Diversity in Data
Ethnicity coding standards working group

Importance of standardising ethnicity coding – why we are here

13th January 2022

Professor Kamlesh Khunti & Ashley Akbari
Why we need good quality ethnicity data

• Measuring and understanding unwarranted variation in health and social care outcomes and determinants by ethnic group is vital [1] for:

✓ Informing clinical practice.

✓ Planning of health and social care services.

✓ Targeting and determining the impact of service improvement initiatives.

✓ Evaluating national interventions and policy changes.

✓ The allocation of research funding and resources.

• There is also a moral and legal requirement for reducing health inequalities (e.g. Health and Social Care Act 2012), as well as policy requiring ethnicity data collection (i.e. 2020 letter from NHS England and NHS Improvement).
Renewed focus on an old issue

- COVID-19 has highlighted significant ethnic inequalities in health outcomes within ethnic minority (EM) groups, and between White and EM.

However it is concerning that:

✓ Many COVID-19 clinical trials are not recording or presenting ethnicity data [2].
✓ A lack of standardised ethnicity collection and reporting has limited data pooling in meta analyses etc. [3].
✓ The ability to examine COVID-19 mortality rates by ethnic group has been hindered by the lack of recording of ethnicity on death certificates in England [1].
✓ In some countries such as the US, there was variable capture of ethnicity data by region which limited timely analyses on COVID-19 outcomes and determinants [4].

“Poor data about ethnicity has obscured the true extent of ethnic disparities in the impact of the pandemic…”

Research report June 2021

Ethnicity coding in English health service datasets

Sarah Scobie, Jonathan Spencer, Viena Raleigh

Supported by
What do we require of ethnicity data

Ideally we require ethnicity data to be:

✓ Complete i.e. limited proportion of missing entries.
✓ Useable i.e. low proportion of non-valid entries (e.g. Not known or not stated).
✓ Accurate i.e. high concordance between self-reported and HCP collected data.
✓ Consistent i.e. same grouping recorded on multiple visits, or between different data sources (e.g. primary & secondary care health records and administrative data).
✓ Granular i.e. most specific ethnic grouping possible recorded (with clear mappings to higher level aggregate groupings for standardised reporting).
✓ Comparable i.e. coding groups standardised across data sources.
Historically poor data quality

• Completeness and usability of primary care ethnicity data was historically poor, but improved after the introduction of QOF incentives in 2006 (see right) [5].

• Secondary care data has varied, but improved since late 2000’s.

• However the 2021 NHS Race observatory analysis of secondary care data [1] reported:

  ✓ Systematic biases in data quality – location, age, length of stay.
  ✓ Incomplete coding and inconsistent use of codes.
  ✓ Records of patients from minority ethnic groups were less likely to be recorded consistently over time.
Variation in collection processes

- Some literature suggests there is variation in data collection processes related to [1, 6,7]:
  - **Individual factors** e.g. patient, healthcare staff.
  - **Organisational factors** e.g. resource limitations, computer system design.
  - **Policy factors** e.g. coding standards/norms.
- For example, a comparison of self-reported ethnicity and (HES) coded ethnicity in England found that misclassification varied by a small amount between ethnic groups, but varied by a greater degree between hospitals (range: 67% to 100%) [8].
Standardised coding

- Work is ongoing by Government Statistical Service (GSS) to produce harmonised standards, based on the 18 ONS categories [9].

- However not all data sources will have capacity or user need for full standards, therefore minimum categorisation standards are required - see Khunti et al. [10]

- Some organisations (e.g. NHS) continue to use ethnic group coding based on the 2001 Census – which differs on categorisation for Chinese, Gypsy and traveller, and Arab groups – this requires addressing [11].

- We also need to look at how developed standards harmonise with those developed in other countries.
Areas to target/improve

- Staff (HCP’s, GP’s etc.) knowledge and willingness needs to be supported through standardised training and resource support.

- Cross-sector guidelines/protocols for data collection methods require developing to sit alongside harmonised standards.

- UK harmonised standards should be comparable with international standards.

- Ethnicity data reporting standards should be developed to promote transparency on the impact of data quality on interpretation of analysis/study findings.

- Mechanisms/methods developed to utilise existing data collections handling for varying levels of completeness, quality and consistency between data sources and overtime.

- Working with public and stakeholders to understand barriers/challenges in reporting & collection of ethnicity data, and key priority areas for research & intelligence.
References


