

# Transcript: Exploring Social Inequalities through National Datasets



[0:00:00]

Jools Kasmire:

Hello, everyone. I'm Dr Jools Kasmire from the UK Data Service Computational Social Science Training Team based in Manchester.

Today I'm going to talk to you about the UK Data Service and how it can contribute to your exploration of inequalities. I'm going to start by giving a quick summary of what the UKDS is, then an overview of the data we hold, and finally how you can get started finding and accessing our data.

UKDS is a comprehensive resource funded by the Economic and Social Research Council, one of many funded by UK Research and Innovation. As such, we provide access to the largest collection of social, economic and population data in the UK. Data providers deposit their data with us and then we make this available to users. As well as providing access to all of that data, the UKDS provides support, training and guidance to help researchers find, access and use this data.

So, who is it for? Well, we'd like to say it's for everyone. This includes, but is not limited to, academic researchers and students, government analysts, charities and foundations, business consultants, independent research centres, think tanks and more.

And our data comes from many different sources, such as official agencies, largely the central government, international statistical time series, research institutions, individual academics who make their work, their data, available through us as part of their research grants,

market research agencies, public records and historical sources and more.

And we hold a lot of different types of data. We hold quite a lot of survey data, both cross-sectional and longitudinal. We also hold aggregate data, international macro data, census data and qualitative and mixed methods data as well as a few things that don't fit really well into any of these categories.

So, now I'll just quickly cover how you can find and access all of this data from the UK Data Service.

If you don't know how to get started, we recommend you click on Find Data, the tab right at the top. This directs you to a few common options, as well as some links to tutorials and more. We'll come onto the catalogue search later, but let's start by browsing here, just in the centre, browse and access key data.

Theme is the first browsing option presented to you. Inequality is not a theme in its own right, but some of the themes are clearly relevant, such as economics, education or ethnicity. There are more themes if you scroll down and if you keep scrolling down you find you can also browse by data type, and if you keep scrolling down even more, you can browse by teaching data sets and then a few other categories that we have grouped together under general. Geography might be quite interesting if you're looking at inequalities in terms of geographic distribution.

Okay, we're back at the Find Data tab, this one right here at the top, and this time we're going to search the catalogue. You can find some guidance on how to search the catalogue, but a good option is just to enter a keyword into the search bar and hit search. I've entered here "inequality". Just to be clear, this same search bar is on the home

page, but whether you start here through find data or on the homepage and use the search bar, you will get some results.

[0:03:48]

The catalogue search tool is our most popular way to find data because it's very intuitive to search by keyword or topic, but you can also use the search bar to find data by study name or number.

When you have your results, you can refine them using the filters on the left hand side. So, you can filter by date or topic or type, anything that you feel will help you drill down to get the most relevant data for your research.

Next, I recommend you check the Variable and Question Bank which you can access in a few ways, but the easiest is just here from the results under the search bar. And this helps you find particular variables within data sets that are relevant to your keywords. So, for example, if you type "inequality" into the Variable and Question Bank, you get a different set of results than you did from the basic data catalogue search bar and that's because these are the results that contain a variable talking about inequality. Just a note on the Variable and Question Bank though. It doesn't contain all of the data sets we currently hold, so it's worth searching the catalogue for earlier or later versions of a data set that's returned under the Variable and Question Bank search just to make sure and see if that same variable was included in previous or later editions of a particular search result.

The options on the left hand side here, you can expand with these little plus signs. And again, they allow you to filter your results and get very relevant data back.

You can of course search in the QualiBank and this is useful if you're interested in qualitative data as it allows you to search key terms within the different types of qualitative resources. Again, you can drill down using the filters to find your relevant results.

Finally, we have the HASSET Thesaurus. HASSET stands for Humanities and Social Science Electronic Thesaurus. I say that because it may not be obvious what it stands for. It wasn't to me. This search tool contains key social science terms and related concepts. For example, I found equality here and it is categorised under the right to non-discrimination and that itself is under civil and political rights, which is under human rights, which is under well-being (society), and it's clear you can find this from the individual results, but you can also search for it using this hierarchy tab. Obviously the screenshot here is of the alphabetical listing which is pretty self-explanatory.

The thesaurus is useful to show you other terms that might be useful to search in either the QualiBank or the Variable and Question Bank or the main catalogue and you can search these other terms, maybe that fall higher or lower in the hierarchy, or that appear under the related concepts.

Now just a quick point on the different data access levels. Starting at the most restricted types, we have controlled data which is also called secure access data. This data can only be accessed through a safe room, secure lab or within the UK or at the SafePod Network.

