Methodological considerations around the use of routinely collected data to examine health inequalities

Rohini Mathur
Centre for Statistical Methodology Symposium
November 2021

Outline

Describe

- The history, context, and meaning of ethnicity in the UK

Introduce

- Standard coding systems for ethnicity and the quality of UK ethnicity data

Consider

- Key issues when conducting ethnicity-focused research

Ethnicity vs. Race

❖Historically, ethnicity used synonymously with race - the belief that humankind is made up of biologically distinct sub-groups

❖Theories about ‘race’ used to justify imperialism, eugenics and slavery: Now discredited

❖Genetic diversity within so called races is greater than that between races.

❖This does not preclude the existence of important genetic variation in health outcomes.

United Nations Statement on Problems of Race- 1951

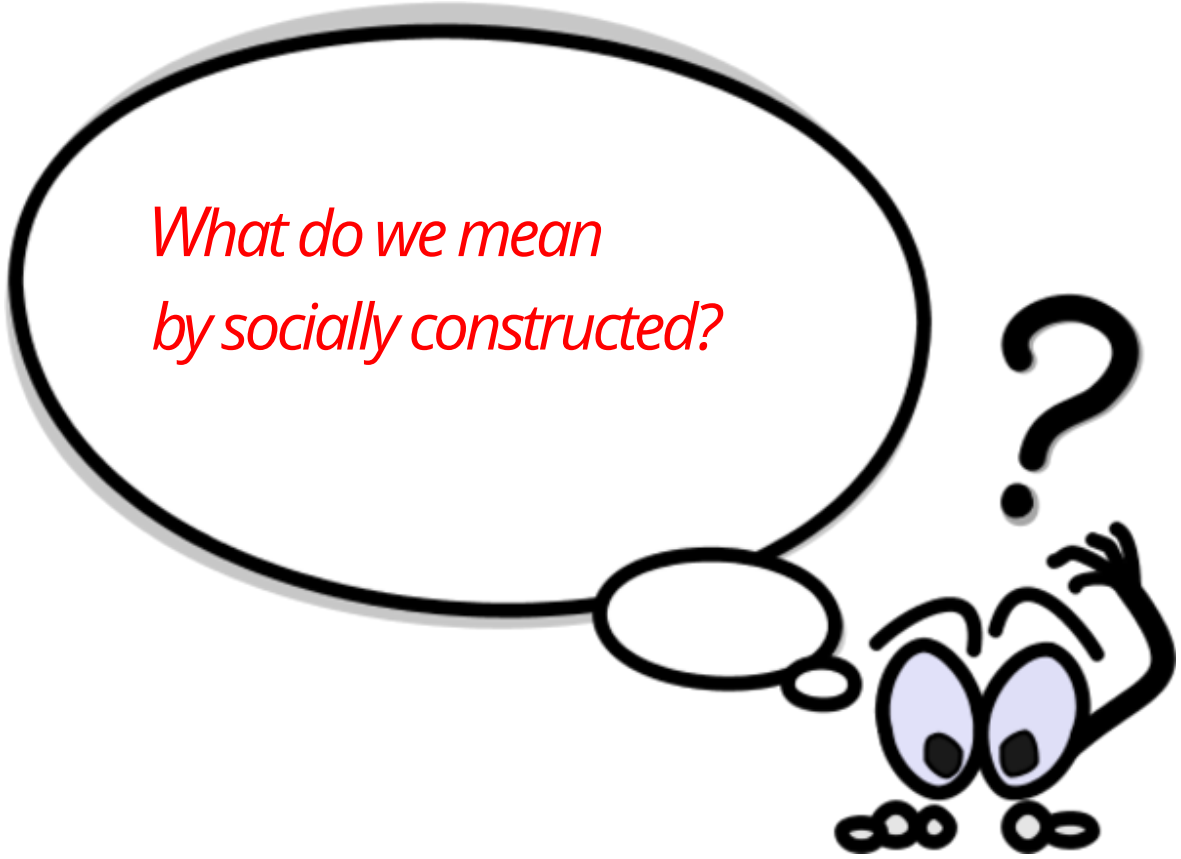
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

"National, religious, geographic, linguistic and cultural groups do not necessarily coincide with racial groups: and the cultural traits of such groups have no demonstrated genetic connection with racial traits. Because serious errors of this kind are habitually committed when the term 'race' is used in popular parlance,

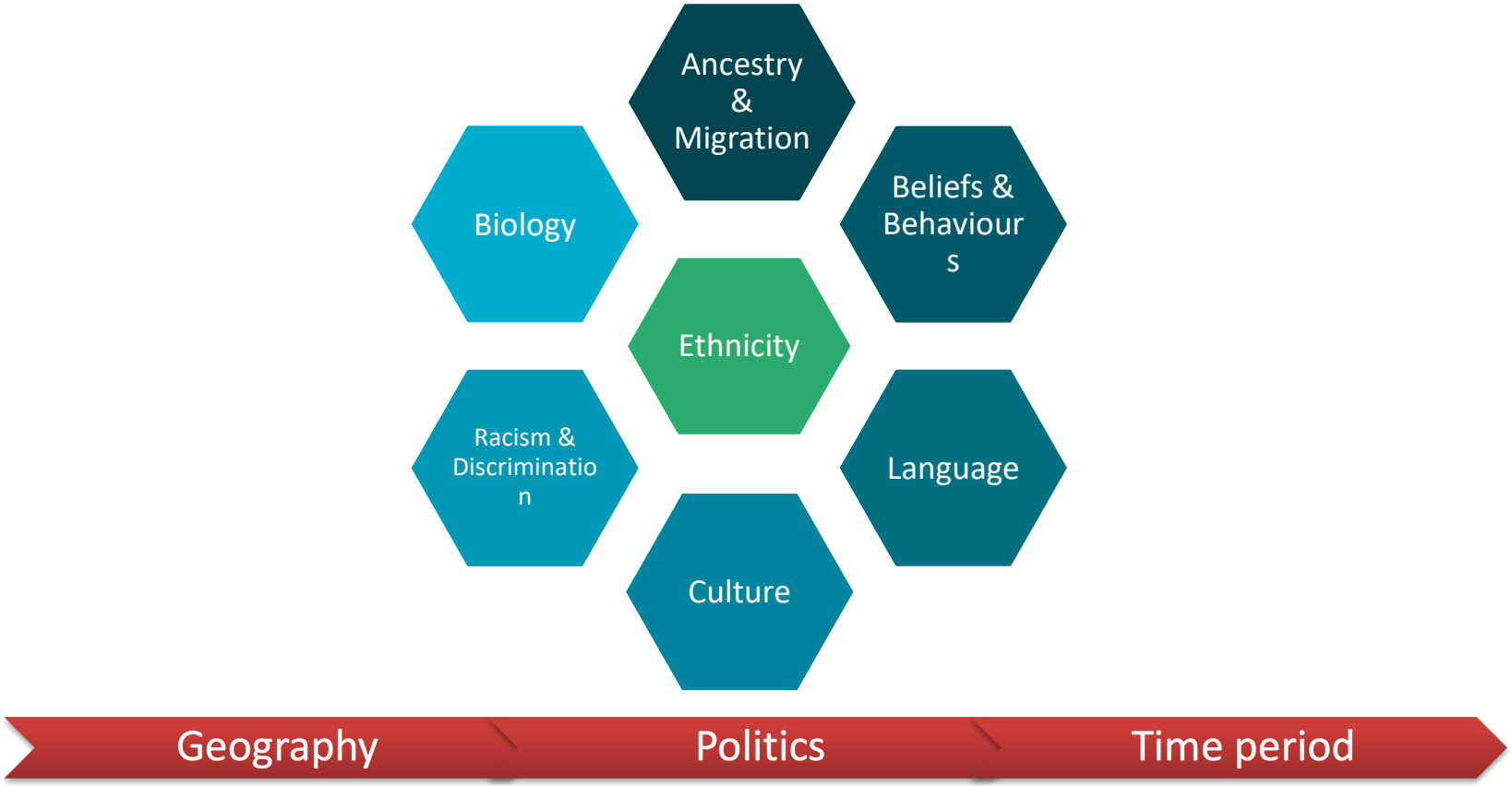
it would be better when speaking of human races to drop the term 'race' altogether and speak of 'ethnic groups'.

What is ethnicity?

Ethnicity reflects an individual's own ***self-identification***, which encompasses a broad range of ***socially constructed*** characteristics



What is ethnicity?

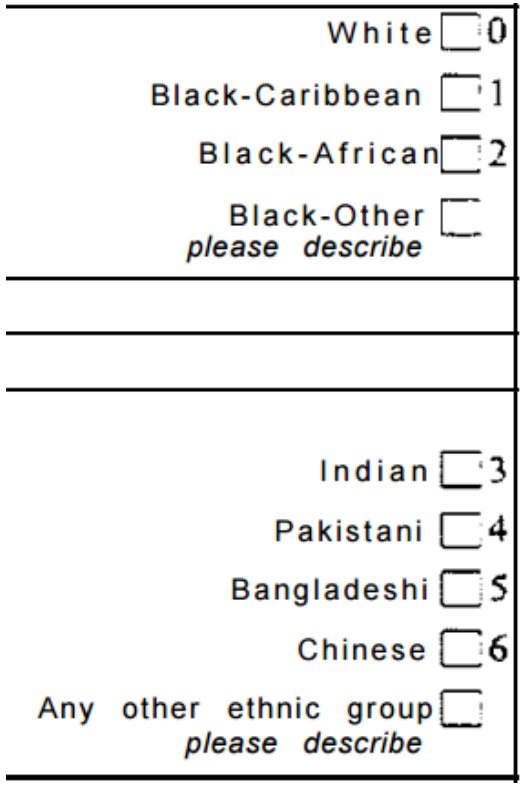


- The distribution of factors manifest unequally in different population groups, and can be conceptualized as ethnic differences.
- The relationship between ethnicity and health is complex and relationships do not always go in a single direction.

There is no one universally accepted definition of ethnicity:

- The meaning and interpretation of ethnicity is context dependent and fluid as it is inherently tied to the social, cultural, and political context in which it is used
  - In order to effectively investigate ethnicity- it must be operationalized into practical categories which are understood to encompass a rich variety of concepts
  - The categories we use are not “natural” – rather they reflect groupings which are considered relevant to each particular time period and context
  - The ethnic categories used in official statistics are, to some extent, arbitrary and have been selected primarily for pragmatic reasons

1991 | 2001 | 2011 | 2021



8 What is your ethnic group?

Choose ONE section from A to E, then ✓ the appropriate box to indicate your cultural background.

A White
- British
- Irish
- Any other White background, *please write in*

B Mixed
- White and Black Caribbean
- White and Black African
- White and Asian
- Any other Mixed background, *please write in*

C Asian or Asian British
- Indian
- Pakistani
- Bangladeshi
- Any other Asian background, *please write in*

D Black or Black British
- Caribbean
- African
- Any other Black background, *please write in*

E Chinese or other ethnic group
- Chinese
- Any other, *please write in*

16 What is your ethnic group?

Choose **one** section from A to E, then tick **one** box to best describe your ethnic group or background

A White
- English/Welsh/Scottish/Northern Irish/British
- Irish
- Gypsy or Irish Traveller
- Any other White background, write in

B Mixed/multiple ethnic groups
- White and Black Caribbean
- White and Black African
- White and Asian
- Any other Mixed/multiple ethnic background, write in

C Asian/Asian British
- Indian
- Pakistani
- Bangladeshi
- Chinese
- Any other Asian background, write in

D Black/African/Caribbean/Black British
- African
- Caribbean
- Any other Black/African/Caribbean background, write in

E Other ethnic group
- Arab
- Any other ethnic group, write in

15 What is your ethnic group?

Choose **one** section from A to E, then **tick one box** to best describe your ethnic group or background

A White
- Welsh, English, Scottish, Northern Irish or British
- Irish
- Gypsy or Irish Traveller
- Roma
- Any other White background, write in

B Mixed or Multiple ethnic groups
- White and Black Caribbean
- White and Black African
- White and Asian
- Any other Mixed or Multiple background, write in

C Asian or Asian British
- Indian
- Pakistani
- Bangladeshi
- Chinese
- Any other Asian background, write in

D Black, Black British, Caribbean or African
- Caribbean
- African background, write in below
- Any other Black, Black British or Caribbean background, write in

E Other ethnic group
- Arab
- Any other ethnic group, write in

Challenges to measuring ethnicity

Categories in epidemiological variables must be meaningful and discrete in order to be interpretable

Meaning of ethnicity is context and time specific

- Ie. The ethnic group Asian has a very different meaning in the UK compared to North America.

