









Causal inference in multi-cohort studies using the target trial framework to identify and minimize sources of bias

Marnie Downes^{*1,2} , Meredith O'Connor^{1,2,3} , Craig A. Olsson^{1,2,4} , David Burgner^{1,2} , Sharon Goldfeld² ,
Elizabeth A. Spry^{1,2,4} , George Patton^{1,2} , Margarita Moreno-Betancur^{1,2} 

¹Department of Paediatrics, The University of Melbourne, 50 Flemington Road, Parkville, Victoria, 3052, Australia

²Murdoch Children's Research Institute, 50 Flemington Road, Parkville, Victoria, 3052, Australia

³Faculty of Education, The University of Melbourne, 100 Leicester Street, Carlton South, Victoria, 3053, Australia

⁴Centre for Social and Early Emotional Development, School of Psychology, Faculty of Health, Deakin University, 75 Pigdons Rd, Waurn Ponds, Victoria, 3216, Australia

*Corresponding author: Marnie Downes, Department of Paediatrics, The University of Melbourne, Reception Level 2, West Building, Royal Children's Hospital, 50 Flemington Road, Parkville, Victoria, 3052 (marnie.downes@unimelb.edu.au)

Abstract

Longitudinal cohort studies, which follow a group of individuals over time, provide the opportunity to examine the causal effects of complex exposures on long-term health outcomes. Utilizing data from multiple cohorts has the potential to add further benefit by improving the precision of estimates through data pooling and by allowing examination of effect heterogeneity through replication of analyses across cohorts. However, the interpretation of findings can be complicated by biases that may be compounded when pooling data or contribute to discrepant findings when analyses are replicated. The “target trial” is a powerful tool for guiding causal inference in single-cohort studies. Here we extend this conceptual framework to address the specific challenges that can arise in the multi-cohort setting. By representing a clear definition of the target estimand, the target trial provides a central point of reference against which biases arising in each cohort and from data pooling can be systematically assessed. Consequently, analyses can be designed to reduce these biases and the resulting findings appropriately interpreted in light of potential remaining biases. We use a case study to demonstrate the framework and its potential to strengthen causal inference in multi-cohort studies through improved analysis design and clarity in the interpretation of findings.

Key words: causal inference; target trial; bias; multi-cohort; data pooling; replication; cohort study; meta-analysis.

Introduction

Causal inference, understood as the examination of the impact of potential interventions,¹ is a common goal in health research, where the ultimate intent is to inform future action that will improve patient or population outcomes. Randomized controlled trials (RCTs) are considered the “gold standard” for causal inference, but it is often not feasible to implement an RCT design.² For example, in child and adolescent health, there is a pressing need to identify targets for preventive intervention³ to counter risk associated with challenges such as obesity,⁴ mental disorders,⁵ and allergic diseases⁶ that can track forward to adult noncommunicable diseases.^{7–11} However, RCTs are limited in their capacity to provide this evidence because of the long time frame for outcomes and ethical issues in randomizing risks like childhood adversity or poverty.

Existing longitudinal cohort studies that follow individuals over long periods of time offer a viable alternative to address these causal questions.^{12–14} However, relative to RCTs, observational studies may suffer from higher risks of bias, particularly confounding, which may threaten the validity of causal inferences. For this reason, there is a long history of describing sources of bias that may arise in analyses of cohort studies that aim

for causal interpretation.^{15–19} Building on this foundation, recent methodological advances have developed improved understanding of types of bias as well as novel concepts and methods to tackle them that are now widely used in cohort research.¹ The “target trial,” defined as the hypothetical randomized experiment that would answer the causal question of interest,² is one such example, representing a powerful concept for guiding the planning, conduct, and interpretation of causal analyses in observational data, including single-cohort studies.^{20–25}

Beyond the potential for causal biases, single-cohort studies may present limitations in terms of sample size, particularly to investigate rare events and exposures,²⁶ as well as effect modification, among subgroups. Also, the focus of single cohorts on specific settings, populations, and/or epochs may raise questions about the generalizability of findings to other contexts.^{27,28} Furthermore, it is possible that estimated effect sizes, even in large, well-designed studies, may be exaggerated¹² and not replicable when investigations are repeated on new data.²⁹

Given their capacity to address these issues, multi-cohort studies—studies involving analyses of individual-level data from multiple independent cohorts—have become increasingly common,^{30–39} as well as more feasible with platforms arising to

Received: November 28, 2022. Accepted: August 14, 2024

© The Author(s) 2024. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

support data access.⁴⁰ There are 2 key motivations for multi-cohort designs. First, more precise estimation of a causal effect can be achieved by the integration of harmonized data from multiple cohorts into a single data set, on which analyses are performed directly (*pooled data analysis, also known as 1-step individual participant meta-analysis*),^{26,41} or by synthesis of effect estimates obtained from analyses of individual-level data from each cohort separately (*2-step individual participant meta-analysis*).^{42,43} Second, investigation of effect heterogeneity across contexts, defined here as a difference in true causal effects that are distinct in only 1 aspect of their definition (eg, target populations considering different time periods or places, or outcomes measured at different time points), may be conducted if this difference is a key distinguishing feature across the cohorts available. This may be achieved by obtaining and comparing estimates in each cohort separately (*replication of analyses with comparison*), by random-effect 2-step individual participant meta-analysis with the estimate of the between-study variance providing a measure of heterogeneity, or when using outcome regression as a confounding adjustment approach by pooled data analysis, including a cohort-by-exposure interaction term. Elsewhere,²⁸ we have summarized these approaches and recognized other motivations (and associated methods) for multi-cohort studies, such as the capacity to address interrelated components of a complex theoretical model.

It is now timely to draw together these 2 major shifts in cohort-based health research: the target trial framework and the use of multi-cohort designs. As with any causal problem, the reliability of causal inferences arising from the application of any multi-cohort analytic approach is dependent on a thorough understanding of sources of bias. Compared with single-cohort studies, undertaking causal inference in multi-cohort studies faces additional challenges in this regard. When pooling data to improve precision, biases arising in each cohort may be compounded and new biases introduced. Meanwhile, when replicating analyses in each cohort separately to investigate effect heterogeneity, interpretation may be complicated as different biases within each cohort may contribute to and thus distort differences in estimated effects. This makes a clear definition of the target estimand and systematic consideration of sources of bias critically important in the conduct of multi-cohort studies. There is currently little guidance on how to do this, with the need for thorough planning recently highlighted.⁴⁴

The aim of this article is to propose how the target trial framework can be extended to address the specific challenges that arise when undertaking causal inference in multi-cohort studies. We review the approach in single-cohort studies, describe the extension to multi-cohort studies, and then use a case study to demonstrate its value in understanding potential biases associated with multi-cohort studies to inform analysis design and interpretation. We conclude by offering some guidance on the selection of existing analytic methods for addressing identified biases.

The target trial framework in single-cohort studies

The target trial framework involves 2 steps. The first is to specify the target trial, defined as the hypothetical randomized experiment that would ideally be implemented to answer the causal research question of interest.² A detailed description is developed by specifying key protocol components, including eligibility criteria, treatment strategies, assignment procedures, follow-up period, and outcome. Articulation of the target trial yields refined

definitions of the research question and corresponding estimand (ie, the causal effect of interest).²⁰

The second step is to consider the assumptions under which one may emulate the target trial with the observational data available to obtain an unbiased estimate of the causal effect. Bias refers to a discrepancy between the estimate of the causal effect of interest from a study (on average over replications) and its true value (which is typically unknown in practice). In emulating each of the target trial protocol components, there are corresponding analysis decisions to consider (analytic sample selection, treatment/exposure measure, selection of confounders, timing of measures, outcome measure), each of which can reduce or introduce bias depending on the assumed causal relationships between the study variables. A causal diagram or directed acyclic graph (DAG) that describes these relationships helps to develop a detailed understanding of potential biases to consider in the target trial emulation.¹ Armed with this understanding, analysis decisions can be made to (1) reduce or counter potential biases, (2) avoid introducing new biases, and (3) allow for thoughtful interpretation of findings considering potentially remaining biases.

