Examining the use of ethnicity at different levels of aggregation using I-MAIHDA: A methodological exploration & reflection

Joseph Lam, Aaron Koay, Mario Cortina Borja,
Robert Aldridge, Ruth Blackburn, Katie Harron

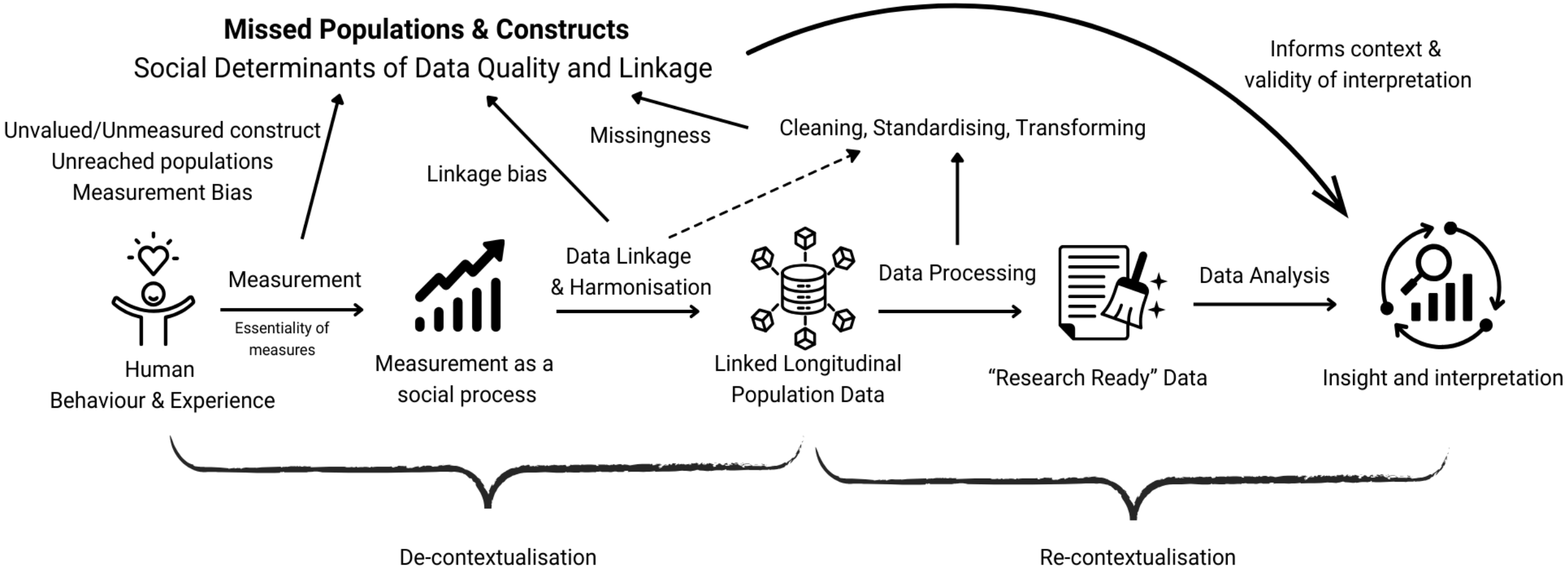
Jo_Lam_

Joseph.Lam.18@ucl.ac.uk

CHIG child health informatics group

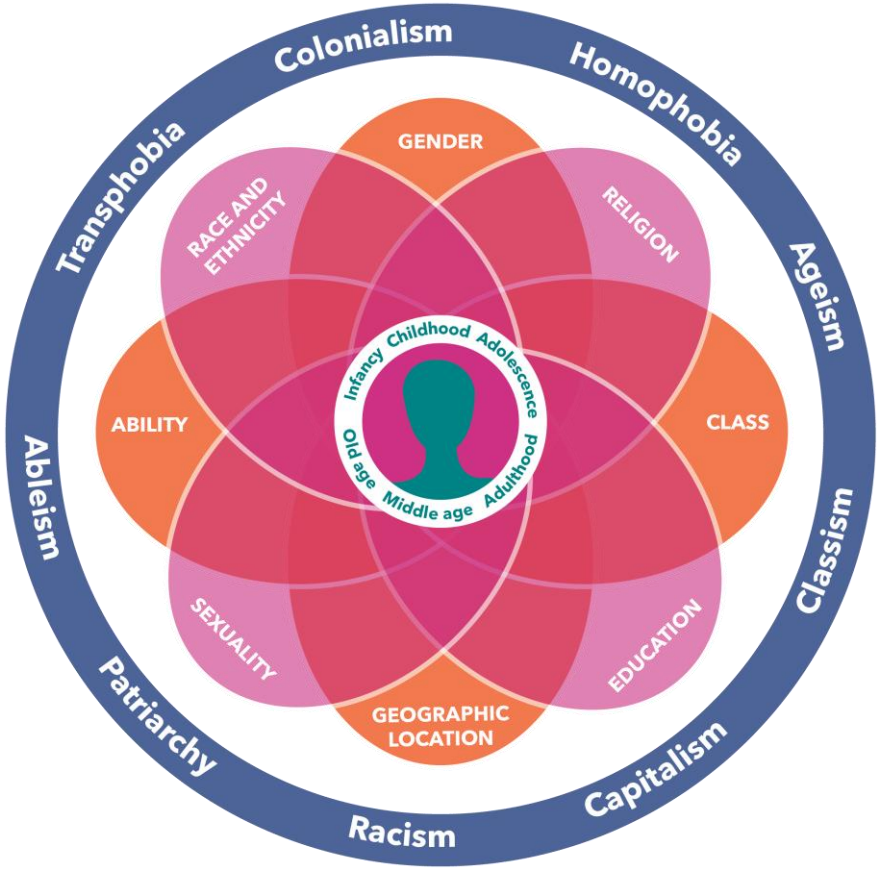
GREAT ORMOND STREET INSTITUTE OF CHILD HEALTH