Quantity and definition of ethnic categories continually evolving

Ethnic categories hide underlying heterogeneity

- Ie. Within the South Asian community, diet, behaviour and religious practices can vary greatly between Indian, Pakistani and Bangladeshi communities.

Researchers must recognize the limits of ethnic categories and approach their use critically

Ethnicity recording in UK Primary and Secondary Care Data Sources

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Ethnic categorization in the UK

Primary care:
>300 Read codes
>650 SNOMED codes

Secondary Care:
16 category ethnicity only

White
- British
- Irish
- Other White

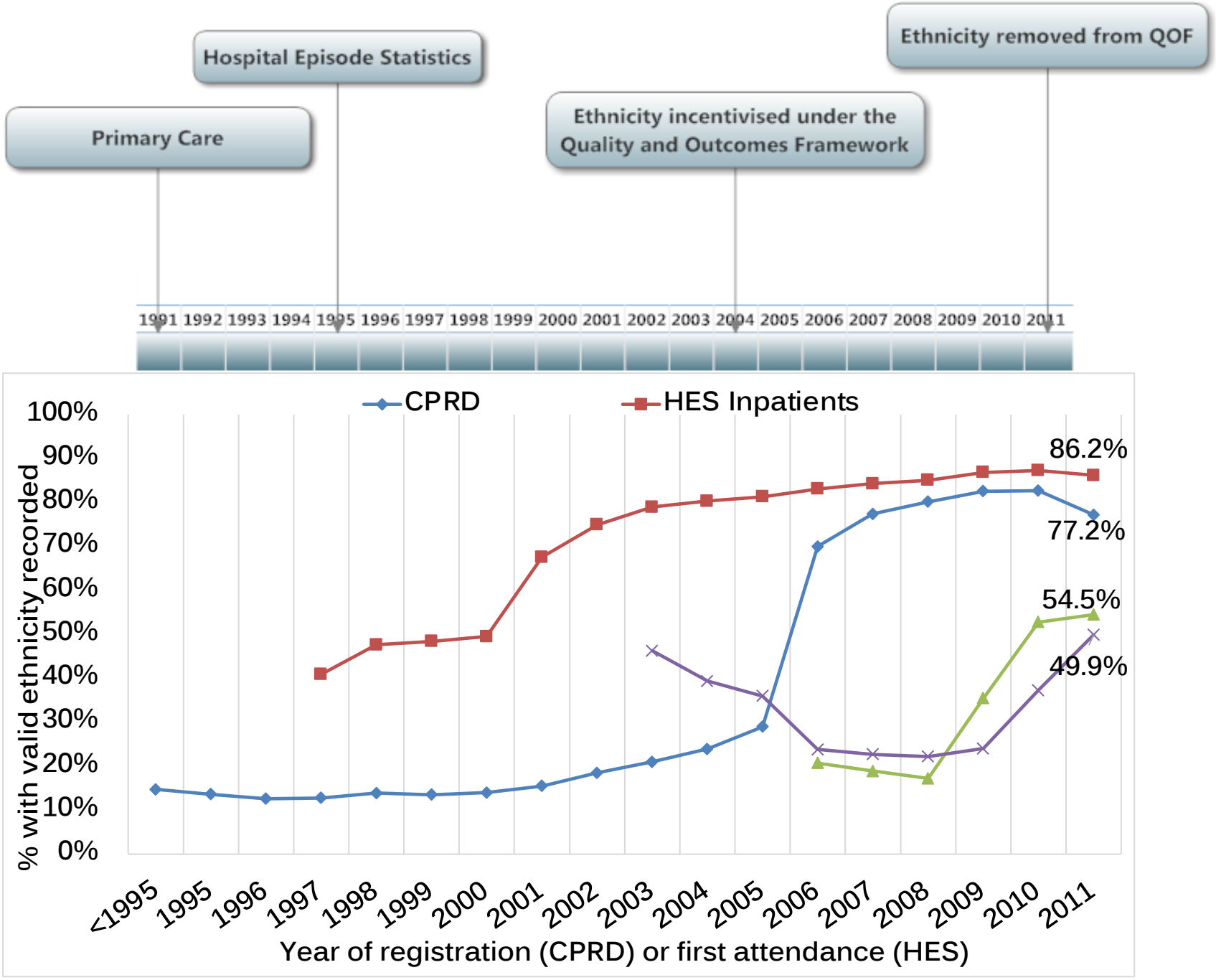
Mixed
- White + Caribbean
- White + African
- White + Asian
- Other mixed

Asian or Asian British
- Indian
- Pakistani
- Bangladeshi
- Other Asian

Black or Black British
- Caribbean
- African
- Other Black

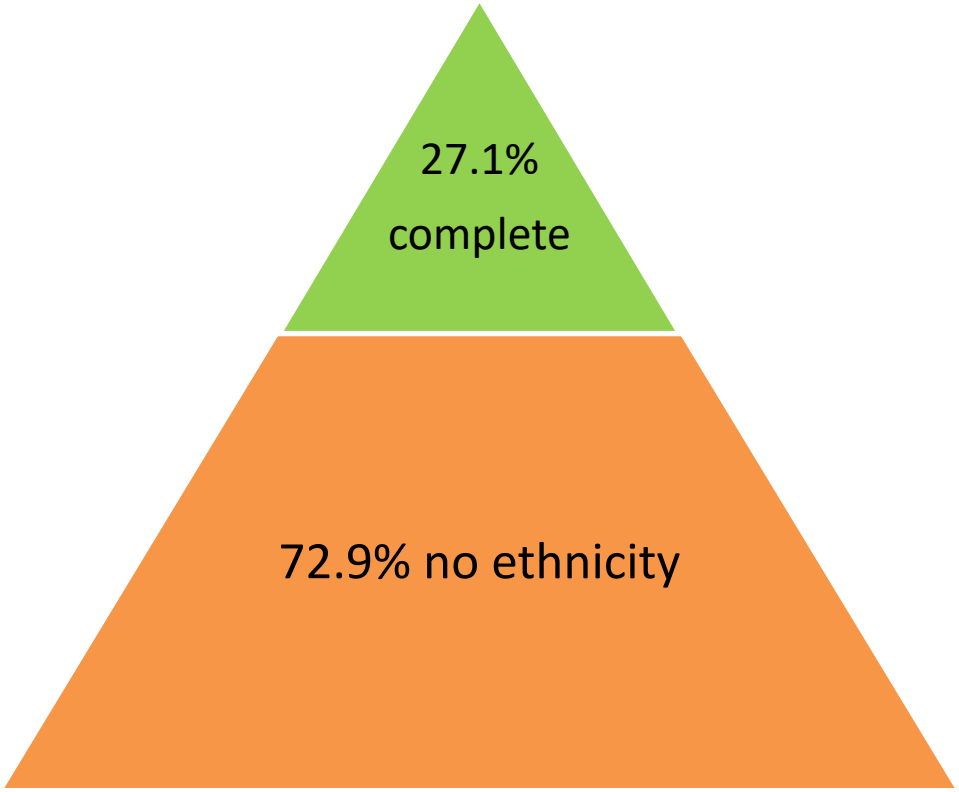
Other
- Chinese
- Other

Timeline of ethnicity recording in UK primary care (CPRD)



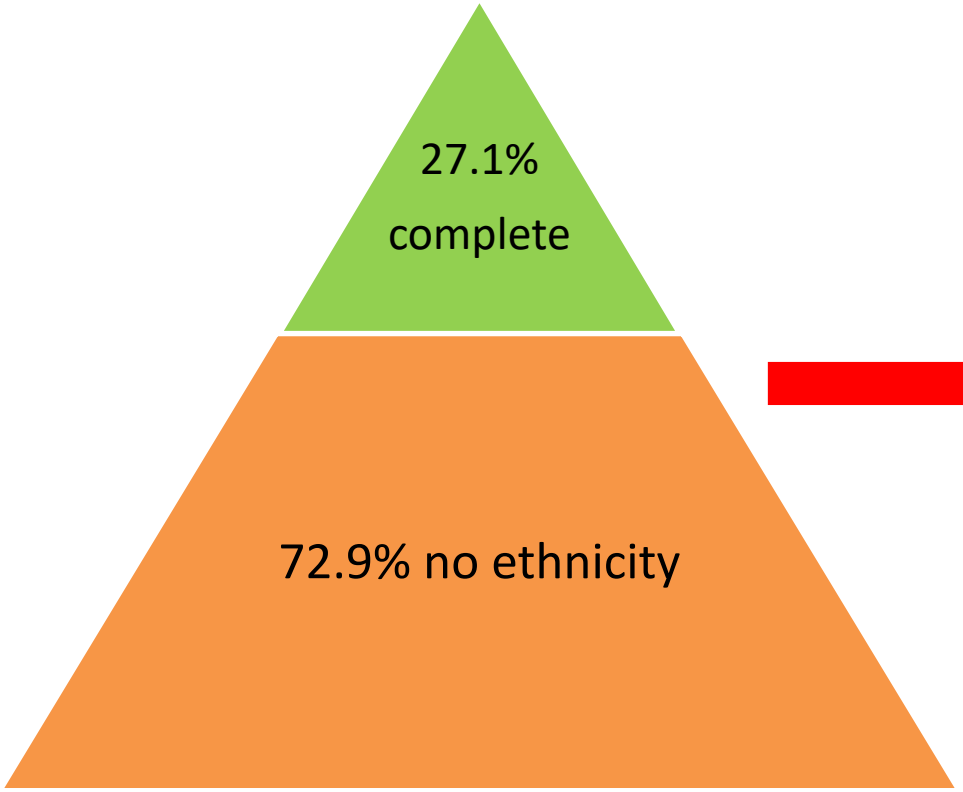
Overall completeness of ethnicity in CPRD