Specifically, there are 3 key causal biases to consider: selection bias, confounding bias, and measurement bias (Table 1). In the DAGs in Figures 1A, 2A, and 3A, these biases are represented as arising from biasing paths between exposure and outcome that introduce a noncausal association. While in practice, it is generally preferable to develop a single DAG describing the assumed relationships between all important variables, here we have chosen to present 3 separate simplified DAGs to demonstrate how DAGs can be used to understand sources of each type of causal bias. We note also that here and in subsequent sections, we focus on the simple but common setting of a binary or categorical point treatment (exposure) and a binary or continuous point outcome. We do not consider the more complex settings of sustained treatment strategies and time-to-event outcomes, as well as related issues like immortal time bias, which are described in depth elsewhere for single-cohort studies²¹ and represent key areas in which the target trial framework is widely known to be helpful in limiting bias risks.

The target trial framework in multi-cohort studies

We first consider application of the target trial framework when combining cohorts with the purpose of improving precision using pooled data analysis and then highlight how specification of the target trial changes when the framework is applied to investigate effect heterogeneity by replicating analyses in each cohort separately. Although we do not specifically focus on this, it is worth noting that many of the important considerations when combining data from multiple cohorts are also relevant when combining data from multiple RCTs.⁴⁸

When pooling cohort data to improve precision, the initial step, as in single-cohort studies, is to define the causal estimand by specifying the target trial. The next step is to consider emulating the target trial within each cohort separately, where each emulation strategy is tailored specifically to the individual cohort's design features and observed data available. Indeed, we propose that the critical point to the application of this framework in the multi-cohort setting is that the specified target trial provides the central point of reference for the identification of potential sources of biases that inform the emulation strategy in each cohort, as opposed to comparing one cohort to another, which

Table 1. A summary of the 3 key causal biases that are important to consider when working through a target trial emulation: selection bias, confounding bias, and measurement bias.

Bias	Description	Target trial protocol component where emulation can give rise to bias (related analytic design aspect)
Selection bias	<p>Occurs when the sample used for analysis (the analytic sample) is not representative of the population for whom inference is sought (the target population) due to, for example, individuals with certain characteristics being more likely to not participate or be lost from the study over time. Bias can arise in the selection of the cohort sample (the sample of participants recruited into the cohort, ie, those who are both eligible and consent to participate) from the target population and also in the selection of the analytic sample from the cohort sample.</p> <p>Lu et al⁴⁵ distinguish between 2 types of selection bias:</p> <ul style="list-style-type: none"> • Type 1 selection bias arises due to restricting to or conditioning on 1 (or more) level(s) of a common effect of 2 variables (known as a "collider"), one of which is either the exposure or a cause of the exposure, and the other is either the outcome or a cause of the outcome. Formally, in a causal diagram or directed acyclic graph (DAG), type 1 selection bias is represented as arising from a noncausal (biasing) path that becomes open after restricting to or conditioning on a collider (see Figure 1A). • Type 2 selection bias arises due to restricting to 1 (or more) level(s) of an effect measure modifier (a variable by which the magnitude of the effect of the exposure on the outcome differs across strata). It is not possible to depict by way of biasing paths in a DAG as for the other biases. 	Eligibility criteria (analytic sample selection)
Confounding bias	<p>Arises from differences between the exposure groups in terms of individual, pre-exposure characteristics that are also related to the outcome. Formally, in a DAG, confounding bias results from an open noncausal (biasing) "backdoor" path between the exposure and the outcome that remains even if all arrows pointing from the exposure to other variables (the descendants of the exposure) are removed (ie, the path has an arrow pointing into the exposure)¹ (eg, see Figure 2A).</p>	Assignment procedures (confounder selection)
Measurement bias	<p>Refers to bias that arises as a result of measurement error, which reflects a discrepancy between the measured value of a quantity and its true value, as determined by a gold standard instrument. This includes misclassification in the case of categorical data. Measurement bias can be depicted in a DAG in many ways; Figure 3A provides an example of bias arising from measurement error in the outcome, with a noncausal (biasing) path that becomes open when using the measured outcome.</p>	<p>Treatment strategies (exposure measurement)</p> <p>Follow-up period (timing of measurements)</p> <p>Outcome (outcome measurement)</p>

is the usual tendency. For example, specifying the eligibility criteria characterizes the target population for whom inference is sought. This is usually a broad population from which the sample participants of each cohort study have been sourced. Each study sample will only imperfectly represent the target population, and the deviation (and thus the potential for selection bias) may differ across cohorts. Recognizing this discrepancy is only possible through specification of a single target trial with a clear target population.

More generally, there are 2 distinct sources of bias when pooling data from multiple cohorts. First, there may be causal biases in the emulation of the target trial in each cohort. We refer to these as "within-cohort" biases. These biases may operate in the same or opposite directions, so combining multiple cohorts together can result in a compounding or reduction in bias overall. Second, additional biases may be introduced through systematic differences between cohorts in terms of design features such as

calendar period, geographic location, recruitment and assessment methods, and personnel. We refer to these as "across-cohort" biases. These biases may take the form of selection bias due to additional common causes of study participation or missing data and the outcome, confounding bias due to additional (measured or unmeasured) common causes of exposure and outcome, and measurement bias due to measurement error being different between the cohorts, which may be exacerbated in the harmonization process required to create a single integrated data set (see Figures 1B, 2B, and 3B, respectively).

Examination of within-cohort biases allows evaluation of how well the target trial is emulated in each cohort, while consideration of across-cohort biases enables appraisal of the additional biases arising as a result of data pooling. This dissection of sources of bias would not be possible if cohorts are compared to each other, with no clarity about the causal estimand. Using this approach, if it is deemed that the emulation of the target

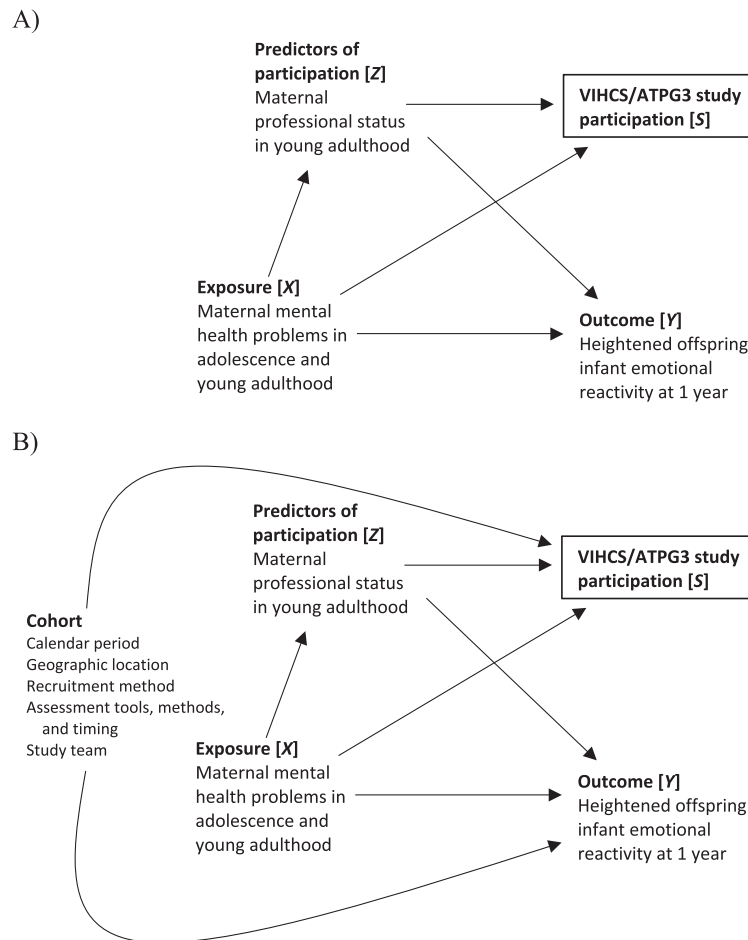


Figure 1. Directed acyclic graphs (DAGs) depicting examples of (A) type 1 selection bias in a single cohort (within-cohort selection bias) and (B) additional type 1 selection biases in pooled data analyses of multiple cohorts (across-cohort selection bias).

trial is more problematic in one cohort than another, or in other words, if there is considerably more potential for bias in one cohort than another (perhaps due to important study design features that do not meaningfully align or harmonization decisions that require a substantial loss of information), then it may be more appropriate to proceed with a single-cohort study or, in the case of several cohorts, drop one and go ahead with a reduced multi-cohort study. For example, a multi-cohort study using pooled data analysis to increase precision will, in practice, be appropriate only when the source populations from which the individual cohorts were recruited are relatively well aligned with the target population.

