

International Journal of Population Data Science

Journal Website: www.ijpds.org



Reporting and analysing ethnicity in populational health data and linkage research: A bibliographical review

Joseph Lam¹, Robert Aldridge², Ruth Blackburn², and Katie Harron¹

¹UCL Great Ormond Street Institute of Child Health, London, United Kingdom

²UCL Institute of Health Informatics, London, United Kingdom

Improved availability of population-based data via data linkage enables researchers to develop deeper insight into racial health inequities in the UK. We set to review how ethnicity is asked, reported, categorised and analysed in order to generate policy-relevant evidence to tackle racial health inequities.

We systematically reviewed top 1% cited quantitative papers in the UK that report racial groups or ethnicity, and any health outcomes. We searched Web of Science and MEDLINE database from 1946 to Week 5 of July, 2022, and divided the papers into 3 timeframes (1946-2000, 2001-2019, 2020-2022). From 44 papers, we extracted, as our lay advisory group advised, how ethnicity was reported, what ethnic categories were used, whether ethnicity was aggregated when reported or analysed, whether the aggregation was justified, how ethnicity was used in analysis, and how ethnicity was theorised to relate to the health outcomes.

Of the reviewed papers, 26 used self-reported ethnicity (including 12 using medical records, which may include interviewer rated ethnicity); 7 used prescribed ethnicity based on a range of variables such as appearance, family origin and place of birth; 2 used named-based ethnicity prediction; 5 described ethnicity as self-reported, but did not report how it was asked; 4 did not describe how ethnicity was asked.

Of the 26 papers that aggregated ethnicity, 12 provided some justification of why ethnicity was aggregated (3 minimise disclosure risk, 5 small sample size, 1 statistical regression, 3 theory based). Only 9 papers explicitly theorised the role of ethnicity in their analysis, and how it related to the relevant health outcomes. Missing, mixed or other ethnicity were treated variably across studies.

Ethnicity is a multi-dimensional construct. Researchers should communicate clearly how ethnicity is operationalised for their studies, with appropriate justification for clustering and analysis that is meaningfully theorised. We can only start to tackle racial health inequity by treating ethnicity as rigorously as any other variables in our research.

