# Transcript: Exploring Ethnicity through National and European Datasets

[0:00:00]

**Stephen Jivraj:** Hello, my name is Stephen Jivraj, and I'm from University College London, and I'll be joined in this presentation today by Momoko Nishikido, who is from Queen's University Belfast, which is where there is a project that Momo and I are currently working on called the Geographies of Ethnic Diversity and Inequalities.

Today, Momo and I will be talking to you about how ethnic group is measured in censuses and surveys in England and Wales.

Now, I'd like to start with some information that the UK Data Service has asked us to present to you about information that's included amongst their resources that has ethnic group in those data. So, primarily, if you look at many of the UK large nationally representative surveys, let's take the Health Survey for England as an example, it includes information about ethnic group. But for those who are interested in more detailed information about the employment, the health, the discrimination experience that some ethnic minority groups might be having, then the EVENS is perhaps the quintessential dataset with which to use.

The Health Survey for England, EVENS data, other data that include ethnic group information like the Crime Survey for England and Wales, the Annual Population Survey, all of these are nationally representative but cross-sectional surveys, so repeated at different time points in some of them, although EVENS has actually only been collected once.

And there are other cross-national surveys, like the European Social Survey, that include information about ethnic group, and other international databases, lots of which are available through the UK Data Service.

Today, we're going to focus in our presentation on talking you through how the Census measures ethnic group and what the data on ethnic group show us, and then in a later presentation by Lucinda Platt, Lucinda is going to talk to you more about longitudinal resources that include information about ethnic group.

So, to start with, in terms of large social nationally representative studies in England and Wales that include information about ethnic group, so let's take the Health Survey for England as an example, although the Health Survey for England has a large sample each time the data is collected, about 10,000 people, that doesn't allow you to do analysis of very specific ethnic groups.

So, take the Chinese group, about 1% of the population nationally. So, on average in any given health survey for England, you would only expect around a hundred people in the Chinese group, and that's not enough to conduct analysis by age, sex and other characteristics.

So, in these nationally representative surveys with these large, but in terms of doing ethnic group specific analysis, quite small sample sizes, the way to go about doing analysis is either to use reduced categories of ethnic group or to pull data together across survey years where it has been collected at multiple time points to provide a sufficient enough sample with which to do more finely grained subgroup analysis by ethnicity.

In terms of the Census, which is what we're going to talk most about today, there are two assets that we'd like to make you aware of that are available from the Census.

So, there is aggregate data, which many of you may have used through the Office for National Statistics website or maybe Nomis, and that aggregate data is amazing, aggregate in the sense that it's provided for geographies, but it can provide it to very low level geographies. So, you can find in your

neighbourhood the distribution of people across the categories that are included in the Census in terms of ethnic group.

[0:03:54]

As well as aggregate data, there is also microdata available from the Census, much less used, but even more powerful. And by microdata, we mean individual data that you can access from each of the censuses going back to 1991, both through the UK Data Service as well as through other services as well. There is also longitudinal microdata available from the Census, but I'll leave Lucinda to talk more about that in her presentation.

Now, if you want to learn more about how the Census includes information about ethnicity as well as all of the other resources that UK Data Service hold, then their learning hub is the place to go to.

But what I'd like to turn us to now is to talk about how ethnic group has been measured in the Census. So, ethnic group was first included in the Census in 1991, where the question asked people to describe their ethnic or racial group in relation to their ancestry. And there were, by comparison to the later Census, a more reduced set of response categories for people to describe themselves as. And we should say that on the ethnic group question in 91, as it's been for all the censuses after that point, people are always asked to mark one ethnic group that they identify with.

Fast forward ten years from 1991 to 2001, then there was quite a big change in how ethnic group was asked. The question was changed in terms of its framing, so the question asked about people's cultural background rather than their ancestry or their racial group. And also, there were a number of categories added to the response options that people could pick, most notably, the mixed group was brought in 2001 and it was felt that prior to this point, people would typically identify with their minority identity. By 2001, it was thought more important to collect information about this mixed group.

Fast forward another ten years to 2011, very little change to the response categories. In fact, the only difference is that an Arab category was added to the other group and the Chinese group, which was previously in the other group, got switched to the Asian group. But a slightly more fundamental change between 2001 and 2011 was the framing of the question. So, in 2011, people were asked what is their ethnic group, and there was no other context provided about that as there were previously in 2001 with cultural background and 91 with ancestry. And in 2021, the question was exactly the same as it was in 2011 and the response categories were near identical with very little alteration.

So, what did the data show from 1991 up to 2021 in terms of these ethnic group? Well, what we present here is a reduced set of the categories that were available in 1991, enabling us to look forward. But essentially what we're looking at here and the entire size of these bars is the total number of people who identified as having a non-white identity at each of the censuses.

So, in 1991, this stood at around three million, or about 10% of the population, but by 2021, it had grown to more than ten million people and more than 20% of the population in one of these non-white ethnic groups.

