American Journal of Epidemiology **Submitted Manuscript**

Special Collection: n/a

Title: Building transparency and reproducibility into the practice of pharmacoepidemiology and outcomes research

Authors: Shirley V Wang, Anton Pottegård

ORCiD IDs: 0000-0001-7761-7090, 0000-0001-9314-5679

Correspondence Address: 1620 Tremont St, Suite 303 Boston, MA 02120 (email: swang1@bwh.harvard.edu)

Joint Authorship: Shirley V Wang (1, 2), Anton Pottegård (3)

Affiliations: 1. Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, Boston, MA; 2. Department of Medicine, Harvard Medical School, Boston, MA; 3. Clinical Pharmacology, Pharmacy and Environmental Medicine, Department of Public Health, University of Southern Denmark, Odense, Denmark

Key words: Transparency, reproducibility, real-world evidence, open science, pharmacoepidemiology

Acknowledgments[1]**:** We would like to acknowledge Daniel Prieto-Alhambra, Vin Lo Re, Miyuki Hsieh, and Asbjørn Hróbjartsson for reading and providing early comments on this work.

Funding: None

Conflict of Interest: SVW reports ad hoc consulting for Veracity Healthcare Analytics, Exponent Inc, and MITRE a Federally Funded Research and Development Center for the Centers for Medicare and Medicaid Services.

Disclaimer: n/a

Data Availability Statement: This commentary does not involve data analysis. Therefore, there is no protocol to pre-register, nor is there code or data to share. There are no applicable reporting guidelines for this commentary. The sources of all open science badges are provided in figure footnotes.

Abstract word count: 100

Manuscript word count: 3,625

[1] Study investigators, conference presentations, preprint publication information, thanks.

Abstract:

Real-world evidence (RWE) studies are increasingly used to inform policy and clinical decisions. However, there remain concerns about the credibility and reproducibility of RWE studies. Observational researchers should highlight the level of transparency of their studies by providing a succinct statement addressing study transparency with the publication of every paper, poster, or presentation that reports on a RWE study. In this paper, we propose 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.

Key messages:

- While there is universal agreement on the critical importance of transparent and reproducible science, the building blocks of open science practice that are common across many disciplines have not yet been built into routine workflows for pharmacoepidemiology and outcomes researchers.
- We propose a framework for explicitly stating the level of transparency that 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 transparency statement outlined in the present paper can be used by research teams to proudly display the open science practices that were used to generate evidence designed to inform public health policy and practice. While transparency does not guarantee validity, such a statement signals confidence from the research team in the scientific choices that were made.

Real-world evidence (RWE) studies are increasingly used to inform policy and clinical decisions. However, there remain concerns about the credibility of RWE studies. One such area of concern, which has been consistently articulated by various stakeholders[1-5], is a need for greater transparency and reproducibility in database study conduct.

Biomedical journals and large decision-making organizations are currently working on and may adopt heterogeneous policies and procedures regarding transparency requirements for database studies. However, investigators do not have to wait for mandates to follow transparent research practices. Rather, they can highlight such practices with each paper that they produce.

In this paper, we propose a framework for explicitly stating the level of transparency that a given RWE study has achieved across five key domains: 1) protocol, 2) pre-registration, 3) data, 4) code sharing, and 5) reporting checklists.

Transparency statement for researchers

Researchers can highlight the transparency of their research by providing a succinct statement addressing the five domains of study transparency with the publication of every paper, poster, or presentation that reports on a database study, e.g., pharmacoepidemiology or health outcomes research.

The proposed transparency statement for researchers is based on the Transparency and Openness Promotion (TOP) guidelines developed a decade ago for journals to encourage more transparent and reproducible scientific practices.[6] TOP has garnered commitment from over 5,000 journals, mostly in the social sciences, to review and implement eight modular open science standards with increasing levels of stringency over time as part of the policies and procedures required for publication. Among these journals, there is increasing use of positive reinforcement methods such as the display of open science badges on publications. [7-9]

The TOP guidelines focus on actions and policies that journals can take. However, the uptake of the TOP guidelines in clinical and epidemiology journals has been slow compared to the social sciences, where a heavily publicized "reproducibility crisis" provided an impetus for rapid change.[10] Therefore, we borrow from the TOP framework and put it in the hands of researchers to declare and display the levels to which they have built transparent and reproducible research practices into each study. Example sentences to build a transparency statement are provided in Figure 1.

