UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Mechanism and Mediation: Counterfactual Strata Effects for Perinatal Epidemiology

Permalink

https://escholarship.org/uc/item/5m34j2p6

Author

Gupta, Shalika

Publication Date

2024

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Mechanism and Mediation: Counterfactual Strata Effects for Perinatal Epidemiology

by

Shalika Gupta

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Epidemiology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Maya L. Petersen, Chair
Associate Professor Laura B. Balzer
Associate Professor Patrick T. Bradshaw
Professor Mark J. van der Laan

Spring 2024

Mechanism and Mediation: Counterfactual Strata Effects for Perinatal Epidemiology

Copyright 2024
by
Shalika Gupta

Abstract

Mechanism and Mediation: Counterfactual Strata Effects for Perinatal Epidemiology

by

Shalika Gupta

Doctor of Philosophy in Epidemiology

University of California, Berkeley

Professor Maya L. Petersen, Chair

In epidemiology, causal effects are often of interest within a post-exposure subgroup and the exposure may affect membership into this subgroup. In the presence of censoring, loss to follow-up, non-compliance in randomized trials, or when exposure affects a precondition necessary for the outcome of interest to occur, subgroup effects may be particularly relevant for formulating causal parameters that answer meaningful questions about the specific mechanisms by which an exposure drives changes in an outcome. In this dissertation, we use the Causal Roadmap to introduce and illustrate the application of counterfactual strata effects in these types of settings. In perinatal studies, where an exposure may affect whether a live birth occurs and infant outcomes are only defined among live births, these types of counterfactual strata effects can be used to understand 1) the effect of an exposure on infant outcomes among live births if everyone were to receive the treatment as compared to the control (i.e., a counterfactual strata total effect) and 2) the effect of an exposure on infant outcomes among live births absent any exposure-induced changes to the composition of people who had live births and the timing of births (i.e., a counterfactual stratum direct effect).

Using the SEARCH Study as a motivating example, we first discuss the challenges of formulating causal effects in perinatal settings and then apply counterfactual strata effects to understand the impact of the SEARCH intervention on perinatal HIV transmission among women living with HIV at study baseline. Lastly, we describe how counterfactual strata effects can be more generally formulated to ask and answer questions when effects are of interest within post-exposure strata in simple settings with a single mediator of subgroup membership, as well as in more complex settings with multiple, possibly longitudinal mediators.

To my parents, Niraj and Saranga Gupta,

and, in loving memory of my mausi, Neena Goyal.

Contents

Contents ii

List of Figures iv

List of Tables v

Introduction 1

1 When Exposure Affects Subgroup Membership: Framing Relevant Causal Questions in Perinatal Epidemiology 5
- 1.1 Introduction . . . 5
- 1.2 Previous Mediation-Based Approaches . . . 7
- 1.3 Counterfactual Stratum Direct Effects in Perinatal Epidemiology . . . 8
- 1.4 Discussion . . . 14

2 Infant HIV-Free Survival and Perinatal HIV Transmission in the SEARCH Study 17
- 2.1 Introduction . . . 17
- 2.2 Study Setting and Context . . . 20
- 2.3 Methods: The Causal Roadmap . . . 21
- 2.4 Results . . . 28
- 2.5 Discussion . . . 33

3 Counterfactual Strata Effects 36
- 3.1 Introduction . . . 36
- 3.2 Application of the Causal Roadmap to Counterfactual Strata Effects in the Point Treatment Setting . . . 38
- 3.3 Extensions to Longitudinal Settings . . . 46
- 3.4 Simulations . . . 53
- 3.5 Discussion . . . 57

Discussion 60

Bibliography **62**

A Chapter 1 **68**
A.1 Structural Causal Model 68
A.2 Summary of Natural Direct Effect Parameters 69

B Chapter 2 **71**
B.1 Causal Estimands 71
B.2 Identification 73
B.3 Estimation 74

List of Figures

1 DAG illustrating treatment-affected confounder of mediator and outcome (L). . 3

1.1 Directed Acyclic Graph (DAG) with pregnancy as a mediator of the effect of an exposure on infant outcomes. For illustration, we focus on the setting of a randomized trial and omitting unmeasured factors. 6

1.2 DAG with pregnancy as a mediator of a randomized exposure on infant outcomes including time-varying confounders at two time points (and omitting unmeasured factors). 6

1.3 (A) A simplified DAG with post-treatment covariates and mediators at one time point in the SEARCH Study. (B) A post-intervention DAG illustrating a controlled direct effect parameter which enforces maternal survival and live birth. (C) A post-intervention DAG illustrating a stochastic direct effect parameter where in the hypothetical intervention, mediators are drawn conditional only on baseline covariates. (D) A post-intervention DAG representing a stochastic direct effect parameter where the hypothetical intervention on the live birth process is drawn conditional on the baseline covariates and time-varying past. 12

2.1 Simplified representation of the mechanisms by which the SEARCH Study may have impacted infant HIV-free survival. 19

3.1 Directed Acyclic Graph (DAG) with baseline covariates (W), randomized treatment (A), post-treatment covariates (L), mediator representing indicator of subgroup membership (Z), and outcome ($Y1$). 40

List of Tables

1.1 Definition and interpretations of causal estimands proposed for the perinatal epidemiologic setting. . . . 15

2.1 Summary of baseline and select post-baseline variables from the SEARCH Study overall and by study arm. . . . 29
2.2 Total effect estimates of infant HIV-free survival using TMLE with and without Super Learner . . . 30
2.3 Total effect estimates of perinatal HIV transmission using TMLE with and without Super Learner . . . 30
2.4 Stochastic direct effect estimated using TMLE under the observed mediator distribution in the intervention arm. . . . 32
2.5 Stochastic direct effect estimated using TMLE under the observed mediator distribution in the intervention arm. . . . 33

3.1 Simulation results for the counterfactual stratum direct effect (risk difference), using a well-supported intervention on the mediators. . . . 56
3.2 Simulation results for the counterfactual stratum direct effect (risk difference), corresponding to a controlled direct effect. . . . 56
3.3 Simulation results for the counterfactual stratum direct effect (risk ratio), using a well-supported intervention on the mediators. . . . 57
3.4 Simulation results for the counterfactual stratum direct effect (risk ratio), corresponding to a controlled direct effect. . . . 57

A.1 Definition and interpretations of alternative target natural direct effect parameters. 70

B.1 Stochastic direct effect estimated using TMLE under the observed mediator distribution in the control arm. . . . 75
B.2 Stochastic direct effect estimated with TMLE under a well-supported user-specified mediator distribution. . . . 76
B.3 Stochastic direct effect estimated using TMLE under the observed mediator distribution in the control arm. . . . 76
B.4 Stochastic direct effect estimated with TMLE under a well-supported user-specified mediator distribution. . . . 77

Acknowledgments

My heart is filled with gratitude for all of the people who have made this possible. First, to my advisor Maya Petersen, thank you for your patience, guidance, and support throughout graduate school. Your dedication to teaching and conducting research with real-world impact has been such an inspiration to me. Thank you for believing that I could write this dissertation even when I thought it wasn't possible. I would also like to thank my dissertation committee members, Laura Balzer, Mark van der Laan, and Patrick Bradshaw, for their encouragement and help in completing this work. Patrick, I feel so lucky to have had the opportunity to take several of your classes at Berkeley. I have learned so much from your teaching.

Many advisors and mentors have supported me in my journey to, and through, graduate school, in particular Chelsea Specht, Patina Mendez, Elizabeth Noth, Katherine Karriker-Jaffe, Maureen Lahiff, Kristine Madsen, and Hannah Thompson. Chelsea, thank you for welcoming me into your lab as an undergraduate at Berkeley; it was only with your encouragement that I considered pursuing a PhD. Maureen, your dedication to your students is unmatched, and I could not have made it through the master's program without your support. Kris and Hannah, thank you for the opportunity to work on the Fit Study and for introducing me to the world of cluster randomized trials. To Lauren Krupa and Jordan Elmowitz, thank you both for all the work you do to support the students in the epidemiology and biostatistics divisions.

To my classmates at the School of Public Health, especially Partow Imani, Erika Brown, Shelley Facente, Rebecca Hyde, Bria Hiebert-Crape, Xing Gao, Alejandra Benitez, and Lina Montoya, I have learned so much from you all. Partow, we made it through coursework, qualifying exams, and now, writing our dissertations. Thank you for your friendship and support throughout this entire process. Erika, you are an incredible epidemiologist and an even more incredible friend. Thank you for the many sweet treats and mid-day coffee breaks. To Rebecca, my first friend at Berkeley Public Health, thank you for understanding my indecisiveness and helping me think through so many decisions. And, to Lina, thank you for so generously sharing your time and knowledge, without which I could not have completed this dissertation.

To all of my family and my friends outside of graduate school, none of this would have been possible without your unwavering support. Extra special thanks to Deidre Patterson, Lauren Kerr, Alyssa Beck, Vanessa Au, SiSi Shan, Asha Vergis, Caitlin Farnsworth, and Jason Brown. To Deidre, thank you for understanding everything always. You, Zara, Caleb, Victor, and Sokka have made Berkeley feel like home. Lauren, I am forever grateful for your friendship and perspective. Alyssa, thank you for your encouragement from the very beginning of this Berkeley adventure.

To my parents, thank you for your unconditional love and support. You have given everything to help your children achieve their dreams. And, to my dadiji, Sudarshan Gupta, no one was more excited than you for me to earn this PhD. We finally did it! Miss you, love you, thank you.

This research was generously supported by the US National Institute of Allergy and Infectious Diseases (NIAID) (awards U01AI099959, UM1AI068636, and R01 AI074345-06A1); the National Heart, Lung, and Blood Institute (NHLBI); and the National Institute of Mental Health (NIMH) and co-funded under award number U01AI150510. It was also supported by the President's Emergency Plan for AIDS Relief and Gilead Sciences. We would like to acknowledge the Ministry of Health of Kenya, Ministry of Health of Uganda, the SEARCH Study collaboration, and all study communities and participants.

Introduction

In this work, we introduce a new class of causal estimands, counterfactual strata effects, which quantify causal effects when impacts are of interest within a subgroup and exposure may affect membership into this subgroup. We first propose these causal estimands in the setting of perinatal epidemiologic studies where causal questions may involve understanding the effect of an exposure, which may affect the probability of a live birth, on infant outcomes, which are only defined if and when a live birth occurs. Total and mediated effect parameters have been proposed for use when live birth mediates the relationship between an exposure and infant outcomes; in Chapter 1, we describe the limitations of current approaches and present, along with their identification results, two alternative causal estimands: 1) a counterfactual strata total effect and 2) a counterfactual stratum direct effect. We use the motivating example of a randomized controlled trial designed to reduce perinatal HIV transmission, the Sustainable East Africa Research in Community Health (SEARCH) Study, to illustrate the relevance of these causal parameters in perinatal settings.

In Chapter 2, we apply the proposed counterfactual strata effects to the SEARCH Study to understand the impact of this community-level intervention on infant HIV-free survival and perinatal HIV transmission over a three year follow-up period. We found that the SEARCH intervention increased the probability of infant HIV-free survival and reduced the risk of perinatal HIV transmission among live births in the population of women living with HIV at study baseline. Moreover, estimates of the counterfactual stratum direct effects of the SEARCH intervention on infant outcomes, using different choices of stochastic interventions on the live birth process, demonstrated that these improvements in outcomes persisted after accounting for intervention-induced changes to the types of people who had live births and the timing of live births over the follow-up period. In other words, if we were to keep the composition of people who had live births and the timing of their births the same, the SEARCH intervention resulted in a higher probability of infant HIV-free survival and lower probability of HIV transmission among live births than the control condition.

In Chapter 3, we generalize the formulation of counterfactual strata effects presented earlier and discuss the relevance of these parameters for asking and answering questions where causal effects are of interest within post-exposure subgroups more broadly, including survival settings, randomized trials where compliance and censoring are concerns, or when exposure affects a precondition necessary for an outcome to be observed or defined. We also demonstrate the performance of a longitudinal maximum likelihood estimator of the

counterfactual stratum direct effect in simulations. We conclude with a discussion of possible extensions of this work. Laura B. Balzer, Moses R. Kamya, and Diane V. Havlir were co-authors on Chapter 1; Maya L. Petersen was a co-author on Chapters 1, 2, and 3. The SEARCH collaboration, including Theodore Ruel, James Ayieko, Laura B. Balzer, Moses R. Kamya, Diane V. Havlir contributed to Chapter 2. The direct effect parameter which is proposed in this work is based on mediated effect estimands formulated using randomized interventions on mediating variables. Below we provide a brief review of mediated effects, with a particular focus on mediated effects formulated using stochastic interventions on mediating variables.

Longitudinal Mediation

The Baron and Kenny approach to mediation, in which a series of three regression models is used to evaluate the presence of mediation, is the most common method for mediation analyses;[1] however, limitations of this approach have been described extensively in the mediation literature, such as dependence on linear models and the inability to detect mediation when an indirect and direct effect exist but are of opposite sign.[2,3,4,5] In addition, the approach described by Baron and Kenny applies only to mediation questions in the point treatment setting and does not take into account the presence of time-varying confounding among the exposure, mediating, and outcome variables.[1] Although researchers have since developed and extended this approach to longitudinal[6] and hierarchical[7] settings, many of these methods are based in the same tradition of linear structural equation modeling as the Baron and Kenny approach, and as a result, share some of the same limitations.[8]

Robins and Greenland[9] defined causal mediation more generally within a counterfactual framework using potential outcomes. Pearl[10] extended these ideas, using a counterfactual framework to define causal parameters, controlled direct effects and natural (in)direct effects, in the point treatment setting and to describe the assumptions necessary for identifiability of these parameters in the general case of a non-parametric model. A controlled direct effect captures the effect of an exposure A on an outcome Y that is not mediated by an intermediate variable Z; it corresponds to a hypothetical intervention to change A from some value a to a' while intervening to set the value of a mediator Z to some value z.[9,10] Alternatively, a natural direct effect parameter describes the effect of changing A from a value of a to a' on Y while holding the value of a mediator Z at the value it would take in the absence of an intervention, and a natural indirect effect describes the effect of setting A to some value a and changing the value of the mediator from the value it would take in the presence of the exposure to the value it would take in the absence of the exposure for each individual.[9,10] When compared to linear structural modeling approaches, this counterfactual framework is compatible with a wider range of statistical models, including non-parametric and semi-parametric estimators.[8]

In the point treatment setting, identification of a total effect requires independence of the counterfactual outcome $Y(a)$ and the exposure A given baseline covariates W. Identi-

fication of controlled direct effects additionally requires independence of the counterfactual outcome $Y(a, z)$ and the mediator of interest Z given past exposure and baseline covariates. Identification of natural direct and indirect effects requires assumptions beyond those required for the identification of a total effect or controlled direct effects.[11] More specifically, identification of natural effects relies on the independence of the counterfactual mediator $Z(a)$ and the exposure A given the baseline covariates W, as well as on the conditional independence of the counterfactual outcome $Y(a, z)$ and the counterfactual mediator $Z(a')$ given baseline covariates.[11] The latter assumption, also known as the cross-world assumption, is unique when compared to the other assumptions required for identifiability because it involves the independence of two counterfactual quantities under two different exposures of interest, $A = a$ and $A = a'$.[12]

In the NPSEM framework, if the counterfactual outcome $Y(a, z)$ is independent of the mediator given past exposure and covariates, then the cross-world assumption will be satisfied as long as there exists no variable that is affected by the exposure A and also confounds the mediator-outcome relationship (Figure 1).[11,13] In the longitudinal setting, the existence of variables that are both affected by exposure and confounders of the mediator-outcome relationship are commonplace; in the extensions of the counterfactual framework for mediation to settings with post-treatment mediator-outcome confounders, the target causal parameters of the natural direct and indirect effects are expressed using stochastic interventions on the mediator(s).[14,15,16,17] Didelez et al.[18] proposed formulations of direct and indirect effects using random interventions (possibly conditional on covariates); in the random interventional approach, an intervention on the mediator involves randomly drawing its value from a distribution rather than deterministically setting the mediator equal to the value it would have taken for a particular individual under a specific intervention on the exposure variable.[18] van der Laan and Petersen[19] proposed an alternative definition of a direct effect parameter that was generalized to any choice of conditional distribution that can be used to average over the z-specific individual controlled direct effects and provided identifiability results and approaches for estimation of this type of parameter.

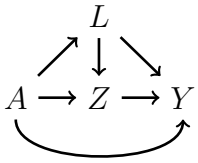


Figure 1: DAG illustrating treatment-affected confounder of mediator and outcome (L).

While VanderWeele and Tchetgen Tchetgen[15] and Zheng and van der Laan[14] explicitly consider stochastic interventions on longitudinal mediators that allow for multiple time-varying covariates and mediators over time, Rudolph et al.[16] and Diaz et al.[17] discuss the formulation and estimation of target parameters in the setting with mediators and post-treatment mediator-outcome confounders measured at a single time point. In this work,

we use this type of stochastic interventional approach to formulate a counterfactual stratum direct effect that can be used in settings with one or multiple possibly time-varying mediators, to define causal estimands with relevance to causal questions in epidemiology broadly when exposure effects are of interest within post-exposure subgroups.

Chapter 1

When Exposure Affects Subgroup Membership: Framing Relevant Causal Questions in Perinatal Epidemiology

1.1 Introduction

Perinatal epidemiology often evaluates the effects of exposures on infant outcomes when the exposure precedes or occurs during pregnancy and may affect the probability of live birth (Figure 1.1). This poses a profound challenge to specifying relevant and interpretable causal effects of exposures on infant outcomes. Among individuals at risk of pregnancy, infant outcomes are only defined once a live birth occurs. Importantly, the timing of birth and the risk of adverse infant outcomes due to baseline characteristics of the individuals who give birth may differ by exposure status, even when the exposure is randomized. This fundamental challenge, although particularly acute in perinatal epidemiology, occurs whenever causal effects are of interest within a subset of the target population, and membership in that subset is affected by the exposure.[20]

In perinatal epidemiology, Chiu et al.[21] proposed defining total effects based on a composite endpoint of live birth and infant outcomes. This definition implicitly assumes that all individuals in the population are pregnancy seeking (or all are not pregnancy seeking), which may be applicable in some settings, but not all.[21] It further includes any exposure effects on the live birth process, which may or may not be of interest. A range of mediation-based causal estimands have been proposed in perinatal epidemiology to disentangle these effects, including controlled direct, separable, and principal stratum direct effects.[21,22,23] While useful in some settings, each has limitations.[22,24]

In this chapter, we first discuss an alternative definition of total effects in perinatal studies, using a composite outcome as in Chiu et al.,[21] but instead focusing on a counterfactual

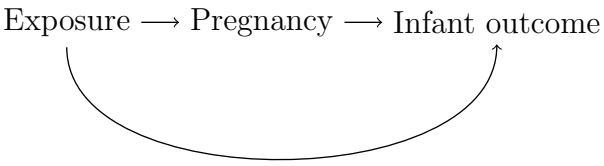


Figure 1.1: Directed Acyclic Graph (DAG) with pregnancy as a mediator of the effect of an exposure on infant outcomes. For illustration, we focus on the setting of a randomized trial and omitting unmeasured factors.

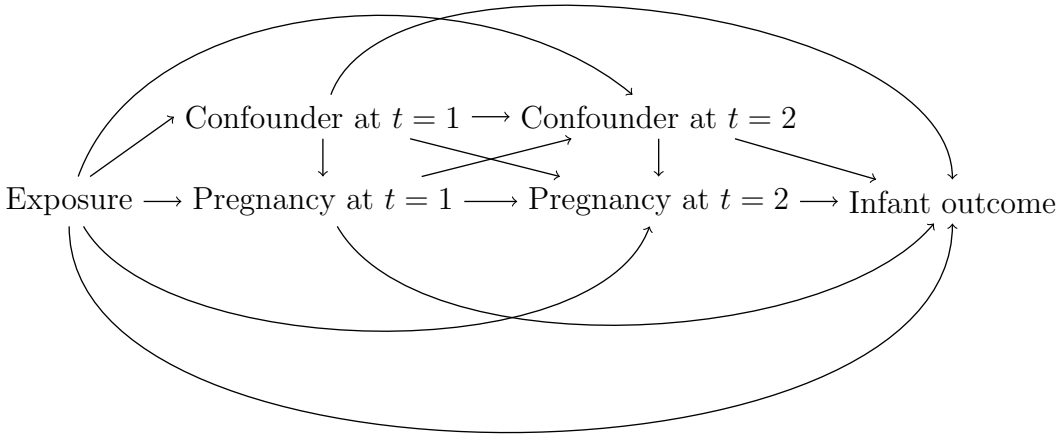


Figure 1.2: DAG with pregnancy as a mediator of a randomized exposure on infant outcomes including time-varying confounders at two time points (and omitting unmeasured factors).

strata total effect (CSTE) among counterfactual live births. The proposed counterfactual strata total effect includes by definition any impact of an intervention that operates by changing the types of people who give birth and timing of live births. Such differences are not a failure of causal identification; instead, these differences are a result of the paths by which an exposure can affect infant outcomes in settings where group membership may be affected by treatment.

Next, in order to isolate the exposure effect on infant outcomes not mediated by the birth process, we present an alternative class of conditional mediation causal parameters that are identifiable, interpretable, and which answer distinct questions in perinatal epidemiology from those previously proposed. Following the causal roadmap,[25] we consider a longitudinal data structure with time-varying covariates and a time-varying mediator (live birth process) (Figure 1.2). We present counterfactual stratum direct effect (CSDE) parameters formulated

using stochastic interventions on the live birth process, which allows us to isolate the exposure effect on infant outcomes *not* due to any effects on the types of people who give birth and timing of live births (i.e., within the same counterfactual population at-risk), and discuss identification assumptions for these parameters. The SEARCH Study provides a detailed illustration. Our approach is applicable to a wide range of perinatal epidemiologic research, as well as more generally to the broad class of studies in which the causal effect of interest is defined among a subset of the population, and membership in this subset may be affected by the exposure.

The SEARCH Study

The SEARCH Study was a cluster randomized trial that evaluated the effectiveness of universal antiretroviral treatment (ART) and person-centered HIV care on population health in rural Kenya and Uganda.[26,27] The study tested 90% of adults for HIV at baseline[28] and ascertained vital status and HIV status after three years for infants born during the study. In pre-specified analyses, the SEARCH intervention reduced perinatal HIV transmission and increased infant HIV-free survival among women of reproductive age living with HIV.[29] However, the extent to which this total effect may have operated through intervention-induced changes to the counterfactual population of infants at-risk, due to timing of births (in the context of time-varying transmission risk across the pre and postnatal risk periods) and/or the types of people who gave birth (e.g., through changes to maternal survival, fertility, or behavior) remains unclear. CSDEs provide an approach to answering this question.

1.2 Previous Mediation-Based Approaches

A number of mediation-based perinatal epidemiology parameters have been previously proposed. In particular, prior authors[21,22] have sought to exclude from the total effect of the exposure on infant outcomes any impact of the exposure on the live birth process (the mediator) through: 1) hypothetical static interventions on the live birth process (i.e., controlled direct effects); 2) restriction to the group of individuals who would have had a live birth irrespective of exposure (i.e., principal stratum direct effects); and, 3) partitioning of the exposure into distinct components that operate in different ways on the live birth process and infant outcomes (i.e., separable direct effects).

Each of these approaches has limitations. First, a controlled direct effect evaluates the exposure effect on infant outcomes if all individuals in the target population were to give birth. This type of hypothetical intervention may be both hard to interpret and poorly supported by the data, particularly if the probability of live birth is low and/or many are not seeking pregnancy. A controlled direct effect quantifies the impact of the SEARCH intervention on infant HIV-free survival if all reproductive age women living with HIV at study baseline were to have a birth during the follow-up period, a counterfactual scenario that is neither realistic nor informative (Figure 1.3B).

Second, identification of principal stratum direct effects requires strong assumptions, such as an assumption that no 'defiers' exist. In a perinatal setting with live birth as a mediator, this corresponds to assuming that there are no individuals who would have a live birth if unexposed, but not if exposed. Moreover, if exposure affects the timing of births within the principal stratum, this direct effect parameter would still include any impact that operates through time at risk. For example, in SEARCH this would include any effect due to changes in timing of birth during follow-up and corresponding shifts in the distribution of infant exposure to pre versus and postnatal transmission risk.

Third, separable direct effect parameters are predicated on the decomposition of the exposure into two components, one that acts directly on the outcome without affecting the mediator (live birth process) and a second that operates only on the (infant) outcome through the mediator. Often, this decomposition, which requires specifying mechanisms by which exposure affects infant survival but not maternal survival, may not be feasible. In the SEARCH trial, where the outcome of interest involves infant HIV-free survival and the mediator includes maternal survival, we cannot disentangle the portion of the intervention that impacts only the outcome without affecting the mediator.

To answer many questions in perinatal epidemiology, alternative approaches are thus needed that account for the live birth process while addressing interpretable causal questions that are identified under plausible assumptions.

1.3 Counterfactual Stratum Direct Effects in Perinatal Epidemiology

We propose an alternative approach to defining causal parameters in perinatal epidemiology, using direct effects based on stochastic interventions on the mediating variables. Stochastic interventions consider a hypothetical intervention on the mediators where their values are randomly drawn from a distribution (possibly conditional on covariates) rather than deterministically set to a particular value.

Direct effect parameters formulated using stochastic interventions on mediating variables[14,15,18,19,30,31] have several advantages, including that they define more realistic causal parameters (by allowing for variation in mediator values), reduce positivity concerns, naturally generalize to longitudinal settings, and remain identifiable in the common setting of post-treatment confounding of the mediator-outcome relationship. In perinatal settings, stochastic direct effects are particularly appealing because they can be formulated to exclude from the total effect any intervention impact that affects infant risk that is mediated by the live birth process. In the remainder of the chapter, we discuss the use of counterfactual stratum direct effects to frame and identify causal questions in perinatal epidemiology, using the SEARCH Study as a running example.

