Pharmacoepidemiology and Drug Safety

WILEY

LETTER TO THE EDITOR

Ten Must-Read Papers on Transparency and Reproducibility in Pharmacoepidemiology

Anton Pottegård

Clinical Pharmacology, Pharmacy and Environmental Medicine, University of Southern Denmark, Odense, Denmark

Correspondence: Anton Pottegård (apottegaard@health.sdu.dk)

Received: 14 November 2024 | **Revised:** 31 January 2025 | **Accepted:** 12 February 2025

Funding: The author received no specific funding for this work.

We hear this all the time: Real-world evidence is changing how we make decisions, yet there is still hesitation around trusting these studies. What can we do to remedy this?

I think the answer lies in pushing our field towards a more transparent future, where study design and implementation are better documented to allow not only regulators but also our fellow academics to scrutinize our work.

So, what can we do? We as pharmacoepidemiologists must champion this cause ourselves, and we can do this by pushing to increase the transparency and reproducibility of our own work.

With this article, I provide my suggestion for a brief "transparency and reproducibility core readings." The list comprise 10 papers, subjectively selected from the rapidly expanding literature, that represent both best practices and new developments within the domains of transparency and reproducibility in pharmacoepidemiological studies.

1 | Documentation of the Need for Reproducibility and Transparency

Wang et al. Nat Commun. 2022 Aug 31;13(1):5126 [1].

The seminal REPEAT study documents the need to do better in terms of transparency and reproducibility through a large-scale replication effort of 150 database studies in top clinical and epidemiology journals—using the same healthcare databases as original investigators. While most results were closely reproduced, a subset was not. The paper also highlights areas where greater methodological transparency could further improve reproducibility and validity assessment, such as the temporality of measurement of key study characteristics, code algorithms used to define and characterize the population, and operational algorithms used to define the duration of follow-up and censoring criteria.

2 | Design and Analysis Framework

Desai et al. BMJ. 2024 Feb 12:384:e076460 [2].

The PRINCIPLED framework proposes a stepwise process to systematically consider key choices for study design and data analysis to foster the generation of reliable and reproducible evidence. These steps include (1) formulating a well-defined causal question via specification of the target trial protocol; (2) describing the emulation of each component of the target trial protocol and identifying fit-for-purpose data; (3) assessing expected precision and conducting diagnostic evaluations; (4) developing a plan for robustness assessments including sensitivity analyses and quantitative bias analyses; and (5) inferential analyses. Additional inspiration for structuring the workflow around protocol development is provided by Muntner et al. [3] also outlining a stepwise process with data checks and feasibility assessments conducted within a "clean room" where restricted access to data allows preliminary analyses to be conducted without revealing how decisions affect the subsequent analysis. This operationalizes the increasing recognition of the value of documented data checks during protocol development [4].

3 | Description of Data Source

Gini et al. Pharmacoepidemiol Drug Saf. 2024 May;33(5):e5787 [5].

Transparent reporting of the characteristics of the data sources used in a given study is essential to interpret study findings and, in particular, differences in findings between different settings. The DIVERSE project reviewed current practices for reporting data source characteristics and identified nine domains that are used to describe data sources, including for example, data originator, data dictionary, and healthcare system and culture. These domains provide useful guidance on aspects to consider when describing the data sources used in a study.

4 | Protocol Template

Wang et al. Pharmacoepidemiol Drug Saf. 2023 Jan;32(1):44–55 [6].

The HARPER protocol template was developed by a joint task force between the Professional Society for Health Economics and Outcomes Research (ISPOR) and the International Society for Pharmacoepidemiology (ISPE). It builds off existing templates and is designed to create a shared understanding of intended scientific decisions through a common text, tabular, and visual structure. The template provides a set of core recommendations for clear and reproducible RWE study protocols and is intended to be used as a backbone throughout the research process from developing a valid study protocol to registration, through implementation and reporting on those implementation decisions.

5 | Creating and Sharing Codelists

Matthewman et al. NIHR Open Res. 2024 Sep 18:4:20 [7].