Framework for Evidence Generation



Intersectionality

- Intersectionality is an *emerging* critical social theory (Collins, 2019)
- Systems of power are **interdependent**. They interlock to generate complex intergroup and intragroup differences in (dis)advantages (Collins, 1990; Combahee River Collective, 1977; Crenshaw, 1989)
- The whole is greater than the sum of its parts
- **How to operationalise (in epidemiology/public health)?**



Available at: https://justassociates.org/big-ideas/intersectionality

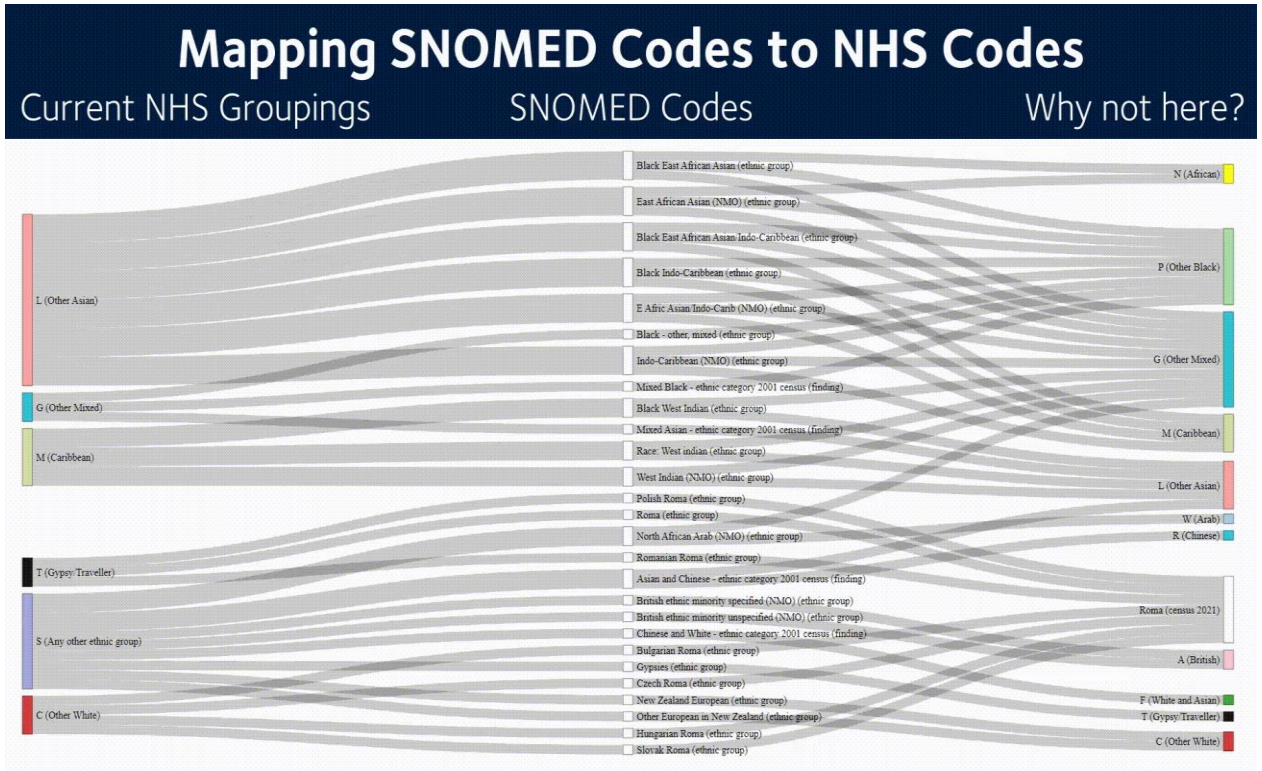
"Once You've Blended the Cake, You Can't Take the Parts Back to the Main Ingredients"

(Bowleg, 2012)

Are more disaggregated ethnicity categories always “better”?