In the following sections we discuss why each domain is important, briefly describe current practices, and suggest how to structure a statement that covers each aspect of study transparency, altogether forming a transparency statement for inclusion in the scientific manuscript.

Five building blocks for transparency statement

1. Protocol

A well-documented protocol is essential for all database studies, regardless of whether the study is exploratory or confirmatory, or whether the goal is to conduct descriptive, predictive, or inferential analyses. A thorough study protocol serves several important purposes, describing the research question, study design, data sources, and how study parameters will be measured and analyzed. The study protocol also often serves as a "contract" between lead investigator, collaborators and study staff. When the scope is agreed upon and documented in the protocol, this not only help prevent "scope creep" but also facilitates efficient study conduct by providing clear guidance to the analyst. Finally, concerns about fishing or "data dredging" can also be ameliorated by clear documentation and definition of primary analyses in a protocol.

Protocols will, however, often require updates. Working with secondary data that was not collected for research means that there will almost always be unexpected quirks that will necessitate amendments to the original plan outlined in the protocol. Further, unexpected findings might trigger additional analyses, either as exploratory post hoc analyses or as independent sub-investigations in their own right. The fact that a protocol was drafted should never be used as an argument against improving ones methodology whenever possible or recognizing and pursuing new insights or even new hypotheses in the material. Any such improvements or additions should, however, be transparently argued and documented. Therefore, a good version control system with a clear, contemporaneous record of amendments, including documentation of what, when and importantly why they were made is important. Additionally, a log describing the learnings that led to amendments not only justify the changes to the initial protocol but also document important insights that support future studies in the same database or within the same topic.

While some form of protocol may be used by most investigators, those who review protocols often express that they see a great deal of variation in quality. Further, not all groups within pharmacoepidemiology and outcomes research have established a practice of finalizing study protocols, with a final signoff from all investigators and with documentation of subsequent amendments.

Protocol quality can be raised by using already existing protocol templates that incorporate pharmacoepidemiology good practice guidance. Indeed, post-authorization studies requested by the European Medicines Agency are already required to use such templates. Recently, four existing protocol guidance documents[11-14] were harmonized to incorporate study background/planning/rationale with clear communication of the operational details of implementation that are necessary for reproducibility, which led to the creation of the HARmonized Protocol to Enhance Reproducibility, HARPER,[15] which was endorsed by the International Society of Pharmacoepidemiology and the International Society for Pharmacoeconomics and Outcomes Research .

Transparency statement: If a protocol was not created or is not available, then authors could state so (level 0). If the protocol exists yet is only available upon request, this should be stated (level 1). Note that "available upon request" is not considered sufficient to earn an open protocol badge. If the research team can provide a link to a protocol (including amendments) stored on an open-access registry or open repository that can provide a persistent identifier with time stamps (level 2) then they will have earned an open science badge (**Figure 1**). Note that protocols can be made public outside of the context of pre-registration. The research team can attain an even higher level of transparency if their protocol was developed using a recognized structured template such as HARPER[15] (level 3).

2. Pre-registration

Public registration of database studies has several benefits for the research enterprise as a whole. The key value of pre-registration is publishing a time-stamped protocol. A repository of conducted database studies could reduce research waste and publication bias.[16] Further, the knowledge that a pre-registered protocol will be turned into a public document will encourage a more thorough and thoughtful discussion within the investigative team about planning and documenting analyses.

That said, pre-registration is not of equal importance for all database studies. For studies that aim to evaluate a hypothesis about a treatment effect, pre-registration of a well-documented protocol prior to conduct of inferential analysis can help to address concerns about results-driven analyses or selective presentation of findings.[17] False statements are always possible but may be deterred if research teams are required to sign off on a public attestation of registration prior to inferential analyses. For descriptive, exploratory, or predictive studies, pre-registration is less critical. For these types of studies, however, registration remains a means for promoting open science, by providing a stable repository for publicly shared protocols and making research more accessible. Opponents of pre-registration may fear being "scooped", that pre-registration of a protocol may interfere with publication, or that pre-registration may stifle scientific discovery if results from the initial plan are held to have higher value than unplanned results from analyses that are deviations from the protocol. However, investigators who fear being scooped may choose a study registration site with an embargo feature that allows the date-time stamped material to be preserved in a lock box until the embargo date has passed. With medical journals increasingly supporting publication of papers with pre-prints, a pre-registered protocol is unlikely to interfere with future publication and can be useful supplementary material to reference. For those who fear stifling of scientific discovery by registering a protocol, we refer to the previous section where we discussed the importance of a contemporaneous record of amendments. Again, we emphasize that a pre-registered protocol does not mean that no changes may be made, nor that results obtained after deviations to

the plan should be ignored. In the end, the protocol should provide a roadmap from where the investigators started to where they ended.