Structural Causal Model

In many perinatal applications, the observed data consist of baseline covariates, an indicator of exposure, time-varying covariates and mediators, and infant outcome (e.g., HIV-free survival in SEARCH). A longitudinal structural causal model (SCM) can be used to formalize knowledge about the data-generating process.[32] For simplicity, we focus on an individual-level SCM (noting that this framework generalizes to hierarchical settings[33,34]). We use $\bar{V}_t$ to denote $(V_1, \ldots, V_t)$, which represents the values of the variable V from time 1 through time t. L_0 represents baseline covariates (e.g., maternal age, prior ART use, and community HIV prevalence). A denotes exposure (e.g., if the individual was randomized to receive the intervention). L_t is the set of time-varying covariates (e.g., maternal ART status and infant vital status). Z_t represents the mediator(s) of interest, which in SEARCH consist of $Z1_t$ (maternal mortality by time t) and $Z2_t$ (live birth status at time t). Let $Y_{\tau+1}$ denote the final outcome. In SEARCH, $Y_{\tau+1}$ denotes an indicator of an infant being born and surviving HIV-free until the end of follow-up, a composite of $Z2_\tau$ (an indicator of live birth by time τ, $Y1_{\tau+1}$ (infant survival), $Y2_{\tau+1}$ (infant HIV-infection status) by the end of follow-up. The exogenous nodes, representing unmeasured variables, are: $U = (U_{L_0}, U_A, U_{L_t}, U_{Z1_t}, U_{Z2_t}, U_{Y1_{\tau+1}}, U_{Y2_{\tau+1}})$ for $t = 1, \ldots, \tau$. The endogenous nodes are: $X = (L_0, A, \bar{L}_t, \bar{Z}_t, Y_\tau)$ for $t = 1, \ldots, \tau$. In randomized trials such as SEARCH, the exposure assignment mechanism is known. Less is known about the time-varying covariates, mediator and outcome processes, which, in particular, may have been affected by exposure and in turn may have affected subsequent covariates, mediators, and outcomes (Appendix A.1). Figure 1.3A illustrates (using fewer timepoints) the mechanisms of interest by which the SEARCH intervention may have impacted infant HIV-free survival, which include the intervention's effects through changes to maternal ART initiation, maternal mortality, and live birth.

Causal Estimands

The relevance of counterfactual stratum direct effect parameters for perinatal epidemiology, as well as the applicability of the CSDE approach for a much broader class of counterfactual strata causal effects, is clarified by first revisiting the definition of total effects in these settings.

Counterfactual Strata Total Effects

Epidemiology commonly studies total effects, defined using marginal summaries of the distribution of counterfactual outcomes. However, in many settings, including perinatal epidemiology, conditional effects defined using summaries of the distribution of counterfactual outcomes *among a counterfactual subset of the population* are often of greater interest. To formulate such conditional total effect parameters, as in Chiu et al.,[21] we use a composite outcome of live birth and infant outcomes. However, in contrast to Chiu et al.,[21] who define

the total effect as the effect of the exposure on the probability of the composite event of live birth and infant outcomes, we instead define a counterfactual strata total effect (Table 1.1):

$$\begin{aligned} &P[Y1_{\tau+1}(a=1)=1, Y2_{\tau+1}(a=1)=1 | Z2_{\tau}(a=1)=1] \\ &\qquad - P[Y1_{\tau+1}(a=0)=1, Y2_{\tau+1}(a=0)=1 | Z2_{\tau}(a=0)=1] \\ &= \frac{P[Y_{\tau+1}(a=1)=1]}{P[Z2_{\tau}(a=1)=1]} - \frac{P[Y_{\tau+1}(a=0)=1]}{P[Z2_{\tau}(a=0)=1]} \end{aligned}$$

In the SEARCH Study, this counterfactual strata total effect evaluates the effect of the exposure on the outcome of live birth and infant HIV-free survival ($Y1_{\tau+1}(a)=1, Y2_{\tau+1}(a)=1$), but only among counterfactual live births ($Z2_{\tau}(a)=1$). By the definition of conditional probabilities, the conditional causal effect can alternatively be expressed by contrasting the joint probability of the composite counterfactual outcome of live birth and infant outcomes divided by the probability of counterfactual live birth for each exposure of interest.

Conditioning on the probability of counterfactual live birth will exclude from the traditional total effect parameter any effect of the intervention on infant outcomes that operates through changes to the marginal probability of live birth by the end of follow-up; however, this counterfactual strata total effect parameter does not ensure equivalence of other aspects of the live birth process under differing exposures, and thus will still include any effects due to exposure-induced changes to the types of people who have a live birth and the timing of births (i.e., composition of the counterfactual population at-risk). For example, in SEARCH, the counterfactual strata total effect of the intervention on infant HIV-free survival would include both any direct reduction in vertical transmission (e.g., by improving HIV viral suppression) and any increase in transmission due to facilitating live birth among persons at higher risk of transmission (e.g., resulting from barriers to care) or changing the time of birth (e.g., such that a higher proportion of follow-up time is prenatal). Such counterfactual strata total effects are often of real-world interest in that they reflect the total change in infant outcomes due to multiple mechanisms expected were an intervention to be implemented. Nonetheless, in many cases it remains of interest to isolate intervention effects on outcomes among the same types of people with the same timing of births.

Counterfactual Stratum Direct Effect

Counterfactual stratum direct effect parameters allow us to quantify the effect of an exposure on infant outcome absent any effect of the intervention on the live birth process, or in other words, the exposure effect in populations with the same types of people (Table 1.1). We build on our definition of a counterfactual strata total effect to define such a counterfactual stratum direct effect. In contrast to a controlled direct effect parameter, which would compare hypothetical interventions in which exposure is varied but all persons in the target population survive to give birth (Figure 1.3B), stochastic direct effect parameters,[18,19,30] compare hypothetical interventions in which exposure is varied but mediator

values Z_τ are drawn from a distribution that is held constant, denoted here as $\bar{G}(L_0)$. Let $\bar{G}(\bar{Z}_\tau|L_0) \equiv \prod_{t=1}^{\tau} G_t(Z_t|\bar{Z}_{t-1}, L_0)$, refer to a user-specified conditional distribution of the mediators, $\bar{Z}_\tau$ given baseline covariates L_0, where $G_t(Z_t|\bar{Z}_{t-1}, L_0)$ represents the probability distribution of the mediator at time t given the mediator past and baseline covariates. For ease of notation, we will sometimes use $\bar{G}$ to refer to this conditional distribution. Given some choice of mediator distribution $\bar{G}(L_0)$, we use this approach to define the following counterfactual stratum direct effect parameter:

$$\begin{aligned}
&P[Y1_{\tau+1}(a=1,\bar{G})=1, Y2_{\tau+1}(a=1,\bar{G})=1|Z2_\tau(\bar{G})=1] \\
&\qquad - P[Y1_{\tau+1}(a=0,\bar{G})=1, Y2_{\tau+1}(a=0,\bar{G})=1|Z2_\tau(\bar{G})=1] \\
&= \frac{P[Y_{\tau+1}(a=1,\bar{G})=1]}{P(Z2_\tau(\bar{G})=1)} - \frac{P[Y_{\tau+1}(a=0,\bar{G})=1]}{P(Z2_\tau(\bar{G})=1)}
\end{aligned}$$

As with the CSTE, this CSDE is defined among infants born to persons in the target population under each exposure level of interest; however, in contrast to the counterfactual strata total effect, the hypothetical intervention to draw maternal survival and live birth from some fixed distribution $\bar{G}$ ensures that the exposure effect is evaluated among the same counterfactual population. For example, in SEARCH, the CSDE corresponds to the difference in counterfactual probability of HIV-free survival among infants born to persons with HIV under intervention or control, had maternal mortality and live birth over time, $\overline{Z1}_\tau$ and $\overline{Z2}_\tau$, respectively, been drawn from the same underlying distribution.

In contrast to the principal stratum direct effect, which evaluates the impact of the SEARCH intervention in the group who would give birth regardless of treatment assignment, the stochastic direct effect parameter can evaluate the direct effect of an exposure among different groups of people, depending on the choice of the distribution $\bar{G}$ from which mediator values are drawn. A variety of CSDEs can be formulated using different choices of stochastic interventions on the mediator variables to help understand how the intervention effect varies by the timing of births and composition of people who give birth.

Choice of Mediator Distribution: Interpretation and Identification

In specifying a distribution $\bar{G}$ from which the mediator value(s) are drawn, there are several options, with implications for both parameter identification (or ability to rewrite the causal parameter as a parameter of the observed data distribution and thus estimate it) and interpretation. One approach[18,19,35] is to define the (potentially longitudinal) mediator distribution as the unknown counterfactual distribution in the absence of exposure, possibly conditional on baseline covariates (also referred to as a marginal mediator or fixed stochastic direct effect in the context of non-conditional effect[15,16] (Appendix A.2)). In perinatal settings, such a natural direct effect formulated conditional on baseline covariates holds constant the types and timing of births (Figure 1.3C) across exposure conditions; however, in the presence of time-varying confounding, it will not decompose the original total

Figure 1.3: (A) A simplified DAG with post-treatment covariates and mediators at one time point in the SEARCH Study. (B) A post-intervention DAG illustrating a controlled direct effect parameter which enforces maternal survival and live birth. (C) A post-intervention DAG illustrating a stochastic direct effect parameter where in the hypothetical intervention, mediators are drawn conditional only on baseline covariates. (D) A post-intervention DAG representing a stochastic direct effect parameter where the hypothetical intervention on the live birth process is drawn conditional on the baseline covariates and time-varying past.

effect.[14,15] Identification of this parameter requires no unmeasured confounding of the 1) exposure-outcome, 2) exposure-mediator, and 3) mediator-outcome relationships at all time points (but avoids a stronger "cross-world" assumption required by earlier definitions[9,10]). Identification requires a positive probability of 1) the values the mediator takes under the control condition given each exposure of interest within all values of the mediator-outcome confounders (baseline and time-varying covariates) at each time t and 2) of each possible exposure within all values of baseline covariates.

Another choice for the stochastic intervention on the mediator(s) is to draw mediator values from the counterfactual control distribution conditional both on baseline covariates and time-varying history, as proposed by Zheng and van der Laan.[14] While this conditional mediator formulation of the natural direct effect is appealing because it provides a decomposition of the average treatment effect,[14] in perinatal settings, this type of parameter does not ensure that the timing of births and types of people who give birth under different exposures will be the same (i.e., does not ensure equivalent counterfactual populations at-risk across differing exposures). In perinatal settings (Figure 1.3D), this corresponds to a hypothetical intervention where live birth is a function of both the baseline covariates and the time-varying past (e.g., maternal time-to-ART start). Because the time-varying past is affected by the static intervention on A, the types of people who have a live birth under an intervention to assign everyone to receive the intervention may not be the same as those who would have a live birth if everyone were to be assigned to the control condition. This parameter also requires stronger assumptions for identifiability than the natural direct effect using marginal mediator distributions (i.e., strong sequential randomization and its associated positivity assumptions).[14]

Alternatively, counterfactual stratum direct effect parameters can be formulated such that mediator values are drawn from a distribution, conditional only on baseline covariates (Figure 1.3C), or from a corresponding distribution that is estimated from the observed data (i.e., a data-adaptive parameter as proposed by Rudolph et al.[16,36]). Data-dependent choices for the stochastic intervention on the mediator include the estimated mediator distribution in the intervention or comparator conditions or an intervention that is well-supported in the study sample. In SEARCH, for example, we could formulate a data-adaptive counterfactual stratum direct effect parameter which compared infant HIV-free survival among live births if everyone received the intervention versus the control condition and if maternal mortality and time-to-live birth were drawn from their observed distribution in the intervention arm conditional on baseline covariates, such as maternal age, community-specific HIV prevalence, and baseline ART initiation status.

Such stochastic direct effect parameters, in which the distribution from which the mediators is drawn is known or estimated, require fewer assumptions for identifiability than mediated effects formulated using either the marginal or conditional mediator approaches described previously. More specifically, identification requires no unmeasured confounding of the 1) exposure-outcome and 2) mediator-outcome relationships at all time points. In addition, identification requires positive probabilities of 1) each exposure of interest given baseline covariates and 2) of the values the mediator can take under its specified known

distribution given each exposure within all past values of mediator-outcome confounders (baseline and time-varying factors) at each timepoint.

The assumption of no unmeasured confounding of the exposure and outcome will be satisfied by design in a randomized controlled trial such as the SEARCH Study. Randomization of the exposure variable is not sufficient to satisfy the second assumption of no unmeasured confounding of the mediator-outcome relationship(s). In the SEARCH trial, the second assumption will be satisfied if we believe that all confounders of the live birth process (maternal mortality and live birth at time t) and outcome (infant HIV-free survival) at each time point were measured. In SEARCH, a potential set of such variables includes maternal ART initiation through time t, mediator history (maternal mortality and live birth through time $t-1$), and baseline covariates. Measurement of additional time-varying covariates, such as maternal health (e.g., CD4 cell count and viral load) over time, makes it more likely that this assumption will be satisfied. In general, measuring a rich set of baseline and time-varying covariates will improve the plausibility of both of these assumptions on confounding.

The counterfactual strata total effect will be identified if there is 1) no unmeasured exposure-outcome confounding and 2) a positive probability of each exposure within all baseline covariate values. Because the outcome of interest is defined as the composite event of the mediator live birth and infant outcomes, in SEARCH, this randomization assumption more specifically corresponds to assumptions of no unmeasured confounding of the exposure-live birth ($A-Z2_\tau$), exposure-infant HIV status ($A-Y1_\tau$), and exposure-infant vital status ($A-Y2_\tau$) relationships. Unlike the CSTE, the CSDE does not require an assumption of no unmeasured confounding of the exposure-live birth relationship because the mediator, the live birth process ($\bar{Z}_\tau$), is intervened upon in the formulation of the CSDE parameter.[37] As a result, in observational settings, it is possible that the CSDE may be identified even when the CSTE is not, due to unmeasured exposure-mediator confounding. (In the setting of a randomized trial, randomization assumptions involving the exposure will be satisfied by design.)

1.4 Discussion

A common challenge of causal effect estimation in perinatal epidemiology is that outcomes are only measured among infants born alive, but exposures typically predate birth and may affect the live birth process, resulting in different types of people who have live births. While particularly notable in the perinatal epidemiology literature, this challenge also occurs ubiquitously across settings in where an exposure may affect the composition of the group that is at risk for the study outcome.[38] We add to existing literature addressing this challenge by first explicitly defining and interpreting counterfactual strata total effects in this setting, highlighting two components of the total effect, one due to due to exposure effects among counterfactual group members, and a second due to exposure effects on counterfactual group membership.

Table 1.1: Definition and interpretations of causal estimands proposed for the perinatal epidemiologic setting.

Parameter	Interpretation (General)	Interpretation (SEARCH)	Counterfactuals (SEARCH)	Hypothetical Intervention
Total effect parameter $\frac{P[Y_{\tau+1}(a=1)=1]}{P[Z2_\tau(a=1)=1]} - \frac{P[Y_{\tau+1}(a=0)=1]}{P[Z2_\tau(a=0)=1]}$ $= P[Y1_{\tau+1}(a=1)=0, Y2_{\tau+1}(a=1)=0 \mid Z2_\tau(a=1)=1]$ $- P[Y1_{\tau+1}(a=0)=0, Y2_{\tau+1}(a=0)=0 \mid Z2_\tau(a=0)=1]$	• The total effect of an intervention on infant health outcomes among live births • Includes: any effect of the intervention operating through changes to the timing and types of people who give birth • Excludes: any effect of the intervention on infant health outcomes through changes to the marginal probability of live birth by the end of the follow-up period	• The total effect of the SEARCH intervention on infant HIV-free survival among live births	• $Y_{\tau+1}(a)$: counterfactual outcome of live birth and infant HIV-free survival at time $\tau+1$ when $A=a$ • $Y1_{\tau+1}(a)$: counterfactual infant vital status at time $\tau+1$ when $A=a$ • $Y2_{\tau+1}(a)$: counterfactual infant HIV status at time $\tau+1$ when $A=a$ • $Z2_\tau(a)$: counterfactual live birth status by time τ when $A=a$	• Compare intervening to set $a=1$ (i.e., receive the SEARCH intervention) for everyone to setting $a=0$ (i.e., receive country standard of care) for everyone
Stochastic direct effect parameters $\frac{P[Y_{\tau+1}(a=1,\bar{G})=1]}{P(Z2_\tau(\bar{G})=1)} - \frac{P[Y_{\tau+1}(a=0,\bar{G})=1]}{P(Z2_\tau(\bar{G})=1)}$ $= P[Y1_{\tau+1}(a=1,\bar{G})=0, Y2_{\tau+1}(a=1,\bar{G})=0 \mid Z2_\tau(\bar{G})=1]$ $- P[Y1_{\tau+1}(a=0,\bar{G})=0, Y2_{\tau+1}(a=0,\bar{G})=0 \mid Z2_\tau(\bar{G})=1]$ where $\bar{G} \equiv \bar{G}_\tau$ and $\bar{G}_t(\bar{z}_t \mid L_0) \equiv P(\bar{Z}_t = \bar{z}_t \mid L_0)$ is a known distribution (conditional on baseline covariates)[a]	• A direct effect of an intervention on infant health outcomes where the live birth process (a person's probability of giving birth at each time point) is drawn from the same known distribution conditional on baseline covariates • Includes: any effect of the intervention on the outcome that operates through mechanisms other than the live birth process • Excludes: any effect of the intervention on infant health outcomes through changes to the mediator: timing/occurrence of live births and to the types of people who give birth	• The direct effect of the SEARCH intervention on infant HIV-free survival if the live birth process was drawn from some known distribution conditional on the baseline covariates of a woman in the target population	• $Y_{\tau+1}(a,\bar{G})$: counterfactual outcome of live birth and infant HIV-free survival at $t=\tau+1$ under an intervention to set $A=a$ and draw maternal survival and time-to-live birth from a known distribution that is conditional on observed baseline covariates • $Y1_{\tau+1}(a',\bar{G})$: counterfactual infant vital status at $t=\tau+1$ under an intervention to set $A=a$ and draw maternal survival and time-to-live birth from the known mediator distribution • $Y2_{\tau+1}(a',\bar{G})$: counterfactual infant HIV status at $t=\tau+1$ under an intervention to set $A=a$ and draw maternal survival and time-to-live birth from the known mediator distribution • $Z2_\tau(\bar{G})$: counterfactual live birth status at $t=\tau$ under the known stochastic intervention on the mediators	• Compare intervening to set $a=1$ to setting $a=0$ for all persons in the target population, while drawing the live birth process from the same known distribution conditional on baseline covariates

[a] In the specific case where the known distribution of the mediators conditional on baseline covariates is equivalent to the true marginal mediator distribution conditional on baseline covariates in the control arm, this stochastic direct effect parameter will be equal to the natural direct effect parameter formulated using marginal mediator distributions (and will provide a decomposition of a type of total effect).

We then suggest the use of a class of counterfactual stratum direct effects to disentangle these two exposure pathways and discuss setting-specific advantages over other types of direct effect parameters described in the perinatal literature. Such direct effect parameters are formulated using stochastic interventions on the mediator process conditional on baseline covariates. The choice of stochastic intervention on the mediator(s) requires careful consideration as it determines the subgroup in which the intervention is assessed. In perinatal settings, the choice of stochastic intervention will determine the timing of births and composition of people who have live births, and will thus affect the interpretation of the stochastic direct effect parameter. This flexibility also represents an advantage of this approach; applying different choices of stochastic interventions on the mediating variables can reveal how the direct effect of the intervention is mediated by the live birth process that occurs in perinatal studies, as well as providing a straightforward approach to transporting (under assumptions) perinatal exposure effect estimates to new settings with differing live birth processes, or more generally, of transporting stochastic direct effects among selected subgroups to new settings in which the group selection process my differ.

Approaches for the estimation of stochastic direct effect parameters have been developed by van der Laan and Petersen,[19] Rudolph et al.,[16] Zheng and van der Laan,[14] VanderWeele and Tchetgen Tchetgen,[15] and Diaz et al.,[17] among others, and include a range of inverse probability weighted and semiparametric efficient multiply robust estimators. However, to date, while stochastic mediation effects have been applied in perinatal studies previously,[31,39] to the best of our knowledge, they have not been used to formulate counterfactual stratum direct effect parameters that account for exposure-induced changes to the live birth process in these types of settings, nor more generally to account for exposure-induced changes in subgroup membership when conditional causal effects are of interest. This chapter aims to provide a bridge to the wider use of these methods in settings where they are perhaps uniquely positioned to address the underlying question of interest. In the next chapter, we will apply these ideas to our motivating example to understand perinatal HIV transmission in the SEARCH Study.

Chapter 2

Infant HIV-Free Survival and Perinatal HIV Transmission in the SEARCH Study

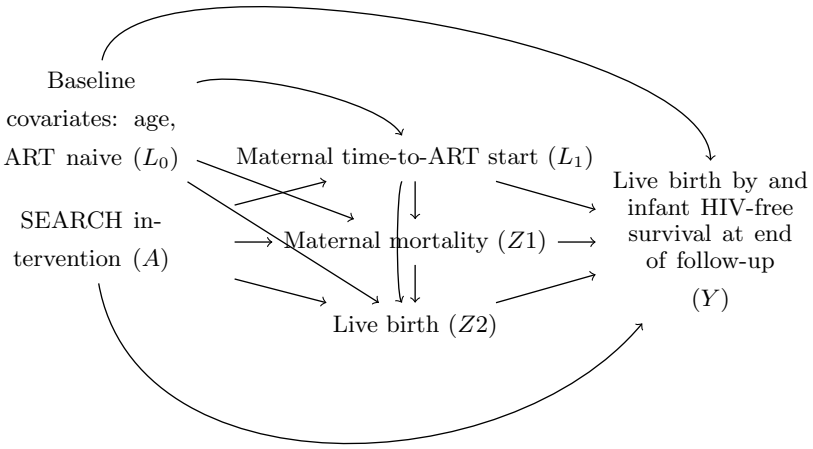
2.1 Introduction

Perinatal HIV transmission occurs when HIV is passed from a pregnant person living with the virus to their child during pregnancy, labor, delivery, or breastfeeding.[40] Although the rate of perinatal HIV transmission may be as high as 45% in the absence of any intervention,[40] transmission rates can be reduced to less than 5% with treatment, which is comprised primarily of antiretroviral therapy (ART) for the birth parent and child.[41] In 2022, there were an estimated 130,000 new infections among children globally, the majority of which occurred in sub-Saharan Africa.[42] The consequences of perinatal HIV infection can be severe. Without treatment, an estimated one-third of HIV-positive children will die by one year of age, and an estimated one-half of HIV-positive children will die by age two.[43] Children living with HIV, including those who receive ART, are also at higher risk of additional medical complications[44] such as respiratory impairment,[45] stunting, wasting,[46] and other metabolic disorders.[47,48] Many HIV-positive children do not receive ART, and as a result, while globally children compose 4% of people living with HIV, they account for 13% of deaths from AIDS-related causes.[42] Interventions and improvements in HIV care and delivery are needed to reduce the burden of perinatal HIV transmission and achieve global elimination goals.[42]

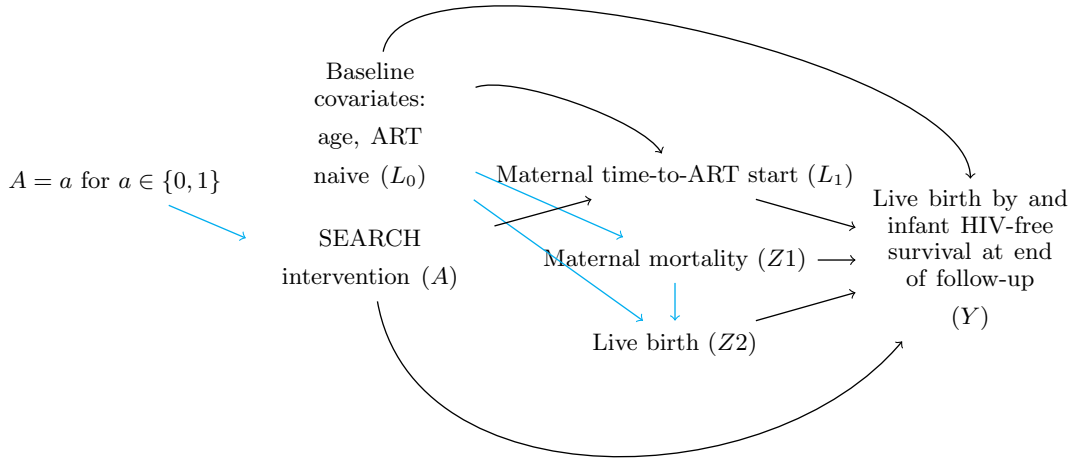
The Sustainable East Africa Research in Community Health (SEARCH) Study was a parallel-arm pair-matched cluster randomized trial designed to evaluate the effectiveness of a multi-component, patient-centered care delivery intervention on population health, including perinatal HIV transmission, in rural Kenya and Uganda.[26,49] We found, on average, infant HIV-free survival was significantly higher and the rate of perinatal HIV transmission significantly lower in communities that received the SEARCH intervention as compared to control communities after three years.[29]

The SEARCH intervention may have operated in different ways to drive the observed improvement in infant outcomes. The primary hypothesized pathway is by increasing the probability that the mother was treated effectively with ART throughout her pregnancy and during breastfeeding. However, in addition, it is possible that the intervention impacted the population of infants among whom the outcome was measured, via impacts on the live birth process (when considering the outcome of infant HIV-free survival) and the live birth process and infant survival (when considering the outcome of perinatal HIV transmission) (Figure 2.1A). The SEARCH intervention may have affected maternal survival and the probability of live birth (i.e., the "live birth process"), consequently impacting the types of people who had a live birth and the timing of live births over the course of the follow-up period.

As described in Chapter 1, mediational parameters can be applied to better understand the mechanisms by which an intervention such as SEARCH operated to affect outcomes. A counterfactual strata total effect quantifies impact of the SEARCH intervention on infant HIV-free survival among live births if everyone was assigned to treatment as compared to if everyone was assigned to control. In addition, counterfactual stratum direct effect parameters (CSDE) formulated using stochastic interventions on mediating variables (e.g., the live birth process) can be used to understand the impact of the SEARCH intervention absent treatment-induced changes to the live birth process. Stochastic interventions correspond to drawing mediating variables from distributions rather than hypothetically intervening to set mediators to a particular value. In this chapter, we use counterfactual stratum direct effects estimated using longitudinal targeted maximum likelihood estimation (TMLE) to estimate the effect of the SEARCH intervention on infant HIV-free survival and perinatal HIV transmission. We also revisit the definition and estimation of counterfactual strata total effects (CSTE) as applied to these infant health outcomes in the SEARCH Study.



