

RESEARCH NOTE

Open Access



Pragmatic trials: ignoring a mediator and adjusting for confounding

Paraskevi Pericleous*

Abstract

Objectives: In pragmatic trials, the new treatment is compared with usual care (heterogeneous control arm) that makes the comparison of the new treatment with each treatment within the control arm more difficult. The usual assumption is that we can fully capture the relations between different quantities. In this paper we use simulation to assess the performance of statistical methods that adjust for confounding when the assumed relations are not true. The true relations contain a mediator and heterogeneity with or without confounding, but the assumption is that there is no mediator and that confounding and heterogeneity are fully captured. The statistical methods that are compared include multivariable logistic regression, propensity score, disease risk score, inverse probability weighting, doubly robust inverse probability weighting and standardisation.

Results: The misconception that there is no mediator can cause to misleading comparative effectiveness of individual treatments when a method that estimates the conditional causal effect is used. Using a method that estimates the marginal causal effect is a better approach, but not for all scenarios.

Keywords: Mediator, Confounder, Heterogeneity, Pragmatic, Trials

Introduction

Randomised controlled trials are held in an artificial environment with a carefully selected type of patients and placebo as the control group to the new treatment [1–3]. Pragmatic trials are held in real clinical practice, compare new treatment with usual care which consists of multiple treatments (heterogeneous control group) [4–6]. Comparing the new treatment with the usual care can be unbiased due to the randomization, but this is not true when comparing the new treatment with individual control treatments. Heterogeneity in the patient control group, which could lead to confounding, makes the comparison more complex. This paper considers a pragmatic trial, and thus there is heterogeneity in the control arm.

The usual assumption when dealing with pragmatic trials is that heterogeneity and confounding are fully captured or that the treatment causes the outcome [1]. The aim of this paper is to examine what happens when this assumption that is perceived as true, is false. The

scenarios considered examine whether the treatment does not directly cause the outcome, but via a mediator in cases where heterogeneity and confounding are fully and partially captured. Different methods are compared for adjusting for measured confounding when there is the mistaken assumption that there is no mediator when heterogeneity with or without confounding are fully or partially captured. We will investigate this via simulations using multivariable logistic regression (Logistic), propensity score (PS), disease risk score (DRS), inverse probability weighting (IPW), doubly robust inverse probability weighting (DRIPW) and standardization (ST) [7, 8]. These methods are widely used and estimate the potential outcome if all the patients were on the same treatment [7]. The effect that the treatment has on the potential outcome is called causal effect, if it comes from a conditional model is called ‘conditional causal effect’ and if it comes from a marginal model is called ‘marginal causal effect’. Estimating the causal effects requires exchangeability which is ensured by randomization and collapsibility which is ensured by adjusting for a fully captured confounding [7, 9].

*Correspondence: skevi.pericleous@ons.gov.uk
Data Science Campus, Office for National Statistics, Newport, UK



The motivation for using a multivariable logistic regression is because it is widely used when the interest is to predict a binary outcome, e.g. whether the patient is dead or alive, hospitalized or not, cancer metastasis or not within a specific time-period. A recent example of that can be found in Agarwal et al. [10] where they used hospital admission as a binary outcome with a logistic regression model.

Main text

Methods

We use the same notation as in Pericleous et al. [1]. Z is the treatment allocation ($Z=0, 1, 2$ denote the new treatment, baseline control treatment and second control treatment respectively). $n = \sum_{k=0}^2 n_k$, is the total number of patients participating and n_k the number of patients within each treatment group. $Y, C \sim \text{Bernoulli}(0.5)$ and $U \sim N(0, 0.64)$ are the binary outcome, observed and unobserved heterogeneity. We use no intercepts for the logit models to simplify the models. Assuming no refusals, patients are assigned to usual care using:

$$\text{logit}(P[Z = 2|C, U]) = \alpha_1 C + \alpha_2 U + \alpha_3 CU, \quad (1)$$

$$(Z = 1 \text{ otherwise})$$

where $I[\cdot]$ is the indicator function. We assumed a linear relationship of the mediator depending on the treatment:

$$M | V, Z = \beta_0 V + \beta_1 I[Z = 0] + \beta_2 I[Z = 2], \quad (2)$$

$$\text{where } V \sim N(0, 1)$$

The binary outcome (alive or dead) Y is given by:

$$\text{logit}(Y|M, U) = \beta_3 C + \beta_4 U + \beta_5 M. \quad (3)$$

We are interested in the scenarios shown in Table 1. More details for the scenarios are shown in Additional file 1: Figures S1 and S2. Additional file 1: Figure S1 shows the true relations, while Additional file 1: Figure S2 shows the assumption made. We want to examine the asymptotic behaviour of the models and thus, we use 10,000 patients and 1000 replications for each scenario. Having 10,000 patients in a pragmatic trial is a rare scenario. However, choosing a large number for patients and replication we ensure that we examine the asymptotic properties of the models and that the results are

not due to chance. We need to clarify that if the asymptotic behaviour of the model is problematic and cannot provide unbiased results, then having a small number of patients will not change that. We are mainly interested in estimating β_1 and β_2 .

We applied the most widely used methods for adjusting for confounding, as in Pericleous et al. [1]. These include the ones that calculate the conditional causal effect: (1) multivariable logistic regressions adjusted for confounding; (2) propensity score (PS), (3) disease risk score adjustment (DRS), (4) doubly robust inverse probability weighting (DRIPW), and the ones that calculate the marginal causal effect, (5) inverse probability weighting (IPW) and (6) standardisation [1, 7, 8].

