Protocol for pharmacoepidemiology simulation study practices: A review of the literature

Ryan Muddiman[a], Mary Walsh[a], Fiona Boland[b], Teresa Perez[c], Florencia Ines Aiello Battan[c], John Tazare[d], Anna Schultze[e], Frank Moriarty[a]

[a]School of Pharmacy and Biomolecular Sciences, RCSI, Ireland
[b]Data Science Centre, School of Population Health, RCSI, Ireland
[c]Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, Spain
[d]Department of Medical Statistics, London School of Hygiene & Tropical Medicine, London, UK
[e]Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

December 12, 2024

1. **Introduction**

This is the protocol for a review that we will conduct focusing on simulation studies published in the journal Pharmacoepidemiology and Drug Safety. Simulation studies are often used to measure the performance of an estimation approach in data-intensive fields. In pharmacoepidemiology the main source of data is observational in nature, often from health records, and the robustness of findings is highly dependent on the ability of study design and analysis to address bias. Causal or descriptive analysis are the main approaches used to draw conclusions about the observed population. Using simulations allows one to have control over the data characteristics and therefore allow a performance of inferential measurements to be obtained using the underlying data-generating mechanism as a ground truth.

We are interested in several aspects of simulation studies, including the data generation mechanism and inferential methods used and how their performance is measured. We are inspired by previous work which analysed code-sharing practices in pharmacoepidemiology[1] which we aim to emulate with respect to simulation studies. We will apply a more specific definition of simulation in our review. Our definition of simulation is those computer simulations that aim to quantify the performance of a statistical method. We note that many of our characteristics of interest were previously defined in a review of simulation studies in medicine.[2] We used a framework from that study that structures reporting based on aims, data generating mechanisms, estimand, methodology and performance measure (ADEMP). The ADEMP framework was used to inform the types of data and information we will extract. Once the first version of this protocol is finalised, it will be posted to Open Science Framework, and subsequent changes to the protocol will be documented by uploading subsequent versions.

The aim of the review is to synthesise evidence from simulation studies published in the Journal Pharmacoepidemiology and Drug Safety, and to determine the major commonalities and differences among computer simulation studies in pharmacoepidemiology with a view to informing a later simulation study on deprescribing.

2. **Search strategy**

Tazare et al.[1] conducted a literature review identifying all studies published in Pharmacoepidemiology and Drug Safety (PDS) between 2017 and 2022, to investigate code sharing in the field of pharmacoepidemiology. They used the following search :

“Pharmacoepidemiology and drug safety”[Journal] **AND**
((“2017/01/01”[Date - Publication] : “2022/12/31”[Date - Publication]))

and classified 37 papers as reporting simulation studies as part of the review.

We will extend that search by including articles published in PDS on or after 1 January 2023 and before the 31 October 2024. The updated search will use the PubMed database and the following search string:

“Pharmacoepidemiology and drug safety”[Journal] **AND**
((“2022/12/31”[Date - Publication] : “2024/10/31”[Date - Publication]))

Before finalising the analysis, we will rerun this search for the period 1 November 2024 to 31 December 2024 to include any additional simulation studies. We will use the R package easyPubMed for searching the literature. All records retrieved will be stored in a CSV file. We will then screen all abstracts using Covidence and include those based on the following criteria.

3. **Eligibility criteria**

Original research articles or brief reports, where computer simulated data is used will be included. We will exclude letters, commentaries and review articles. We will also exclude any simulation studies that are not computer simulations (i.e., to prevent including healthcare simulation) and exclude articles that do not measure the performance of an inferential method (i.e., forward modelling using real world parameters to simulate a process without using inferential methods).

4. **Data Screening**

Simulation studies listed in Tazare et al. will be extracted and screened with those identified from our search. Titles and abstracts of each of the articles identified in the updated search will be screened for eligibility by two reviewers. Second, full-text screening will be conducted to assess eligibility. Excluded records will be checked by a second reviewer.

The resulting list of articles will then be saved and shared on the Open Science Framework and the eligibility process displayed using a flow diagram.

5. **Data extraction**

We will then manually extract the following from the full text of all included sources (which will be completed by a single reviewer):

Datum	**Values**
DOI	e.g. 10.1002/pds.5446

PMID	e.g. 35505471
Publication year	e.g. 2018, 2019 etc.
Publication type	e.g. Article, brief report

Table 1 Publication information

Datum	**Values**
Aims	
Clinically focused	Yes, no (i.e. whether the study examines the question with reference to a specific condition, drug, and/or intervention, or from a generic perspective)
Clinical condition examined	e.g. Cancer, cardiovascular disease, etc
Type of intervention examined	Drug, vaccine, other
Drug class or vaccine	e.g. The specific drug class or vaccine involved in the simulation
Simulation type	Monte Carlo, plasmode simulation, bootstrap, other
Data-generating mechanisms	
Data generation mechanism	Stochastic process, random sampling, other
Time parameterisation	Continuous, discrete
Covariate dependency	Yes, no
Time-varying covariates	Yes, no
Time-varying effects	Yes, no
Source for data generation	Real-world data, synthetic data, both
Factoring used (variation in static parameters)	Yes, no
Factor variables	Means, covariances, etc.
Number of simulations (n_{sim})	Integer
Number of observations per simulation (n_{obs})	Integer
Estimand/Target of analysis	

Datum	Values
Primary quantity of interest	Estimand, p-value, other
Methods	
Primary inferential method	Maximum Likelihood Estimation, Partial Likelihood Method, Restricted Maximum Likelihood Estimation, other
Performance measures	
Performance measure for assessing inferential method	RMSE, Bias, Confidence Interval, other
Uncertainty analysis on performance measure	Standard error, coverage, other
Other	
Programming language	R, Python, MATLAB, SAS, SPSS, Other
Published code is linked and/or referenced in the text	Yes, no, available on request
Shared code use	Data-generation, statistical analysis, both, NA

Table 2 Simulation Parameters

We have pre-specified the parameters we believe to be of relevance for simulation studies in pharmacoepidemiology; however, if further parameters are identified during the review that are deemed to be relevant, these will also be extracted and in the final report we will identify that these were not pre-specified.

6. Data analysis

Descriptive statistics, tables and graphs will be used to explore and summarize the data described in Table 2. We will summarise the data using descriptive statistics (frequencies and percentages) of each extracted variable and present a bar chart where applicable. The raw and summarised data will be saved as a CSV file and shared on the Open Science Framework.

References

1. Tazare, J., et al., *Sharing Is Caring? International Society for Pharmacoepidemiology Review and Recommendations for Sharing Programming Code.* Pharmacoepidemiol Drug Saf, 2024. **33**(9): p. e5856.
2. Morris, T.P., I.R. White, and M.J. Crowther, *Using simulation studies to evaluate statistical methods.* Stat Med, 2019. **38**(11): p. 2074-2102.