Are Target Trial Emulations (TTEs) the "gold standard" for observational studies?

Neil Pearce[1], Jan P. Vandenbroucke[1,2,3]

[1] Department of Medical Statistics
London School of Hygiene and Tropical Medicine, UK

[2] Leiden University Medical Center
Dept. Clinical Epidemiology
PO Box 9600
2300 RC Leiden, The Netherlands

[3] Department of Clinical Epidemiology, Aarhus University, Denmark.

Correspondence to:
Professor Neil Pearce
Department of Medical Statistics
London School of Hygiene and Tropical Medicine, UK
Email: neil.pearce@lshtm.ac.uk
Phone: 020-79588151

November 2022
(Revised January 2023)

Abstract

There has been considerable debate in epidemiology about whether the randomized controlled trial (RCT) is the “gold standard” for epidemiological studies. In particular, it has been argued that observational studies that are intended to address a causal question should be based on a hypothetical “target trial” which the observational study should attempt to emulate. Some studies take this approach further, and use epidemiological data sets to create matched-TTE, i.e. a cohort in which the “exposed” and “non-exposed” are made alike as much as possible, by matching “exposed” and “non-exposed” on a number of key variables. In this paper, we argue that although target trial emulations are appropriate and valid in some circumstances, that other approaches involving observational data may be more valid and appropriate in other situations. Our main concern about the TTE approach, is that in similar way as standard observational analyses, it cannot adequately deal with situations in which confounding is intractable, whereas other approaches can deal with this situation much more validly. In addition, there are two specific disadvantages of the matched-TTE design in that: (i) it is likely to yield similar effect estimates, but wider confidence intervals, compared to “standard” analysis which adjusts for the (matched) confounders using all of the available data; (ii) it makes it more difficult, if not impossible, to assess the likelihood, and likely strength and direction, of residual confounding. More generally, we argue that causal inference usually involves an approach that incorporates evidence from a wide variety of study designs and populations, rather than focussing on a single “ideal study”. The target trial framework can be very useful when thinking about the design of a single study, in the particular circumstances that allow for a TTE approach (e.g. the required data are available and confounding can be controlled with statistical adjustment). However, thereby it does not provide the starting point, nor the “gold standard” analysis, for causal inference in general.

There has been considerable debate in epidemiology about whether the randomized controlled trial (RCT) is the “gold standard” for epidemiological studies; some of the debate is old, but some is new [1, 2]. Almost all are agreed that a well-conducted RCT, where possible and appropriate, is to be preferred because an experimental set-up enables control of several key features of a study, and randomization ensures that any differences in outcome risk between the groups being compared are due to chance. It is thus tempting to apply the RCT paradigm to observational studies, i.e. to propose that if an RCT is not possible for a particular issue, then an observational study that closely mimics the RCT approach is the most preferable.

Therefore, it has been argued that observational studies that are intended to address a causal question should be based on a hypothetical “target trial” which the observational study should attempt to emulate[3], and recent years have seen the publication of a number of “target trial emulation” (TTE) studies. Some studies take this approach further, and use epidemiological data sets to create a cohort in which the “exposed” and “non-exposed” are made alike as much as possible, by matching “exposed” and “non-exposed” on a number of key variables [4]. We use the term “matched-TTE”, which to our knowledge has not been used previously, to specifically refer to this matched approach, which is a subgroup of the more general TTE approach.

Many TTE studies are excellent and have produced apparently valid findings[5]. This approach can be applied in a variety of settings including pharmacoepidemiology[6], and health policy[7]. However, there are several key limitations of this approach which mean that it cannot be universally adopted as the “gold standard” for observational studies. In particular, we argue that although target trial emulations are appropriate and valid in some circumstances, that other approaches involving observational data may be more valid and appropriate in other situations. More generally, we argue that causal inference usually involves an approach that incorporates evidence from a wide variety of study designs and populations, rather than focussing on a single “ideal study”[1, 2]. Although the target trial framework can be very useful when thinking about the design of a single study, this does not mean that it provides the starting point, or the “gold standard” analysis, for causal inference more generally. Indeed, the TTE approach considerably narrows the methodological framework, and the methods considered to be valid, for causal inference. This is exemplified by ‘scoring systems’ for reviews that are based on a how closely a study mimics a RCT[8]. These take the TTE approach to its logical conclusion in that a single ideal study design is seen as the “gold standard” whereas other types of evidence are either scored lower or ignored. In particular, triangulation [9] (which we consider to be the best approach for many situations) does not feature in the TTE methodological framework[8, 10].

