



Online first

# Reporting Studies Conducted Using Open Access Data (ROAD) Guideline Statement

Saif Aldeen Alryalat, M.D.<sup>1,2</sup>, Lna Malkawi, M.D.<sup>3,4</sup>, Randa Farah, M.D.<sup>5</sup>

<sup>1</sup> Department of Ophthalmology, University of Jordan, <sup>2</sup> Department of Ophthalmology, Houston Methodist Hospital, <sup>3</sup> Department of Radiology and Nuclear Medicine, University of Jordan, <sup>4</sup> Department of Diagnostic and Interventional Imaging, The University of Texas Health Science Center at Houston, <sup>5</sup> Department of Internal Medicine, University of Jordan

Keywords: Open Data, Guideline, Writing, EQUATOR, ROAD

<https://doi.org/10.59707/hymrUNIE2175>

---

## High Yield Medical Reviews

---

The growing availability of open-access data presents numerous opportunities for researchers, but also raises challenges in terms of adequately reporting methods and findings. This article presents the Reporting of Studies Conducted using Open Access Data (ROAD) guidelines: a comprehensive, practical framework developed to standardize and improve the reporting of research using open-access data. The guidelines were built upon existing principles for observational studies, tailored specifically to address the context of open data use. Their development involved an extensive review of published open data studies, and input from a diverse panel of experts through a series of consensus meetings. The ROAD guidelines encompass various aspects of study reporting, including specifying the original dataset, articulating study design and setting, detailing participant selection and variables, and acknowledging data providers. By enhancing transparency and reproducibility, these guidelines aim to improve the quality of research reports, ensure accurate interpretation of results, and foster more effective use of open-access data in the scientific community. We invite feedback and further refinement from researchers and practitioners to ensure the continued relevance of the ROAD guidelines in the dynamic landscape of open data research.

## INTRODUCTION

In the era of information technology, the availability of open-access data has radically transformed the landscape of scientific research. This transformation is fueled by an increasing number of publicly available datasets across various disciplines, democratizing access to vast pools of information and accelerating scientific discovery.<sup>1</sup> With this ready access to information, researchers can conduct studies without having to collect new data, leading to cost savings, increased efficiency, and the ability to investigate questions that would otherwise be impracticable.<sup>2</sup>

While the benefits of open-access data are profound, these developments have also raised new challenges concerning reporting studies conducted using these resources. The process of interpreting and drawing conclusions from open datasets requires an intricate understanding of the data's context, collection methods, and limitations.<sup>3</sup> Moreover, ensuring these studies' transparency, reproducibility, and integrity necessitates comprehensive reporting standards that are suitably tailored to this unique research paradigm. However, the current reporting standards for studies based on open-access data exhibit limitations. There is a lack of uniformity in these standards across different disciplines, leading to inconsistencies in the quality and reliability of published studies. Moreover, crucial information such as the precise description of the dataset used, data processing steps, and methodological considerations often

remain undisclosed, hampers these studies' reproducibility and verifiability.<sup>4</sup>

Recognizing these gaps, we propose the Reporting of studies conducted using Open Access Data (ROAD) guidelines. The ROAD guideline aims to provide a comprehensive and standardized approach for reporting based on open-access data, thereby enhancing these studies' clarity, transparency, and interpretability. This paper focuses on the formulation of these guidelines, their potential benefits, and the path toward their adoption.

## DEVELOPMENT OF ROAD GUIDELINE

Our approach in developing the ROAD guideline was to build on a solid foundation of pre-existing guidelines for observational studies, tailoring and extending them to better fit the unique context of open data use. This involved the incorporation of specific recommendations and requirements that address the challenges and opportunities inherent in open-access data. In doing so, we expect the ROAD guidelines to significantly improve the clarity and quality of study reporting, ensuring that references to open-access data are precise and unambiguous. The methodology adopted for developing the Open Data research checklist encompassed a series of structured steps:

**Preliminary Review and Requirement Gathering:** Initially, the author's team undertook a rigorous review of several research studies that used datasets from the Specimen

and Data Repository Information Coordinating Center (BioLINCC) repository.<sup>5</sup> This database was chosen for its richness of open data across diverse areas of bioinformatics.<sup>6</sup>

These studies were selected using a random selection tool, [random.org](http://random.org), to eliminate selection bias and ensure a broad representation of approaches in open data use. Based on this review, a preliminary list of critical reporting requirements was drafted, setting the groundwork for best practice guidelines in reporting research using open data.