All patients 1990-2012
N=12,099,672

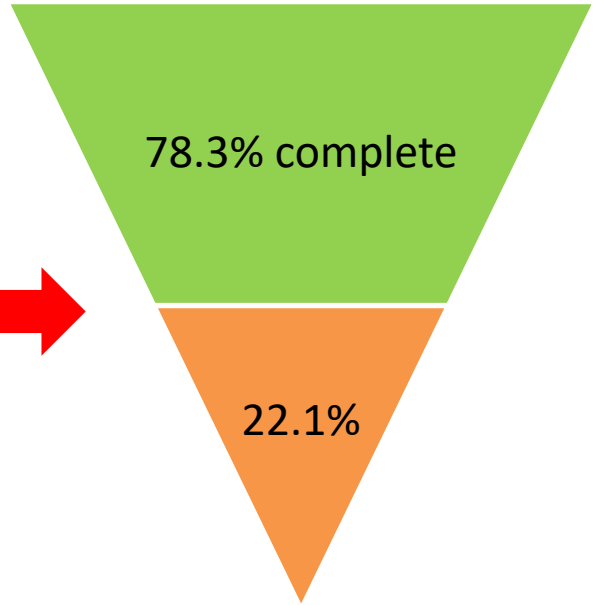


Overall completeness of ethnicity in CPRD

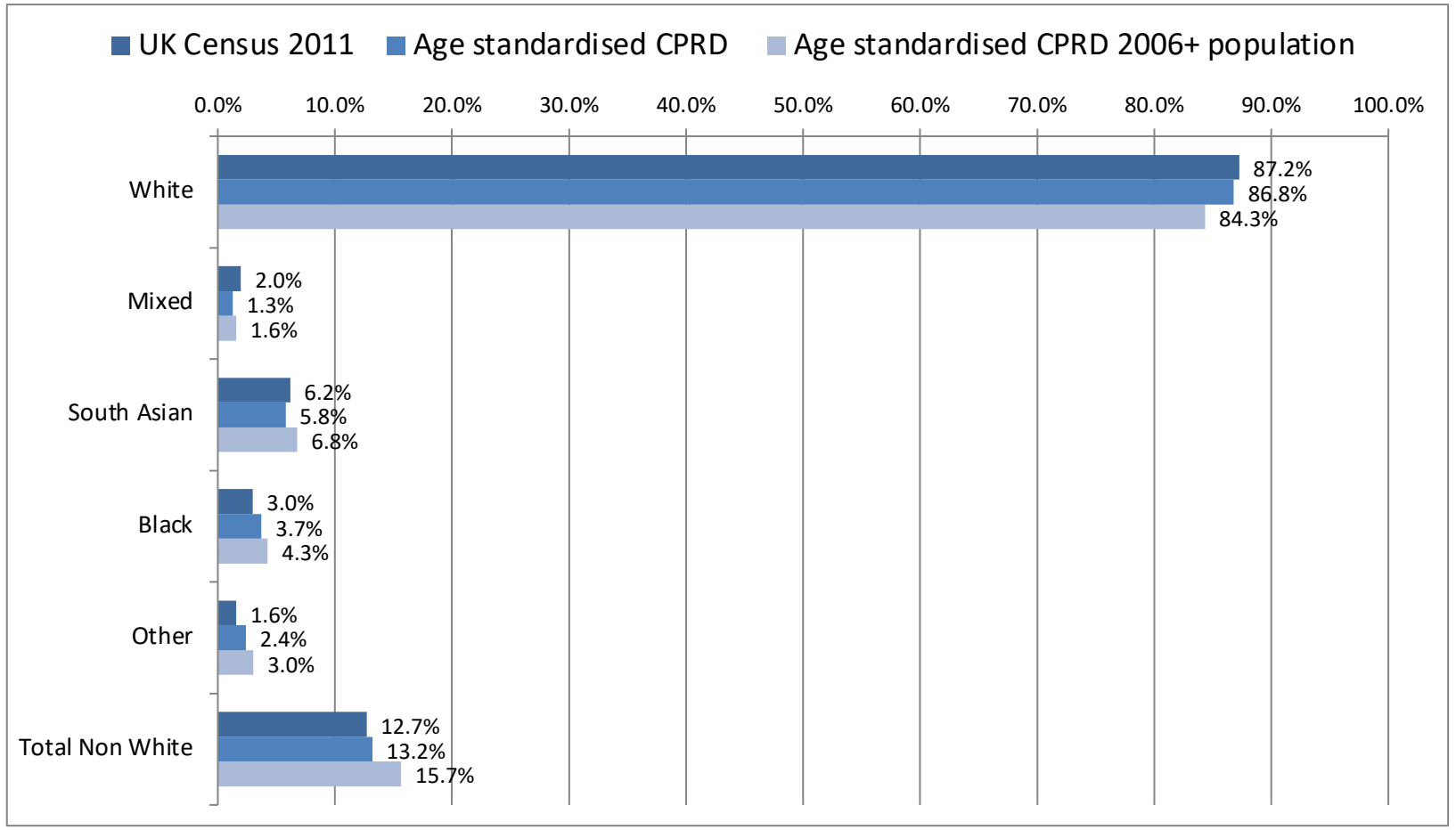
All patients 1990-2012
N=12,099,672



Patients registered from April 2006
N=2,201,065



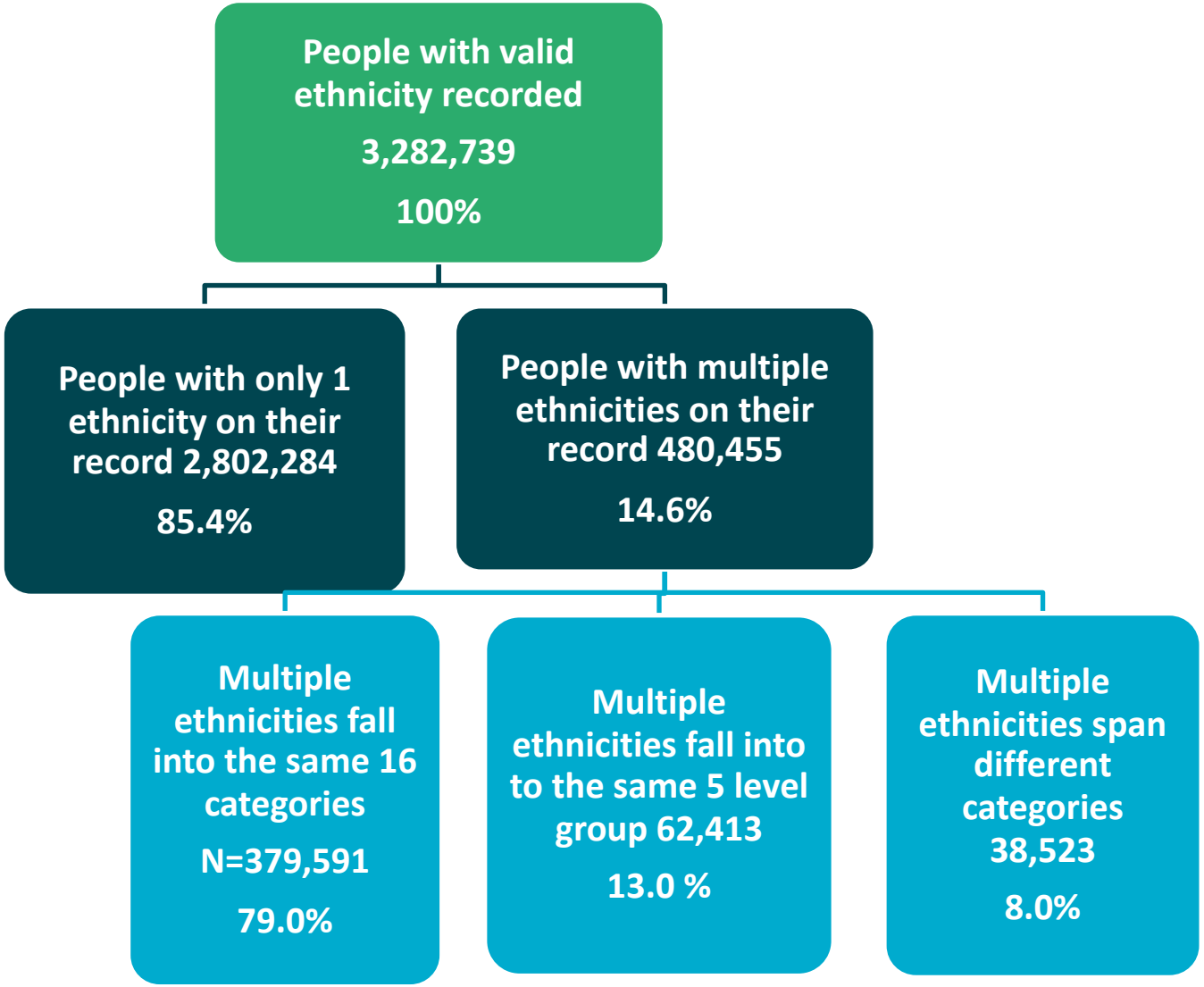
Ethnic breakdown of the CPRD and UK Populations on March 27th 2011



The ethnic breakdown of the UK population in the 2011 census was very similar to the whole CPRD population on that date, both before and after age standardization

Pragmatic methods for assigning ethnicity in UK data sources

Consistency of ethnicity recording in CPRD



Scenario 1: Multiple ethnicities are the same

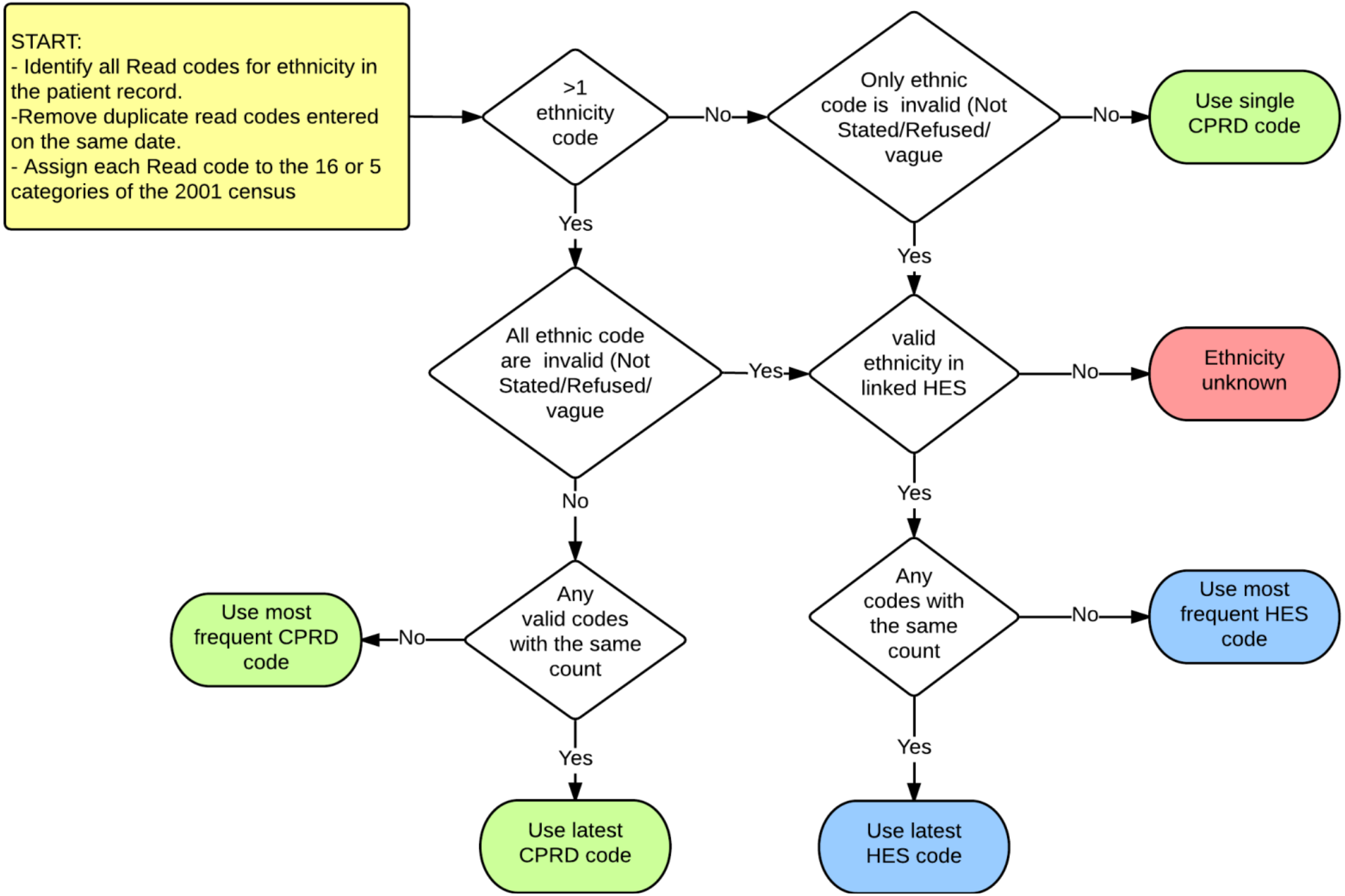
sysdate	desc	ethcode
04dec2008	White	9S1
05nov2009	White	9S1
29oct2010	White	9S1
16nov2011	White	9S1

Scenario 2: Multiple ethnicities are categorically the same

sysdate	desc	ethcode	eth5
15jan2007	Bangladeshi	9S8	South Asian
25jan2007	Other Asian ethnic group	9SH	South Asian
25jan2007	Indian	9S6	South Asian
03may2007	Other Asian ethnic group	9SH	South Asian
10nov2007	Other Asian ethnic group	9SH	South Asian
12dec2008	Ethnic group not recorded	9SE	Not Stated
17jan2011	Tamil	9iA5	South Asian

