BMJ Open

Current trends in the application of causal inference methods to pooled longitudinal non-randomised data: a protocol for a methodological systematic review

Edmund Yeboah,[1] Nicole Sibilla Mauer [ORCID] ,[1] Heather Hufstedler,[1] Sinclair Carr,[1,2] Ellicott C Matthay,[3] Lauren Maxwell,[1] Sabahat Rahman,[4] Thomas Debray,[5,6] Valentijn M T de Jong,[5] Harlan Campbell,[7] Paul Gustafson [ORCID] ,[7] Thomas Jänisch,[1,8,9] Till Bärnighausen[1,10]

For numbered affiliations see end of article.

Correspondence to
Heather Hufstedler;
heather.hufstedler@uni-heidelberg.de

ABSTRACT

Introduction Causal methods have been adopted and adapted across health disciplines, particularly for the analysis of single studies. However, the sample sizes necessary to best inform decision-making are often not attainable with single studies, making pooled individual-level data analysis invaluable for public health efforts. Researchers commonly implement causal methods prevailing in their home disciplines, and how these are selected, evaluated, implemented and reported may vary widely. To our knowledge, no article has yet evaluated trends in the implementation and reporting of causal methods in studies leveraging individual-level data pooled from several studies. We undertake this review to uncover patterns in the implementation and reporting of causal methods used across disciplines in research focused on health outcomes. We will investigate variations in methods to infer causality used across disciplines, time and geography and identify gaps in reporting of methods to inform the development of reporting standards and the conversation required to effect change.

Methods and analysis We will search four databases (EBSCO, Embase, PubMed, Web of Science) using a search strategy developed with librarians from three universities (Heidelberg University, Harvard University, and University of California, San Francisco). The search strategy includes terms such as 'pool*', 'harmoniz*', 'cohort*', 'observational', variations on 'individual-level data'. Four reviewers will independently screen articles using Covidence and extract data from included articles. The extracted data will be analysed descriptively in tables and graphically to reveal the pattern in methods implementation and reporting. This protocol has been registered with PROSPERO (CRD42020143148).

Ethics and dissemination No ethical approval was required as only publicly available data were used. The results will be submitted as a manuscript to a peer-reviewed journal, disseminated in conferences if relevant, and published as part of doctoral dissertations in Global Health at the Heidelberg University Hospital.

Strengths and limitations of this study

► This systematic review employs a search strategy which has been rigorously built and piloted in collaboration with three university library scientists.
► The reviewers will use the blinded review platform provided by Covidence.
► Limitations of the study include its restrictions to the English language and to three select publication years.

INTRODUCTION

At the heart of the causal theory is the concept of counterfactuals, where, ideally, patients could be assigned to two distinct treatment regimens (exposed (1) and unexposed (0) to treatment), without the patient under treatment (1) being exposed to non-treatment condition (0), and vice versa. Because this is impossible to create in the real world, scientists have built on theories such as counterfactuals to develop study designs or analysis methods with which to infer causal relationships. One key underlying assumption of causal methods is *exchangeability*. In epidemiology, exchangeability means that the counterfactual risk of those unexposed is equal to the observed risk of those exposed.[1] Randomisation produces such exchangeability conditions and is thus often considered the gold standard design in medicine to infer causality.

Randomization, however, is not always practical or ethical, for example, the random assignment of individuals to smoking studies to investigate the long-term health consequences of smoking. Measuring the long-term effects of this exposure of interest (smoking)

is generally only possible through longitudinal observational studies. The lack of exchangeability in observational studies is thereby a threat to one's ability to derive causal effects. Together, the Neyman-Rubin model[2] and Pearl's work with Directed Acyclic Graphs (DAGs)[3] have had an extensive impact on epidemiology, extending the breadth and reach of causal statistical inference for observational data. Single-stage regression-based adjustment (RBA) is perhaps the most well-known and most common way researchers approach causality. With RBA, researchers can adjust for measured confounders and several other threats to causality. However, RBA methods are inadequate in controlling for confounders, which are simultaneously mediators or colliders,[4] or in instances when unmeasured confounders exist. To overcome these limitations, scientists across different disciplines have, sometimes concurrently, developed other methods to support causal inferences from non-randomised, observational data. For the sake of clarity, we will refer to these non-RBA methods throughout the protocol as 'causal methods'.