And what's interesting is not only do we see a big increase in this non-white group over this time point, but we see an increase in the diversity of ethnic diversity over this time point, because you can see that all of these bars pretty much grow in about the same proportional size, and in fact, the group that's largest by 2021 is the other group, which is the most ethnically diverse of all of these groups. So, over this time point, there's been a big increase in the diversity of diversity.

Now, if we look at more local levels, what we can see is some of the places, particularly in London, which are incredibly diverse, what we might describe as super diverse. And the quintessential example of this is Newham in East London, which is a remarkable place in terms of its ethnic group distribution

because there is no group that accounts for more than about a sixth of the population. And in fact, there are probably six or seven ethnic minority groups that each account for more than 10% of the population. So, it's an incredibly diverse place.

But many other places in London also have very diverse ethnic profiles, as you can see listed on the slide there. Now, outside of London, there is no place anywhere near as diverse as Newham, but there are many places where the non-white group of people are greater than the entire white group.

Now, that gained a lot of traction in the press with the publication of both the 2011 and the 2021 Census data, in that there are now, what we describe, as these non-white majority cities. But we think that's a real misclassification of what's going on here, because although on all of the places you can see listed here on the slide, all of these local authorities, there is a higher percentage of people in ethnic minority groups than there is the ethnic majority white British group, in almost all of these places, that makeup of those ethnic minority groups is incredibly diverse. And there isn't one group that's dominating the entire population in that local authority. So, it's a bit of a misnomer to describe these places being ethnic minority dominated places. They are just places which are incredibly ethnically diverse and include people from many different ethnic groups, although a minority of their population is in the white British group.

I'm going to hand over to Momo next, who's going to pick up some information on these more locally, local information about ethnic group and ethnic group deprivation.

Momoko Nishikido:     So, I'll be continuing this presentation with this figure here, which shows the ethnic diversity index compared with the share of a local authority districts white British population.

So, each dot is a local authority district. On the X-axis is the share of a district's white British population, and the ethnic diversity measure is on the Y-axis with a higher value here meaning more ethnic diversity. And what we see is that the smaller the share of a local authority district's white British population, the greater the ethnic diversity of that district. And districts where the white British group makes up less than 50% of the population are mostly in London, but there are a few districts outside of London, like Leicester, Slough, Watford, Luton, Manchester, as well as Birmingham.

As the white British population drops below the 50% threshold, it's hard to predict ethnic diversity weakens somewhat, which we can see in this figure with the dots becoming more scattered below 50%. This could mean that other factors like the share of other ethnic groups drive this relationship.

And as Stephen just mentioned, rather than using the term minority/majority, which emphasises one ethnic group as a threshold, and it's usually the white group. The absence of white people in a place might be better suited as a signal of ethnic diversity and places being characterised as minority/majority should instead maybe be thought of as ethnically diverse spaces.

So, so far, we've taken you through some evidence that shows increasing ethnic diversity and residential mixing in neighbourhoods in England and Wales. While ethnic diversity is increasing, residential segregation in England and Wales has consistently been on the decline. We see that here in this figure, with the X-axis showing the census year and the Y-axis showing the segregation measure used, a very common one called the index of dissimilarity, this index here represents the spatial unevenness between one ethnic group and all others. So, for instance, the spatial unevenness between the Bangladeshi group and all others that are not Bangladeshi.

[0:11:47]

For all ethnic groups, each represented by a coloured line in the figure, we see steady declines in spatial unevenness or segregation between 1991 and 2021. The largest values from 1991 from the Bangladeshi and Pakistani groups, with values around 0.8, reflect historic immigrant settlement patterns.

Taking the most recent periods as 2011 to 2021, the greatest reductions in segregation were for the black African and black Caribbean ethnic groups. White segregation from minority groups also declined in this period and the largest proportion of decline in residential segregation between 1991 and 2021, which is the whole time period here, was for the black African group.

The Bangladeshi, black Caribbean, and Indian ethnic groups also experienced large decreases in neighbourhood residential segregation between 1991 and 2021. So, over time, there's been more ethnic diversity and mixing in residential spaces across England and Wales.

So, I'm going to switch it up a little bit here. From here, I'll talk about the Ethnic Group Deprivation Index, or the EGDI, which offers a novel perspective by considering for the first time how deprivation differs by ethnic group and domain or type of deprivation in neighbourhoods across England and Wales. And this index is calculated using census data and is made up of four dimensions. So, that's employment, education, housing, and health. And ethnic groups in an LSOA, or lower layer super output area, had to have at least 30 people in that ethnic group in that LSOA across all four domains to be included in the calculation of the EGDI.

And this map here shows the minority ethnic groups with the largest EGDI value for LSOAs meaning that they're the most deprived ethnic group in this neighbourhood relative to other groups that they live amongst.

