Transparency in epidemiological analyses of cohort data:

A case study of the Norwegian Mother, Father, and Child cohort study (MoBa)

Timo Roettger[3], Adrian Dahl Askelund[1,2], Viktoria Birkenæs[5], Ludvig Daae Bjørndal[3], Agata Bochynska[3], Bernt Damian Glaser[2,3], Tamara Kalandadze[6], Max Korbmacher[4,7,8]*, Ivana Malovic[3], Julien Mayor[3], Pravesh Parekh[5], Daniel S. Quintana[3,9,10] and Laurie J. Hannigan[1,2,11]

1Psychiatric Genetic Epidemiology group, Research Department, Lovisenberg Diaconal Hospital, Oslo, Norway. 2PsychGen Centre for Genetic Epidemiology and Mental Health, Norwegian Institute of Public Health, Oslo, Norway. 3University of Oslo, Oslo, Norway. 4Department of Neurology, Haukeland University Hospital, University of Bergen, Bergen, Norway. 5Centre for Precision Psychiatry, Division of Mental Health and Addiction, University of Oslo and Oslo University Hospital, Oslo, Norway. 6Østfold University College, Halden, Norway. 7Department of Health and Functioning, Western Norway University of Applied Sciences, Førde, Norway. 8Mohn Medical Imaging and Visualization Centre (MMIV), Bergen, Norway. 9KG Jebsen Center for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway. 10Oslo University Hospital, Oslo, Norway. 11Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

Background

- Secondary analysis of cohort data is an important tool in epidemiological research
- Practices to ensure the transparency and reproducibility of secondary data analysis are paramount to ensuring the robustness and usefulness of findings
- The uptake of such practices has not yet been systematically assessed

Approach

- We i) assess the prevalence of the reproducible practices in publications from a specific cohort study (MoBa) between 2007-2023 and ii) – based on the results and current standards for reproducibility – perform exemplar MoBa analyses and demonstrate recommended practices for reproducible reporting

i) Systematic assessment of reproducible and transparent practices in secondary data analysis using the MoBa cohort

Methods

- Design and data collection plan was uploaded as a preregistration prior to data collection (https://osf.io/ef4t5/)
- Data extraction based on abstracts and full-texts, assessing preregistration, accessibility, data and code sharing, conflicts of interests, and other criteria

Results

- 1005 articles that have used the MoBa dataset (obtained from the MoBa website) coded by six raters (91% agreement across criteria)
- Of those articles that were in principle accessible, 733 articles were coded as presenting empirical data and using the MoBa corpus (91.7%).
- Figure 1 summarises the results in relation to these articles' year of publication. Green represents transparent practices, and grey represents non-transparent practices.