Causal methods include those, which address observed time-varying confounders, such as G-methods (including inverse probability-weighted marginal structural models, g estimation of a structural nested model, and the g formula),[5 6] and methods to control for both measured and unmeasured confounding (also frequently known as quasi-experimental methods), such as difference-in-difference estimations (DiD),[7] interrupted time series (ITS),[8] regression discontinuity design (RDD),[9] fixed-effect RBA models,[10] and instrumental variables (IV).[11] Applying causal methods to longitudinal observational studies can strengthen numerous domains by allowing for better confounder control. However, even with the development of these methods, many researchers are still cautious to claim causality in their research questions, let alone in their conclusions, due to the understanding that no results derived from causal methodology can completely sustain or deny causality. This often leads to authors' employment of euphemisms such as '*link*' and '*association*' in their publications.

In contrast with randomised studies, which are often too time-consuming or costly to ensure a large enough sample size to make population-level inferences, observational study designs can result in increased sample sizes at lower costs. However, causal inference methods employed with non-randomised data are often data-hungry, requiring enormous sample sizes to which researchers may not have access. Pooling data from multiple studies is a solution researchers can employ to satisfy the sample size requirements of the causal methods. The pooling across multiple studies also has the potential to increase the diversity of study populations and enhance external validity. Studies pooling aggregate data are valuable,[12 13] but pooling individual patient data (IPD) can allow for better control of confounders, leading to enhanced internal validity.[14–16] Implementing causal methodologies with IPD from several studies is similar to multicentre single studies but more complex due to differences in types and measurements of variables captured in each study, greater study composition heterogeneity, or missing data.[17 18]

The transparent reporting of any applied research methods is a crucial component of good scientific practice, allowing authors to defend their findings and readers to understand the rigour of the approach and allow results to be replicated. The literature can be distorted when authors do not report critical details, such as eligibility or exclusion criteria for participants, the definition and composition of all variables included in the analysis, or how missing data and potential sources of bias were addressed. Reporting standards, however, are not uniform across disciplines or outcomes of interest and are generally not tailored to causal methods. This review will highlight the current trends in the application and the reporting of causal methods in longitudinal observational studies pooling individual data from multiple studies.

Existing research on how causal methods are used and reported across disciplines is sparse. Two reviews reported the causal methods used to account for time-dependent confounding with non-randomised exposure data from randomised controlled trials (RCTs).[4 19] Our team conducted a similar methodological review focused on infectious disease cohorts,[20] but, to our knowledge, no study has systematically reviewed the implementation of causal methods and differences in reporting across disciplines in studies that pool individual-level, observational, longitudinal data from multiple studies focused on health outcomes. Furthermore, it has not yet been addressed if and how the implementation and reporting of causal methods may vary across time, geography and academic disciplines. Investigating this is of interest, as it is widely acknowledged that researchers still tend to use dominant methods in their home disciplines instead of what is indicated by the data. Suppose such behaviour, or other trends in the application of methods, are indeed present. In that case, the results of this investigation could inform future conversations and consensus in the use and reporting of causal inference methods for health outcomes. This project seeks to fill the above-mentioned knowledge gaps by conducting a descriptive analysis on (1) the causal methods implemented with pooled IPD across studies, (2) detail and quality of reporting of these methods, and (3) if/how it varies across time, geography or disciplines.

This protocol follows the Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidelines for systematic review protocols.[21]

METHODS

Researchers listed here have been developing the methods for this methodological systematic review since mid-2019. Due to a COVID-19 pandemic-related delay in the screening process, the analysis and final manuscript are expected to be completed and submitted in Winter 2021.

Search strategy

The following databases will be searched using a combination of Medical Subject Headings (MeSH) and text terms, for example, "pool*", harmoniz*", "cohort*", "observational", and variations on 'individual-level data' (see online supplemental appendix 1), tailored to four databases (EBSCO (PsycINFO, Academic Search Complete, Business Source Premier, CINAHL, EconLit), Embase, PubMed and Web of Science). Due to capacity, the review will be limited to articles published in 3 years, 5 years apart: 2009, 2014, 2019. In order to capture the methods implemented in the pre-COVID era, we stopped the review in 2019. This protocol is registered with PROSPERO (CRD42020143148).