And as a side note, this map is a cartogram, which makes it easier for us to see what's going on inside areas that may be smaller in size, but consist of larger populations, basically more densely populated areas like London. We

see that in parts of the east of London, the Bangladeshi group has the highest deprivation levels, towards the south of the river, there are a few places where the black Caribbean or the black African groups have the highest levels of deprivation. But overall, I think this map shows how complex this is. There are patches and concentrations of areas in LSOAs, which are the most deprived by a particular ethnic group.

So, while ethnic residential segregation has been declining over recent decades, as I just showed, it doesn't necessarily mean that people who are sharing neighbourhoods also share the same levels of deprivation given that we see here that certain ethnic groups do stand out in neighbourhoods as having relatively higher levels of deprivation.

And as I mentioned, the EGDI is made-up of four domains. In this figure here, from top to bottom, employment, education, housing and health. Along the X-axis is the district under which a given LSOA sits. The dot is an ethnic group/LSOA combination. So, each dot represents an LSOA and the first dot that you see in each panel represents the LSOA and ethnic group with the highest value of deprivation on that given domain. And the colour of each dot represents an ethnic group.

So, these charts here only pick up the top 50 most deprived ethnic group/LSOA combinations. And you see there's lots of colours on the graph, lots of ethnic groups, even in the top 50 deprivation levels for each of the four domains.

[0:15:26]

And so, what we really see here is the diversity of deprivation by ethnic group and by neighbourhood. There's also a lot of variability in the geography represented, although it might be a bit difficult to see closely in this figure on the slide here. Actually, there are lots of local authority districts included here, even in this very small extract from the EGDI.

If you look at the four plots, you'll see that the colours represented, and therefore the ethnic groups that those colours correspond to, they tend to vary. In education, the second panel down, for example, you see certain groups being more predominant than in, for example, health, which is at the very bottom. What this shows us is that it's important to remember that deprivation varies considerably by characteristic. So, the way that members of the given group are deprived will vary geographically, but will also vary by characteristic. A group that may be housing deprived doesn't necessarily mean they are also deprived in health or vice versa. And considering these complexities are really important. The combined deprivation measure then, or the EGDI value, is the sum in some sense of these different forms of deprivation, like in the index of multiple deprivation.

One of the most important outputs of the EGDI are the range values which captures the difference in EGDI values between the most and least deprived groups in an area. And this is what we've mapped here.

In simple terms, if there are very small differences – in this case, the range is small and closer to zero – the most deprived group and the least deprived group are very similar, meaning they have similar levels of deprivation. These neighbourhoods are shown in red on this map.

Also, something to note here is that this map only shows places that have more than one ethnic group present, which makes sense since this map shows the difference in deprivation level between the most and the least deprived groups.

So, all the reds are places where there are very small ethnic inequalities. It may be that the deprivation experience of different ethnic groups is almost identical. Those places that are in blue, and to some degree those in green and yellow, have large ethnic inequalities. So, there are places where the difference between the most and least deprived groups are very large. And we see these places, for example, in the east of London, places like Tower Hamlets, that come up in our research have very large range values. So, in

these cases, you have one ethnic group, which is in the bottom decile by deprivation, so it's the least deprived 10%, and at least one other group which is in the top 10%, the top decile, so the most deprived 10% overall. In these areas in blue, there are particularly pronounced differences between the most and least deprived ethnic groups within given neighbourhoods or LSOAs. And as you can see, there are lots of these examples, particularly in urban areas, but not solely. Places like London and Birmingham, for example, where there are a lot of localities, are places where we see large differences in deprivation. Places with these blue, green and yellow colours have differences in these cases of at least four deciles which is quite significant. That said, in some places there are differences as large as eight or nine deciles, so really profound ethnic inequalities and differences in deprivation which could be masked by any measure that captures the population as a whole, like the IMD. It's still a very rich and, of course, a very important measure.

But breaking apart the population like this into ethnic groups does reveal the fact that an overall deprivation measure may indicate moderate deprivation or even low deprivation, whereas when we start to look at this by ethnic group, we begin to understand that actually there may be pockets of deprivation in places where one group is really highly deprived, even within the same neighbourhood, where others are less deprived. So, interventions or support may not be targeted to members of that group.

[0:19:13]

If you're interested in learning more about the increasing ethnic diversity in England and Wales or about the EGDI, members of the GEDI Project have published open access articles on these two topics and the links to access them will be provided. Thank you.

Lucinda Platt: Hello, I'm Lucinda Platt, and in this presentation, I'm going to be talking about some of the longitudinal data that is available to us to study questions relating to ethnicity. I'm going to talk about these opportunities and I'm going to provide some examples from my own research.

First, I want to start by talking a little bit about the value of longitudinal data. Longitudinal data allow us to study short range dynamics, such as moves into and out of poverty, as well as long range processes, such as social mobility between generations, between parents and children, and across life courses of individuals.

Longitudinal data allow us to study transitions that are important in the life course, for example, from education into work or from work into retirement and the characteristics and circumstances associated with these.

