Research Article

Michael P. Fay* and Dean A. Follmann

Mediation analyses for the effect of antibodies in vaccination

https://doi.org/10.1515/jci-2022-0043
received June 30, 2022; accepted February 14, 2024

Abstract: We review standard mediation assumptions as they apply to identifying antibody effects in a randomized vaccine trial and propose new study designs to allow the identification of an estimand that was previously unidentifiable. For these mediation analyses, we partition the total ratio effect (one minus the vaccine effect) from a randomized vaccine trial into indirect (effects through antibodies) and direct effects (other effects). Identifying λ, the proportion of the total effect due to an indirect effect, depends on a cross-world quantity, the potential outcome among vaccinated individuals with antibody levels as if given placebo, or *vice versa*. We review assumptions for identifying λ and show that there are two versions of λ, unless the effect of adding antibodies to the placebo arm is equal in magnitude to the effect of subtracting antibodies from the vaccine arm. We focus on the case when individuals in the placebo arm are unlikely to have the needed antibodies. In that case, if a standard assumption (given confounders the potential mediators and potential outcomes are independent) is true, only one version of λ is identifiable, and if not neither is identifiable. We propose alternatives for identifying the other version of λ, using experimental design to identify a formerly cross-world quantity. Two alternative experimental designs use a three-arm trial with the extra arm being passive immunization (administering monoclonal antibodies), with or without closeout vaccination. Another alternative is to combine information from a placebo-controlled vaccine trial with a placebo-controlled passive immunization trial.

Keywords: controlled vaccine efficacy, correlates of protection, identifiability, indirect effect, mediation assumptions, sequential ignorability

MSC 2020: 62D99, 62P10

1 Introduction

Our goal is identifying the proportion of the vaccine effect from a randomized placebo-controlled vaccine trial that is due to the antibodies induced by the vaccine, say λ. In the mediation literature, the total effect is partitioned into an indirect effect that acts through the mediator (e.g., the vaccine effect that acts through the antibody response) and a direct effect that acts directly on the outcome (e.g., the rest of the vaccine effect).[1] Mediation analyses typically make certain positivity and independence assumptions in order to identify

[1] Unfortunately, in the vaccine literature, the term indirect vaccine effect is used to describe another issue: the protective vaccine effect for a non-vaccinated individual due to nearby vaccinated individuals being less likely to be infectious, see, e.g., Halloran et al. [1], Section 2.8. This article is not about that type of indirect vaccine effect.

*** Corresponding author: Michael P. Fay,** Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Rockville, Maryland, United States of America, e-mail: mfay@niaid.nih.gov
Dean A. Follmann: Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Rockville, Maryland, United States of America

indirect and direct effects (see VanderWeele [2] or Section 3.1). This article explores those assumptions in detail, specifically focusing on the application to this antibody/vaccine problem. For ease of exposition, unless stated otherwise, the term antibodies refers to only the particular type of antibody induced by the experimental vaccine. Also for simplicity, we focus on the placebo-controlled vaccine trial, but many of the same issues may apply when another type of control arm is used instead of placebo (e.g., an arm that is given a control vaccine for another disease). Our focus is on the case when individuals in the placebo (or other control) arm do not have detectable levels of the antibody of interest, which is often the case for new or rare diseases.

The antibody/vaccine mediation problem is important as seen by its application to Coronavirus disease (COVID-19) vaccination. Large placebo-controlled vaccine trials showed impressive vaccine efficacies during the stages of the pandemic before the omicron variant was prevalent. It would be useful to be able to evaluate modifications of the vaccines (e.g., dose changes, adjuvant changes, updating the antigen to current variants) under new populations and new exposures without having to re-run such large trials. Instead, we could use antibodies to help predict vaccine efficacy (VE). Gilbert et al. [3] comprehensively explored the effects of antibodies on the VE of the Moderna (mRNA-1273) COVID-19 vaccine, including a mediation analysis using methods described in Benkeser et al. [4].

However, Benkeser et al. [4] focused on details of estimation and only briefly mentioned standard mediation assumptions; in this study, we focus on identifiability and more explicitly discuss the implications of the standard assumptions to the vaccine/antibody mediation problem. The major difficulty of this identifiability problem is two cross-world quantities: the potential outcome of a vaccinee but with mediator values as if the vaccinee had gotten placebo, or *vice versa*. These two quantities lead to two versions of indirect effects, which we call the *subtracting* indirect effect (due to subtracting the antibody effect from the vaccinees) or the *adding* indirect effect (due to adding antibody effect to the placebo participants). (In the mediation literature, these are called the *total* indirect effect and *pure* indirect effect, respectively [5].) We explore the typical mediation assumptions except without the often untenable assumption that the placebo participants have detectable antibodies specific to the infectious agent. We show that the subtracting indirect effect and λ_s (its version of λ) are identifiable, but the adding indirect effect and λ_a (its version of λ) are not. We further show that a key assumption for identifying the subtracting indirect effect is that the potential outcomes are independent of the potential mediator responses conditional on confounders. We show that in a simple case without confounders, the identification of λ_s by that assumption is equivalent to identifying a correlation of two potential random variables as 0 by assuming their independence.

Some major contributions of this study are new ways to identify the adding indirect effect by supplementing the randomized placebo-controlled vaccine trial with extra experimental information. This can be performed by adding a randomized passive immunization arm to the randomized placebo-controlled vaccine trial, where passive immunization is performed by infusing (or injecting) individuals with monoclonal antibodies. We discuss identifiability in that three-arm trial, with and without a closeout vaccination in the passive immunization arm. Alternatively, we can combine the placebo-controlled vaccine trial with a second randomized trial of passive immunization versus placebo to identify the adding indirect effect. We first identify controlled vaccine effects from the placebo versus vaccine trial and then the controlled protective effects (which act like adding indirect effects) from the placebo versus passive immunization trial. The latter effect does not require the conditional independence assumption between potential outcomes and mediators, since that independence can be met by randomization to antibody values in the passive immunization arm.

Because this article is partially about explaining mediation analysis to vaccine researchers and vaccine trials to mediation experts, we necessarily give extensive background on the causal effects, vaccine biology, and mediation definitions (Section 2). Section 3 details typical mediation assumptions applied to this problem giving explicit identifiability results. The binary mediator problem is addressed in Section 4 to give some intuition about the assumptions. We describe identifying the adding indirect effect by supplementing the main placebo-controlled vaccine trial with a passive immunization arm or experiment in Section 4.4 (binary mediator case) and Section 5 (general mediator case).