Types of studies to be included

Inclusion criteria: studies that

1. Have pooled longitudinal data from more than one cohort or RCT, if the exposure or treatment is not the factor that was randomised (methodological papers which have an applied example that fits this and the following additional criteria will also be considered). Nested trial designs that allow causal inferences to be drawn from a randomised trial to a target population are not included, as we do not classify them as distinct cohorts due to the preplanned study design.
2. Clearly state in the abstract level that data was pooled at the individual level,
3. Evaluated any type of health outcome (eg, body mass index, blood pressure, mortality),
4. Address a causal question (judged by four reviewers by carefully looking at study objective, discussion, and conclusion in the abstract; euphemisms for causality were used as indications, such as effect, impact, benefit, increase, decrease),
5. Estimated a causal effect size related to the said causal question,
6. Are published in the English language in the years 2009, 2014, or 2019 (the electronic publication date will be considered),
7. Are available in full-text through open access (through the journal), university license, or another collaborator on the project.

   Exclusion criteria: studies that
1. Have solely included the randomised variables from the RCTs as the exposure or treatment variables in their causal analysis,
2. Analysed only data from a single cohort study (whether single site or multisite),
3. Included either non-longitudinal designs (cross-sectional or repeated cross-sectional studies), or case studies in the pooled analysis,
4. Did not estimate an effect size corresponding to a causal research question (eg, focused on description, prediction, or prognostics).
5. If the study uses data from multiple registries from a Nordic country (Denmark, Finland, Iceland, Norway, and Sweden), they will be considered one cohort due

to the standardised nature of their nationwide registries.[22]
6. Furthermore, case studies, non-human studies, reviews, commentaries, corrections, editorials, erratums and grey literature (eg, protocols, abstracts, conference abstracts, dissertations) will be excluded.

Condition or domain being studied

For each study meeting the inclusion criteria, we will record the methods (including parameters such as study design, type of statistical analysis, methods to account for missing data, approaches to test assumptions required to infer causality; see online supplemental appendix 2 for detailed information on parameters included) that are used to estimate causal effects in longitudinal, observational studies pooling data from more than one cohort. We will also capture the primary discipline of the study. Disciplines will be determined in a multistep process based on the journal of publication: first, the journal name will be entered into either Journal Citation Reports or the National Library of Medicine to check the classification of discipline; second, if more than one discipline is assigned to the journal, the level of the impact factor or another comparable metric will be used to determine the discipline.

Participants/population

Studies will be restricted to human populations. No restrictions will be applied with regard to disease, age, gender, ethnicity, geography, or other characteristics of the study population.

Main outcomes

This study is a methodological systematic review designed to establish the implementation and reporting of methods to infer causality in studies that use individual-level data from multiple cohorts (such as pooled cohort studies and IPD meta-analyses) across disciplines. Expected outcomes of the review are to establish how and what the authors report with respect to:

1. Methods applied to infer causality, with a focus on, but not exclusively, non-regression-based adjustment (eg, IV, RDD, ITS, DiD, G-methods), analytic methods (study design, statistical analysis), and the motivating factor(s) in their selection.
2. Approaches to control for clustering of outcomes across cohorts.
3. Approaches to account for differences in data quality, such as variable measurement or standardisation, across individual cohorts.
4. Approaches to account for missing data (eg, imputation, omission).
5. Discussion and testing or evaluation of assumptions required for the chosen study design and analytic approach to isolate the causal effect of interest.
6. Covariates included in the adjustment set, for example, sociodemographic information (sex, age, education), individual health-related characteristics, and

whether they were labelled as confounders or mediators of the causal relationship.

7. Whether the conducting of sensitivity analyses was discussed.
8. Trends of data collection, analysis, and application of different causal methods across time, geography, and disciplines.
9. Justification for methods used.
10. Discussing potential effects of heterogeneity on the generalisability of results.

Data extraction (selection and coding)

Search results from all four databases will be uploaded to EndNote and deduplicated. All remaining results will be uploaded to and screened in Covidence. Four researchers (EY, NM, SC and HH) will conduct title/abstract screening and full-text review, and discrepancies will be resolved by discussion or a fifth reviewer. All efforts will be made to access the full text through databases, university access, collaborators' connections, or by contacting the authors. Data extraction of full-text articles will be completed independently and subsequently cross-checked (see full data extraction form in online supplemental appendix 2). Discrepancies will be resolved as in the previous step. The process will be documented in a PRISMA flow chart.

Data analysis

In addition to identifying and mapping the application of causal methods across time, geography and disciplines, studies will receive points based on the quality of their reporting: if any of the items from the data extraction sheet were not mentioned, 0 points; if the item was alluded to but not clearly addressed, 0.5; if the item was clearly addressed, 1 point. The awarding of points will err on the side of generosity. The results will be depicted graphically for the reader.