Next, we have safeguarded data which is available through the end user licence. Access to this kind of data requires registration. Some of this data is also special licence data which has additional requirements.

Then there's reshare data which is self-deposited by data creators or owners. Access to and controls on this data can vary.

Finally, the least restricted type is open data. This is data that can be accessed by anyone, even without the need to register.

[0:07:42]

Importantly, most of the data in our collection can be accessed free of charge. You can simply register with your institutional login. If you don't have an institutional login, you can apply to get a UKDA user name.

Thank you for listening and do let me know if you have any questions.

Nissa Finney:

I'm Nissa Finney and I'm going to talk about the Evidence for Equality National Survey and how these data can be used to research inequalities.

So, what is EVENS, the Evidence for Equality National Survey? It's the largest and most comprehensive survey of ethnic and religious minorities in the UK. The data were collected in 2021 and there are over 14,000 participants in the survey, including 10,000 who identify as ethnic minorities. So, this data set is really very novel and rich for understanding inequalities between ethnic groups.

It's also notable that EVENS has used a very innovative approach to getting more and more representation of ethnic minorities in the survey. So, it's a non-probability survey.

EVENS was motivated by a context of ethnic inequalities in the UK that have been shown to be stark and persistent and many commentators have argued that these ethnic inequalities were exacerbated during the pandemic. Yet researchers were really calling for attention to the data about ethnic inequalities and ethnic groups and particularly emphasising the ethnicity data gap or the problems with the data available. So, this was the second motivation for EVENS to try to address this ethnicity data gap because existing surveys tend to represent a limited number of broad ethnic groups and the sampling

for ethnic boost samples tends to be in areas of residential clustering of ethnic minorities. So, there are certain biases in survey data that we try to address in the EVENS survey.

General surveys also, whilst having a great deal of very rich information, don't usually have questions that are particular or bespoke to the concerns of ethnic minorities and while we have fantastic coverage of the population, including ethnic groups in census data and other largescale administrative data, the topic coverage is very limited.

So, EVENS is one of the very first examples of the application of a non-probability methodology to try to address problems of representation and to improve the quality of survey data for understanding inequalities, particularly with minority populations.

The data for EVENS were collected in 2021 via a 30 minute questionnaire, which was developed with partners and this was available online via an open web link and by telephone interview in 14 languages that are shown here. It was an non-probability responsive recruitment design, which means that there is no sampling frame, an open invitation was issued in anybody who considered themselves to be an ethnic or religious minority could take part. There were some eligibility criteria. People had to be resident in England, Wales or Scotland, for example, and all who completed the survey received a £10 voucher.

The questions in the survey were developed from other surveys together with some original questions and the survey was administered by Ipsos under full ethical approval of the University of Manchester.

And the recruitment was done very much in partnership with voluntary sector partners that are shown here. And these partnerships with

these organisations were really central to enabling EVENS to recruit a diverse range of people across ethnic groups and across parts of the country.

[0:12:15]

So, this recruitment resulted in a sample of over 14,000 and you can see here that there are fairly large samples for many of the ethnic minority groups. So, 1,300 for India and 1,000 for Black African, almost 900 for Pakistani, almost 700 for white other identifying participants. And we have a large Jewish sample as well, nearly 700. And the survey also has fairly large samples across religious groups and regions of Britain.

There are large numbers in the young and mid-adult ranges, smaller numbers in the older age group. So, care has to be taken in using the older minority samples. And thus, it's important that there are weights in the data set, which indeed there are, both for coverage and for selection bias, which is particularly important with a non-probability sample. And the coverage weights are baselined to Census 2021. So, it's really vital when using the EVENS data to use the weights in your analysis.

So, what can you do with the data? What is in the data? There are lots of topics. You can find out about socioeconomic and financial circumstances, identity, housing, demographics, mental and physical health, political involvement in movements such as Black Lives Matter, social cohesion, belonging, attitudes to the police, COVID-19 compliance, trust in government and racism and discrimination.

And there are two particular modules that are quite rich in EVENS compared to other data sets. The first of these is the ethnic identity variables in the survey, which includes an open response question

where people could write in their own words their description of their ethnic background together with other categorical variables as standard, information on synagogue membership, the importance of ethnic identification and ways of participating in ethnic activities. There's also a country of birth in there.

And on racism, there's a really rich module including experience of being insulted, having damage to property, physical attack, unfair treatment in a range of settings, treatment by neighbours, worry about being harassed and the change in these experiences of racism particularly during the COVID-19 pandemic. And these questions are asked in relation to a number of time periods of the respondent's life, so it's possible to build up a longitudinal sense of experiences of racism. There's also a very interesting question on how people respond to experiences of racism.