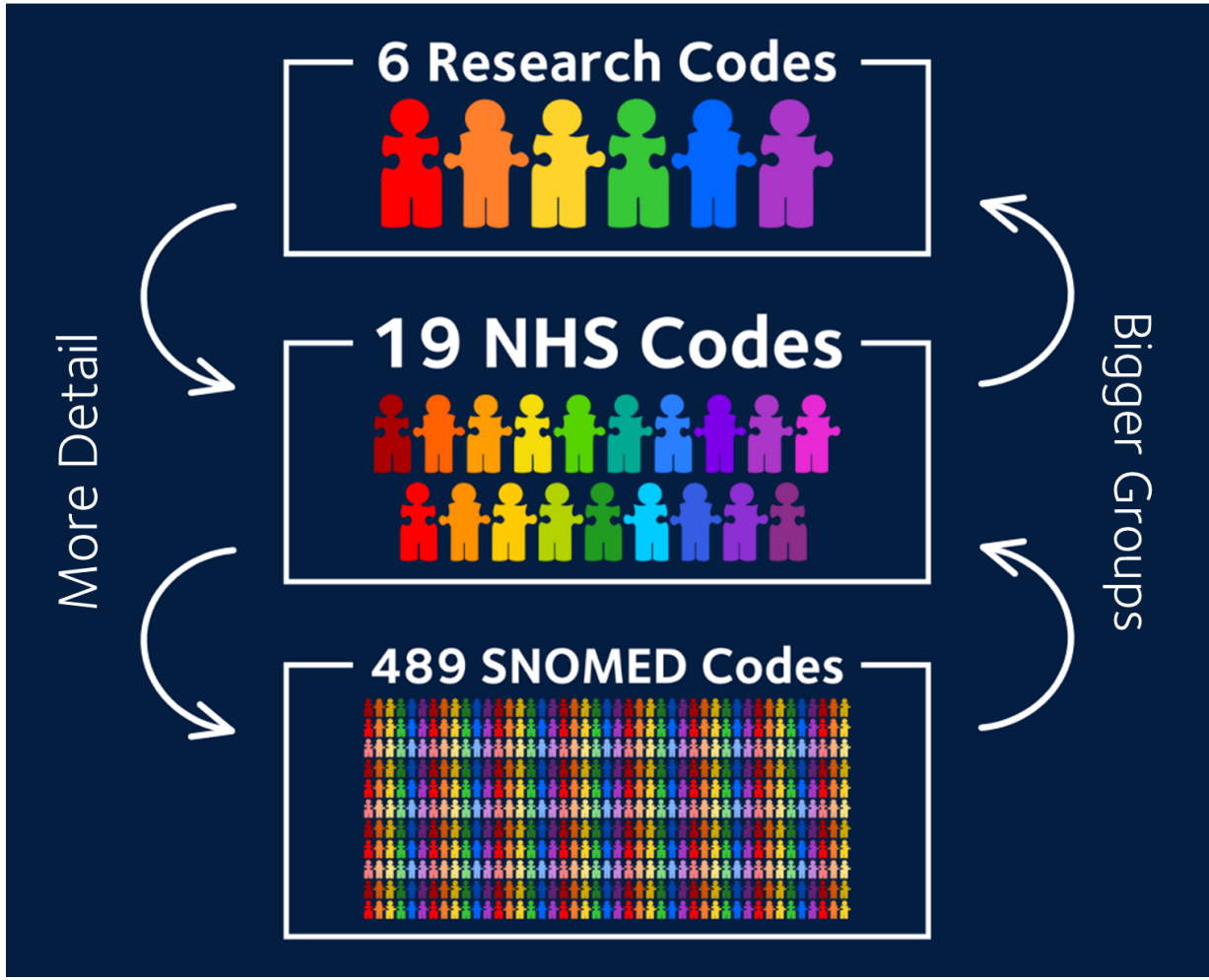
Evidence from Equality National Survey: A Survey of Ethnic Minorities During the COVID-19 Pandemic (EVENS)

Aggregation & granularity



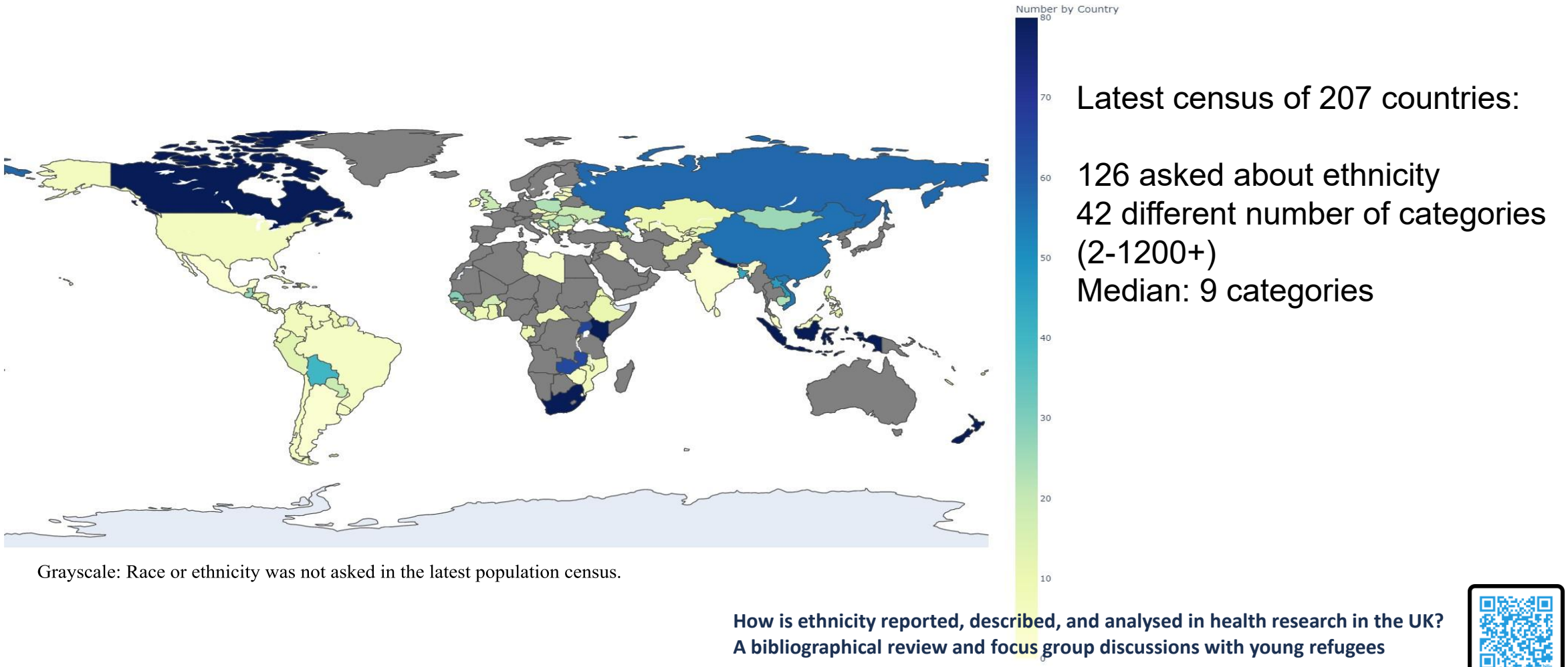
Pineda-Moncusí, M., Allery, F., Delmestri, A. *et al.* Ethnicity data resource in population-wide health records: completeness, coverage and granularity of diversity. *Sci Data* **11**, 221 (2024). https://doi.org/10.1038/s41597-024-02958-1

Different Levels of Ethnicity Codes



Ethnic categories are not universal

Number of reported ethnic groups in latest population census



Grayscale: Race or ethnicity was not asked in the latest population census.