(A) Directed Acyclic Graph (DAG) with maternal survival and live birth (together, composing the "live birth process") as mediators of the effect of the SEARCH intervention on infant HIV-free survival illustrated using one post-baseline timepoint.



(B) Post-intervention DAG representing hypothetical intervention on live birth process if values of maternal survival and live birth were drawn from distributions conditional on baseline covariates and past mediators.

Figure 2.1: Simplified representation of the mechanisms by which the SEARCH Study may have impacted infant HIV-free survival.

2.2 Study Setting and Context

The SEARCH Study was a pair-matched, cluster randomized trial implemented in 32 communities in East Africa between 2013-2016. In all 32 communities, residents were eligible to participate in multi-disease health fairs, including HIV testing, at baseline; baseline HIV testing coverage reached 90% of adult residents. In the 16 control communities, ART was prescribed on the basis of the national standard of care guidelines, which at study start restricted ART eligibility. In the 16 intervention communities, additional health fairs were conducted annually between 2013 and 2016; all residents who tested positive for HIV at any during follow-up were eligible for immediate ART treatment (i.e. "treat all").[26,49,50] Additionally, a streamlined person-centered care delivery model was implemented in intervention communities to improve health outcomes by reducing barriers to care (e.g., long waiting times), fostering better relationships between patients and clinic staff (e.g., by increasing the accessibility of clinicians), and improving knowledge and reducing stigma relating to ART and HIV.[50]

Our study sample consisted of all women living with HIV of reproductive age (defined as age between 12 and 52 years of age) at study baseline who did not outmigrate over the course of the follow-up period. At the end of follow-up, women who self-reported having given birth in the prior four year period were asked to provide each infant's name, date of birth, birth status, and vital status. Based on maternal reporting of infant names and dates of birth, all live births reported by women living with HIV were linked to study identifiers for infants born after the start of the baseline community health clinic and before the end of follow-up in year three of the study.

Because reported names and birth dates were inexact across different data sources, we performed fuzzy matching based on birth names and dates using the *fuzzyjoin* package[51] in `R`[52] to link reported births to infants assigned SEARCH study identifiers. Matching was first performed at the household level by fuzzy joining infants in each household with all reported births by women in the same household. Any remaining unmatched births from the household-level matching step were included in a pool for matching to infants in the same community. Matches were manually reviewed by at least one person after each fuzzy matching step, blinded to intervention arm, and any matches of poor quality were excluded. A manual search within households for infants with similar names and dates of birth was performed to resolve any remaining unmatched live births.

Infant vital status was based on maternal self-report, and live births were linked to HIV testing data that was collected at the end of the follow-up to ascertain infant infection status. HIV testing data from year three community health campaigns and tracking efforts were used to assess HIV-infection status for infants when available. For infants less than 18 months of age at the date of testing, DNA PCR tests and negative antibody tests determined HIV status; positive antibody tests at less than 18 months of age were not considered to be an indication of infant HIV status because maternal HIV antibodies may persist among infants in this age group.[53] For infants greater than 18 months of age at the date of testing, DNA PCR tests and/or any antibody tests were used to determine HIV infection status. Mother-

infant pairs were excluded from our analyses if infants were reported to be alive at the end of the study but missing HIV infection status.

2.3 Methods: The Causal Roadmap

Causal Questions

To understand the overall effect of the SEARCH intervention on infant HIV-free survival, we used a counterfactual strata total effect to quantify the impact of the SEARCH intervention on infant HIV-free survival among counterfactual live births if all women of reproductive age living with HIV at study baseline were assigned to receive the intervention as compared to if all were assigned to the control condition. Because of possible treatment-induced changes to the types of women who have a live birth and the timing of births, we additionally use counterfactual stratum direct effects to summarize the effect of the SEARCH intervention absent any treatment-induced changes to the live birth process. In other words, a counterfactual stratum direct effect answers the question: what is the effect of the SEARCH intervention on infant HIV-free survival if all women in the target population were assigned to the intervention as compared to the control and if the live birth process (i.e., maternal survival and live birth), were drawn from the same distribution under both treatment assignments? To formulate a counterfactual stratum direct effect parameter, we draw mediator values (i.e., the distributions of maternal survival and live birth) from a specified distribution, for example the observed distribution of the live birth process in the intervention arm of the study. The choice of this mediator distribution determines the composition of the group among which the effect of the SEARCH intervention on infant outcomes will be evaluated; other choices of this mediator distribution are discussed in greater detail in Section 2.3.

Structural Causal Model (SCM)

We adopt the the structural causal model (SCM), representing background knowledge about our data generating process, described in Chapter 1, and have summarized it below for ease of reference. We use a longitudinal causal model that accommodates time-varying covariates and mediators at multiple time points. For ease of notation and presentation, we describe this structural causal model as a pooled individual-level data structure; this SCM generalizes easily to a hierarchical SCM accounting for clustering at the community level.[33,34] The SEARCH Study was conducted over a three-year period; we use τ to denote the final timepoint of the study and each time t represents one year of study follow-up. For any time-varying variable, V, we use $\bar{V}_t$ to represent the vector of covariate values $(V_1, \ldots, V_t)$.

The baseline covariates, L_0, consist of maternal age, baseline maternal ART status (treatment-naive vs. treatment-experienced), and community-level HIV prevalence. A is an indicator of whether or not an individual lived in a community that was randomized to

receive the SEARCH intervention. For the outcome of infant HIV-free survival, time-varying covariates, L_t, consist of maternal ART status and infant vital status at each time t.

For the outcome of perinatal HIV transmission, maternal ART status is a time-varying covariate, L_t. We define an additional variable $Z3_t$ to represent the composite variable of live birth and infant survival at time t (i.e., an indicator of live birth ($Z2_t$) and infant survival ($Y1_t$)). In a slight abuse of notation, we use Z_t to denote both the vector of two variables, $Z1_t$ and $Z2_t$, representing maternal survival and live birth, respectively, at each time t (for the outcome of infant HIV-free survival), as well as the vector of two variables, $Z1_t$ and $Z3_t$, representing maternal survival and the composite variable of live birth and infant survival, respectively, at each time t (for the outcome of perinatal HIV transmission). Together, the mediating variables $\bar{Z}1_\tau$ and $\bar{Z}2_\tau$ comprise the "live birth process." The vector of the mediators $Z1_t$ and $Z3_t$ represents the process of live birth and infant survival that mediates the relationship between the exposure (i.e., the SEARCH intervention) and the outcome of infant HIV-infection ($Y2_{\tau+1}$) at the end of follow-up.

Maternal survival is a time-varying confounder of the relationship between live birth and the infant outcomes (i.e., infant vital status and infant HIV infection status) because it deterministically affects the live birth process, as well as potentially affects infant survival and subsequent risk of infection. Thus, we account for it as an endogenous node in our SCM. Hypothetical interventions on the mediator of live birth may be poorly defined if we do not jointly intervene upon both maternal survival and live birth when formulating the counterfactual stratum direct effect parameters. If maternal survival is treated as a time-varying covariate rather than a mediator, it is possible to draw values of live birth we know to be incompatible with the value of maternal survival when the hypothetical stochastic intervention on live birth is drawn conditional only on baseline covariates.[14] Similarly, to avoid a poorly defined hypothetical intervention on the mediators when evaluating the outcome of perinatal HIV transmission, we intervene upon maternal survival, live birth, and infant survival together; a hypothetical intervention on infant survival without consideration of maternal survival and live birth could result in values of infant survival which we know to be inconsistent with the values of maternal survival and time-to-live birth.

Infant HIV-free survival among counterfactual live births is defined using two variables, $Y1_{\tau+1}$ and $Y2_{\tau+1}$, to represent infant survival and absence of infant HIV infection, respectively. Infant outcomes will be missing if no live birth occurred; $Y_{\tau+1}$ represents the composite event of live birth ($Z2_\tau = 1$) and infant HIV-free survival ($Y1_{\tau+1} = 1$, $Y2_{\tau+1} = 0$) at the final timepoint. $M_{\tau+1}$ represents the composite event of live birth by and infant survival ($Z3_\tau = 1$) with HIV infection ($Y2_{\tau+1} = 1$) at the time τ. The time-varying covariates of maternal time-to-ART initiation, time-to-live birth, and infant vital status over time, are continuous variables that we have discretized into three time points ($\tau = 3$). We define each of these variables as an indicator of whether the event (e.g., maternal ART initiation) occurred by the end of the time interval t. Maternal time-to-ART initiation and infant vital status are time-to-event variables; once the event occurs, all subsequent values of the same variable are constant.

To preserve the time-ordering of these variables after discretizing the continuous data, we

made the exclusion restrictions that time-varying covariates do not affect one another or time-varying mediators within the same time interval t. In addition, because the SEARCH Study was a randomized controlled trial, we also made the exclusion restriction that the baseline covariates L_0 do not affect the intervention A. $U = (U_{L_0}, U_A, U_{\bar{L}_t}, U_{\bar{Z}1_t}, U_{\bar{Z}2_t}, U_{Y1_{\tau+1}}, U_{Y2_{\tau+1}})$, $t = 1, \ldots, \tau$ are the exogenous nodes representing unmeasured variables and have joint distribution P_U. A simplified version of this SCM, which only considers the post-treatment covariates and mediators at one time point, is represented as a DAG in Figure 2.1A. This figure illustrates the mechanisms of interest by which the SEARCH intervention may have impacted infant outcomes, all of which are captured within the counterfactual strata total effect; it also includes a post-intervention DAG illustrating the effect pathways captured by a counterfactual stratum direct effect of the SEARCH intervention on infant HIV-free survival.

Causal Estimands

Counterfactual Strata Total Effects

In the SEARCH Study, we can use a counterfactual strata total effect to summarize the difference in the risk of counterfactual infant HIV-free survival among counterfactual live births if all women of reproductive age living with HIV at study baseline were assigned to receive the SEARCH intervention as compared to if everyone was assigned to the control. We can also formulate a CSTE parameter to describe the expected difference in counterfactual perinatal HIV transmission (i.e., the difference in counterfactual infant HIV infection among counterfactual live births for which the infant survived until the end of the study) if everyone in the population were assigned to receive the SEARCH intervention as compared to the control condition. Formal definition of these parameters can be found in Appendix B.1.

Counterfactual Stratum Direct Effects

The counterfactual strata total effect of the SEARCH intervention on infant HIV-free survival includes effects of the SEARCH intervention that operate through changes to the types of women who had a live birth as well as the timing of live births. In contrast, the counterfactual stratum direct effect summarizes the difference in infant HIV-free survival among live births if all women of reproductive age living with HIV at study baseline were assigned to receive the SEARCH intervention versus the comparator condition and if the live birth process, which involves both maternal survival ($\bar{Z}1_\tau$) and live birth ($\bar{Z}2_\tau$), were drawn from a constant stochastic distribution (Appendix B.1). This parameter corresponds to a static intervention to set the exposure variable $A = 1$ as compared to $A = 0$ and a stochastic intervention on the mediators to draw their values $\bar{Z}_\tau$ from a distribution $\bar{G}_{L_0}$, that is possibly conditional on baseline covariates ($\bar{Z}_\tau \sim \bar{G}(L_0)$). $\bar{G}(\bar{Z}_\tau|L_0) \equiv \prod_{t=1}^{\tau} G_t(Z_t|\bar{Z}_{t-1}, L_0)$ refers to a user-specified conditional distribution of the mediators, $\bar{Z}_\tau$ given baseline covariates L_0, and $G_t(Z_t|\bar{Z}_{t-1}, L_0)$ represents the probability distribution of the mediator at time t given the

mediator past and baseline covariates. We will sometimes use $\bar{G}$ to refer to the conditional distribution $\bar{G}_{L_0}$ for ease of notation. By formulating this parameter such that the values of the live birth process are drawn randomly from the same distribution irrespective of the intervention on the exposure variable A, this parameter excludes any impact of the SEARCH intervention on infant HIV-free survival that operates by changing the timing of live births or the composition of the group that had live births over the follow-up period.

In addition, we use a counterfactual stratum direct effect of the SEARCH intervention on perinatal HIV transmission to understand the difference in perinatal HIV transmission among infants who survived until the end of follow-up if all women of reproductive age living with HIV at study baseline were assigned to receive the SEARCH intervention versus the control condition, and if maternal survival, live birth, and infant survival, were all drawn from the same distribution under treatment and control (Appendix B.1). Unlike the CSTE of SEARCH on perinatal HIV transmission, this parameter does not include any effect of the SEARCH intervention on HIV transmission that occurred due to changes to the timing of live births across the follow-up period nor to the types of people who had infants who survived until the end of the study.

In the special case that the hypothetical intervention on mediating variables used to formulate the counterfactual stratum direct effect corresponds to setting the mediators to deterministically to a constant value, the counterfactual stratum direct effect will be equal to a controlled direct effect. However, in the SEARCH example, and many other mediation problems, the choice of the mediator distribution, $\bar{G}$, used to formulate the CSDE will impact the true parameter value and the composition of the group among which infant outcomes are evaluated. Selecting a mediator distribution, $\bar{G}$, that is estimated from the observed data results in a data-adaptive target parameter.[36] In this chapter, we estimate different CSDE formulated using different data-dependent choices of $\bar{G}$, noting that because the data-adaptivity only occurs due to estimation of the mediator distribution, identification and estimation of such parameters can proceed treating the sample-specific estimate of the mediator distribution as known, providing valid inference of the resulting sample-specific parameter.[36] Alternatively, the true counterfactual mediator distribution under treatment or control can be used to formulate a counterfactual stratum direct effect parameter, as discussed in more detail in Chapter 1; this type of CSDE requires alternative assumptions for identifiability and approaches for estimation from those discussed in this chapter

Observed Data

The observed data for each individual includes individual-level baseline covariates, a community-level indicator of treatment, individual-level time-varying covariates and mediators, and individual-level infant outcomes, which can be represented as:

$$O = (L_0, A, \bar{L}_t, \bar{Z}_t, Y1_{\tau+1}, Y2_{\tau+1}, Y_{\tau+1}, M_{\tau+1})$$

Treating the data as pooled individual-level data, the observed data consists of n inde-

pendent and identically distributed (i.i.d.) copies of O. Because treatment A was randomly assigned, the treatment mechanism is known, and the statistical model is semi-parametric.

Identification

We can identify the counterfactual strata total effects of the SEARCH intervention on infant HIV-free survival and perinatal HIV transmission under assumptions of 1) no unmeasured exposure-outcome confounding and 2) a positive probability of each treatment of interest conditional on the past within all values of baseline covariates (Appendix B.2). The outcomes, $Y_{\tau+1}$ and $M_{\tau+1}$, are composite outcomes consisting of the live birth, infant vital status, and infection status variables. Consequently, the assumption of no unmeasured exposure-outcome confounding requires assumptions of conditional independence of 1) the exposure-live birth relationship, 2) the exposure-infant vital status relationship, and 3) the exposure-infant infection status relationship at the final timepoint. Because the SEARCH Study was a randomized trial, these assumptions are satisfied by design.

Additional assumptions are required to identify counterfactual stratum direct effects. Assuming no unmeasured 1) exposure-outcome and 2) mediator-outcome confounding and the associated positivity assumptions (i.e., a positive probability of the known mediator distribution within all values of baseline and time-varying covariates), each of the counterfactual stratum direct effects are identifiable (Appendix B.2). The first randomization assumption will be satisfied in a randomized study such as SEARCH.

The second assumption of no mediator-outcome confounding will be satisfied if the measured baseline and time-varying covariates can control for confounding between the mediator and outcome. With respect to the outcome of infant HIV-free survival, this second assumption will be satisfied if we believe maternal age, baseline maternal ART status, community-level HIV prevalence, time-to-maternal ART start, and time-varying infant vital status sufficiently control from confounding between the live birth process (i.e., maternal survival and live birth) and infant HIV-free survival at the end of the study. For the outcome of perinatal HIV transmission, this randomization will be satisfied if we believe the baseline covariates and time-varying maternal ART status account for confounding between the process of infant survival (i.e., maternal survival, live birth, and infant survival) and infant HIV-infection at the end of the study.

Estimands

If the assumptions to identify the counterfactual strata total effect of the SEARCH intervention on infant HIV-free survival are satisfied, the causal estimand (represented as a risk difference) is identified as the following statistical estimand:

$$\frac{E_0\left[E_0(Y_{\tau+1}|A=1,L_0)\right]}{E_0\left[E_0(Z2_{\tau}|A=1,L_0)\right]} - \frac{E_0\left[E_0(Y_{\tau+1}|A=0,L_0)\right]}{E_0\left[E_0(Z2_{\tau}|A=0,L_0)\right]}$$

Similarly, if the assumptions to identify the counterfactual strata total effect of the SEARCH intervention on perinatal HIV transmission are satisfied, the causal estimand (represented as a risk difference) is identified as:

$$\frac{E_0\left[E_0(M_{\tau+1}|A=1,L_0)\right]}{E_0\left[E_0(Z3_\tau|A=1,L_0)\right]} - \frac{E_0\left[E_0(M_{\tau+1}|A=0,L_0)\right]}{E_0\left[E_0(Z3_\tau|A=0,L_0)\right]}$$

The numerators of the statistical estimands are the g-computation estimands for the composite outcomes, $Y_{\tau+1}$ and $M_{\tau+1}$. The denominators of the statistical estimands are the g-computation estimands for the outcome of live birth by time τ, $Z2_\tau$. In addition, if the assumptions to identify the counterfactual stratum direct effects are satisfied, then the causal estimand for the counterfactual stratum direct effect for the outcome of infant HIV-free survival is identified as the following statistical estimand:

$$\Psi(P_0) = \frac{\begin{array}{c}\sum\limits_{l_0,\bar{l}_\tau,\bar{z}_\tau}(E_0[Y_{\tau+1}|\bar{Z}_\tau=\bar{z}_\tau,\bar{L}_t=\bar{l}_t,A=1,L_0=l_0]\times\bar{G}(\bar{Z}_\tau=\bar{z}_\tau|L_0=l_0)\times \\ P_{0,L_0}(L_0=l_0)\prod\limits_{t=1}^{\tau}P_{0,L_t}(L_t=l_t|\bar{Z}_{t-1}=\bar{z}_{t-1},\bar{L}_{t-1}=\bar{l}_{t-1},A=1,L_0=l_0))\end{array}}{\begin{array}{c}\sum\limits_{l_0,\bar{z1}_\tau,\bar{z2}_{\tau-1}}G_\tau(Z2_\tau=1|\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z2}_{\tau-1}=\bar{z2}_{\tau-1},L_0=l_0)\times \\ \bar{G}(\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z2}_{\tau-1}=\bar{z2}_{\tau-1}|L_0=l_0)\times P_{0,L_0}(l_0)\end{array}}$$

$$- \frac{\begin{array}{c}\sum\limits_{l_0,\bar{l}_\tau,\bar{z}_\tau}(E_0[Y_{\tau+1}|\bar{Z}_\tau=\bar{z}_\tau,\bar{L}_t=\bar{l}_t,A=0,L_0=l_0]\times\bar{G}(\bar{Z}_\tau=\bar{z}_\tau|L_0=l_0)\times \\ P_{0,L_0}(L_0=l_0)\prod\limits_{t=1}^{\tau}P_{0,L_t}(L_t=l_t|\bar{Z}_{t-1}=\bar{z}_{t-1},\bar{L}_{t-1}=\bar{l}_{t-1},A=0,L_0=l_0))\end{array}}{\begin{array}{c}\sum\limits_{l_0,\bar{z1}_\tau,\bar{z2}_{\tau-1}}G_\tau(Z2_\tau=1|\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z2}_{\tau-1}=\bar{z2}_{\tau-1},L_0=l_0)\times \\ \bar{G}(\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z2}_{\tau-1}=\bar{z2}_{\tau-1}|L_0=l_0)\times P_{0,L_0}(l_0)\end{array}}$$

Similarly, for the outcome of perinatal HIV transmission, the counterfactual stratum direct effect is identified as the following statistical estimand:

$$\Psi(P_0) = \frac{\begin{array}{c}\sum_{l_0,\bar{l}_\tau,\bar{z}_\tau}(E_0[M_{\tau+1}|\bar{Z}_\tau=\bar{z}_\tau,\bar{L}_t=\bar{l}_t,A=1,L_0=l_0]\times\bar{G}(\bar{Z}_\tau=\bar{z}_\tau|L_0=l_0)\times \\ P_{0,L_0}(L_0=l_0)\prod_{t=1}^{\tau}P_{0,L_t}(L_t=l_t|\bar{Z}_{t-1}=\bar{z}_{t-1},\bar{L}_{t-1}=\bar{l}_{t-1},A=1,L_0=l_0))\end{array}}{\begin{array}{c}\sum_{l_0,\bar{z1}_\tau,\bar{z3}_{\tau-1}}G_\tau(Z3_\tau=1|\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z3}_{\tau-1}=\bar{z3}_{\tau-1},L_0=l_0)\times \\ \bar{G}(\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z3}_{\tau-1}=\bar{z3}_{\tau-1}|L_0=l_0)\times P_{0,L_0}(l_0)\end{array}}$$

$$-\frac{\begin{array}{c}\sum_{l_0,\bar{l}_\tau,\bar{z}_\tau}(E_0[M_{\tau+1}|\bar{Z}_\tau=\bar{z}_\tau,\bar{L}_t=\bar{l}_t,A=0,L_0=l_0]\times\bar{G}(\bar{Z}_\tau=\bar{z}_\tau|L_0=l_0)\times \\ P_{0,L_0}(L_0=l_0)\prod_{t=1}^{\tau}P_{0,L_t}(L_t=l_t|\bar{Z}_{t-1}=\bar{z}_{t-1},\bar{L}_{t-1}=\bar{l}_{t-1},A=0,L_0=l_0))\end{array}}{\begin{array}{c}\sum_{l_0,\bar{z1}_\tau,\bar{z3}_{\tau-1}}G_\tau(Z3_\tau=1|\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z3}_{\tau-1}=\bar{z3}_{\tau-1},L_0=l_0)\times \\ \bar{G}(\bar{Z1}_\tau=\bar{z1}_\tau,\bar{Z3}_{\tau-1}=\bar{z3}_{\tau-1}|L_0=l_0)\times P_{0,L_0}(l_0)\end{array}}$$

The denominator of the statistical estimand for the counterfactual stratum direct effect of the SEARCH intervention on infant HIV-free survival is equal to the mean probability of live birth under the known distribution $\bar{G}$ from which maternal survival and live birth are drawn. For the perinatal HIV transmission outcome, the denominator of the statistical estimand of this direct effect is equal to the mean probability of live birth and infant survival under the known distribution $\bar{G}$ from which maternal survival, live birth, and infant survival are drawn.

Estimation and Inference

Targeted maximum likelihood estimation

We used longitudinal targeted maximum likelihood estimation to estimate both the counterfactual strata total and direct effect parameters.[54] TMLE can improve the precision and reduce bias of parameter estimates relative to other estimation methods and can leverage machine learning algorithms for estimation of nuisance parameters while retaining valid inference.[55]

Counterfactual strata total effects were estimated using the `R` package `ltmle`[56]. Estimation of each total effect required estimation of two conditional means, the first representing the expected value of the composite outcome (Y_τ or M_τ) and the second representing the expected value of the denominator, either live birth by time τ for the outcome of infant HIV-free survival or infant survival at time τ for the outcome of perinatal HIV transmission (Appendix B.3). TMLE also used the estimated treatment mechanism (i.e., the probability of treatment A given the past). We estimated the conditional means and the conditional probabilities of treatment using Super Learner[57] with a small library of algorithms (`SL.mean`

for simple means, `SL.glm` for main terms linear models, and `SL.gam`[58,59,60] for generalized additive models). In addition, for comparison, we also estimated nuisance parameters (e.g., conditional expectations and treatment mechanisms) using main terms logistic regression.

Estimation of the counterfactual stratum direct effect required estimation of a series of nested conditional expectations of the composite outcome of infant HIV-free survival (or perinatal HIV transmission) given the past (Appendix B.3).[16,17,61] Super Learner was also incorporated into estimation of the longitudinal TMLE used to estimate counterfactual stratum direct effects under different choices of mediator distributions.

We accounted for possible dependence among observations within the same community in the estimation of the standard errors using the hierarchical TMLE approach described in Benitez et al.[34] Inference for both total and direct effect estimates was obtained by applying the functional delta method to estimate the influence curves of the counterfactual strata total effect and counterfactual stratum direct effect estimators after accounting for clustering by community.

2.4 Results

As can be seen in Table 2.1, in our sample, there were $n = 7280$ women living with HIV at study baseline who met the inclusion criteria for the study (i.e., were between the ages of 12 and 52, did not outmigrate during the study). Of these, 923 (12.7%) gave birth over the course of the follow-up period, and, overall, 903 of the infants survived through the end of the study. The probability of live birth was slightly higher in the control arm (13.2%) than in the intervention arm (12.3%). There were 19 HIV infections among infants who survived until the end of follow-up, with 17 of these infections occurring in the control arm and 2 infections in the intervention arm of the study.

Table 2.1: Summary of baseline and select post-baseline variables from the SEARCH Study overall and by study arm.

