

# Quality of Administrative Data: A new Application of Representativity-Indicators

**Nina Sommerland, Kim Warne**

Methodologist

Methodological Research hub

Methodology and quality directorate

**17 May 2024**



# Acknowledgement

- The original method and code developed by Professor Natalie Shlomo and Dr Sook Kim of Manchester University, was developed and tested in RStudio and is available on the GitHub.
- <https://github.com/sook-tusk/qualadmin>

# Presentation overview

1. What are R-indicators?
2. Trial application to administrative data in Office for National Statistics.
3. Future research and developments.

# ONS research objectives

- The UK does not have a population register.
- Research in Office for National Statistics (ONS) exploring **quality indicators for administrative data**.
- How do we know that the administrative data is representative of the population?
- Currently published a [working paper](#) on the ONS website, but research is still ongoing.

# Why use R-indicators?

- R-Indicators quantify how **representative** a dataset is for select categorical variables of the population.
- Needs **comparative population information**, can come from a census or a probability-based reference sample.
- Previously used for non-response in surveys, but has been developed for measuring **representativeness in administrative data**.

# Our application

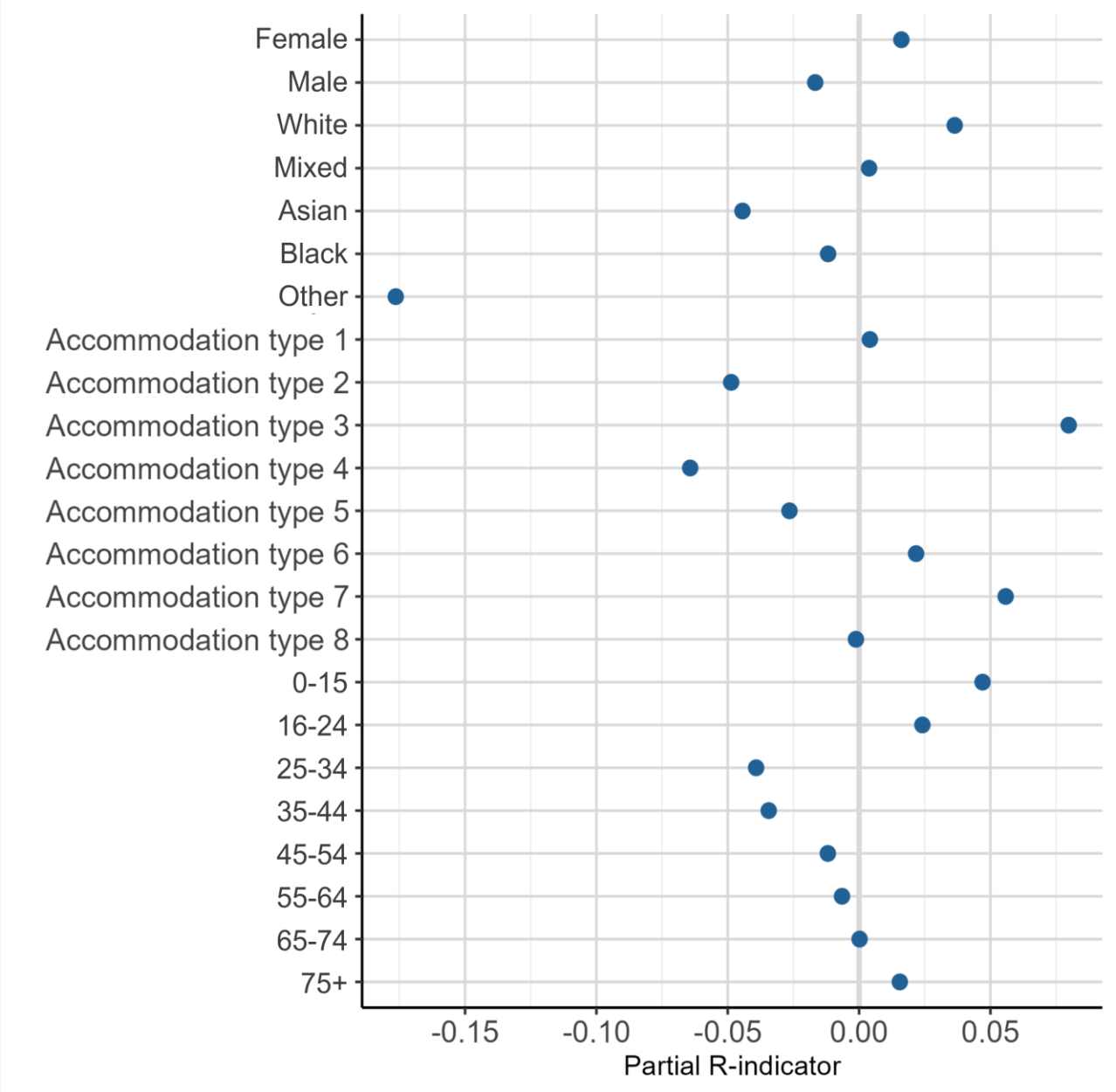
- Population: A local authority in the UK.
- Admin data: Composite dataset of admin data with, using the variables age group, sex, ethnicity and accommodation type.
- Auxiliary population information obtained from Census 2021.

# Method specification

- The model quantifies the **expectation of a record (person) being included in the administrative dataset**, given a certain variable category.
- The partial R-indicator measures the amount of **variation of the participation propensities** between the categories of a variable.
- The larger the between-category variation is, the stronger the impact of the variable on a lack of representativeness.

# Result partial R-indicators (ONS, 2023)

- Ethnicity “Other” is underrepresented in the admin data.
- "Accommodation type 3" is slightly overrepresented.
- Overall results shows a promising use for R-indicators for admin data.





# Conclusion and future developments

- We recommend further research into both R-indicators to develop **thresholds** that indicate when any lack of representativeness found is critical, and to what degree.
- More research on **distance metrics** (see working paper).
- We're planning research on **mode effects for surveys**.
  
- Please share with us any methods or applications of interest!

# Full link to working paper

- <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/qualityindicatorsforrepresentativenessinadministrativedataindicatorsanddistancemetrics>