There are some variables in EVENS that also worth noting that are quite unusual to the data set in some ways. Type of accommodation, access to sanitation and water, outdoor space, working arrangements, financial circumstances, loneliness, mental health, very specific physical health conditions, receipt of care, experiences of COVID, income and immigration status.

So, after all that introduction about the data, let's see the kinds of work that can be undertaken with it. So, you can find out a lot more about the results of the survey and the possibilities in the EVENS book which is available by the CODE website [ethnicity.co.uk](http://ethnicity.co.uk) as a free eBook, and this is a thematically organised book that presents a lot of findings based on the EVENS survey about ethnic inequalities.

I'll pick out just a few illustrations to show and the first is on experiences of racism which lots can be done with the data and I refer you to this forthcoming paper to see some examples, but a headline result is that 80% of EVENS participants from minoritised ethnic



groups have experienced some form of racial discrimination. So, none of the data set can really give us this kind of information and it really does emphasise that racism is the common experience of ethnic minorities in Britain.

[0:17:12]

This chart shows the likelihood of being bereaved by ethnic groups, so there's a row for each ethnic group and the green point on each row indicates the odds of having been bereaved relative to the white British population which is indicated by the red vertical line. So, if the green dot is to the right of the red vertical line, it indicates that that ethnic group has higher odds of having been bereaved compared to the white British.

If you think about what this means for a moment, being bereaved, someone close to you has died recently and the impact that can have on your life. This is a very significant indicator of people's life circumstances and we can see that whether it's bereavement due to COVID, which is in the top chart here, or for other reasons, which is the bottom chart, the likelihood of being bereaved is far greater for most ethnic minority groups than it is for the white British group. And I think this is a very stark and unusual measure of ethnic inequalities that really causes us to think about the daily lived experience and the difference between ethnic groups.

This chart shows the rates of having financial difficulties again with a row for each ethnic group and again relative to the white British. And again, we see that there are higher rates of experiencing financial difficulties for most ethnic minority groups compared to the white British group. If we take that Arab group, for example, towards the bottom of the chart, they have an over 20% higher rate of experiencing financial difficulties than the white British group.

We can also look at belonging with this survey. And this chart shows the particularly high likelihood of having a strong sense of local belonging for Indian, Pakistani and Bangladeshi populations in the UK which raises really interesting issues about community, about community resilience and sustainability and about social cohesion which could be explored in relation to other aspects of the survey.

And we also see in this chart here that most ethnic groups have very strong feelings of being part of English and British society. So, this is the probability of feeling part of society and you can see that this is over 90% probability of feeling part of society for most ethnic groups. The exception here, which we should take note of, being the Roma population.

So, that's just a snapshot of some of the possibilities and results from EVENS in a descriptive sense and this is an example of some of the current projects that are ongoing using EVENS. There's work going on to look at life course experiences of racism; the connection between racism, ethnicity and loneliness; protective effects of religion for loneliness during the pandemic; mental disorders during the pandemic across ethnic groups; local neighbourhood belonging and ethnicity; how ethnic identity is articulated and what this tells us about ethnic group categorisation; political trust and how this connects with COVID-19 compliance; working on flexible contracts and how this connects to ethnic and gender inequalities; and methodological work about producing robust non-probability survey data that can better represent minoritised populations.

So, please do consider using the EVENS data. It's available from the UK Data Service, the full data set and also a teaching data set which is a simplified version and excellent for using in any classes or workshops.

And thanks here to the EVENS team, to the ESRC, to CODE and to all of the EVENS partners shown here who made the survey possible.

[0:21:44]

Bożena Wielgoszewska: Good morning. My name is Bożena and I'm a senior research fellow at the Centre for Longitudinal Studies. Today I'll talk about the project that, together with my colleagues from across the Social Research Institute, we worked on over the past few years with the aim to better understand gender wage gap.

This is the project on which Alex, Heather, Dave, Francesca and I have been working on since September 2019. This project has been funded by Economic and Social Research Council and aims to investigate the gap between men's and women's hourly pay over the life course and across generations. You can access a lot of information about this project, including the links to the papers that I'll talk about today, on our project website which you can access via the QR code here. And today I'll give a brief overview of the key findings from this project.

But before I get into the details about the project, I would like to show you some more general aggregate trends that are based on aggregated data. These plots show the percentage gender wage gap on the Y axis, so the gap of say 20% means that women's median pay is 20% lower than men's.

So, the first graph shows the evolution of gap over the life course of successive cohorts showing that it tends to be relatively small for younger adults, but gets bigger as people age, generally peaking around late 30s, early 40s, before it decreases again. It also shows that the gap has been smaller for successive cohorts.

