UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS
UNECE Expert Meeting on Statistical Data Collection – Towards a New Normal?
26 to 28 October 2022, Rome, Italy

Session 3
28 October 2022

Understanding data collection quality, inclusivity and representativeness at source

Dr Karina Williams (ONS)
Methods.Research@ons.gov.uk

Abstract

In the Methodology and Quality directorate, at the Office for National Statistics, we aim to optimise the collection of data to better inform our society through producing statistics for public good. We will present our approaches on:

1 Exploring inclusivity and representativeness in administrative data for our statistical purposes.
2 Understanding administrative data quality for our statistical purposes at the start of the data journey.
3 Whether we can gain greater insight and understanding of administrative data quality for our statistical purposes using qualitative approaches.

We are carrying out innovative research on, and placing importance in, collecting, and assessing administrative data input quality, inclusivity, and representativeness at source for our statistical purposes.

We are exploring inclusivity and representativeness from group representatives (as gatekeepers) and directly with the public. We conduct qualitative interviews to gain in-depth understanding in how specific population groups interact with services which contribute to administrative data. This gains insight in how inclusive and representative these sources are for our statistical purposes.

Assessing quality further along the data journey, we are conducting research to understand quality of specific administrative data from the perspective of administrative staff that collate and process the data.

Products, for use across statistical organisations and wider, from our research programme include: developing tools and frameworks to aid assessment of administrative data quality and to assist conversations with data suppliers to help improve the quality for our statistical purposes.

Background

In the Methodology and Quality directorate, at the Office for National Statistics (ONS), we aim to optimise the collection of data to better inform our society through producing statistics

for public good. We are carrying out research exploring possible options of the type of data we can use in ONS and how we can use it.

Within the UK, many organisations collect a large amount of potentially useful information during standard operations. This information can be considered a valuable resource to analysts if used securely and in an agreed and informed way, as it will typically reduce the time and costs associated with survey data collection and is called “administrative data”.

In ONS we want to understand more the quality of these administrative data at source for our statistical purposes. As part of understanding the data quality we also want to understand how inclusive and representative this data is at source too for our statistical purposes. The reason for this is to understand what methods, outputs or additional sources are needed to transparently improve and communicate quality, inclusivity and representativeness in the statistics ONS produce that use these data.

The research in this paper contributes towards the UK’s National Statistician’s 2023 recommendations and the UK’s National Statistician’s Inclusive Data Taskforce (IDTF) implementation plan. In 2023, the National Statistician will deliver recommendations to Government on the future of the population, migration and social statistics system in England and Wales. Regarding the National Statistician’s Inclusive Data Taskforce (IDTF) implementation plan: In October 2020 the UK’s National Statistician established the IDTF to ensure inclusivity in UK data in a broad range of areas including protected characteristics, areas associated with sustainable development goals and equalities. The taskforce identified 46 recommendations, aligned to 8 inclusive data principles, which are required to ensure UK data and evidence is inclusive.

Following publication of the IDTF recommendations report, ONS have led the production of the IDTF implementation plan which draws together the varied, complex and far reaching workstreams required across the UK statistical system to deliver these recommendations. This includes the research and analysis needed across government to assess whether the data we use in our statistics are inclusive.

The research presented within this paper mainly falls within principle 6 to ‘broaden the range of methods that are routinely used and create new approaches to understanding experiences across the population of the UK’:

> *“ONS is researching the coverage of specific administrative datasets to better understand how certain groups within the population are represented. Qualitative research methods are also being developed to give us a greater insight into any inclusivity issues for such data sources.”(Inclusive Data Taskforce implementation plan, 2022)*

The IDTF (ONS, 2021) highlighted that for administrative data to be used effectively and responsibly for research purposes, analysts require an understanding of the extent to which it can be considered inclusive and representative. Throughout this paper we will refer to both inclusion and representation as concepts. In ONS these concepts are defined as:

- Inclusivity - the extent to which groups or individuals are included in administrative data. An example of a lack of inclusivity would be members of a group / groups not being present on administrative data.
- Representativeness - the extent to which administrative data reflect groups or individuals' characteristics. Examples of a lack of representation would be (a) being present in administrative data without having characteristics recorded and/or (b) being present on administrative data, but members of the group / groups or public being classified differently to how they should be recorded.