**Consensus Meeting:** After the review and requirement gathering phase, the author's team met to discuss the initial draft checklist. The objective of this meeting was to foster a shared understanding, consolidate different perspectives, and reach a consensus on the components of the initial reporting checklist.

**External Peer Review:** We sent the consensus checklist, accompanied by a selection of Open Data publications that drew on BioLINCC repository data, to several active researchers with substantial experience and expertise in Open Data research. The aim was to review these materials, focusing on criticism, revision, and suggesting improvements to the checklist.

**Final Author Meeting and Checklist Formulation:** In the final phase of the process, the authors met again to discuss the feedback and suggestions received from the expert committee. Drawing on this input, the final version of the Open Data research checklist was formulated. This meeting synthesized all the information, finalized the guidelines, and established a concrete path forward for their adoption and dissemination.

The following are detailed ROAD guidelines. Moreover, the supplementary table provides a brief checklist for the ROAD guideline.

## ROAD GUIDELINE

### 1. Title and Abstract

- Standard: a) Indicate the study's design with a commonly used term in the title or the abstract. b) Provide an informative and balanced summary of the study in the abstract.
- ROAD: c) Avoid using the original data name or acronym in the title as per some data providers' requirements. d) Mention the original database used in the study within the methods section of the abstract.

### 2. Introduction

- Background and Rationale: a) Explain the scientific background and rationale for the investigation Objectives: a) State specific objectives, including any pre-specified hypotheses. . (For ROAD), b) give a brief about the original dataset used before stating the aim of the study.

### 3. Methods

- Study Design: a) Present key elements of study design early in the paper. (For ROAD), b) focus

on describing your study design rather than the original study's design.

- Linkage: (for ROAD) Briefly describe and cite the original dataset in the references and provide a link to it.
- Setting: a) Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection. (For ROAD) b) Use the original study setting briefly and state if your study was different, providing reasons for any differences in recruitment, exposure, follow-up, and data collection. Participants: (For ROAD) Briefly describe the original dataset population and cite the study describing the details, then describe your study population and detail the criteria behind selecting your sample. For studies that use multiple datasets, briefly describe each study and cite the reference for details.
- Variables: Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. (For ROAD) State all these variables included in your study from the original dataset.
- Data Sources/Measurement: Describe the sources of data and methods of assessment for each variable of interest.
- Bias: Describe any efforts to address potential sources of bias.
- Study Size: (For ROAD) The explanation of how the study size was arrived at is not necessary.
- Quantitative Variables: Explain how quantitative variables were handled in the analyses.
- Statistical Methods: a) Describe all statistical methods, including those used to control for confounding. b) Describe any methods used to examine subgroups and interactions

### 4. Results

- Participants: a) Report the number of individuals at each stage of the study and give reasons for non-participation at each stage. b) Consider the use of a flow diagram.
- Descriptive Data: Provide characteristics of study participants, including the number of participants with missing data for each variable of interest and information on exposures and potential confounders.
- Outcome Data: Report numbers of outcome events or summary measures.
- Main Results: Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included. Report category boundaries when continuous variables were categorized.
- Other Analyses: Report any additional analyses done, such as subgroup analysis, interactions, and sensitivity analyses.

5. Discussion
  - Key Results: Summarize key results with reference to study objectives.
  - Limitations: Discuss the limitations of your study and (for ROAD) discuss the limitations of the original data, considering sources of potential bias or imprecision.
  - Interpretation: Give a cautious overall interpretation of the findings while considering the goals, restrictions, variety of analyses, outcomes, and other pertinent data.
  - Generalizability: Discuss the generalizability (external validity) of the study results.
6. Other Information
  - Funding: (For ROAD) State the source of funding for your study, irrespective of the original dataset's sources unless required by the original data provider.
  - Accessibility of Protocol, Raw Data, and Programming Code: (For ROAD) a) Cite the data repository and any supplementary materials related to the data. b) Provide information on how to access any supplementary information such as the study protocol, raw data, or programming code relevant to the study.
  - Authors and Acknowledgments: (For ROAD) a) Acknowledge the original data provider. b) Some data providers may require being listed as an author.
  - Data Access: (for ROAD) Describe how other researchers can access the dataset.

## DISCUSSION

The ROAD guidelines serve as a comprehensive framework for researchers to accurately report their research findings using open-access data. The objective is to foster transparency, reproducibility, and credibility in research using such data while respecting the rights and stipulations of the original data providers.

The ROAD guidelines incorporate essential modifications to conventional research reporting protocols, tailoring them to effectively capture the unique challenges and opportunities of studies using open-access data. The guidelines offer researchers a structured approach to credibly present their research design, methodology, and findings while explicitly acknowledging the original data source.

