

Nuffield Trust and NHS Race and Health Observatory workshop on ethnicity coding

Introduction

The Nuffield Trust and NHS Race and Health Observatory (RHO) held a workshop on 30 March 2021, as part of our project on the quality of ethnicity coding.

The workshop involved participants from the Office for National Statistics (ONS), Public Health England (PHE), NHS Digital, the Race Disparity Unit, NHS England and Improvement, Department of Health and Social Care, the Strategy Unit, and the Wellcome Trust.

The aims of the workshop which were to:

- Share findings from the current [Nuffield Trust analysis of the quality of ethnicity coding in NHS data sets](#), and hear about ongoing work on the quality of ethnicity coding by the Office of National Statistics and Public Health England
- Agree ways to facilitate an understanding of how existing data can be used, it's limitations notwithstanding, and how a consistent analytical approach can be adopted across organisations
- Identify actions needed to improve the quality of ethnicity coding in health care

Analysis work on ethnicity coding

At the workshop we heard about work from the Nuffield Trust, Public Health England (PHE) and the Office for National Statistics (ONS).

- The Nuffield Trust presented analysis of the completeness, validity and consistency of coding, in order to identify issues which users of data needed to be aware of, and problems with data quality which need to be addressed. The Nuffield Trust and NHS RHO will be reporting on this work in May 2021.
- PHE presented the methodology they are planning to use to assign ethnicity in public health data analysis, to address problems with ethnicity coding, and improve the face validity of outputs. They are planning to implement this methodology during May.
- ONS presented the results of analysis linking the 2011 census data on ethnicity to HES data from 2011 and 2019. They identified challenges in linking data due to gaps in coverage (for example non-response in the census) and false negatives in the linkage (due to HES data quality). Although linkage has improved over time due to improvements in HES data quality, there are biases introduced by the linkage process, because coverage and false negatives affect different groups in the population to differing degrees. ONS do not currently plan to publish this work.

There were a number of common themes from the analysis presented. In particular, all the projects found that the impact of coding issues had a disproportionate impact on data

quality for ethnic minority patients. This reinforces the importance of addressing data quality issues at source, and during analysis.

In addition to the presentations, participants also contributed details of other relevant work underway:

- ONS noted that the method to assign ethnicity to death certificates was still under discussion.
- NHS England recognised that urgent action was needed to improve the quality of data. Actions were planned to improve the data in GP records, and a *Unified Information Standard for Protected Characteristics* was being established to support data quality improvements.
- NHS Digital are undertaking work to develop more complete ethnicity coding, using records from primary care (GPES) as well as hospital care
- Understanding Patient Data are undertaking work to understand patient and staff's [perspectives on recording of ethnicity](#)

Discussion

The discussion focused on two related areas:

1. Developing a consistent approach to analysis, given limitations of the data
2. Actions to improve data quality (given lack of time, this question received relatively less discussion)

Within the discussion on developing a consistent approach to analysis, workshop participants focused on the following themes:

- Assessing validity: this is a challenge, given that the most credible gold standard, the census, is now 10 years out of date. While reassigning ethnic groups can improve the face validity of data (for example mortality rates), this is not the same as validation. What would standards for best practice in validation look like?
- Access to datasets: different organisations have access to different datasets – and in particular most NHS organisations don't have access to GP data, and only ONS has access to the census. Could ONS provide a validation service for others?
- Specific ethnic group analyses: most of the analysis so far as been at the broad ethnic group level. Given the challenges with "other" categories and the specific health needs of different ethnic groups (eg Indian v Bangladeshi people), more focus is needed on specific groups
- Ethical considerations: is the principle of self-certification of ethnicity is compromised if the ethnicity of records is reassigned during analysis? Factors to consider here include: the purpose of the analysis (eg population health v individual treatment); whether the individual has expressed a wish not to state ethnicity v whether this is unknown

- Comparison of alternative approaches to reassigning data: ONS, PHE and NHS Digital are adopting different approaches, eg ONS make a distinction between ethnic group being reported as unknown v not-stated, and some methods use the most recent ethnic code for an individual whereas others use the modal code.
- Transparency: transparency in report the approach to addressing data quality, including publishing code, can help to address both validation aspects and ethical concerns.

Regarding improving data quality at source, the following key points were made:

- Action is needed to address data quality at the point of collection – by the time it gets to HES it is too late
- There is variation between organisations and systematic biases – this highlights the potential to learn from organisations which are achieving better data quality, and to improve guidance and address systematic factors which impact on data quality
- The NHS is not using the most recent ONS ethnicity codes. ONS has a role to support harmonisation of data, but action is needed to address the different data collection requirements in the NHS
- The last guidance on ethnicity coding issued to NHS trusts dated back to 2001. Further, it is unclear whether GPs had ever been given guidance on how ethnicity should be requested and recorded, and whether GPES data had been cross-validated (as with HES).

Next steps

The Nuffield Trust and Race and Health Observatory will draw on the discussion at the workshop in formulating recommendations from their analysis. We will also contact participants later in the year with a view to convening a further workshop to discuss progress on the issues discussed, and potential for further action.