Actually, the idea that one should think about the corresponding RCT, when designing an observational study is not new. As early as 1953, Dorn[11] recommended that one should ask the question "how would the study be conducted if it were possible to do it by controlled experimentation?" (quoted by Cochran in 1965[12]). We agree, and have used this approach to solve difficult design problems in our own research, e.g. when designing studies of the role of beta agonists in asthma death[13] – and the value of prophylactic administration of antibiotics before dental treatment in persons with pre-existing heart valve disease [14]. Thinking about the corresponding RCT can be useful in terms of thinking about eligibility for the study, definitions of exposure, selection of controls, and other issues. However, this does not mean at all that observational studies should be closely tailored and restricted to mimic an RCT, nor that this should involve a cohort study design. In fact, our own useful applications of RCT-thinking to design an observational study[13, 14] as with others[15], were actually case-control studies; imagining an RCT in the relevant source population and risk period made it clearer how to sample cases and controls and how to define exposures.

In this paper, we focus on a key shortcoming of the TTE approach in terms of how it limits conceptions of possible study designs, and thereby negates alternative approaches that may be more valid in particular circumstances. We then offer some specific criticisms of the matched-TTE approach.

General limitations of the TTE approach

The key disadvantage of the TTE approach is that it assumes that any causal analysis of observational data has to be framed within this RCT paradigm, focussing on a single "ideal study". In contrast, we would argue that there are many situations where confounding in a TTE study, in similar way as standard observational analyses, may simply be intractable – while other types of approach will work.

By way of example, a TTE analysis which compares the effectiveness of the Pfizer and Moderna vaccines[4] is unlikely to have substantial residual confounding, because all study participants have been vaccinated, and (conditional on factors such as age and time) it is essentially a natural experiment as to whether the study participants received one or the other vaccine. However, the situation is quite different when comparing vaccinated and unvaccinated. Then, there will usually be substantial differences in lifestyle, socioeconomic factors, and health-seeking behaviour between the two groups, and also between those who get tested for COVID-19 and those who do not. These are likely to be impossible to entirely or substantially eliminate with standard adjustment methods, e.g. using multiple regression and adjusting for measured confounders individually or with propensity scores.

One recent example is the study of Whiteley et al [16], which used a standard cohort study analysis, adjusting for potential confounders. The authors compared the association of Covid-19 vaccines ChAdOx1 and BNT162b2 with major venous, arterial and thrombocytopenic events. For those aged <70 years, the adjusted hazard ratios (HRs) were close to 1.0 for both vaccines for venous thromboses and arterial thromboses, but strong protective effects were reported for those aged 70 years or more. These findings are unlikely biologically, and the authors commented that *"adjusted associations in this study may still be biased by unmeasured confounding by patient characteristics that predict both vaccination and thrombosis and that are difficult to ascertain in electronic health records"*. While this was a standard cohort study, the elements that constitute a TTE design, however, would not have altered anything to this intrinsic intractability of confounding in those data, as the final analysis would have met the same problems.

Similar examples of apparently intractable confounding include studies of breast cancer recurrence among women receiving adjuvant chemotherapy[17], the association of religious service attendance with mortality[18], the effects of vegetarian diet[19], anti-hypertensives[20], screening colonoscopy[21], and vitamin E supplements[22]. In most instances, the findings of observational studies of these issues have been contradicted by findings from RCTs and/or from Mendelian Randomization studies. Similar problems arise due to confounding by indication when studying effects of treatments [17, 23, 24], with a key distinction being between studies of 'intended and non-intended effects'[25]. Recent papers have shown that the 'intractability' of confounding by indication still exists, despite the application of the most modern forms of analyses[17, 18]. One notable recent example is the study of Danaei et al[20] who use a a TTE design and found the expected protective effects of statins on mortality (that is: as expected from the results of RCTs), but found that the use of anti-hypertensives increased mortality, in contrast with the established findings of many RCTs.

