Working with the National Pupil Database in the ONS Secure Research environment

Alexey Bessudnov

IDSAI seminar, University of Exeter

2022-10-26

Why secure data?

- ▶ Open access data in the UK Data Service (https://ukdataservice.ac.uk)
- ▶ Some data cannot be in open access for data protection reasons (both administrative and survey data)

Examples of secure data

- Annual Population Survey
- 2011 census data (a 10% sample)
- National Pupil Database
- Growing Up in England (National Pupil Database linked with the census)
- NHS Test and Trace
- See https://ons.metadata.works/domain/index.html

Secure data research infrastructure

- ONS Safe Rooms (London, Titchfield, Newport)
- SafePods (a number of locations, including Exeter)
- Workplace access (AOC - Assured Organisation Connectivity), including remote access
- To get access you need to be accredited as a researcher and pass training
- Access is given for a research project (needs to be approved), specific research questions, specific variables in specific data sets
- Outputs are subject to statistical control

National Pupil Database

- ▶ Maintained by the Department for Education
- ▶ Data from annual school censuses (state schools only)
- ▶ Attainment data (key stage 1 to 5 tests)
- ▶ Exclusions, absences, pupil referral units etc.
- ▶ Each pupil has a unique ID that allows to link data over time

Question

- ▶ The dynamic of ethnic gaps in school attainment (test scores) in England
- ▶ The discussion of ethnic inequalities in education is often framed in terms of disadvantage of ethnic minorities
- ▶ But we know that some ethnic minorities outperform the White majority in education
- ▶ This area is under explored (although there are some studies) as
  - ▶ the data are rarely available (survey samples are not large enough and are based on self-reporting)
  - ▶ the topic is politically sensitive
  - ▶ Some data are published by the government: https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training

Data

- The annual spring school census data (for the data on ethnicity) for 2007 and 2018
- The school census includes every pupil in state schools in England (therefore excluding 6-7% of pupils in independent schools)
- Key stage 2 reading and maths attainment data (Year 6, age 10-11)
- About 588,000 children in 2007 and 620,000 children in 2018

Method

- ▶ Descriptive analysis, mostly with visualisations
- ▶ Focus on the entire distributions rather than the means (as the data size allows for this)
- ▶ Relative distribution (a method for comparing two distributions) (Handcock and Morris 2006)
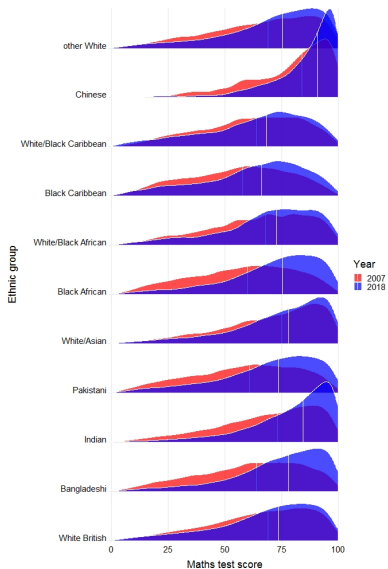
Ethnic distributions in 2007 and 2018 (Year 6)

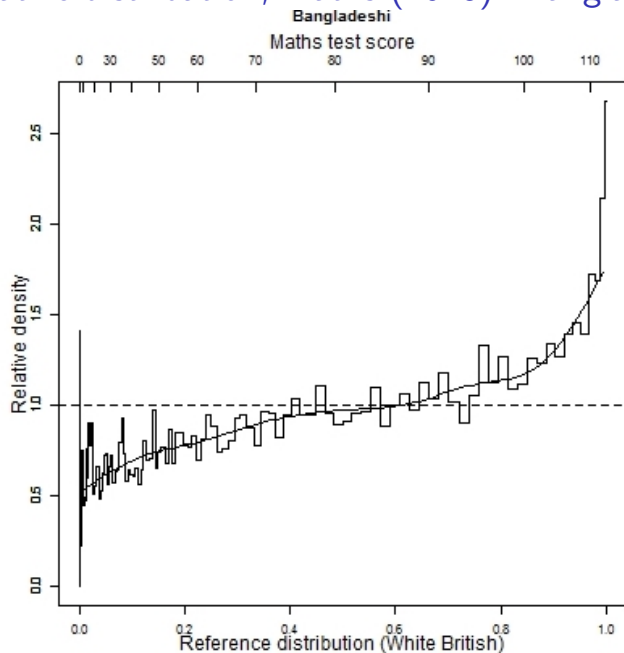| Ethnic group	2007	2018
White British	448,344	416,495
Pakistani	17,352	27,312
other White	15,102	38,459
Black African	14,499	25,202
Indian	13,140	18,162
Black Caribbean	8,403	7,234
Bangladeshi	7,494	10,959
White/Black Caribbean	6,830	9,601