Pre-registration of RWE studies is possible via several repositories, with the main options being EU-PAS (https://www.encepp.eu/encepp/studiesDatabase.jsp), Clinicaltrials.gov and the RWE Registry (OSF.io/registries/rwe). Currently, public sharing of database study protocols is very uncommon, with an unknown (but expected to be small) proportion of pharmacoepidemiology and outcome studies being registered. Among studies on the EU-PAS register for observational studies, 57% do not include a protocol.[18] The recently established registry at OSF.io/registries/rwe requires upload of a protocol in order to complete registration, however, it allows the materials to be embargoed while the study is being conducted.[19]

Transparency statement: If the study was not pre-registered then authors could state so (level 0). If the research team did not include a protocol in their registration (e.g. pre-registration provided only title, aims and summary) (level 1), then they may state so. If they registered a full study protocol prior to conducting analyses on an open access registry or repository that can provide a persistent identifier with time stamps (level 2) then they will have earned an open science badge (**Figure 1**).

3. Data

Most database studies use individual-level patient health information such as insurance claims or electronic health record data. While researchers are able to access these data through strict data use agreements, legal restrictions in these agreements will generally prevent sharing of analytic data or derivates thereof. However, there are a couple of alternatives to publicly sharing data for researchers who work with protected access healthcare data. Researchers could provide guidance for other investigators on how to access the source databases underlying their work. Importantly, such guidance must include details on how to contact and contract with the data holder(s), which version/release of the data was accessed and/or the date that data extraction was performed. While this is often prohibited for database studies, when data use contracts permit, analytic datasets could be stored and shared via protected access data repositories.[20] In addition, or alternatively, researchers could provide synthetic data[21] that protects patient privacy, retains the statistical properties of the original data, and facilitates testing and use of shared code. Importantly, any data that is shared must follow FAIR principles (Findable, Accessible, Interoperable, Reusable), including metadata about how to use the shared materials, a persistent unique identifier, clear access processes and procedures that account for data protection and data privacy requirements, among other characteristics.[22]

Transparency statement: If data are not available, then authors could state so (level 0). If data are only available upon request, this should be stated (level 1). As before, stating "available upon request" is not considered sufficient to earn an open science badge. If the research team provides a link to de-identified data stored on

an open access repository OR if they provide an appendix with detailed information on contacts, contracting process, and version/ETL OR provide access to a synthetic dataset along with analysis code (level 2) then they will have earned an open science badge for protected access data (**Figure 1**). Without these details, i.e., when only referencing the data holder by vendor name, it will in most cases be impossible to recreate that data request and therefore impossible to reproduce the analysis.

4. Code

While a well-documented protocol is central to the transparency and reproducibility of a study, computational reproducibility generally further requires the sharing of data with analytic code and other supporting materials such as a readme file, data dictionary, and protocol. Such materials provide the steps needed to exactly reproduce the registered study.

Sharing of study materials for database studies, however, is currently limited. First, it requires considerable effort to maintain analytical code so that it is understandable for investigators outside the project or organization. Similarly, it can be demanding to document the necessary meta data. However, transparency is not achieved by posting a link to an unorganized dump of materials, where relevant information is difficult to find or interpret. As an example, sharing long scripts of analytic code without clear annotation of the decisions that are being implemented gives a false sense of transparency without any real impact on reproducibility and ability to evaluate study quality. Git can be a useful tool for creating reproducible code workflows with a well-documented version-control system.[23] Second, some may consider their research procedures (e.g. code, algorithms) intellectual property that they would not want in the public domain without some form of attribution or recognition. Furthermore, some data vendors do not permit public sharing of code that could reveal the underlying data model of protected access data. The effort to create and share well-documented research materials will, however, arguably benefit both the researcher as well as consumers of their work. Algorithms, code, and other materials can be citable resources, with high quality, findable, and interpretable materials being cited more than data dumps.[24] Importantly, the Open Science Framework registries are linked to researcher ORCID identifiers and provide digital object identifiers (DOI) to facilitate identification and citation of posted study materials. Accessible research materials will help database researchers learn and build from each other's work, thus reducing waste and accelerating growth and innovation.

