

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

**Nina Sommerland, PhD**

Methodologist

Methodological Research hub

Methodology and quality directorate

**13 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

- 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-Indicator methods quantify how representative a dataset is for select categorical variables of the population.
- Population information can come from a Census or a weighted survey.
- Previously used for missingness in surveys, but has been developed for measuring representativeness administrative data.

# Application

- 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 unconditional 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.

- The estimate for the overall r-indicator for administrative data source is

$$\hat{R}_{\hat{\rho}^P} = 1 - \hat{S}_{\hat{\rho}^P}$$

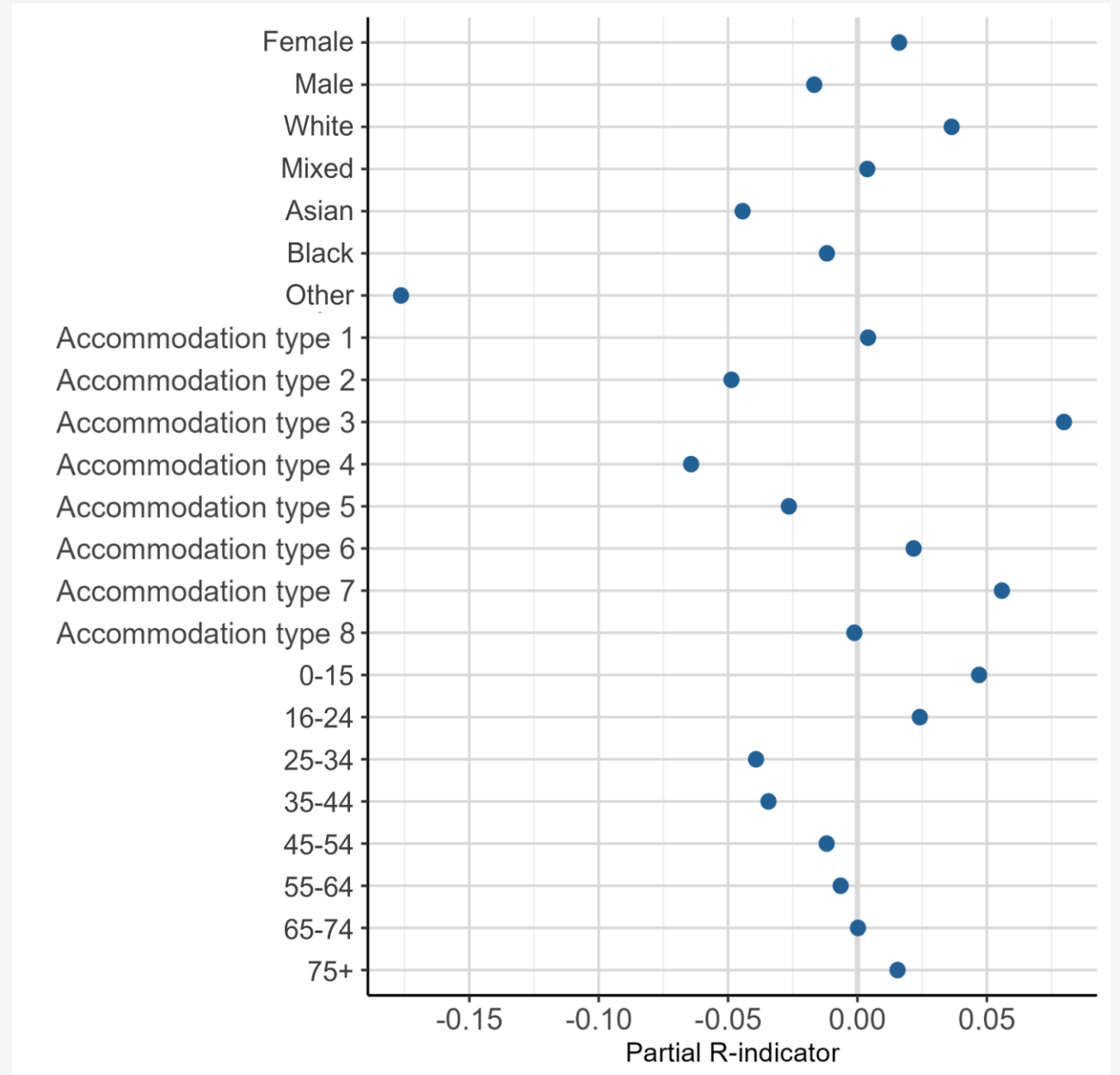
- The estimate for the unconditional partial R-indicator for

variable  $\mathbf{x}_k$  is:  $R_U(\mathbf{x}_k) = \sqrt{\frac{1}{N} \sum_{h=1}^H m_h (\hat{\rho}_h - \hat{\rho})^2}$ . (add citation)



## 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

- To help decision making, 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!