		Full sample ($n = 7280$)		Control ($n = 3360$)		Intervention ($n = 3920$)	
Baseline							
		n	%	**n**	%	**n**	%
ART-experienced	No	3137	43.09	1505	44.79	1632	41.63
	Yes	4143	56.91	1855	55.21	2288	58.37
		Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Community HIV prevalence		0.16	0.08	0.15	0.08	0.16	0.07
Maternal age		33.66	9.16	33.64	9.09	33.67	9.22
Post-Baseline							
		n	%	**n**	%	**n**	%
Live birth by end of study	No	6357	87.32	2918	86.85	3439	87.73
	Yes	923	12.68	442	13.15	481	12.27
Live birth and infant survival through end of study	No	6377	87.6	2929	87.17	3448	87.96
	Yes	903	12.4	431	12.83	472	12.04
Live birth and infant HIV-free survival at end of study	No	6396	87.86	2946	87.68	3450	88.01
	Yes	884	12.14	414	12.32	470	11.99
Live birth and infant HIV infection at end of study	No	7261	99.74	3343	99.49	3918	99.95
	Yes	19	0.26	17	0.51	2	0.05

Counterfactual strata total effects in the SEARCH Study

Infant HIV-free survival

The estimated relative risk of the SEARCH intervention on infant HIV-free survival using Super Learner was 1.04 (95% CI: $1.02, 1.07$) (Table 2.2). Because the assumption of no unmeasured confounding of the exposure-outcome relationship is satisfied in a randomized trial such as SEARCH, we can interpret this estimate as the expected increase in the probability of infant HIV-free survival among live births if everyone was assigned to receive the SEARCH intervention as compared to if everyone was assigned to the control condition. In other words, the SEARCH intervention increased the probability that infants born during follow-up to mothers living with HIV at study baseline would survive and remain HIV uninfected by the end of the study by 4%. On an absolute scale, the intervention reduced the probability that an infant born to a woman living with HIV would die or become HIV infected by 3.9% (95% CI: $1.5\%, 6.3\%$), a substantive improvement given the 6% baseline risk of infant death or HIV infection among infants born in the control communities.

Table 2.2: Total effect estimates of infant HIV-free survival using TMLE with and without Super Learner

	Using main terms logistic regression		Using Super Learner*	
Infant HIV-free Survival	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth and infant HIV-free survival	0.1195	(0.1023, 0.1367)	0.1203	(0.1028, 0.1378)
Estimated probability of live birth	0.1224	(0.1046, 0.1401)	0.1232	(0.1052, 0.1412)
Estimated probability of infant HIV-free survival among live births	0.9767	(0.9621, 0.9915)	0.9763	(0.9621, 0.9908)
A=0 (Control)				
Estimated probability of live birth and infant HIV-free survival	0.1236	(0.1080, 0.1392)	0.1228	(0.1073, 0.1383)
Estimated probability of live birth	0.1319	(0.1148, 0.1489)	0.1310	(0.1140, 0.1479)
Estimated probability of infant HIV-free survival among live births	0.9374	(0.9181, 0.9572)	0.9375	(0.9184, 0.9570)
Risk Difference				
$\hat{\Psi}_{RD}$	0.0393	(0.0149, 0.0636)	0.0389	(0.0151, 0.0626)
Risk Ratio				
$\hat{\Psi}_{RR}$	1.0419	(1.0156, 1.0689)	1.0414	(1.0158, 1.0678)

* Super Learner library consists of: SL.mean, SL.glm, SL.gam

Perinatal HIV transmission

The estimated relative risk of the SEARCH intervention on perinatal HIV transmission using Super Learner was 11% (95% CI: 2%, 75%), meaning the expected risk of HIV infection among infants who survived through the end of the study if everyone was assigned to receive the SEARCH intervention was was 11% of the risk of perinatal HIV transmission to if everyone was assigned to the control condition (Table 2.3). The estimates of the counterfactual strata total effects for both outcomes were very similar if the conditional expectations and treatment mechanisms were estimated using main terms regression instead of Super Learner, as expected given the small Super Learner library considered, the limited number of communities, and the small number of covariates.

Table 2.3: Total effect estimates of perinatal HIV transmission using TMLE with and without Super Learner

	Using main terms logistic regression		Using Super Learner*	
Perinatal HIV Transmission	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0005	(0.0000, 0.0015)	0.0005	(0.0000, 0.0015)
Estimated probability of live birth and infant survival	0.1200	(0.1025, 0.1375)	0.1208	(0.1030, 0.1385)
Estimated probability of infant HIV infection among infants alive at end of study	0.0042	(0.0006, 0.0290)	0.0042	(0.0006, 0.0285)
A=0 (Control)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0050	(0.0029, 0.0071)	0.0050	(0.0029, 0.0071)
Estimated probability of live birth and infant survival	0.1286	(0.1120, 0.1451)	0.1278	(0.1113, 0.1443)
Estimated probability of infant HIV infection among infants alive at end of study	0.0390	(0.0269, 0.0565)	0.0394	(0.0272, 0.0572)
Risk Difference				
$\hat{\Psi}_{RD}$	-0.0347	(-0.0514, -0.0181)	-0.0352	(-0.0518, -0.0186)
Risk Ratio				
$\hat{\Psi}_{RR}$	0.1083	(0.0151, 0.7748)	0.1074	(0.0155, 0.7461)

* Super Learner library consists of: SL.mean, SL.glm, SL.gam

Counterfactual stratum direct effects in the SEARCH Study

Infant HIV-free survival

We estimated the counterfactual stratum direct effect of the SEARCH Study on infant HIV-free survival using three different data-adaptive choices of mediator distributions: 1) the observed mediator distribution estimated from the intervention arm of the trial; 2) the mediator distribution estimated from the control arm of the trial; and 3) the estimated mediator distribution conditional on baseline covariates in the overall sample. We regressed the mediator on all past mediators and baseline covariates among 1) intervention arm participants, 2) control arm participants, and 3) the overall sample, respectively, to estimate the predicted probabilities of each mediator at time t conditional on the mediator past and baseline variables. Under the observed mediator distribution in the intervention arm, we estimated the relative risk of the SEARCH intervention on infant HIV-free survival absent any intervention-induced changes to the live birth process to be 1.05 (95% CI: $1.03, 1.08$) (Table 2.4). If there is no unmeasured confounding of the exposure-outcome and mediator-outcome relationships, this point estimate can be interpreted as the expected change in the probability of infant HIV-free survival among live births if everyone received the SEARCH intervention as compared to the comparator condition and if maternal survival and live birth were drawn from the same distribution in either case (i.e., if the types of people and timing of live births remained the same).

Estimates of the counterfactual stratum stochastic direct effect of the SEARCH intervention on infant HIV-free survival using the other choices of known distributions for the mediators were similar to the results using the estimated mediator distribution from the intervention arm (see Appendix B.3 for full results). For the outcome of infant HIV-free survival, estimated standard errors were similar across all three choices of mediator distributions used to formulate the direct effect parameter; however, for the perinatal HIV transmission outcome, the estimated standard error using the estimated mediator distribution from the overall sample was smaller than the estimated standard errors using both the estimated mediator distribution in the intervention arm and the estimated mediator distribution in the control arm.

Table 2.4: Stochastic direct effect estimated using TMLE under the observed mediator distribution in the intervention arm.

	Using main terms logistic regression		Using Super Learner*	
Infant HIV-free Survival	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth and infant HIV-free survival	0.1185	(0.1161, 0.1210)	0.1184	(0.1160, 0.1209)
Estimated probability of live birth	0.1218	(0.1188, 0.1249)	0.1218	(0.1188, 0.1249)
Estimated probability of infant HIV-free survival among live births	0.9728	(0.9554, 0.9904)	0.9722	(0.9553, 0.9894)
A=0 (Control)				
Estimated probability of live birth and infant HIV-free survival	0.1123	(0.1086, 0.1161)	0.1123	(0.1089, 0.1159)
Estimated probability of live birth	0.1218	(0.1188, 0.1249)	0.1218	(0.1188, 0.1249)
Estimated probability of infant HIV-free survival among live births	0.9219	(0.8997, 0.9446)	0.9223	(0.9021, 0.9429)
Risk Difference				
$\hat{\Psi}_{RD}$	0.0509	(0.0234, 0.0784)	0.0499	(0.0244, 0.0755)
Risk Ratio				
$\hat{\Psi}_{RR}$	1.0552	(1.0247, 1.0865)	1.0541	(1.0259, 1.0831)

* SuperLearner library consists of: SL.mean, SL.glm, SL.gam

Perinatal HIV transmission

Similar to the outcome of infant HIV-free survival, the relative risk of the SEARCH intervention on perinatal HIV transmission excluding any intervention-induced changes to the process of live birth and infant survival was estimated using three different choices of data-adaptive mediator distributions $\bar{G}$. Using the observed mediator distribution of maternal survival, live birth, and infant survival in the intervention arm, the relative risk of the SEARCH intervention on perinatal HIV transmission was estimated to be 0.15 (95% CI: $0.05, 0.39$) (Table 2.5). Under the assumptions of no unmeasured confounding of the exposure-outcome and mediator-outcome relationships, this estimate can be interpreted as the expected change in the risk of perinatal HIV transmission if everyone received the SEARCH intervention as compared to the control condition and if maternal survival, live birth, and infant survival were drawn from the same distribution in either case (i.e., if the types of people and pattern of live birth and infant survival remained the same). Refer to Appendix B.3 for the estimates of the CSDE of the SEARCH intervention on perinatal HIV transmission using other choices of mediator distributions. Estimates of the CSDE using main terms logistic regression were very similar to those using Super Learner for both infant outcomes.

Table 2.5: Stochastic direct effect estimated using TMLE under the observed mediator distribution in the intervention arm.

	Using main terms logistic regression		Using Super Learner*	
Perinatal HIV Transmission	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0006	(0.0001, 0.0029)	0.0008	(0.0003, 0.0020)
Estimated probability of live birth and infant survival	0.1193	(0.1171, 0.1215)	0.1193	(0.1171, 0.1215)
Estimated probability of infant HIV infection among infants alive at end of study	0.0049	(0.0010, 0.0241)	0.0064	(0.0025, 0.0165)
A=0 (Control)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0054	(0.0041, 0.0072)	0.0052	(0.0040, 0.0068)
Estimated probability of live birth and infant survival	0.1193	(0.1171, 0.1215)	0.1193	(0.1171, 0.1215)
Estimated probability of infant HIV infection among infants alive at end of study	0.0456	(0.0344, 0.0605)	0.0437	(0.0334, 0.0571)
Risk Difference				
$\hat{\Psi}_{RD}$	-0.0408	(-0.0559, -0.0256)	-0.0373	(-0.0506, -0.0240)
Risk Ratio				
$\hat{\Psi}_{RR}$	0.1064	(0.0208, 0.5434)	0.1460	(0.0542, 0.3930)

* SuperLearner library consists of: SL.mean, SL.glm, SL.gam

2.5 Discussion

The results of our analysis demonstrate that the SEARCH intervention improved infant HIV-free survival after three years among the population of women of reproductive age living with HIV at study baseline (risk difference: 3.89%, 95% CI: 1.51%, 6.26%); relative risk: 1.04, 95% CI: 1.02, 1.07). In addition, the SEARCH intervention also reduced the risk of perinatal HIV transmission in this population (risk difference: −3.52%, 95% CI: −5.18%, −1.86%); relative risk: 0.11, 95% CI: 0.02, 0.75). Because of the design of the SEARCH Study, where the intervention was randomized at the community level, our target causal counterfactual strata total effect parameters are identifiable under assumptions that hold by design and our parameter estimates have a causal interpretation.

Unlike a total effect parameter that is formulated as a contrast in the probabilities of the counterfactual composite outcome of live birth and infant outcomes, as proposed in earlier work,[21] counterfactual strata total effect parameters describe the expected change in the conditional probabilities of counterfactual infant outcomes among counterfactual live births. In SEARCH, the estimated composite outcome of live birth and infant HIV-free survival is higher in the control arm than in the intervention arm (Table 2.2); however, interpretation of this parameter is complicated by the inability to disentangle the intervention effect on infant outcomes and survival among live births from intervention effects on the probability of live birth. Indeed, in the SEARCH Study, our results suggest that the intervention may have changed both 1) the probability of live birth among women living with HIV and 2) the probability of infant outcomes and survival among live births. In contrast, use of a counterfactual strata total effect allows us to quantify the effect of the intervention in terms of conditional probabilities (i.e., the probability of infant HIV-free survival *among* live births). Using this formulation, we find that the SEARCH intervention increased infant HIV-free survival among infants born to women living with HIV compared to the control condition.

In the SEARCH Study, individuals were only at risk for the study outcome of infant HIV-free survival after a live birth occurred; further, the SEARCH intervention may plausibly have affected the types of people who had a live birth and the timing of births through its effects on maternal survival, fertility, and/or behavior. In this work, we also estimated counterfactual stratum direct effect parameters to understand the impact of the SEARCH intervention 1) on infant HIV-free survival absent any intervention-induced changes to the live birth process (i.e., the composition of people who had live births and the timing of live births) and 2) on perinatal HIV transmission excluding any intervention-induced changes to the live birth process and infant survival (i.e., the composition of people who gave birth to infants who survived until the end of study follow-up).

We used three different choices of mediator distributions to estimate these counterfactual stratum direct effects of the impact of the SEARCH Study intervention on infant HIV-free survival and perinatal HIV transmission. The choices of mediator distribution determined the composition of the group (with respect to baseline covariates) among which the outcomes were evaluated and the timing of live births within these groups. Using the estimated mediator distribution in the intervention arm: 1) infant HIV-free survival was estimated to be 1.05 (95% CI: $1.03, 1.08$) times higher under the SEARCH intervention as compared to the control and 2) the risk of perinatal HIV transmission under the SEARCH intervention was estimated to be 0.15 (95% CI: $0.05, 0.39$) times the risk of transmission under the control condition. Under all three choices of mediator distributions, estimates of the impact of the SEARCH intervention were similar.

While the assumption of no unmeasured exposure-outcome confounding is satisfied in the SEARCH Study, the assumption of no unmeasured mediator-outcome confounding is not guaranteed with respect to either outcome. As a result, causal interpretation of the estimates of the direct effect of the SEARCH intervention on infant HIV-free survival and perinatal HIV transmission depends on the plausibility of the absence of unmeasured variables that may confound the mediator-outcome relationship. Potential unmeasured mediator-outcome confounders include measures of time-varying maternal health, which could impact live birth as well as infant health outcomes. We were limited in our ability to adjust for time-varying confounding in this analysis due to limited measurement of interim variables in the control communities. Annual health fairs where additional post-baseline data were collected were held only in intervention communities; health fairs were only conducted at baseline and study end in control communities. In addition, in this analysis, we excluded mother-infant pairs where the infant survived until the end of the study but was missing HIV infection status. If missing infection status was differential by treatment arm, excluding these observations could introduce bias into our estimates.

In this work we illustrate how counterfactual strata total and direct effect parameters can be used to better understand the impact of interventions in perinatal settings, and more generally, when exposures affect the composition of the subgroup that is at risk of study outcomes. Moreover, although rates of perinatal HIV transmission have decreased over the past 25 years, this progress towards eliminating new HIV infections among children has recently slowed as ART coverage among pregnant and breastfeeding people has stagnated.[42]

The results from the SEARCH Study demonstrate that universal HIV test-and-treat policies are effective strategies for improving infant HIV-free survival and reducing perinatal HIV transmission. In the next chapter, we will discuss how counterfactual strata effects can be applied more broadly in epidemiology, using the LEADER trial and revisting the SEARCH Study as motivating examples.

Chapter 3

Counterfactual Strata Effects

3.1 Introduction

In epidemiology, stratification is used to address systematic biases, understand effect heterogeneity, and account for censoring resulting from loss to follow-up or the occurrence of competing events.[62] It may also be used when exposure affects whether an outcome is observed, for example in the study of disease sequelae when exposure affects risk of developing the disease. While stratification on covariates that precede exposure can address concerns of confounding and exposure heterogeneity with respect to pre-exposure covariates, stratification on covariates that mediate the exposure-outcome relationship requires more careful consideration when causal effects are of interest.

Stratifying on, or adjusting for, post-exposure variables to identify causal effects in longitudinal settings where exposures and covariates may be time-varying (i.e., in settings of time-dependent confounding) is well-studied.[63] Stratification may also be used when investigating 1) questions of effect heterogeneity with respect to post-exposure covariates that are affected by exposure, and/or, 2) the effects of an exposure in a particular post-exposure stratum or subgroup. Traditional approaches to these types of questions often rely on stratifying on the observed post-exposure variable. In trial settings where there may be concerns about non-compliance or non-adherence to randomized treatment, per-protocol approaches are often used, which stratify on observed compliance or adherence to exclude the impacts of non-compliance or non-adherence to randomized treatment from the intention-to-treat effect.[64]

Stratification on observed post-exposure variables also often occurs in survival analysis settings, where time-varying survival is affected by the exposure and the hazard is evaluated among those who were observed to survive through each time interval.[20,62] In instances where outcomes are only observed, measured, and/or primarily of interest among a subset of the target population, whether due to non-compliance, loss to follow-up, censoring, or exposure affecting a precondition necessary for the outcome to be defined (e.g., live birth when infant outcomes are of interest), stratification on the observed post-exposure variables may not be

necessary to formulate an identifiable causal estimand that evaluates the causal effect within a post-exposure stratum. Moreover, stratification on the observed post-exposure variable may restrict the types of causal effects that can be identified and estimated in these types of settings.

Alternatively, we can formulate target parameters where we stratify on the counterfactual, instead of the observed, values of post-exposure variables. In this chapter, we discuss how parameters formulated in this way, referred to as counterfactual strata effects, can be used to understand the effect of a treatment within a subset of a larger target population, where exposure affects composition of this subgroup, thus making the (sub)population(s) of interest possibly counterfactual. Counterfactual strata total effects summarize the effect of treatment on outcomes conditional on counterfactual group membership. As described in Chapter 1, total effects formulated in this way exclude any impact of the exposure on the marginal probability of selection into a counterfactual stratum (subgroup), in contrast to total effects formulated in terms of a composite outcome of subgroup membership and the outcome of interest within the subgroup. However, such total effects will include effects of treatment that operate by changing the composition of the individuals selected into each counterfactual stratum under different treatments. Even when exposure was randomized, the composition of the individuals within each counterfactual stratum (i.e., those who would be in the subgroup if treated versus those who would be in the subgroup if assigned to control) may differ on the basis of exposure.

Counterfactual stratum direct effect parameters are designed to evaluate the impact of exposure on study outcomes absent any exposure-induced changes to the types of individuals in the post-exposure strata. Informed by causal estimands from the literature on stochastic mediation,[14,15,16,18,19,35] these direct effect parameters are formulated such that the hypothetical intervention on treatment is static and the hypothetical intervention on mediating variables (subgroup membership) is stochastic. By drawing subgroup membership from some distribution (possibly conditional on baseline covariates), the types of people in the counterfactual stratum of interest with respect to baseline characteristics remains the same, regardless of the hypothetical static intervention on the treatment variable.

In longitudinal settings where subgroup membership may vary over time and the timing of subgroup membership may affect the outcome of interest, counterfactual stratum direct effects can be defined to exclude effects of treatment on the outcome that operate through changes to both 1) the types of people selected into the counterfactual stratum and 2) the timing of selection into a counterfactual stratum. The purpose of the counterfactual stratum direct effects is not necessarily to inform real-world interventions on the mediating variables. Rather, these types of mediated effects provide a better understanding of the mechanisms through which an exposure operates. They can be used to answer different causal questions than those associated with other causal mediation estimands used in similar settings, including controlled,[9] separable,[22,23,65,66] and principal stratum direct effects.[67]

In this chapter, we apply the Causal Roadmap[25] to illustrate how counterfactual strata effects can be applied broadly using the examples of concomitant medication use in the LEADER trial and perinatal HIV transmission in the SEARCH Study. We first illustrate

how counterfactual strata effects can be applied in a point treatment setting with a single mediator representing subgroup membership and then extend this approach to more complex, longitudinal settings. We use a longitudinal targeted maximum likelihood estimator (TMLE) to estimate counterfactual stratum direct effects. We also demonstrate performance of this estimator in simulations with different choices of stochastic interventions on mediating variables.

3.2 Application of the Causal Roadmap to Counterfactual Strata Effects in the Point Treatment Setting

Causal questions

Broadly, counterfactual strata total effects can be used to answer the causal question: what is the effect of the treatment on the outcome of interest within a subgroup if everyone in the target population was assigned to the intervention as compared to the control condition? In contrast to counterfactual strata total effects, counterfactual stratum direct effects do not include effects of the intervention that operate through changing the composition of the counterfactual subgroup. These direct effects can be used to answer the causal question: what would be the effect of the treatment on the outcome among counterfactual group members if everyone were assigned to the intervention as compared to the control condition and if subgroup membership was drawn from the same distribution (possibly conditional on baseline covariates) in both cases? To illustrate how counterfactual strata effects can be applied in the point treatment setting, we use two examples, the LEADER trial and the SEARCH Study. Counterfactual strata effects can be used to better understand the mechanisms by which the intervention operates in both of these studies, though subgroup effects are of interest in each study for different reasons.

The LEADER trial The LEADER trial was a randomized study designed to evaluate the effect of a medication (liraglutide) as compared to placebo on cardiovascular outcomes in a population of individuals with diabetes and high cardiovascular risk.[61] In this trial, differential medication use outside of the primary treatment of interest (i.e., concomitant medication use) by treatment arm following randomization complicates the interpretation of an average treatment effect if concomitant medication use was higher for those assigned to the control as compared to treatment arm and if the use of additional medications lowered cardiovascular risk. For example, individuals in the control arm may be more likely to be prescribed concomitant medications like insulin to manage their glucose levels than individuals randomized to the primary treatment because of the primary treatment's effect on lowering blood glucose levels.[61] Part of the total effect of the primary treatment on cardiovascular outcomes would be mediated by concomitant mediation use, and we may underestimate the

effect of the primary treatment on cardiovascular risk relative to a setting where concomitant medication use was the same under treatment and control conditions.[61]

A counterfactual strata total effect can be formulated to answer the question: what is the effect of the primary treatment on cardiovascular outcomes in the group that did not take additional medications if everyone was assigned to treatment as compared to the placebo? By comparing cardiovascular outcomes in the group without concomitant medication use under treatment versus control, the effect of the primary treatment that operates through changes to the probability of concomitant medication use will be excluded. However, this total effect includes effects of the primary treatment on cardiovascular outcomes that operate through changes to the types of people with post-baseline concomitant medication use. To understand the effect of the primary treatment on cardiovascular outcomes among those without concomitant medication use absent any treatment-induced changes to the types of people in this subgroup, we can use a counterfactual stratum direct effect parameter to ask: what is the effect of primary treatment on cardiovascular outcomes in the group that did not take additional medications if everyone was assigned to treatment as compared to the placebo and the distribution of concomitant use was drawn from the same distribution in both cases?

The SEARCH Study We will also use the example of infant HIV-free survival in the SEARCH Study, a large cluster randomized trial that investigated the effectiveness of a patient-centered, universal HIV test-and-treat intervention on community health, to demonstrate how counterfactual stratum parameters can be used in settings where the outcome of interest may be undefined for a subset of the target population. In the SEARCH trial, 32 communities in rural Kenya and Uganda were randomized to receive either community-wide HIV testing at baseline and the standard of care on the basis of national guidelines or annual universal community-based multi-disease screenings with integrated HIV testing and immediate ART treatment for all who tested positive for HIV.[26,27] Infant HIV-free survival among infants born to women living with HIV at study baseline was a study outcome of interest in this trial. Infant survival and infection status are only defined in the subgroup of live births, a post-intervention variable, and this live birth process may differ on the basis of study arm, as a result of the SEARCH intervention's potential impact on maternal survival, fertility, or behavior. A counterfactual strata total effect could be used to ask: what is the effect of the SEARCH intervention on infant HIV-free survival among live births if everyone was assigned to receive the SEARCH intervention as compared to if everyone was assigned to the control arm of the study?

Because part of the effect of the SEARCH intervention on infant HIV-free survival may operate through changes to the types of people who have live births, we can also define a counterfactual stratum direct effect to better understand the effect of the SEARCH intervention under a hypothetical intervention to keep the types of people, with respect to baseline covariates, who had live births the same, regardless of assignment to the intervention or control condition. More specifically, we can formulate a counterfactual stratum direct effect

parameter to ask: what is the effect of the SEARCH intervention on infant HIV-free survival if everyone was assigned to treatment as compared to the control arm and if the probability of live birth was was drawn from the same distribution regardless of treatment assignment?

Structural causal model (SCM)

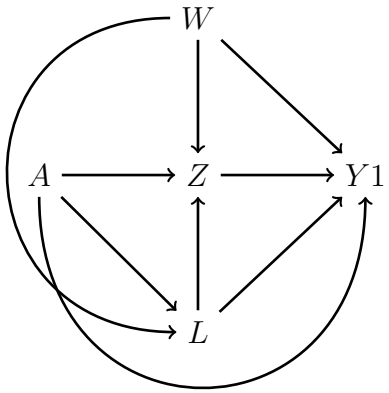


Figure 3.1: Directed Acyclic Graph (DAG) with baseline covariates (W), randomized treatment (A), post-treatment covariates (L), mediator representing indicator of subgroup membership (Z), and outcome ($Y1$).

The structural causal model we use in the point treatment setting consists of baseline covariates (W), a randomized single time point exposure (A), post-treatment covariates (L) that may confound the mediator-outcome relationship, an indicator of subgroup membership (Z), and an outcome ($Y1$) (Figure 3.1). A composite outcome, Y is a composite variable of subgroup membership (Z) and the outcome $Y1$.

$$
\begin{aligned}
W &= f_W(U_W) \\
A &= f_A(U_A) \\
L &= f_L(W, A, U_L) \\
Z &= f_Z(W, A, L, U_Z) \\
Y1 &= f_{Y1}(W, A, L, Z, U_{Y1}) \\
Y &= Z \times Y1
\end{aligned}
$$

The LEADER trial In the LEADER trial (simplifying the study to the point treatment setting), baseline covariates W include concomitant use of medications, such as insulin, at the beginning of the study, treatment A consists of randomization to the primary treatment

(liraglutide), the mediator of subgroup membership Z represents the absence of post-baseline concomitant medication use, and the outcome $Y1$ describes cardiovascular outcomes. The composite outcome Y represents the joint outcome of subgroup membership (no post-baseline concomitant drug use) and the primary cardiovascular outcome $Y1$.

The SEARCH Study In the SEARCH Study, baseline covariates W include maternal age, community-level HIV prevalence, and baseline antiretroviral (ART) therapy use. A is an indicator of randomization to the SEARCH intervention, L is the set of post-baseline covariates such as maternal ART initiation status, Z represents subgroup membership (live birth), and $Y1$ represents infant HIV-free survival. $Y1$ is missing (undefined) in the absence of a live birth. Y is the composite outcome of live birth and infant HIV-free survival.

In the point treatment setting, we address time-varying confounding by variables L but defer until Section 3.3 discussion of additional complexity that arises in the special case of this model where L is an indicator of survival. This type of time-varying confounding is a concern in both the LEADER trial and the SEARCH Study. In Section 3.3, we describe how time-varying survival can be treated as a variable within a vector of mediators Z to ensure that the hypothetical interventions on subgroup membership are well-defined when we formulate our causal estimands.