Transparency statement: If analytic code is not available, then authors could state so (level 0). If the code is only available upon request, this should be stated (level 1). Stating that analytic code is "available upon request" is not sufficient to earn an open science badge. If the research team can provide a link to analytic code that at minimum creates tables, figures, and analysis results from a derived analytic dataset on an open access

repository (e.g. github) (level 2) then they will have earned an open science badge (**Figure 1**). The research team can attain an even higher level of transparency if they also share code or workflows used to generate the derived analytic dataset from source data warehouses (level 3). Importantly, sharing code comes with the responsibility of sharing well-structured scripts with sufficient annotation/comments for it to be digestible by other researchers.

5. Reporting checklists

The purpose of publications and reports are to communicate research findings. Unlike protocols, they include key points, results, and interpretation. Research reporting checklists increase the value of research communications by providing guidance on the most critical elements that must be included. Such checklists can be used after conducting the study to create an outline for drafting a clearly written paper, or in the very early phases of research to quickly outline and evaluate a research plan for discussion with an advisor, collaborator, or decision-maker before proceeding with development of a full protocol.

Several checklists are available to support the reporting of RWE studies and these reporting checklists are also required by many journals. The strengthening the reporting of observational studies in epidemiology (STROBE)[25] checklist is one of the most well recognized. However, within subdisciplines of observational studies, there are substantive differences in the types of information on methodology that are crucial to report. To cover a broad base, some of the items in STROBE are necessarily generic and/or of less relevance for database studies (e.g. provide reasons for non-participation). Other items that are highly relevant are not addressed (e.g. use of a design diagram to provide clarity on temporality of assessment windows). An alternative to STROBE, is the "reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology" (RECORD-PE)[26] a reporting checklist that builds off of STROBE and is specifically tailored for database studies.

Transparency statement: If the manuscript was not prepared in accordance with a reporting checklist, then this could be stated (level 0). If the research team used a relevant reporting checklist and share the checklist with the publication (level 2) then they will have earned the open science reporting badge (**Figure 1**). In the absence of specific requirements by a journal, the RECORD-PE checklist should be preferred.

Examples

Here, we provide three 'transparency statements' as examples (Figure 2). These are constructed based on three previous papers from our own research teams and are selected to exemplify the evolution of transparent and reproducible research practices over time in our own work. Of note, even the most recent example does

not attain full transparency according to the present framework. Thus, these examples also highlight areas to focus on to build greater transparency into our pharmacoepidemiology and health outcomes research.

These examples highlight implementation of proposed wording to cover each of the five domains. However, this language - the exact wording of the transparency statement can and should be adjusted to the individual case. For example, the statement for examples #2 and #3 combines the statement on pre-registration with the statement on protocol availability, as the two are interlinked for these particular papers. This may not always be the case. Similarly, highlighting transparency domains not covered by a given paper (the domains at 'level 0') could be considered optional.

Discussion and conclusions

While there is universal agreement on the critical importance of transparent and reproducible science, the building blocks of open science practice that are universal across disciplines have not yet been built into routine workflows for many pharmacoepidemiology and outcomes researchers.

The ideas that are summarized in the proposed transparency statement are not new. Regulators, HTAs, and other decision-making organizations either have or are currently considering imposing pre-registration, protocol, and other requirements for certain types of database studies submitted to their organizations.[5,18,30,31] Clinical and epidemiology journals as well as professional conferences are also considering policies, processes, standards, or mandates to improve transparency and reproducibility. At the same time, there are ongoing discussions about adoption of positive reinforcement methods that have been successful in other fields[9] such as the display of open science badges on publications, posters and abstracts. Display of such badges prompt discussion and can lead to cultural shift by inspiring others to earn similar recognition. Future potential incentives can easily be envisioned, e.g., peer reviewers being more likely to accept review tasks for papers with badges, or other types of promotion, e.g. via social media or open science metrics analogous to the h-index for papers having earned badges. Development of processes and infrastructure to support follow through on declarations, such as recent new policies on data, protocol and code sharing from funders like the National Institutes of Health[32] and Patient Centered Outcomes Research Institute[33], are important as a declaration does not guarantee action and it has been observed that in spite of modest increases in declarations of data or code sharing for medical research in recent years, actual sharing of such materials has not necessarily increased.[34]