2 Background

2.1 Causal vaccine effects

Vaccines work by exposing the vaccinee to an antigen, a protein on an outward facing part of the infectious agent. The adaptive immune system of the vaccinee then produces activated antigen-specific B cells and helper T cells. Those cells are part of the germinal center that creates more antigen-specific B cells, which in turn develop into either antibody-producing plasma cells or other types of cells such as memory B cells. The plasma cells create antibodies specific to that antigen that travel to other areas of the body through the blood. These antigen-specific antibodies are usually detected from assays performed on blood samples, and those assays are of different types (e.g., an ELISA where the antibodies bind to proteins in a plate, or a functional assay that measures how much the antibody neutralizes the agent). In this article, we will often refer to the mediator as "antibodies," and a more precise description would be the measured level of those antigen-specific antibodies collected at a specific study time from a blood sample using one assay. Although the antibodies measured in blood are one product of the vaccination, some of those antigen-specific antibodies may be in other parts of the body (e.g., mucosal surfaces) and will not be detected by a blood sample and hence would not be measured as the mediator. Furthermore, there are many other responses to the vaccine that may lead to other mechanisms that may be helpful in protecting the vaccinee from developing disease in response to a future exposure from the infectious agent (for more details, see, e.g., [6]).

Vaccination is a way of preparing the immune system to be ready for future exposures. Vaccination is often especially helpful for individuals who have not previously been exposed, since those who have previously been exposed and survived will often have developed natural immunity. The gold standard for estimating VE is a controlled randomized clinical trial. Prior to the availability of approved vaccines (e.g., for human immunodeficiency virus (HIV), or in the early stages of the COVID-19 pandemic), the primary efficacy analysis from vaccine randomized trials typically exclude individuals that already have antibodies protective against the infectious agent of interest, since that population does not have the greatest need of vaccination. Furthermore, because it takes time for the vaccinee to develop antibodies, vaccine trials also typically exclude individuals from the primary efficacy analysis data set that became diseased before the time when the peak antibody level is expected to be reached, usually 1 to 4 weeks after the last dose of the vaccination (see e.g., [7], Table 1).

2.2 VE using potential outcomes

We introduce formal causal notation, first for VE, then for mediation. Because this article will focus on identifiability not estimation, we do not need subscripts to differentiate individuals in the population. For example, we let $A = 1$ if the ith participant is randomized to vaccine, and $A = 0$ if they are randomized to placebo, and no i subscript is needed. Thus, A represents the allocated arm of a typical (i.e., randomly selected) individual. Let $Y = 1$ if the typical participant gets the event (e.g., has confirmed disease) during the time they are at risk during the study, and $Y = 0$ if they do not. Write the potential outcomes for the typical participant as both Y_0, the outcome if they were randomized to placebo, and Y_1, the outcome if they were randomized to vaccine. Although we can only observe one of the two potential outcomes per individual, we can compare the expected response in both arms, because by randomization, each individual has a positive and known probability of being in either of the arms.

VE is defined as 1 minus a ratio effect (e.g., incidence ratio, hazard ratio, odds ratio) (see e.g., [1], Table 2.2), and in this article, we define it as, $VE = 1 - E(Y_1)/E(Y_0)$. For vaccine mediation, sometimes odds ratios are used (see, e.g., [8,9]). Fortunately, when only a small percentage of the placeboes have the event, there is little difference between the different ratios [10]. The VE estimand we use is reasonable in a randomized trial because the exposure processes are approximately balanced between the two arms [11]. For example, although exposure may change by geographic area, the distributions of geographic areas by arm are balanced by randomization. Similarly, although exposure will vary from individual to individual because of differential calendar times joining the study or censoring times leaving the study, those differences will be approximately equally distributed in the two arms as long as the rate of infection is not large, and the study entry times and censoring times are independent of the potential outcomes.

Unlike a difference effect (e.g., $E(Y_1) - E(Y_0)$), a ratio effect is relatively invariant to differences in exposure, so it is useful for a vaccine trial since we can rarely predict well how many participants will be exposed during the course of the trial. To see this invariance, we reexpress the potential outcomes. Let Y_a^* be the indicator of whether the typical participant would obtain the event when they were in arm a if they were exposed ($Y_\mathrm{a}^* = 1$ is yes, $Y_\mathrm{a}^* = 0$ is no). Let $Z = 1$ represent if they were exposed during the study (1 = yes, 0 = no), so that $Y_\mathrm{a} = ZY_\mathrm{a}^*$. If we assume the exposure is independent of whether they would be protected if exposed (which is a reasonable assumption in a blinded randomized trial), then

$$VE = 1 - \frac{E(Y_1)}{E(Y_0)} = 1 - \frac{E(ZY_1^*)}{E(ZY_0^*)} = 1 - \frac{E(Z)E(Y_1^*)}{E(Z)E(Y_0^*)} = 1 - \frac{E(Y_1^*)}{E(Y_0^*)}. \quad (1)$$

The VE does not depend on the exposure rate, but only on the differential in the potential to be protected given exposed, which is the scientific effect of interest.

2.3 Defining direct and indirect effects

Now, consider mediation. Let M_a be the potential mediator for a typical individual when $A = a$. In our example, M_1 is the antibody level (e.g., IC50 titer at day 29 after vaccination) if randomized to vaccine, and M_0 that level if randomized to placebo. Just as for the outcomes, we can only observe one of the potential antibody levels for each individual. To explicitly denote the effect of the antibody on the outcome, we now write the outcome for the typical individual with $A = a$ as Y_{aM_a}. More generally, we let Y_am be a potential outcome variable, where A is set to a and M is set to m (and m does not need to equal M_a). This notation implies that only the values of a and m matter in the response, not the manner in which those values are assigned (e.g., if the mediator response for individual i is m, it does not matter that the mediator value of m occurred naturally in response to a vaccine or placebo (i.e., $M_\mathrm{a} = m$), or was it set to m by some other external intervention). This is known as the consistency assumption [12], and we will assume that for now, but revisit it later. Under consistency, the observed response, Y, is $Y = Y_{AM_A}$ and the observed mediator, M, is $M = M_A$. In this notation,

$$VE = 1 - \theta_\mathrm{T} = 1 - \frac{E[Y_{1M_1}]}{E[Y_{0M_0}]}, \quad (2)$$

where θ_T represents the total ratio effect.

Typically, mediation effects are defined as differences (see, e.g., [13]), but for vaccines, we define mediation effects in terms of ratio effects. We partition the total ratio effect (θ_T) into the product of an indirect ratio effect (θ_I) and a direct ratio effect (θ_D). Let $\theta_\mathrm{I} = \theta_\mathrm{T}^{\lambda}$ and $\theta_\mathrm{D} = \theta_\mathrm{T}^{(1-\lambda)}$, giving $\theta_\mathrm{T} = \theta_\mathrm{I}\theta_\mathrm{D}$. Thus, on the log scale $\log(\theta_\mathrm{T}) = \log(\theta_\mathrm{I}) + \log(\theta_\mathrm{D})$, indirect and direct effects are additive, and λ is the proportion of the total log-ratio effect due to indirect log-ratio effects. By algebra, $\lambda = \frac{\log(\theta_\mathrm{I})}{\log(\theta_\mathrm{T})}$. The proportion-mediated effect was defined this way in the study of Gilbert et al. [3] (see, e.g., Table S9), although its explicit form and motivation was not given.