Scenario 3: Multiple ethnicities are categorically different

sysdate	desc	ethcode	eth5
20jul2004	White British	9S10	White
06jan2004	Ethnic group not given - patient refused	9SD	Not Stated
17jun2004	Indian	9S6	South Asian
17mar2005	Other white ethnic group	9S12	White
20jun2005	Vietnamese	9SC	Other
06jan2006	White	9S1	White
06jan2006	Ethnic group not recorded	9SE	Not Stated
06jan2006	Ethnic group not given - patient refused	9SD	Not Stated
14aug2006	Black Caribbean	9S2	Black
04sep2006	White	9S1	White
05feb2007	White	9S1	White
14sep2009	English	9i20	White
28jan2011	Any other group	9iFK	Other
13jul2011	Other Asian ethnic group	9SH	South Asian
26sep2011	Kosovan	9i2J	White
11nov2011	Black, other, non-mixed origin	9S4	Black


START:
- Identify all Read codes for ethnicity in the patient record.
-Remove duplicate read codes entered on the same date.
- Assign each Read code to the 16 or 5 categories of the 2001 census
>1 ethnicity code
No
Only ethnic code is invalid (Not Stated/Refused/ vague
No
Use single CPRD code
Yes
Yes
All ethnic code are invalid (Not Stated/Refused/ vague
Yes
valid ethnicity in linked HES
No
Ethnicity unknown
No
Yes
Any valid codes with the same count
No
Use most frequent CPRD code
Any codes with the same count
No
Use most frequent HES code
Yes
Yes
Use latest CPRD code
Use latest HES code

Considerations for using routinely recorded ethnicity data in observational research

Key considerations

1. The social context in which ethnicity is defined

2. Biases in capture of ethnicity data in healthcare settings

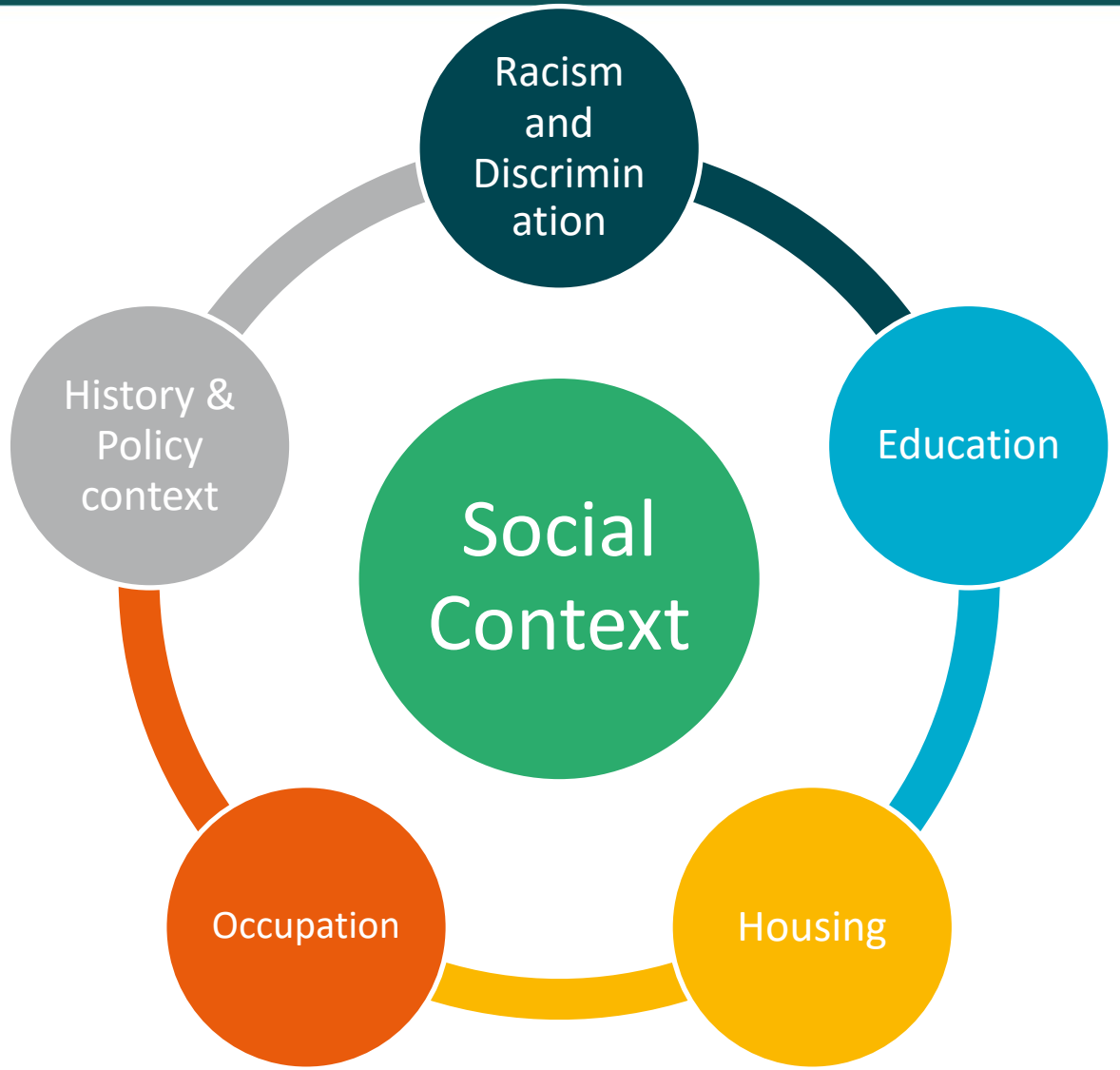
3. Categorization of ethnicity codes

4. Conceptualization of ethnicity in epidemiological models

5. Methods for dealing with missing ethnicity

1. The social context

- Ethnicity classifications reflect how particular groups of people have been **racialized** - ie/ how their ethnic identity has been shaped by historical and political forces
- Understanding health differences between ethnic groups requires an understanding of how political and historical context shape the wider determinants of health
- Need to consider how structural, institutional, and interpersonal racism shape the experience of minority ethnic groups



https://www.health.org.uk/publications/long-reads/how-to-interpret-research-on-ethnicity-and-covid-19-risk-and-outcomes-five

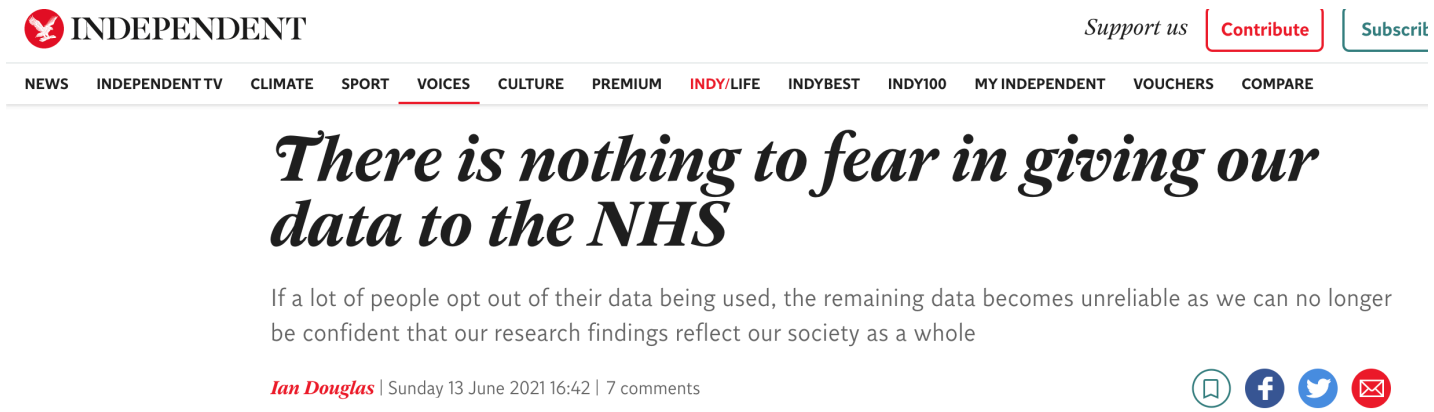
2.Biases in the capture of ethnicity data

Data quality and completeness affected by:

- Who attends primary care
- Whether a person feels represented by the categories
- Whether ethnicity recording is financially incentivised

People who don't have their ethnicity recorded may be:

- Less likely to attend health services
- May be healthier (younger, fewer conditions which require regular monitoring)
- More mobile (students, homeless, asylum seekers)
- Less able to access care
- Less likely to be asked about their ethnicity



INDEPENDENT

Support us | Contribute | Subscrit

NEWS | INDEPENDENT TV | CLIMATE | SPORT | VOICES | CULTURE | PREMIUM | INDY/LIFE | INDYBEST | INDY100 | MY INDEPENDENT | VOUCHERS | COMPARE

There is nothing to fear in giving our data to the NHS

If a lot of people opt out of their data being used, the remaining data becomes unreliable as we can no longer be confident that our research findings reflect our society as a whole

Ian Douglas | Sunday 13 June 2021 16:42 | 7 comments

3. Categorisation of ethnicity codes

- Standard census categories may not be most appropriate for your specific research question
- People who don't feel represented by categories may respond as 'Other', 'Unknown' or 'Refused'
- Overuse of the 'Other' categories means that ethnicity is not being recorded optimally for every individual
- Bespoke categories may be required
  - Important to consult with communities and healthcare providers

QUALITY IMPROVEMENT REPORT

Improving MMR vaccination rates: herd immunity is a realistic goal

Philippa Cockman *general practitioner and clinical lead for immunisation*[1], Luise Dawson *public health nurse*[2], Rohini Mathur *research fellow*[3], Sally Hull *senior clinical lecturer*[3]

South Asian group had the highest proportion of children vaccinated within the COVER period (93.61%). Focus group work by Tower Hamlets Primary Care Trust before the intervention indicated concerns in the Somali community about the MMR vaccine.[23] Feedback from practices also suggested that Somali children were hard to reach. When Somali children were treated as a separate ethnic group, we found that only 56% of Somali children had been vaccinated within the COVER period.

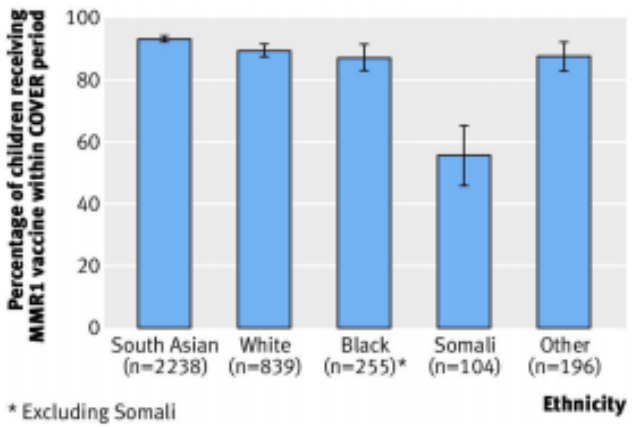


Fig 4 Crude proportion of children receiving the MMR1 vaccine within the COVER period, by ethnicity