Patient and public involvement

There is no patient or public involvement in the design, conduct reporting or dissemination of the study, as this research is based on previously published data.

ETHICS AND DISSEMINATION

No ethical approval was required as only publicly available data will be used. The results will be submitted as a manuscript to a peer-reviewed journal, disseminated in conferences if relevant, and published as part of doctoral dissertations in Global Health at the Heidelberg University Hospital.

Author affiliations
[1]Heidelberg Institute of Global Health, Heidelberg University, Heidelberg, Germany
[2]Center for Interdisciplinary Addiction Research, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3]Center for Health and Community, University of California San Francisco, San Francisco, California, USA
[4]University of Massachusetts Medical School, University of Massachusetts, Worchester, Massachusetts, USA
[5]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands
[6]Cochrane Netherlands, University Medical Center Utrecht, Utrecht, The Netherlands
[7]Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada
[8]Department of Epidemiology, Colorado School of Public Health, Aurora, Colorado, USA
[9]Center for Global Health, Colorado School of Public Health, Aurora, Colorado, USA
[10]Harvard Center for Population and Development Studies, Harvard University, Cambridge, Massachusetts, USA

ORCID iDs
Nicole Sibilla Mauer http://orcid.org/0000-0002-0465-2300
Paul Gustafson http://orcid.org/0000-0002-2375-5006

REFERENCES

1 Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;60:578–86.
2 Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974;66:688–701.
3 Pearl J. *Probabilistic Reasoning in intelligent systems: networks of plausible inference*. San Franciso, California: Morgan Kaufman Publisher, Inc, 1988.
4 Farmer RE, Kounali D, Walker AS, *et al*. Application of causal inference methods in the analyses of randomised controlled trials: a systematic review. *Trials* 2018;19:23.
5 Doosti-Irani A, Mansournia MA, Collins G. Use of G-methods for handling time-varying confounding in observational research. *Lancet Glob Health* 2019;7:e35.
6 Naimi AI, Cole SR, Kennedy EH. An introduction to g methods. *Int J Epidemiol* 2017;46:756–62.
7 Snow J. *The cholera near golden square, and at Deptford*. 9. Medical Times and Gazette, 1854.
8 Lopez Bernal J, Cummins S, Gasparrini A. Difference in difference, controlled interrupted time series and synthetic controls. *Int J Epidemiol* 2019;48:2062–3.
9 Thistlethwaite DL, Campbell DT. Regression-discontinuity analysis: an alternative to the ex post facto experiment. *J Educ Psychol* 1960;51:309–17.
10 Gunasekara FI, Richardson K, Carter K, *et al*. Fixed effects analysis of repeated measures data. *Int J Epidemiol* 2014;43:264–9.

11 Baiocchi M, Cheng J, Small DS. Instrumental variable methods for causal inference. *Stat Med* 2014;33:2297–340.

12 Tierney JF, Fisher DJ, Burdett S, *et al*. Comparison of aggregate and individual participant data approaches to meta-analysis of randomised trials: an observational study. *PLoS Med* 2020;17:e1003019.

13 Tudur Smith C, Marcucci M, Nolan SJ. Individual participant data meta-analyses compared with meta-analyses based on aggregate data. *Cochrane Database Syst Rev* 2016;9:MR000007.

14 Thompson A. Thinking big: large-scale collaborative research in observational epidemiology. *Eur J Epidemiol* 2009;24:727–31.

15 Debray TPA, Moons KGM, van Valkenhoef G, *et al*. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods* 2015;6:293–309.

16 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:c221.

17 Lesko CR, Jacobson LP, Althoff KN, *et al*. Collaborative, pooled and harmonized study designs for epidemiologic research: challenges and opportunities. *Int J Epidemiol* 2018;47:654–68.

18 Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Stat Methods Med Res* 2018;27:1634–49.

19 Clare PJ, Dobbins TA, Mattick RP. Causal models adjusting for time-varying confounding-a systematic review of the literature. *Int J Epidemiol* 2019;48:254–65.

20 Hufstedler H, Matthay EC, Rahman S, *et al*. Current trends in the application of causal inference methods to pooled longitudinal observational infectious disease studies-A protocol for a methodological systematic review. *PLoS One* 2021;16:e0250778.

21 Moher D, Shamseer L, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.

22 Furu K, Wettermark B, Andersen M, *et al*. The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol* 2010;106:86–94.