When replicating analyses in each cohort separately to investigate effect heterogeneity, the description of the target trial must explicitly state the specific varying characteristic of interest (eg, a difference in geographic location of the target population should be included in the eligibility criteria of the target trial). While technically this defines multiple target trials, each one identical except for the specific varying characteristic of interest, for brevity, we henceforth continue to refer to a single-target trial. Cohort differences with regard to the specific characteristic of interest will not be identified as within-cohort biases, while cohort differences relative to the target trial in all other aspects (eg, recruitment in different calendar periods) could be identified as within-cohort biases. An estimated difference in the causal effects across cohorts may thus arise due to effect heterogeneity (an actual difference in the true causal effects), different

within-cohort biases in each target trial emulation, or random variability. Unfortunately, it is often extremely challenging to pinpoint the source of discrepancies. Use of the target trial framework will help identify and thus plan to minimize within-cohort biases as much as possible, as well as outline potential remaining biases to inform interpretation.

Example case study

We consider the published study by Spry et al,⁴⁹ which utilized data from 2 Australian longitudinal cohort studies (Table 2) to examine the extent to which preconception maternal mental health in adolescence and young adulthood affects children's early life behavioral outcomes. We do not conduct new analyses here. Rather, by articulating a relevant but previously unspecified target trial and considering the emulation strategies implicit in the analytic decisions adopted in the publication, we illustrate how the framework (1) sheds light on potential within- and across-cohort biases, (2) enables critique of the analysis methods used regarding the extent to which biases were addressed, and (3) illuminates the interpretation of published findings. Although the framework can be used in this way to appraise published studies, it is preferably applied prospectively when planning analyses for a multi-cohort study.

The study aimed to obtain precise estimates of the causal effects of preconception maternal mental health problems in both adolescence and young adulthood on infant emotional reactivity at 1 year postpartum through a pooled data analysis of

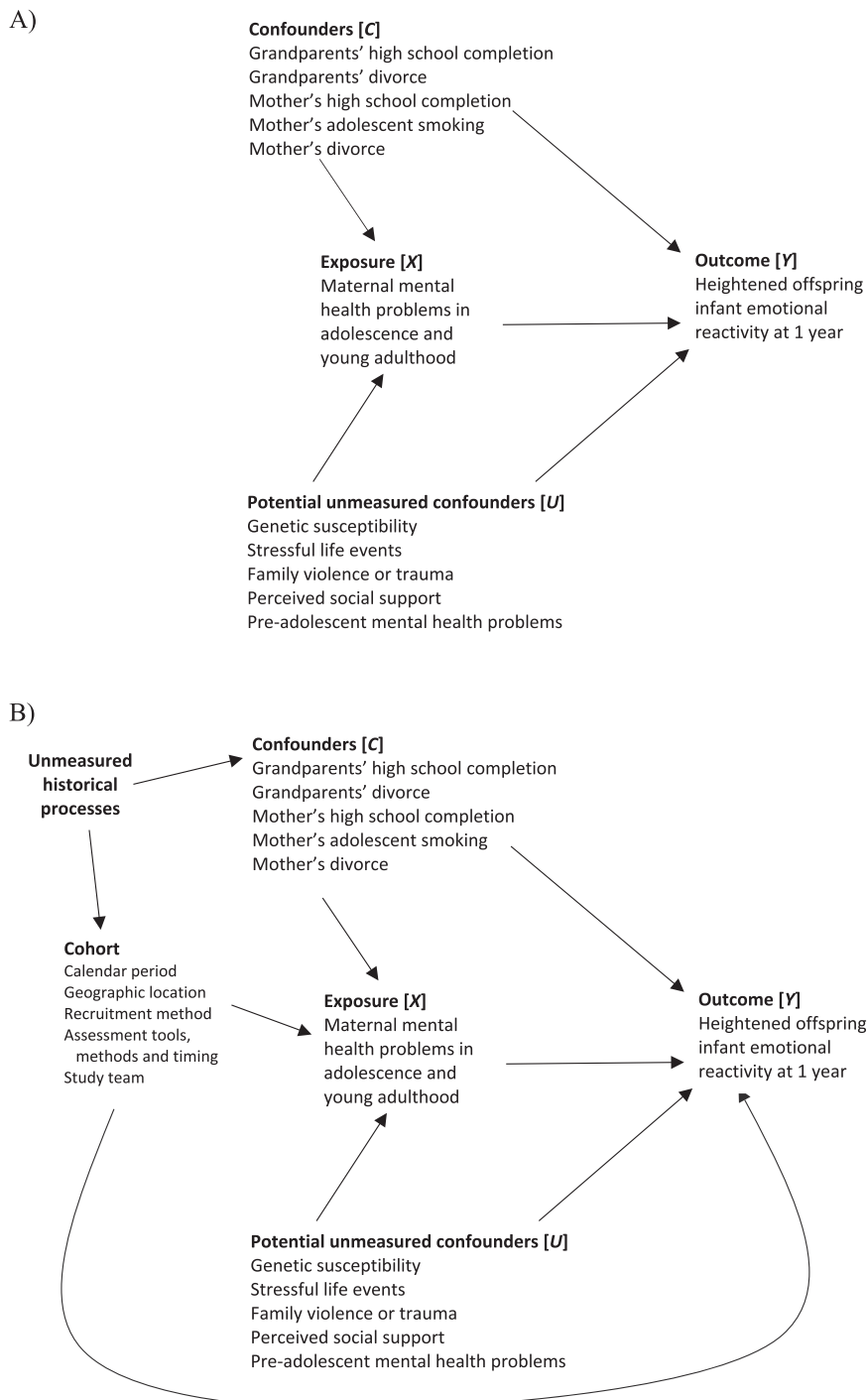


Figure 2. Directed acyclic graphs (DAGs) depicting examples of (A) confounding bias in a single cohort (within-cohort confounding bias) and (B) additional confounding biases in pooled data analyses of multiple cohorts (across-cohort confounding bias). (Following the work of VanderWeele and Robinson⁴⁶ and Moreno-Betancur et al,⁴⁷ we include a node in B to represent the unmeasured historical processes that are common causes of the cohort characteristics and the individual characteristics considered as confounders.)

the 2 cohorts. Estimates were obtained from a logistic regression model adjusted for cohort and selected confounders, fitted within a generalized estimating equation framework to account for within-family clustering. These estimates can be interpreted as conditional causal odds ratios under the assumption of no effect modification by any of the adjusting variables and that the outcome regression model is correctly specified,¹ in addition to the usual causal assumptions of conditional exchangeability,

consistency, and positivity, including no bias due to measurement error or missing data. Results of the pooled data analysis were reported as the primary finding, and as is good practice, replication of analyses was reported as secondary analyses to explore the consistency of findings across cohorts. As expected, pooled data analysis achieved superior precision, and some degree of discrepancy between the cohort-specific causal effect estimates was observed (Table 3).

Table 2. A summary of the 2 Australian longitudinal cohort studies utilized in the published multi-cohort study by Spry et al⁴⁹ to examine the extent to which preconception maternal mental health in adolescence and young adulthood affects children's early life behavioral outcomes.

	VIHCS	ATPG3
Type of study	A prospective longitudinal study ⁵⁰ arising from an existing cohort study, the Victorian Adolescent Health Cohort Study. ⁵¹	A prospective longitudinal study ⁵² arising from an existing cohort study, the Australian Temperament Project. ⁵³
Original cohort study		
Recruitment	VAHCS commenced in 1992 and recruited a sample of mid-secondary school students aged 14 to 15 years (<i>n</i> = 1943; 1000 female) from Victoria, Australia.	ATP commenced in 1983 and recruited a sample of infants aged 4 to 8 months (<i>n</i> = 2443; 1170 female), along with their parents, through maternal and child health centers in urban and rural local government areas in Victoria, Australia.
Follow-up	Participants were assessed at 6-monthly intervals during adolescence and 3 times in young adulthood.	Families were invited to complete mail surveys every 1 to 2 years until 19 to 20 years of age and every 4 years thereafter.
Subsequent intergenerational cohort study		
Screening and recruitment	Six-monthly screening of VAHCS female participants between 2006 and 2013 (age 29-35 years) for pregnancies or recently born infants.	Six-monthly screening of ATP female participants between 2012 and 2018 (age 29-35 years) for pregnancies or recently born infants.
Follow-up	VIHCS participants were asked to complete telephone interviews in trimester 3, as well as 2 months and 1 year postpartum, for each infant born during screening, with follow-up assessments continuing into offspring childhood and adolescence.	ATPG3 participants were asked to complete interviews in trimester 3, as well as 2 months and 1 year postpartum, for each infant born during screening, with follow-up assessments continuing into offspring childhood.