An indirect effect is the effect of changing only the mediator while holding the rest of the effect of vaccine constant. There are two main ways to define an indirect ratio effect. The first way is $\theta_{\mathrm{I_a}} = \frac{E[Y_{0M_1}]}{E[Y_{0M_0}]}$, which measures the effect on the placebo arm (denominator) of changing the mediator to the value it would have after vaccination but without actually vaccinating participants (numerator). The “a” subscript denotes “adding” antibody to unvaccinated participants at the value they would have gotten if vaccinated (i.e., M_1). The associated direct effect is part of the total effect that does not go through the mediator, $\theta_{\mathrm{D_a}} \equiv \frac{\theta_\mathrm{T}}{\theta_{\mathrm{Ia}}} = \frac{E[Y_{1M_1}]}{E[Y_{0M_1}]}$. Sections 4.4 and 5 focus on the identification of $\theta_{\mathrm{I_a}}$. The second way to define an indirect effect is $\theta_{\mathrm{I_s}} = \frac{E[Y_{1M_1}]}{E[Y_{1M_0}]}$, which measures the effect on the vaccine arm (numerator) of changing the mediator to the value it would have if the individual had been in the placebo arm (denominator). The "s" subscript denotes "subtracting" the extra antibodies in the vaccinated participants, so that the antibody levels are what would have been seen in the placebo arm (i.e., M_0). The associated direct effect is $\theta_{\mathrm{D_s}} \equiv \frac{\theta_\mathrm{T}}{\theta_{\mathrm{I_s}}} = \frac{E[Y_{1M_0}]}{E[Y_{0M_0}]}$. The mediation analysis in Gilbert et al. [3] estimated $\theta_{\mathrm{I_s}}$ and its associated λ value. Figure 1 shows how the total ratio effect can be partitioned these two ways:

$$\theta_\mathrm{T} = \theta_{\mathrm{I_a}}\theta_{\mathrm{D}_a} = \theta_{\mathrm{I_s}}\theta_{\mathrm{D_s}}. \tag{3}$$

Let λ_a be λ under the first partition, and λ_s be its value under the second.

Robins and Greenland [5] used different terminology: $\theta_{\mathrm{I_a}}$ is the *pure* indirect effect, $\theta_{\mathrm{D_a}}$ is the *total* direct effect, $\theta_{\mathrm{I_s}}$ is the *total* indirect effect, and $\theta_{\mathrm{D_s}}$ is the *pure* direct effect. Alternatively to equation (3), we can partition the total effect into three parts: pure indirect effect, the pure direct effect, and an interaction effect (say, ξ), $\theta_\mathrm{T} = \theta_{\mathrm{I_a}}\theta_{\mathrm{D_s}}\xi$, where $\xi = \frac{E[Y_{1M_1}]E[Y_{0M_0}]}{E[Y_{1M_0}]E[Y_{0M_1}]}$. If $\xi = 1$, then there is no interaction, $\theta_{\mathrm{I_a}} = \theta_{\mathrm{I_s}}$ and $\theta_{\mathrm{D}_a} = \theta_{\mathrm{D_s}}$, the dotted quadrilateral in Figure 1 is a parallelogram, and all ways of partitioning the direct and indirect ratio effects are the same. There are many other ways to define interaction (see, e.g., [2], Section 7.6).

3 Mediation analysis using the sequential ignorability assumptions

3.1 Sequential ignorability assumptions

In estimating direct or indirect effects, the difficult parameters to identify are either $E[Y_{0M_1}]$ (for identifying $\theta_{\mathrm{I_a}}$) or $E[Y_{1M_0}]$ (for identifying $\theta_{\mathrm{I_s}}$), because we observe no participant with those kinds of responses. Those responses are called "cross-world" quantities because the intervention is in one world (e.g., where the

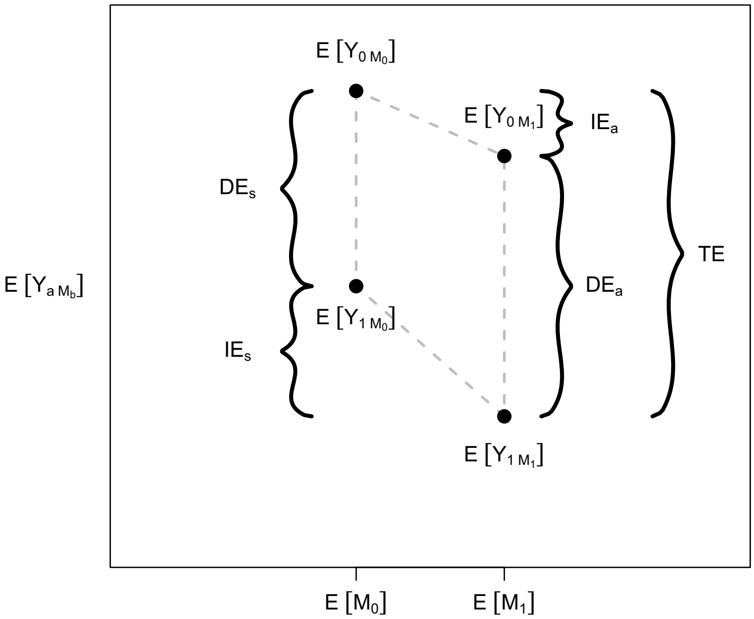

Figure 1: Graphic of different ways of partitioning the total effect (TE) into an indirect effect (IE) and direct effect (DE). The vertical axis is on the log scale for ratio effects, and on the arithmetic scale for difference effects. Moving southeast from $E[Y_{0M_0}]$ to $E[Y_{0M_1}]$ corresponds to adding in antibodies, while moving northwest from $E[Y_{1M_1}]$ to $E[Y_{1M_0}]$ corresponds to subtracting out antibodies. The Robins and Greenland [5] terminology is *pure* indirect effect (IE_a), *total* direct effect (DE_a), *total* indirect effect (IE_s), and *pure* direct effect (DE_s).

participant was allocated to arm $A = a$) but the mediator is in another world (e.g., where the participant has mediator value as if allocated to arm a', $M_{a'}$, where $a \neq a'$). Thus, in order to identify $E[Y_{1M_0}]$ or $E[Y_{0M_1}]$ (and hence, to identify λ_s and λ_a, respectively), we need to make some assumptions that are not fully testable from the study data. Pearl [14] showed that using a set of assumptions, we can identify additive direct and indirect effects using a simple equation that has become known as the mediation formula (see, e.g., [15]). In this article, since we deal with ratio effects, we focus on the piece of the mediation formula that estimates $E[Y_{1M_0}]$ or $E[Y_{0M_1}]$ (equation (4) in the following). Imai et al. [16] give two assumptions, which they called sequential ignorability assumptions, and those two assumptions are effectively a different statement of the assumptions of Pearl (see, [2], p. 200–201). In this article, we use those sequential ignorability assumptions, but for completeness we list the Pearl assumptions in Appendix S1. The situation the assumptions address is more general than the randomized trial situation, but it is important to review them since these set of assumptions are applied widely (see e.g., [2,9]), including in vaccine applications (see, e.g., Cowling et al. [8] and Gilbert et al. [3]).

