Protocol title: Protocol pre-registration of observational studies-is this happening?

Team:
Sonia Gran[1]
Sophie Leducq[1,2]
Loes Hollestein[3,4]
Christian Apfelbacher[5]
Rishabh Mazmudar[6]
Faaris Zaki[7]

Affiliations:
[1] Centre of Evidence Based Dermatology, School of Medicine, Faculty of Medicine & Health Sciences, University of Nottingham, United Kingdom
[2] Department of Dermatology, University Hospital of Tours, Tours, France
[3] Department of Dermatology, Erasmus University Medical Center, Rotterdam, The Netherlands
[4] Department of Research, Netherlands Comprehensive Cancer Organisation (IKNL), Utrecht, The Netherlands
[5] Institute of Social Medicine and Health Systems Research, Otto von Guericke University Magdeburg, Magdeburg, Germany
[6] Case Western Reserve University School of Medicine, Cleveland, OH, United States of America
[7] Oxford University Hospital NHS Foundation Trust, Oxford, United Kingdom

Background

Observational studies are important to identify risk or protective factors for disease incidence or other clinical outcomes, contribute knowledge on long-term adverse events and inform clinical trials. Despite their important role, observational studies, even with ethical review, are not subject to the same requirements such as publishing a pre-study plan as in randomised controlled trials.

Clinical trial registration is mandatory by the International Committee of Medical Editor Journals and most journals require evidence of pre-registered trial protocols that are published before participant recruitment starts to be submitted alongside trial reports.[1] However, when it comes to publishing observational studies, study registration and adhering to a protocol is not necessary for most journals although observational studies are vulnerable to bias and selective reporting. The main benefit of pre-registered protocols for observational studies is to avoid multiple analysis and false positive findings (the "cherry picked").[2] Having pre-specified questions and a pre-specified protocol overcome selective reporting outcome bias.

Research objective

The aim of this study is to determine the proportion of observational studies published in the last year that were registered or had a protocol, across leading peer-reviewed journals medical journals defined as the top 5 highest impact factors journals in 2022. We have chosen high impact journals as we expect high standards in these journals.

Study design

A retrospective review of published papers.

Methods

We will use MEDLINE to search for observational studies. The search strategy is shown in Appendix 1.

We will identify observational studies published between January 1, 2022 and December 31st, 2022. We will look at studies published in the 5 highest-impact factor general medicine journals (*The Lancet*, *British Medical Journal*, *Journal of the American Medical Association*, *New England Journal of Medicine* and *Annals of Internal Medicine*).

Primary outcome
We will calculate
(i) the proportion of observational study papers that have been registered (no restriction on the type of registers)
(ii) the proportion of observational study papers that have a protocol available in the public domain

For each outcome, we will collect if the study or protocol have been pre-registered defined as:

- a study/protocol registered before collecting the data (i.e. before patients' enrolment) for a prospective study
- a study/protocol registered before analysing the data for a retrospective study or for a study using data that have already been collected (e.g. administrative healthcare records, cohorts that have been started before)

Secondary outcomes
(i) Proportion of studies with a protocol that adhered to the protocol (i.e. primary outcome was the same and adherence to prespecified subgroups analysis), (ii) proportion of protocols that included a statistical analysis plan (SAP). These outcomes will be stratified by whether the protocol was pre-registered or not.

For each paper, we will extract the following information:
- Country of the first author
- Journal
- Title
- First author
- Specialty
- Study design (cohort study, case-control study, cross-sectional study, case series study)
- Data source e.g. electronic healthcare records, review of patient notes, questionnaires etc.
- Study start date
- Main conclusion: positive outcome (significant effect hypothesis) (at least 1 positive outcome reported) Y/N/Not applicable
- Date of publication of the article
- Registration of the study on a platform or website Y/N (based from paper)
- Name of platform or website where study registered
- Date of registration
- Pre registration of the study Y/N
- Protocol in the public domain Y/N (based from paper)
- Name of journal or platform or website
- Date of publication of the protocol where protocol is available
- Pre registration of the protocol Y/N
- Protocol adherence (if protocol written) for primary outcome Y/N/Not applicable
- Protocol adherence (if protocol written) for prespecified subgroups analysis Y/N/Not applicable
- If there is deviation from the protocol, explanation and justification provided by the authors in the paper Y/N/Not applicable
- SAP included in the protocol Y/N

We will also extract the authors' guidelines of each journal regarding registration of protocol and submission of protocol with the manuscript.

Inclusion criteria: We will include observational studies (cohort study, case-control study, cross-sectional study, case series) published in the leading general medicine journals in 2022. Case reports will be excluded as protocols are not required for these studies.

Database: we will use OVID MEDLINE database.

Study selection: This will be done by two reviewers independently (using the software Rayyan): titles and abstracts and then full texts will be screened to exclude non-observational studies, case reports and reviews. Differences will be resolved by a third reviewer. If more than 200 articles are included after titles/abstracts/full texts selection, we will randomly select 200 articles to accommodate resources available.

Data extraction: This will be conducted by two reviewers independently using a piloted structured data extraction form developed to collect the relevant information from the selected articles (using the software Airtable®). Differences will be resolved by a third reviewer.
Before the study starts, three reviewers will evaluate a set of 10 papers and will resolve any differences in extraction and ensure interpretation of the data extraction tool is the same for all reviewers.
Data will be extracted based on the included article and its protocol (if available). Registers will not be searched if no information is available in the paper.

Analysis
A narrative synthesis will be conducted to describe and summarize the information from the included studies. Descriptive data will be expressed with number (and %) for categorical data.

Anticipated start date
28/03/2023

Planned dissemination
We plan to publish the results in a methodological peer-reviewed journal, e.g. Journal of Clinical Epidemiology.

References

1. Internet: https://www.icmje.org/ [Last accessed 1 st November 2022]
2. Swaen GG, Teggeler O, van Amelsvoort LG. False positive outcomes and design characteristics in occupational cancer epidemiology studies. Int J Epidemiol. 2001;30:948-54.

Appendix 1

Equation research

1.	Case-Control Studies/ or Control Groups/ or Matched-Pair Analysis/ or retrospective studies/ or ((case* adj5 control*) or (case adj3 comparison*) or control group*).ti,ab,kw.
2.	cohort studies/ or longitudinal studies/ or follow-up studies/ or prospective studies/ or retrospective studies/ or cohort.ti,ab. or longitudinal.ti,ab. or prospective.ti,ab. or retrospective.ti,ab.
3.	Cross-Sectional Studies/ or Prevalence/ or (cross-sectional or prevalence or transversal).ti,ab,kw.
4.	Epidemiologic Studies/
5.	Incidence/ or incidence.ti,ab,kw.
6.	1 or 2 or 3 or 4 or 5
7.	bmj.jn.
8.	lancet.jn.
9.	"new england journal of medicine".jn.
10.	jama.jn.
11.	"annals of internal medicine".jn.
12.	7 or 8 or 9 or 10 or 11
13.	6 and 12
14.	limit 13 to yr='2022'