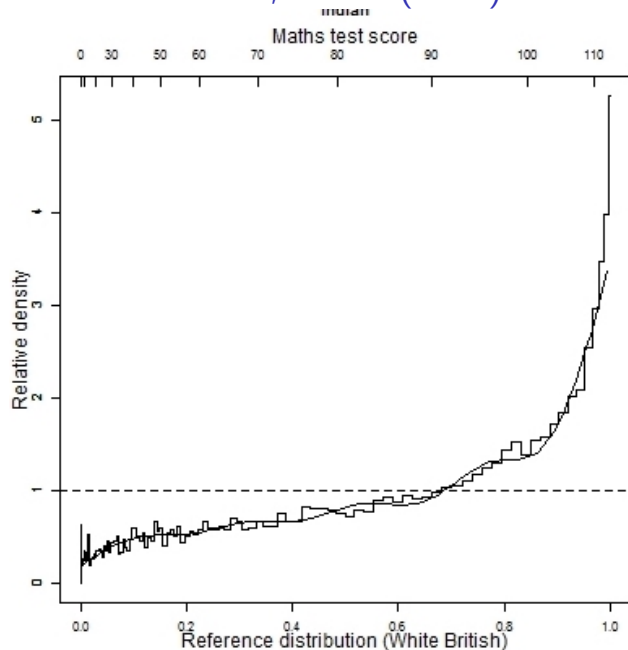# Test score distributions: Reading
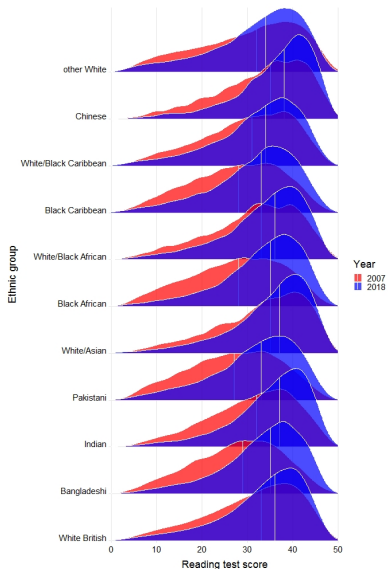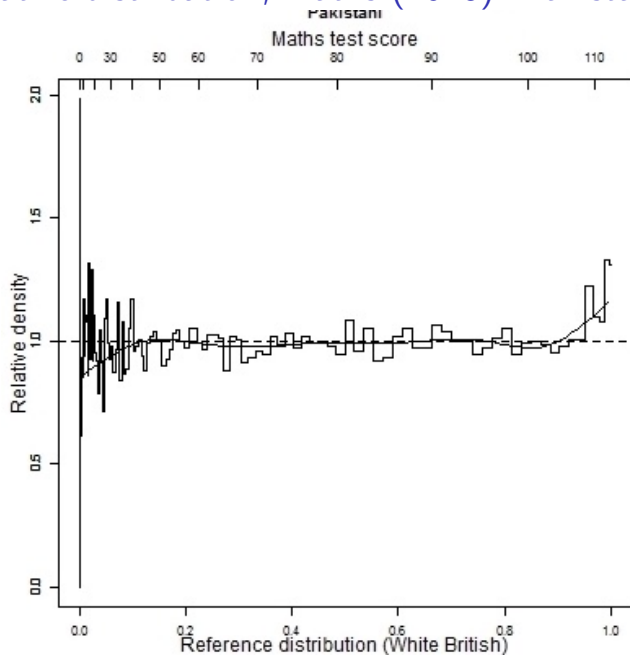
Test score distributions: Mathematics