Latest census of 207 countries:

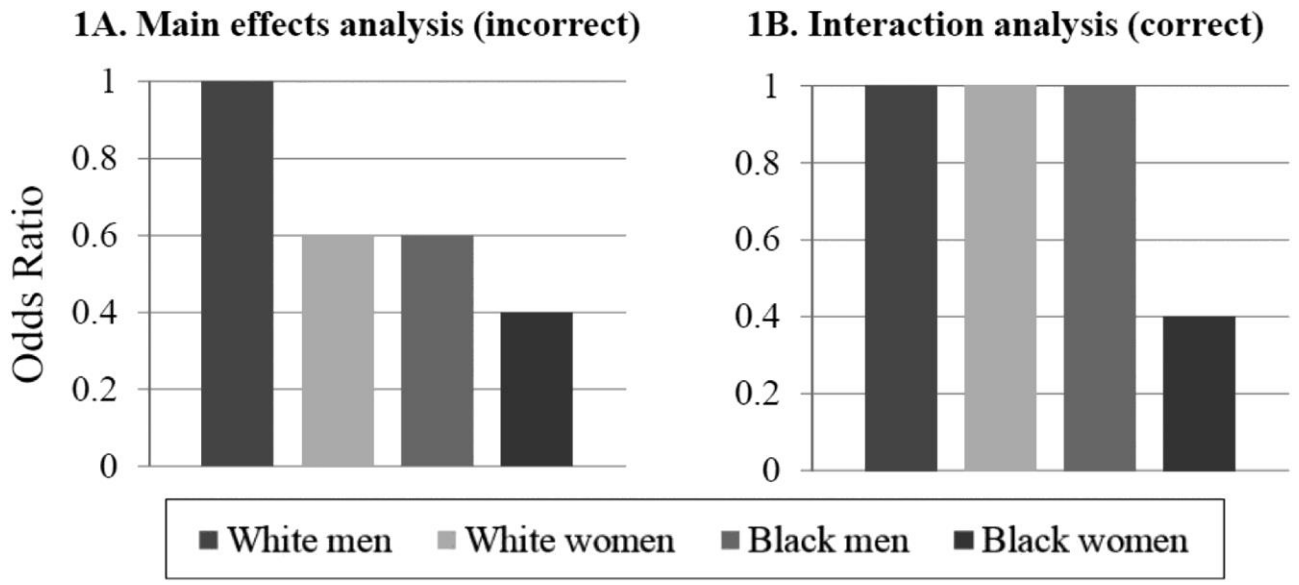
126 asked about ethnicity
42 different number of categories (2-1200+)
Median: 9 categories

How is ethnicity reported, described, and analysed in health research in the UK? A bibliographical review and focus group discussions with young refugees

Intersectionality and interaction

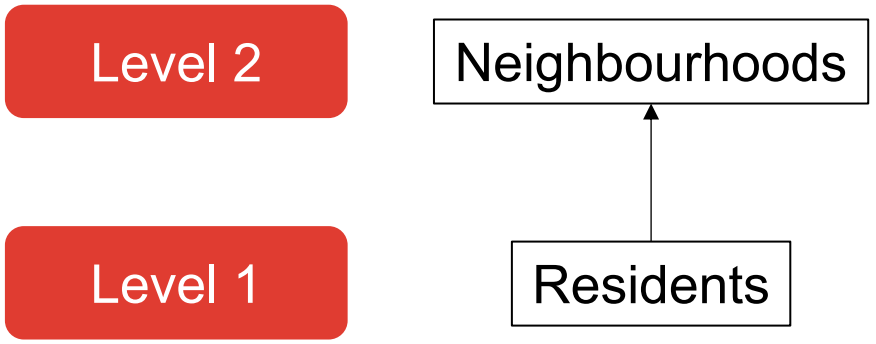
- **Intersectionality**: Interlocking systems of power
- **Statistics**: Combined, multiplicative effects of two or more exposure variables on outcome

The whole is greater than the sum of its parts



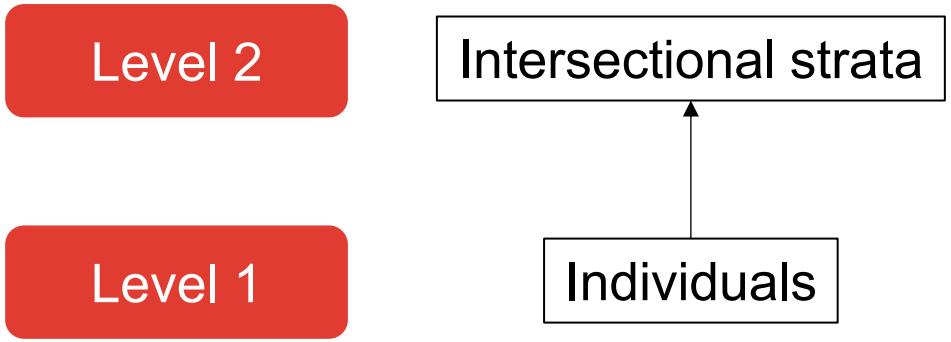
(Bowleg and Bauer, 2016)