Longitudinal data allow us to take account of circumstances in earlier life for later life outcomes and experiences. For example, the circumstances that children experience in early childhood, for later life childhood, for adulthood and even for later life.

Longitudinal data allow us to use the temporal ordering that some things necessarily happen before others and can be observed to happen before others to aid causal arguments. For example, about the relationship of attitudes and behaviours, do certain behaviours lead to a change in attitudes, for example, does the birth of a child lead to differences in genderal attitudes or do attitudes shape behaviours?

And it also allows us to take account of individual heterogeneity. By having repeat observations of individuals over time, we can take account of these fixed effects, these things that are specific to an individual that we're never going to be able to measure fully and take account of that in order to look at patterns across characteristics we do observe across multiple individuals.

There are a number of questions that are particularly interesting for ethnicity research that longitudinal data allow us to study. For example, given the heightened poverty risks of many ethnic groups, longitudinal data allow us to study how that is experienced, the duration of poverty moves in and out, and whether there are systematic differences in these dynamics by ethnic group that contribute to these cross-sectional patterns we observe and what the policy implications are of these.

For example, if moves in are more likely, then how do we avoid those moves into poverty? And if duration differs a lot, then that might lead to us to consider that poverty is more severe for some groups than others.

We can look at questions and whether there are differences in educational and occupational mobility across different groups and we can try and unpack what's driving those and how those apply to our understanding of mobility processes more generally.

We can look at whether there are differences in occupational aspirations themselves and how those help us to understand differences in labour market outcomes. Is it differing aspirations, differing motivations and intentions to end up in particular occupations or labour market statuses that lead to the cross-sectional differences we observe in the data? Or is the demand side more important than the supply side? And we can also look at the differing influence of parents, peers and community by these means.

[0:23:05]

We can look at whether different ethnic groups have different employment chances and wages as they transition from education to work and whether their patterns of transition from work into retirement also differ, and again, what the implications of these are.

Unemployment scarring has been much studied in the literature, that is, that the fact of being unemployed at one point in life affects labour market

opportunities and outcomes later on. And we can see whether this helps us to explain, observe differences in ethnic employment gaps.

And we can look at whether the strength or nature of ethnic or national identity changes over time and whether we can relate any such changes over time to patterns of socioeconomic circumstances and how those changes or patterns of socioeconomic circumstances thereby affect the demographic regularities in ethnicity and identity that we observe in the data.

But longitudinal analysis and longitudinal data do come with costs. It's costly to achieve large overall samples of different ethnic groups given the relatively small proportions of minorities in the population as a whole and it's costly to ensure high quality boost samples for existing survey data where there aren't pre-existing sampling frames, which is the case for most population groups.

And as a result, up to the early 2000s, there was little availability of such data. There were occasional studies which used specially accessed administrative data, for example, a study of mine which used housing benefit and council tax benefit from one local authority to look at poverty dynamics or low income dynamics, but that was the exception rather than the rule.

However, since the early 2000s, there has been a growth in longitudinal resources available for ethnicity research. And here I list just four of those, but the Millennium Cohort Study is a good example where they oversampled areas with relatively high proportions of lower income and ethnic minority families to enable separate analysis by certain ethnic groups and this is a rich data source that started around 2001 and we have information on or from the child at ages nine months, three years, five years, seven years, 11 years, 14 years, 17 years, and most recently, 23 years old. So, that gives us key information at regular transition points in the child's life.

Next Steps is also a cohort study. So, it took all those who were in Year 9 in England in 2004 as the basic population from which to sample. But again, it oversampled those from ethnic minority backgrounds enabling distinct research to consider ethnic differences. in these across the annual survey that took place between ages 13/14 and 19/20 and then subsequent follow-ups at ages 25 and 32.

Understanding Society is a household panel survey. So, here we have information across all age groups and for all household members in the households of the individuals surveyed. This started in 2009 and explicitly included an ethnic minority boost sample at the beginning. So, this boost sample enables separate analysis by key ethnic groups and there have been further studies, further boost samples, carried out at Wave 6 and another one planned again for Wave 20 in order to keep the population numbers up in this large study.

And then there were also studies that are not so much survey studies, but uses of data that provide rich longitudinal resources such as the ONS Longitudinal Study. So, this is a 1% sample of the Census population for 1971 that's followed up over time. And for all those who appear in this study who have also appeared in the Census from 1991 or subsequent censuses, we have ethnic group information. So, we can use that to chart back their experience, for example, the experience of those when they were growing up in 1971 and 1981, as well as looking forward to outcomes in 2001, 2011, and now most recently, 2021.

[0:27:44]

So, I'm going to give three examples of my research using three of these studies to show how they can be used in different ways and to address different questions that are very salient in ethnicity research.