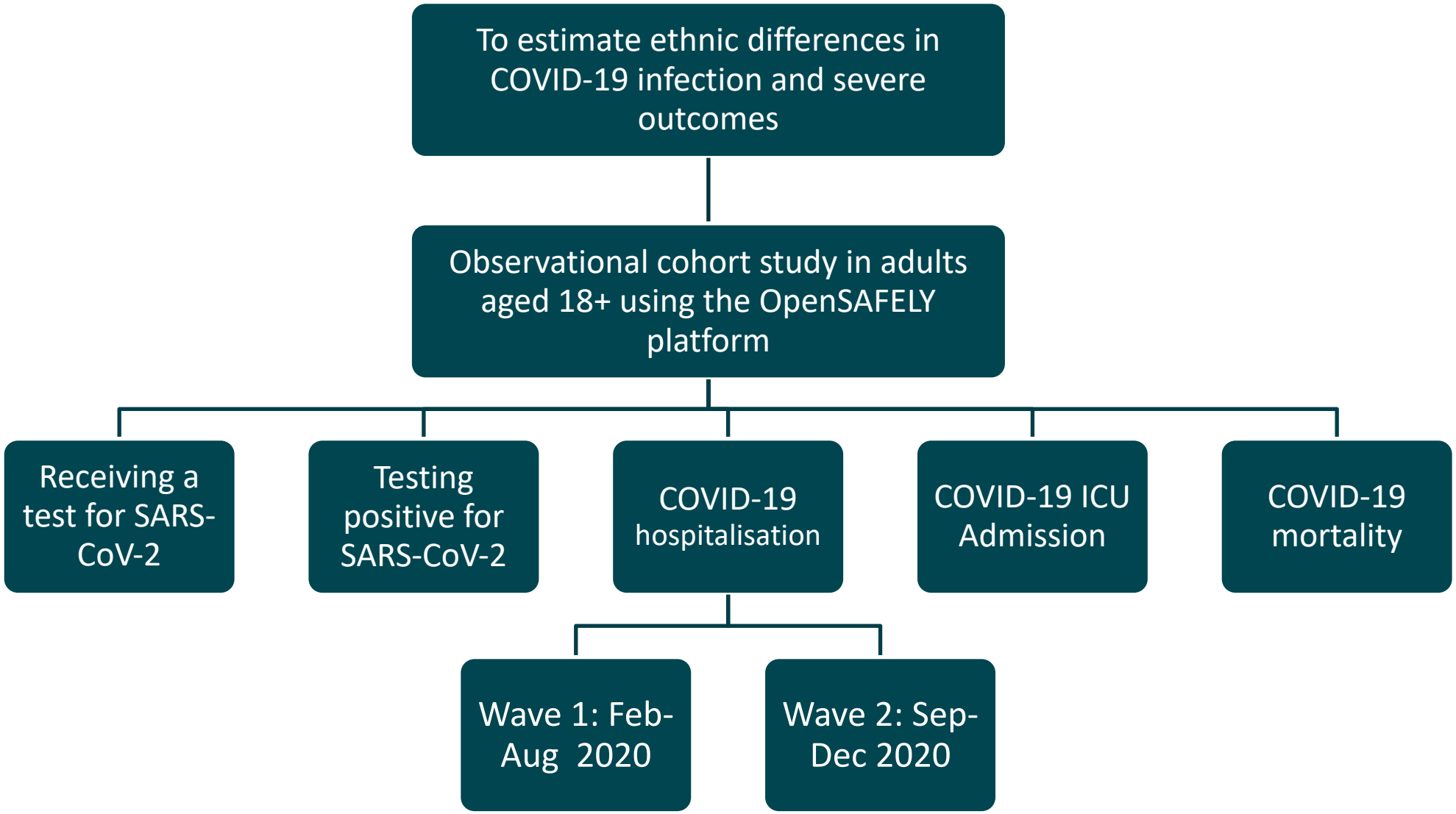
Example study:

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

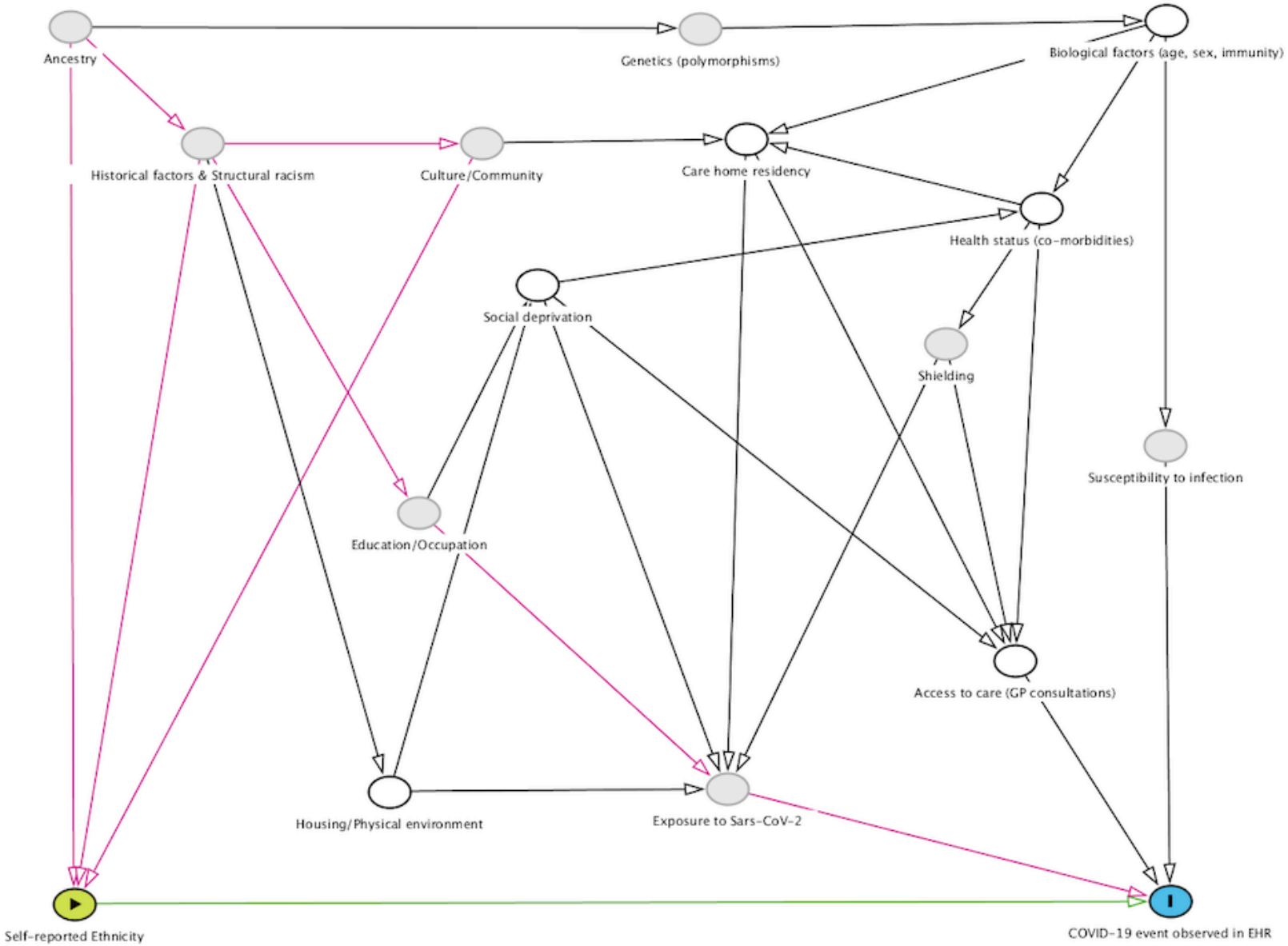
Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: an observational cohort study using the OpenSAFELY platform

Rohini Mathur, Christopher T Rentsch*, Caroline E Morton*, William J Hulme, Anna Schultze, Brian MacKenna, Rosalind M Eggo, Krishnan Bhaskaran, Angel Y S Wong, Elizabeth J Williamson, Harriet Forbes, Kevin Wing, Helen I McDonald, Chris Bates, Seb Bacon, Alex J Walker, David Evans, Peter Inglesby, Amir Mehrkar, Helen J Curtis, Nicholas J DeVito, Richard Croker, Henry Drysdale, Jonathan Cockburn, John Parry, Frank Hester, Sam Harper, Ian J Douglas, Laurie Tomlinson, Stephen J W Evans, Richard Grieve, David Harrison, Kathy Rowan, Kamlesh Khunti, Nishi Chaturvedi, Liam Smeeth†, Ben Goldacre†, for the OpenSAFELY Collaborative*

Study Aims


To estimate ethnic differences in COVID-19 infection and severe outcomes
Observational cohort study in adults aged 18+ using the OpenSAFELY platform
Receiving a test for SARS-CoV-2
Testing positive for SARS-CoV-2
COVID-19 hospitalisation
COVID-19 ICU Admission
COVID-19 mortality
Wave 1: Feb-Aug 2020
Wave 2: Sep-Dec 2020

4. Conceptualization of ethnicity in causal models



https://codelists.opensafely.org/

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

OpenSAFELY Codelists

Ethnicity

Download CSV

About | Definition | Full list | Tree

Coding system
CTV3 (Read V3)

Project
OpenSAFELY

Codelist ID
opensafely/ethnicity

Version
2020-04-27

Edit metadata

Create new version

Versions

- 2020-04-27

Description

A list of ethnicity codes in use in UK general practice including aggregate grouping at two levels.

Methodology

An initial long list of ethnicity codes was provided from a paper on the Completeness and Usability of Ethnicity Data in UK-based Primary Care and Hospital Databases. This list was further iterated by aggregating ethnicity to parent ethnicity at two levels based on a list provided by Rohini Mathur, the first author on the above paper. This task was completed independently by Ieva Lipsa, Alex J Walker and Brian MacKenna. Any divergence was resolved through discussion between AJW and BMK.