That said, we do not have to wait for new incentives, mandates or policies to be developed across different stakeholder organizations. The transparency statements outlined in the present paper can be used by research teams to proudly display the open science practices that were used to generate evidence designed to inform

public health policy and practice. Such a statement signals confidence from the research team in the scientific choices that were made and indicates that the research team is not simply asking the evidence consumer to trust that they know what they are doing; rather they are willing to show their work, to show how the evidence was generated, and make it possible for the results to be reproduced or replicated.

It is important to note that transparency and validity are often conflated. While the statement proposed in this paper is focused on transparency, it is possible to produce highly transparent research that is intractably biased. Conversely, one can also produce valid and robust causal inferences from database studies without being transparent about how that evidence was generated. Importantly, we see a move toward greater transparency as a move toward validity not because transparency equals validity, but because transparency enables researchers, reviewers, clinical, regulatory, and coverage decision-makers to better assess validity . Transparency on the five building blocks will provide evidence consumers with the materials needed to differentiate the highly valid database studies whose evidence they would want to use to inform decision making from the flawed studies whose evidence they want to disregard.

Akin to the transformation among trialists two decades ago, the power to realize a future where real-world evidence is recognized as consistently transparent, reproducible, and fit-for-decision-making lies in the hands of current and future researchers. We both hope and expect to see transparency statements and open science badges become ubiquitous in pharmacoepidemiology and outcomes research over the years to come.

References

1. United States Food and Drug Administration. Framework for FDA's Real World Evidence Program. 2018. Dec 2018. Accessed 1/31/2019. https://www.fda.gov/media/120060/download
2. European Medicines Agency. The ENCePP Code of Conduct For Scientific Independence and Transparency in the Conduct of Pharmacoepidemiological and Pharmacovigilance Studies. European Medicines Agency. Accessed 4/14/2020, http://www.encepp.eu/code_of_conduct/documents/ENCePPCodeofConduct.pdf
3. Makady A, Ham RT, de Boer A, et al. Policies for Use of Real-World Data in Health Technology Assessment (HTA): A Comparative Study of Six HTA Agencies. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*. Apr 2017;20(4):520-532. doi:10.1016/j.jval.2016.12.003
4. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Pharmacoepidemiology and drug safety*. 2017;26(9):1033-1039.
5. Hampson G, Towse A, Dreitlein WB, Henshall C, Pearson SD. Real-world evidence for coverage decisions: opportunities and challenges. *Journal of comparative effectiveness research*. Dec 2018;7(12):1133-1143. doi:10.2217/cer-2018-0066
6. TOP Factor. Center for Open Science. Accessed 11/4/2022, https://www.cos.io/initiatives/top-guidelines

