Epidemiology Publish Ahead of Print

DOI: 10.1097/EDE.0000000000001638

Transparency and Rigor: Target Trial Emulation Aims to Achieve Both

Bianca L De Stavola[1], Manuel Gomes[2], Michail Katsoulis[3]

[1] Great Ormond Street Institute of Child Health, University College London, UK

[2] Institute of Epidemiology and Health, University College London, UK

[3] MRC Unit for Lifelong Health and Aging, University College London, UK

Corresponding author: Bianca L De Stavola UCL Great Ormond Street Institute of Child Health, 30, Guildford Street London WC1N 1EH, UK b.destavola@ucl.ac.uk

Running title: Transparency and Rigor

Conflicts of interest: the three authors have no conflicts of interest

Financial support: none

No data were used

Editor's Note: Related articles can be found on pages XXX, XXX, and XXX.

ABOUT THE AUTHORS

Bianca L De Stavola is Professor of Medical Statistics at the Great Ormond Street Institute of Child Health, University College London. Her current research involves the implementation of causal inference methods for investigating the impact of educational interventions on health outcomes in children suffering from chronic conditions, as well as studying mediating pathways of genetic and social disadvantage.

Manuel Gomes is a health econometrics methodologist and Associate Professor at the Institute of Epidemiology and Health Care, University College London. His main research interests are in the development, assessment, and translation of causal inference methods for comparative

effectiveness research and health policy evaluation. He is recipient of a NIHR Advanced Fellowship that is concerned with exploiting target trial emulation for health technology assessment using real-world data for regulatory and reimbursement purposes.

Michail Katsoulis is an Associate Professor in Biomedical Statistics at the MRC Unit for Lifelong Health and Aging, University College London. His research mainly focuses on the application of causal inference methods in observational data arising from Electronic Health Records and longitudinal cohorts.

Commentary

Pearce and Vanderbroucke argue in their commentary [1] that the target trial emulation target trial emulation framework is being hyped as a new gold standard for causal investigations in observational epidemiology, with undesirable consequences for the training of newer generations of epidemiologists. Their view is that other approaches, such as triangulation and quasi-experimental studies, are to be preferred in many situations. While such approaches clearly play an important role in causal inference, we find that Pearce and Vandenbroucke draw their conclusions from some of the less rigorous applications of the target trial emulation framework, and thus overlook the potential this framework can offer. In this commentary, we expand on this, motivated by the desire to rebalance their concerns for the benefit of the next generation of epidemiologists. We conclude with an invitation to replace the classification of research frameworks as "gold-standard" (or otherwise) with one that assesses them in terms of **transparency** and **rigor**.

What does the target trial emulation framework entail?

There is nothing conceptually new in target trial emulation for experienced epidemiologists. What is gained from systematically adopting this approach, however, is the explicit specification of all the design and analytical components of a study within a protocol, ahead of the actual data gathering. This is standard practice for the design of prospective observational studies but less so when the data already exist, for example when administrative and medical databases are routinely collected and linked [2]. The target trial emulation protocol specifies the inclusion criteria, how outcome and exposure are to be measured, and (if applicable) when follow-up starts, ends, and its frequency. Unlike the usual design of, say, cohort studies, it also includes the causal contrast of interest (e.g., the 10-year mortality risk difference between exposure groups)

and the corresponding statistical analysis plan [3]. This structure mimics ("emulates") the entries of a protocol for an ideal pragmatic randomized controlled trial (RCT) that would address the same causal question, leading to the target trial emulation label.

Outlining all the design elements of an observational study, especially when the data already exist, helps avoiding unwitting introductions of new biases. For example, defining eligibility criteria and exposure in line with the aims of the study helps excluding individuals that would not be pertinent for the causal question (e.g., prevalent hormone replacement therapy (HRT) users when estimating the impact of HRT initiation [4]), or avoiding inappropriate assignment of baseline exposure from post-baseline information which would lead to immortal time bias [5]. Moreover, declaring the causal contrast of interest ahead of the statistical analysis plan avoids selective reporting of results and adds transparency of intent (contrary to what is traditionally done). This also matters because there may be a choice of causal contrasts (also referred to as estimands [6]), each corresponding to different notions of causal effects. Finally, the statistical analysis plan completes the protocol and identifies the estimation methods for the selected causal contrast, while attempting to deal with the biases that most likely affect observational data: confounding, selection (possibly due to missing data or missed linkage), and measurement error. Ideally the analysis plan should also outline sensitivity analyses of the robustness of the results to alternative assumptions, also known as quantitative bias analysis [7, 8].

Are Pearce and Vandenbroucke right in their criticisms?