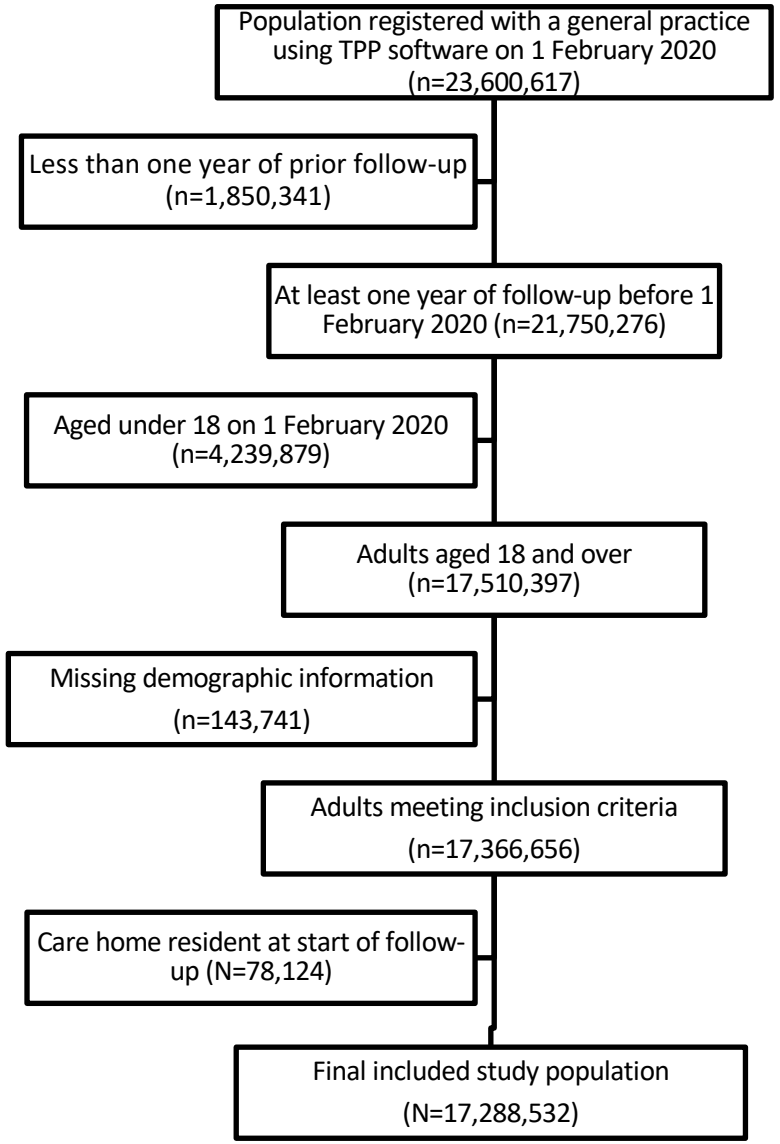
References

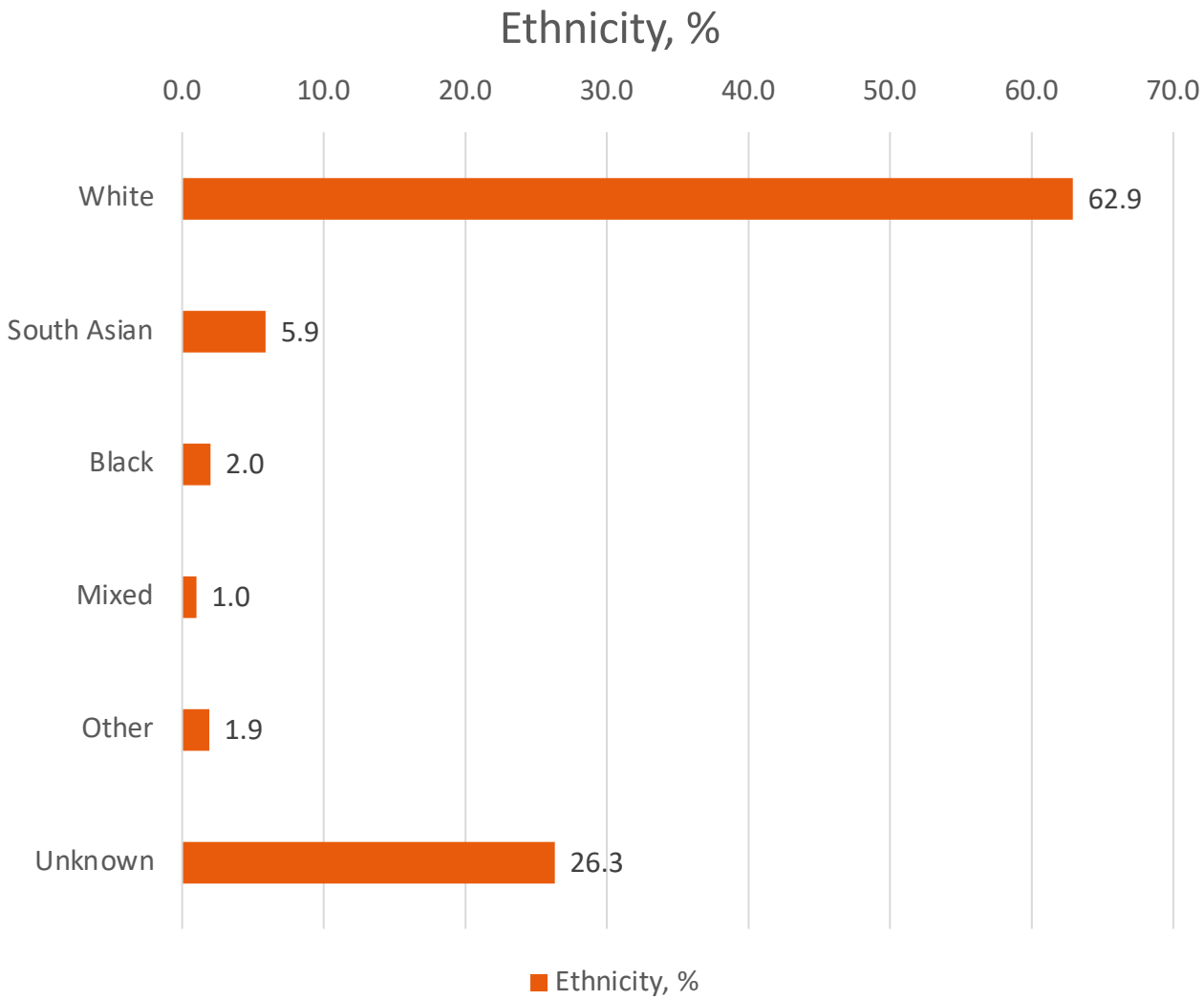
- GitHub discussion
- Paper: Completeness and Usability of Ethnicity Data in UK-based Primary Care and Hospital Databases

Signed off by

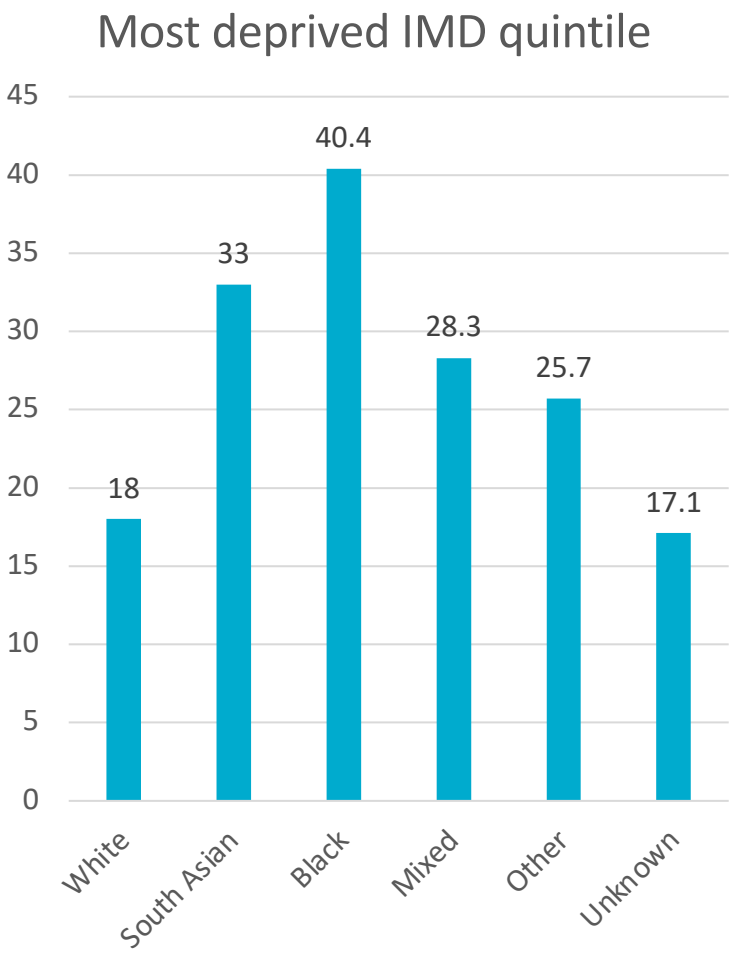
- Brian MacKenna, EBM Datalab (April 22, 2020)
- Alex Walker, EBM Datalab (April 22, 2020)

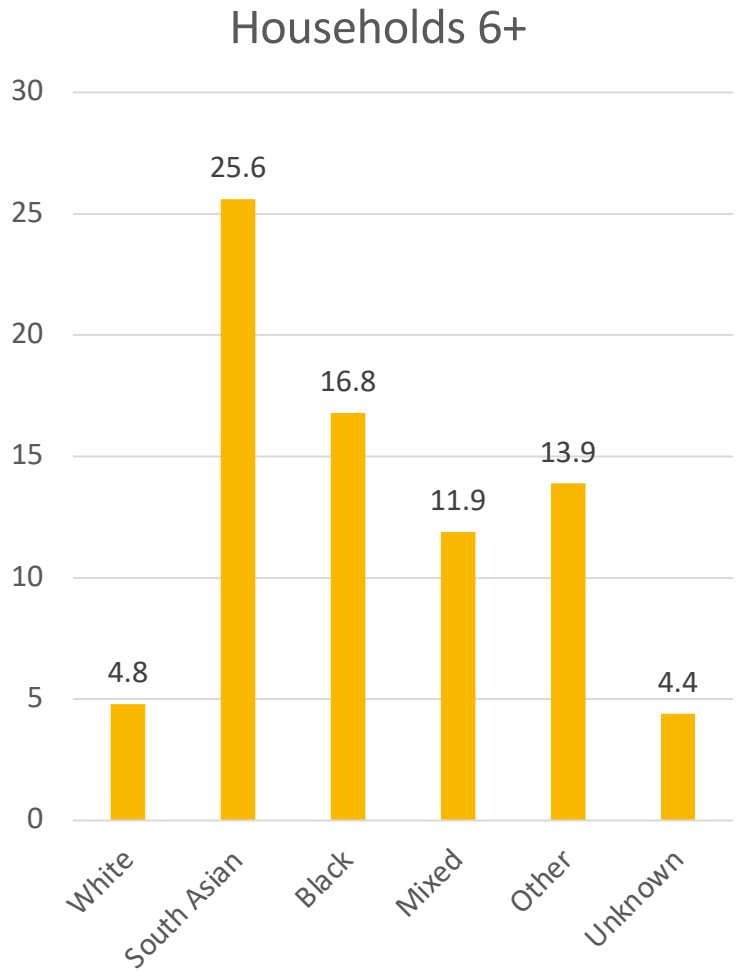
Study Population

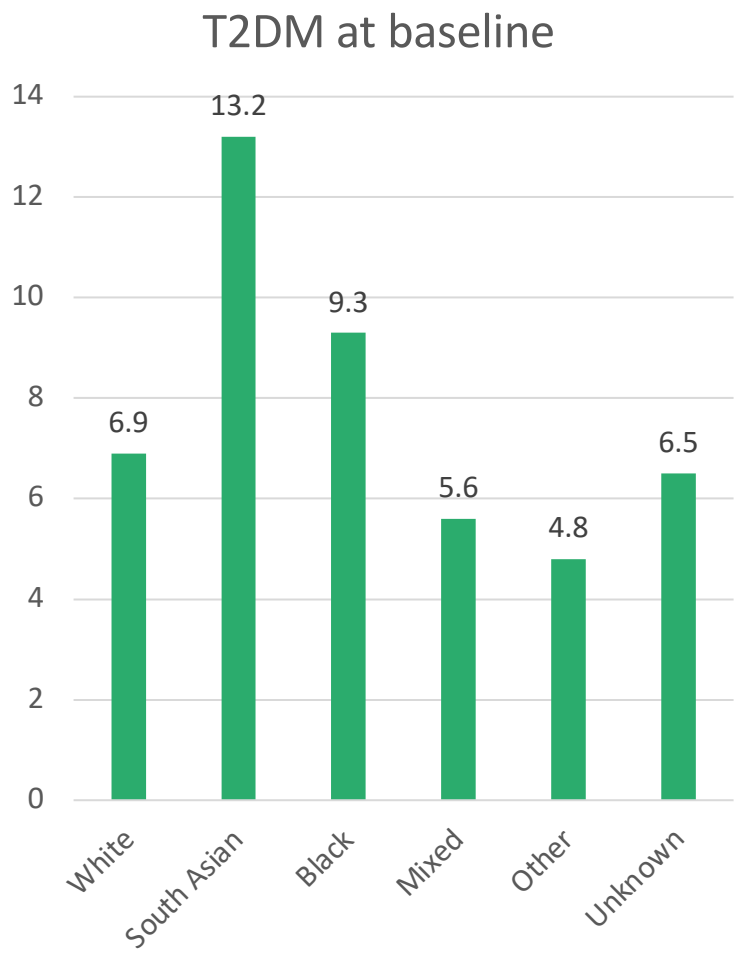




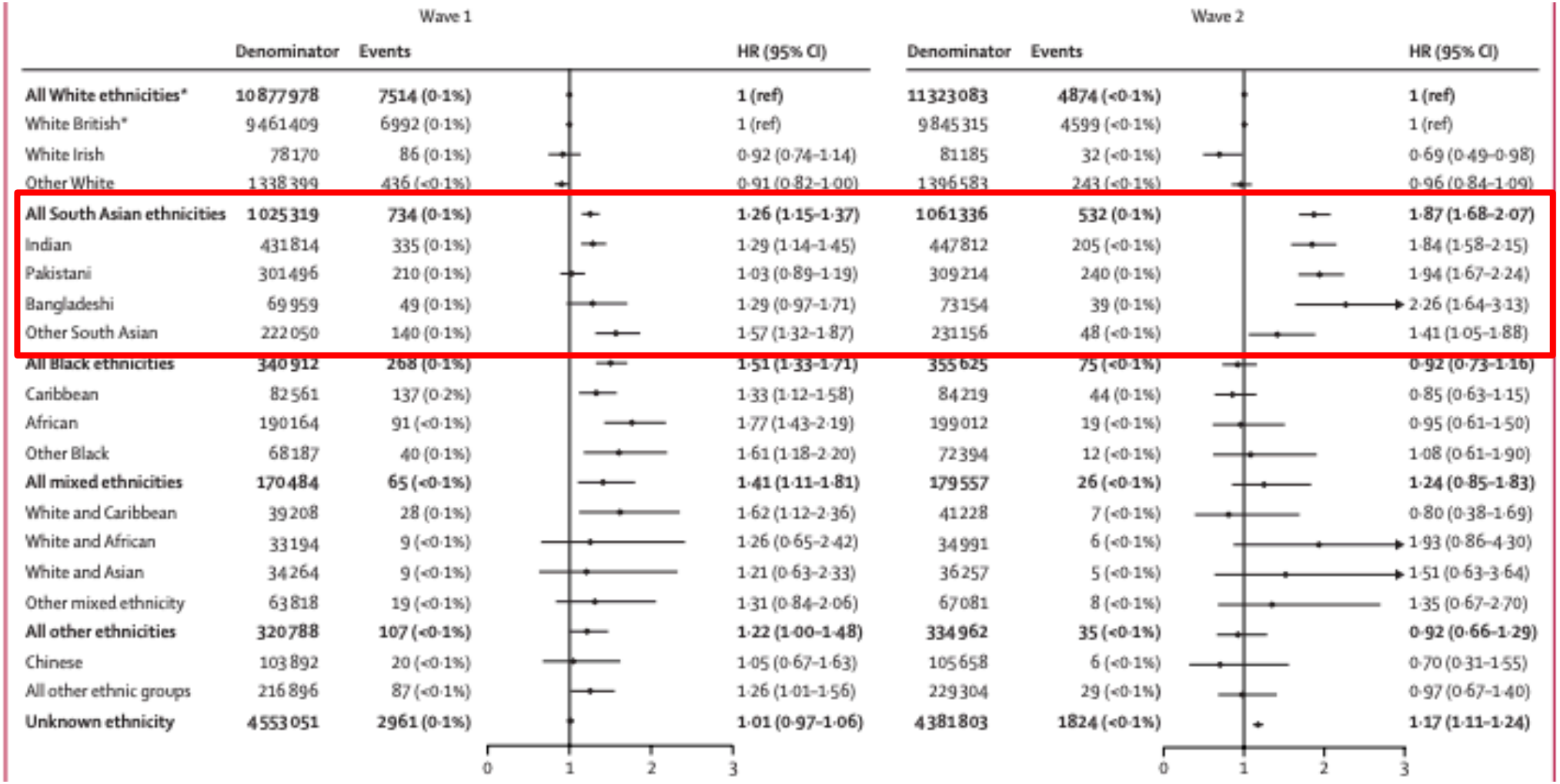
Are people with unknown ethnicity different?