ATP, Australian Temperament Project; ATPG3, Australian Temperament Project Generation 3; VAHCS, Victorian Adolescent Health Cohort Study; VIHCS, Victorian Intergenerational Health Cohort Study.

Application of the target trial framework to the case study

Table 4 outlines a proposed target trial and corresponding emulation strategies for the Victorian Intergenerational Health Cohort Study (VIHCS) and the Australian Temperament Project Generation 3 (ATPG3) implicit in the analysis approach described in Spry

et al.⁴⁹ Considering each protocol component in turn, we first identify the within-cohort biases (eg, Figures 1A, 2A, 3A), which may have been compounded in the pooled data analysis and may explain observed discrepancies in the replication of analyses. We then describe additional across-cohort biases that may have arisen in the pooled data analysis (eg, Figures 1B, 2B, 3B). In these DAGs, an additional node representing a cohort indicator is

Table 3. Estimated causal effects (expressed as odds ratios with 95% CIs) of preconception maternal mental health problems in adolescence and young adulthood on heightened offspring infant emotional reactivity in pooled data analysis and replication of analyses reported in Spry et al,⁴⁹ including adjustment for the identified confounders listed in Table 4 and Figure 2.

	Odds ratio	95% CI
<i>Pooled data analysis</i>		
No preconception maternal mental health problems (unexposed)		
Preconception maternal mental health problems (exposed)		
In adolescence only	1.3	(0.9, 2.0)
In young adulthood only	1.3	(0.7, 2.1)
In adolescence and young adulthood	2.1	(1.4, 3.1)
<i>Replication of analyses</i>		
VIHCS		
No preconception maternal mental health problems (unexposed)		
Preconception maternal mental health problems (exposed)		
In adolescence only	1.2	(0.6, 2.3)
In young adulthood only	1.8	(0.8, 3.9)
In adolescence and young adulthood	2.4	(1.3, 4.2)
ATPG3		
No preconception maternal mental health problems (unexposed)		
Preconception maternal mental health problems (exposed)		
In adolescence only	1.5	(0.8, 2.7)
In young adulthood only	0.9	(0.4, 2.0)
In adolescence and young adulthood	1.9	(1.1, 3.4)

Abbreviations: ATPG3, Australian Temperament Project Generation 3; VIHCS, Victorian Intergenerational Health Cohort Study.

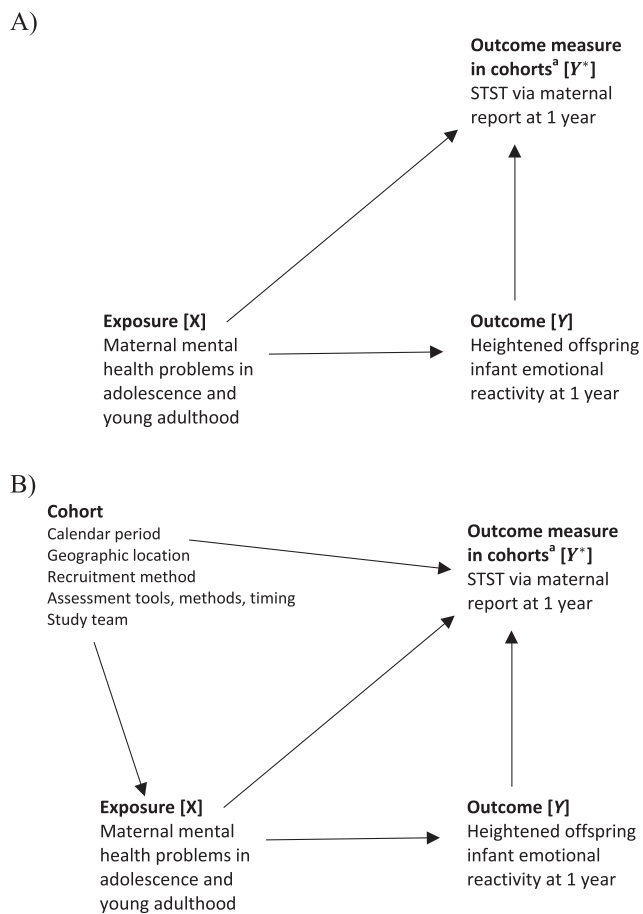


Figure 3. Directed acyclic graphs (DAGs) depicting examples of (A) measurement bias in a single cohort (within-cohort measurement bias) and (B) additional measurement biases in pooled data analyses of multiple cohorts (across-cohort measurement bias). ^aFor the case of k cohorts, there could be up to k different outcome measures Y_i^* , $i = 1, \dots, k$, from which a pooled outcome measure Y^{**} would be derived. STST, Short Temperament Scale for Toddlers.

introduced as a proxy for all those cohort characteristics, whether measured or unmeasured, that are fixed within a cohort but vary across cohorts.

A second case study⁵⁴ is described in [Appendix S1](#), and [Table S1](#) highlights, in the context of this example, additional considerations in the application of the framework when aiming to investigate effect heterogeneity using replication of analyses.

Eligibility criteria

The target population of interest is defined as young adolescent females (13 years) in Victoria, Australia, in the 1990s. The original cohorts of the Victorian Adolescent Health Cohort Study (VAHCS) and the Australian Temperament Project (ATP) were designed to recruit close-to-representative samples of this target population, but individuals from families who could not speak English were excluded. Hence, there is potential for within-cohort type 2 selection bias (see [Table 1](#)) if the effect of the exposure on the outcome in mothers from non-English-speaking families is different from the effect in those from English-speaking families.

There is further potential for within-cohort selection bias, not only because some eligible individuals might not have consented to participate in the original cohorts, but also because those who did may still not have been eligible to participate in the subsequent intergenerational studies. Specifically, eligibility for

recruitment to VIHCS/ATPG3 required female participants of the VAHCS/ATP to have fallen pregnant and had a live birth during the screening period. There is potential for type 1 selection bias (see [Table 1](#)) if becoming pregnant and having a live birth outside the age range of 29 to 35 years is a common effect of the exposure and unmeasured variables (eg, problematic maternal drug use during adolescence and young adulthood) that are also causes of the outcome. Note that in settings like this, where the outcome is defined after a post-treatment index event (here, live birth), the choice of estimand is a complex problem often without a completely satisfactory solution.⁵⁵ The target trial specified here examines so-called total effects, but whether this is the most meaningful estimand is debatable. Another potential source of selection bias is that eligible individuals might not have consented to participate in the VIHCS/ATPG3. The DAG in [Figure 1A](#) depicts one possible example of type 1 selection bias due to nonparticipation, with maternal professional status in young adulthood a common cause of VIHCS/APG3 study participation and the outcome. Conditioning on study participation leads to a biasing path between the exposure and the outcome via maternal professional status.

Additionally, given the intergenerational nature of these studies, and despite high retention rates, there is potential for within-cohort type 1 selection bias due to missing data in any analysis variable, for example, due to study dropout, if the analysis is restricted to participants without missing data (ie, “complete cases”). This could be depicted in a DAG similar to [Figure 1A](#), with the study participation node replaced with a “complete case” indicator. Spry et al⁴⁹ used multiple imputation to handle missing data,⁵⁶ which can address potential selection bias if auxiliary variables that predict both missingness and dropout are included.