Codelists are used to operationalize the definition of key study parameters such as exposure, population, outcome, and confounders when conducting studies with routinely collected health data. This paper provides a 'best practice' framework in codelist development and further emphasizes the value of sharing codelists once developed. While it will not always be feasible to adhere to the full framework, it provides useful inspiration to adopt a more rigorous practice when defining study variables.

6 | Protocol Pre-Registration

Orsini et al. Value Health. 2020 Sep;23(9):1128–1136 [8].

This paper argues that study protocols, in particular for hypothesis-evaluating treatment effectiveness (HETE) studies, should be pre-registered. Such pre-registration facilitates the transparent reporting not only of study planning and implementation but also of the rationale for subsequent amendments to the protocol. This recommendation led to the establishment of the Real-World Evidence Registry within the Open Science Framework, publicly available at osf.io/registries/rwe, providing an open repository to share study resources such as protocols, data, and analytical code.

7 | Visualizations

Gatto et al. Pharmacoepidemiol Drug Saf. 2022 Nov;31(11):1140–1152 [9].

This paper provides guidance on how to use visualizations throughout the life cycle of a pharmacoepidemiology study, from the initial study design to the final report. A list of specific suggestions for plot types is provided that not only ensures clear communication of study findings but also documents decision-making about study design and implementation, such as the "design diagrams" proposed by Schneeweiss [10]. An easy-to-use creator of these diagrams can be found at presc.sdu.dk/repeat-diagrams [11].

8 | Code Sharing

Tazare et al. Pharmacoepidemiol Drug Saf. 2024 Sep;33(9):e5856 [12].

This study documents that programming code sharing is very rare (although slightly increasing) in pharmacoepidemiology studies. It also provides recommendations for sharing programming code, including the use of permanent digital identifiers, appropriate licenses, and, where possible, adherence to good software practices around the provision of metadata and documentation, computational reproducibility, and data privacy.

9 | Reporting Guidelines

Langan et al. BMJ. 2018 Nov 14:363:k3532 [13].

Transparent reporting requires you to ensure that you have got all the necessary pieces of information in your manuscript. This article describes the RECORD-PE checklist (also available on www.record-statement.org) and explains each checklist item with examples of good reporting on pharmacoepidemiological research using non-randomized, routinely collected data.

10 | Transparency Statement

Wang and Pottegård. Am J Epidemiol. 2024 May 23:kwae087 [14].

When transparency and reproducibility was a priority in your study conduct, you should proudly highlight these efforts. This paper proposes a framework for an explicit transparency statement that declares the level of transparency a given RWE study has achieved across five key domains: (1) protocol, (2) pre-registration, (3) data, (4) code sharing, and (5) reporting checklists. The framework allows researchers to declare and display the levels to which they have built transparent and reproducible research practices into their studies. Further, the framework can serve as inspiration for editors of scientific journals as to what they can require of authors reporting on real-world studies.

As an example, the framework was recently added to the author instructions of Pharmacoepidemiology and Drug Safety.

Towards More Trustworthy Real-World Evidence

The transparency and reproducibility agenda has received increasing attention in recent years. Nevertheless, I believe that we are still in the early days of formalizing best practices on these aspects in pharmacoepidemiology.

New practices and standards will certainly be established over the next few years. Likely (and hopefully) the list above will be outdated as new papers push this agenda forward. Thus, it is not my intention to provide a final list of resources. Rather, I hope that this list can be used to make this agenda more accessible and accelerate its adoption.

The full implementation of study conduct and reporting that supports transparency and reproducibility will require other stakeholders such as regulators and scientific journals to establish incentives that promote best practice. Meanwhile, we as researchers must develop the tools and practices needed and support the new generation of pharmacoepidemiologists who, through the adoption of these practices, will show the way towards a future with more transparent, reproducible, and thus trustworthy real-world evidence generation.

Acknowledgments

I would like to thank Christopher T. Rentsch, Juan Hincapie-Castillo, and Rishi J. Desai for valuable comments and suggestions to the list and its presentation.

Conflicts of Interest

The author declares no conflicts of interest.