Thus, there are important situations where even the best possible TTE study that relies on emulating conditional randomization cannot answer a causal question. Of course, this shortcoming also applies to "standard" cohort and case-control studies, that do not necessarily involve a TTE paradigm, but the formalism of the TTE approach may suggest a level of methodological rigour that simply cannot be obtained in such situations. This point is acknowledged by Hernan and Robins[3] for situations in which an active treatment is compared with no treatment or usual care, but they do not acknowledge the implication, i.e. that this may be a situation where the TTE approach simply won't work (e.g. in the case of anti-hypertensives[20]), no matter how well it is applied. In contrast, other well-established methods can be tried to validly assess causality in such situations. For example, the test-

negative design (TND)[26] has been shown to work well, in assessing factors such as vaccine effectiveness or risk factors for Covid-19[27, 28] and in other conditions[26], since any biases in terms of health-seeking behaviour will usually apply reasonably equally to the test-positive cases and the test-negative controls. Thus, even in a single study, a test-negative design may produce a reasonably valid effect estimate when a TTE approach would suffer from intractable confounding due to differences in health-seeking behaviour. Other approaches that make use of external interventions on groups, e.g. a difference-in-differences approach [4], may be useful to study COVID-19 interventions such as mask-wearing (where one expects similar intractability of confounding as with vaccination)[29]. Useful other approaches in single studies include the use of instrumental variables, regression discontinuity, and Mendelian Randomization[10]. Some of these approaches can be mapped onto a target trial framework[5-7, 30]. However, we would argue that these other approaches go beyond a TTE framework, and provide additional information and checks for bias [10] which cannot be achieved within the TTE framework. Moreover, in still other situations, any single study may struggle to yield valid causal assessments, but approaches that make comparisons across studies, such as triangulation [9] (i.e. the comparison of studies in which the expected biases are likely to be in opposite directions), may be of great value for identifying and controlling for confounding that is apparently intractable[31].

Specific problems of the matched-TTE design

There are also two specific limitations of the matched-TTE design which should be considered.

The first disadvantage of the matched-TTE approach is that it involves throwing away a great deal of useful data in the process of matching the exposed and non-exposed groups. Matching can be employed in cohort studies, whether or not they follow a TTE paradigm, although it has different implications than in case-control studies in which it is used more often. Matching in cohort studies [32, 33] is usually only done in specific situations, for example when the number of study participants on which particular information can be obtained is limited (e.g. for reasons of logistics or cost), which is then solved by selecting an equal number of exposed and non-exposed subjects at baseline for each age-sex stratum and only in these persons the additional information is collected[33]. Even in this situation, matching can sometimes harm efficiency, although it introduces no bias[32, 33]. However, it is of greater concern when the use of matching involves throwing away potential study participants for whom all relevant information has already been collected, because standard adjusted analyses would result in equal validity and greater precision.

To take a hypothetical example, suppose that matching was simply on age, which had two categories – old and young. If the exposed group was mostly old, and the non-exposed group was mostly young, then this can be handled in a conventional analysis adjusting for age. A matched-TTE approach would involve matching by age, which would involve discarding many exposed older persons (since matches would not be available for all of them), and also discarding many non-exposed younger persons. The matched analysis would therefore obtain virtually identical effect estimates to the age-adjusted analysis (i.e. no gain in validity), but with wider confidence intervals because of participants being discarded. This simplified example shows that in essence the matched analysis is the same analysis as stratification by age, but with lesser numbers. These problems become more acute as the number of matching factors increases.

For example, in their study of the comparative effectiveness of BNT162b2 and mRNA-1273 vaccines in US veterans, Dickerman et al[4] reduced the analysed participants from 764,803 to 439,684, a reduction of 43%, in the process of producing 219,842 pairs (comparing the two vaccines), matched on calendar date, age, sex, race, urbanicity of residence, and geographical location. The matched pairs population experienced 2016 SARS-Cov-2 infections, 559 of which were detected as symptomatic, 411 resulted in hospitalization, 125 in ICU admission, and 81 in death. The corresponding figures for the 43% of potential participants who were excluded are not reported, but assuming that their rates of infection are similar, this means that a substantial number of events were excluded from the analysis, including for categories (e.g. death) where the number of events was relatively rare. As a result, the 95% confidence intervals for these outcomes were relatively wide (for death, the risk ratio comparing the two vaccines was 1.11; 95%CI 0.69-1.91).

