http://mc.manuscriptcentral.com/jph

The need for improved collection and coding of ethnicity in health research

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

CORRESPONDENCE

The need for improved collection and coding of ethnicity in health research

Kamlesh Khunti, Professor of Primary Care Diabetes & Vascular Medicine[1]

Ash Routen, Postdoctoral Research Associate[1]

Amitava Banerjee, Associate Professor in Clinical Data Science and Honorary Consultant Cardiologist [2]

Manish Pareek, Associate Clinical Professor in Infectious Diseases[3]

[1] Diabetes Research Centre, Leicester General Hospital, University of Leicester, Leicester, UK

[2]Institute of Health Informatics, University College London, UK

[3]Department of Respiratory Sciences, University of Leicester, Leicester, UK

Correspondence to:

Professor Kamlesh Khunti

Email: kk22@leicester.ac.uk

Phone: (0)116 258 4005

Word Count: 496

References: 9

Covid-19 has exposed the need for improved collection and coding of ethnicity, despite years of debate on the issue (1, 2). A review found only 7% of Covid-19 research papers and surveillance reports presented ethnicity-disaggregated data (3). Complete, detailed and accurate ethnicity data is required for clinicians, researchers and policy makers to further examine outcomes and their causes, and manage healthcare needs related to Covid-19 in ethnic minority populations.

The completeness of ethnicity data within healthcare and routine databases has been poor historically. For example, only 27% of patients in Clinical Practice Research Datalink had ethnicity recorded between 1990-2012 (4). In addition, accuracy is often low, with 20-35% error in coding of major ethnic minority groups in NHS hospital records when compared to self-reported ethnicity (5). Reasons for poor quality ethnicity data may include a lack of understanding on the importance of the data, reluctance of staff to ask for data, fears over patient reactions, and confusion about categorisation (2).

The Covid-19 evidence base suggests risk factors, infection risk, and survival differ by ethnic group. Yet the coding of ethnicity has varied between studies. A study of hospital mortality predictors used 5 ethnicity categories (White, Mixed, Asian/Asian British, Black, Other) (6), whereas Public Health England disaggregated mortality data by 16 categories (7).

In particular, the use of crude ethnicity labelling requires attention, as ethnic groups are not homogenous. The label "South Asian" (i.e. Indian sub continent ancestry), for example, is widely used in Covid-19 research despite heterogeneity in culture,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

behaviours, disease status and risk within this 'group' (1). Indeed, coronary heart disease risk factors – some of which are associated with adverse Covid-19 outcome - vary among South Asians, with differences in smoking, lipids, cholesterol, obesity etc. observed between Indians, Pakistanis and Bangladeshis (8). Disaggregation of ethnic groups is critical in all future Covid-19 and wider public health research.

An important first step in improving the use of ethnicity categorisation is to ensure labels are sufficiently granular as to capture important heterogeneity, and that they are employed systematically to decrease confusion around categorisation, and permit easier data collection and pooling. The Office for National Statistics (ONS) has published a guide on the collection and classification of ethnicity, national identity and religion (9). ONS suggest an 18-label classification system (see Box 1) to increase consistency and comparability of statistical outputs (9).

For some research scenarios this level of disaggregation is not practical (e.g. small sample size, existing databases). However it may be possible for prospective studies to collect such detailed data, but sample sizes will need to be large to detect between group differences. We suggest all research and new routine datasets should collect and report ethnicity at a minimum of five levels, and where possible use nine (Table 1). Increased consistency and comparability of ethnicity categorisation will provide clinicians, researchers and policy makers with the best possible data to inform mitigation strategies to protect minority groups from Covid-19, and improve the wider study of ethnicity and health.

Table 1.

Recommendation	Coding
Office for National Statistics (Recommended for England, there is slight variance for Wales, Scotland and Northern Ireland)	**White** 1. English/Welsh/Scottish/Northern Irish/British 2. Irish 3. Gypsy or Irish Traveller 4. Any other White background **Mixed/Multiple ethnic groups** 5. White and Black Caribbean 6. White and Black African 7. White and Asian 8. Any other Mixed/Multiple ethnic background **Asian/Asian British** 9. Indian 10. Pakistani 11. Bangladeshi 12. Chinese 13. Any other Asian background **Black/ African/Caribbean/Black British** 14. African 15. Caribbean 16. Any other Black/African/Caribbean background **Other ethnic group** 17. Arab 18. Any other ethnic group
Suggested categorisation	1. **White** (British, Irish, Any other white Background) 2. **Mixed** (White and Black Caribbean, White and Black African, White and Asian, Any other Mixed background) 3. **Indian** 4. **Pakistani** 5. **Bangladeshi** 6. **Chinese** 7. **Black Caribbean** 8. **Black African** 9. **Any other ethnic group**
Minimum categorisation	1. **White** (British, Irish, Any other white Background) 2. **Mixed** (White and Black Caribbean, White and Black African, White and Asian, Any other Mixed background) 3. **South Asian** (Indian, Pakistani, Bangladeshi) 4. **Black** (Black Caribbean, Black African, Any other Black) 5. **Any other ethnic group**

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Conflicts of interest

KK is Director of the University of Leicester Centre for Black and Minority Ethnic Health, Trustee of the South Asian Health Foundation and national lead for NIHR Applied Research Collaboration for Ethnicity and Diversity.

Acknowledgements

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1.	Bhopal R. Glossary of terms relating to ethnicity and race: for reflection and debate. Journal of Epidemiology and Community Health. 2004;58(6):441.

2.	Iqbal G, Gumber A, Johnson M, Szczepura A, Wilson S, Dunn J. Improving ethnicity data collection for health statistics in the UK. Diversity and Equality in Health and Care. 2009;6(4):267-85.

3.	Pareek M, Bangash MN, Pareek N, Pan D, Sze S, Minhas JS, et al. Ethnicity and COVID-19: an urgent public health research priority. The Lancet. 2020;395(10234):1421-2.

4.	Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. Journal of Public Health. 2013;36(4):684-92.

5.	Saunders CL, Abel GA, El Turabi A, Ahmed F, Lyratzopoulos G. Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey. BMJ Open. 2013;3(6):e002882.

6.	Williamson E, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. medRxiv.2020: 2020.05.06.20092999.

7.	PHE. Disparities in the risk and outcomes of COVID-19. London; 2020.

8.	Bhopal R, Unwin N, White M, Yallop J, Walker L, Alberti KG, et al. Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. BMJ. 1999;319(7204):215-20.

9.	ONS. Ethnic group, national identity and religion London2016 [Available from: https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicg roupnationalidentityandreligion.