First, I'm going to look at how I studied social mobility across ethnic groups using the ONS Longitudinal Study. So, here I count out a number of studies

looking at outcomes using the data released on the 1991 Census link, the 2001 Census link and the 2011 Census link and I'm going to give an example from the most recent study, the 2011, which looked at outcomes for 2011, which I carried out with Carolina Zuccotti, and here we looked both at educational mobility, so the chances of reaching degree level qualifications from those from different class backgrounds, and then the consequent occupational mobility. So, what were the labour market outcomes and the occupational outcomes across different ethnic groups and how far these differed by educational level and therefore were driven by educational outcomes.

So, importantly for this study, something that I implemented when I first looked at this, was using a prospective measure of mobility. So, here we measure the characteristics of families and parents, parental social origins, when the children were living there with their parents in childhood. And then we follow up the children who are the ONSLS members over time. And this is important because if we're looking at those with parents, immigrant parents, then it's important to measure the social class they had when they were already in the UK rather than mixing that with their potential occupational class prior to migration which may have differed in crucial ways.

Because of the extent and scale and richness of the data, it's possible to explore labour market outcomes for different groups and for more groups over the added decades as the study matures and what these studies have shown is the distinctiveness of the UK case when looking at questions with social mobility by comparison with other countries where there's been much more of a focus on using social class background to explain lack of educational mobility and lack of social mobility rather than as here exploring how there is greater educational mobility and consequently social mobility across minority groups.

And because social mobility is inherently a long span question, having information every decade is an advantage, having it over this long time sweep rather than an issue.

So, just briefly some findings. This shows that educational outcomes differ for all groups according to their social class origins. So, those from manual class origins are less likely to achieve degree qualifications than those from advantaged or what we call service class origins. While this is true across all groups, there is a class gap, a class effect. The class effect is much narrower for some groups, and in addition, the educational achievement of those from manual class origins is higher for all the minorities compared to the white British majority. So, in some cases, for example, you can see for Indians and Bangladeshi men, they are achieving degrees from manual class backgrounds at higher rates than white British from service class backgrounds which raises questions about how we understand class processes and what it is that's going into enabling or constraining achievement for those from different class origins.

And when we turn to labour market outcomes, if we concentrate just on the occupational outcomes that are in the right-hand side of the screen, we can see here in turn that to those with more educational qualifications are likely to do much better in terms of achieving professional managerial jobs, as we might expect, and therefore these qualifications have knock-on effects. But you can see that even here there's some additional advantage appears in occupational outcomes for men from some of the groups and the white British are doing relatively not so well in terms of occupational outcomes which was a kind of fresh insight.

And we see similar patterns for women, that those from highly educated outcomes are doing much better, and also that those with higher education qualifications have much more similar patterns of employment and economic activity and that the gaps occur much more for those with lower levels of qualifications. But in terms of occupational outcomes, the patterns are very similar for women.

The second example comes from Understanding Society, and here we were looking at the exit and entry among women of different ethnic origins to try and understand some of those labour market participation differences or gaps. I've shown that they're partly different by education, but they've been subject to lots of interest and attention as to why we have these differences in participation across women from different ethnic groups, and so we tried to come at this by looking at transitions into and out of participation by pooling pairs of waves from Understanding Society from the six first waves of data.

And we made use of the detailed information on, not only ethnic group, but also on own parental and grandparental country of birth to define the origins of each of the groups of women.

And we used a rich set of data, including around genderal attitudes, religiosity, family economic circumstances, to try and look at those factors that might be triggers or drivers of different patterns of entry and exit.

And here the size of the study and the richness of the variables were crucially important and the annual data enabled this approach of looking at the transitions, pair transitions, over the relatively short timescale that we needed.

In terms of findings controlling for other factors, we did find that there were some differences in patterns of entry and exit across groups, though there were also some groups where there was no difference from the white majority women in terms of their entry and exit probabilities. And we also found that there were some differences in the sensitivity to particular triggers, for example, Caribbean women's transitions were less sensitive to partnership changes than other women's and while traditional gender role attitudes were associated with lower entry and higher exit across groups, they didn't explain the differences in exit and entry for Pakistani and

Bangladeshi women compared to white majority women. So, that was in the face of claims that what's driving these differences are differences in traditional attitudes.

And finally, a short example from the Millennium Cohort Study where we examined the evolution of aspirations across three surveys, those carried out at age 7, 11 and 14, using repeat measures from the children's own questionnaire and combining that with information from parental interviews about the wider circumstances and parental attitudes in the household.

So, we matched this data also to labour force survey data. So, we could match the occupational information on the aspired to job to the objective characteristics on pay from the labour force survey to enable us to see how far children were aspiring to high value or less high value jobs.

So, the advantage here was that we could use the repeat measures of children at different ages to factor out individual heterogeneity that I mentioned earlier was one of the advantages of longitudinal data and by using growth curve models. And we could also look at whether aspirations evolved from the earlier ages to the later ages to both become more realistic in terms of what children aspire to, but also were they more influenced by peers than parents at those ages.

So, occupational aspirations and their implications are relatively under-researched compared to the wealth of data on educational aspirations. So, this was also a helpful new departure, and I say, we'd try and set some of these patterns on the map where we previously knew very little.