Let X be a vector of baseline covariates. The sequential ignorability assumptions are as follows:

SI1: $\{Y_{\mathrm{am}}, M_{a'}\} \perp\!\!\!\perp A|X = x$ and
SI2: $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}|A = a, X = x$,

for all a, a' and m, assuming the following positivity assumptions:

PosA: $\Pr[A = a|X = x] > 0$, for all a, x,
PosM0: $\Pr[M_0 = m|A = 0, X = x] > 0$, for all x and $m \in \mathcal{S}_M$, and
PosM1: $\Pr[M_1 = m|A = 1, X = x] > 0$, for all x and $m \in \mathcal{S}_M$,

where $\mathcal{S}_M$ is the support for M.

Often the "sequential ignorability assumptions" refers to not just SI_1 and SI_2, but also to the positivity assumptions, which are implicitly assumed or listed as one positivity assumption [16]. In this study, it will be important to separate SI_1 and SI_2 (which we call the sequential ignorability assumptions) and the different positivity assumptions. Later, as suggested by a referee, we explore modifications to the supports in the positivity assumptions. Under the positivity assumptions listed earlier and the sequential ignorability assumptions (and consistency), the cross-world conditional expected potential outcomes are given by

$$E[Y_{aM_{a'}}|X = x] = \sum_m E[Y|A = a, M = m, X = x]\Pr[M = m|A = a', X = x] \tag{4}$$

(see e.g., [2], p. 465).

3.2 Application to the vaccine/antibody mediation analysis

Now, consider identifying λ_s or λ_a from a randomized placebo-controlled vaccine trial. First, the assumptions SI_1 and *PosA* are met for any randomized trial for any set of baseline variables X. In this section, we discuss the harder assumptions to justify for our scenario, which are consistency, SI_2,PosM_0, and PosM_1.

Let the support of M_0 and M_1 be $\mathcal{S}_M = \{0^*, \mathcal{S}_{\mathrm{detect}}\}$, where $M = 0^*$ represents M below the limit of detection, and $M > 0^*$ means $M \in \mathcal{S}_{\mathrm{detect}}$, and $\mathcal{S}_{\mathrm{detect}}$ is the set of possible detectable antibody values. Consider the case of a new pathogen for which it is possible that some individuals will have $M_0 = 0^*$. Then, a violation of PosM_1 occurs when the vaccine induces antibodies for all the individuals with $X = x$, so that $\Pr[M_1 = 0^*|A = 1, X = x] = 0$. A solution is to change the time of antibody measurement to sooner after vaccination so that there are some participants with $M_1 = 0^*$ within each x and we can assume that PosM_1 is not violated, and then, the resulting λ_s estimate can be used as a lower bound for the originally defined λ_s (i.e., the one using the original time of antibody measurement) under a reasonable monotonicity assumption (see e.g., [3]). Changing the timing of the antibody measurement has a downside in that it may be less associated with the response, Y_{AM_A}, since the original timing is typically designed to be at peak antibody levels.

Another problem, and a more common issue with vaccine trials, is the violation of PosM_0 so that the analysis data set will have $\Pr[M_0 > 0^*|A = 0, X = x] = 0$ for some x. In many cases, that probability is 0 for all

x. For example, we could have all $M_0 = 0^*$ for a new infectious agent, because no one would have been previously exposed. Another example is if we exclude from the study those who have antibody detectable at baseline as well as those exposed to the infectious agent before the time of measurement of the antibody (i.e., study time when M is measured). Consider equation (4) with $a = 0$ and $a' = 1$, giving the cross-world mean representing adding antibody to the unvaccinated:

$$E[Y_{0M_1}|X = x] = \sum_m E[Y|A = 0, M = m, X = x]\Pr[M = m|A = 1, X = x]. \tag{5}$$

The expression $E[Y|A = 0, M_0 = m, X = x]$ for $m > 0^*$ in equation (5) is not identifiable when $\Pr[M_0 > 0^*|A = 0, X = x] = 0$, because in that case, we do not observe Y given $A = 0$ and $M_0 = m$ for any $m > 0^*$ on any individual. When $a = 1$ and $a' = 0$, both sequential ignorability assumptions hold, and all positivity assumptions hold except for allowing $\Pr[M_0 > 0^*|A = 0, X = x] = 0$ (i.e., no detectable antibody in the placebo arm), then equation (4) can still be used (see Section S2) to give

$$E[Y_{1M_0}|X = x] = E[Y|A = 1, M = 0^*, X = x]. \tag{6}$$

The right-hand side of equation (6) is just the disease rate among vaccinated individuals with undetectable antibody among study participants with baseline $X = x$. Using equation (6), we can identify $E[Y_{1M_0}]$ as

$$E[Y_{1M_0}] = \sum_x E[Y|A = 1, M = 0^*, X = x]\Pr[X = x], \tag{7}$$

where the summation is over the possible values of X, and $E[Y|A = 1, M = 0^*, X = x]$ is estimated with the sample mean of Y among the vaccinated with undetectable antibody responses and $X = x$.

Consider the consistency assumption [12] in this antibody/vaccine mediation scenario. Suppose we define the antibody mediator as the amount of antibody measured at a particular time post vaccination, and an indirect effect is through that antibody mediator and the rest of the total effect is a direct effect. Consider the plasma cells that produce antibodies which give the antibody-mediated vaccine effects. After the antibody mediator is measured, the plasma cells (especially if they are long-lived plasma cells [see 6]) may play a role in protection by producing more antigen-specific antibodies in the future. By the strict definition, those latter antibodies would be seen as direct effects. If we wish to apply the mediation analysis results to predict the proportion of total vaccine effect due to that antibody mediator from a similar vaccine, then the consistency assumption may approximately hold because the second similar vaccine will likely also produce plasma cells that remain after the antibody is measured. In contrast, if we wish to apply the mediation results from the original vaccine to predict the proportion of the total effect due to antibodies in the case when the intervention is infusion with monoclonal antibodies, then the consistency assumption will likely be violated (i.e., the two mediation analyses will be estimating different effects). It is likely that an individual's outcome under monoclonal antibodies will be different than if they had an antibody response due to a vaccine, because the vaccine will induce plasma cells, which may continue to make antibodies, and hence have a larger direct effect. This violation may not be a problem if the amount of antibodies detected produces overwhelming protection and additional antibodies would not affect risk much. Even if the consistency assumption was violated, the mediation analysis might be useful to estimate bounds on an indirect effect if we could assume that the vaccine-induced antibodies are at least as effective as the addition of externally produced monoclonal antibodies.