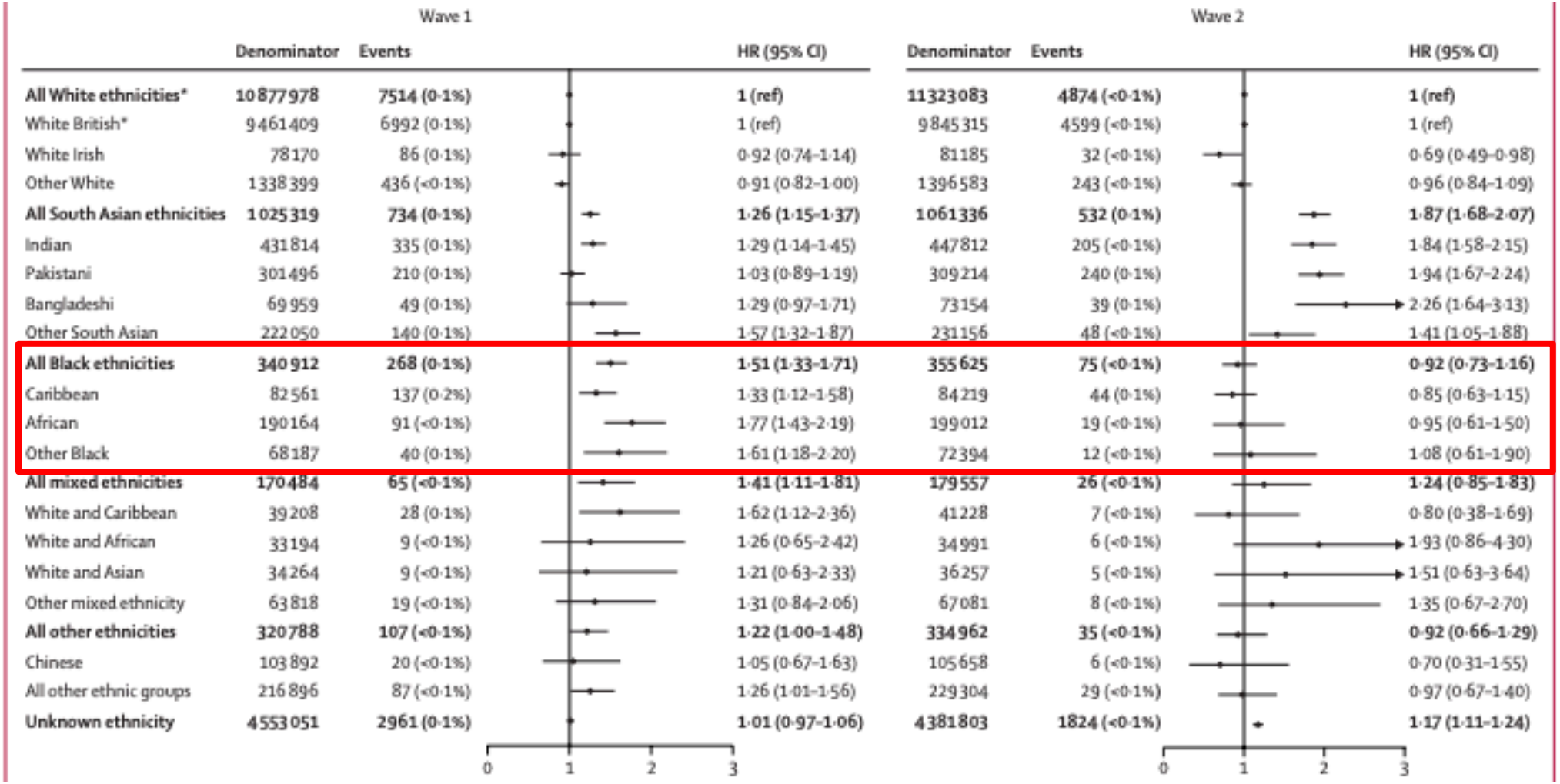


Findings: COVID-19 related mortality



	Wave 1			Wave 2		
	Denominator	Events	HR (95% CI)	Denominator	Events	HR (95% CI)
All White ethnicities*	**10 877 978**	**7514 (0·1%)**	**1 (ref)**	**11 323 083**	**4874 (<0·1%)**	**1 (ref)**
White British*	9 461 409	6992 (0·1%)	1 (ref)	9 845 315	4599 (<0·1%)	1 (ref)
White Irish	78 170	86 (0·1%)	0·92 (0·74–1·14)	81 185	32 (<0·1%)	0·69 (0·49–0·98)
Other White	1 338 399	436 (<0·1%)	0·91 (0·82–1·00)	1 396 583	243 (<0·1%)	0·96 (0·84–1·09)
All South Asian ethnicities	**1 025 319**	**734 (0·1%)**	**1·26 (1·15–1·37)**	**1 061 336**	**532 (0·1%)**	**1·87 (1·68–2·07)**
Indian	431 814	335 (0·1%)	1·29 (1·14–1·45)	447 812	205 (<0·1%)	1·84 (1·58–2·15)
Pakistani	301 496	210 (0·1%)	1·03 (0·89–1·19)	309 214	240 (0·1%)	1·94 (1·67–2·24)
Bangladeshi	69 959	49 (0·1%)	1·29 (0·97–1·71)	73 154	39 (0·1%)	2·26 (1·64–3·13)
Other South Asian	222 050	140 (0·1%)	1·57 (1·32–1·87)	231 156	48 (<0·1%)	1·41 (1·05–1·88)
All Black ethnicities	**340 912**	**268 (0·1%)**	**1·51 (1·33–1·71)**	**355 625**	**75 (<0·1%)**	**0·92 (0·73–1·16)**
Caribbean	82 561	137 (0·2%)	1·33 (1·12–1·58)	84 219	44 (0·1%)	0·85 (0·63–1·15)
African	190 164	91 (<0·1%)	1·77 (1·43–2·19)	199 012	19 (<0·1%)	0·95 (0·61–1·50)
Other Black	68 187	40 (0·1%)	1·61 (1·18–2·20)	72 394	12 (<0·1%)	1·08 (0·61–1·90)
All mixed ethnicities	**170 484**	**65 (<0·1%)**	**1·41 (1·11–1·81)**	**179 557**	**26 (<0·1%)**	**1·24 (0·85–1·83)**
White and Caribbean	39 208	28 (0·1%)	1·62 (1·12–2·36)	41 228	7 (<0·1%)	0·80 (0·38–1·69)
White and African	33 194	9 (<0·1%)	1·26 (0·65–2·42)	34 991	6 (<0·1%)	1·93 (0·86–4·30)
White and Asian	34 264	9 (<0·1%)	1·21 (0·63–2·33)	36 257	5 (<0·1%)	1·51 (0·63–3·64)
Other mixed ethnicity	63 818	19 (<0·1%)	1·31 (0·84–2·06)	67 081	8 (<0·1%)	1·35 (0·67–2·70)
All other ethnicities	**320 788**	**107 (<0·1%)**	**1·22 (1·00–1·48)**	**334 962**	**35 (<0·1%)**	**0·92 (0·66–1·29)**
Chinese	103 892	20 (<0·1%)	1·05 (0·67–1·63)	105 658	6 (<0·1%)	0·70 (0·31–1·55)
All other ethnic groups	216 896	87 (<0·1%)	1·26 (1·01–1·56)	229 304	29 (<0·1%)	0·97 (0·67–1·40)
Unknown ethnicity	**4 553 051**	**2961 (0·1%)**	**1·01 (0·97–1·06)**	**4 381 803**	**1824 (<0·1%)**	**1·17 (1·11–1·24)**

Findings: COVID-19 related mortality



	Wave 1			Wave 2		
	Denominator	Events	HR (95% CI)	Denominator	Events	HR (95% CI)
All White ethnicities*	**10 877 978**	**7514 (0·1%)**	**1 (ref)**	**11 323 083**	**4874 (<0·1%)**	**1 (ref)**
White British*	9 461 409	6992 (0·1%)	1 (ref)	9 845 315	4599 (<0·1%)	1 (ref)
White Irish	78 170	86 (0·1%)	0·92 (0·74–1·14)	81 185	32 (<0·1%)	0·69 (0·49–0·98)
Other White	1 338 399	436 (<0·1%)	0·91 (0·82–1·00)	1 396 583	243 (<0·1%)	0·96 (0·84–1·09)
All South Asian ethnicities	**1 025 319**	**734 (0·1%)**	**1·26 (1·15–1·37)**	**1 061 336**	**532 (0·1%)**	**1·87 (1·68–2·07)**
Indian	431 814	335 (0·1%)	1·29 (1·14–1·45)	447 812	205 (<0·1%)	1·84 (1·58–2·15)
Pakistani	301 496	210 (0·1%)	1·03 (0·89–1·19)	309 214	240 (0·1%)	1·94 (1·67–2·24)
Bangladeshi	69 959	49 (0·1%)	1·29 (0·97–1·71)	73 154	39 (0·1%)	2·26 (1·64–3·13)
Other South Asian	222 050	140 (0·1%)	1·57 (1·32–1·87)	231 156	48 (<0·1%)	1·41 (1·05–1·88)
All Black ethnicities	**340 912**	**268 (0·1%)**	**1·51 (1·33–1·71)**	**355 625**	**75 (<0·1%)**	**0·92 (0·73–1·16)**
Caribbean	82 561	137 (0·2%)	1·33 (1·12–1·58)	84 219	44 (0·1%)	0·85 (0·63–1·15)
African	190 164	91 (<0·1%)	1·77 (1·43–2·19)	199 012	19 (<0·1%)	0·95 (0·61–1·50)
Other Black	68 187	40 (0·1%)	1·61 (1·18–2·20)	72 394	12 (<0·1%)	1·08 (0·61–1·90)
All mixed ethnicities	**170 484**	**65 (<0·1%)**	**1·41 (1·11–1·81)**	**179 557**	**26 (<0·1%)**	**1·24 (0·85–1·83)**
White and Caribbean	39 208	28 (0·1%)	1·62 (1·12–2·36)	41 228	7 (<0·1%)	0·80 (0·38–1·69)
White and African	33 194	9 (<0·1%)	1·26 (0·65–2·42)	34 991	6 (<0·1%)	1·93 (0·86–4·30)
White and Asian	34 264	9 (<0·1%)	1·21 (0·63–2·33)	36 257	5 (<0·1%)	1·51 (0·63–3·64)
Other mixed ethnicity	63 818	19 (<0·1%)	1·31 (0·84–2·06)	67 081	8 (<0·1%)	1·35 (0·67–2·70)
All other ethnicities	**320 788**	**107 (<0·1%)**	**1·22 (1·00–1·48)**	**334 962**	**35 (<0·1%)**	**0·92 (0·66–1·29)**
Chinese	103 892	20 (<0·1%)	1·05 (0·67–1·63)	105 658	6 (<0·1%)	0·70 (0·31–1·55)
All other ethnic groups	216 896	87 (<0·1%)	1·26 (1·01–1·56)	229 304	29 (<0·1%)	0·97 (0·67–1·40)
Unknown ethnicity	**4 553 051**	**2961 (0·1%)**	**1·01 (0·97–1·06)**	**4 381 803**	**1824 (<0·1%)**	**1·17 (1·11–1·24)**