Causal estimands

Counterfactuals

The counterfactuals we use to define our causal estimands consist of $Z(a)$, $Y1(a)$, and $Y(a)$, which represent counterfactual subgroup membership, the counterfactual outcome of interest, and the counterfactual composite outcome of $Y1$ and subgroup membership, respectively, under a static intervention to set the exposure variable A to some value a. The formulation of our counterfactual stratum direct effects corresponds to a hypothetical stochastic intervention on the mediator Z, in which the value of Z is drawn randomly from some distribution $G(W)$ (for ease of notation, often referred to simply as G), where $G(W) \equiv P(Z = 1|W)$. $Z(G)$ denotes counterfactual subgroup membership under this type of stochastic intervention, and $Y1(a, G)$ describes the counterfactual outcome value under the static intervention to set $A = a$ and the stochastic intervention to draw subgroup membership from the distribution G. $Y(a, G)$ represents the counterfactual composite outcome of subgroup membership under the intervention to set $A = a$ and draw the mediator from $G(W)$. We also use $Y1(a, z)$ to denote the counterfactual outcome if we were to set A to a and the mediator Z to the value z.

Counterfactual strata total effects

The counterfactual strata total effect in the point treatment setting, here represented as a risk difference, summarizes the expected difference in the counterfactual conditional probabilities of the outcome $Y1$ given subgroup membership if everyone was assigned to receive the intervention as compared to the control condition:

$$\begin{aligned}\Psi^F(P_{U,X}) &= E[Y1(a=1)|Z(a=1)=1] - E[Y1(a=0)|Z(a=0)=1] \\ &= \frac{E[Y(a=1)]}{E[Z(a=1)]} - \frac{E[Y(a=0)]}{E[Z(a=0)]}\end{aligned}$$

In the LEADER trial, this type of total effect describes the difference in the expected counterfactual cardiovascular outcome among those who do not take concomitant medications if everyone was assigned to the primary treatment as compared to if everyone was assigned to placebo. In the SEARCH Study, this total effect summarizes the difference in counterfactual infant HIV-free survival among live births if everyone was assigned to receive SEARCH intervention as compared to the control condition. These total effects will include any intervention-induced changes to subgroup membership; the composition of the counterfactual subgroup under treatment as compared to control may not be the same. Despite claims in the literature,[23,68] this is still a causal estimand, just one with a specific interpretation.

Counterfactual stratum direct effects

The counterfactual stratum direct effect describes the expected difference in the counterfactual conditional probabilities of the outcome $Y1$ among subgroup members if everyone was assigned to receive the intervention as compared to the control and if subgroup membership was drawn from the same distribution G under both static interventions on A:

$$\begin{aligned}\Psi^F(P_{U,X}) &= E[Y1(a=1,G)|Z(G)=1] - E[Y1(a=0,G)|Z(G)=1] \\ &= \frac{E[Y(a=1,G)]}{E[Z(G)]} - \frac{E[Y(a=0,G)]}{E[Z(G)]}\end{aligned}$$

In the LEADER trial, this direct effect represents the difference in the expected counterfactual cardiovascular outcome among those with no concomitant medication use if everyone was assigned to the primary treatment as compared to the control and if concomitant drug use was drawn from the same distribution G in either case. This direct effect excludes effects of the treatment on cardiovascular outcome resulting from treatment-induced changes to the types of people, with respect to baseline covariates, with concomitant medication use. In the SEARCH Study, this direct effect summarizes the difference in expected counterfactual infant HIV-free survival among counterfactual live births if everyone had been assigned to

the intervention as compared to the control and if live birth was drawn from the same distribution, irrespective of assignment to treatment or control. This counterfactual stratum direct effect excludes impacts of the SEARCH intervention on infant HIV-free survival that operate through changes to the types of people, in terms of baseline covariates, who have a live birth. Regardless of the hypothetical intervention on the exposure variable, the composition of the counterfactual subgroups, with respect to the baseline covariates W, under treatment and control will remain the same.

The choice of the distribution $G(W)$ from which selection into the subgroup is drawn will determine the composition of the counterfactual stratum in which treatment effects will be evaluated. Possible choices for this distribution include the observed distribution of the mediators (possibly conditional on baseline covariates) in either trial arm, a counterfactual mediator distribution (e.g., the distribution under the control distribution, corresponding to a type of natural direct effect), or another well-supported mediator distribution (e.g., the mediator distribution conditional on baseline covariates in a study population). In this chapter, we consider only hypothetical interventions on the mediating variables where the distribution G is treated as known or is estimated from the observed data; a causal parameter formulated such that the distribution G is the true, unknown mediator distribution under a particular level of treatment requires alternative assumptions for identification and different approaches for estimation than presented in this chapter. If the mediator distribution $G(W)$ is estimated from the observed data, the causal estimand will be a data adaptive parameter.[36] For example, as described in Chapter 2, we could estimate the predicted probabilities of the mediator at each time t conditional on the mediator past and baseline covariates using logistic regression. If the mediator distribution is defined as a counterfactual distribution, additional identification assumptions, beyond those described below will be required (as discussed in greater detail in Chapter 1).

Observed data

The observed data consist of baseline covariates, an indicator of treatment, post-treatment mediator-outcome confounders, mediator, and the outcome variables.

$$O = (W, A, L, Z, Y1, Y)$$

Assuming random assignment of treatment A, the statistical model, $\mathcal{M}$, is semi-parametric.

Identification

The counterfactual strata total effect is identified under assumptions of no unmeasured confounding of the exposure effect on the composite variable of subgroup membership and the outcome of interest $Y1$ (3.1); and a positive probability of receiving each treatment level of interest, a, irrespective of the values of baseline covariates (3.2).

$$Y1(a), Z(a) \perp A|W \tag{3.1}$$

$$\min_{a \in \mathcal{A}} P_0(A = a|W) > 0 \; \forall \; w \text{ for which } P_0(W = w) > 0. \tag{3.2}$$

The counterfactual stratum direct effect will be identified under assumptions of conditional independence of the exposure A and the counterfactual outcome $Y1(a)$ (3.3), conditional independence of the mediator Z and the counterfactual outcome $Y1(a)$ (3.4), a positive probability of each treatment a within all values of the baseline covariates (3.5), and a positive probability of subgroup membership (i.e., Z=1) for any values of W that occur with positive probability in the sample and for which probability of subgroup membership is greater than zero under the user-specified mediator distribution $G(W)$, within all values of the baseline and post-treatment covariates that occur with some positive probability (3.6).

While the randomization assumptions that allow for identification of the counterfactual stratum direct effect are the same as those used to identify a controlled direct effect, the positivity assumption is generally weaker for the counterfactual stratum direct effect. More specifically, if there are certain strata of the baseline covariates where the probability of subgroup membership ($P(Z = 1|W = w)$) is equal to zero, the controlled direct effect will not be identified. However, if the probability of subgroup membership under the stochastic intervention to draw the value of the mediator Z from the distribution G is also equal to zero for all strata of covariates where the probability of subgroup membership is equal to zero in the observed data, the positivity assumption for the counterfactual stratum direct effect will be satisfied and the counterfactual stratum direct effect may still be identifiable. In the special case where the stochastic intervention on the mediator is to set the value of Z equal to 1, the counterfactual stratum direct effect will be equal to a controlled direct effect, and the positivity assumptions used to identify both parameters will be the same.

$$Y1(a, z) \perp A|W \tag{3.3}$$

$$Y1(a, z) \perp Z|L, A, W \tag{3.4}$$

$$\min_{a \in \mathcal{A}} P_0(A = a|W) > 0 \; \forall \; w \text{ for which } P_0(W = w) > 0 \tag{3.5}$$

$$P_0(Z = 1|L, A, W) > 0 \; \forall \; l, a, w \tag{3.6}$$
$$\text{for which } (P_0(L = l, A = a, W = w) > 0 \text{ and } G(Z = 1|W = w) > 0)$$

The assumption of conditional independence of the exposure and the counterfactual outcome $Y1(a)$ that is used to identify both the counterfactual strata total effect and the counterfactual stratum direct effect will be satisfied by design in a randomized trial, like the LEADER and SEARCH trials. In a randomized study, the assumption of conditional independence of the exposure and subgroup membership $Z(a)$ that is used to identify the counterfactual strata total effect will also be satisfied by design; this assumption is not needed for identification of the counterfactual stratum direct effect parameter.

The counterfactual stratum direct effect additionally requires conditional independence of the mediator Z and the counterfactual outcome $Y1(a, z)$. This assumption is not guaranteed, even if treatment A was randomized. This assumption will be satisfied if the post-treatment covariates L and baseline covariates W block all back door paths between the mediator and outcome. In the LEADER trial, to satisfy this assumption, we would need to account for confounding of the relationship between concomitant drug use and cardiovascular risk. In the SEARCH study, this randomization assumption will be satisfied if we believe the baseline covariates W and post-treatment covariates L control for confounding between live birth and infant HIV-free survival.

Estimand

We can express the counterfactual strata total effect parameter as a function of the observed data if the assumptions for identification of the parameter are satisfied:

$$\Psi(P_0) = \frac{E_0\left[E_0(Y|A=1,W)\right]}{E_0\left[E_0(Z|A=1,W)\right]} - \frac{E_0\left[E_0(Y|A=0,W)\right]}{E_0\left[E_0(Z|A=0,W)\right]}$$

If the assumptions to identify the counterfactual stratum direct effect parameter are satisfied, then this mediated effect is identified as the statistical estimand:

$$\Psi(P_0) = \frac{\begin{array}{r}\sum_{w,l}(E_0[Y|Z=1,L=l,A=1,W=w]\times G(Z=1|W=w)\times \\ P_{0,L}(L=l|A=1,W=w)\times P_{0,W}(W=w))\end{array}}{\sum_w G(Z=1|w)\times P_{0,W}(w)} - \frac{\begin{array}{r}\sum_{w,l}(E_0[Y|Z=1,L=l,A=0,W=w]\times G(Z=1|W=w)\times \\ P_{0,L}(L=l|A=0,W=w)\times P_{0,W}(W=w))\end{array}}{\sum_w G(Z=1|w)\times P_{0,W}(w)}$$

Estimation

Both counterfactual strata effect parameters can be estimated using TMLE.[55] We use $\hat{\psi}$ to denote the estimator of the parameter Ψ. We can estimate the counterfactual strata total effect using TMLE to estimate the expected outcome Y, $\hat{\psi}_{Y,a}$, and the probability of subgroup membership Z, $\hat{\psi}_{Z,a}$, under each treatment of interest $a \in \mathcal{A}$. We can estimate the risk difference by taking the difference of the ratios of the estimated expected value of Y and the expected value of Z for each treatment level or a risk ratio by taking the ratio of the ratios of these two quantities. TMLE uses estimates of the conditional expectation and the treatment mechanism (i.e., probability of treatment) to estimate each parameter of interest. When the treatment mechanism is known, as in a randomized trial, TMLE is a consistent estimator (i.e., as sample size increases, the parameter estimate will approach the true parameter value).[55] If we also specify a consistent estimator of the conditional

expectation, then TMLE will also be efficient (i.e., achieve the lowest possible asymptotic variance among reasonable estimators).[55]

$$\hat{\psi}_{CSTE} = \frac{\hat{\psi}_{Y,a=1}}{\hat{\psi}_{Z,a=1}} - \frac{\hat{\psi}_{Y,a=0}}{\hat{\psi}_{Z,a=0}}$$

The counterfactual stratum direct effect can be estimated using longitudinal TMLE[69,70] to estimate $\hat{\psi}_{Y,a,G}$, which denotes the expected composite outcome Y under each treatment level of interest a and the known stochastic intervention to draw mediator values from the distribution G.[16,17,61] This TMLE requires estimation of 1) the conditional distributions of the covariates at each time given the past, 2) the conditional probability of the mediator, and 3) the treatment mechanism. The denominator of this direct effect estimator, $\hat{\psi}_{Z,G}$, denotes the expected mean of the probability of subgroup membership if subgroup membership was drawn from the distribution G. It is estimated by taking the mean probability of subgroup membership under the distribution G, $\hat{\psi}_{Z,G} = \frac{1}{n}\sum_{i=1}^{n} G(Z = 1|W_i)$.

$$\hat{\psi}_{CSDE} = \frac{\hat{\psi}_{Y,a=1,G}}{\hat{\psi}_{Z,G}} - \frac{\hat{\psi}_{Y,a=0,G}}{\hat{\psi}_{Z,G}}$$

Influence-curved based inference for estimators of the counterfactual strata total effect and counterfactual stratum direct effect can be obtained using the functional delta method.

3.3 Extensions to Longitudinal Settings

Causal questions

In this section, we will apply the Causal Roadmap[25] to a general longitudinal setting that can accommodate time-varying covariates and multiple mediators. We revisit the example of infant HIV-free survival in the SEARCH Study from Section 3.2 but with additional consideration of the longitudinal nature of the live birth process and the impacts of an intervention that operate through changes to the timing of subgroup selection.

Counterfactual strata total effect

In settings where we have time-varying covariates and (possibly multiple) mediators measured at more than one time point, counterfactual strata total effects can be used to answer the question: what is the effect of a treatment on an outcome within subgroup members if everyone was assigned to the intervention as compared to the control? If the mediator that represents selection into the counterfactual subgroup is longitudinal, then this counterfactual strata total effect will include the impact of the intervention on changes to the timing of selection into the subgroup.

The SEARCH Study In SEARCH, a counterfactual strata total effect could be used to ask: what is the effect of the SEARCH intervention on infant HIV-free survival among live births that occurred by the end of follow-up if everyone was assigned to the SEARCH intervention as compared to the control condition? As discussed in Section 3.2, this total effect will include the effect of the SEARCH intervention that operates by changing the types of women who had a live birth in the intervention arm as compared to the control arm of the study. In addition, the SEARCH intervention may have also operated through intervention-induced changes to the timing of live births across the follow-up period (and the interaction of the two?).

If the SEARCH intervention shifted the distribution of time-to-live birth under the intervention as compared to the control condition, the cumulative amount of time-at-risk would vary as a function of exposure. Moreover, if the risk of perinatal HIV transmission varied across the antepartum, intrapartum, and postpartum periods and/or the risk of infant mortality varied over the period between birth and the end of follow-up, then the total effect would also include any mechanisms by which the SEARCH intervention impacted infant HIV-free survival through differential changes in the distribution of time-at-risk in each of these critical periods between the intervention and control arms.

Counterfactual stratum direct effect

Like in the point treatment setting, counterfactual stratum direct effects do not include effects of the intervention that operate through changing the composition of the group that is selected into the counterfactual subgroup in longitudinal settings. Moreover, these direct effects also exclude any effects of the intervention operating through changes to the timing of selection of different types of people, with respect to baseline covariates, into the counterfactual stratum of interest. Counterfactual stratum direct effects allow us to answer questions about the effect of a treatment on outcomes among counterfactual group members if everyone was assigned to the intervention as compared to the control and if the timing of group membership (and possibly other mediating variables) was drawn from the same distribution.

The SEARCH Study Counterfactual stratum direct effects can be used to understand the effect of the SEARCH intervention on infant HIV-free survival absent intervention-induced changes to the types of women who had a live birth and the timing of births over the follow-up period. They ask: what is the effect of the SEARCH intervention on infant HIV-free survival among live births if the "live birth process" (i.e., maternal survival and live birth over time) was drawn from the same distribution under the treatment and control conditions?

Structural causal model

We use a longitudinal structural causal model to represent a data generating process from baseline through the final time point, $\tau + 1$. This SCM consists of baseline covariates

(L_0), a randomized single time point exposure (A), time-varying covariates (L_t) that include possible mediator-outcome confounders, time-varying mediators including (but not limited to) an indicator of subgroup membership at each time t (Z_t) from time $1, \ldots, \tau$, and an outcome ($Y1_{\tau+1}$). For time-varying covariates, V, the vector $(V_1, \ldots, V_\tau)$ is denoted by $\bar{V}_t$. We use $Y_{\tau+1}$ to denote the composite outcome of subgroup membership at time τ and the outcome of interest at the final time point $Y1_{\tau+1}$. In a longitudinal setting, the composite outcome of interest may consist of an outcome $Y1_{\tau+1}$ and subgroup membership a time point earlier than τ; for simplicity throughout, we will focus on the composite outcome of $Y1_{\tau+1}$ and subgroup membership at time τ. Z_0 is equal to the empty set.

The variable Z_t may represent only of the variable of subgroup membership at each time $1, ..., \tau$, or, it may describe a vector of variables (e.g., $Z_t = (Z1_t, Z_t^*)$) when it is of interest to jointly intervene upon more than one mechanism by which treatment affects outcome. We will use Z_t^* to represent the variable describing subgroup membership within the vector of all mediating variables Z_t. It may be necessary to define Z_t as a vector consisting of more than one variable to ensure that the hypothetical intervention on subgroup selection is well-defined.[14] For example, in the SEARCH Study, where maternal survival confounds the relationship between live birth and infant HIV-free survival, intervening jointly upon maternal survival and live birth accounts for time-varying confounding by maternal survival while avoiding a hypothetical stochastic intervention on live birth that may be inconsistent with the value of maternal survival, as described in Chapter 2. We may also include other time-varying covariates in Z_t, instead of in the set L_t, if we are interested in asking a question about the direct effect of an exposure on outcomes within a subgroup while keeping the composition of the subgroup the same, with respect to these additional time-varying covariates as well as baseline covariates, under different hypothetical static interventions on the exposure A.

$$
\begin{aligned}
L_0 &= f_{L_0}(U_{L_0}) \\
A &= f_A(U_A) \\
L_t &= f_{L_t}(L_0, A, \bar{L}_{t-1}, \bar{Z}_{t-1}, U_{L_t}) \\
Z_t &= f_{Z_t}(L_0, A, \bar{L}_t, \bar{Z}_{t-1}, U_{Z_t}) \\
Y1_{\tau+1} &= f_{Y1_{\tau+1}}(L_0, A, \bar{L}_\tau, \bar{Z}_\tau, U_{Y1_{\tau+1}}) \\
Y_{\tau+1} &= Z_\tau^* \times Y1_{\tau+1}
\end{aligned}
$$

The SEARCH Study In SEARCH, the baseline covariates L_0 consist of age at baseline, community-specific HIV prevalence, and ART initiation status at the beginning of the study, A is an indicator representing receipt of the SEARCH intervention, the time-varying covariates L_t consist of maternal ART initiation status at time t and infant vital status at time t, the mediators Z_t consist of maternal survival ($Z1_t$) and live birth (Z_t^*) at time t, and the outcome at the final time point $Y1_{\tau+1}$ represents infant HIV-free survival at time $\tau + 1$.

$Y1_{\tau+1}$ will be missing for observations for which no live birth occurred by the end of the study. $Y_{\tau+1}$ denotes the composite outcome of live birth by time τ (Z^*_τ) and infant HIV-free survival at time $\tau+1$ ($Y1_{\tau+1}$).

If we were interested in asking a different research question about the effect of the SEARCH intervention on infant HIV-free survival among live births keeping the composition of the live births the same with respect to the baseline maternal age, community-specific HIV prevalence, and baseline as well as time-varying maternal ART initiation status, we could alternatively define our vector of mediators at each time t (Z_t) to consist of maternal ART initiation status, maternal survival, and live birth at time t. In addition to excluding impacts of the SEARCH intervention that operate through changes to the timing and occurrence of maternal survival and live birth, this direct effect would also exclude effects of the intervention that operate through changes to the timing of maternal ART initiation. Formulating a direct effect parameter in this way would keep the composition of the subpopulation of live births the same with respect to the baseline covariates, time-varying maternal survival, time-varying maternal ART initiation status, and the timing of live births.

Causal estimands

Counterfactuals

We use $Z^*_\tau(a)$, $Y1_{\tau+1}(a)$, and $Y_{\tau+1}(a)$, which represent counterfactual subgroup membership, the counterfactual outcome of interest, and the counterfactual composite outcome of $Y1$ at time $\tau+1$ and subgroup membership at time τ, respectively, under a static intervention to set the exposure variable A to some value a. $\bar{G}(L_0)$ denotes the distribution, which may be conditional on baseline covariates, from which mediator values are drawn under the hypothetical intervention on mediating variables, where $\bar{G}(\bar{Z}_\tau|L_0) \equiv P(\bar{Z}_\tau = \bar{z}_\tau|L_0) = \prod_{t=1}^{\tau} G_t(Z_t|\bar{Z}_{t-1}, L_0)$. $G_t(Z_t|\bar{Z}_{t-1}, L_0)$ is used to represent the probability distribution of the mediator at time t given the mediator past and baseline covariates, from time 1 through time τ. For ease of notation, we use $\bar{G}$ to refer to the conditional distribution $\bar{G}(L_0)$.

$Z^*_\tau(\bar{G})$ describes counterfactual subgroup membership under this type of stochastic intervention, and $Y1_{\tau+1}(a, \bar{G})$ represents the counterfactual outcome value when we hypothetically set $A = a$ and draw subgroup membership from the distribution $\bar{G}$. $Y_{\tau+1}(a, \bar{G})$ denotes the counterfactual composite outcome of subgroup membership under the static intervention to set $A = a$ and the stochastic intervention to draw the mediators $\bar{Z}_\tau$ from $\bar{G}$, while $Y1_{\tau+1}(a, \bar{z}_\tau)$ represents the counterfactual outcome setting $A = a$ and $\bar{Z}_\tau = \bar{z}_\tau$.

Counterfactual Strata Total Effect

A counterfactual strata total effect summarizes the expected difference in the outcome of interest among counterfactual subgroup members if everyone was assigned to the intervention as compared to the control condition:

$$\begin{aligned}\Psi^F(P_{U,X}) &= E[Y1_{\tau+1}(a=1)|Z^*_\tau(a=1)=1] - E[Y1_{\tau+1}(a=0)|Z^*_\tau(a=0)=1] \\ &= \frac{E[Y_{\tau+1}(a=1)]}{E[Z^*_\tau(a=1)]} - \frac{E[Y_{\tau+1}(a=0)]}{E[Z^*_\tau(a=0)]}\end{aligned}$$

In the SEARCH Study, this type of total effect can be used to represent the difference in infant HIV-free survival among counterfactual live births if all women of reproductive age living with HIV at study baseline were assigned to receive the SEARCH intervention as compared to if everyone was assigned to the control.

Counterfactual Stratum Direct Effect

A counterfactual stratum direct effect represents the expected difference in the outcome of interest among counterfactual group members if the mediators $\bar{Z}_\tau$, which include but are not limited to subgroup membership over time, were drawn from the same distribution under each static intervention in the exposure variable A. We use $\bar{G}$ to denote the distribution that is possibly conditional on some or all of the baseline covariates, L_0, and from which the values of the mediating variables, $\bar{Z}_\tau$, are drawn.

$$\begin{aligned}\Psi^F(P_{U,X}) &= E[Y1_{\tau+1}(a=1,\bar{G})|Z^*_\tau(\bar{G})=1] - E[Y1_{\tau+1}(a=0,\bar{G})|Z^*_\tau(\bar{G})=1] \\ &= \frac{E[Y_{\tau+1}(a=1,\bar{G})]}{E[Z^*_\tau(\bar{G})]} - \frac{E[Y_{\tau+1}(a=0,\bar{G})]}{E[Z^*_\tau(\bar{G})]}\end{aligned}$$

We can formulate a counterfactual stratum direct effect to summarize the difference in infant HIV-free survival among counterfactual live births if all women of reproductive age living with HIV at study baseline were assigned to receive the SEARCH intervention versus the comparator condition and if the live birth process (maternal survival and the timing of live births) was drawn from the same distribution, $\bar{G}$. This parameter excludes any effects of the SEARCH intervention that operate through changes to the live birth process (i.e., maternal survival and the timing of live births) and keeps the types of women who had live births, with respect to baseline covariates and time-varying maternal survival, the same under the intervention and control conditions.

Observed data

The observed data consist of baseline covariates, an indicator of treatment, time-varying covariates and mediators, and the final composite outcome of interest.

$$O = (L_0, A, \bar{L}_t, \bar{Z}_t, Y1_{\tau+1}, Y_{\tau+1})$$

We assume that the treatment A was randomly assigned, therefore the treatment mechanism is known, and the statistical model, $\mathcal{M}$, is semi-parametric.

Identification

We can identify the counterfactual strata total effect under the randomization assumption of no unmeasured exposure-composite outcome confounding and the assumption of a positive probability of each treatment of interest, a, within all values of baseline covariates.

$$Y1_{\tau+1}(a), Z_\tau(a) \perp A|L_0 \tag{3.7}$$

$$\min_{a\in\mathcal{A}} P_0(A=a|L_0) > 0 \; \forall \; l_0 \text{ for which } P_0(L_0=l_0) > 0 \tag{3.8}$$

The randomization assumptions correspond to conditional independence of treatment (A) with the counterfactual outcome $Y1_{\tau+1}(a)$ (3.7) and conditional independence of treatment (A) and counterfactual group membership at the final time point ($Z_\tau(a)$) (3.8). When treatment is randomized, these assumptions for identification of the counterfactual strata total effect will be satisfied by design.