Next, consider the SI_2 assumption. For this, we once again simplify the immunology. Suppose we can partition the antigen-specific antibody response into two parts: the creation of the antibodies, and the subsequent protective function of the antibodies. Some protective functions of the antibodies are [6, Table 2.1]: (i) binding to toxins produced by the infectious agent to stop their deleterious effect; (ii) stopping replication of a virus by binding to the virus and preventing the virus from entering cells of the host; (iii) opsonizing, where the antibodies mark pathogens outside cells so that other immune cells (e.g., macrophages or neutrophils) may clear them; and (iv) activating other processes (e.g., the complement cascade). The SI_2 assumption (i.e., within each arm, conditional on baseline covariates, X, the potential antibody response, $M_{a'}$, is independent on the potential outcomes for the study, Y_{am}, for all a, a', and m), implies that within levels of X, the processes that create antibodies are independent from the processes that lead to their protective functions. Because the body has evolved to have a useful immune system, it is possible that there would be a dependence between those two processes; however, there may not be a high dependence (e.g., Goel et al. [17] showed that antibody level post-boost did not have any substantial correlation with post-boost memory B-cells, the latter of which are suspected to be related to protection). If there is such a dependence, in order to meet assumption SI_2, we would need to find baseline covariates, X, such that conditional on X, the potential antibody response vector is independent of the potential outcome vector. For example, suppose the baseline covariates can be used to classify individuals into k immune classes, where some classes have healthier immune systems than others. If *within each class*, we can assume $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}$, then SI_2 would be met. Consider an example with only two classes, healthy and immunocompromised, and suppose X could be used to classify each individual into those two types. Furthermore, suppose that within both the healthy and the immunocompromised types, the process for creating antibodies is independent from the protective processes of the antibodies. Under those suppositions, SI_2 would be met.

In summary, we have detailed the main assumptions used in most, if not all, vaccine mediation analyses up to this point, including a recent important vaccine mediation analysis (see, e.g., [3,4]). We have explicitly shown that when PosM_0 fails such that $\Pr[M_0 > 0^*|A = 0] = 0$, then we can still identify the subtracting indirect effect (i.e., total indirect effect), as long as the randomization is valid and the SI_2 and PosM_1 assumptions are met. We have focused on binary potential outcomes instead of time-to-event outcomes to avoid extra complications such as the timing of the outcomes and censoring, since those extra complications are peripheral to our main focus on identifiability.

4 Binary mediators and responses

From the previous section, we showed that a key difficult assumption is SI_2. In this section, we study the simple case when the mediators are binary to gain intuition about how SI_2 is affecting the identifiability of the estimators of λ. In this section, let $M_A = 1$ if $m \geq \tau$, and $M_A = 0$ if $m < \tau$, where τ is some specified threshold. The specified threshold τ, does not need to be the limit of detection, it could alternatively be some positive value of m related to the antibody activity. To keep the exposition simple, we will ignore baseline covariates, but it is straightforward (although notationally cumbersome) to apply the models of this section separately to each level of the baseline covariates if the analogous conditional independence assumptions are appropriate.

4.1 Base model

To start, we gather together some assumptions that appear reasonable for the placebo-controlled randomized vaccine trial. As a shorthand, we call this set of assumptions the "base model." Robins and Greenland [5] studied the mediator problem with binary mediators and binary potential outcomes. Under this model, there are $2^6 = 64$ different types of possible vectors of potential mediator responses and potential outcomes (two levels for the mediator and four levels for the paired outcomes). Robins and Greenland [5] made three assumptions that reduce the 64 types to 18, which we apply and translate to our vaccine/antibody example:

RG1: Vaccination cannot block antibodies that would have been present under the placebo arm. This means we never have types where $M_1 = 0$ and $M_0 = 1$.
RG2: Vaccination cannot cause the disease. This means we never have types where $Y_{1m} = 1$ and $Y_{0m} = 0$.
RG3: Antibodies cannot cause the disease. This means we never have types where $Y_{a1} = 1$ and $Y_{a0} = 0$ for any a.

These assumptions are often called monotonicity assumptions (see, e.g., [18]). These seem reasonable assumptions for many vaccines, although there is one known case of a violation of RG_2 (a vaccine for Dengue virus for one type enhancing the probability of disease in another type [see, e.g., 19]). Additionally, RG_2 excludes cases where vaccination increases the risk of disease. For example, some individuals may change their behavior to increase their risk of infection if they suspect they have been vaccinated and believe that the vaccination works. If that behavior change results in a disease case, that would count as the vaccination causing the disease. Because of this, in vaccine trials, double-blinding is especially important for justifying RG_2. Related to that, the reactogenicity of the vaccine may unintentionally unblind some participants. To alleviate that problem, sometimes the control arm uses a vaccine for a different disease than that of the experimental vaccine.

For the vaccine example, because we exclude individuals that had antibodies at baseline, and because we exclude anyone with detectable disease before the antibody is measured, it is very unlikely that anyone left in the analysis data set in the placebo arm will have a positive antibody response. So we assume that we never have types where $M_0 = 1$. This reduces the types of interest to 12.

These types are listed in Table 1, with each type described in the last column. The first column gives the notation for the proportion of the population with potential outcomes $\mathbf{Y} = [Y_{11}, Y_{10}, Y_{01}, Y_{00}]$ and potential mediators $\mathbf{M} = [M_1, M_0]$, which is $\pi^{\mathbf{M}}_{\mathbf{Y}}$. For example, π^{10}_{0101} is the proportion with $\mathbf{M} = [10]$ (i.e., $M_1 = 1$ and $M_0 = 0$) and $\mathbf{Y} = [0101]$ (i.e., $Y_{11} = 0$, $Y_{10} = 1$, $Y_{01} = 0$, and $Y_{00} = 1$). For a person with that type, Y_{10} and Y_{01} are the cross-world potential outcomes.

Under a randomized trial, A is independent of both the potential outcome vector (i.e., $A \perp\!\!\!\perp \mathbf{Y}$) and the potential mediator response vector (i.e., $A \perp\!\!\!\perp \mathbf{M}$). Altogether, the assumptions for the *base* model described in this section are RG_1, RG_2, RG_3, $\Pr[M_0 = 1] = 0$, $A \perp\!\!\!\perp \mathbf{Y}$, and $A \perp\!\!\!\perp \mathbf{M}$. We feel that often these assumptions will be quite reasonable.

To find identifiable parameters, we first re-express the 12 proportions of Table 1 as Table 2, which lists the 12 proportions in a 4 × 2 table with observable margins, and a 4 × 6 table with one of the observable columns partitioned into five sub-columns. The six observable (and hence identifiable) margins are denoted with ϕ_{amy} parameters, where the *amy* subscripts represent, respectively, A (v = vaccine, p = placebo), M_A (a = antibody response positive, n = no antibody response), and Y_{AM_A} (f = failure [had disease], s = success [did not have disease]). For example, ϕ_{vaf} is the proportion of the population that if randomized to **v**accine would produce **a**ntibodies and have an event (i.e., would **f**ail). Recall, in the base model, $\Pr[M_0 = 1] = 0$ so $\phi_{pas} = 0$ and $\phi_{paf} = 0$ and are not listed in Table 2.

