Diversity in Data - Ethnicity coding working group

Workshop Report – 6th April 2022

On 6th April 2022, we held our second Diversity in Data - Ethnicity coding working group session, chaired by Ashley Akbari and Kamlesh Khunti. The session provided an opportunity for stakeholders working in this area to share their research and intelligence experiences, with relevant talks from Rose Drummond & Cameron Razieh (Office for National Statistics) and Sara Khalid (University of Oxford). About 40 participants attended the meeting, with people from research and custodian communities, along with members of the public. An agenda can be found in Appendix.

Existing work, ongoing challenges and opportunities were discussed around ethnicity data, with suggestions for possible future priorities to be taken forward by this group.

During the interactive breakout sessions, participants covered the following topics:

- Data Collection.
- Training.
- Terminology.
- Building pubic trust.

Below we summarise the main points highlighted during the meeting and associated discussion sessions.

Recording and capturing information about ethnicity

We focussed on how ethnicity is currently captured and how information obtained from different data sources can vary. The main discussion points included:

- Classification can be context specific - The setting in which the data is being collected could result in the questions being answered very differently (cultural, health care, education, financial settings).
- Data capture can vary depending on the computer systems / coding dictionaries and rules available.
- Method of data collection is important – by practitioner or self-report. There can be issues with both. If the data collector does not feel comfortable asking for this information, they may either guess the person's ethnicity, or they may not report this information. If self-reported, a person may choose to opt-out of providing this information for various reasons.
- A person may change their declared ethnicity over time, and the system that captures it may also change the options for recording it over time.
- There might be generational differences in the interpretation of specific terms used when referring to ethnicity.

- As a society are we moving towards self-identification? Many people don’t want to fit neatly into a ‘box’.
- We are still missing ethnicity information for many people, and we don’t have a framework for how it is collected systematically – this framework is highly needed if we don’t want to see high variation in ethnicity data.
- We need to think about how ethnicity can be affected by migration. Where someone was born, where they grew up and where they live now. e.g., how does this influence the way people report information about themselves?
- We need to consider that people might feel uncomfortable being asked about ethnicity too often.
- Researchers tend to just want as much data as possible at all times – what really is there a benefit of capturing it at every opportunity? How often is ‘good enough’ to support services and projects without becoming a burden which impacts the provision of care and services?

Ideas on how capture of ethnicity data could be improved

- **Training and support for data collectors**
  - Improving the conversation between data collector and participant. How we ask patients for this information is important – We should consider how to make patient feel at ease so they feel comfortable providing this information. This could lead to better quality data.
  - Data collectors could offer support when patients are filling this information. Helping the public to understand the categorisations could be useful. Some people may not know their ethnicity or may feel restricted by the categories we currently use.
  - Honouring the right for patients to opt-out of providing their data – How is this recorded? This is not ‘missing data’ or ‘Unknown’, and should be recognised as a specific category when collecting.
  - A framework to define how to collect ethnicity data in a standardised way would be helpful for clinical practice and research– Address different ways of collecting and reporting ethnicity data across all settings.
  - Ensure the data collector understands the importance of capturing ethnicity data. If they don’t see the importance of it, they may just decide not to collect it.
- **Training and support for patients and the public**
  - There is a need for a wider education activities to help people to understand key differences.
  - It is important to explain to the public, the value of collecting ethnicity data for research.
  - There have been examples of data being used towards deliverables the public was not aware of, or may not have been informed enough about which can lead to public concern and mistrust. Transparency and continuous engagement is important in restoring and maintaining ongoing public trust.
- **Mapping** - Determine how codes are defined and applied. This could be improved by using an Index which describes what each group is.
- **Framework** – to improve the collection of Ethnicity data.

How simplification/aggregation of ethnicity capture in groups could lead to under or over representation

Some individuals in the group felt the ethnicity categories used by the NHS were too broad. For instance it was noted that some ethnic groups are underrepresented (eg: South Asian), which can cause heath inequalities. Main discussion points are highlighted below:

- We need to consider where we can simplify and aggregate data and where we might need to keep the granular detailed information.
- Simplification is great from a national statistics perspective, as grouping people down into a smaller number of categories makes it easier for the data to be analysed. However, this can do a disservice when you start looking at specific groups.
- Sometimes breaking data down too far means that the data is not as useful as there may be too many small categories.
- We need to be able to flex the level of aggregation to suit different research needs – granular data should be available but can also offer more high-level simplified data if suitable.

Clarifying the difference between ethnicity and race

- We agreed the distinction between ethnicity and race is not clear. There is also some confusion as to what we mean by using terms such as Lineage, Nationality, Country of birth, Family origins, migrant status, language, religion etc. People might interpret this terminology differently.
- Confusion likely comes from the side of ethnicity as a social construct.
- Several people looked up definitions of ethnicity and race – race won't change, ethnicity may - but race may change *given context.*
- Updated definitions are required.

Actions

The group highlighted a number of possible areas of work:

1. Improve standardisation of ethnicity data - Draft a **framework** outlining the main recommendations around definitions and terminology, to discuss at next meeting.
2. There are many coding standards. **Mapping** will help to provide a guide to what codes to use when analysing data. Index of coding is needed to define codes. Resources can be shared via HDR UK Futures.
3. **Survey** public groups around reporting ethnicity coding information. Some work has recently been carried out in this area. Some newly published research from Understanding Patient Data "New research: insights from Black & South Asian people on patient data"(full report here). Could we collaborate on this work?
4. **Training** material – We could create a dedicated space on HDR Futures? Some videos are already in development. We could link to other resources.
5. Practitioners - **guidance** on how to communicate and explain why that data is needed. Importance of capturing that information.

We will hold at least one more working session to focus on specific aspects of working with ethnicity data.

The next ethnicity coding working group session will take place in September 2022. The exact date is to be decided. Invites will be sent out in due-course.

Appendix

Agenda

Time	Item	Lead and Slides
1pm	Welcome and introductions	Chairs
1.10pm	Assessing the quality of ethnicity data in health admin sources in England	Rose Drummond & Cameron Razieh (Office for National Statistics)
1.20pm	**Exploring Ethnicity Data in the NHS Digital TRE**	Sara Khalid (Oxford)
1.30pm	**Q&A and Intro to breakout**	Chairs - All
1.40pm	**Breakout session:** • **the need of harmonising terminology and definitions when working with ethnicity data** • **standards for data collection**	Facilitators
2.25pm	**Discussion**	Chairs
2.45pm	**AOB and Close**	

Outputs

Video Recording of the session

Assessing the quality of ethnicity data in health admin sources in England – Presented by Rose Drummond & Cameron Razieh (Office for National Statistics)