In fact, there is no benefit in terms of validity of this approach, and there may be a considerable disadvantage in terms of precision. It is age-old epidemiologic knowledge that matching in a cohort study removes confounding by the matching factors at time zero, although it may still be necessary to control for the matching factors because of differential censoring[33]. However, as in the simplified example above, the same aim can be achieved by not matching, using the whole data set and simply controlling for the same factors in the data analysis, usually with a gain in precision. This will usually produce an almost identical point estimate (e.g. the risk ratio) to the matched-TTE approach (effect estimates may not be exactly the same because of a different distribution of effect modifiers), but with better precision. The matching that produces similarities at baseline is nice ‘optically’ but does in essence not differ from any other stratified analysis by multiple strata. One needs to stratify on the relevant factors in order to do the matching, and all the matching does is “throw away” some of the data from each strata.

A second disadvantage of the matched-TTE approach is that the assessment of residual confounding becomes much more difficult. In a standard observational study, one usually presents effect estimates both unadjusted and adjusted for potential confounders. This 3enables us to see the direction and magnitude of the changes with adjustment (we leave aside, issues of collapsibility and the validity of the change-in-estimate approach[34, 35], and focus on the situation where the risk ratio is the main effect estimate), which in turn provides clues as to the likelihood of residual confounding. For example, if the unadjusted risk ratio for the main exposure is 2.0, and this only reduces to 1.9 after confounder adjustment, it is unlikely that there is substantial residual confounding, at least with regards to variables which are associated with the confounders that have been adjusted for. On the other hand, if the main effect estimate changes from 2.0 to 1.4 after confounder adjustment, it is highly likely that it would having changed further (perhaps going right down to 1.0) if there had been better information on the confounders (i.e. less non-differential misclassification).

Of course, such sensitivity analyses may not be conclusive, e.g. when all measured confounders are poorly measured or proxied. However, at least we can do such sensitivity analyses, and assess the results, in an unmatched analysis. In a3 matched-TTE it is impossible to make this sort of assessment since the confounders have been matched between exposed and non-exposed, and the unadjusted and adjusted effect estimates will be virtually identical. One is then left with the need to simply assume that there is no residual confounding, with few means to assess this – Hernan and Robins[3] mention several possible approaches such as using “reversed” strategies, or negative control outcomes, but do not consider the more straightforward approach of assessing the effects of confounder adjustment – something that is not possible in a matched-TTE design.

Discussion

A key feature of the TTE approach, and particularly the matched-TTE approach, is that it “looks like” an RCT, as persons in the intervention and control group have been made as much alike as possible, and it is therefore easier to market to policy makers as providing valid and useful findings. The TTE approach has produced useful and apparently valid findings, e.g. in re-analyses of data on postmenopausal hormone therapy and coronary heart disease[36], although it is notable that very similar findings have been produced with a more conventional observational study analysis, adjusting for potential confounders and for time since first exposure[37, 38].

Our concern is that although the target trial framework can be very useful when thinking about the design of a single study, in the particular circumstances that allow for a TTE approach (e.g. there required data are available and confounding can be controlled with

matching or standard adjustment methods), that this does not provide the starting point, or the “gold standard” analysis, for causal inference more generally. In particular, starting with the TTE approach considerably narrows the context, and the methods considered to be valid, for causal inference.

A key concern about the TTE approach is that it cannot adequately deal with situations in which confounding is intractable, whereas other approaches (test-negative design, difference-in-differences, triangulation) can deal with particular confounding scenarios much more validly.

In addition, there are at least two specific disadvantages of the matched-TTE design in that: (i) it is likely to yield similar effect estimates, but wider confidence intervals, compared to a more “standard” analysis which adjusts for the (matched) confounders using all of the available data; (ii) it makes it more difficult, if not impossible, to assess the likelihood, and like strength and direction, of residual confounding.

The idealization of the TTE in teaching as well as in scoring systems sends a strong signal to newer generations of epidemiologists, in that it places the TTE as the first study design to always think about, or even as the “ideal” study design. This will preclude them of seeing other opportunities for interesting and possibly valid comparisons. These include TND studies and other imaginative case-control approaches [39], and natural experiments[4, 8]. In cohort applications, it may also limit studies to those in which follow-up is available from ‘first exposure onwards’, which would seriously handicap chronic disease epidemiology [40], while quite satisfactory methods are available to adjust for time since first exposure when this is appropriate.

In conclusion, the TTE approach, and particularly the matched-TTE approach, has considerable appeal in terms of “marketing” observational studies to policy makers and other non-epidemiologists. However, it cannot deal with situations of intractable confounding - as standard analyses never could - in contrast with other commonly used study designs. In addition, the matched-TTE approach has little scientific benefit since it will usually produce the same findings as an unmatched-TTE approach, or a more conventional data analysis, while having lower precision, and less ability to assess residual confounding. The TTE approach is a useful approach in some circumstances, and complements other important methods, but it is not “the solution”.

