Confounding adjustment methods in longitudinal observational data with a time-varying treatment: a mapping review

Stan R W Wijn,1 Maroeska M Rovers,2 Gerjon Hannink1

ABSTRACT

Objectives To adjust for confounding in observational data, researchers use propensity score matching (PSM), but more advanced methods might be required when dealing with longitudinal data and time-varying treatments as PSM might not include possible changes that occurred over time. This study aims to explore which confounding adjustment methods have been used in longitudinal observational data to estimate a treatment effect and identify potential inappropriate use of PSM.

Design Mapping review.

Data sources We searched PubMed, from inception up to January 2021, for studies in which a treatment was evaluated using longitudinal observational data.

Eligibility criteria Methodological, non-medical and cost-effectiveness papers were excluded, as were non-English studies and studies that did not study a treatment effect.

Data extraction and synthesis Studies were categorised based on time of treatment: at baseline (interventions performed at start of follow-up) or time-varying (interventions received asynchronously during follow-up) and sorted based on publication year, time of treatment and confounding adjustment method. Cumulative time series plots were used to investigate the use of different methods over time. No risk-of-bias assessment was performed as it was not applicable.

Results In total, 764 studies were included that met the eligibility criteria. PSM (165/201, 82%) and inverse probability weighting (IPW; 154/502, 31%) were most common for studies with a treatment at baseline (n=201) and time-varying treatment (n=502), respectively. Of the 502 studies with a time-varying treatment, 123 (25%) used PSM with baseline covariates, which might be inappropriate. In the past 5 years, the proportion of studies using PSM is the most frequently used method that used PSM over IPW increased.

Conclusions PSM is the most frequently used method to correct for confounding in longitudinal observational data. In studies with a time-varying treatment, PSM was potentially inappropriately used in 25% of studies. Confounding adjustment methods designed to deal with a time-varying treatment and time-varying confounding are available, but were only used in 45% of the studies with a time-varying treatment.

INTRODUCTION

The increasing availability of real-world data derived from electronic health records, registries, wearables and surveys can be a valuable source of data to evaluate the effectiveness of a treatment. Deriving inference directly from real-world data can be challenging as it is prone to confounding. To adjust for confounding, researchers use methods such as propensity score matching (PSM) to create two comparable groups in which both the treated- and untreated patients have similar observable characteristics (like age, pain scores, weight, etc.) similar to a randomised trial.

Although these methods can be sufficient when a patient is treated at the start of a study (baseline), more advanced methods might be required when dealing with longitudinal data and time-varying or repeated treatments. Adjustment at baseline in the presence of longitudinal data and time-varying treatment might not include possible changes that occurred over time. These can include
changes in treatment regimens or disease progression, but can also comprise weight changes, pain scores or changes in behaviour (eg, stopped smoking). These changes can alter the balance between treated and untreated patients and can result in different estimates of the treatment effect (see box 1).

Methods like time-dependent PSM and the g-methods (inverse probability weighting (IPW), parametric g-formula or g-estimation) can incorporate time-varying covariates and time-varying treatments and can take feedback between the treatment and outcome over time into account. It is however unclear if these methods are regularly used in practice when dealing with longitudinal observational data with a time-varying treatment. Therefore, this mapping review aimed to identify and describe which methods have been used to adjust for confounding bias in longitudinal observational data and identify potential inappropriate use of baseline adjustment methods (like PSM).

Figure 1

Figure 1 Forest plot displaying the results of the two empirical examples (left: meniscectomy, right: intra-articular corticosteroid injection (IAC)). Four methods were compared using baseline covariates, four methods using time-dependent covariates and time-varying treatment and one without correction. CCA, conventional covariate adjustment; IPW, inverse probability weighting; PSM, propensity score matching; tPSM, time-dependent propensity score matching.
Study selection and data extraction were performed by one reviewer (SRWW). Any issues during study selection, data extraction or analysis were discussed and resolved by all authors. No risk of bias assessment was performed because the scope of this paper targets the statistical methods that have been used in these papers, and therefore a risk of bias assessment was not applicable.

Analysis
Study selection was performed in Rayyan.11 Study characteristics (author, publication year, journal), time of treatment (at baseline, time-varying or unclear) and confounding adjustment method were extracted and analysed in R (V.4.1.0, The R Foundation for Statistical Computing, Vienna, Austria). Studies were sorted based on publication year, time of treatment and confounding adjustment method and described using descriptive statistics. If a study used multiple adjustment methods or a combination of methods, we included all methods, that is, more methods than papers could be identified. Cumulative time series plots were used to investigate the use of different methods over time for treatments at baseline and time-varying treatments using the Plotly package.12

RESULTS
Our search identified 2140 articles of which eventually 764 met the eligibility criteria after title and abstract review, and subsequent full-text review (see also figure 2).