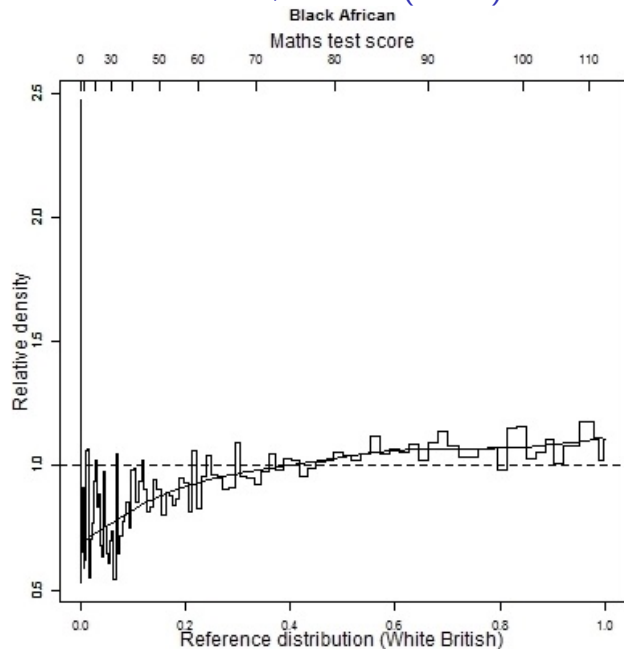# Relative distribution, maths (2018): Bangladeshi
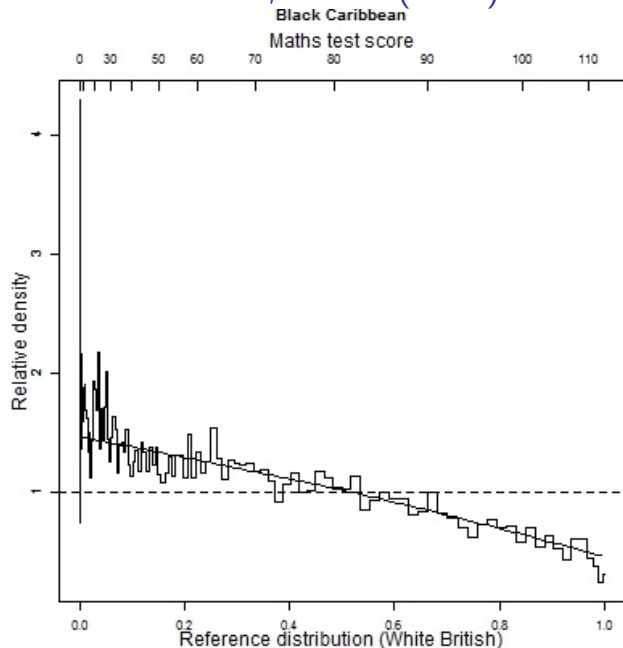
Relative distribution, maths (2018): Indian

Relative distribution, maths (2018): Pakistani

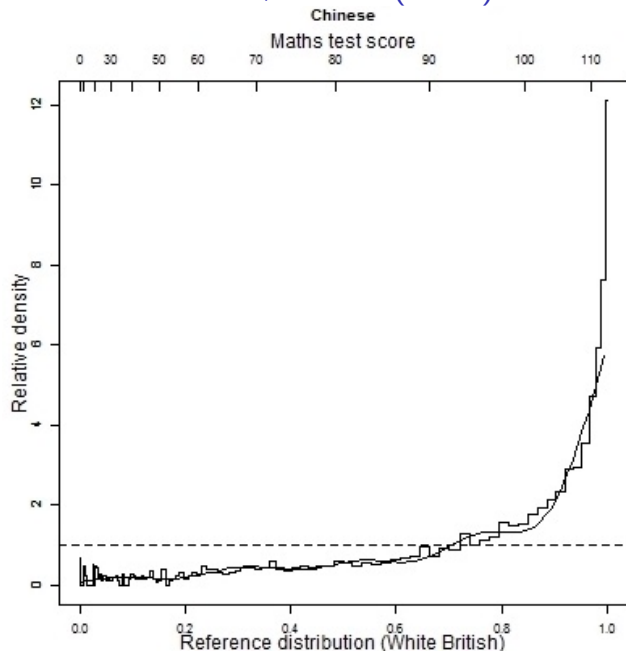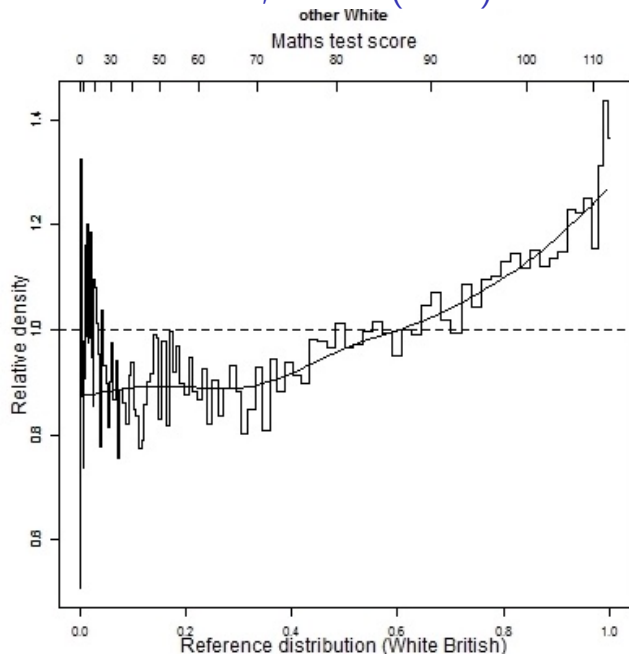# Relative distribution, maths (2018): Black African

