very cautious with the words I use, because it's difficult to get a level of certainty, but that's definitely something that I think should be explored going forward, when the continuation of the LS and more Pakistani and Bangladeshi descendants enter ages of like 40 and 40 and beyond, that's something that I think we need to continue to study.

Matt Wallace 12:54

And the thing is, as well, on that point is the more we add to the LS, the more Census is added to it, the more we update all of this kind of life course information, the more valuable the LS comes to migrant health, because fundamentally, migrants are aging, but there's still quite a young group in many, in many ways, and problems with health are rare, and the risk of mortality is rare. But as these groups age, and we can capture that in the LS, we can really start to say more detailed things with a lot more certainty.

Christine Garrington 13:23

Yeah, it's really important to draw that link, isn't there? Okay, it's, you know, it's brilliant for the research, but it's brilliant potentially, for the populations that you're studying and the individuals and the groups that you're looking to help fundamentally, with the research you're doing. Yeah, fantastic. And now, Matt, as we said earlier, you really do have an impressive track record of publications using the LS. I wonder if you'd be happy to highlight a couple of key things that we've learned from your work. I hate to sort of drill it down to one or two, but and why you think they're important or have helped you know greater understanding.

Matt Wallace 13:55

Yeah, of course, I always like to qualify these things these days by saying that migration is inherently complex and diverse and heterogeneous. And some of the things I'm about to say don't apply to all migrants or all migrant groups. You know the drill. But fundamentally, what have I kind of managed to say with the LS so far is that migrants, international migrants, with, of course, some variation by country of birth live longer than non-migrants, generally, that owes to lower mortality from non-communicable diseases, like cancers, cardiovascular diseases, certain respiratory diseases. And that low mortality cannot be explained by problems with the data. Now, not problems with the LS, but one of the very common explanations for lower migrant mortality and better migrant health is our inability to really capture migrants in the data that we're using. These data sources are not set up to capture the mobility of migrants, and there are very many potential sources of bias that we can introduce into our

data. But I've hopefully done enough with the LS at that point to kind of show that bias cannot explain the lower mortality of migrants. And the most interesting finding to me is that the children of migrants do not live longer than migrants or non-migrants, and is often the case that they actually have higher risks of death, particularly in young adulthood. To me, that represents a very worrying kind of intergenerational reversal in mortality from advantage to disadvantage. It says a lot about the experiences of migrants after they arrive, and certainly the experiences of their children with respect to social inequality, racism, discrimination, integration, and these are all things you kind of hope to raise awareness about and inform as you continue to produce research on migrant health.

Christine Garrington 15:49

Okay, so Matt, I'm keen for you to talk a bit about some research that you did with Fran Darlington Pollock, which showed that migrants appear to be living longer lives, but with higher levels of limiting long term illness. And I know that's also feeding into some new work with LS that you're planning, so talk us through that.

Matt Wallace 16:06

Yeah, of course. So this, was actually my most recent finding with Fran from the LS, that migrants might be living longer lives in worse health than non-migrants are. Particularly if migrants were born in low middle income countries like India, Pakistan and Bangladesh and the Caribbean, but also especially if they were women as well. This was really interesting finding, it kind of adds considerable nuance to the migrant health debate, and it suggests that the links between the health and mortality of migrants aren't as strong as we thought they were. This finding served as a really important foundational finding for a big funding bid that I put together, and it is thanks in part to that LS research that 1.3-million-pound funding bid was initially awarded by the European Research Council and then funded by UK research and innovation. We'll kind of spend the next five years looking at why migrants are living longer lives in worse health in the Nordic region and in the UK as well, at the Centre for Research on Inclusive Society at the University of Salford, and ultimately, look to fit another piece of the migrant health puzzle in place. I think what's really important in the context of the LS is it shows the power of the LS, both in the ability of the LS to do really complex, high level migrant health research, but equally, the role that LS research plays in helping to advance its users careers by showing to potential funders and reviewers of funding bids at this kind of high level research is possible.

Christine Garrington 17:28

You both, I think, continue to use the LS. What opportunities then, does the 2021 census data provide you and others with to dig deeper into these important questions, as you say, big issues Matt.

Matt Wallace 17:43

Absolutely. I mean, as a kind of we lent upon it a little bit before, right? But fundamentally, it's about extending the LS and getting that kind of certainty of another Census point to enumerate people again, to know where they are, to know where they're living, to be able to look at changes in the various characteristics between the 2011 Census and the 2021 Census, whether that's health itself, whether that's self-reported health limiting long term illness, whether that's changes in people's education, changes in people's labour market circumstances, partnership circumstances, and it provides another 10 years of data for us to analyse and look at how health has changed over time, look at how migration has changed over time, and the links between the two.

Christine Garrington 18:25

Joe I'm guessing you can't wait to get your hands on it.

Joe Harrison 18:29

Yes, if it comes out prior to acceptance of a paper I've got under review at the moment, then I'll probably will update my results to extend to 2021, especially if the linkage to the NHS has been done and the death records are in place. So yeah, that'll be good it'll be a more accurate, more correct, and more time relevant to the research I've done. Even if it's just an extra three or four years, it's better in reality. And I can compare the same to Norway, I can extend the data there too. There's also, I think we alluded to it before, the aging of the second generation, getting more information about their circumstances, especially the interaction with socio-economic success that we know health has, and whether that is part of the reason that this migrant mortality advantage is being reversed. So we'll have more information about whether these groups of migrants are employed to the same rate, is the same sort of occupations, the same education level as the majority population, and whether that's a mechanism that is causing the reversal of this advantage, so the more data better.

Christine Garrington 19:31

Final question to you both, if I may, as champions of the LS, how would you or do you encourage other researchers, your colleagues, your peers, to work with the data. What would you say to persuade them additionally of its merits? I feel like we've done a lot of that through this conversation, I hope. But I wonder if there's anything else that you would say that would perhaps push people to give it a go and try it out and go check it out.

Joe Harrison 19:55

Yeah, I would say, if you're considering it, you think it could have merit in your research for any topic, not just migrants and migrant health, I would reach out to Celsius. Have a look at the data dictionaries. Jemima Stockton and Ali Sizer have been part of my journey and my PhD, and they've been so helpful. Like Matt says, the support from Celsius is, yeah, unbelievable. They're willing to give information all the time, and the release of outputs has actually been very quick. I've been super, super happy with how quickly it's come. So, yeah, I would reach out to them, because you need them to start the process and but they'll help you also develop your ideas, and they can reality check you if your idea maybe isn't feasible, but also guide you in a way that is like, maybe you could do this.

Matt Wallace 20:40

I mean, the LS is representative, it's reliable, it's robust. There's a wealth of high-quality kind of metadata information there to learn more about it. There are dedicated support teams in place, the level of support, I think, is unparalleled for any other secondary data resource in the UK. It's just a really fantastic, flexible, rich data source that people should be using. And I still talk about it, any chance I get. You know, I've worked in Sweden until recently, for the last five and a half years. And of course, this the total population registers there, which are often seen as this, this panacea of quantitative secondary data, but the LS itself it's a fantastic data source, and it doesn't come too far behind in my opinion. People should be using it, we need to still keep trying to raise awareness about it and telling people about the merits of the LS and it's one of the UK's largest and best quantitative data sources, that's the bottom line I guess. People need to use it.

Christine Garrington 21:43

Amazing data, amazing support. Matt and Joe, thank you so much for being champions of the ONS Longitudinal Study and for your contribution to research in the area of the health of immigrants. 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 Podcast, you can follow us on x at @CelsiusNews, and you can also find us on LinkedIn. Thanks for listening.