Comparisons with USA

Racial and ethnic disparities for SARS-CoV-2 positivity in the United States: a generalizing pandemic

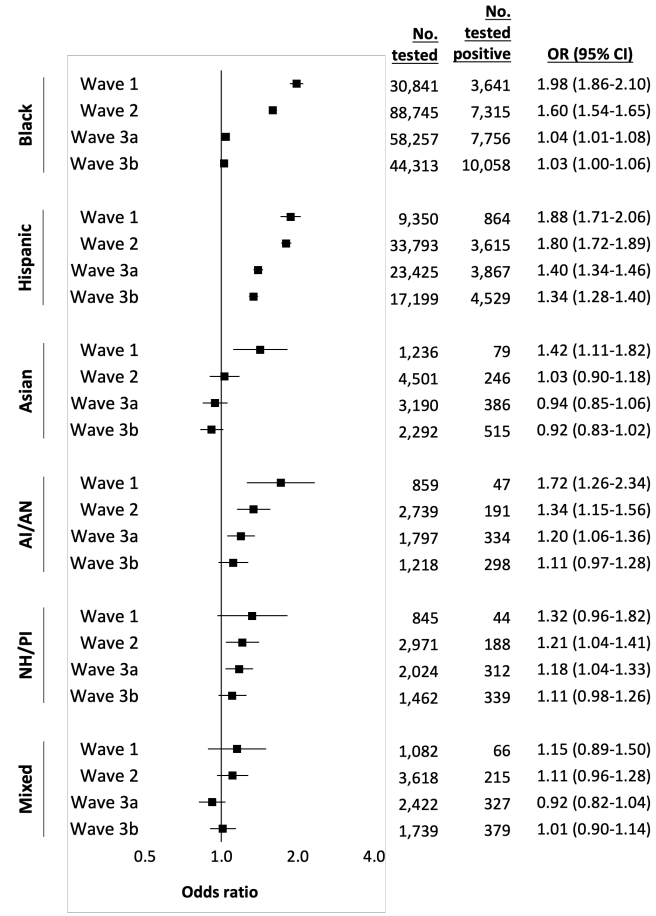
Keywords: health disparities, COVID-19, SARS-CoV-2, race, ethnicity, testing

Jacqueline M. Ferguson, PhD[1,2*]; Amy C. Justice[3,4,5]; Thomas F. Osborne, MD[1,6]; Hoda S. Abdel Magid, PhD[1,7,8]; Amanda L. Purnell, PhD[9]; Christopher T. Rentsch[3,5,10]

https://doi.org/10.1101/2021.04.27.21256215

Excess risk of COVID-19 highest in Wave 1 and attenuated over time in all ethnic minority groups

Figure 2. Racial and ethnic disparities in testing positive for SARS-CoV-2 between February 12, 2020 and February 12, 2021, by wave of the pandemic



Notes: Wave 1 (February 12 – May 31, 2020); Wave 2 (June 1 – September 30, 2020); Wave 3a (October 1 –December 11, 2020); and Wave 3b (December 12, 2020 – February 12, 2021). Referent group for all comparisons was White. Number tested/number tested positive for White individuals were as follows: 72,249/3,254 for Wave 1; 247,151/11,863 for Wave 2; 174,293/26,391 for Wave 3a; 117,797/24,793 for Wave 3b. Models conditioned on site of care and adjusted for other demographics, baseline comorbidity, substance use, and medication history.

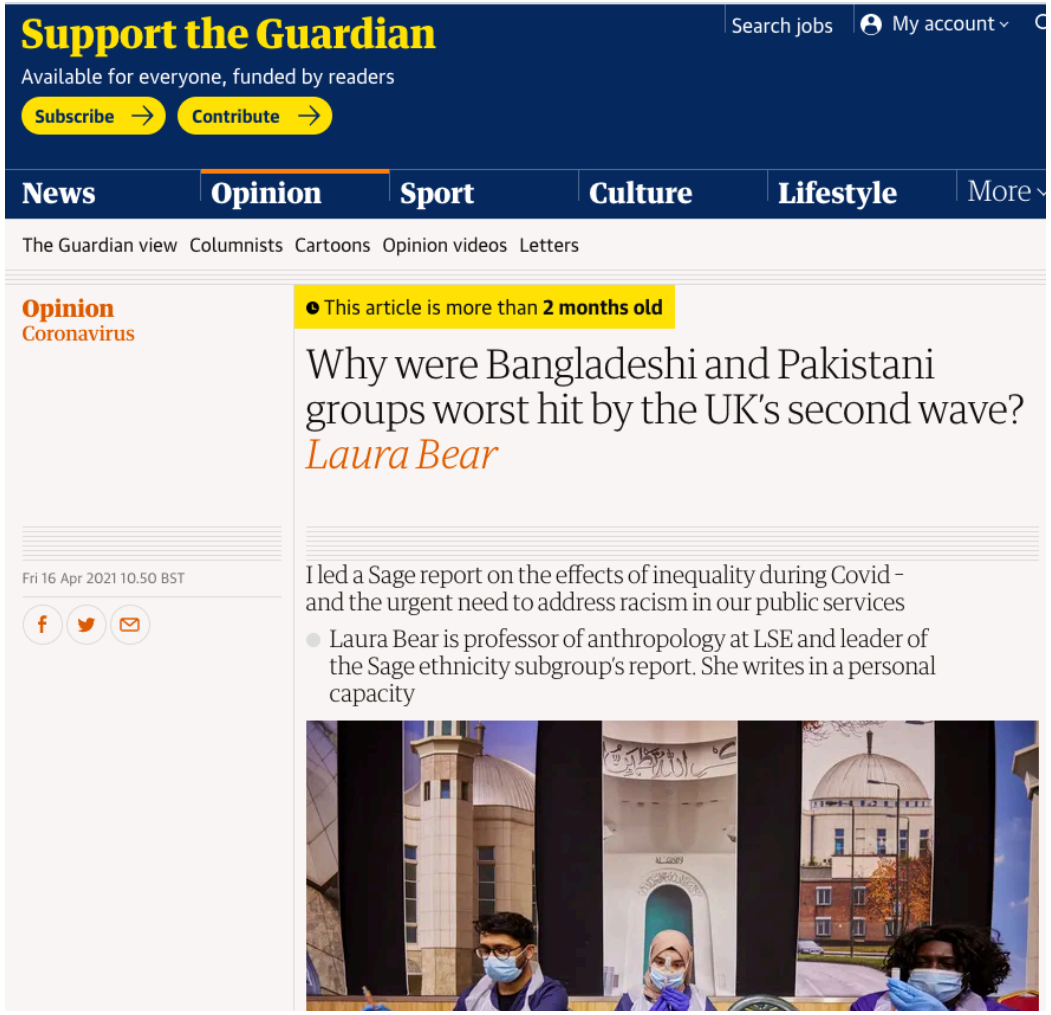
Abbreviations: OR, odds ratio; CI, confidence interval; AI/AN, American Indian/Alaska Native; NH/PI, Native Hawaiian/Pacific Islander.

Qualitative insights into Wave 1 vs. Wave 2

Poorer outcomes in Pakistani and Bangladeshi groups in wave 2 due to **the amplifying interaction of:**

1. health inequalities
2. disadvantages associated with occupation and household circumstances
3. barriers to health care access, and
4. potential influence of policy and practice on COVID-19 health seeking behaviour

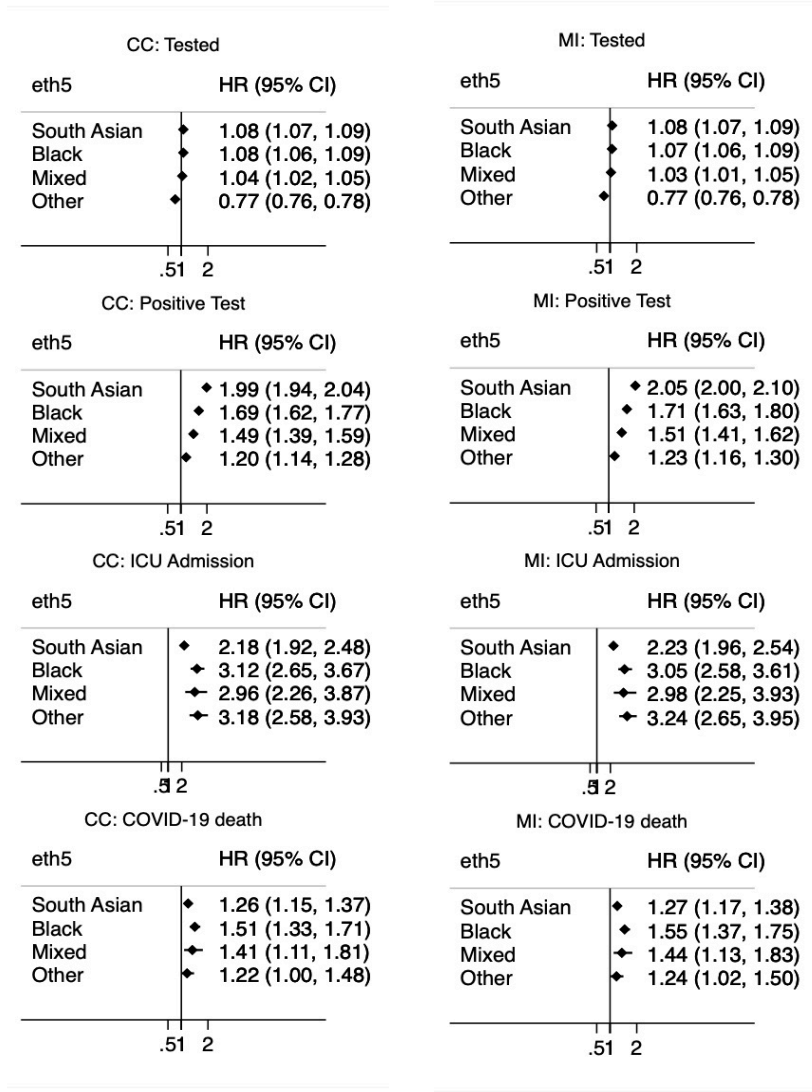
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/976030/S1168_Ethnicity_Subgroup_Wave_1_and_2_qual_comparison.pdf



5. Multiple imputation for unknown ethnicity

- MI is a useful tool for exploring biases related to missing ethnicity
- Results from MI mirrored those of complete case analysis.
- No strong bias introduced by limiting study to those with recorded ethnicity

*CC = complete case MI=Multiple Imputation



Final thoughts

- Reducing ethnic inequalities will need **action across a broad range of measures** such as addressing the wider adverse effects of disadvantage and structural discrimination, reducing within and between household transmission, and improving control of clinical conditions and risk factors.

- Better and more readily available **linked data** are necessary for characterising ethnic disparities in more detail and targeting public health initiatives for the prevention and management of COVID-19.

- **Engaging with ethnic minority communities** to understand their lived experiences will be essential for generating evidence to prevent further widening of inequalities in a timely and actionable manner.

Summary

- Ethnic monitoring is important to uncover potential patterns of disadvantage and racism which may otherwise remain insidious
- Ethnicity provides a useful lens through which to identify population groups between whom disparities may exist
- When hypothesizing about and interpreting the mechanisms through which ethnicity is related to health, it is essential to be clear that health outcomes are determined by factors associated with ethnicity, not ethnicity itself

Thank you for listening!