7. Open Science Badges. Center for Open Science. 10/20/2022. https://www.cos.io/initiatives/badges
8. Open Code Badge. CoMSES network. Accessed Jan 10, 2023. https://www.comses.net/resources/open-code-badge/
9. Kidwell MC, Lazarevic LB, Baranski E, et al. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLoS biology*. May 2016;14(5):e1002456. doi:10.1371/journal.pbio.1002456
10. Korbmacher M, Azevedo F, Pennington CR, et al. The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology*. 2023/07/25 2023;1(1):3. doi:10.1038/s44271-023-00003-2
11. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *Bmj*. Jan 12 2021;372:m4856. doi:10.1136/bmj.m4856
12. Agency EM. Guidance for the format and content of the protocol of non-interventional post-authorisation safety studies. Updated 26 Sep 2012. Accessed Jan 4 2022, https://www.ema.europa.eu/en/documents/other/guidance-format-content-protocol-non-interventional-post-authorisation-safety-studies_en.pdf
13. Guidelines for good pharmacoepidemiology practices (GPP). Guideline. The International Society of Pharmacoepidemiology. Updated Feb. Accessed 2, 17. http://www.ncbi.nlm.nih.gov/pubmed/17868186
14. Technology NESfh. *National Evaluation System for health Technology Coordinating Center (NESTcc) Methods Framework A Report of the Methods Subcommittee of the NEST Coordinating Center – An initiative of MDIC*. 2020. https://nestcc.org/data-quality-and-methods/
15. Wang SV, Pottegard A, Crown W, 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*. Oct 10 2022;doi:10.1002/pds.5507
16. Williams RJ, Tse T, Harlan WR, Zarin DA. Registration of observational studies: is it time? *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. Oct 19 2010;182(15):1638-42. doi:10.1503/cmaj.092252
17. Orsini LS, Monz B, Mullins CD, 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. *Pharmacoepidemiology and drug safety*. Nov 2020;29(11):1504-1513. doi:10.1002/pds.5079
18. Kurz X, Perez-Gutthann S, the ESG. Strengthening standards, transparency, and collaboration to support medicine evaluation: Ten years of the European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). *Pharmacoepidemiology and drug safety*. 2018/03/01 2018;27(3):245-252. doi:10.1002/pds.4381
19. RWE Registry. Open Science Framework. Accessed 10/20/2022, https://osf.io/registries/rwe/discover
20. Kuntz RE AE, Califf RM, Ingelfinger JR, Krumholz HM, Ommaya A, Peterson ED, Ross JS, Waldstreicher J, Wang SV, Zarin DA, Whicher DM, Siddiqi SM, Lopez MH. *Individual patient-level data sharing for continuous learning: A strategy for trial data sharing. NAM Perspectives.* 2019.
21. Gonzales A, Guruswamy G, Smith SR. Synthetic data in health care: A narrative review. *PLOS Digit Health*. Jan 2023;2(1):e0000082. doi:10.1371/journal.pdig.0000082
22. García-Closas M, Ahearn TU, Gaudet MM, et al. Moving Toward Findable, Accessible, Interoperable, Reusable Practices in Epidemiologic Research. *American journal of epidemiology*. 2023;192(6):995-1005. doi:10.1093/aje/kwad040

23. Weberpals J, Wang SV. The FAIRification of research in real-world evidence: A practical introduction to reproducible analytic workflows using Git and R. *Pharmacoepidemiology and drug safety*. 2024/01/03 2024;n/a(n/a)doi:https://doi.org/10.1002/pds.5740
24. García-Closas M, Ahearn TU, Gaudet MM, et al. Moving towards FAIR practices in epidemiological research. *American journal of epidemiology*. 2023;doi:10.1093/aje/kwad040
25. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS medicine*. Oct 16 2007;4(10):e296. doi:10.1371/journal.pmed.0040296
26. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). 2018;363:k3532. doi:10.1136/bmj.k3532 %J BMJ
27. Pottegard A, Meegaard PM, Holck LH, Christensen R, Madsen H, Hallas J. Concurrent use of tramadol and oral vitamin K antagonists and the risk of excessive anticoagulation: a register-based nested case-control study. *European journal of clinical pharmacology*. Mar 2013;69(3):641-6. doi:10.1007/s00228-012-1363-x
28. Arana A, Pottegard A, Kuiper JG, et al. Long-Term Risk of Skin Cancer and Lymphoma in Users of Topical Tacrolimus and Pimecrolimus: Final Results from the Extension of the Cohort Study Protopic Joint European Longitudinal Lymphoma and Skin Cancer Evaluation (JOELLE). *Clin Epidemiol*. 2021;13:1141-1153. doi:10.2147/CLEP.S331287
29. Kildegaard H, Lund LC, Hojlund M, Stensballe LG, Pottegard A. Risk of adverse events after covid-19 in Danish children and adolescents and effectiveness of BNT162b2 in adolescents: cohort study. *Bmj*. Apr 11 2022;377:e068898. doi:10.1136/bmj-2021-068898
30. *Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products*. Vol. FDA-2021-D-1214. 2021. Accessed 10/20/2022. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug
31. *Real-world evidence framework*. 2022. Accessed 10/20/2022. https://www.nice.org.uk/corporate/ecd9/chapter/overview
32. NIH Data Management and Sharing Policy. 1/3/2024. https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy
33. Policy for Data Management and Data Sharing. Accessed 1/2/204, https://www.pcori.org/about/governance/policy-data-management-and-data-sharing
34. Hamilton DG, Hong K, Fraser H, Rowhani-Farid A, Fidler F, Page MJ. Prevalence and predictors of data and code sharing in the medical and health sciences: systematic review with meta-analysis of individual participant data. *Bmj*. 2023;382:e075767. doi:10.1136/bmj-2023-075767