And the second graph shows the same data, so it's based on the Annual Survey of Hours and Earnings and its predecessor the New Earnings Survey, but for different age groups over historic times. Here

you can see that the gaps were larger in the past and for older age groups, but have been decreasing over time.

So, such aggregate data is great for showing these trends, but often it doesn't have information about people's circumstances or jobs that they work in, which is why we need to look into other studies such as the British Cohort Studies.

And this topic has been studied by other scholars too, and here you can see findings from some of the studies published over the last 50 years. All of the studies estimated the raw gender pay gap, shown here in blue, and the adjusted pay gap, shown in grey. The adjustment here is for human capital, which is generally understood as given a person's educational attainment and work experience and here the gaps are presented as female to male ratios. So, if the pay of men and women were equal, these ratios will be one and anything below one indicates a male premium. In other words, a gap of 0.8 means that women on average earn 80p for every pound earned by a man.

As you can see, over time these ratios are closer to one. So, in 1971, married women and about 50% of male wage, but this increases for women overall to 82 by 2015. And taking account of differences in human capital between men and women always reduces the gap, bringing the ratios closer to one.

Over time, there have been a number of policies that were introduced with the view of increasing gender equality and some of the main ones that were introduced during that period that's covered by our research are shown on the timeline here.

The first one, Equal Pay Act, which was introduced over 50 years ago, mandated equal pay for the same work. The amendment introduced in 1983 extended this not just to the same work, but to work of equal value. Some of these policies were not aimed specifically at pay, but

at increasing employment rates of women. There have been a number of policies related to childcare which typically affects female employment rates, such as Employment Protection Act that introduced statutory paid maternity leave and more recently the right to share parental leave between both parents. Some others were aimed at flexible working and part-time working where part-time workers are also typically female. For example, the policy setting the minimum hourly wage and giving employees the right to request flexible working.

[0:26:14]

In our project we used the data from the British Cohort Studies as these studies have a rich set of indicators about the person's characteristics, circumstances and employment histories and etc, etc, which allows us to better understand the reasons behind the pay gap over this period of time.

Four of the course studies shown here are hosted at the Centre for Longitudinal Studies. That's the National Child Development study, shown in light blue; the 1970 British Cohort Study, shown in green; Next Steps, shown in grey; and the Millennium Cohort Study, shown in dark blue. This graph also includes the timeline of earlier cohort National Study of Health Development who were born in 1946 and the dots on these arrows reflects the timing of their main sweeps of data collection.

The first publication that I would love to talk about is referenced here, is focused on the wages of the participants of the National Child Development Study when they were between 23 and 55 years old. This graph shows the gap as female to male ratio where the black line shows the raw estimates for different ages. The grey line, which is partially hidden behind the dark blue line, shows estimates adjusted for education only. The dark blue line shows estimates adjusted for

presence of children and partner. And the light blue line on top shows the estimate adjusted additionally for work experience.

Here again the gaps are U-shaped, the raw gap growing initially until the peak at age 42, but closing in later life. The adjusted gap is peaking earlier at age 33 as experienced differences account for more of a gap from age 42 onwards. At age 23, the estimates from all our models are similar, implying that differences in education, experience and family situation, although small at this age, make little difference. Over time the adjustment for work experience makes substantial difference, especially in later life.

The differences in experience were particularly large for experiences in full-time jobs and part-time experience contributed very little to women's pay.

In the subsequent publication referenced here, we extended the analysis to the younger cohort, the cohort born in 1970, to compare the differences. This paper also extended the analysis to account for attrition of cohort members over time and also for selection into employment given that their differences in employment participation rates over time, especially for women, especially once they have children.

Here the solid U-shaped lines show female to male ratios that were shown previously. The red line is the raw gap that you saw on the previous slide. The black line, which is not much different from the red line is adjusted for attrition. The green additionally accounts for selection into employment. So, here we took into consideration what the potential wages of those who are not observed in paid work would be. We did this by allocating them a pay of the nearest neighbour, so a person in our sample with similar characteristics that has earnings.

As you can see, this makes little difference in later life, but the green lines are below the red and black line which shows positive selection into employment, meaning that the gaps would be even larger if more young women were employed. The dotted lines above show the estimates for the BCS 70. The gaps in this cohort are similar for similar ages, follow the similar pattern over time but are smaller.

[0:30:02]

We then focused on young adults, even though the gap at this age is relatively small. This precedes the acquisition of family responsibilities which means that we can sidestep the impact that parenting has on wages. So, this paper compares how the pay gap in mid-20s, although for slightly different ages in different cohort, has changed over the four cohorts.