Multi-level Modelling



- Shared contexts → **clustering effects**
- Multi-level modelling allows us to estimate
  - Average differences
  - *Within*-cluster variation
  - *Between*-cluster variation

Intersectional Multi-level Analysis of Heterogeneity and Discriminatory Accuracy (I-MAIHDA)

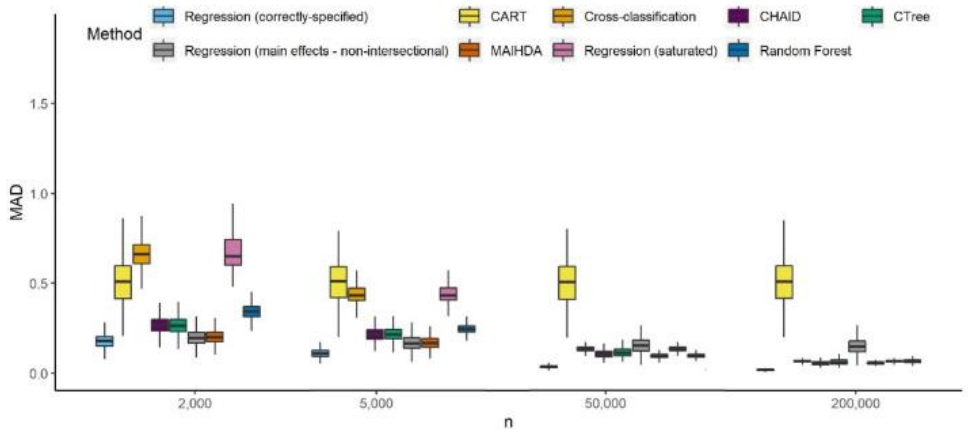


- Evans et al. (2018)
- Based on **abstract social clustering**
- Individual social identity/positions as proxies for social contexts
- Theoretically-engaged descriptive approach

Methodological Advantages

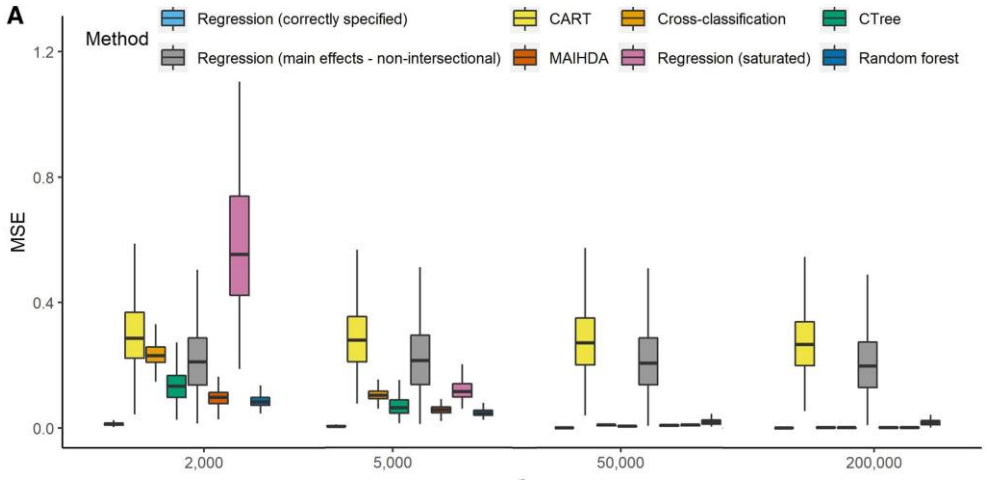
Estimation Accuracy

Binary outcome with categorical inputs



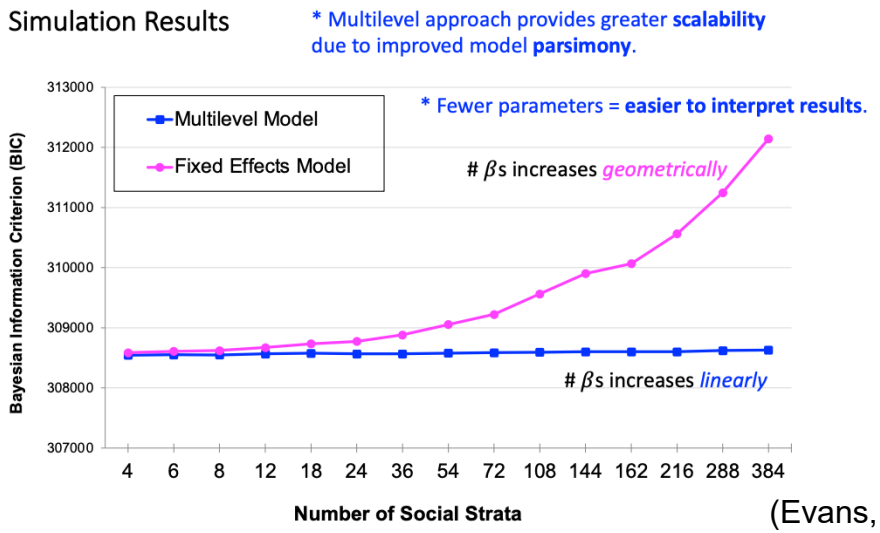
(Mahendran et al., 2022a)