There is also the potential for across-cohort type 1 selection biases due to systematic between-cohort differences in factors that predict both the VIHCS/ATPG3 study participation and the outcome. For example, the 2 original cohorts utilized different recruitment strategies: the VAHCS in adolescence through secondary schools and the ATP in infancy through maternal-child health services. Different factors during these distinct life stages may have influenced family decisions to consent to participate in the VAHCS/ATP, which was a prerequisite to participating in the VIHCS/ATPG3. Furthermore, other between-cohort differences such as period of recruitment may be predictors of the outcome. Together, this creates a further biasing path between exposure and outcome via cohort that is introduced by the restriction to VIHCS/ATPG3 participants ([Figure 1B](#)). In published supplementary analyses of complete cases, cohort differences in factors that are common causes of missing data and the outcome, such as study personnel and assessment tools/methods, may have introduced additional type 1 selection bias (again similar to the DAG in [Figure 1B](#), with the VIHCS/ATPG3 study participation node replaced with a “complete case” indicator).

Treatment strategies

The exposure of interest is preconception maternal mental health problems in adolescence (13–18 years) and young adulthood (19–29 years), considered here for simplicity as a categorical point exposure. When considering the target trial, this is an example of an imprecisely defined intervention. While an important intervention target, it is not clear how mental health problems might be changed or treated. It could be medication, individual therapy, and/or population-level interventions. This lack of precise articulation within the target trial complicates the interpretation of

Table 4. Proposed target trial and emulation strategies implicit in the statistical analysis approach of Spry et al⁴⁹ for precise estimation of the causal effects of preconception maternal mental health problems in both adolescence and young adulthood on offspring infant emotional reactivity.

Protocol component	Target trial	Emulation strategies
A) Eligibility criteria	<p>Target trial</p> <p>Target population: Young adolescent females (13 years of age) in Victoria, Australia, in the 1990s</p>	<p>ATPG3</p> <p>Analytic sample selection: Female ATP study participants (recruited through rural and urban Victorian Maternal and Child Health centers at 4-8 months of age in 1983), who subsequently reported pregnancy or recently born infant between 2012 and 2018 (29-35 years) when screened and consented to participate in the ATPG3.</p> <p>Approach to handling missing data and other potential sources of selection bias: All ATPG3 participants were retained in the sample regardless of missing data via use of multiple imputation. There was no attempt to address selection bias due to conditioning on ATP participation, having a live birth, and ATPG3 participation.</p> <p>Treatment/exposure measure:</p> <p><i>Intervention arms:</i> The presence of any mental health problems at ≥ 1 wave in: 1. Adolescence (13-18 years; ATP waves 10-12) only 2. Young adulthood (19-28 years; ATP waves 13-15) only 3. Adolescence (ATP waves 10-12) and young adulthood (ATP waves 13-15)</p> <p><i>Comparator arm:</i> No mental health problems at any wave in adolescence or young adulthood (ATP waves 10-15) Mental health problems measure: Wave 10: SMFQ ≥ 11 or RBPQSF mean > 1 Waves 11-12: SMFQ ≥ 11 or KCMAS mean > 1 Waves 13-15: DASS-21, Depression ≥ 7, Anxiety ≥ 6, or Stress ≥ 10</p>
B) Treatment strategies	<p>Treatment arms in the trial:</p> <p><i>Intervention arms:</i> Preconception maternal mental health problems in: 1. Adolescence (13-18 years) only 2. Young adulthood (19-29 years) only 3. Adolescence (13-18 years) and young adulthood (19-29 years)</p> <p><i>Comparator arm:</i> No preconception maternal mental health problems in adolescence (13-18 years) or young adulthood (19-29 years)</p>	<p>VIHCS</p> <p>Analytic sample selection: Female VAHCS study participants (a sample of 1992 Victorian mid-secondary school students aged 14-15 years), who subsequently reported pregnancy or recently born infant between 2006 and 2013 (29-35 years) when screened and consented to participate in the VIHCS.</p> <p>Approach to handling missing data and other potential sources of selection bias: All VIHCS participants were retained in the sample regardless of missing data via use of multiple imputation. There was no attempt to address selection bias due to conditioning on VAHCS participation, having a live birth, and VIHCS participation.</p> <p>Treatment/exposure measure:</p> <p><i>Intervention arms:</i> The presence of any mental health problems at ≥ 1 wave in: 1. Adolescence (14-18 years; VAHCS waves 2-6) only 2. Young adulthood (19-29 years; VAHCS waves 7-9) only 3. Adolescence (VAHCS waves 2-6) and young adulthood (VAHCS waves 7-9)</p> <p><i>Comparator arm:</i> No mental health problems at any wave in adolescence or young adulthood (VAHCS waves 2-9) Mental health problems measure: Waves 2-7: CIS-R ≥ 12 Waves 8-9: GHQ-12 ≥ 3</p>
C) Assignment procedures	<p>Randomization strategy: Randomization at recruitment without blind assignment</p>	<p>Selection of confounders: Confounder (harmonized self-reported measure)</p> <ul style="list-style-type: none"> • Mother's parents' high school completion (neither parent vs at least 1 parent completed) • Mother's parents' divorce/separation during or before adolescence (ever vs never divorced/separated) • Mother's high school completion (ever vs never completed) • Mother's adolescent smoking (daily smoking at ≥ 1 adolescent wave vs no daily smoking) • Mother's history of divorce/separation (ever vs never divorced/separated) <p>Approach to confounding adjustment: Outcome regression</p> <p>Timing of measures: Start: VAHCS wave 2 (mother aged 14-15 years old) End: VAHCS wave 3 (child aged 1 year old)</p> <p>Outcome measure: STST via maternal report, mean score ≥ 4</p>
D) Follow-up period	<p>Start and end times: Start: At randomization (mother aged 13 years) End: Child aged 1 year</p>	<p>Outcome regression</p> <p>Timing of measures: Start: ATP wave 10 (mother aged 13-14 years old) End: ATPG3 wave 3 (child aged 1 year old)</p> <p>Outcome measure: STST via maternal report, mean score ≥ 4</p>
E) Outcome	<p>Outcome measure: Heightened offspring infant emotional reactivity determined through triangulation of parent-report, clinician ratings, and direct observation of infant behavior</p>	<p>Outcome regression</p> <p>Timing of measures: Start: VAHCS wave 2 (mother aged 14-15 years old) End: VAHCS wave 3 (child aged 1 year old)</p> <p>Outcome measure: STST via maternal report, mean score ≥ 4</p>
F) Causal contrasts of interest and causal effect measure	<p>Conditional odds ratio comparing risk of heightened offspring infant emotional reactivity in each intervention arm relative to the comparator arm in the target population</p>	<p>Outcome regression</p> <p>Timing of measures: Start: ATP wave 10 (mother aged 13-14 years old) End: ATPG3 wave 3 (child aged 1 year old)</p> <p>Outcome measure: STST via maternal report, mean score ≥ 4</p>

Abbreviations: ATP, Australian Temperament Project; ATPG3, Australian Temperament Project Generation 3; CIS-R, Clinical Interview Schedule-Revised; DASS-21, Depression Anxiety Stress Scales; GHQ-12, General Health Questionnaire; RBPQSF, Revised Behavior Problem Checklist Short Form; KCMAS, Revised Children's Manifest Anxiety Scale; SMFQ, Short Mood and Feelings Questionnaire; STST, Short Temperament Scale for Toddlers; VAHCS, Victorian Adolescent Health Cohort Study; VIHCS, Victorian Intergenerational Health Cohort Study.

findings and the selection of confounders.⁴⁶ While a common issue in studies asking complex causal questions of this nature, it does not preclude the examination of such questions; rather, it reinforces the need for clear causal thinking and, in particular, the benefits of using the target trial framework to highlight study strengths and limitations.²⁰

Beyond the issue of imprecisely defined interventions, measuring mental health problems is inherently challenging. It is not directly observable, nor can any informant provide the “true” perspective.⁵⁷ Symptoms can also manifest differently over contexts and time.⁵⁸ While well-regarded assessment tools are available, the best such tools can hope to achieve is to approximate the intended underlying construct, with some degree of measurement error unavoidable. In this example, different measurement tools were used across cohorts and also across waves within each cohort, and it is possible that these capture overlapping but slightly different constructs. Additionally, both cohorts relied on mothers’ self-report only, which, while valuable, may be influenced by feelings of guilt, shame, or embarrassment.⁵⁹ Such measurement issues in the exposure could lead to within- and across-cohort measurement biases, as is described in more detail below for the outcome measure.