[0:36:43]

And here's just some illustrations of what we found. This is for boys of different ethnic groups and in each case the minority groups compared with the white majority you can see there's quite a lot of similarity in their patterns. They did tend to decline in the value of the job aspired to as they potentially became more realistic in their aspirations as they grew a bit older.

And there were a few differences across the boys to the extent that there were any, they seemed to be in the line of boys from minority groups tending to aspire to slightly more higher paid jobs than the majority.

And similarly with girls, a general pattern across all groups, they, in this case, were aspiring to slightly more highly paid jobs over time, but still not reaching the level of the aspirations of boys in terms of the value of the job. And again, to the extent that there were differences among the girls, these were in terms of the minority girls aspiring to higher value jobs compared to the white majority girls. And this again goes against some of the expectations that labour market disadvantages of some groups of minority girls might be associated with having kind of lower aspirations.

So, that's all my examples. Just to conclude, longitudinal data are very rich resources. We have a good set of longitudinal studies in the UK that offer potential exploring salient questions in ethnicity research and these resources compare well with those in all other countries, even those countries which have registered data in terms of having these surveys, which enable us to ask questions, tackling questions of attitudes and aspirations as well as objective measures.

And longitudinal research is important because it can address questions not otherwise answerable and it can shed fresh light on existing known irregularities that we pick up from cross-sectional measures. And they can also help us to think newly about mechanisms and social processes by showing us how, for example, social class background meets, it appears to mean different things for different ethnic groups, it raises questions about what the mechanisms are driving that.

Longitudinal surveys are though costly to run. It's a challenge to retain sufficient sample sizes over time because you're going back to the same people over time, you face attrition, and high quality and ethnic minority boost samples often add substantially to the costs.

Longitudinal data are often complicated to set up for analysis. They require a high degree of initial investment in preparing the data and thinking about the analysis and then using appropriate techniques. And taking account of that attrition that occurs in appropriate ways can also sometimes be a challenge.

Developments in linked administrative data offer some possibilities for new areas of research and also for resolving some of the challenges associated with survey research, but it's likely that survey data is always going to be necessary to address these specific questions we are interested in relation to ethnicity. There's a lot of potential still left, a lot more we can do.

Here are the datasets mentioned that you can learn more about and you have access to these slides so you can follow up these links and here are the studies I referred to. And I'll leave it there. Thank you very much for joining me.

Kitty Lymperopulou:     Hi, my name is Kitty Lymperopulou. I'm a senior research fellow at the University of Plymouth and ADR UK fellow.

In this presentation, I will be introducing the DataFirst criminal justice datasets and discussing insights and opportunities for researching ethnicity and the criminal justice system.

[0:40:37]

I'll start by introducing why we're interested in ethnicity and the criminal justice system. England and Wales have the highest rate of imprisonment in Western Europe where ethnic minority people make up 27% of prisoners compared to 13% of the population. Government statistics show that there are ethnic disparities across different stages of the criminal justice system. For example, the latest statistics show that black individuals are more than four times as likely to be stopped and searched by the police, 2.2 times as likely to be arrested and receive 23% longer custodial sentences than white individuals.

The latest figures from the Crime Survey for England and Wales show that 60% of black people and 49% of black Caribbean and white and black Caribbean people expressed confidence in the police, compared to 68% of the population. This highlights significant trust deficits within certain ethnic communities, ethnic minority communities, and the issue of trust is compounded by systemic issues such as those identified by Dame Casey in her 2023 review which concluded that the Metropolitan Police was institutionally racist.

These statistics highlight the need for research evidence to guide reforms that promote equity and justice within the justice system and to build public trust and accountability. The Ministry of Justice DataFirst datasets, which I will talk about today, were developed in response to a government-commissioned review into the treatment and outcomes of black, Asian, and minority ethnic individuals in the criminal justice system, published by David Lammy in 2017, which highlighted significant data gaps and called for evidence-based explanations of ethnic disparities in the criminal justice system.

DataFirst is a pioneering data linking programme made by the Ministry of Justice and funded by ADR UK and the ESRC based on administrative datasets from criminal justice jurisdictions in England and Wales which have been linked at a person level to enable analysis of justice system users, the pathways, outcomes and interactions with other jurisdictions.

The criminal justice datasets include HMCTS or Courts and Tribunals Service case management data from magistrates' courts and the Crown Courts, as well as records from prison and probation services for management of people under their supervision or custody.

A linking dataset allows these to be joined on person and, in the case of the courts data, on individual case level. More recently, the Ministry of Justice released family court and civil court data that can be linked with the criminal justice datasets using a newly released cross justice linking dataset.

The DataFirst programme also involves linking of justice data with other government departments, including the Ministry of Justice DfE data share, which links the national police computer with the national pupil database. This is a separate linkage, and at the moment, it is not possible to link it with the other datasets.