And here you can see the convergence in employment and education rates over the successive cohorts in the 20s. The employment rates for young men have decreased over time, while the rates of female employment rose from 46% in 1972 to 77% in 2015.

There's also been substantial increase in the female educational attainment, we have women catching up and even overtaking men.

Here you can see the raw and covariate adjusted female coefficients on log hourly pay where anything below zero indicates female penalty. For all in black, higher education graduates in dark grey, and non-graduates in light grey for each cohort. The black estimates for all showed that female penalty was quite large in 1972 and is smaller for more recent cohort, even though the adjustment doesn't make that much difference. However, we can also see opposite trends amongst graduates and non-graduates where the female penalty for graduates, so the light grey estimates over time gets closer to zero, but the darker grey penalty for graduates actually increases. And this is happening,

bearing in mind, during the time when more women attained higher education degrees.

While we were working on this project, we also experienced the COVID-19 pandemic and at this time the Centre for Longitudinal Studies issued a special set of data collections across all cohorts that took place in three waves between March 2020 and March 2021. And we also took this opportunity to investigate how this was related to gender inequality and employment rates and earnings.

In the first paper, we focused on the employment rates, mainly those who stopped working due to furlough or other reasons and this graph shows the likelihood on being on paid leave, including furlough for men and women with different family arrangements. So, we have women who live with partner and children, women with partner but no children, single women and lone mothers. And then men with partner but no children and single men. And our reference category are men with partner and children.

Here I also split this into two panels, those who worked in occupations considered as essential, so the key workers, and those who didn't. And the blue dots and whiskers show the raw gaps, then we adjust this for some basic characteristics shown in red and for the job characteristics as shown in green.

As you can see, the two bottom rows do not show any significant differences, meaning that for men, men were equally likely to be furloughed regardless of whether they have partner or children. However, women in all types of families were more likely to be furloughed if they were not key workers. And this is also true for mothers who were key workers, even if they had a partner.

In the subsequent paper, which is still in progress, we focus on those who were employed in March 2020 and remain employed in March



2021. The major change experience during this time was the increase of working from home and hybrid work. So, we look at the differences in pay across the work locations. This graph shows there are differences in pay between men in red and woman in green where the size of the dot reflects the size of the sample in our data. Here we showed pooled estimates from across the three Cohorts, Next Steps, NCDS and BCS. As you can see, the average pay of men is higher than of women across all work locations, however, the pay of those who work from employers' premises is generally lower than that of those who work from home or hybrid. And the pay gaps are particularly wide amongst those who work from home for at least some of the time.

[0:34:30]

In our paper, we also adjusted differences for a number of covariates and since we're interested in whether these differences emerged since the outbreak of COVID or whether these patterns were present before, we also adjust for pay before the pandemic. And this analysis showed that despite remote work being relatively uncommon before the pandemic, average pay of men and women working from different location is largely explained by the pay prior to the outbreak, which implies that selection into working from home occurred in line with pre-existing inequalities.

And finally, since the pandemic, our centre has also been working on the new sweeps of data collection from the Cohorts, the data most recently released last month is from Next Steps cohort at age 32. And we used this data to look into some initial findings in terms of the gender pay gap in this cohort at this age. And this graph shows employment rates for men and women who had and didn't have children. As you can see, the employment rates are generally quite high in this cohort, however, mothers are the only group that really stands out as compared to men and women with no children, as well

as fathers. They're more likely to work part-time, as shown in light pink, and to look after family as their main activity, as shown in dark green.

So, we look at the differences in pay across those groups, bearing in mind that we already exclude quite a large proportion of mothers that are not in paid work. The raw estimates here are shown in blue, which we subsequently adjust for a number of individual characteristics shown in red and as well as job characteristics as shown in green. And here the reference category are men with no children.

As you can see, the pay of women who do not have children by this age is between 6% 8% lower on average, depending on the adjustment. The pay of mothers, however, is over 20% lower. The gaps close a little bit with the adjustment for individual characteristics, but open up again once we account for job characteristics. And although the pay of fathers is a little bit lower, these differences are not statistically significant once we take into consideration their characteristics. So, the residual gap here is what we would, that is especially large for mothers is something that we would like to explain. And the fact that we're not explaining it very well taking into consideration a large range of individual and job characteristics including here self-perceived job quality, points to the importance of other factors that we don't control for such as perhaps preferences, social norms, or discrimination.

So, that was a very short summary of the key findings from our project and there is a lot more information available on the project website.

Thank you very much for your attention and please let me know if you have any comments or questions.

[End of Transcript]