We now turn to Pearce and Vandenbroucke's main criticisms:

"The target trial emulation approach considerably narrows the methodologic framework"

Pearce and Vandenbroucke consider the idea of adopting a target trial emulation approach as restrictive. We appreciate that certain causal questions cannot be addressed within the

counterfactual approach target trial emulation belongs to. For example questions that involve exposures such as social inequality, are ill-defined in the sense that it is difficult to justify how a given level of that exposure, most likely achieved by different means, would have the same impact on the outcome [9]. In such situation there would be no clear estimand and the corresponding target trial could not be specified. Though this seems to be a limitation, being explicit about the definition of the causal question, is an advantage because one is forced to be precise about what the causal contrast is, for example by replacing the ill-defined comparison of the impact of social inequality levels on birth weight with a more proximal exposure, say smoking during pregnancy.

More specifically, Pearce and Vandenbroucke state at the start of *General limitations of the target trial emulation approach:* "There are many situations where confounding in a target trial emulation study, in similar way as standard observational analyses, may simply be intractable", with similar remarks repeated in other parts of the commentary. This seems to imply that the no-unmeasured confounding assumption is crucial to how the target trial emulation framework is implemented. It is true that many target trial emulation studies use regression adjustment or propensity score-based methods which rely on this assumption (e.g., [10]). However, other estimation approaches that do not rely on this assumption, such as instrumental variables (IVs) and difference in differences, are also used in the target trial emulation literature (e.g. in [11] and [12]). Of note, with these estimation methods the no unmeasured confounding assumption is replaced by other unverifiable assumptions (e.g., exclusion restriction with monotonicity) which may be more plausible in certain settings. When unmeasured confounding is expected and estimation relies on the no unmeasured confounding assumption, sensitivity analyses, for example that exploit negative and positive control outcomes (e.g. [13]), could help assess the

extent of the induced bias. Importantly, the choice of estimation method should be guided by the causal contrast that one wishes to estimate because alternative estimation methods may target different estimands (and different populations [14]). For instance, difference-in-differences typically targets average treatment effects on the treated (ATT), whilst standard IV-based methods might target (depending on the assumptions made) the ATT or the local average treatment effects, the latter being the effect of the exposure in the subset of the population that would comply with the assigned exposure level.

In summary, adopting the target trial emulation framework does not restrict our choice of estimation approach, contrary to what PV suggest. Furthermore, outlining estimands and estimation methods within the target trial emulation protocol maps the causal question of interest onto the appropriate estimand and guides the choice of estimation method.

"The matched target trial emulation approach has little scientific benefit"

There are three main related comments we wish to make on this. The first concerns PV's argument that matched target trial emulation suffers from loss of power. We do not disagree with this. However, PV raise additional concerns about the representativeness of the results when data are "thrown away" because of matching. This is highlighted in their simple example where (binary) age is the only confounder and matching on age excludes several exposed older individuals, whereas regression adjustment would keep them all. Yet, regression adjustment would have a different disadvantage deriving from the fact that the age distribution in exposed and unexposed does not overlap (a case of lack of positivity [6]). In such setting, it would not be possible to identify the best fitting regression model for the outcome because any parametrization would rely on extrapolating the relationship of exposure and confounders with the outcome where there are no data. As the consequence of an incorrect model specification

could be the introduction of bias in the estimated parameters, the choice between matching and regression adjustment becomes a choice between bias due to extrapolating beyond the data and loss of representativeness.

Their second argument is that the direction of confounding bias can be detected by comparing adjusted and unadjusted estimates, a comparison that is not possible with matched data. We respectfully disagree. Pearce and Vandenbroucke use two examples to support this claim where an unadjusted estimate of 2.0 becomes after adjustment either 1.9 or 1.4 and conclude that in the first case it is improbable that there is substantial residual confounding, unlike in the second case. This reasoning does not hold because, if there is unmeasured confounding, the direction of such confounding could be in either direction or there could be unaccounted effect [15]).

Another criticism of matched target trial emulation studies aired by Pearce and Vandenbroucke concerns the inability of matching, when carried out with respect to baseline confounders, to deal with censoring bias (a bias that affects RCTs as well as observational studies). Clearly such bias would need addressing whatever design and estimation method is adopted and there are many approaches available to deal with it. Pearce and Vandenbroucke mention that regression adjustment would work, but they do not seem to consider that the drivers of loss to follow-up may not be included in the original set of baseline confounders. Besides, if the interest is in the causal contrast in the absence of censoring, inverse probability of censoring weights (IPCW) would be a more flexible approach (also for matched target trial emulation studies) since it has the added advantage of separating how confounding and censoring are dealt with, the first by matching and the second by weighting.

Gold standard?