Continuous outcome with categorical inputs



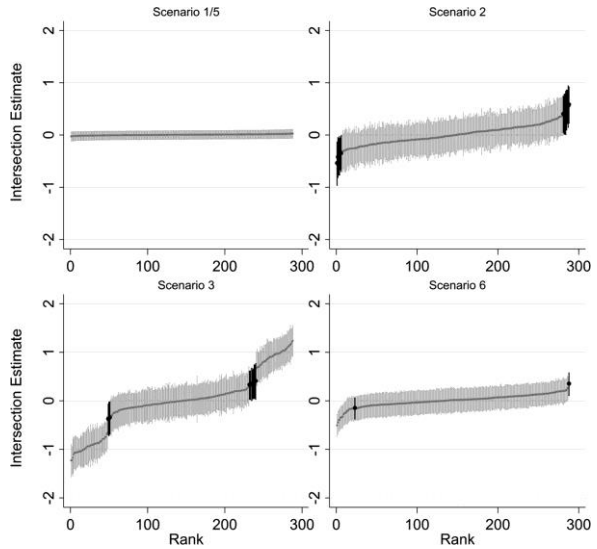
(Mahendran et al., 2022b)

Parsimony



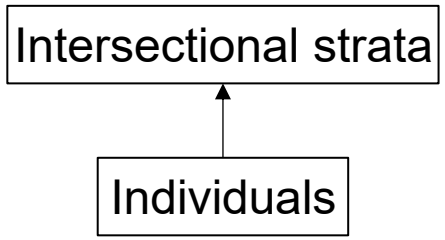
(Evans, 2018)

Shrinkage



(Bell et al, 2019)

I-MAIHDA Modelling



- Two-level random-intercept logistic regression models

- **Model A: Null model**
  - The extent to which intersectional strata can explain outcome heterogeneity
  - **Variance partition coefficient (VPC)**: % total between-stratum variance

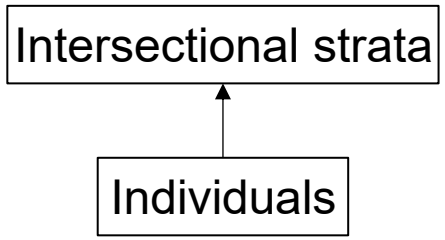
$$y_{ij} \sim \text{Bernoulli}(\pi_j)$$

$$\text{logit}\left(\pi_j\right) \equiv \log\left(\frac{\pi_j}{1-\pi_j}\right) = \underbrace{\beta_0}_{\text{Fixed}} + \underbrace{u_j}_{\text{Random}}$$

$$u_j \sim N(0, \sigma_u^2)$$

$$\text{VPC} = \frac{\sigma_{u\,(\text{Model A})}^2}{\sigma_{u\,(\text{Model A})}^2 + 3.29} \text{ x } 100$$

I-MAIHDA Modelling



- **Model B: Main effects model**
  - Simultaneously adjust for variables making up the intersectional strata as main effects

$$\text{logit}\,(\pi_j) \equiv \log\left(\frac{\pi_j}{1-\pi_j}\right) = \underbrace{\beta_0 + \beta_1 x_{ij} + \cdots + \beta_p x_{pj}}_{\text{Fixed}} + \underbrace{u_j}_{\text{Random}}$$

u_j ~ interaction effects

  - **VPC**: % between-stratum variance attributable to interaction effects
  - Proportional change in variance (**PCV**): % reduction in between-stratum variance attributable to additive main effects

$$\text{PCV} = \frac{\sigma^2_{u(\text{Model A})} - \sigma^2_{u(\text{Model B})}}{\sigma^2_{u(\text{Model A})}}$$

  - Estimate **predicted probability**

$$\pi_{ij} = \text{logit}^{-1}\left(\beta_0 + \beta_1 x_{ij} + \cdots + \beta_p x_{pj} + u_j\right)$$

    - *Decompose total effects into main effects and interaction effects*

The Problem of Aggregation:
And how it presents in an Intersectional Model.

Sex

Age

UK nationality

"Asian", "Black", "White"," Mixed", "Other"

Ethnicity

21 Categories

Experience of Racism

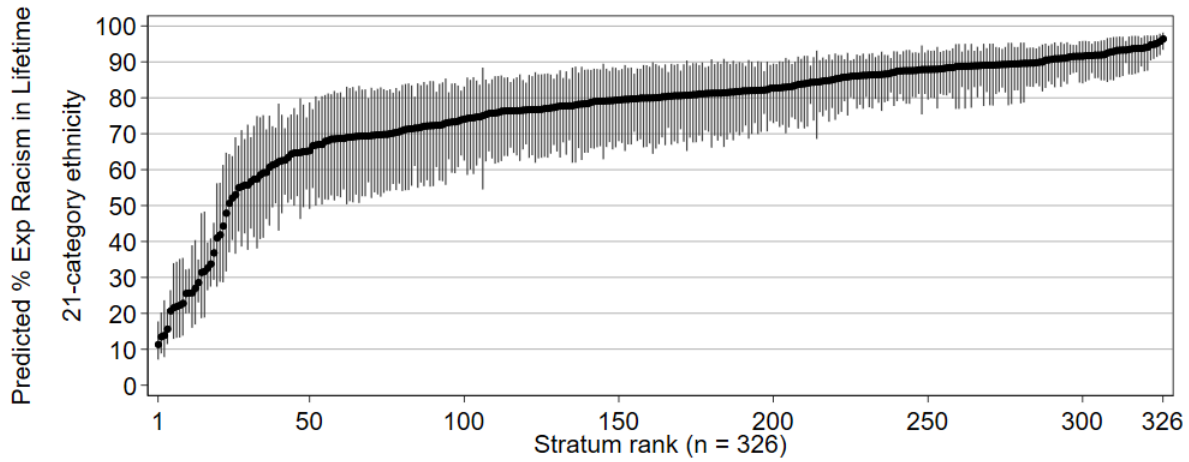
Intersectional Relationships between Age, Sex, Ethnicity, Nationality and Experience of Racism in the UK Using Different Ethnicity Categorisations: A Comparative Study Using Survey Data