So, these datasets offer major opportunities, but they also come with some limitations, and especially when it comes to ethnicity. So, to illustrate the strengths and limitations and also insights that can be drawn, I'll focus on the criminal courts datasets which I've been using for my research.

There are two types of criminal courts in the UK. So, magistrates' courts and the Crown Court. All criminal cases start in a magistrates' court which deal mainly with summary or less serious offences. More serious offences are sent to the Crown Court for trial or sentencing. One of the key strengths of administrative data for examining ethnicity is its large size. These datasets contain records on all defendants appearing in criminal courts in England and Wales between 2011 and 2023 providing large samples to carry out research on ethnicity and justice involvement.

[0:46:09]

So, for example, the magistrates' courts dataset contains 18.1 million records and include approximately 110 variables giving details on characteristics of defendants, such as age, gender and ethnicity, offence proceedings and outcomes in criminal cases dealt with by the magistrates' court in England and Wales from 2011 to 2023.

The Crown Court datasets contain 1.2 million records, and each record captures one defendant case giving details on defendant characteristics, again, age, gender at least, but also residence, as well as offences, proceedings, and outcomes dealt with in the Crown Court. This dataset covers the period of 2013 to 2023, so a ten year period. The Crown Court

dataset includes over 140 variables, including the date of hearings, types of offences, arrests and court proceedings such as plea, bail and remand.

Before we can meaningfully analyse ethnicity with this data, I mean, it's important to understand how ethnicity is recorded in the datasets and how complete these records are. So, measuring ethnicity is sensitive and challenging. Ethnicity is fluid, subjective and multidimensional relating broadly to a social group or group identity and it draws on notions of shared origins or ancestry. And as researchers, we often rely on standardised ethnicity measures which classify individuals into fixed ethnic groups that are readily available in the datasets that we use. The ethnicity variable in the datasets is based on the ONS 16+1 ethnicity classification used in the 2001 Census. Self-identified ethnicity in the Crown Court's data is collected during an interview by the police and is entered into police administrative systems and passed to the Courts and Tribunals Service.

So, the datasets also include a measure of ethnicity which is also collected by the police and so it is ascribed by a police officer based on visual appearance which identifies individuals as white, black, Asian or other.

And you may have seen published criminal justice statistics based on police-defined ethnicity. It is important to know how the two relate to each other. So, a comparison between self-identified and police-defined ethnicity in the criminal courts' datasets shown here shows that there's a high degree of consistency between the aggregated self-identified ethnicity and police-defined ethnicity for the white, black and Asian groups, but this is not the case for people with a mixed ethnic background or for other ethnic minority groups not individually reported.

For example, around 40% of defendants in the white and black African and white and black Caribbean groups have been classified in the black ethnic group by the police and 30% to the other group.

We also observe an inconsistent recording of ethnicity by the police for the Chinese and other groups. Criminal justice statistics based on the two ethnicity classifications may therefore give a different picture about justice involvement and outcomes of ethnic minority groups. Police defined ethnicity should be considered as less reliable given that it is prone to inaccurate or false attribution and that it fails to capture the complex and multifaceted dimensions of ethnicity.

Another important feature of administrative data for ethnicity research is its coverage. A common challenge we encounter in ethnicity research is having insufficient numbers of or incomplete information on ethnicity.

Ethnicity coverage is also an issue in criminal justice research. So, when we look at the coverage of ethnicity in magistrates' courts data, where the majority of cases relate to less serious offences and defendants are therefore less likely to have been charged by the police, so this is when ethnicity information is collected, we see that around two-thirds of defendants had missing ethnicity. In comparison, less than a quarter of defendants appearing in the Crown Court's dataset had missing ethnicity.

[0:51:48]

Analysis of missing ethnicity reveals that ethnicity coverage varies across defendants. For example, it is less complete for females compared to males and for older than younger defendants. And it also varies across case characteristics. By far, the largest variation in the coverage of ethnicity of defendants in courts relates to offence. So, the figures here show ethnicity coverage by offence in magistrates' courts and the Crown Court.

The figures, so each bar in the figures represent an offence category. It shows that in magistrates' courts, it is ethnicity coverage in summary offences. So, these are offences that make up the majority of cases in magistrates' courts. That explains missing ethnicity. So, around three-quarters of defendants in summary offences are missing ethnicity compared

with less than a quarter of defendants in cases related to other offences such as drugs, possession of weapons, theft and violent offences.

In the Crown Court, there are lower levels of missing ethnicity across all offences and this illustrates that despite the large size and richness of these datasets, it may not always be possible to disaggregate outcomes by ethnicity because high levels of missing ethnicity can reduce the statistical power of a study and can produce biased estimates and inaccurate conclusions.

I will now turn to some research findings drawing on examples from my research on ethnic disparities in the criminal justice system. The research draws on the DataFirst criminal courts data and defendant appearances in the Crown Court over a four year period from 2017 to 2020 and excludes summary offences.