Defining the study protocol ahead of data gathering is where epidemiologists have something to learn from trialists. This does not mean that randomized control trials (RCTs) would always be better at answering the same question compared to target trial emulation studies. For example, RCTs may suffer from limited generalizability and inability to study long-term outcomes. Nor should the target trial emulation framework be seen as the gold standard. There is no need, and we believe no general desire, to label any particular approach as the best suited to address causal questions using observational data: this is because the combination of causal question and available data requires mapping that question into a corresponding causal contrast, and assessing which estimation method for that estimand would invoke the more plausible assumptions. The target trial emulation protocol gives a formal structure for bringing together these considerations. Additionally, the adoption of this protocol's principles helps increase transparency and rigor in our work and thus has the potential to strengthen the evidence we gather from non-experimental data [16]. The principle of increased transparency is in line with what is invoked by the open science movement ([17]; https://www.cos.io/), which is now increasingly influencing how epidemiologic research is carried out, with analyses pre-specified and research tools such as computer code made accessible in publicly accessible platforms (e.g. https://www.cos.io/initiatives/prereg and GitHub). Our advice for the next generations of epidemiologists therefore is, where appropriate, to use the target trial emulation framework to formalize your causal investigations as this will be help you put transparency and rigor at the core of your research endeavor.

References

[1] N. Pearce and J. Vandenbroucke, "Are Target Trial Emulations (TTEs) the “gold standard” for observational studies?," *Epidemiology*, 2023.

[2] K. Harron, "Data Linkage in Medical Research," *BMJMED*, 2022.

[3] M. Hernan and J. Robins, "Unsing Big Data to Emulate a Target Trial When a Randomized Trial Is not Available," *American Journal of Epidemiology*, pp. 758-764, 2016.

[4] F. Grodstein, J. Manson, G. Colditz and e. al, "A Prospective Observational Study of Postmenopausal Hormone Therapy and Primary Prevention of Cardiovascular Disease," *Ann Internal Medicine,* pp. 933-941, 2000.

[5] M. Hernán, B. Sauer, S. Hernández-Díaz, R. Platt and I. Shrier, "Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses.," *J Clin Epidemiol.*, pp. 70-75, 2016.

[6] E. C. S. S. B. E. I. Goetghebeur, "Formulating causal questions and principled statistical answers. ," *Statistics in Medicine,* p. 4922– 4948, 2020.

[7] X. Zhang, D. Faries, H. Li, J. Stamey and G. Imbens, "Addressing unmeasured confounding in comparative observational research.," *Pharmacoepidemiol Drug Saf. ,* p. 373– 382, 2018.

[8] T. Lash, M. Fox, R. MacLehose, G. Maldonado, L. McCandless and S. Greenland, "Good practices for quantitative bias analysis," *Int J Epidemiol,* pp. 1969-1985, 2014.

[9] M. Hernán, "Does water kill? A call for less casual causal inferences.," *Ann Epidemiol.*, pp. 674-680, 2016.

[10] B. Dickerman, H. Gerlovin, A. Madenci and K. Kurgansky, "Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans," *N Engl J Med.*, pp. 105-115, 2022.

[11] S. Moler-Zapata, R. Grieve, D. Lugo-Palacios and A. Hutchings, "Local Instrumental Variable Methods to Address Confounding and Heterogeneity when Using EHR: an applicaiton to emergency surgery," *Medical Decision Making*, pp. 1010-1026, 2022.

[12] E. Ben-Michael, A. Feller and E. Stuart, "A Trial Emulation Approach for Policy Evaluations with Group-level Longitudinal Data.," *Epidemiology.*, pp. 533-554, 2021.

[13] J. Desai, C. Hyde, S. Kabadi, M. St Louis, V. Bonato, A. Katrina Loomis, A. Galaznik and M. Berger, "Utilization of Positive and Negative Controls to Examine Comorbid Associations in Observational Database Studies.," *Med Care.* , pp. 244-251, 2017.

[14] S. Swanson, "Instrumental variable analyses in pharmacoepidemiology: what target trials do we emulate?," *Curr Epidemiol Rep*, pp. 281-287, 2017.

[15] F. Aguayo, J. Muñoz, F. Perez-Dominguez and e. al., " High-Risk Human Papillomavirus and Tobacco Smoke Interactions in Epithelial Carcinogenesis.," *Cancers*, p. 2201, 2020.

[16] M. Hernan, "Methods of public health research- strengthening causal inference from observational data," *NEJM*, pp. 1345-1348, 2021.

[17] M. Munafò, B. Nosek, D. Bishop and e. al, " A manifesto for reproducible science.," *Nat Hum Behav.* , 2017 .