In the ROAD checklist, we emphasize the significance of accurately acknowledging the source of the original open-access data, tailoring the study design to the unique context of open-access data, and outlining potential limitations that may arise from the original data. The checklist also guides researchers to maintain transparency in data handling, variable selection, bias mitigation, and statistical methods.

We strongly encourage researchers to adhere to these guidelines when using open access data. They should cite the original dataset, state their reasons for any differences

in recruitment, exposure, follow-up, and data collection, and detail the criteria behind selecting their study sample.

In addition, authors should refrain from using the original data name or acronym in the study's title unless requested by the original data providers. They should also discuss any limitations of their study, considering both the limitations of their research design and the original data's potential shortcomings.

The ROAD guidelines recognize the importance of open-access data in promoting scientific discovery and standardized reporting to strengthen the research community's trust in studies that make use of these resources. Adherence to these guidelines will ensure that the richness of open-access data is maximized, empowering researchers to produce robust, high-quality, and reliable research.

As open data continues to gain prominence in research, these guidelines offer a roadmap for researchers to navigate this landscape, promoting rigorous, transparent, and ethical reporting of studies using open-access data.

## LIMITATION

While the ROAD guidelines provide a comprehensive framework for reporting research using open-access data, they have some limitations. The diversity and variability of open-access data sources, each with unique policies, quality standards, and potential biases, can pose challenges that the guidelines may not comprehensively address. The guidelines provide a general direction, but specific issues related to unique data sources may not be fully covered.

Furthermore, the dynamic nature of open access data, with constant updates and the release of new datasets, means that the guidelines, although based on current best practices, may need timely revisions to address emerging trends and issues in the open data landscape.

Ethical and legal considerations form another limitation; at the same time, the guidelines recommend appropriate citation and acknowledgment of original data providers, but they may not cover all ethical and legal considerations, such as those concerning privacy and data protection. Hence, researchers must ensure compliance with all relevant ethical and legal requirements when using open-access data. Finally, it must be emphasized that while the ROAD guidelines aim to enhance transparency and reproducibility in research using open-access data, they are not intended to replace rigorous study design, robust data analysis, and critical interpretation of results. Adherence to these fundamental principles is crucial in conducting high-quality research.

## CONFLICT OF INTEREST

None.

Submitted: June 06, 2023 AST



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

## REFERENCES

1. BaHammam AS, Chee MWL. Publicly Available Health Research Datasets: Opportunities and Responsibilities. *Nat Sci Sleep*. 2022;14:1709-1712. doi:[10.2147/nss.s390292](https://doi.org/10.2147/nss.s390292)
2. Aldeen AlRyalat S. Open data are a boon for underfunded researchers. *Nature*. 2018;563(7730):184. doi:[10.1038/d41586-018-07310-3](https://doi.org/10.1038/d41586-018-07310-3)
3. Wang W, Liu M, He Q, et al. Data source profile reporting by studies that use routinely collected health data to explore the effects of drug treatment. *BMC Med Res Methodol*. 2023;23(1):95. doi:[10.1186/s12874-023-01922-8](https://doi.org/10.1186/s12874-023-01922-8)
4. Kahn MG, Brown JS, Chun AT, et al. Transparent Reporting of Data Quality in Distributed Data Networks. *eGEMs*. 2015;3(1):1052. doi:[10.13063/2327-9214.1052](https://doi.org/10.13063/2327-9214.1052)
5. Coady SA, Mensah GA, Wagner EL, Goldfarb ME, Hitchcock DM, Giffen CA. Use of the National Heart, Lung, and Blood Institute Data Repository. *N Engl J Med*. 2017;376(19):1849-1858. doi:[10.1056/nejmsa1603542](https://doi.org/10.1056/nejmsa1603542)
6. AlRyalat SA, Farah RI, Shehadeh B, et al. Biomedical researchers and students knowledge about predatory journals. *J Acad Librariansh*. 2019;45(5):102056. doi:[10.1016/j.acalib.2019.102056](https://doi.org/10.1016/j.acalib.2019.102056)

## SUPPLEMENTARY MATERIALS

### **Supplementary table**

Download: [https://hymr.scholasticahq.com/article/85154-reporting-studies-conducted-using-open-access-data-road-guideline-statement/attachment/179952.docx?auth\\_token=jwRRFQJ5dlnmxU6A1PNm](https://hymr.scholasticahq.com/article/85154-reporting-studies-conducted-using-open-access-data-road-guideline-statement/attachment/179952.docx?auth_token=jwRRFQJ5dlnmxU6A1PNm)

---