Acknowledgements

We thank Matt Fox and Elizabeth Williamson for comments on the draft manuscript, and George Davey Smith for useful discussions of these issues.

References

1. Broadbent A, Vandenbroucke JP, Pearce N, *Formalism or pluralism? A reply to commentaries on 'Causality and causal inference in epidemiology'.* International Journal of Epidemiology, 2016: p. 1841-1851.
2. Vandenbroucke J, Broadbent A, Pearce N, *Causality and causal inference in epidemiology - the need for a pluralistic approach.* International Journal of Epidemiology, 2016: p. 1776-1786.
3. Hernán MA, Robins JM, *Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available.* Am J Epidemiol, 2016. **183**(8): p. 758-64.
4. Dickerman BA, Gerlovin H, Madenci AL, Kurgansky KE, Ferolito BR, Figueroa Muñiz MJ, Gagnon DR, Gaziano JM, Cho K, Casas JP, Hernán MA, *Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans.* N Engl J Med, 2022. **386**(2): p. 105-115.
5. Labrecque JA, Swanson SA, *Target trial emulation: teaching epidemiology and beyond.* Eur J Epidemiol, 2017. **32**(6): p. 473-475.
6. Swanson SA, *Instrumental Variable Analyses in Pharmacoepidemiology: What Target Trials Do We Emulate?* Curr Epidemiol Rep, 2017. **4**(4): p. 281-287.
7. Ben-Michael E, Feller A, Stuart EA, *A Trial Emulation Approach for Policy Evaluations with Group-level Longitudinal Data.* Epidemiology, 2021. **32**(4): p. 533-540.
8. Pearce N, Vandenbroucke JP, *Arguments about face masks and Covid-19 reflect broader methodologic debates within medical science.* Eur J Epidemiol, 2021. **36**(2): p. 143-147.
9. Lawlor DA, Tilling K, Davey Smith G, *Triangulation in aetiological epidemiology.* International Journal of Epidemiology, 2016. **45**: p. 1866-1886.
10. Pearce N, Vandenbroucke J, Lawlor D, *Causal inference in environmental epidemiology: old and new approaches.* Epidemiology, 2019: p. 311-316.
11. Dorn HF, *Philosophy of ifnerences from retrospective studies.* American Journal of Public Health., 1953. **43**: p. 677-683.
12. Cochran WG, *The planning of observational studies of human populations.* Journal of the Royal Statistical Society (Series A), 1965. **128**: p. 234-266.
13. Pearce N, Hensley MJ, *Epidemiologic studies of beta agonists and asthma deaths.* Epidemiol Rev, 1998. **20**(2): p. 173-86.
14. Van der Meer JT, Van Wijk W, Thompson J, Vandenbroucke JP, Valkenburg HA, Michel MF, *Efficacy of antibiotic prophylaxis for prevention of native-valve endocarditis.* Lancet, 1992. **339**(8786): p. 135-9.
15. Dickerman BA, García-Albéniz X, Logan RW, Denaxas S, Hernán MA, *Emulating a target trial in case-control designs: an application to statins and colorectal cancer.* Int J Epidemiol, 2020. **49**(5): p. 1637-1646.
16. Whiteley WN, Ip S, Cooper JA, Bolton T, Keene S, Walker V, Denholm R, Akbari A, Omigie E, Hollings S, Di Angelantonio E, Denaxas S, Wood A, Sterne JAC, Sudlow C, *Association of COVID-19 vaccines ChAdOx1 and BNT162b2 with major venous, arterial, or thrombocytopenic events: A population-based cohort study of 46 million adults in England.* PLoS Med, 2022. **19**(2): p. e1003926.
17. Bosco JL, Silliman RA, Thwin SS, Geiger AM, Buist DS, Prout MN, Yood MU, Haque R, Wei F, Lash TL, *A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies.* J Clin Epidemiol, 2010. **63**(1): p. 64-74.
18. Li S, Stampfer MJ, Williams DR, VanderWeele TJ, *Association of Religious Service Attendance With Mortality Among Women.* JAMA Intern Med, 2016. **176**(6): p. 777-85.
19. Vandenbroucke JP, *Observational research, randomised trials, and two views of medical science.* PLoS Med, 2008. **5**(3): p. e67.