To be identifiable under the base model is to be able to express a parameter in terms of the ϕ parameters. From Table 2, we see that both $E(Y_{1M_1})$ and $E(Y_{0M_0})$ are identifiable:

Table 1: Twelve types left after excluding types based on assumptions RG_1, RG_2, and RG_3, and those with $M_0 = 1$

$\pi^{M_1M_0}_{Y_{11}Y_{10}Y_{01}Y_{00}}$	M_1	M_0	Y_{11}	Y_{10}	Y_{01}	Y_{00}	Y_{1M_1}	Y_{1M_0}	Y_{0M_1}	Y_{0M_0}	Type description	
π^{00}_{0000}	0	0	**0**	0	**0**	0	0	0	0	0	Ab nonresp.	uninfectable
π^{00}_{0001}	0	0	**0**	0	**0**	1	0	0	1	1	Ab nonresp.	either alone prot.
π^{00}_{0011}	0	0	**0**	0	**1**	1	0	0	1	1	Ab nonresp.	direct alone prot.
π^{00}_{0101}	0	0	**0**	1	**0**	1	1	1	1	1	Ab nonresp.	Ab alone prot.
π^{00}_{0111}	0	0	**0**	1	**1**	1	1	1	1	1	Ab nonresp.	need both
π^{00}_{1111}	0	0	**1**	1	**1**	1	1	1	1	1	Ab nonresp.	totally doomed
π^{10}_{0000}	1	0	0	**0**	**0**	0	0	**0**	**0**	0	Ab resp.	uninfectable
π^{10}_{0001}	1	0	0	**0**	**0**	1	0	**0**	**0**	1	Ab resp.	either alone prot.
π^{10}_{0011}	1	0	0	**0**	**1**	1	0	**0**	**1**	1	Ab resp.	direct alone prot.
π^{10}_{0101}	1	0	0	**1**	**0**	1	0	**1**	**0**	1	Ab resp.	Ab alone prot.
π^{10}_{0111}	1	0	0	**1**	**1**	1	0	**1**	**1**	1	Ab resp.	need both
π^{10}_{1111}	1	0	1	**1**	**1**	1	1	**1**	**1**	1	Ab resp.	totally doomed

The bold values are not observable in the randomized trial (i.e., are cross-world counterfactuals). The notation for the proportion of the population for each type is $\pi^{\mathbf{M}}_{\mathbf{Y}}$, where $\mathbf{M} = [M_1, M_0]$ and $\mathbf{Y} = [Y_{11}, Y_{10}, Y_{01}, Y_{00}]$. In the description, the “either” or ”both” refer to the antibody (Ab) and/or the direct effect, prot. = protected and resp. = responder.

Table 2: Proportions of the 12 types of responses in the base model and the marginal combinations of them

	$\boldsymbol{A = 0, M_0 = 0,}$	$\boldsymbol{A = 0, M_0 = 0, Y_{00} = 1}$					
	$\boldsymbol{Y_{00} = 0}$						**Observable marginal**
	[0000]	**[0001]**	**[0011]**	**[0101]**	**[0111]**	**[1111]**	
$A = 1, M_1 = 0, Y_{10} = 0$	π^{00}_{0000}	π^{00}_{0001}	π^{00}_{0011}				ϕ_{vns}
$A = 1, M_1 = 0, Y_{10} = 1$				π^{00}_{0101}	π^{00}_{0111}	π^{00}_{1111}	ϕ_{vnf}
$A = 1, M_1 = 1, Y_{11} = 0$	π^{10}_{0000}	π^{10}_{0001}	π^{10}_{0011}	π^{10}_{0101}	π^{10}_{0111}		ϕ_{vas}
$A = 1, M_1 = 1, Y_{11} = 1$						π^{10}_{1111}	ϕ_{vaf}
observable marginal	ϕ_{pns}			ϕ_{pnf}			1

Let $\mathbf{M} = [M_1, M_0]$ and $\mathbf{Y} = [Y_{11}, Y_{10}, Y_{01}, Y_{00}]$. The proportions are written as $\pi^{\mathbf{M}}_{\mathbf{Y}}$. For example, π^{10}_{0101} is the proportion of the population with potential mediators equal to $M_1 = 1$ and $M_0 = 0$, and potential outcomes equal to $Y_{11} = 0$, $Y_{10} = 1$, $Y_{01} = 0$, and $Y_{00} = 1$. We only observe six marginals in Table 2 (4 from the treatment arm [$A = 1$], and 2 from the control arm [$A = 0$]), so those combinations of types are identifiable, as well as certain functions of them (e.g., the sum of the marginals for the first two rows is $\Pr[M_1 = 0]$, which is identifiable). The observable marginals are written as ϕ_{amy}, where the $_{\text{amy}}$ subscripts refer to, respectively, A ($a = v$ is **v**accine arm, $a = p$ is **p**lacebo arm), M_A ($m = a$ is detectable **a**ntibodies, $m = n$ is **n**o detectable antibodies), and Y_{AM_A} ($y = f$ is **f**ailure, $y = s$ is **s**uccess).

$$\begin{aligned} E[Y_{1M_1}] &= \phi_{vaf} + \phi_{vnf}, \\ E[Y_{0M_0}] &= \phi_{pnf}. \end{aligned} \tag{8}$$

On the other hand, the cross-world expectations are not identifiable without further assumptions, so are given with a combination of identifiable and other parameters (Tables 1 and 2):

$$E[Y_{1M_0}] = \phi_{vaf} + \phi_{vnf} + \pi^{10}_{0101} + \pi^{10}_{0111} = E[Y_{1M_1}] + \pi^{10}_{0101} + \pi^{10}_{0111}, \tag{9}$$

$$E[Y_{0M_1}] = \phi_{vaf} + \phi_{vnf} + \pi^{00}_{00\cdot 1} + \pi^{10}_{0011} + \pi^{10}_{0111} = E[Y_{1M_1}] + \pi^{00}_{00\cdot 1} + \pi^{10}_{0011} + \pi^{10}_{0111}, \tag{10}$$

where $\pi^{00}_{00\cdot 1} = \pi^{00}_{0001} + \pi^{00}_{0011}$, and the “·” denotes summations over an index. From equation (9), we see that to identify $E[Y_{1M_0}]$, we need to be able to identify $\pi^{10}_{0101} + \pi^{10}_{0111}$. By inspection of the definition of the ϕ parameters in Table 2, we see that no combination of the ϕ parameters are able to identify $\pi^{10}_{0101} + \pi^{10}_{0111}$ (see third row, second column). Therefore, $E[Y_{1M_0}]$ is not identifiable from the base model. Furthermore, under the base model, the proportion of the total effect due to an indirect effect, λ, is not identifiable. We formally show this in Theorem 1.

Theorem 1. *Any* $\lambda_{\text{s}} \in [0, 1]$ *is compatible with the base model. Similarly, any* $\lambda_{\text{a}} \in [0, 1]$ *is compatible with the base model.*

The proof is in Section S3. Theorem 1 shows that the base model is inadequate to obtain any information about λ (either λ_{s} or λ_{a}) from the data; more assumptions are needed when λ is the interest.

4.2 Model with potential mediators independent of potential outcomes

We previously discussed in Section 3 the sequential ignorability assumptions, and how they may be applied even when $\Pr[M_0 = 1|A = 0] = 0$ to estimate $E[Y_{1M_0}]$ (see equation (7), which additionally adjusts for baseline covariates). The base model essentially already includes SI_1, and now, we explore adding SI_2 to the base model. In Section 3.2, we discussed why it is difficult to justify SI_2 in the vaccine/antibody mediation analysis; nevertheless, because the sequentially ignorability assumptions are common and allow some identifiability, we discuss them here. In Section 4, we have simplified the exposition, leaving off the baseline variables, and write the independence expression in SI_2 without explicitly conditioning on $A = a, X = x$ as $Y_{\text{am}} \perp\!\!\!\perp M_{a'}$; it means that each mediator potential response is independent of each potential outcome. Let $\mathcal{M}_2$ represent the model defined as the base model with the added assumption from SI_2 that $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}$.