Figures

Figure 1 Constructing a 5-piece transparency declaration* for a real-world evidence study

	Level 0	Level 1	Level 2	Level 3
Protocol	There was no protocol for this study. OR The protocol for this study is not available.	The protocol for this study is available upon request.	The initial protocol and amendments for this study are available at (*provide link*[†])	The protocol was developed using a structured template. The initial protocol and amendments are available at (*provide link*[†])
Pre-registration	This study was not pre-registered on an institutional registry.	This study title and aims were pre-registered prior to conducting analyses at (*provide link*[†])	This study protocol was pre-registered prior to conducting analyses at (*provide link*[†])	n/a
Data	The data are not available.	The data are available upon request.	The data are available by (*provide contact information and steps to acquire a data use agreement AND/OR a link to synthetic data AND/OR a information on how to access data in a protected access repository*[†])	n/a
Code	The analytic code for this study is not available.	The analytic code for this study is available upon request.	The analytic code to create the tables, figures and analysis results for this study is available at (*provide link*[†])	The code to create the analytic dataset from the source data warehouse and code to generate the tables, figures, and analysis results for this study is available at (*provide link*[†]).
Reporting	The manuscript is not prepared in accordance with a reporting checklist.	n/a	The manuscript is aligned with guidance from the reporting checklist <name>. The checklist is provided in the appendix.	n/a

* Declarations at levels 2+ qualify for an open science badge.

Badges were developed by the Open Science Foundation (https://osf.io/5fghy) and CoMSES (https://www.comses.net/resources/open-code-badge/) and are covered by CC BY license.

[†]Link must be to an open access institutional registry or repository which provides a persistent identifier and is time-stamped.

For example, RWE Registry (https://osf.io/registries/rwe/), EU-PAS Register (https://www.encepp.eu/encepp/studiesDatabase.jsp), GitHub or GitLab (for programming code), etc.

Figure 2 Example transparency statements

A. Example 1

Using a case-control study design nested within users of the anticoagulant drug warfarin, we investigated the association between use of tramadol compared to non-use and the risk of bleeding.[27] The transparency statement could have appeared as:

> This study was not pre-registered on an institutional registry. The protocol for this study is available upon request. Neither analytical code nor data for this study is available. The manuscript was not prepared in accordance with a reporting checklist.

B. Example 2

Using a cohort design, we conducted a regulator-mandated multinational phase IV study on use of topical tacrolimus and pimecrolimus and the risk of skin cancer.[28] The transparency statement could have appeared as:

> This study is registered in the EU Electronic Register of Post-Authorisation Studies (EUPAS21769), including the initial protocol and subsequent amendments. Neither analytical code nor data for this study is available. The manuscript is aligned with guidance from the reporting of studies conducted using observational routinely collected health data (RECORD) statement.
>
> 
>

C. Example 3

In a population-based cohort study, we assessed the risk of acute and post-acute adverse events after SARS-CoV-2 infection in children and adolescents and evaluated the effectiveness of the BNT162b2 mRNA vaccine among adolescents.[29] The transparency statement could have appeared as:

> This study is registered in the Real-World Evidence Registry (https://osf.io/7ejh5), including the initial protocol and subsequent amendments. The analytical code to create the tables, figures and analysis results for this study is available at https://gitlab.sdu.dk/pharmacoepi/sars-cov-2-children. Deidentified data can be made available for authorised researchers after application to Forskerservice at the Danish Health Data Authority. The final data used for this study were extracted January 2022. Information on the application process can be found here: https://sundhedsdatastyrelsen.dk/da/forskerservice. The manuscript is aligned with guidance from the reporting of studies conducted using observational routinely collected health data (RECORD) statement.
>
> 
>

Badges were developed by the Open Science Foundation (https://osf.io/5fghy) and CoMSES (https://www.comses.net/resources/open-code-badge/) and are covered by CC BY license.