Under the randomization assumptions of no unmeasured exposure-outcome (3.9) and mediator-outcome confounding (3.10) and the associated positivity assumptions, the counterfactual stratum direct effect will also be identified. In the setting of a randomized trial, the first randomization assumption of no exposure-outcome confounding will be satisfied. However, randomization of treatment does not guarantee the assumption of an absence of mediator-outcome confounding used to identify the counterfactual stratum direct effect. If we believe that the baseline covariates (L_0) and time-varying covariates (L_t) sufficiently control for confounding between the mediators (Z_t) and the outcome ($Y1_\tau$), then this second randomization assumption will be satisfied. Unlike for identification of the counterfactual strata total effect, identification of the counterfactual stratum direct effect is possible in the presence of unmeasured confounders of treatment and group membership (Z^*_τ) at the final time point because the mediator Z^*_τ is an intervention node in this direct effect parameter formulation.[55]

$$Y1_{\tau+1}(a, \bar{z}_\tau) \perp A|L_0 \tag{3.9}$$

$$Y1_{\tau+1}(a, \bar{z}_\tau) \perp Z_t|\bar{L}_t, \bar{Z}_{t-1}, A, L_0 \; \forall \; t \text{ in } 1, \ldots, \tau \tag{3.10}$$

$$\min_{a\in\mathcal{A}} P_0(A=a|L_0) > 0 \; \forall \; l_0 \text{ for which } P_0(L_0=l_0) > 0 \tag{3.11}$$

$$\min_{z_t\in\mathcal{Z}_t} P_0(Z_t=z_t|\bar{L}_t, \bar{Z}_{t-1}, A, L_0) > 0 \; \forall \; \bar{l}_t, \bar{z}_{t-1}, a, l_0 \text{ for which} \tag{3.12}$$

$$(P_0(\bar{L}_t=\bar{l}_t, \bar{Z}_{t-1}=\bar{z}_{t-1}, A=a, L_0=l_0) > 0 \text{ and } G(Z_t=z_t|\bar{Z}_{t-1}=\bar{z}_{t-1}, L_0=l_0) > 0)$$

The first positivity assumption (3.11) states that there is a positive probability of each treatment level a in $\mathcal{A}$ within all values of the baseline covariates L_0. This is the same positivity assumption used in the point treatment setting. The second positivity assumption (3.12) states that there must be a positive probability of the mediator values that are possible under the hypothetical intervention on the mediators $\bar{G}(L_0)$ within all values of the time-varying past.

Estimands

If the assumptions to identify the counterfactual strata total effect are satisfied, the parameter is identified as the following statistical estimand:

$$\Psi(P_0) = \frac{E_0\left[E_0(Y_{\tau+1}|A=1,L_0)\right]}{E_0\left[E_0(Z^*_\tau|A=1,L_0)\right]} - \frac{E_0\left[E_0(Y_{\tau+1}|A=0,L_0)\right]}{E_0\left[E_0(Z^*_\tau|A=0,L_0)\right]}$$

If the assumptions to identify the counterfactual stratum direct effect are satisfied, then this mediated effect is identified as the statistical estimand:

$$\Psi(P_0) = \frac{\begin{array}{c}\sum\limits_{l_0,\bar{l}_\tau,\bar{z}_\tau}(E_0[Y_{\tau+1}|\bar{Z}_\tau=\bar{z}_\tau,\bar{L}_t=\bar{l}_t,A=1,L_0=l_0]\times\bar{G}(\bar{Z}_\tau=\bar{z}_\tau|L_0=l_0)\times \\ P_{0,L_0}(L_0=l_0)\prod\limits_{t=1}^{\tau}P_{0,L_t}(L_t=l_t|\bar{Z}_{t-1}=\bar{z}_{t-1},\bar{L}_{t-1}=\bar{l}_{t-1},A=1,L_0=l_0))\end{array}}{\sum\limits_{l_0,\bar{z}^-_\tau}G_\tau(Z^*_\tau=1|\bar{Z}^-_\tau=\bar{z}^-_\tau,L_0=l_0)\times\bar{G}(\bar{Z}^-_\tau=\bar{z}^-_\tau|L_0=l_0)\times P_{0,L_0}(l_0)}$$

$$- \frac{\begin{array}{c}\sum\limits_{l_0,\bar{l}_\tau,\bar{z}_\tau}(E_0[Y_{\tau+1}|\bar{Z}_\tau=\bar{z}_\tau,\bar{L}_t=\bar{l}_t,A=0,L_0=l_0]\times\bar{G}(\bar{Z}_\tau=\bar{z}_\tau|L_0=l_0)\times \\ P_{0,L_0}(L_0=l_0)\prod\limits_{t=1}^{\tau}P_{0,L_t}(L_t=l_t|\bar{Z}_{t-1}=\bar{z}_{t-1},\bar{L}_{t-1}=\bar{l}_{t-1},A=0,L_0=l_0))\end{array}}{\sum\limits_{l_0,\bar{z}^-_\tau}G_\tau(Z^*_\tau=1|\bar{Z}^-_\tau=\bar{z}^-_\tau,L_0=l_0)\times\bar{G}(\bar{Z}^-_\tau=\bar{z}^-_\tau|L_0=l_0)\times P_{0,L_0}(l_0)}$$

We use $\bar{Z}^-_\tau$ to denote the vector of mediators $\bar{Z}_\tau$, excluding the indicator of subgroup membership at time τ, Z^*_τ. Note that we use this notation because subgroup membership $Z^*_\tau = 1$ may be compatible with multiple possible mediator vectors, including multiple possible values of other mediators (other than the final indicator of subgroup membership) at time τ.

Estimation

As in the point treatment setting, we can estimate the counterfactual strata total effect using TMLE to estimate the expected outcome $Y_{\tau+1}$, $\hat{\psi}_{Y_{\tau+1}a}$, and the probability of subgroup membership Z^*_τ, $\hat{\psi}_{Z^*_\tau,a}$, under each treatment of interest $a \in \mathcal{A}$. We can contrast the ratios of the estimated expected value of the composite outcome $Y_{\tau+1}$ and the estimated expected value of subgroup membership by time τ, $Z^*\tau$, to formulate either a risk ratio or a risk difference.

$$\hat{\psi}_{CSTE} = \frac{\hat{\psi}_{Y_{\tau+1},a=1}}{\hat{\psi}_{Z^*_\tau,a=1}} - \frac{\hat{\psi}_{Y_{\tau+1},a=0}}{\hat{\psi}_{Z^*_\tau,a=0}}$$

The counterfactual stratum direct effect can be estimated using longitudinal TMLE.[69,70] We estimate, $\hat{\psi}_{Y_{\tau+1}a,\bar{G}}$, which represents the expected composite outcome under each treatment of interest $a \in \mathcal{A}$ and the stochastic intervention to draw the mediator values $\bar{Z}_\tau$ from the distribution $\bar{G}$.[16,17,61] As described in Section 3.2, to estimate the expected composite outcome we use the estimated 1) conditional distributions of the covariates at each time given the past, 2) conditional probabilities of the mediators, and 3) treatment mechanism. In a randomized trial, the treatment mechanism is known and does not need to be estimated, though estimation of this quantity can improve precision.[55,69] $\hat{\psi}_{Z^*_\tau,\bar{G}}$, denotes the expected mean of the probability of subgroup membership at time τ if all mediating variables $\bar{Z}_\tau$, which include but are not limited to subgroup membership over time, were drawn from the distribution $\bar{G}$. The mean probability of subgroup membership under the distribution $\bar{G}$ is represented by $\hat{\psi}_{Z^*_\tau,\bar{G}}$, which is equal to $\frac{1}{n}\sum_{i=1}^{n}\sum_{\bar{z}_\tau^-} G(Z^*_\tau = 1|\bar{Z}_\tau = \bar{z}_\tau, L_{0,i}) \times \bar{G}(\bar{Z}_\tau = \bar{z}_\tau|L_{0,i})$.

$$\hat{\psi}_{CSDE} = \frac{\hat{\psi}_{Y_{\tau+1},a=1,\bar{G}}}{\hat{\psi}_{Z^*_\tau,\bar{G}}} - \frac{\hat{\psi}_{Y_{\tau+1},a=0,\bar{G}}}{\hat{\psi}_{Z^*_\tau,\bar{G}}}$$

3.4 Simulations

We performed simulations to evaluate the performance of the longitudinal TMLE for the estimation of counterfactual stratum direct effects using a data generating process motivated by the example of infant HIV-free survival in the SEARCH Study. Longitudinal TMLE was implemented 1) using main terms regressions to estimate the conditional expectations and 2) using Super Learner with a small library of algorithms (`SL.mean` for simple means, `SL.glm` for main terms linear models, and `SL.gam`[58,59,60] for generalized additive models). We considered two types of counterfactual stratum direct effects for three different values of τ. The first type of direct effect parameter was formulated such that the stochastic intervention on the mediating variables (i.e., maternal survival and live birth) involved drawing the mediator values from the observed distribution of the mediating variables at each time t conditional on baseline covariates and the past mediators through time $t-1$ in a large sample from the data generating process. We regressed each mediator at every time t on the mediator past and baseline covariates to estimate the predicted probabilities of the mediator at each time point; we refer to this direct effect as the counterfactual stratum direct effect that uses a well-supported intervention on the mediators. The second direct effect was formulated such that the stochastic intervention on the mediating variables enforced maternal survival and live birth at all time points (i.e., a controlled direct effect). We ran 500 simulations with a sample size of $n = 5000$ for each setting to evaluate the bias, variance, mean squared error (MSE), and 95% confidence interval coverage of the TMLE estimators of the risk difference and risk ratio.

Data generating process

The data generating process for the simulations is motivated by the example of infant HIV-free survival in the SEARCH Study. The variables in this model consist of two binary baseline covariates $L1_0$ and $L2_0$, randomized treatment A, a time-to-event covariate $L1_t$ representing maternal ART initiation status over time, and a time-varying covariate $Y1_t$ representing infant survival over time. $Y2_t$ represents time-varying infant HIV infection status, which is only observed at time $\tau + 1$. L_t is equal to $\{L1_t, Y1_t\}$, where $Y1_0$, $Y1_1$, $Y2_0$, and $Y2_1$ are empty sets. The vector of mediators Z_t consists of two variables $Z1_t$ and $Z2_t$, representing maternal mortality and live birth at time t. $Y1_t$ is missing if a live birth has not occurred by time t; $Y2_{\tau+1}$ is only non-missing if a live birth has occurred by time τ ($Z_\tau = 1$) and the infant is alive at time $\tau + 1$ ($Y1_{\tau+1} = 1$).

$$
\begin{aligned}
L1_0 &\sim Bernoulli(0.25)\\
L2_0 &\sim Bernoulli(0.4)\\
A &\sim Bernoulli(0.5)\\
L1_t &\sim \begin{cases} 1, & \text{if } L_{t-1}=1\\ Bernoulli(\text{expit}(-0.5-0.1\times L2_0+0.5\times A- \\ \qquad 0.05\times \bar{Z}1_{t-1}+0.05\times \bar{Z}2_{t-1}- \\ \qquad \sum_{i=1}^{t-1} 1(Y1_i\neq NA)\times Y1_i)), & \text{otherwise}\end{cases}\\
Z1_t &\sim \begin{cases} 1, & \text{if } Z1_{t-1}=1\\ Bernoulli(\text{expit}(-2\times L1_0-1\times L2_0- \\ \qquad 0.5\times A-2\times \bar{L}_t- \\ \qquad 0.1\sum_{i=1}^{t-1} 1(Y1_i\neq NA)\times Y1_i)), & \text{otherwise}\end{cases}\\
Z2_t &\sim \begin{cases} 1, & \text{if } Z2_{t-1}=1\\ Bernoulli(\text{expit}(-2.5-0.1\times L1_0+0.5\times L2_0+ \\ \qquad 0.5\times A-0.1\times \bar{L}_t)), & \text{otherwise}\end{cases}\\
Y1_{t+1} &\sim \begin{cases} NA, & \text{if } Z2_t=0\\ 1, & \text{if } Y1_t=1\\ Bernoulli(\text{expit}(-L1_0+L2_0+0.5\times A-\bar{L}_t+ \\ \qquad 0.25\sum_{i=1}^{t} 1(Y2_i\neq NA)\times Y2_i)), & \text{otherwise}\end{cases}\\
Y2_{t+1} &\sim \begin{cases} NA, & \text{if } Y1_{t+1}=1\\ 1, & \text{if } Y2_t=1\\ Bernoulli(\text{expit}(-0.5-L1_0+0.1\times L1_0-A- \\ \qquad \bar{L}_t-\bar{Z}1_t)), & \text{otherwise}\end{cases}
\end{aligned}
$$

Results

The results of the simulations that evaluated the performance of the direct effect estimators are presented below for both the risk difference and the risk ratio. The true values of the parameter in each setting are denoted by Ψ in each table. While performance metrics were similar for both choices of stochastic interventions on the mediators in the setting where $\tau = 1$, the mean squared error (MSE) was about two times greater for the risk difference, and about four times greater for the risk ratio, under a stochastic intervention correspond-

ing to a controlled direct effect compared to a stochastic intervention using a well-supported user-supplied mediator distribution for $\tau = 3$.

We may expect to see a larger difference in estimator performance between the estimators of the direct effect parameter formulated using a well-supported intervention on mediating variables as compared to the direct effect parameter corresponding to a controlled direct effect if positivity was a greater concern in this data generating process (i.e., if the probability of either mediator, $Z1_t$ or $Z2_t$, was low in particular strata of baseline covariates). For both stochastic interventions on mediating variables, confidence interval coverage decreased as the number of time points increased.

Table 3.1: Simulation results for the counterfactual stratum direct effect (risk difference), using a well-supported intervention on the mediators.

	Mean	Bias	Variance	MSE	95% CI Coverage
$\tau = 1$ ($\Psi = 0.3465$)					
Main terms regression	0.3478	0.0014	0.0016	0.0016	0.960
Super Learner	0.3404	-0.0061	0.0016	0.0017	0.956
$\tau = 2$ ($\Psi = 0.3612$)					
Main terms regression	0.3427	-0.0185	0.0011	0.0014	0.924
Super Learner	0.3390	-0.0221	0.0010	0.0015	0.930
$\tau = 3$ ($\Psi = 0.3666$)					
Main terms regression	0.3428	-0.0238	0.0010	0.0016	0.916
Super Learner	0.3417	-0.0250	0.0010	0.0016	0.900

Table 3.2: Simulation results for the counterfactual stratum direct effect (risk difference), corresponding to a controlled direct effect.

	Mean	Bias	Variance	MSE	95% CI Coverage
$\tau = 1$ ($\Psi = 0.3521$)					
Main terms regression	0.3582	0.0061	0.0016	0.0016	0.962
Super Learner	0.3480	-0.0041	0.0017	0.0018	0.956
$\tau = 2$ ($\Psi = 0.3979$)					
Main terms regression	0.3682	-0.0297	0.0018	0.0027	0.928
Super Learner	0.3669	-0.0310	0.0019	0.0028	0.920
$\tau = 3$ ($\Psi = 0.4164$)					
Main terms regression	0.3838	-0.0326	0.0019	0.0030	0.896
Super Learner	0.3841	-0.0324	0.0021	0.0032	0.894

Table 3.3: Simulation results for the counterfactual stratum direct effect (risk ratio), using a well-supported intervention on the mediators.

	Mean	Bias	Variance	MSE	95% CI Coverage
$\tau = 1$ ($\Psi = 1.8408$)					
Main terms regression	1.8586	0.0177	0.0261	0.0264	0.966
Super Learner	1.8292	-0.0116	0.0247	0.0248	0.954
$\tau = 2$ ($\Psi = 1.8306$)					
Main terms regression	1.7788	-0.0518	0.0143	0.0170	0.932
Super Learner	1.7652	-0.0654	0.0127	0.0170	0.920
$\tau = 3$ ($\Psi = 1.8132$)					
Main terms regression	1.7459	-0.0673	0.0116	0.0162	0.910
Super Learner	1.7388	-0.0744	0.0111	0.0167	0.900

Table 3.4: Simulation results for the counterfactual stratum direct effect (risk ratio), corresponding to a controlled direct effect.

	Mean	Bias	Variance	MSE	95% CI Coverage
$\tau = 1$ ($\Psi = 1.873$)					
Main terms regression	1.9126	0.0397	0.0298	0.0313	0.970
Super Learner	1.8737	0.0007	0.0319	0.0319	0.950
$\tau = 2$ ($\Psi = 2.0683$)					
Main terms regression	1.9696	-0.0988	0.0368	0.0466	0.922
Super Learner	1.9601	-0.1083	0.0389	0.0506	0.892
$\tau = 3$ ($\Psi = 2.1327$)					
Main terms regression	2.0190	-0.1137	0.0466	0.0595	0.874
Super Learner	2.0203	-0.1124	0.0470	0.0596	0.864

3.5 Discussion

The challenge of formulating causal effects when exposure affects selection into a subgroup and causal effects are of interest within this subgroup has previously been characterized as an issue of selection bias. Effect measures such as the hazard ratio have been described as suffering from a "built-in selection bias" because the types of people that remain at-risk of the study outcome (i.e., in the subgroup of interest at time t) may change as a result of competing events and/or occurrence of the study outcome, which removes individuals from the population-at-risk over the course of the follow-up period.[20,62,71] The resulting differences

in risk that arise between unexposed and exposed groups over time in survival settings or when differential loss to follow-up or censoring occurs has been described as a type of bias induced by exposure or bias arising from conditioning on a collider variable.[20,62,71]

By contrast, we contend that the process of selection is a mechanism through which the exposure affects the outcome. While the effects of the exposure on selection may be unintended consequences of an intervention or treatment, this type of selection is not necessarily a systematic bias that distorts an underlying causal effect, rather, depending on the question, it may compose part of the causal effect of an exposure on the outcome. Conditioning or stratifying on the observed variable that is affected by treatment may induce selection bias in the presence of unmeasured confounding of the mediator (i.e., subgroup membership) and outcome relationship. However, 1) in the absence of exposure-outcome and mediator-outcome confounding, a controlled direct effect may be identifiable when we condition on the observed mediator,[9] and 2) stratifying on the observed post-exposure variable may not be necessary to formulate and estimate identifiable causal effects in these types of settings. It is often of greater substantive interest to formulate causal parameters, like the counterfactual strata total and counterfactual stratum direct effects, that stratify on counterfactual rather than observed variables. For example, in Chapter 1, we described the limitations of controlled direct effects in perinatal settings, including the SEARCH Study, where the probability of a live birth may be low overall, or within particular baseline covariate strata, and a parameter corresponding to a hypothetical intervention where everyone has a live birth may not answer a meaningful research question.

Principal stratification is an alternative approach that has been applied in settings where exposure affects selection into a post-exposure subgroup of interest, including when censoring or non-compliance are of concern.[11,67,72] Whereas counterfactual stratum direct effects consider the effect of an intervention within a counterfactual stratum defined using single-world counterfactuals, in the principal stratification approach, the impact of an intervention is evaluated within a counterfactual stratum defined using joint potential outcomes.[11,67] As a result, counterfactual stratum direct effect parameters require weaker, single-world assumptions for identification while parameters formulated using principal stratification often require stronger assumptions to be identified.[11,67,72]

As discussed in Chapter 1, principal stratum direct effect parameters may be of interest in perinatal settings, such as the SEARCH Study, where exposure affects the occurrence of a live birth.[21] Principal stratum direct effects quantify the causal effect of an exposure on an outcome within the subgroup that would have been selected into the subgroup irrespective of exposure assignment. In SEARCH, this principal stratum corresponds to the group who would have had a live birth regardless of assignment to the intervention or control arm of the study; limitations of this parameter in perinatal settings are described in greater detail in Chapter 1.

In randomized studies where there may be concerns of non-compliance or non-adherence, local average treatment effects (i.e., complier average causal effects), a type of principal stratum parameter, have been applied; these parameters describe the effect of an intervention in the group that would always comply with treatment assignment.[11,72,73,74,75,76,77] The issue of

noncompliance in randomized trials is similar to the challenges raised by concomitant medication use in the LEADER trial, where differences in concomitant medication use by trial arm may complicate the interpretation of a traditional total effect. In this trial, principal stratum direct effects can be used to describe the causal effect of the primary treatment on cardiovascular outcomes in the group of individuals who would not take any concomitant medications post-baseline irrespective of whether they were assigned to receive the primary treatment or placebo. When the subgroup of interest may not be the principal stratum, counterfactual stratum direct effect parameters provide a way to understand how an intervention affects an outcome within a subgroup that can be flexibly defined using different choices of stochastic interventions on mediating variables, which include but are not limited to subgroup membership.

Moreover, counterfactual stratum direct effects are easily adapted from settings where the mediator of subgroup selection occurs at a single time point to longitudinal settings with multiple mediators and subgroup membership that varies over time. By extending counterfactual stratum direct effects to longitudinal settings, they can be formulated to exclude effects of an exposure on the outcome that arise from intervention-induced changes to the timing of selection into the subgroup. Principal stratum effects are more difficult to define when the variable representing subgroup membership has more than two levels and/or when selection is modeled as a longitudinal process.[11]

Causal effects, such as counterfactual strata effects, can be defined in settings where exposure influences selection, and, depending on the formulation of these causal estimands, they can summarize the impact of an exposure including or excluding exposure-induced changes to the composition of the subgroup of interest. Counterfactual strata total effects are formulated in a way that excludes any impact of the intervention on the marginal probability of selection into a subgroup; these total effects still include any effect of the intervention that operates through changes to the types of people within the subgroup under different exposures. Alternatively, counterfactual stratum direct effects are formulated to exclude exposure-induced changes to the types of people within the subgroup of interest and quantify the causal effect of the exposure on the outcome of interest if the composition of the counterfactual stratum remained the same (regardless of the value of the exposure).

In this chapter, we have demonstrated how counterfactual strata total effects and counterfactual stratum direct effects can be applied to understand the effects of interventions within post-exposure strata using the Causal Roadmap,[25] with particular attention to the effects of timing on subgroup selection. These parameters require weaker assumptions for identification than many other causal estimands that are often used in these settings and can be estimated using longitudinal targeted maximum likelihood estimation. They offer alternatives to existing approaches for handling issues of censoring, loss to follow-up, noncompliance, and the formulation of causal effects when exposure affects the precondition necessary for observation of study outcomes.

Discussion

Causal effects are often of interest within strata of post-exposure variables that may be affected by exposure and mediate the exposure-outcome relationship. In these settings, formulating causal effects that stratify on counterfactual post-exposure variables, instead of their observed values, provides a different understanding of the mechanisms by which an exposure operates to change an outcome variable as compared to other types of causal effects that have been used in these settings, including controlled and principal stratum effects. In Chapter 1, we introduced two causal effects that are formulated in this way, counterfactual strata total effects and counterfactual stratum direct effects, and described their relevance in perinatal settings, where exposure may affect both the probability of live birth as well as the probability of infant outcomes, which are only defined if and when a live birth occurs.

In Chapter 2, we applied these parameters to the specific example of infant HIV-free survival and perinatal HIV transmission in the SEARCH Study. We found evidence that the SEARCH Study significantly increased the probability of infant HIV-free survival and reduced the risk of perinatal HIV transmission among live births to women living with HIV at the beginning of the study. In addition, when we estimated the effects of the SEARCH intervention excluding any impacts of the intervention that operated through changes to the types of people who had live births and the timing of live births, we similarly found that the SEARCH intervention resulted in significant improvements in infant outcomes as compared to the control condition, across different choices of stochastic interventions on the mediating variables (i.e., the live birth process). The magnitude of the improvements in infant HIV-free survival and perinatal HIV transmission were greater for the estimated counterfactual stratum direct effects than the estimated counterfactual strata total effects.

In Chapter 3, we discussed how counterfactual strata effects differ from traditional approaches to handling issues that arise in epidemiology, from censoring and loss to follow-up to the formulation of causal effects when exposure may affect who becomes at-risk for a study outcome (e.g., who has a live birth when infant outcomes are of interest) or who remains at-risk for a study outcome (e.g., in survival settings where competing events or occurrence of the study outcome may change the composition of the population-at-risk). We used the Causal Roadmap[25] to illustrate how counterfactual strata effects can be used when subgroup membership is a single mediating variable, as well as in more complex, longitudinal settings where subgroup membership may vary over time or it may be of interest to formulate direct effect parameters that intervene on more than one time-varying mediator. We demonstrated

the performance of a longitudinal TMLE for this type of direct effect parameter in simulations, considering different types of hypothetical interventions on the mediating variables.

Future directions of this work include the application of counterfactual strata effects to understand both the effect of an exposure on infant outcomes among live births (a conditional probability), as well as the effect of an exposure on infant outcomes absent any intervention-induced changes to the process of live birth (i.e., maternal survival and live birth) to answer other questions in perinatal epidemiology. In addition, in Chapter 2, while we accounted for the cluster randomized design of the SEARCH Study, we also "broke the match," as SEARCH was a pair-matched community-randomized trial. Explicit incorporation of the pair-matched design of the SEARCH Study offers the potential to increase precision. Further, the two-stage adaptive pre-specification approach of Balzer et al.[78] could be adapted for the current parameters.

As described in Chapter 3, counterfactual strata effects are broadly applicable outside of perinatal settings as well; these parameters can be used to understand the effects of exposures on outcomes within post-exposure strata generally and offer an alternative to stratifying on observed post-exposure variables, which often occurs in the presence of censoring, loss to follow-up, non-compliance, or when outcomes are only defined for a subset of the target population. The approach for formulating counterfactual stratum direct effects to exclude intervention-induced changes to the population at-risk for study outcomes may be particularly relevant in survival settings to formulate and estimate an alternative type of hazard ratio, which could quantify the impact of an intervention on outcomes if the population-at-risk under each treatment were to have the same composition.

Bibliography

[1] Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

[2] MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation Analysis. *Annual review of psychology*, *58*, 593.

[3] Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206.

[4] Pearl, J. (2012). The causal mediation formula–a guide to the assessment of pathways and mechanisms. *Prevention Science: The Official Journal of the Society for Prevention Research*, *13*(4), 426–436.

[5] Pardo, A., & Román, M. (2013). Reflections on the Baron and Kenny model of statistical mediation. *Anales de Psicología*, *29*(2), 614–623
(Spain).

[6] Goldsmith, K. A., MacKinnon, D. P., Chalder, T., White, P. D., Sharpe, M., & Pickles, A. (2018). Tutorial: The Practical Application of Longitudinal Structural Equation Mediation Models in Clinical Trials. *Psychological methods*, *23*(2), 191–207.

[7] Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233.

[8] Imai, K., Keele, L., & Tingley, D. (2010). A General Approach to Causal Mediation Analysis. *Psychological methods*, *15*(4), 309–334.

[9] Robins, J. M., & Greenland, S. (1992). Identifiability and Exchangeability for Direct and Indirect Effects. *Epidemiology*, *3*(2), 143–155.

[10] Pearl, J. (2001). Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, 411–420.

[11] VanderWeele, T. (2015). *Explanation in causal inference : Methods for mediation and interaction*. Oxford University Press.

[12] Naimi, A. I., Kaufman, J. S., & MacLehose, R. F. (2014). Mediation misgivings: Ambiguous clinical and public health interpretations of natural direct and indirect effects. *International Journal of Epidemiology*, *43*(5), 1656–1661.

[13] Avin, C., Shpitser, I., & Pearl, J. (2005). Identifiability of Path-Specific Effects.