In this slide, you can see findings on patterns of ethnic disparities in the courts pipeline, so remand, plea, sentencing outcomes and offence types using the relative rate index which provides a standardised ratio measure of disproportionality that allows us to compare outcomes between ethnic groups. So, it is calculated by dividing the rate of an ethnic minority group in a given outcome by the rate of the white British group. So, a relative rate index greater than one indicates the outcome is more likely in a given ethnic minority group compared to the white British group.

So, the table shows that people from ethnic minority groups are more likely to be sent to the Crown Court for trial to plead not guilty and to be remanded in custody when they appear in the Crown Court compared to the white British group. We can also see that among those convicted in the Crown Court, defendants from the black Caribbean, Chinese and other white group were more likely to receive a custodial sentence than the white British group and the ethnic disproportionality is much more pronounced among young male defendants, so black Caribbean young males are more likely to

receive a custodial sentence than the white, their counterparts, and it is more pronounced particular types of offences such as drugs offences.

One of the key findings of this research is that the extent of disproportionality varies considerably between ethnic subgroups within the Asian, black, mixed and white ethnic groups. Published government statistics and existing evidence to date have been based on aggregated ethnic groups which mask this variation.

The data also allows us to go beyond descriptive statistics to determine whether ethnic disparities in the criminal justice system are unwarranted. In other words, whether ethnic differences in criminal justice outcomes cannot be attributed to differences in legally relevant factors which would provide an indication of differential treatment based on ethnicity.

So, the results here show, so they're based on multi-level modelling and they examine the effect of ethnicity on the likelihood of a prison sentence. They show that defendants from nearly all ethnic minority groups were more likely to receive a custodial sentence. So, the odds of imprisonment were 41% higher for Chinese defendants and between 16% and 21% higher for Asian defendants. They were also between 9% and 90% higher for black defendants while the odds of imprisonment were 22% higher for white and black African defendants compared to white British defendants after adjusting for other characteristics which includes offence type and severity and previous convictions.

[0:57:36]

In contrast, when we look at the effect of ethnicity on sentence length, what we see is that ethnic disparities are largely explained by legal factors. So, after adjusting for these factors, observed differences between most ethnic minority groups and the white British disappear. The exception is defendants from Pakistani, Bangladeshi and black Caribbean groups who

are shown to have worse sentencing outcomes than the white British, receiving between 4% and 11% longer sentences than the white British.

The implications of these findings are that ethnic differences which remain after adjusting for other factors are indicative of biases in the criminal justice system and differential treatment of ethnic minority groups. So, the lower extent of any disparities in sentence length compared to imprisonment suggests that ethnic minority groups are treated more equally at later stages of the sentencing process which reflects, to some extent, the lower discretion of judges in passing decisions about sentence length which are primarily determined by sentencing guidelines and mandatory minimum sentences passed by Parliament.

The research findings and implications, which are part of the project Ethnic Inequality in the Criminal Justice System, are available in a range of publications, including a datacomic and a policy briefing you can see here from the ADR UK website and the project website, ethnicitycriminaljustice.co.uk.

My study is one example of what can be done with the datasets, but there are wider opportunities for researching ethnicity with the DataFirst datasets, and they include, I think, opportunities through linking with other criminal justice datasets, so linking the criminal courts data with probation and prisons to understand pathways and trajectories of people with justice involvement and to follow individuals over time to understand pathways and interactions in the criminal justice pipeline, for example, from arrests, court appearances, prison release, recall and reappearances. And there are also opportunities to gain new insights into some of the processes underlying racial disparities and racial bias, for example, my current studies using the cross-justice linkages to understand ethnic disparities across stages of the justice system and how early decisions contribute to cumulative disadvantage and impact later outcomes in the criminal justice system.

The ability to link different datasets also enables filling in gaps. So, for example, additional information on nationality and religion is included in the prison data that can be combined with ethnicity information in the courts datasets to define ethnoreligious groups or other ethnic national identities. And information can be drawn from other datasets, such as information on alcohol and drug abuse, including the probation data, can be used to better understand the circumstances of those impacted by the justice system and how they relate to criminal justice outcomes.

And finally, linking criminal justice data with external socioeconomic datasets, such as the Indices of Deprivation, enables researchers to situate criminal justice disparities in a broader socioeconomic context revealing how pre-existing inequalities, such as poverty and unemployment, influence justice involvement and outcomes.

I will finish the presentation with some information about the applying for access to the DataFirst datasets. So, access to the DataFirst dataset is either through the ONS Secure Research Service or the SAIL Databank and requires a project and a researcher accreditation. So, to access datasets from the SRS, applicants will need to be accredited under the ONS Accredited Researcher Scheme and you will need to apply for access to the specific data that's required for a research project from relevant data owners as well. So, this is done by completing separate forms, depending on how you're planning to access the data. And you will need to submit a research project application for accreditation through the Project Accreditation Panel, also known as RAP.

[1:03:10]

Thank you for listening.

[End of Transcript]