PS is the probability to receive a specific treatment given the observed covariates using a multivariable logistic regression. Then, used as a covariate or as weights to predict the binary outcome [8]. In our case, we used it as a covariate. DRS is the probability of the binary outcome using a logistic regression with the treatment as a covariate. Then, used as a covariate to predict the binary outcome [8]. IPW uses weights which calculates by dividing the probability of the observed treatment exposure with the probability of the binary outcome using a Logistic regression given the confounders [7]. DRIPW uses the IPW weights to model the outcome, but the confounders are also used on the same model [7]. Standardisation: expands the dataset, modelling the outcome, getting the prediction and standardizing by averaging [7] (standardizes the mean outcome to the confounder distribution). It is mathematically equivalent to IPW [7]. For more details on the methods see Hernan and Robins [7], Pericleous et al. [1] and Schmidt et al. [8].

Results

The first four methods estimate the conditional causal effect (Logistic, DRIPW, PS, DRS) and the final two methods estimate the marginal causal effect (IPW and ST). All methods used are based in the misconception that there is no mediator. In Scenario 1, the methods that estimate the conditional causal effect do not perform well and provide biased results. The methods that estimate the marginal causal effect provide unbiased estimates of both β_1 and β_2 (Additional file 1: Figure S3). In Scenario 2, it seems that partially captured confounding leads to

Table 1 Simulation scenarios

Scenario	Parameter settings
(a) No unmeasured heterogeneity or confounding	$\alpha_2 = \alpha_3 = \beta_5 = 0, \alpha_1 = \beta_0 = \beta_1 = \beta_2 = \beta_4 = \beta_6 = 1$
(b) Unmeasured heterogeneity with confounding	$\alpha_1 = \alpha_2 = \alpha_3 = \beta_0 = \beta_1 = \beta_2 = \beta_4 = \beta_5 = \beta_6 = 1$
(c) Unmeasured Heterogeneity without confounding	$\alpha_2 = \alpha_3 = \beta_5 = 0, \alpha_1 = \beta_0 = \beta_1 = \beta_2 = \beta_4 = \beta_6 = 1$

biased results from all methods (Additional file 1: Figure S4). Mathematically, not considering a mediator that exists creates a partially captured heterogeneity, and thus Scenario 3 could be considered as having two different kind of heterogeneity. This is the possible reason why all methods (except ST that is doing slightly better in Scenario 1) perform relatively the same in Scenario 1 and Scenario 3 (Additional file 1: Figure S5).

Discussion

Ignoring a mediator and adjusting for confounding lead to biased estimates of the conditional causal effect, even though the heterogeneity and confounding are fully captured. IPW and Standardisation, however, provide unbiased estimates of the marginal causal effect under these circumstances. In cases where there is unobserved heterogeneity standardization is not as good as IPW in estimating the marginal causal effect, but it provides a rather good estimate. In cases, however, where confounding and heterogeneity are not fully captured all the methods provide biased results for both the conditional and the marginal causal effect. In conclusion, ignoring a mediator can lead to misleading conclusions about the comparative effectiveness of individual treatments when using methods that estimate the conditional causal effect. It is advised to use methods that calculate the marginal causal effect such as IPW that provide unbiased estimators when ignoring a mediator. However, in the case where unobserved heterogeneity and confounding exist then all the methods provide biased estimates.

Limitations

The limitations in this study are: we assume no treatment refusals, that there is only one mediator and that the relationship between the mediator and the treatments is linear.

Additional file

[Additional file 1.](#) Additional figures.

Authors' contributions

The author has done all the work related to this study. The author read and approved the final manuscript.

Acknowledgements

Not applicable.

Competing interests

The author declares that she has no competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

There is no need for ethical approval.

Funding

This work received no funding.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 24 August 2018 Accepted: 13 March 2019

Published online: 20 March 2019

References

- Pericleous P, Van Staa T, Sperrin M. The effects of heterogeneity in the comparative effectiveness of individual treatments in randomised trials. *Stud Health Technol Inf.* 2017;235:221–5.
- Zwarenstein M, Oxman A. Why are so few randomized trials useful, and what can we do about it? *J Clin Epidemiol.* 2006;59(11):1125–1126.
- Zwarenstein M, Treweek S. What kind of randomised trials do patients and clinicians need? *Evid Based Med.* 2009;14(4):101–103.
- Segal JB, Weiss C, Varadhan R. Understanding heterogeneity of treatment effects in pragmatic trials with an example of a large, simple trial of a drug treatment for osteoporosis. White Paper: Center for Medical Technology and Policy; 2012.
- Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, Schwartz JS. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. *Med Pub Issues Ann Int Med.* 2009;206:209.
- van Staa TP, Dyson L, McCann G, Padmanabhan S, Belatry R, Goldacre B, Cassell J, Pirmohamed M, Torgerson D, Ronaldson S, Adamson J, Taweel A, Delaney B, Mahmood S, Baracaia S, Round T, Fox R, Hunter T, Gulliford M, Smeeth L. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess.* 2014;18(43):1–146.
- Hernán MA, Robins JM. Causal inference. Boca Raton: Chapman & Hall/CRC, Forthcoming; 2019.
- Schmidt AF, Klungel OH, Groenwold RHH. Adjusting for confounding in early post-launch settings: going beyond logistic regression models. *Epidemiology.* 2016;27(1):133–42.
- Greenland S, Pearl J. Adjustments and their consequences—collapsibility analysis using graphical models. *Int Stat Rev.* 2011;79(3):401–26.
- Agarwal P, Mukerji G, Desveaux L, Ivers NM, Bhattacharyya O, Hensel JM, Shaw J, Bouck Z, Jamieson T, Onabajo N, Cooper M, Marani H, Jeffs L, Bhatia RS. Mobile app for improved self-management of type 2 diabetes: multicenter pragmatic randomized controlled trial. *JMIR Mhealth Uhealth.* 2019;7(1):e10321.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