[14] Zheng, W., & van der Laan, M. (2017). Longitudinal Mediation Analysis with Time-varying Mediators and Exposures, with Application to Survival Outcomes. *Journal of Causal Inference*, *5*(2).

[15] VanderWeele, T. J., & Tchetgen Tchetgen, E. J. (2017). Mediation analysis with time varying exposures and mediators. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *79*(3), 917–938.

[16] Rudolph, K. E., Sofrygin, O., Zheng, W., & van der Laan, M. J. (2017). Robust and Flexible Estimation of Stochastic Mediation Effects: A Proposed Method and Example in a Randomized Trial Setting. *Epidemiologic methods*, *7*(1), 20170007.

[17] Díaz, I., Hejazi, N. S., Rudolph, K. E., & van der Laan, M. J. (2021). Nonparametric efficient causal mediation with intermediate confounders. *Biometrika*, *108*(3), 627–641.

[18] Didelez, V., Dawid, P., & Geneletti, S. (2006). Direct and Indirect Effects of Sequential Treatments. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence* (pp. 138–146). AUAI Press.

[19] van der Laan, M. J., & Petersen, M. L. (2008). Direct Effect Models. *The International Journal of Biostatistics*, *4*(1), Article 23.

[20] Hernán, M. A. (2010). The Hazards of Hazard Ratios. *Epidemiology*, *21*(1), 13–15.

[21] Chiu, Y.-H., Stensrud, M. J., Dahabreh, I. J., Rinaudo, P., Diamond, M. P., Hsu, J., Hernandez-Diaz, S., & Hernan, M. A. (2020). The Effect of Prenatal Treatments on Offspring Events in the Presence of Competing Events An Application to a Randomized Trial of Fertility Therapies. *Epidemiology*, *31*(5), 636–643.

[22] Young, J. G., & Stensrud, M. J. (2021). Identified Versus Interesting Causal Effects in Fertility Trials and Other Settings With Competing or Truncation Events. *Epidemiology*, *32*(4), 569–572.

[23] Stensrud, M. J., Robins, J. M., Sarvet, A., Tchetgen, E. J. T., & Young, J. G. (2023). Conditional Separable Effects. *Journal of the American Statistical Association*, *118*(544), 2671–2683.

[24] Snowden, J. M., Reavis, K. M., & Odden, M. C. (2020). Conceiving of Questions Before Delivering Analyses Relevant Question Formulation in Reproductive and Perinatal Epidemiology. *Epidemiology*, *31*(5), 644–648.

[25] Petersen, M. L., & van der Laan, M. J. (2014). Causal models and learning from data: Integrating causal modeling and statistical estimation. *Epidemiology*, *25*(3), 418–426.

[26] Havlir, D. V., Balzer, L. B., Charlebois, E. D., Clark, T. D., Kwarisiima, D., Ayieko, J., Kabami, J., Sang, N., Liegler, T., Chamie, G., Camlin, C. S., Jain, V., Kadede, K., Atukunda, M., Ruel, T., Shade, S. B., Ssemmondo, E., Byonanebye, D. M., Mwangwa, F., Owaraganise, A., Olilo, W., Black, D., Snyman, K., Burger, R., Getahun, M., Achando, J., Awuonda, B., Nakato, H., Kironde, J., Okiror, S., Thirumurthy, H., Koss, C., Brown, L., Marquez, C., Schwab, J., Lavoy, G., Plenty, A., Mugoma Wafula, E., Omanya, P., Chen, Y.-H., Rooney, J. F., Bacon, M., van der Laan, M., Cohen, C. R., Bukusi, E., Kamya, M. R., & Petersen, M. (2019). HIV Testing and Treatment with the Use of a Community Health Approach in Rural Africa. *New England Journal of Medicine*, *381*(3), 219–229.

[27] Ayieko, J., Petersen, M. L., Charlebois, E. D., Brown, L. B., Clark, T. D., Kwarisiima, D., Kamya, M. R., Cohen, C. R., Bukusi, E. A., Havlir, D. V., & Van Rie, A. (2019). A Patient-Centered Multicomponent Strategy for Accelerated Linkage to Care Following Community-Wide HIV Testing in Rural Uganda and Kenya. *Journal of Acquired Immune Deficiency Syndromes*, *80*(4), 414–422.

[28] Chamie, G., Clark, T. D., Kabami, J., Kadede, K., Ssemmondo, E., Steinfeld, R., Lavoy, G., Kwarisiima, D., Sang, N., Jain, V., Thirumurthy, H., Liegler, T., Balzer, L. B., Petersen, M. L., Cohen, C. R., Bukusi, E. A., Kamya, M. R., Havlir, D. V., & Charlebois, E. D. (2016). A hybrid mobile approach for population-wide HIV testing in rural east Africa: An observational study. *The Lancet HIV*, *3*(3), e111–119.

[29] Gupta, S., Kabami, J., Chamie, G., Sang, N., Kwarisiima, D., Black, D., Balzer, L., Ayieko, J., Cohen, C. R., Bukusi, E. A., Kamya, M., Havlir, D., Petersen, M., & Ruel, T. (2020). Population-level HIV-free infant survival in the SEARCH trial. Conference on Retroviruses and Opportunistic Infections (CROI).

[30] Geneletti, S. (2007). Identifying direct and indirect effects in a non-counterfactual framework. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *69*(2), 199–215.

[31] VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator-outcome confounder. *Epidemiology*, *25*(2), 300–306.

[32] Pearl, J. (2010). An Introduction to Causal Inference. *The International Journal of Biostatistics*, *6*(2), 7.

[33] Balzer, L. B., Zheng, W., van der Laan, M. J., & Petersen, M. L. (2019). A new approach to hierarchical data analysis: Targeted maximum likelihood estimation for the causal effect of a cluster-level exposure. *Statistical Methods in Medical Research*, *28*(6), 1761–1780.

[34] Benitez, A., Petersen, M. L., van der Laan, M. J., Santos, N., Butrick, E., Walker, D., Ghosh, R., Otieno, P., Waiswa, P., & Balzer, L. B. (2023). Defining and estimating effects in cluster randomized trials: A methods comparison. *Statistics in Medicine*, *42*(19), 3443–3466.

[35] van der Laan, M. J., & Petersen, M. L. (2005). Direct Effect Models. *U.C. Berkeley Division of Biostatistics Working Paper Series*.

[36] Hubbard, A. E., Kherad-Pajouh, S., & van der Laan, M. J. (2016). Statistical Inference for Data Adaptive Target Parameters. *The International Journal of Biostatistics*, *12*(1), 3–19.

[37] Pearl, J., & Robins, J. M. (1995). Probabilistic evaluation of sequential plans from causal models with hidden variables. *Uncertainty in Artificial Intelligence: Proceedings of the Eleventh Conference on Artificial Intelligence*, 444–453.

[38] Balzer, L. B., Ayieko, J., Kwarisiima, D., Chamie, G., Charlebois, E. D., Schwab, J., van der Laan, M. J., Kamya, M. R., Havlir, D. V., & Petersen, M. L. (2020). Far from MCAR: Obtaining population-level estimates of HIV viral suppression. *Epidemiology*, *31*(5), 620–627.

[39] Goin, D. E., Rudolph, K. E., Gomez, A. M., & Ahern, J. (2020). Mediation of Firearm Violence and Preterm Birth by Pregnancy Complications and Health Behaviors: Addressing Structural and Postexposure Confounding. *American Journal of Epidemiology*, *189*(8), 820–831.

[40] *Global Guidance on Criteria and Processes for Validation: Elimination of Mother-to-Child Transmission of HIV and Syphilis, 2nd Edition* (tech. rep.). (2017). World Health Organization. Geneva.

[41] WHO — Mother-to-child transmission of HIV. (n.d.).

[42] Joint United Nations Programme on HIV/AIDS. (2023). *The path that ends AIDS: UNAIDS Global AIDS Update 2023.* (tech. rep. No. Licence: CC BY-NC-SA 3.0 IGO). Geneva.

[43] Newell, M.-L., Coovadia, H., Cortina-Borja, M., Rollins, N., Gaillard, P., & Dabis, F. (2004). Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: A pooled analysis. *The Lancet*, *364*(9441), 1236–1243.

[44] Vreeman, R. C., Scanlon, M. L., McHenry, M. S., & Nyandiko, W. M. (2015). The physical and psychological effects of HIV infection and its treatment on perinatally HIV-infected children. *Journal of the International AIDS Society*, *18*(7Suppl 6).

[45] Rylance, J., Mchugh, G., Metcalfe, J., Mujuru, H., Nathoo, K., Wilmore, S., Rowland-Jones, S., Majonga, E., Kranzer, K., & Ferrand, R. A. (2016). Chronic lung disease in HIV-infected children established on antiretroviral therapy. *AIDS*, *30*(18), 2795.

[46] Sofeu, C. L., Tejiokem, M. C., Penda, C. I., Protopopescu, C., Ndongo, F. A., Ndiang, S. T., Guemkam, G., Warszawski, J., Faye, A., Giorgi, R., & study Group, t. A.-P. (2019). Early treated HIV-infected children remain at risk of growth retardation during the first five years of life: Results from the ANRS-PEDIACAM cohort in Cameroon. *PLOS ONE*, *14*(7), e0219960.

[47] Leonard, E. G., & McComsey, G. A. (2005). Antiretroviral therapy in HIV-infected children: The metabolic cost of improved survival. *Infectious Disease Clinics of North America*, *19*(3), 713–729.

[48] Aldrovandi, G. M., Lindsey, J. C., Jacobson, D. L., Zadzilka, A., Sheeran, E., Moye, J., Borum, P., Meyer, W. A., Hardin, D. S., & Mulligan, K. (2009). Morphologic and Metabolic Abnormalities in Vertically HIV-infected Children and Youth. *AIDS*, *23*(6), 661–672.

[49] Sustainable East Africa Research in Community Health - ClinicalTrials.gov. (n.d.).

[50] Kwarisiima, D., Kamya, M. R., Owaraganise, A., Mwangwa, F., Byonanebye, D. M., Ayieko, J., Plenty, A., Black, D., Clark, T. D., Nzarubara, B., Snyman, K., Brown, L., Bukusi, E., Cohen, C. R., Geng, E. H., Charlebois, E. D., Ruel, T. D., Petersen, M. L., Havlir, D., & Jain, V. (2017). High rates of viral suppression in adults and children with high CD4+counts using a streamlined ART delivery model in the SEARCH trial in rural Uganda and Kenya. *Journal of the International AIDS Society*, *20*.

[51] Robinson, D. (2018). *Fuzzyjoin: Join Tables Together on Inexact Matching* (tech. rep.). R package version 0.1.4.

[52] R Core Team. (2023). R: A Language and Environment for Statistical Computing.

[53] World Health Organization. (2011). *Manual on Paediatric HIV Care and Treatment for District Hospitals: Addendum to the Pocket Book of Hospital Care of Children*.

[54] Petersen, M., Schwab, J., Gruber, S., Blaser, N., Schomaker, M., & van der Laan, M. (2014). Targeted Maximum Likelihood Estimation for Dynamic and Static Longitudinal Marginal Structural Working Models. *Journal of causal inference*, *2*(2), 147–185.

[55] van der Laan, M. J., & Rose, S. (2011). *Targeted learning: Causal inference for observational and experimental data*. Springer.

[56] Lendle, S. D., Schwab, J., Petersen, M. L., & van der Laan, M. J. (2017). Ltmle: An R Package Implementing Targeted Minimum Loss-Based Estimation for Longitudinal Data. *Journal of Statistical Software*, *81*(1), 1–21.

[57] Polley, E., LeDell, E., Kennedy, C., & van der Laan, M. J. (2023). SuperLearner: Super Learner Prediction.

[58] Hastie, T. (1991). Generalized additive models. In *Statistical Models in S* (Vol. Chapter 7). Wadsworth & Brooks/Cole.

[59] Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized Additive Models*. CRC Press.

[60] Venables, W. N., & Ripley, B. D. (2003). *Modern Applied Statistics with S*. Springer Science & Business Media.

[61] Rytgaard, H. C., Fong, E., Tarp, J. M., Gerds, T. A., Rasmussen, S., van der Laan, M. J., & Ravn, H. (2024). Hypothetical treatment interventions to handle treatment drop-in in randomized controlled trials.

[62] Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Philadelphia : Wolters Kluwer Health/Lippincott Williams & Wilkins.

[63] Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling*, *7*(9), 1393–1512.

[64] Hernán, M. A., & Hernández-Díaz, S. (2012). Beyond the intention to treat in comparative effectiveness research. *Clinical Trials*, *9*(1), 48–55.

[65] Didelez, V. (2019). Defining causal mediation with a longitudinal mediator and a survival outcome. *Lifetime Data Analysis*, *25*(4), 593–610.

[66] Stensrud, M. J., Young, J. G., Didelez, V., Robins, J. M., & Hernán, M. A. (2022). Separable Effects for Causal Inference in the Presence of Competing Events. *Journal of the American Statistical Association*, *117*(537), 175–183.

[67] Frangakis, C. E., & Rubin, D. B. (2002). Principal Stratification in Causal Inference. *Biometrics*, *58*(1), 21–29.

[68] Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., & Hernán, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*, *39*(8), 1199–1236.

[69] Bang, H., & Robins, J. M. (2005). Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, *61*(4), 962–973.

[70] van der Laan, M. J., & Gruber, S. (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The International Journal of Biostatistics*, *8*(1).

[71] Hernán, M. A., Hernández-Díaz, S., & Robins, J. M. (2004). A structural approach to selection bias. *Epidemiology (Cambridge, Mass.)*, *15*(5), 615–625.

[72] Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association*, *91*(434), 444–455.

[73] Imbens, G. W., & Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, *25*(1), 305–327.

[74] Balke, A., & Pearl, J. (1997). Bounds on Treatment Effects From Studies With Imperfect Compliance. *Journal of the American Statistical Association*, *92*(439), 1171–1176.

[75] Cheng, J., & Small, D. S. (2006). Bounds on causal effects in three-arm trials with noncompliance. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(5), 815–836.

[76] Cuzick, J., Sasieni, P., Myles, J., & Tyrer, J. (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*(4), 565–588.

[77] Little, R. J., Long, Q., & Lin, X. (2009). A comparison of methods for estimating the causal effect of a treatment in randomized clinical trials subject to noncompliance. *Biometrics*, *65*(2), 640–649.

[78] Balzer, L. B., van der Laan, M., Ayieko, J., Kamya, M., Chamie, G., Schwab, J., Havlir, D. V., & Petersen, M. L. (2023). Two-Stage TMLE to reduce bias and improve efficiency in cluster randomized trials. *Biostatistics*, *24*(2), 502–517.

Appendix A

Chapter 1

A.1 Structural Causal Model

A longitudinal structural causal model is used to represent the data generating process of perinatal HIV transmission in the target population of women living with HIV who are at risk for pregnancy over the follow-up period. L_0 denotes the baseline covariates and A the intervention node. L_t represents time-varying covariates (which may be confounders of the mediator-outcome relationship(s)) at each time t. Z_t denotes the set of mediators, $Z1_t$ and $Z2_t$, at each time point. We use two mediators to capture two mechanisms by which the intervention may affect the timing of births and the types of people who give birth: 1) by the effect of the intervention on maternal morality ($Z1_t$) and 2) by the effect of the intervention on the timing of live birth ($Z2_t$). Both $Z1_t$ and $Z2_t$ are time-to-event variables that each take a value of 1 deterministically at time $t+1$ if they are equal to 1 at time t. The final outcome of interest is the composite variable $Y_{\tau+1}$, which takes a value of 1 if, in terms of our motivating example, a live birth occurs and the infant survives and is HIV-free at the end of follow-up and a value of 0 otherwise. Infant survival is represented by the node $Y1_t$; its value is missing if a live birth has not yet occurred by time t. $Y2_{\tau+1}$ denotes whether a live infant was HIV-free at the end of follow-up. Because infection status was only assessed at the end of the study, it is not a time-varying covariate, and its value will be missing in the absence of a live birth or in the event of infant mortality prior to the end of the follow-up period.

$$
\begin{aligned}
L_0 &= f_{L_0}(U_{L_0}) \\
A &= f_A(U_A) \\
L_t &= f_{L_t}(L_0, A, \bar{L}_{t-1}, \bar{Z}_{t-1}, U_{L_t}) \\
Z_t &= \{Z1_t, Z2_t\} \\
Z1_t &= \begin{cases} 1, & \text{if } Z1_{t-1} = 1 \\ f_{Z1_t}(L_0, A, \bar{L}_t, \bar{Z}1_{t-1}, \bar{Z}2_{t-1}, U_{Z1_t}), & \text{otherwise} \end{cases} \\
Z2_t &= \begin{cases} 1, & \text{if } Z2_{t-1} = 1 \\ f_{Z2_t}(L_0, A, \bar{L}_t, \bar{Z}1_t, \bar{Z}2_{t-1}, U_{Z2_t}), & \text{otherwise} \end{cases} \\
Y1_t &\in L_{t+1} \\
Y1_t &= \begin{cases} NA, & \text{if } Z2_{t-1} = 0 \\ 0, & \text{if } Y1_{t-1} = 0 \\ f_{Y1_t}(L_0, A, \bar{L}_{t-1}, \bar{Z}_{t-1}, U_{Y1_t}), & \text{otherwise} \end{cases} \\
Y2_{\tau+1} &= \begin{cases} NA, & \text{if } Z2_\tau = 0 \text{ or } Y1_{\tau+1} = 0 \\ f_{Y2_{\tau+1}}(L_0, A, \bar{L}_\tau, \bar{Z}_\tau, U_{Y2_{\tau+1}}), & \text{otherwise} \end{cases} \\
Y_{\tau+1} &= \begin{cases} 1, & \text{if } Z2_\tau = 1, Y1_{\tau+1} = 1, \text{ and } Y2_{\tau+1} = 1 \\ 0, & \text{if } Z2_\tau = 0 \text{ or } (Z2_\tau = 1 \text{ and } (Y1_{\tau+1} = 0 \text{ or } Y2_{\tau+1} = 0)) \end{cases}
\end{aligned}
$$

A.2 Summary of Natural Direct Effect Parameters

Table A.1: Definition and interpretations of alternative target natural direct effect parameters.

Parameter	Interpretation (General)	Interpretation (SEARCH)	Counterfactuals (SEARCH)	Hypothetical Intervention
Marginal mediator distributions[a] $\frac{P[Y_{\tau+1}(a=1,\bar{G}^{a=0})=1]}{P(Z2_\tau(\bar{G}^{a=0})=1)} - \frac{P[Y_{\tau+1}(a=0,\bar{G}^{a=0})=1]}{P(Z2_\tau(\bar{G}^{a=0})=1)}$ $= P[Y1_{\tau+1}(a=1,\bar{G}^{a=0})=0, Y2_{\tau+1}(a=1,\bar{G}^{a=0})=0 \mid Z2_\tau(\bar{G}^{a=0})=1]$ $= P[Y1_{\tau+1}(a=0,\bar{G}^{a=0})=0, Y2_{\tau+1}(a=0,\bar{G}^{a=0})=0 \mid Z2_\tau(\bar{G}^{a=0})=1]$ where $\bar{G}^{a=0} \equiv (G_1^{a=0},\ldots,G_\tau^{a=0})$ and $G_t^{a=0}(z_t \mid l_0) \equiv P\big(Z_t(a=0)=z_t \mid L_0=l_0\big)$	• A natural direct effect of an intervention on infant health outcomes where the live birth process is drawn from its true distribution of live births conditional on baseline covariates under the control condition • Includes: any effect of the intervention on the outcome that operates through mechanisms other than the live birth process • Excludes: any effect of the intervention on infant health outcomes through changes to the timing/occurrence of live births and to the types of people who give birth	• The natural direct effect of the SEARCH intervention on infant health outcomes where the live birth process is drawn from the true distribution of live births conditional on baseline covariates under the control condition	• $Y_{\tau+1}(a,\bar{G}^{a'})$ represents counterfactual live birth and infant HIV-free survival at $t=\tau+1$ under an intervention to set $A=a$ and draw the mediators from their (unknown) distribution under $A=a'$ conditional on baseline covariates • $Y1_{\tau+1}(a,\bar{G}^{a'})$ represents counterfactual infant vital status at $t=\tau+1$ under an intervention to set $A=a'$ and draw maternal survival and time-to-live birth from their distribution under $A=a'$ conditional on baseline covariates • $Y2_{\tau+1}(a,\bar{G}^{a'})$ represents counterfactual infant HIV status at $t=\tau+1$ under an intervention to set $A=a'$ and draw maternal survival and time-to-live birth from their distribution under $A=a'$ conditional on baseline covariates • $Z2_\tau(\bar{G}^{a'})$ is the counterfactual live birth status at $t=\tau$ under the stochastic intervention	• Compare intervening to set $A=1$ for everyone to setting $A=0$ for everyone and drawing the live birth process from the true, unknown live birth distribution conditional on baseline covariates under the control exposure
Conditional mediator distributions $\frac{P[Y_{\tau+1}(a=1,\bar{\Gamma}^{a=0})=1]}{P(Z2_\tau(\bar{\Gamma}^{a=0})=1)} - \frac{P[Y_{\tau+1}(a=0)=1]}{P(Z2_\tau(a=0)=1)}$ $= P[Y1_{\tau+1}(a=1,\bar{\Gamma}^{a=0})=0, Y2_{\tau+1}(a=1,\bar{\Gamma}^{a=0})=0 \mid Z2_\tau(\bar{\Gamma}^{a=0})=1]$ $- P[Y1_{\tau+1}(a=0)=0, Y2_{\tau+1}(a=0)=0 \mid Z2_\tau(a=0)=1]$ where $\bar{\Gamma}^{a=0} \equiv (\Gamma_1^{a=0},\ldots,\Gamma_\tau^{a=0})$ and $\Gamma_t^{a=0}(z_t \mid \bar{l}_t,\bar{z}_{t-1}) \equiv P\big(Z_t(a=0)=z_t \mid \bar{L}_t(a=0)=\bar{l}_t, \bar{Z}_{t-1}(a=0)=\bar{z}_{t-1}\big)$	• A natural direct effect of an intervention on infant health outcomes where the live birth process is drawn from its true distribution of live births conditional on baseline and time-varying history under the control condition • Includes: any effect of the intervention on the outcome through changes to time-varying covariates which may in turn impact the timing/occurrence of live births and the types of people who give birth • Excludes: any effect of the intervention on infant health outcomes that operating only through the survival of those at risk for pregnancy and/or live birth	• The natural direct effect of the SEARCH intervention on infant HIV-free survival if the live birth process was drawn from its true, unknown distribution conditional on baseline and time-varying covariate history under the control condition	• $Y_{\tau+1}(a,\bar{\Gamma}^{a'})$ is counterfactual live birth and infant HIV-free survival at $t=\tau+1$ under an intervention to set $A=a$ and draw maternal survival and time-to-live birth from their distribution under $A=a'$ conditional on observed baseline and time-varying covariate history • $Y_{\tau+1}(a)$ is counterfactual live birth and infant HIV-free survival at $t=\tau+1$ under an intervention to set $A=a$ • $Y1_{\tau+1}(a,\bar{\Gamma}^{a'})$ is counterfactual infant vital status at $t=\tau+1$ under an intervention to set $A=a$ and draw maternal survival and time-to-live birth from their distribution under $A=a'$ conditional on observed baseline and time-varying covariate history • $Y1_{\tau+1}(a)$ is counterfactual infant vital status at $t=\tau+1$ under an intervention to set $A=a$ • $Y2_{\tau+1}(a,\bar{\Gamma}^{a'})$ is counterfactual infant HIV status at $t=\tau+1$ under an intervention to set $A=a$ and draw maternal survival and time-to-live birth from their distribution under $A=a'$ conditional on observed baseline and time-varying covariate history • $Y2_{\tau+1}(a)$ is counterfactual infant HIV status at $t=\tau+1$ under an intervention to set $A=a$ • $Z2_\tau(a,\bar{\Gamma}^{a'})$ is the counterfactual live birth status at $t=\tau$ under this stochastic intervention conditional on baseline and time-varying covariate history • $Z2_\tau(a)$ is the counterfactual live birth at $t=\tau$ under an intervention to set $A=a$	• Compare intervening to set $A=1$ for everyone to setting $A=0$ for everyone while drawing the live birth process from its true, unknown distribution conditional on baseline and time-varying history

[a] Denominator is equal to $P[Z2_\tau(a=0)=1]$

Appendix B

Chapter 2

B.1 Causal Estimands

Counterfactuals

$Y_{\tau+1}(a)$ denotes the counterfactual outcome of live birth and infant HIV-free survival under a hypothetical intervention to set the value of the exposure A to some value a, and $M_{\tau+1}(a)$ denotes the counterfactual outcome of live birth, infant survival, and infant HIV infection if the value of exposure was set to $A = a$. $Z2_\tau(a)$ describes counterfactual live birth by time τ under an intervention to set exposure A to a. $Z3_\tau$ represents counterfactual live birth by and infant survival at time τ.

For the mediated effects, we also use $Y_{\tau+1}(a, z)$ to denote counterfactual live birth and infant HIV-free survival and $M_{\tau+1}(a, z)$ to denote counterfactual live birth and infant survival with HIV infection status under a hypothetical intervention to set the exposure A to the value a and the mediators Z to the value z. In addition, $Y_{\tau+1}(a, \bar{G})$ describes counterfactual live birth and infant HIV-free survival under a static intervention to set the exposure A to a and to randomly draw the values of the mediators Z (i.e., the live birth process) from some known distribution (possibly conditional on baseline covariates), $\bar{G}$. $Z2_\tau(\bar{G})$ represents counterfactual live birth by the final time point τ if live birth was drawn from this known distribution.

$M_{\tau+1}(a, \bar{G})$ describes counterfactual live birth and infant survival with HIV infection under a static intervention to set the exposure A to a and to randomly draw the values of maternal survival, live birth, and infant survival from some known distribution (possibly conditional on baseline covariates), $\bar{G}$. $Z3_\tau(\bar{G})$ represents counterfactual live birth by and infant survival at the final time point τ if the values of live birth and infant survival over time were drawn from this known distribution.

In a slight abuse of notation, we will use $\bar{G}$ to represent the two different stochastic interventions on mediating variables for each outcome of interest.