20. Danaei G, García Rodríguez LA, Cantero OF, Logan RW, Hernán MA, *Electronic medical records can be used to emulate target trials of sustained treatment strategies.* J Clin Epidemiol, 2018. **96**: p. 12-22.
21. García-Albéniz X, Hsu J, Bretthauer M, Hernán MA, *Effectiveness of Screening Colonoscopy to Prevent Colorectal Cancer Among Medicare Beneficiaries Aged 70 to 79 Years: A Prospective Observational Study.* Ann Intern Med, 2017. **166**(1): p. 18-26.
22. Davey Smith G, Holmes MV, Davies NM, Ebrahim S, *Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues.* Eur J Epidemiol, 2020. **35**(2): p. 99-111.
23. Freemantle N, Marston L, Walters K, Wood J, Reynolds MR, Petersen I, *Making inferences on treatment effects from real world data: propensity scores, confounding by indication, and other perils for the unwary in observational research.* Bmj, 2013. **347**: p. f6409.
24. Salas M, Hofman A, Stricker BH, *Confounding by indication: an example of variation in the use of epidemiologic terminology.* Am J Epidemiol, 1999. **149**(11): p. 981-3.
25. Vandenbroucke JP, *When are observational studies as credible as randomised trials?* Lancet, 2004. **363**(9422): p. 1728-1731.
26. Vandenbroucke JP, Pearce N, *Test-Negative Designs: Differences and Commonalities with Other Case-Control Studies with "Other Patient" Controls.* Epidemiology, 2019. **30**(6): p. 838-844.
27. Vandenbroucke JP, Brickley EB, Pearce N, Vandenbroucke-Grauls C, *The Evolving Usefulness of the Test-negative Design in Studying Risk Factors for COVID-19.* Epidemiology, 2022. **33**(2): p. e7-e8.
28. Vandenbroucke JP, Brickley EB, Vandenbroucke-Grauls C, Pearce N, *A Test-Negative Design with Additional Population Controls Can Be Used to Rapidly Study Causes of the SARS-CoV-2 Epidemic.* Epidemiology, 2020. **31**(6): p. 836-843.
29. Cowger TL, Murray EJ, Clarke J, Bassett MT, Ojikutu BO, Sánchez SM, Linos N, Hall KT, *Lifting Universal Masking in Schools - Covid-19 Incidence among Students and Staff.* N Engl J Med, 2022.
30. Swanson SA, Tiemeier H, Ikram MA, Hernán MA, *Nature as a Trialist?: Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials.* Epidemiology, 2017. **28**(5): p. 653-659.
31. Iarc Working Group on the Identification of Carcinogenic Hazards to Humans, *IARC Monographs on the Identification of Carcinogenic Hazards to Humans*, in *Opium Consumption*. 2021, International Agency for Research on Cancer: Lyon, France.
32. Greenland S, Morgenstern H, *Matching and efficiency in cohort studies.* Am J Epidemiol, 1990. **131**(1): p. 151-9.
33. Lash TL, Vanderweele TJ, Haneuse S, Rotham KJ, *Modern epidemiology, 4th ed*. 2021, Philadelphia: Wolters Kluwer.
34. Greenland S, *Noncollapsibility, confounding, and sparse-data bias. Part 2: What should researchers make of persistent controversies about the odds ratio?* J Clin Epidemiol, 2021. **139**: p. 264-268.
35. Greenland S, *Noncollapsibility, confounding, and sparse-data bias. Part 1: The oddities of odds.* J Clin Epidemiol, 2021. **138**: p. 178-181.
36. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM, *Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease.* Epidemiology, 2008. **19**(6): p. 766-779.
37. Prentice RL, Langer R, Stefanick ML, Howard BV, Pettinger M, Anderson G, Barad D, Curb JD, Kotchen J, Kuller L, Limacher M, Wactawski-Wende J, Women's Health Initiative I, *Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the*

discrepancy between observational studies and the Women's Health Initiative clinical trial. Am J Epidemiol, 2005. **162**(5): p. 404-414.
38. Vandenbroucke JP, *The HRT controversy: observational studies and RCTs fall in line.* Lancet, 2009. **373**(9671): p. 1233-1235.
39. Smith AH, Pearce NE, Callas PW, *Cancer case-control studies with other cancers as controls.* Int J Epidemiol, 1988. **17**(2): p. 298-306.
40. Vandenbroucke J, Pearce N, *Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology?* Am J Epidemiol, 2015. **182**(10): p. 826-33.