What we are doing in Methodology and Quality at ONS is a slightly different approach in how we try to understand quality, inclusivity and representativeness in administrative data for our statistical purposes. We have firstly adopted an innovative approach by looking at data quality during the input stage/ at source. We gain this intelligence by collecting information from members of the public about how they enter their information that we later use as administrative data (we do not collect their actual data in this research). We also collect information from clerical staff that process the data at the input stage to understand more about what decisions and rules are made.

We are also innovative in our choice of methods. The aim is to gain greater depth and insight for our statistical purposes in administrative data quality using a qualitative approach. Research assessing quality in administrative data typically uses quantitative approaches. We are hoping a qualitative approach will provide deeper insight in our understanding of administrative data quality.

Collecting this information at the input stage/ start of the data journey and using in-depth qualitative approaches, we hope, allows ONS to gain an even better understanding of quality of the data we use for statistical purposes. This understanding directly impacts ONS decision making on factors such as how ONS process the data, what methods to use, what additional sources are needed and how to communicate our statistics and findings to the public.

Our programme aims include:

1 Exploring inclusivity and representativeness in administrative data for our statistical purposes.
2 Understanding administrative data quality for our statistical purposes at the start of the data journey.
3 Investigating whether we can gain greater insight and understanding of administrative data quality for our statistical purposes using qualitative approaches.

The next section explains the suite of research we have completed or are in the process of carrying out to fulfil our aims.

Research project 1: Gaining insight from members of the public and group representatives

We are carrying out research to explore how the public interact with and complete questions and information about themselves when registering for public services such as through

government organisations. The aim is to understand more about inclusivity in the administrative data we use at ONS for our statistical purposes.

We are gaining insights by testing key questions that the public complete based on the groups we have sampled. The aim is to explore the impact of questions and the way the information is collected at source. We use a qualitative method; collected through a cognitive interviewing approach.

We ask the public whether they are likely to register their information, how they interact with services to complete their information, and how they update their information. We present example questions as elicitation aids to facilitate discussion. We do not collect their actual data: We just gain insight on how members of the public complete their information.

We are also asking and collecting similar information on group representatives to explore whether they can provide insight on how the community they represent completes their information.

We are in the process of reporting on the findings of this research.

Research project 2: Exploring quality at the input /data entry stage - administrative data case studies

In ONS there have been a series of tests looking into whether we can explore quality of administrative data at the data entry / data input stage focusing on sources of administrative data at the start of the data journey. One piece of research has focused on exploring the Electoral Roll data in England and Wales.

The purpose was to gain an understanding for our statistical purposes of how the data is collected, processed and had updates to Electoral Roll data over time.

The project consisted of three qualitative research strands conducted to understand the data journey of Electoral Roll information. The three strands are:

1. Research with Electoral Roll administration teams and Electoral Roll Registration Officers.
2. Research with members of the public - to assess how they complete their Electoral roll information (we do not collect their actual data).
3. Research with colleagues at ONS who securely receive some of the data for statistical purposes – to explore their processing, their methods and how they communicate with data suppliers on data quality for our statistical purposes.

From these findings we have created a conversation tool kit. This toolkit comprises of questions that can be asked to data suppliers (more information on this is presented in the next section). We will also aim to create an end-to-end data journey map with the Electoral roll as the first case study. Other administrative data case studies would be introduced to this map when explored over time.

It is our goal to explore other administrative data sources using a similar approach in the future. Each will provide us with case study examples on how to collate and process the data to produce statistics.

Research project 3: Development of toolkits to help understand administrative data quality for statistical purposes

There have been toolkits developed as outputs/products to the research projects presented within this paper.

From the Electoral Roll research, we have developed a conversation toolkit which provides questions that analysts can use as aids when communicating with data suppliers to understand more about the data quality at source for their statistical purposes. This will assist analysts to understand more about what methods they need in the processing of data for statistical outputs. We have begun development of this toolkit by designing the questions, conducting an internal expert review of the question designs, applying feedback, as well as carrying out cognitive testing on some of these questions - using the Electoral Roll research as the case study.

We have organised the question bank into themes. The themes incorporate:

- the Data Management Association (DAMA) quality dimensions
- the stages of the data journey

In line with the above, our draft bank currently contains the following themes:

- completeness
- accuracy
- timeliness
- validity
- uniqueness
- content, coverage and purpose
- data collection
- data processing (includes questions on data linkage and 'systems and people')

Other aspects we would look to include are:

- accessibility
- consistency

We are continuing to develop and cognitively test the conversation toolkit with different administrative data case study examples in the future. We are building guidance on how to use the toolkit, guidance on what information is needed from data suppliers including what mode the data was collected in. We are also going to provide guidance on how to carry out desk-based research on the administrative data to help understand quality for statistical purposes.