Assignment procedures

In the target trial, individuals would be randomized to experiencing mental health problems or not. Clearly, however, even if a specific intervention had been defined to achieve this, it would be unethical to randomize it. Focusing on the available data, a number of measured confounders were identified, based on prior evidence in the literature, representing socioeconomic circumstances and adolescent smoking (Figure 2A). Even after adjusting for these confounders, there remains potential for residual within-cohort confounding bias due to biasing paths via unmeasured confounders (also Figure 2A). There is also potential for residual within-cohort confounding bias due to confounder measurement error as a result of using proxies (eg, high school completion and divorce used as proxies for socioeconomic circumstances) and inaccurate reporting of sensitive information (eg, smoking history or family divorce).

There is potential for exacerbation of residual within-cohort confounding bias due to measurement error in the necessary harmonization of confounder variables. For example, measurement of maternal education was not aligned between the VIHCS and ATPG3; hence, there was some loss of information in the blunt harmonized measure of “never completed” vs “ever completed” high school. Finally, some additional confounders may have been available in only 1 of the 2 cohorts, so they were not included in the adjustment set, thus increasing the potential for confounding bias.

Additional across-cohort confounding biases may arise in the pooled data analysis due to additional biasing pathways between exposure and outcome because of systematic between-cohort differences in measured or unmeasured design aspects (Figure 2B). For example, recruitment of mothers and their infants took place in 2006–2013 for the VIHCS and in 2012–2018 for the ATPG3, with period effects on exposure and outcome assessment potentially inducing additional confounding bias.

Follow-up period

Consistent with the target trial, follow-up in both cohorts commenced when mothers were in adolescence and concluded at 1 year postpartum. There is potential for within-cohort measurement error (and thus bias) as a result of the exposure and outcome

measurements not being taken at the exact ages specified in the target trial for all participants.

Additional across-cohort measurement bias is possible due to systematic between-cohort differences in the timing of exposure and outcome measurements. For example, there was variation between (as well as within) cohorts in the timing of measuring the 1-year postpartum outcome.

Outcome

Both cohorts used the Short Temperament Scale for Toddlers (STST) via maternal report at 1 year postpartum, with a mean score of ≥ 4 indicative of heightened offspring infant emotional reactivity, in contrast to the target trial, which would triangulate perspectives from parents and clinicians with direct observation of infant behavior. Therefore, there is potential for measurement error and thus bias in the use of the STST as a proxy for a more comprehensive assessment of infant emotional reactivity. Additionally, Spry et al⁴⁹ identified that maternal report of infant outcomes may be affected by a mother’s mental state such that depressed mothers perceive their infant as more reactive. This measurement bias due to inaccurate reporting is depicted in Figure 3A as a biasing path between the exposure and the outcome via the measured outcome.

There could be additional across-cohort bias in the pooled data analysis if there were systematic between-cohort differences in measurement of the outcome (Figure 3B). This is less of a concern here, however, since infant emotional reactivity was measured consistently in both cohorts. If different outcome measures were used across cohorts, the DAGs in Figure 3A and 3B would become more complex, depicting the cohort-specific outcome measures Y_i^* ($i = 1, \dots, k$) and the pooled outcome measure Y^{**} , respectively.

Guidance for planning and reporting analyses to address identified biases

Once within- and across-cohort biases have been identified via application of the target trial framework, a comprehensive analysis plan to counter them can be developed.⁶⁰ It will not be possible to completely eliminate all causal biases, but the aim is to devise an analysis strategy that diminishes bias threats as much as possible and, importantly, does not introduce new ones, for example, through conditioning on a “collider” (see Table 1). This section provides general guidance on the selection of existing analytic approaches for addressing potential sources of each type of causal bias in multi-cohort studies.

Selection bias

Missing data methods such as multiple imputation^{56,61} and inverse probability weighting (IPW)^{62,63} are widely utilized for countering selection bias due to study participation, loss to follow-up, and other missing data. Modern implementations of multiple imputation^{64,65} provide a flexible approach to handle complex multivariable missingness problems, allowing specification of tailored imputation models, including all analysis variables, in addition to predictors of incomplete variables that may also be predictors of missingness (eg, maternal professional status in Figure 1A). In pooled analyses, the cohort indicator is one such variable (Figure 1B) and, as such, should be included as a covariate in multiple imputation. Alternatively, in replication of analyses, it is recommended that multiple imputation be performed in each cohort separately,⁶⁶ particularly when cohorts represent distinct populations, settings, and/or time periods. This was the approach

taken by Spry et al⁴⁹ and allows imputation models to be tailored to each cohort.

It is worth noting that there is no preferred method for addressing selection bias due to multivariable missing data. The most appropriate approach for a given problem depends primarily on the target estimand and the multivariable missingness assumptions, which determine identifiability. Recent work has used “missingness DAGs” (DAGs expanded to include variable-specific missingness indicators) to depict missingness assumptions in a single-cohort setting and then derive the identifiability of a range of estimands based on these assumptions.^{67,68} These results can in turn be used to guide the method for handling missing data.⁶⁷⁻⁶⁹ Extensions to the multi-cohort setting would be valuable.

Meta-analysis methods for transporting inferences from multiple RCTs to a new target population⁷⁰⁻⁷² may be useful to specifically address type 2 selection bias in pooled data analysis. However, extensions are needed to allow simultaneous adjustment for confounding and measurement bias,^{45,70,71} both of which are important considerations when working with observational data.

Confounding bias

There are 2 classes of analytic approaches for addressing confounding bias: conditioning-based methods (eg, multivariable outcome regression) and standardization-based methods (or “G-methods,” eg, IPW, g-computation).¹ These methods require modeling either the outcome or the exposure based on the selected confounder set. So-called doubly robust methods use models for both processes and reduce the risk of model misspecification bias due to their good performance when at least one of the models is consistently estimated and because they can be coupled with machine learning.⁷³ Regardless of the analytic method, it is critical to include the cohort indicator as an additional confounder in a pooled data analysis to address any across-cohort confounding bias due to systematic differences in study design features. This was the approach taken by Spry et al,⁴⁹ effectively closing the additional noncausal path between exposure and outcome via cohort that is opened by combining the 2 cohorts (Figure 2B).

Measurement bias

Addressing sources of measurement bias is complex, particularly in multi-cohort studies where harmonization of measures to create a single integrated data set is required and may entail simplification of measures (eg, collapsing categories) to find a minimum common ground among cohorts. Specialized methods to handle measurement error such as “regression calibration”⁷⁴ exist, although these usually require strong assumptions and validation samples. As with confounding bias, inclusion of the cohort indicator in a pooled data analysis will address across-cohort measurement biases due to systematic between-cohort differences by closing the additional noncausal path between exposure and outcome via cohort (Figure 3B).

Sensitivity analyses

Sensitivity analyses play an important role in exploring the robustness of findings to assumptions underpinning the chosen analytic approach. This is particularly relevant for multi-cohort studies, where it may be useful to explore the ramifications of data integration by repeating the analysis under different harmonization decisions. Formal quantitative bias analysis approaches^{75,76} that quantify the direction and magnitude of systematic biases may also be valuable. These methods are

increasingly recommended for individual cohort analyses, and we recommend they also be applied to pooled data analyses, with their application guided by DAGs expanded with the cohort indicator (eg, Figures 1B, 2B, 3B).

Conclusion

The target trial framework can be seen as a powerful tool for improving the conduct of causal inference in observational studies through requiring explicit and detailed definition of the target estimand and facilitating systematic assessment of potential sources of bias. From this perspective, we have described the application and utility of this framework in multi-cohort studies, clarifying that the target trial is the central reference point for identifying biases as opposed to comparing studies to each other. Specifically, this approach enables identification of biases within each cohort individually and those that may be introduced when combining data from multiple cohorts. Disentangling biases arising from different sources helps to better harness the risk of bias in analyses and informs the interpretation of findings, particularly discrepant findings across cohorts. As such, use of the target trial framework in multi-cohort studies as described here can help strengthen causal inferences through improved analysis design, transparency in the assumptions, and clarity in reporting and interpretation. We expect continued refinement of this framework as we learn from its application to a wider range of multi-cohort studies in the future.

Acknowledgments

This work was undertaken to inform new cross-cohort developments within the Murdoch Children’s Research Institute (MCRI) LifeCourse program (<https://lifecourse.melbournechildrens.com/>). The article built upon the short course “Observational Studies: Modern Concepts and Analytic Methods,” run by the Clinical Epidemiology and Biostatistics Unit at MCRI.