![PRISMA Flow Diagram](image)

**Figure 2** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the flow of papers in the mapping review. In total, 764 studies were included and categorised according to the time of treatment. CA, covariate adjustment; IPW, inverse probability weighting; PS, propensity score; PSM, propensity score matching; TdPSM, time-dependent propensity score matching.
The main reasons for exclusion were the lack of intervention/treatment (n=405), a scope outside of medicine (n=376), a methodological paper (n=348) or the study did not use longitudinal observational data or did not correct for confounding (n=123). Of all included papers, 201 (26%) had a treatment at baseline, 502 (66%) had a time-varying treatment and 61 (8%) papers had no clearly defined time of treatment. Of the papers with a treatment at baseline, the majority used PSM with baseline covariates (n=165, 82%) as a method to correct for confounding. Studies that had a time-varying treatment most often used IPW (154 papers, 30%), PSM with baseline covariates was used in 123 papers (25%), PSM with baseline covariates combined with time-dependent Cox regression in 69 papers (14%), covariate adjustment using the propensity score in 49 papers (10%), time-dependent PSM in 40 papers (8%), parametric G-formula in 22 papers (4%), propensity score stratification in 18 papers (2%) and G-estimation in 13 papers (3%). Confounding adjustment methods designed to deal with a time-varying treatment and time-varying confounding (IPW, parametric g-formula or g-estimation) were used in 45% of the papers with a time-varying treatment. In the last 5 years, the proportion of studies with a time-varying treatment that used PSM with baseline covariates over IPW increased (199 vs 158 in 2020, for PSM with baseline covariates and IPW, respectively) (figure 3). For papers of which the time of treatment was unclear, PSM at baseline was most frequently used in 28 papers (46%). We added an overview of the most commonly used methods found in our search and when they should be used (figure 4).

**DISCUSSION**

Although advanced methods are available to correct for confounding in longitudinal observational data, we showed that these methods are not always used in studies that have a time-varying treatment. Instead, 25% of the studies that had a time-varying treatment used PSM with baseline covariates to correct for confounding which can potentially result in a biased treatment effect.

Our findings confirm the results by Clare et al whom provided a summary of new methods that have been used in literature to deal with time-varying confounding. They showed that these methods are not always used in studies that have a time-varying treatment. Instead, 25% of the studies that had a time-varying treatment used PSM with baseline covariates to correct for confounding which can potentially result in a biased treatment effect. Some potential limitations should also be discussed. First, the main limitation of a mapping review is the broad descriptive level at which studies are analysed and described. However, it does provide a general overview of the published literature and suggests that methods to deal with confounding in studies with a time-varying treatment are underused. Furthermore, no risk of bias assessment of the included studies was performed and study selection and data extraction were performed by one reviewer. Using a second reviewer throughout the entire study screening process could increase the number of relevant studies identified for use in a systematic review. However, as we targeted the overall trends in data analysis of studies with longitudinal observational data, this would likely not affect our conclusions much. Second, although it is common to search multiple databases in a systematic review, our mapping review was limited to PubMed. We found over 2000 papers in PubMed which was ample
Common methods to correct for confounding

Multiple methods are used to correct for confounding. Here we list the most common types and when they should be used.

Covariate adjustment using propensity score
- The outcome variable is a representation of the indicator variable denoting treatment status and the estimated propensity score (included in the analysis of study).
- Not recommended for eliminating bias or difference as it does not allow balancing of covariates across treated and control groups.

Propensity score stratification
- Stratifies patients into mutually exclusive subsets based on either estimated propensity score (based on design from analysis of study).
- Patients within strata have similar baseline values of the propensity score.
- Not recommended for eliminating bias of differences.

Propensity score matching
- Creating matched sets of treated and untreated patients who share a similar value of the propensity score. Variance reduction from analysis of study.
- PSM is recommended over stratification or covariate adjustment as it eliminates greater proportionality of systematic differences in baseline characteristics between treated and untreated.
- Not recommended for time-varying treatment or time-varying confounding.

Inverse probability weighting
- Generates a pseudo population in which exposures are independent of confounders, analyzing and adjusting for marginal structural model (parametric, covariance regression type).

Parametric G-formula
- Models the joint density of the observed data to generate potential outcomes under different hypothetical treatment strategies included in the analysis of the study.
- Suitable for longitudinal data with time-varying treatments and can adjust for time-varying confounders that are affected by prior exposure.

G-estimation
- Exploits the conditional independence between the exposure and potential outcomes to estimate structural nested models parameters included in the analysis of the study.
- Suitable when time-varying exposures or when personal characteristics strongly influence the conclusions of a study. A direct comparison of different confounding adjustment methods is not recommended to predict how different methods can affect the conclusions of a study. A direct comparison of different methods to correct for confounding is not recommended as this could stimulate selective reporting of (positive) study results. Every analysis of longitudinal observational data should start by selecting the method best suited for the data at hand. Figure 4 provides an overview of the most commonly used methods and can assist researchers to select the most appropriate method available.

CONCLUSION

PSM using baseline covariates is the most used method to correct for confounding in longitudinal observational data, even in the presence of a time-varying treatment. Of the 502 identified studies with a time-varying treatment, 123 (25%) used PSM with baseline covariates, which might be inappropriate. Confounding adjustment methods designed to deal with a time-varying treatment and time-varying confounding (IPW, parametric g-formula or g-estimation) are available, but were only used in 45% of the papers with a time-varying treatment and this can potentially result in biased estimates of the treatment effect.

Twitter Stan R W Wijn @Stan_Wijn and Maroeska M Rovers @MaroeskaRovers

Contributors SRW: conceptualisation, methodology, software, validation, formal analysis, writing - original draft, visualisation. MMR: conceptualisation, writing - review and editing, supervision, project administration, funding acquisition. GH: conceptualisation, methodology, validation, writing - review and editing, supervision, project administration, funding acquisition, guarantor.

Funding This work was supported by the Junior Research project (2018) grant provided by the Rudbod Institute for Health Sciences, Rudbod University Medical Centre, Nijmegen, The Netherlands.
Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. The search strategy is available in the supplemental file and all data extraction documents are available on request to the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Stan R W Wijn http://orcid.org/0000-0003-3782-6677
Maroessa M Rovers http://orcid.org/0000-0002-3095-170X

REFERENCES