We have developed an administrative data quality framework that is specifically tailored to the needs of users, and that aims to walk them through core quality assessment of administrative data for statistical purposes in an accessible way. This will help government organisations working with administrative data to make informed decisions about their administrative data quality and resulting use for their statistical purposes. Based around user needs and 'fitness for purpose', the framework has been shaped by user feedback at all stages of development. It aims to draw together much of the existing and valuable guidance in this area, collating useful tools and recommendations into one, user-friendly resource.

The framework is organised around two main phases, input quality and output quality, so that it is broadly consistent with existing frameworks and because this was a user need.

- The input quality phase focuses on the data you are using: how suitable is it for what you want to do with it?
- The output quality phase focuses on the statistics or analysis you have produced: how well does it meet you and your users' needs?

Each of these phases contains:

1. A description of each quality dimension (eg accuracy, relevance).
2. Context around what issues you might face with each quality dimension in an administrative data context for statistical purposes.
3. Questions and tools to help you decide the relative importance of each quality dimension for your statistical purpose.
4. First steps to think about when assessing the data/output against each quality dimension, with further links to more in-depth methods and resources.

The purpose of the framework is not to answer all the questions for you, but to point you towards what you should be thinking about (and doing) to understand whether your data and/or your output is fit for use for statistical purposes.

We have received feedback from different government departments who have used their research as a case study example in our framework. We have completed prototype publications on this framework and look to disseminate a full publication this Autumn/Winter 2022/2023. In the future we look to add more case study examples of this use. We are also linking the conversation toolkit to our administrative data quality framework

There are quantitative methods we have developed to assess and measure quality in administrative data for our statistical purposes through measurement error, using latent class modelling approaches. We are publishing papers on these research methods this Autumn/Winter 2022/2023. This also includes aiming to publish a catalogue of quantitative methods to measure administrative data quality.

Next steps

1. Carry out more inclusivity and representativeness research using qualitative techniques with members of the public to understand more about administrative data quality for our statistical purposes.
    - Feedback findings from this research to the 2023 recommendations.
    - Share learning with colleagues across the statistical system to progress implementation of the IDTF recommendations.
2. More in depth qualitative research looking at the administrative data journey with different administrative data as case study examples to understand more about administrative data quality for our statistical purposes.
    - Feedback findings from this research to the 2023 recommendations.
    - Share learning with colleagues across the statistical system to progress implementation of the IDTF recommendations.
3. Apply what we have learnt from our research to further develop and test our toolkits.

4. Assess how best to integrate our qualitative research findings with our quantitative research development.
5. Continue to develop quantitative research methods to assess administrative data quality for statistical purposes.
6. Develop standardised qualitative methods for assessing administrative data quality for statistical purposes.

Key outcomes from presenting to the UNECE expert group meeting

There are several aspects we would like to get feedback on from presenting at the UNECE expert group meeting on data collection:

1. Is this type of programme of research something you are interested in carrying out similarly at your institute / organisation?
    a. If it is, which aspects of it?
2. Would you like to collaborate and/or keep in touch going forwards?
3. If you have carried out something similar how do your research methods compare?
4. Is there any feedback you would like to give regarding the methods and pieces of research we have carried out and presented on?
5. At ONS I'm setting up a working group on this topic with similar methods: would you like to be part of this?

Conclusion

Qualitative methods to understanding administrative data quality for statistical purposes does add greater insight and context alongside assessing quality through usual quantitative approaches. Qualitative approaches to assessing administrative data quality can provide further insights on both how to ensure and communicate quality when producing statistics.

There are some new approaches to quantitative research that we are also wanting to explore as mentioned in this report and are in the process of publishing these pieces of research.

As we are at the analysis stage for our inclusivity and representative research we will be reporting on the findings in relation to the impact on statistical processes in due course. We will be presenting on these findings across ONS and across government and we will be continuing to drive IDTF implementations.

We are continuing to carry out innovative research on, and placing importance in, collecting and assessing administrative data input quality, inclusivity, and representativeness at source to understand administrative data quality at the start of the data journey for our statistical purposes at ONS.