Theorem 2. *Under $\mathcal{M}_2$, then*

(*i*) π^{00}_{0000},$\pi^{00}_{00\cdot1}$, $\pi^{00}_{\cdot1\cdot1}$, π^{10}_{0000}, $\pi^{10}_{0\cdot\cdot1}$, *and* π^{10}_{1111} *are identifiable, with* $\pi^{00}_{0000} = \phi_{pns}\phi_{vn}$, $\pi^{00}_{00\cdot1} = \phi_{vns} - \phi_{pns}\phi_{vn}$, $\pi^{00}_{\cdot1\cdot1} = \phi_{vnf}$, $\pi^{10}_{0000} = \phi_{pns}\phi_{va}$, $\pi^{00}_{0\cdot\cdot1} = \phi_{vas} - \phi_{pns}\phi_{va}$, *and* $\pi^{10}_{1111} = \phi_{vaf}$, *and* $\phi_{va} = 1 - \phi_{vn} = \phi_{vaf} + \phi_{vas}$.

(*ii*) $E[Y_{0M_1}]$ *is not identifiable, and*

(*iii*) $E[Y_{1M_0}]$ *is identifiable and equal to* $\frac{\phi_{vnf}}{\phi_{vn}}$.

The proof of Theorem 2 is in Section S4. First, note that Theorem 2(i) allows us to test for certain violations of the assumptions, because $\pi^{00}_{00\cdot1}$ and $\pi^{00}_{0\cdot\cdot1}$ must be positive, or rewriting

$$\pi^{00}_{00\cdot1} \geq 0 \Leftrightarrow \frac{\phi_{vns}}{\phi_{vn}} \geq \phi_{pns} \Leftrightarrow \frac{\phi_{vnf}}{\phi_{vn}} \leq \phi_{pnf},$$
$$\pi^{00}_{0\cdot\cdot1} \geq 0 \Leftrightarrow \frac{\phi_{vas}}{\phi_{va}} \geq \phi_{pns} \Leftrightarrow \frac{\phi_{vaf}}{\phi_{va}} \leq \phi_{pnf}.$$

Second, Theorem 2(ii) shows $E[Y_{0M_1}]$ is not identifiable, which may seem strange because the usual positivity and sequential ignorability assumptions show identifiability (see equation (4)); however, recall that $\Pr[M_0 = 1] = 0$ from the base model, which violates the PosM_0 positivity assumption (Section 3.1). A consequence is that λ_a is not identifiable. Finally, plugging in the value of $E[Y_{1M_0}]$ from Theorem 2(iii), we obtain $\theta_{\mathrm{I_s}} = \frac{E[Y_{1M_1}]}{E[Y_{1M_0}]} = \frac{\phi_{vf}\phi_{vn}}{\phi_{vnf}}$, so that

$$\lambda_\mathrm{s} = \frac{\log\left(\frac{\phi_{vf}\phi_{vn}}{\phi_{vnf}}\right)}{\log\left(\frac{\phi_{vf}}{\phi_{pnf}}\right)}. \tag{11}$$

Thus, assuming $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}$ with the base model determines λ_s. This is a red flag, because we do not want the parameter of interest to be identified entirely by an assumption that is not strongly justified by the subject matter. Below, we express different consequences of SI_2 to better critique its plausibility.

Theorem 3. *The following three statements are equivalent*:

Statement A: *Under the base model,* $E(Y_{1M_0})$ *ranges from* $E(Y_{1M_1})$ *to* $E(Y_{0M_0})$, *and if we additionally assume* $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}$, *then* $E(Y_{1M_0}) = \phi_{vnf}/\phi_{vf}$.

Statement B: *Under the base model,* λ_s *ranges from* 0 *to* 1, *and if we additionally assume* $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}$, *then* λ_s *is given by equation* (11).

Statement C: *Under the base model,* $\rho \equiv \mathrm{Corr}(Y_{1M_0}, M_1)$ *ranges from* $\rho_{\min}$ *to* $\rho_{\max}$, *and if we additionally assume* $Y_{\mathrm{am}} \perp\!\!\!\perp M_{a'}$ (*which implies the assumption that* $Y_{1M_0} \perp\!\!\!\perp M_1$), *then* $\rho = 0$, *where*

$$\rho_{\min} = \frac{\left\{\phi_{pnf} - \frac{\phi_{vnf}}{\phi_{vn}}\right\}\phi_{vn}}{\sqrt{(1-\phi_{vn})\phi_{vn}\phi_{pnf}\{1-\phi_{pnf}\}}},$$
$$\rho_{\max} = \frac{\left\{\phi_{vf} - \frac{\phi_{vnf}}{\phi_{vn}}\right\}\phi_{vn}}{\sqrt{(1-\phi_{vn})\phi_{vn}\phi_{vf}\{1-\phi_{vf}\}}},$$

$\phi_{vn} = \phi_{vns} + \phi_{vnf}$ *and* $\phi_{vf} = \phi_{vnf} + \phi_{vaf}$.

The theorem is proven in Section S5. Statement A restates Theorem 2(iii) by rewriting the conditions implied by the base model that requires $\theta_T \leq \theta_{\mathrm{I_s}} \leq 1$, i.e., that the vaccine cannot cause the event and the subtracting indirect effect cannot be harmful and is not as extreme as the total ratio effect. The usual approach to solving this problem is Statement B, which is just expressing λ_s as a function of $E(Y_{1M_1})$, $E(Y_{1M_0})$, and $E(Y_{0M_0})$. Another approach is Statement C, which is expressing $\rho = \text{Corr}(Y_{1M_0}, M_1)$ as a function of $E(Y_{1M_0})$ and some identifiable parameters, and the range for ρ required by the base model just plugs in $E(Y_{1M_1})$ and $E(Y_{0M_0})$ into the expression for ρ. The main point of Theorem 3 is that the usual approach to solving this problem (Statement B) is like trying to identify a correlation, and assuming independence of the two random variables to conclude that the correlation is 0 (Statement C). Statement C seems like assuming what we are trying to identify because independence and correlation are inherently linked, but our objective is estimating λ_s, which is really $\lambda_s(\rho)$, a complex function of ρ. The function $\lambda_s(\rho)$ may be steep or relatively flat for ρ's of interest, and the steepness depends on a given dataset. Also, $\lambda_s(0)$ depends on identifiable parameters specific to each setting (equation (11)).

The models of this section ignore adjustments for baseline covariates. If those baseline covariates are measured and control for all of the differences between the antibody production processes and the antibody protection process, then SI_2 may hold and equation (7) is justified. Those are the assumptions used in the mediation analysis for the Moderna (mRNA-1273) COVID-19 vaccine [3,4]. If only some of those baseline covariates are used but are assumed to be all of them, the resulting estimator may be improved. A problem is that it is often difficult to measure all variables needed to control for these processes and to know that SI_2 approximately holds.