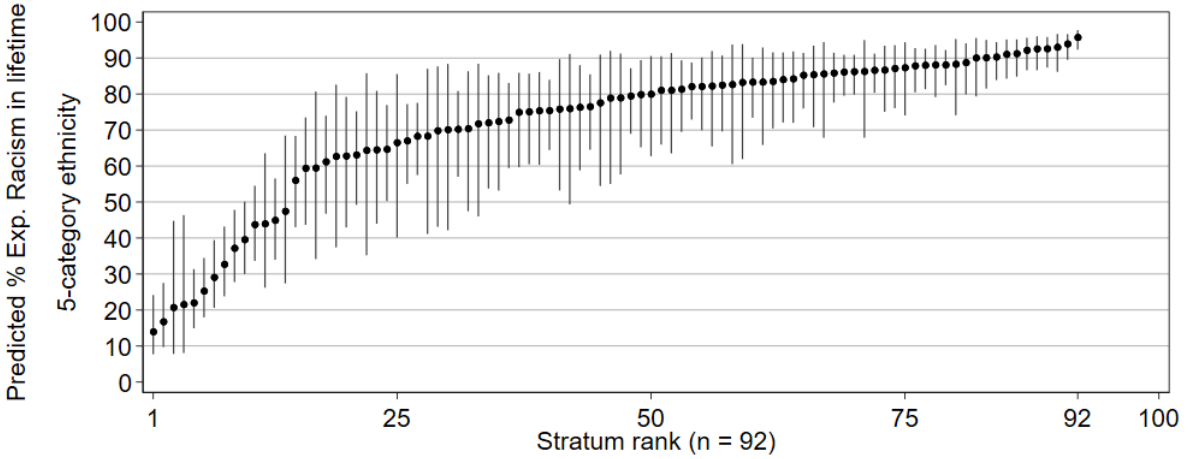
Submitted to Journal of Migration and Health

Predicted percentage of people experiencing racism, by stratum (age, sex, ethnicity, UK nationality)

Overall, 65% of participants reported experiencing racism in their lifetime

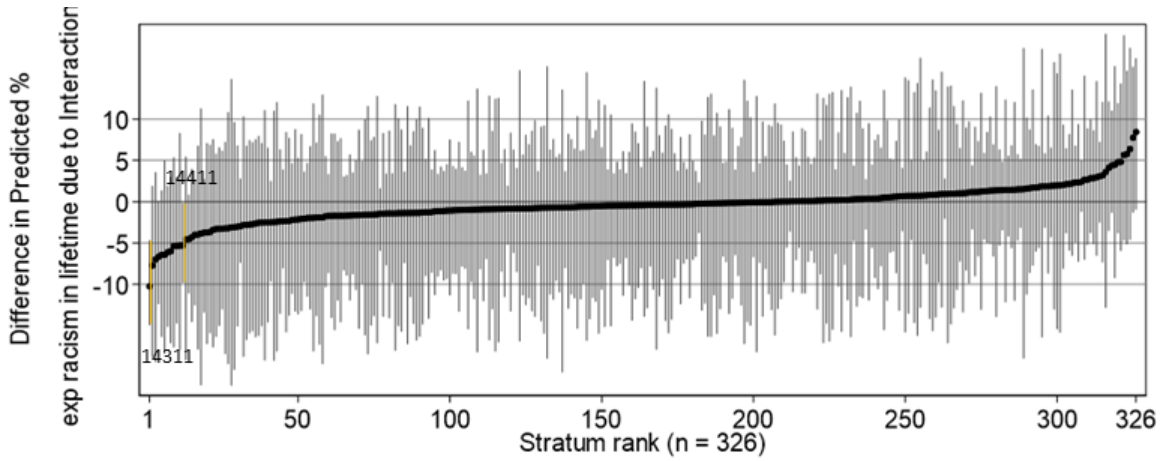
Observation 1: young males and females identifying as having Black Caribbean, Black African (and mixed) ethnic backgrounds, regardless of their UK nationality, have disproportionately high lifetime prevalence of lifetime racism.





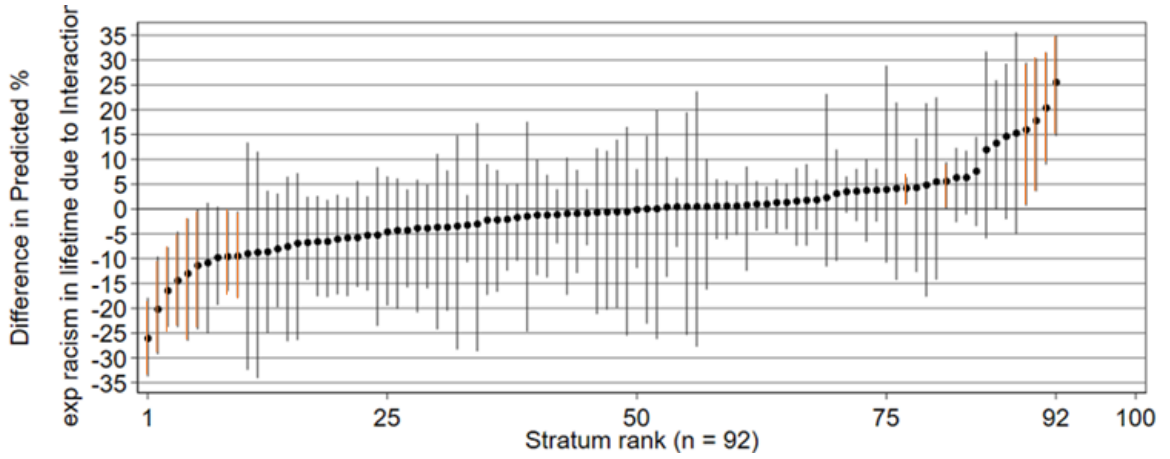
Main Comparison

		21-category Ethnicity				5-category Ethnicity			
		Logistic Model 1A		Logistic Model 1B		Logistic Model 2A		Logistic Model 2B	
		OR	95% CI	OR	95% CI	OR	95% CI	OR	95% CI
Variance Partition Coefficient (%)		25.9		2.4		26.20		7.55	
Proportional Change in Variance (%)				92.8				77.0	