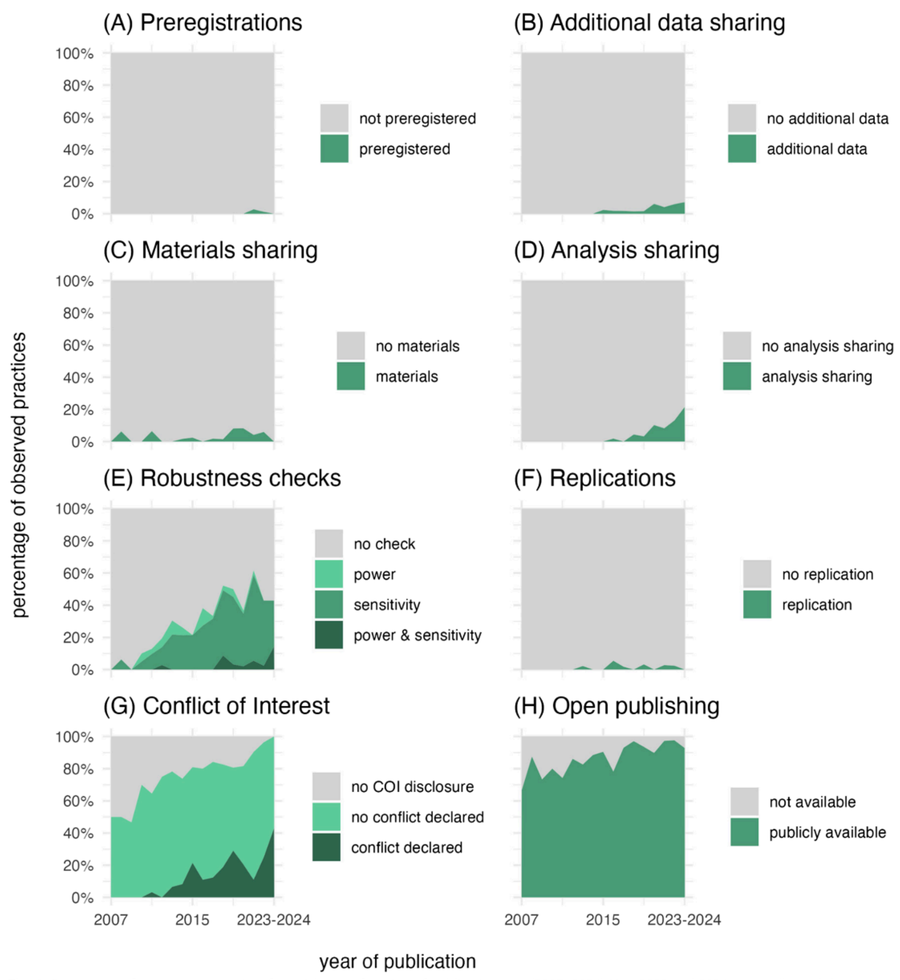

Fig. 1 Percentage of transparent practices for articles as a function of publication year (2007–2024). Year 2023 and 2024 were collapsed due to a small number of *n*

ii) Exemplar analyses

Methods

- We simulated data based on the characteristics of real MoBa data in order to present a realistic example of a reproducibly planned and reported MoBa analysis

Results

- In our hypothetical research project, researchers were interested in testing the independent effects of age (M = 3.46 years, SD = 2.76) and breastfeeding duration (M = 5.47 months, SD = 3.73) on height measured on six occasions during childhood in MoBa and assessing whether these effects vary according to sex (a categorical variable with levels: "Female" and "Male")
- The key features that we emphasise – providing worked examples that can be adapted and re-used by researchers performing secondary data analysis – are:
  - Preregistration
  - Sharing of well-annotated analytic code
  - Detailed reporting of deviations and unregistered steps
  - Robustness checks
  - Sharing of synthetic data to facilitate computational reproducibility

Conclusions

- Some reproducible practices are more common than others, with some practices being virtually absent
- In line with a broader shift towards open science, we observed an increasing use of reproducible research practices in recent years.
- The large amount of analytical flexibility offered by cohorts such as MoBa places additional responsibility on researchers to adopt such practices with urgency
- We demonstrate by example that challenges in implementing reproducible research practices in analysis of secondary cohort data—even including those associated with data sharing—can be meaningfully overcome.

Open access paper

Roettger, T., Askelund, A. D., Birkenæs, V., Bjørndal, L. D., Bochynska, A., Glaser, B. D., Kalandadze, T., Korbmacher, M., Malovic, I., Mayor, J., Parekh, P., Quintana, D. S., & Hannigan, L. J. (2025). Transparency in epidemiological analyses of cohort data a case study of the Norwegian mother, father, and child cohort study (MoBa). BMC Medical Research Methodology, 25(1), 171. https://doi.org/10.1186/s12874-025-02601-6

moba

Lovisenberg Diaconal Hospital

NIPH
PSYCHGEN CENTRE FOR GENETIC EPIDEMIOLOGY AND MENTAL HEALTH

Contact: laurie.hannigan@lds.no