Supplementary material

Supplementary material is available at *American Journal of Epidemiology* online.

Funding

Dr. O’Connor and Dr. Downes were supported by the Melbourne Children’s LifeCourse platform, funded by Royal Children’s Hospital Foundation grant #2018-984D. Prof. Moreno-Betancur is supported by funding from the Australian National Health and Medical Research Council (NHMRC) Investigator Grant 2009572. Dr. Burgner is supported by funding from the Australian NHMRC Investigator Grant APP1175744. Prof. Goldfeld is supported by Australian NHMRC Practitioner Fellowship 1155290. The Murdoch Children’s Research Institute is supported by the Victorian Government’s Operational Infrastructure Support Program.

Conflict of interest

The authors have no conflicts of interest to disclose.

Disclaimer

This content is solely the responsibility of the authors.

Data availability

All data and results reported are from cited sources.

References

- Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC; 2020.
- 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):758-764. <https://doi.org/10.1093/aje/kwv254>
- Shonkoff JP, Boyce WT, McEwen B. Neuroscience, molecular biology, and the childhood roots of health disparities: building a new framework for health promotion and disease prevention. *JAMA*. 2009;301(21):2252-2259. <https://doi.org/10.1001/jama.2009.754>
- Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384(9945):766-781. [https://doi.org/10.1016/S0140-6736\(14\)60460-8](https://doi.org/10.1016/S0140-6736(14)60460-8)
- Patel V, Flisher AJ, Hetrick S, et al. Mental health of young people: a global public-health challenge. *Lancet*. 2007;369(9569):1302-1313. [https://doi.org/10.1016/S0140-6736\(07\)60368-7](https://doi.org/10.1016/S0140-6736(07)60368-7)
- Doll RJ, Joseph NI, McGarry D, et al. *Epidemiology of Allergic Diseases. Allergy and Asthma: The Basics to Best Practices*. Springer International; 2019:31-51.
- Barouki R, Gluckman PD, Grandjean P, et al. Developmental origins of non-communicable disease: implications for research and public health. *Environ Health*. 2012;11(1):42. <https://doi.org/10.1186/1476-069X-11-42>
- Sun C, Burgner DP, Ponsonby A-L, et al. Effects of early-life environment and epigenetics on cardiovascular disease risk in children: highlighting the role of twin studies. *Pediatr Res*. 2013;73(4, pt 2):523-530. <https://doi.org/10.1038/pr.2013.6>
- Mikkelsen B, Williams J, Rakovac I, et al. Life course approach to prevention and control of non-communicable diseases. *Br Med J*. 2019;364:l257. <https://doi.org/10.1136/bmj.l257>
- World Health Organization. *Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020*. World Health Organization; 2013.
- Proimos J, Klein JD. Noncommunicable diseases in children and adolescents. *Pediatrics*. 2012;130(3):379-381. <https://doi.org/10.1542/peds.2012-1475>
- VanderWeele TJ. Can sophisticated study designs with regression analyses of observational data provide causal inferences? *JAMA Psychiatry*. 2021;78(3):244-246. <https://doi.org/10.1001/jamapsychiatry.2020.2588>
- Goldfeld S, Gray S, Azpitarte F, et al. Driving precision policy responses to child health and developmental inequities. *Health Equity*. 2019;3(1):489-494. <https://doi.org/10.1089/hec.2019.0045>
- Lawlor DA, Andersen A-MN, Batty GD. Birth cohort studies: past, present and future. *Int J Epidemiol*. 2009;38(4):897-902. <https://doi.org/10.1093/ije/dyp240>
- Berkson J. The statistical study of association between smoking and lung cancer. *Proc Staff Meet Mayo Clin*. 1955;30(15):319-348.
- Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol*. 1977;106(3):184-187. <https://doi.org/10.1093/oxfordjournals.aje.a112451>
- Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol*. 1980;112(4):564-569. <https://doi.org/10.1093/oxfordjournals.aje.a113025>
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48. <https://doi.org/10.1097/00001648-199901000-00008>
- VanderWeele TJ, Shpitser I. On the definition of a confounder. *Ann Stat*. 2013;41(1):196-220. <https://doi.org/10.1214/12-aos1058>
- Moreno-Betancur M. The target trial: a powerful device beyond well-defined interventions. *Epidemiology*. 2021;32(2):291-294. <https://doi.org/10.1097/EDE.0000000000001318>
- Hernán MA, Sauer BC, Hernández-Díaz S, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol*. 2016;79:70-75. <https://doi.org/10.1016/j.jclinepi.2016.04.014>
- Didelez V. Commentary: should the analysis of observational data always be preceded by specifying a target experimental trial? *Int J Epidemiol*. 2016;45(6):2049-2051. <https://doi.org/10.1093/ije/dyw032>
- García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol*. 2017;32(6):495-500. <https://doi.org/10.1007/s10654-017-0287-2>
- Labrecque JA, Swanson SA. Target trial emulation: teaching epidemiology and beyond. *Eur J Epidemiol*. 2017;32(6):473-475. <https://doi.org/10.1007/s10654-017-0293-4>
- Hernan MA, Wang W, Leaf DE. Target trial emulation: a framework for causal inference from observational data. *JAMA*. 2022;328(24):2446-2447. <https://doi.org/10.1001/jama.2022.21383>
- Curran PJ, Hussong AM. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol Methods*. 2009;14(2):81-100. <https://doi.org/10.1037/a0015914>
- Freese J, Peterson D. Replication in social science. *Annu Rev Sociol*. 2017;43(1):147-165. <http://dx.doi.org/10.1146/annurev-soc-060116-053450>
- O'Connor M, Spry E, Patton G, et al. Better together advancing life course research through multi-cohort analytic approaches. *Adv Life Course Res*. 2022;53:100499. <https://doi.org/10.1016/j.alcr.2022.100499>
- Christensen G, Freese J, Miguel E. *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press; 2019.
- Berger E, Castagne R, Chadeau-Hyam M, et al. Multi-cohort study identifies social determinants of systemic inflammation over the life course. *Nat Commun*. 2019;10(1):773. <https://doi.org/10.1038/s41467-019-08732-x>
- Hughes RA, Tilling K, Lawlor DA. Combining longitudinal data from different cohorts to examine the life-course trajectory. *Am J Epidemiol*. 2021;190(12):2680-2689. <https://doi.org/10.1093/aje/kwab190>
- Kivimaki M, Batty GD, Pentti J, et al. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. *Lancet Public Health*. 2020;5(3):e140-e149. [https://doi.org/10.1016/S2468-2667\(19\)30248-8](https://doi.org/10.1016/S2468-2667(19)30248-8)
- Petrovic D, Haba-Rubio J, de Mestral VC, et al. The contribution of sleep to social inequalities in cardiovascular disorders: a multi-cohort study. *Cardiovasc Res*. 2020;116(8):1514-1524. <https://doi.org/10.1093/cvr/cvz267>
- Robinson O, Carter AR, Ala-Korpela M, et al. Metabolic profiles of socio-economic position: a multi-cohort analysis. *Int J Epidemiol*. 2021;50(3):768-782. <https://doi.org/10.1093/ije/dyaa188>
- Sallis H, Szekely E, Neumann A, et al. General psychopathology, internalising and externalising in children and functional