Observation 2: protective effect of Not being a UK national to experiencing lifetime racism

Observation 3: Interaction in coarse > granular ethnicity



Discussion

1) Inflation of observed intersectionality in 5-category model

2) Protective effect of non-UK national driven by “White: Other White Backgrounds” – did not extend to Black African & Caribbean groups

3) Are more disaggregated categories always better? → Effects may operate at different levels of aggregation.

Similar Study:

Education

Age

US born

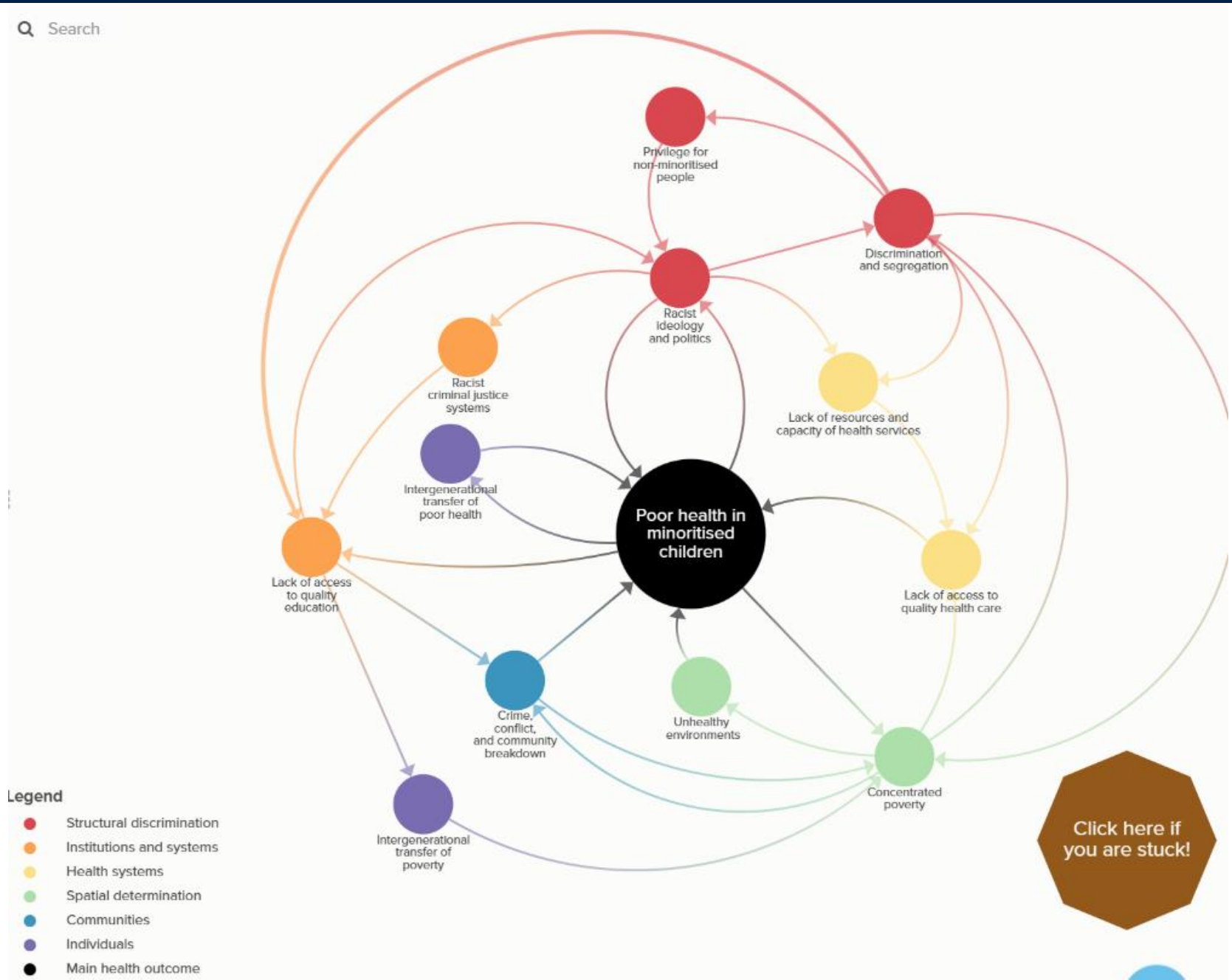
5 Categories:
“Non-Hispanic White”,
“Non-Hispanic Black”,
“Asian/Native Hawaiian & Pacific Islander”,
“Other Non-Hispanic”,
“Hispanic/Latino”

Race and Ethnicity

11 Categories:
Broken down “Hispanic/Latino” to
7 subgroups

Infant Birth Weight

Borrell, Nieves & Evans (2025) – SSM-Population Health



Lancet Commission on Global Racism and Child Health (under review)

Thank you to all EVENS participants.
Thank you to EVENS group & UK Data Service.
Thank you to Evans, Bell & other methodologists.

Pre-print