Relative distribution, maths (2018): Black Caribbean

Black Caribbean

Relative distribution, maths (2018): Chinese

Relative distribution, maths (2018): Other White

Reading test scores (2018)

- I am not showing the graphs to save time
- Generally, there are fewer differences across ethnic groups compared to maths scores
- Compared to the White British:
  - Bangladeshi and Black Africans are performing at the same level (but less likely to be at the bottom and at the very top of the distribution)
  - Indian and in particular Chinese are performing better
  - Black Caribbean, Pakistani and other Whites are performing worse

The White/non-White proportion across the test distributions (2018)
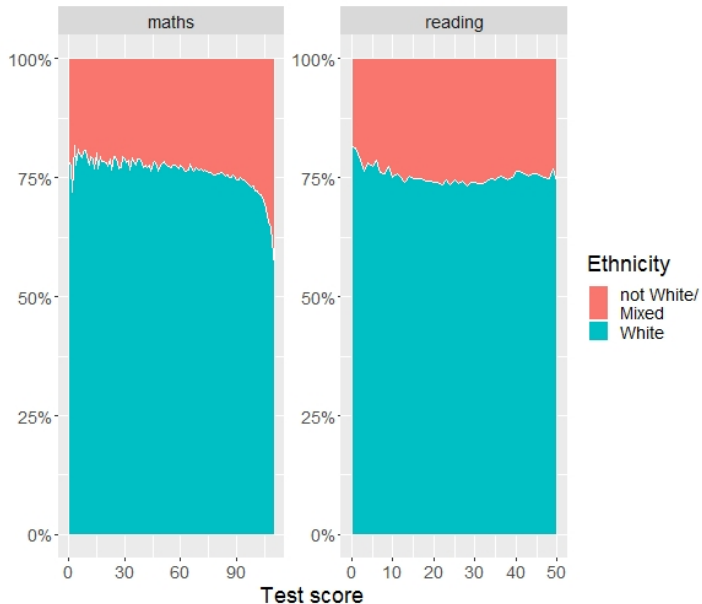
Summary of results

- There is no White advantage in the KS2 test results
- The White British are one of the worst performing groups, especially in maths
- Chinese and Indian pupils perform best across the distribution, both in reading and maths

Limitations

- ▶ Independent schools are missing from the data (but note that the proportion of ethnic minority students in independent schools is about the same as in state schools)
- ▶ The definition and measurement of ethnicity

Explanations

- ▶ Why do we observe ethnic differences?

Explanations

- ▶ Why do we observe ethnic differences?
- ▶ Differential treatment in schools, i.e. discrimination (unlikely: compare the Black Africans and Black Caribbean)

Explanations

- ▶ Why do we observe ethnic differences?
- ▶ Differential treatment in schools, i.e. discrimination (unlikely: compare the Black Africans and Black Caribbean)
- ▶ Socio-economics conditions (but British Indians and Chinese do not come from more advantageous social background)

Explanations

- ▶ Why do we observe ethnic differences?
- ▶ Differential treatment in schools, i.e. discrimination (unlikely: compare the Black Africans and Black Caribbean)
- ▶ Socio-economics conditions (but British Indians and Chinese do not come from more advantageous social background)
- ▶ "Culture" (but where does the culture come from?)

Explanations

- ▶ Why do we observe ethnic differences?
- ▶ Differential treatment in schools, i.e. discrimination (unlikely: compare the Black Africans and Black Caribbean)
- ▶ Socio-economics conditions (but British Indians and Chinese do not come from more advantageous social background)
- ▶ "Culture" (but where does the culture come from?)
- ▶ Inherent group characteristics (territory of danger, no or little empirical evidence)

Explanations

- ▶ Why do we observe ethnic differences?
- ▶ Differential treatment in schools, i.e. discrimination (unlikely: compare the Black Africans and Black Caribbean)
- ▶ Socio-economics conditions (but British Indians and Chinese do not come from more advantageous social background)
- ▶ "Culture" (but where does the culture come from?)
- ▶ Inherent group characteristics (territory of danger, no or little empirical evidence)
- ▶ Selective migration to the UK (my favourite explanation, but needs to be shown empirically)