- outcomes in late adolescence. *J Child Psychol Psychiatry*. 2019;60(11):1183-1190. <https://doi.org/10.1111/jcpp.13067>
36. Sellers R, Warne N, Pickles A, et al. Cross-cohort change in adolescent outcomes for children with mental health problems. *J Child Psychol Psychiatry*. 2019;60(7):813-821. <https://doi.org/10.1111/jcpp.13029>
 37. Silins E, Horwood LJ, Patton GC, et al. Young adult sequelae of adolescent cannabis use: an integrative analysis. *Lancet Psychiatry*. 2014;1(4):286-293. [https://doi.org/10.1016/S2215-0366\(14\)70307-4](https://doi.org/10.1016/S2215-0366(14)70307-4)
 38. Trude ACB, Richter LM, Behrman JR, et al. Effects of responsive caregiving and learning opportunities during pre-school ages on the association of early adversities and adolescent human capital: an analysis of birth cohorts in two middle-income countries. *Lancet Child Adolesc Health*. 2021;5(1):37-46. [https://doi.org/10.1016/S2352-4642\(20\)30309-6](https://doi.org/10.1016/S2352-4642(20)30309-6)
 39. Zondervan-Zwijnenburg MAJ, Veldkamp SAM, Neumann A, et al. Parental age and offspring childhood mental health: a multi-cohort, population-based investigation. *Child Dev*. 2020;91(3):964-982. <https://doi.org/10.1111/cdev.13267>
 40. O'Connor M, Moreno-Betancur M, Goldfeld S, et al. Data resource profile: Melbourne Children's LifeCourse initiative (LifeCourse). *Int J Epidemiol*. 2022;51(5):e229-e244. <https://doi.org/10.1093/ije/dyac086>
 41. Lesko C, Jacobson L, Althoff K, et al. Collaborative, pooled and harmonized study designs for epidemiologic research: challenges and opportunities. *Int J Epidemiol*. 2018;47(2):654-668. <https://doi.org/10.1093/ije/dyx283>
 42. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med*. 2017;36(5):855-875. <https://doi.org/10.1002/sim.7141>
 43. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221. <https://doi.org/10.1136/bmj.c221>
 44. Morris TP, van Smeden M. Causal analyses of existing databases: the importance of understanding what can be achieved with your data before analysis (commentary on Hernán). *J Clin Epidemiol*. 2022;142:261-263. <https://doi.org/10.1016/j.jclinepi.2021.09.026>
 45. Lu H, Cole SR, Howe CJ, et al. Toward a clearer definition of selection bias when estimating causal effects. *Epidemiology*. 2022;33(5):699-706. <https://doi.org/10.1097/EDE.0000000000001516>
 46. VanderWeele TJ, Robinson WR. On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology*. 2014;25(4):473-484. <https://doi.org/10.1097/EDE.0000000000000105>
 47. Moreno-Betancur M, Koplin JJ, Ponsonby AL, et al. Measuring the impact of differences in risk factor distributions on cross-population differences in disease occurrence: a causal approach. *Int J Epidemiol*. 2018;47(1):217-225. <https://doi.org/10.1093/ije/dyx194>
 48. Bangdiwala SI, Bhargava A, O'Connor DP, et al. Statistical methodologies to pool across multiple intervention studies. *Transl Behav Med*. 2016;6(2):228-235. <https://doi.org/10.1007/s13142-016-0386-8>
 49. Spry E, Moreno-Betancur M, Becker D, et al. Maternal mental health and infant emotional reactivity: a 20-year two-cohort study of preconception and perinatal exposures. *Psychol Med*. 2020;50(5):827-837. <https://doi.org/10.1017/S0033291719000709>
 50. Patton GC, Romaniuk H, Spry E, et al. Prediction of perinatal depression from adolescence and before conception (VIHCS): 20-year prospective cohort study. *Lancet*. 2015;386(9996):875-883. [https://doi.org/10.1016/S0140-6736\(14\)62248-0](https://doi.org/10.1016/S0140-6736(14)62248-0)
 51. Patton G, Coffey C, Romaniuk H, et al. The prognosis of common mental disorders in adolescents: a 14-year prospective cohort study. *Lancet*. 2014;383(9926):1404-1411. [https://doi.org/10.1016/S0140-6736\(13\)62116-9](https://doi.org/10.1016/S0140-6736(13)62116-9)
 52. Olsson CA, Letcher P, Greenwood CJ, et al. The Australian Temperament Project Generation 3 study: a population-based multigenerational prospective cohort study of socioemotional health and development. *BMJ Open*. 2022;12(9):e061854. <https://doi.org/10.1136/bmjopen-2022-061854>
 53. Vassallo S, Sanson A. *The Australian Temperament Project: The First 30 Years*. Australian Institute of Family Studies; 2013.
 54. O'Connor M, Ponsonby A-L, Collier F, et al. Exposure to adversity and inflammatory outcomes in mid and late childhood. *Brain Behav Immun-Health*. 2020;9:100146. <https://doi.org/10.1016/j.bbih.2020.100146>
 55. Young JG, Stensrud MJ. Identified versus interesting causal effects in fertility trials and other settings with competing or truncation events. *Epidemiology*. 2021;32(4):569-572. <https://doi.org/10.1097/EDE.0000000000001357>
 56. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377-399. <https://doi.org/10.1002/sim.4067>
 57. Prattley J, Rioseco P, Vittiglia A, et al. *Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC) LSAC Technical Paper No. 30. Agreement between Parent and Child Responses*. Australian Institute of Family Studies; 2023.
 58. Hammarström A, Westerlund H, Kirves K, et al. Addressing challenges of validity and internal consistency of mental health measures in a 27-year longitudinal cohort study—the Northern Swedish Cohort Study. *BMC Med Res Methodol*. 2016;16:4. <https://doi.org/10.1186/s12874-015-0099-6>
 59. Tourangeau R, Yan T. Sensitive questions in surveys. *Psychol Bull*. 2007;133(5):859-883. <https://doi.org/10.1037/0033-2909.133.5.859>
 60. Downes M, Shepherd D, Wijesuriya R, et al. *Statistical Analysis Plan Template for Observational Studies*. 2023. Accessed February 14, 2024. <https://doi.org/10.26188/12471380>
 61. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley; 1987.
 62. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-295. <https://doi.org/10.1177/0962280210395740>
 63. Li L, Shen C, Li X, et al. On weighting approaches for missing data. *Stat Methods Med Res*. 2013;22(1):14-30. <https://doi.org/10.1177/0962280211403597>
 64. Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivar Behav Res*. 1998;33(4):545-571. https://doi.org/10.1207/s15327906mbr3304_5
 65. van Buuren S. *Flexible Imputation of Missing Data*. Boca Raton, FL: Chapman and Hall/CRC; 2012.
 66. Tilling K, Williamson EJ, Spratt M, et al. Appropriate inclusion of interactions was needed to avoid bias in multiple imputation. *J Clin Epidemiol*. 2016;80:107-115. <https://doi.org/10.1016/j.jclinepi.2016.07.004>
 67. Moreno-Betancur M, Lee KJ, Leacy FP, et al. Canonical causal diagrams to guide the treatment of missing data in epidemiologic studies. *Am J Epidemiol*. 2018;187(12):2705-2715. <https://doi.org/10.1093/aje/kwy173>
 68. Zhang J, Dashti SG, Carlin JB, et al. Recoverability and estimation of causal effects under typical multivariable missingness

- mechanisms. *Biom J*. 2024;66(3):e2200326. <https://doi.org/10.1002/bimj.202200326>
69. Lee KJ, Carlin JB, Simpson JA, et al. Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. *Int J Epidemiol*. 2023;52(4):1268-1275. <https://doi.org/10.1093/ije/dyad008>
70. Dahabreh IJ, Petito LC, Robertson SE, et al. Toward causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a new target population. *Epidemiology*. 2020;31(3):334-344. <https://doi.org/10.1097/EDE.0000000000001177>
71. Dahabreh IJ, Robertson SE, Petito LC, et al. Efficient and robust methods for causally interpretable meta-analysis: transporting inferences from multiple randomized trials to a target population. *Biometrics*. 2023;79(2):1057-1072. <https://doi.org/10.1111/biom.13716>
72. Steingrimsdottir JA, Barker DH, Bie R, et al. Systematically missing data in causally interpretable meta-analysis. *arXiv preprint*. arXiv:2205.00610. 2022. Accessed August 22, 2023. <https://arxiv.org/abs/2205.00610>
73. Diaz I. Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics*. 2020;21(2):353-358. <https://doi.org/10.1093/biostatistics/kxz042>
74. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr*. 1997;65(4):1179S-1186S. <https://doi.org/10.1093/ajcn/65.4.1179S>
75. Lash TL, Fox MP, MacLehose RF, et al. Good practices for quantitative bias analysis. *Int J Epidemiol*. 2014;43(6):1969-1985. <https://doi.org/10.1093/ije/dyu149>
76. Dahabreh IJ, Robins JM, Haneuse SJ-PA, et al. Global sensitivity analysis for studies extending inferences from a randomized trial to a target population. *arXiv preprint*. arXiv:2207.09982. 2022. Accessed August 22, 2023. <https://arxiv.org/abs/2207.09982>