Counterfactual stratum total effects

The counterfactual stratum total effect of the SEARCH intervention on infant HIV-free survival, represented as a risk difference, is equal to the difference in the counterfactual probability of infant HIV-free survival ($Y1_{\tau+1} = 1, Y2_{\tau+1} = 0$) among counterfactual live births ($Z2_\tau = 1$). It is given by:

$$\begin{aligned} &P[Y1_{\tau+1}(a=1)=1, Y2_{\tau+1}(a=1)=0 | Z2_\tau(a=1)=1] \\ &\quad - P[Y1_{\tau+1}(a=0)=1, Y2_{\tau+1}(a=0)=0 | Z2_\tau(a=0)=1] \\ &\quad = \frac{E[Y_{\tau+1}(a=1)]}{E[Z2_\tau(a=1)]} - \frac{E[Y_{\tau+1}(a=0)]}{E[Z2_\tau(a=0)]} \end{aligned}$$

The counterfactual stratum total effect of the SEARCH intervention on perinatal HIV transmission can be represented as a risk difference that is equal to the expected difference in the risk of counterfactual infant HIV-free survival if everyone was assigned to the SEARCH intervention ($A = 1$) as compared to the control condition ($A = 0$):

$$\begin{aligned} &P[Y2_{\tau+1}(a=1)=1 | Z3_\tau(a=1)=1] - P[Y2_{\tau+1}(a=0)=1 | Z3_\tau(a=0)=1] \\ &\quad = \frac{E[M_{\tau+1}(a=1)]}{E[Z3_\tau(a=1)]} - \frac{E[M_{\tau+1}(a=0)]}{E[Z3_\tau(a=0)]} \end{aligned}$$

Counterfactual stratum direct effects

The counterfactual stratum direct effect of the SEARCH intervention on infant HIV-free survival summarizes the expected difference in the risk of counterfactual infant HIV-free survival if everyone was assigned to the SEARCH intervention ($A = 1$) as compared to the control condition ($A = 0$) and if maternal survival and live birth were drawn from some known distribution $\bar{G}$ under both static interventions on A.

$$\begin{aligned} &P[Y1_{\tau+1}(a=1,\bar{G})=1, Y2_{\tau+1}(a=1,\bar{G})=0 | Z2_\tau(\bar{G})=1] \\ &\quad - P[Y1_{\tau+1}(a=0,\bar{G})=1, Y2_{\tau+1}(a=0,\bar{G})=0 | Z2_\tau(\bar{G})=1] \\ &\quad = \frac{E[Y_{\tau+1}(a=1,\bar{G})]}{E[Z2_\tau(\bar{G})]} - \frac{E[Y_{\tau+1}(a=0,\bar{G})]}{E[Z2_\tau(\bar{G})]} \end{aligned}$$

By drawing the live birth process (maternal survival and time-to-live birth) from the same distributions, this parameter quantifies the effect of the SEARCH intervention on infant HIV-free survival excluding any impact of SEARCH on the timing of births or the people who had a live birth.

The CSDE of the SEARCH intervention on perinatal HIV transmission describes the expected difference in the risk of counterfactual perinatal HIV transmission. In other words, this parameter summarizes the difference in the counterfactual probability of infant HIV infection ($Y2_{\tau+1}(a, \bar{G})$) among counterfactual infants who survived until the end of study

follow-up ($Z3_\tau(\bar{G})$), if everyone was assigned to the SEARCH intervention ($A = 1$) as compared to the control condition ($A = 0$) and if maternal survival, live birth, and infant survival were drawn from the same known distribution $\bar{G}$ in both cases:

$$P[Y2_{\tau+1}(a = 1, \bar{G}) = 1 | Z3_\tau(\bar{G}) = 1] - P[Y2_{\tau+1}(a = 0, \bar{G}) = 1 | Z3_\tau(\bar{G}) = 1]$$
$$= \frac{E[M_{\tau+1}(a = 1, \bar{G}) = 1]}{E(Z3_\tau(\bar{G}))} - \frac{E[M_{\tau+1}(a = 0, \bar{G}) = 1]}{E(Z3_\tau(\bar{G}))}$$

Because the mediators of maternal survival, live birth, and infant survival are drawn from the same known distribution $\bar{G}$ under both static interventions on A, this parameter represents the effect of the SEARCH intervention on perinatal HIV transmission absent any intervention-induced changes to the timing of live births and the types of people who gave birth to infants who survived through study's end.

B.2 Identification

Counterfactual stratum total effects

For each a in $\mathcal{A} = \{0, 1\}$, the assumptions for identifiability of the counterfactual stratum direct effect of the SEARCH intervention on infant HIV-free survival are given by:

$$Y1_{\tau+1}(a), Z_\tau(a) \perp A | L_0$$

$$\min_{a \in \mathcal{A}} P_0(A = a | L_0) > 0 \; \forall \; l_0 \text{ for which } P_0(L_0 = l_0) > 0$$

Similarly, the assumptions required for identifiability of the CSTE of the SEARCH intervention on perinatal HIV transmission are:

$$Y2_{\tau+1}(a), Z3_\tau(a) \perp A | L_0$$

$$\min_{a \in \mathcal{A}} P_0(A = a | L_0) > 0 \; \forall \; l_0 \text{ for which } P_0(L_0 = l_0) > 0$$

Counterfactual stratum direct effects

The CSDE for the outcome of infant HIV-free survival are identified under the following assumptions:

$$Y_{\tau+1}(a, \bar{z}) \perp A | L_0$$

$$Y_{\tau+1}(a, \bar{z}) \perp Z_t | \bar{L}_t, \bar{Z}_{t-1}, A, L_0 \; \forall \; t \text{ in } 1, \ldots, \tau$$

$$\min_{a \in \mathcal{A}} P_0(A = a | L_0) > 0 \; \forall \; l_0 \text{ for which } P_0(L_0 = l_0) > 0$$

$$\min_{z_t \in \mathcal{Z}_t} P_0(Z_t = z_t | \bar{L}_t, \bar{Z}_{t-1}, A, L_0) > 0 \; \forall \; \bar{l}_t, \bar{z}_{t-1}, a, l_0 \text{ for which}$$

$$(P_0(\bar{L}_t = \bar{l}_t, \bar{Z}_{t-1} = \bar{z}_{t-1}, A = a, L_0 = l_0) > 0 \text{ and } G_t(Z_t = z_t | \bar{Z}_{t-1} = \bar{z}_{t-1}, L_0 = l_0) > 0)$$

Here, Z_t consists of both $Z1_t$ (maternal survival) and $Z2_t$ (live birth). Moreover, assumptions to identify the CSDE of the SEARCH intervention on perinatal HIV transmission are given by:

$$M_{\tau+1}(a, \bar{z}) \perp A | L_0$$

$$M_{\tau+1}(a, \bar{z}) \perp Z_t | \bar{L}_t, \bar{Z}_{t-1}, A, L_0 \; \forall \; t \text{ in } 1, \ldots, \tau$$

$$\min_{a \in \mathcal{A}} P_0(A = a | L_0) > 0 \; \forall \; l_0 \text{ for which } P_0(L_0 = l_0) > 0$$

$$\min_{z_t \in \mathcal{Z}_t} P_0(Z_t = z_t | \bar{L}_t, \bar{Z}_{t-1}, A, L_0) > 0 \; \forall \; \bar{l}_t, \bar{z}_{t-1}, a, l_0 \text{ for which}$$

$$(P_0(\bar{L}_t = \bar{l}_t, \bar{Z}_{t-1} = \bar{z}_{t-1}, A = a, L_0 = l_0) > 0 \text{ and } G_t(Z_t = z_t | \bar{Z}_{t-1} = \bar{z}_{t-1}, L_0 = l_0) > 0)$$

For the perinatal HIV transmission outcome, the mediator Z_t is a vector consisting of $Z1_t$ (maternal survival), $Z2_t$ (live birth), and $Z3_t$ (infant survival).

B.3 Estimation

Estimators

The estimator of the counterfactual stratum total effect of the SEARCH intervention on HIV-free survival, represented as a risk difference, is given by:

$$\hat{\psi}_{CSTE} = \frac{\hat{\psi}_{Y_{\tau+1}, a=1}}{\hat{\psi}_{Z2_\tau, a=1}} - \frac{\hat{\psi}_{Y_{\tau+1}, a=0}}{\hat{\psi}_{Z2_\tau, a=0}}$$

$\hat{\psi}_{Y_{\tau+1}a}$ represents the estimated probability of live birth and infant HIV-free survival $Y_{\tau+1}$ when $A = a$, and $\hat{\psi}_{Z2_\tau, a}$ is the estimated probability of live birth by time τ when $A = a$.

Similarly, the estimator of the CSTE for the perinatal HIV transmission outcome is given by:

$$\hat{\psi}_{CSTE} = \frac{\hat{\psi}_{M_{\tau+1}, a=1}}{\hat{\psi}_{Z3_\tau, a=1}} - \frac{\hat{\psi}_{M_{\tau+1}, a=0}}{\hat{\psi}_{Z3_\tau, a=0}}$$

$\hat{\psi}_{M_{\tau+1}a}$ represents the estimated probability of live birth, infant survival, and infant HIV-infection $M_{\tau+1}$ when $A = a$, and $\hat{\psi}_{Z3_\tau, a}$ is the estimated probability of live birth and infant survival by time τ when $A = a$.

The counterfactual stratum direct effect of the SEARCH intervention on infant HIV-free survival can be estimated using the following estimator, where $\hat{\psi}$ is an estimator of Ψ:

$$\hat{\psi}_{CSDE} = \frac{\hat{\psi}_{Y_{\tau+1}, a=1, \bar{G}}}{\hat{\psi}_{Z2_\tau, \bar{G}}} - \frac{\hat{\psi}_{Y_{\tau+1}, a=0, \bar{G}}}{\hat{\psi}_{Z2_\tau, \bar{G}}}$$

$\hat{\psi}_{Y_{\tau+1}a,\bar{G}}$ represents the estimated probability of live birth and infant HIV-free survival if $A = a$ and under a stochastic intervention to draw the values of maternal survival and live birth from the distribution $\bar{G}$. $\hat{\psi}_{Z2_\tau,\bar{G}}$ is the estimated probability of live birth under the stochastic intervention.

The counterfactual stratum direct effect of the SEARCH intervention on perinatal HIV transmission can be estimated using the following estimator:

$$\hat{\psi}_{CSDE} = \frac{\hat{\psi}_{M_{\tau+1},a=1,\bar{G}}}{\hat{\psi}_{Z3_\tau,\bar{G}}} - \frac{\hat{\psi}_{M_{\tau+1},a=0,\bar{G}}}{\hat{\psi}_{Z3_\tau,\bar{G}}}$$

$\hat{\psi}_{M_{\tau+1}a,\bar{G}}$ represents the estimated probability of live birth, infant survival, and infant HIV infection if $A = a$ and under a stochastic intervention to draw the values of maternal survival, live birth, and infant survival from the distribution $\bar{G}$. $\hat{\psi}_{Z3_\tau,\bar{G}}$ is the estimated probability of live birth and infant survival under this stochastic intervention.

Estimation of Counterfactual Stratum Direct Effects

Infant HIV-free survival

Table B.1: Stochastic direct effect estimated using TMLE under the observed mediator distribution in the control arm.

	Using main terms logistic regression		Using Super Learner*	
Infant HIV-free Survival	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth and infant HIV-free survival	0.1277	(0.1214, 0.1343)	0.1277	(0.1214, 0.1343)
Estimated probability of live birth	0.1312	(0.1236, 0.1392)	0.1312	(0.1236, 0.1392)
Estimated probability of infant HIV-free survival among live births	0.9736	(0.9566, 0.9908)	0.9736	(0.9579, 0.9896)
A=0 (Control)				
Estimated probability of live birth and infant HIV-free survival	0.1217	(0.1140, 0.1299)	0.1216	(0.1141, 0.1297)
Estimated probability of live birth	0.1312	(0.1236, 0.1392)	0.1312	(0.1236, 0.1392)
Estimated probability of infant HIV-free survival among live births	0.9275	(0.9080, 0.9473)	0.9273	(0.9093, 0.9457)
Risk Difference				
$\hat{\Psi}_{RD}$	0.0461	(0.0211, 0.0711)	0.0463	(0.0233, 0.0693)
Risk Ratio				
$\hat{\Psi}_{RR}$	1.0497	(1.0223, 1.0778)	1.0500	(1.0247, 1.0758)

* SuperLearner library consists of: SL.mean, SL.glm, SL.gam

Table B.2: Stochastic direct effect estimated with TMLE under a well-supported user-specified mediator distribution.

	Using main terms logistic regression		Using Super Learner*	
Infant HIV-free Survival	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth and infant HIV-free survival	0.1228	(0.1186, 0.1271)	0.1227	(0.1185, 0.1270)
Estimated probability of live birth	0.1262	(0.1209, 0.1317)	0.1262	(0.1209, 0.1317)
Estimated probability of infant HIV-free survival among live births	0.9732	(0.9563, 0.9904)	0.9725	(0.9564, 0.9888)
A=0 (Control)				
Estimated probability of live birth and infant HIV-free survival	0.1166	(0.1110, 0.1225)	0.1167	(0.1113, 0.1223)
Estimated probability of live birth	0.1262	(0.1209, 0.1317)	0.1262	(0.1209, 0.1317)
Estimated probability of infant HIV-free survival among live births	0.9245	(0.9041, 0.9454)	0.9248	(0.9063, 0.9437)
Risk Difference				
$\hat{\Psi}_{RD}$	0.0487	(0.0230, 0.0743)	0.0477	(0.0241, 0.0713)
Risk Ratio				
$\hat{\Psi}_{RR}$	1.0526	(1.0244, 1.0817)	1.0516	(1.0256, 1.0782)

* SuperLearner library consists of: SL.mean, SL.glm, SL.gam

Perinatal HIV transmission

Table B.3: Stochastic direct effect estimated using TMLE under the observed mediator distribution in the control arm.

	Using main terms logistic regression		Using Super Learner*	
Perinatal HIV Transmission	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0008	(0.0002, 0.0039)	0.0008	(0.0003, 0.0027)
Estimated probability of live birth and infant survival	0.1277	(0.1200, 0.1359)	0.1277	(0.1200, 0.1359)
Estimated probability of infant HIV infection among infants alive at end of study	0.0062	(0.0013, 0.0306)	0.0065	(0.0020, 0.0211)
A=0 (Control)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0059	(0.0042, 0.0082)	0.0056	(0.004, 0.008)
Estimated probability of live birth and infant survival	0.1277	(0.1200, 0.1359)	0.1277	(0.1200, 0.1359)
Estimated probability of infant HIV infection among infants alive at end of study	0.0458	(0.0326, 0.0643)	0.0441	(0.0310, 0.0626)
Risk Difference				
$\hat{\Psi}_{RD}$	-0.0396	(-0.0581, -0.0211)	-0.0375	(-0.0548, -0.0203)
Risk Ratio				
$\hat{\Psi}_{RR}$	0.1362	(0.0268, 0.6929)	0.1484	(0.0436, 0.5046)

* SuperLearner library consists of: SL.mean, SL.glm, SL.gam

Table B.4: Stochastic direct effect estimated with TMLE under a well-supported user-specified mediator distribution.

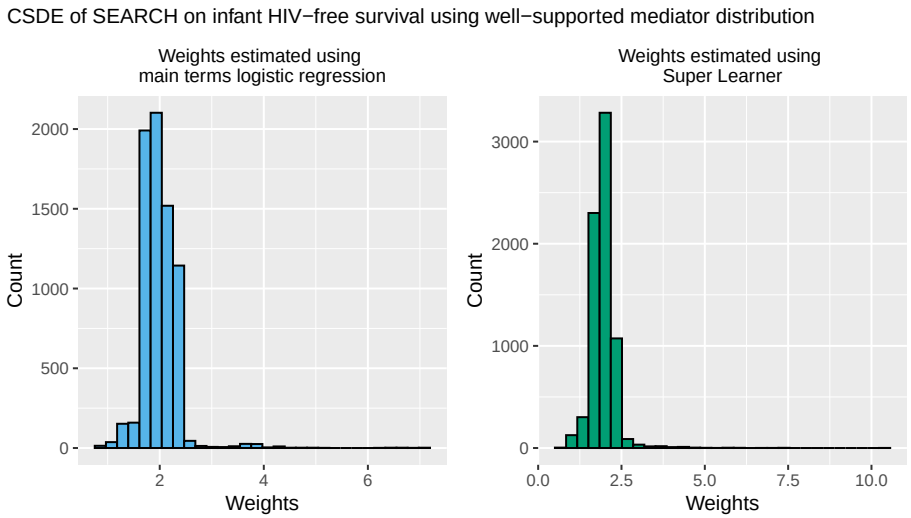
	Using main terms logistic regression		Using Super Learner*	
Perinatal HIV Transmission	Estimate	95% CI	Estimate	95% CI
A=1 (SEARCH intervention)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0007	(0.0001, 0.0034)	0.0009	(0.0003, 0.0023)
Estimated probability of live birth and infant survival	0.1233	(0.1185, 0.1282)	0.1233	(0.1185, 0.1282)
Estimated probability of infant HIV infection among infants alive at end of study	0.0055	(0.0011, 0.0273)	0.0071	(0.0027, 0.0186)
A=0 (Control)				
Estimated probability of live birth, infant survival, and infant HIV infection at end of study	0.0056	(0.0042, 0.0076)	0.0054	(0.0039, 0.0073)
Estimated probability of live birth and infant survival	0.1233	(0.1185, 0.1282)	0.1233	(0.1185, 0.1282)
Estimated probability of infant HIV infection among infants alive at end of study	0.0456	(0.0338, 0.0616)	0.0436	(0.0321, 0.0592)
Risk Difference				
$\hat{\Psi}_{RD}$	-0.0401	(-0.0565, -0.0238)	-0.0364	(-0.0515, -0.0214)
Risk Ratio				
$\hat{\Psi}_{RR}$	0.1205	(0.0235, 0.6168)	0.1638	(0.0598, 0.4486)

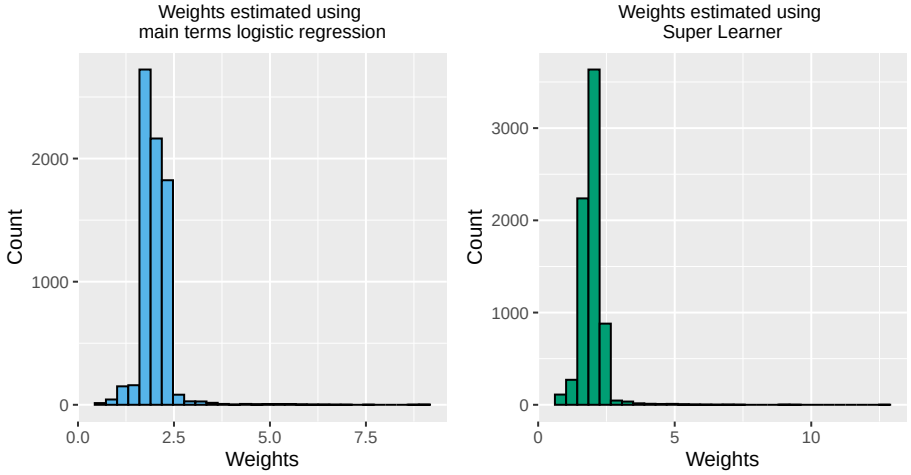
* SuperLearner library consists of: SL.mean, SL.glm, SL.gam

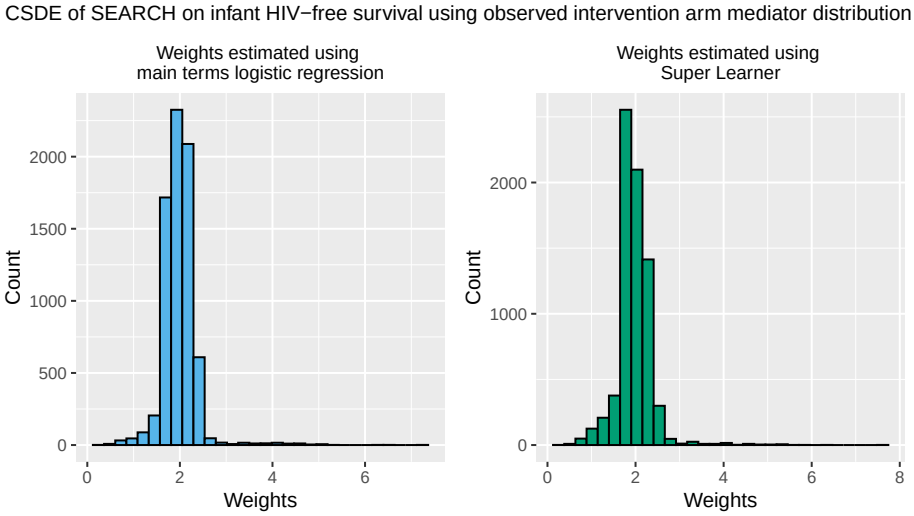
Histograms of estimated weights (probability of treatment and mediators) for counterfactual stratum direct effects

Below are histograms of the estimated weights used in the estimation of the numerator of the counterfactual stratum direct effect parameters for both infant outcomes.

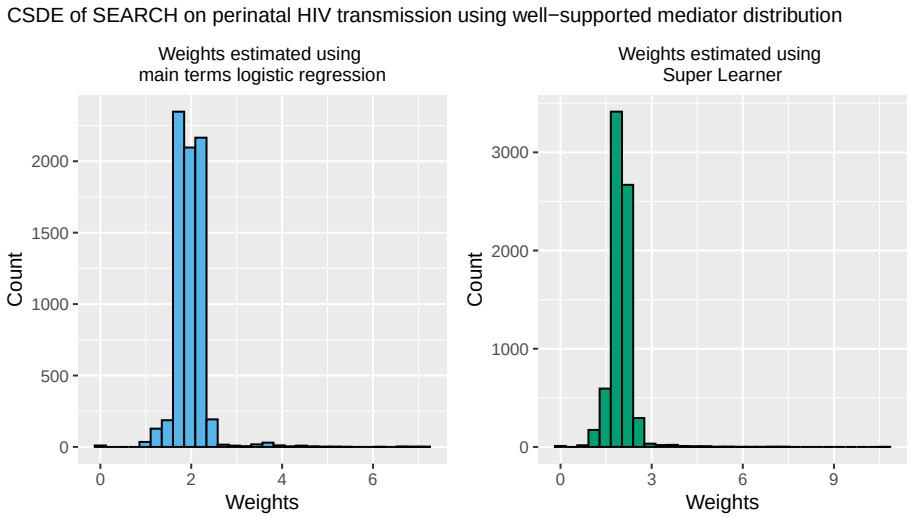
Infant HIV-free survival

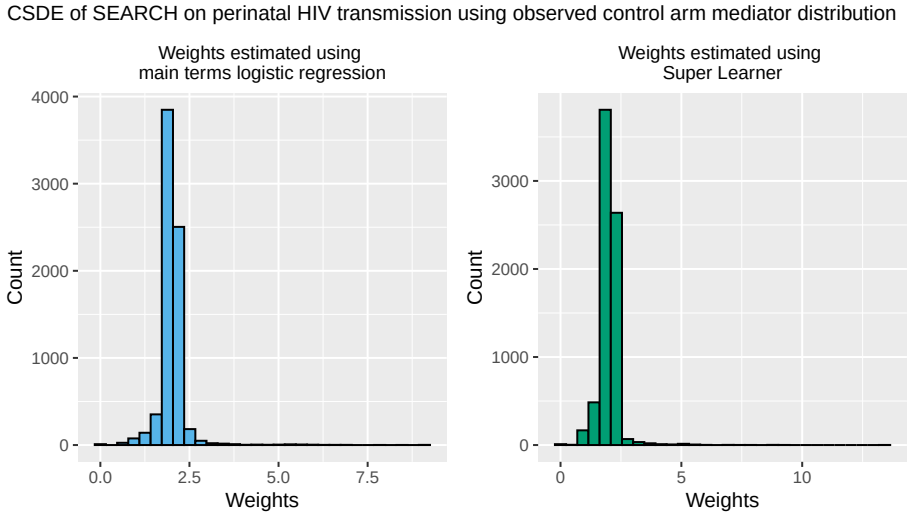


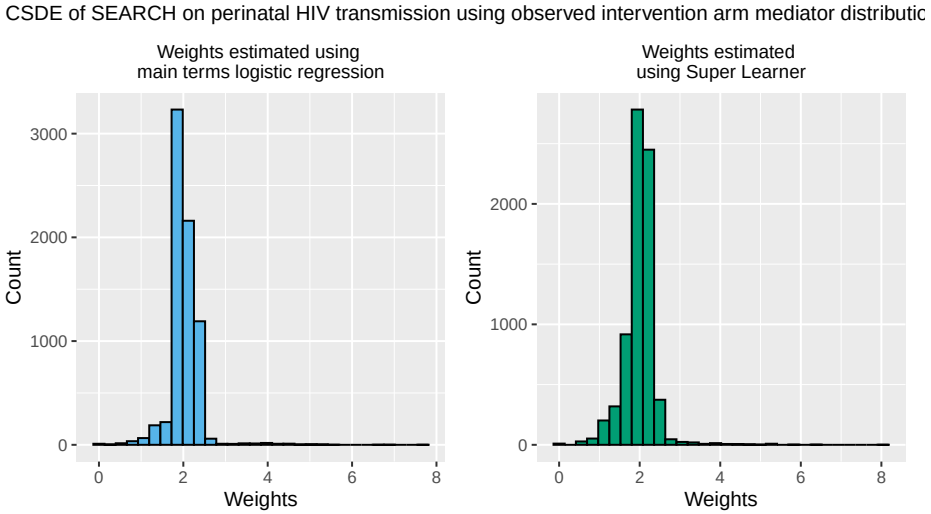




Perinatal HIV transmission




CSDE of SEARCH on perinatal HIV transmission using observed control arm mediator distribution
Weights estimated using
main terms logistic regression
Weights estimated using
Super Learner
Count
Weights
4000
3000
2000
1000
0
0.0
2.5
5.0
7.5
0
5
10


CSDE of SEARCH on perinatal HIV transmission using observed intervention arm mediator distributic
Weights estimated using
main terms logistic regression
Weights estimated
using Super Learner
Count
Weights
3000
2000
1000
0
0
2
4
6
8