Anton Pottegård

References

1. S. V. Wang, S. K. Sreedhara, S. Schneeweiss, and REPEAT Initiative, "Reproducibility of Real-World Evidence Studies Using Clinical Practice Data to Inform Regulatory and Coverage Decisions," *Nature Communications* 13, no. 1 (2022): 5126.

2. R. J. Desai, S. V. Wang, S. K. Sreedhara, et al., "Process Guide for Inferential Studies Using Healthcare Data From Routine Clinical Practice to Evaluate Causal Effects of Drugs (PRINCIPLED): Considerations From the FDA Sentinel Innovation Center," *BMJ* 384 (2024): e076460, https://doi.org/10.1136/bmj-2023-076460.

3. P. Muntner, R. K. Hernandez, S. T. Kent, et al., "Staging and Clean Room: Constructs Designed to Facilitate Transparency and Reduce Bias in Comparative Analyses of Real-World Data," *Pharmacoepidemiology and Drug Safety* 33, no. 3 (2024): e5770, https://doi.org/10.1002/pds.5770.

4. S. V. Wang and S. Schneeweiss, "Data Checks Before Registering Study Protocols for Health Care Database Analyses," *JAMA* 331, no. 17 (2024): 1445, https://doi.org/10.1001/jama.2024.2988.

5. R. Gini, R. Pajouheshnia, H. Gardarsdottir, et al., "Describing Diversity of Real World Data Sources in Pharmacoepidemiologic Studies: The DIVERSE Scoping Review," *Pharmacoepidemiology and Drug Safety* 33, no. 5 (2024): e5787, https://doi.org/10.1002/pds.5787.

6. S. V. Wang, A. Pottegård, W. Crown, et al., "HARmonized Protocol Template to Enhance Reproducibility of Hypothesis Evaluating Real-World Evidence Studies on Treatment Effects: A Good Practices Report of a Joint ISPE/ISPOR Task Force," *Pharmacoepidemiology and Drug Safety* 32, no. 1 (2023): 44–55.

7. J. Matthewman, K. Andresen, A. Suffel, et al., "Checklist and Guidance on Creating Codelists for Routinely Collected Health Data Research," *NIHR Open Research* 4 (2024): 20.

8. L. S. Orsini, M. Berger, W. Crown, et al., "Improving Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing—Why, What, and How: Recommendations and a Road Map From the Real-World Evidence Transparency Initiative," *Value in Health* 23, no. 9 (2020): 1128–1136.

9. N. M. Gatto, S. V. Wang, W. Murk, et al., "Visualizations Throughout Pharmacoepidemiology Study Planning, Implementation, and Reporting," *Pharmacoepidemiology and Drug Safety* 31, no. 11 (2022): 1140–1152.

10. S. Schneeweiss, J. A. Rassen, J. S. Brown, et al., "Graphical Depiction of Longitudinal Study Designs in Health Care Databases," *Annals of Internal Medicine* 170, no. 6 (2019): 398.

11. L. C. Lund, J. Hallas, and S. V. Wang, "Online Tool to Create Publication Ready Graphical Depictions of Longitudinal Study Design Implemented in Healthcare Databases," *Pharmacoepidemiology and Drug Safety* 30, no. 7 (2021): 982.

12. J. Tazare, S. V. Wang, R. Gini, et al., "Sharing Is Caring? International Society for Pharmacoepidemiology Review and Recommendations for Sharing Programming Code," *Pharmacoepidemiology and Drug Safety* 33, no. 9 (2024): e5856.

13. S. M. Langan, S. A. Schmidt, K. Wing, et al., "The Reporting of Studies Conducted Using Observational Routinely Collected Health Data Statement for Pharmacoepidemiology (RECORD-PE)," *BMJ* 14 (2018): k3532.

14. S. V. Wang and A. Pottegård, "Building Transparency and Reproducibility Into the Practice of Pharmacoepidemiology and Outcomes Research," *American Journal of Epidemiology* 193, no. 11 (2024): 1625–1631, https://doi.org/10.1093/aje/kwae087.