4.3 Example

In Table 3, we show the results of a made-up randomized placebo-controlled vaccine trial with 10,000 participants in each arm. To keep it simple, we assume there are no baseline variables measured. We estimate VE as 90% since $\hat{\theta}_T = \frac{\hat{E}[Y\{1,M(1)\}]}{\hat{E}[Y\{0,M(0)\}]} = \frac{10\,/\,10000}{100\,/\,10000} = 0.10$. If the sequential ignorability assumptions are met and $\Pr[M_0 = 1|A = 0] = 0$, then we can use equation (6) with only 1 level of X (which is equivalent to the result of Theorem 2), to obtain, $\hat{E}[Y_{1M_0}] = \hat{E}[Y_{1M_1}|M_1 = 0] = 0.4\%$. This gives us $\hat{\theta}_{I_s} = \frac{\hat{E}[Y\{1,M(1)\}]}{\hat{E}[Y\{1,M(0)\}]} = \frac{0.1\,\%}{0.4\,\%} = 0.25$, and $\hat{\theta}_{D_s} = \frac{\hat{E}[Y\{1,M(0)\}]}{\hat{E}[Y\{0,M(0)\}]} = \frac{0.4\,\%}{1\%} = 0.40$, and $\hat{\lambda}_s = \frac{\log(\hat{\theta}_{I_s})}{\log(\hat{\theta}_T)} = 0.6021$, so that under model $\mathcal{M}_2$, we estimate that about 60.21% of the total ratio effect is mediated through the antibodies.

4.4 Three-arm trial

Of the models studied, only model $\mathcal{M}_2$ achieves identifiability of λ_s, and no model gives identifiability of λ_a. Thus, we consider an experimental solution. Consider trial with three arms: vaccine, placebo, and passive

Table 3: Made-up results from a randomized placebo-controlled vaccine trial with 10,000 participants in each arm

a	$m(a)$	$y(a, m(a))$	Number of participants	ϕ parameter	Conditional event proportions
1	1	1	2	$\hat{\phi}_{vaf} = 0.02\%$	$\hat{E}[Y_{1M_1}|M_1 = 1] = 2/8{,}000 = 0.025\%$
1	1	0	7,998	$\hat{\phi}_{vas} = 79.98\%$	
1	0	1	8	$\hat{\phi}_{vnf} = 0.08\%$	$\hat{E}[Y_{1M_1}|M_1 = 0] = 8/2{,}000 = 0.4\%$
1	0	0	1,992	$\hat{\phi}_{vns} = 19.92\%$	
0	1	1	0	$\hat{\phi}_{paf} = 0$	
0	1	0	0	$\hat{\phi}_{pas} = 0$	
0	0	1	100	$\hat{\phi}_{pnf} = 1\%$	$\hat{E}[Y_{0M_0}|M_0 = 0] = 100/10{,}000 = 1\%$
0	0	0	9,900	$\hat{\phi}_{pns} = 99\%$	

immunization, where passive immunization is infusing (or injecting) participants with antibodies made externally, such as monoclonal antibodies. We start with the assumptions of the base model of Section 4.1 and add the third arm. We make a consistency assumption about the antibodies, assuming that the antibody effects from infusion of monoclonal antibodies will be the same as if those antibodies occurred due to vaccination.

We treat that intervention ($A = 2$) as if it is a placebo with $M = 1$, so we observe Y_{01} as responses from this arm (see Table 4). We label the ϕ parameters as before except that the arm index is now one of three: "v," "p," or "i." Although we can observe Y_{01}, we do not observe M_1 in the arm with the infused mediator. Thus, we can only estimate $\phi_{if} = \phi_{iaf} + \phi_{inf}$ and $\phi_{is} = \phi_{ias} + \phi_{ins}$. Table 4 gives ϕ_{if} and ϕ_{is} in terms of π parameters, and using arguments similar to those used in Section S3, those extra ϕ parameters do not lead to identification of $E[Y_{1M_0}]$ (equation (9)) or $E[Y_{0M_1}]$ (equation (10)).

Suppose that in addition to the base model assumptions, we can perfectly predict from baseline covariates which individuals will have $M_1 = 1$. Then, in the passive immunization arm, the expected proportion of failures among those predicted to have $M_1 = 1$ will be identified as $\Pr[Y_{01} = 1 \text{ and } M_1 = 1] = \phi_{iaf} = \pi^{10}_{0011} + \pi^{10}_{0111} + \phi_{vaf}$, while in the placebo arm, the expected proportion of failure among those predicted to have $M_1 = 0$ will be identified as $\Pr[Y_{00} = 1 \text{ and } M_1 = 0] = \phi_{vnf} + \pi^{00}_{00\cdot1}$. Thus, we identify $E[Y_{0M_1}]$ from the three-arm trial, $E[Y_{0M_1}] = \sum_{m=0}^{1}\Pr[Y_{0m} = 1 \text{ and } M_1 = m]$, allowing us to identify $\theta_{\mathrm{I_a}}$ and λ_{a}. What makes this identification possible is that there are only two levels for M_1 and all individuals randomized to placebo obtain one level, while all those randomized to passive immunization obtain the other level. Identification with more than two levels for M_1 is not as straightforward.

5 Nonbinary mediators and adding passive immunity experiments

In this section, we again allow a more general antibody response such that the support of M_1, $\mathcal{S}_{M_1}$, contains both undetectable (i.e., $M_1 = 0^*$) and any of a set of detectable antibody levels ($M_1 > 0^*$). We continue to assume (similar to the base model of the binary case) that $\Pr[M_0 = 0^*] = 1$. We explore the identification of λ_{a} from

Table 4: Ten strata (four strata in the vaccine arm, four in the placebo arm, and two in the passive immunization arm), where the fifth and sixth rows represent strata that are not observed when $M_0 = 0$ always (as in the base model)

Description	ϕ	π	A	M_A	Y_{AM_A}	M_1	M_0	Y_{11}	Y_{10}	Y_{01}	Y_{00}
Vaccine/Abs/failure	ϕ_{vaf}	π^{10}_{1111}	1	1	1	1	·	1	·	·	·
Vaccine/Abs/success	ϕ_{vas}	$\pi^{10}_{0000} + \pi^{10}_{0\cdot\cdot1}$	1	1	0	1	·	0	·	·	·
Vaccine/no Abs/failure	ϕ_{vnf}	$\pi^{00}_{\cdot1\cdot1}$	1	0	1	0	·	·	1	·	·
Vaccine/no Abs/success	ϕ_{vns}	$\pi^{00}_{0000} + \pi^{00}_{00\cdot1}$	1	0	0	0	·	·	0	·	·
Placebo/Abs/failure	ϕ_{paf}	0	0	1	1	·	1	·	·	1	·
Placebo/Abs/success	ϕ_{pas}	0	0	1	0	·	1	·	·	0	·
Placebo/no Abs/failure	ϕ_{pnf}	$\pi^{00}_{00\cdot1} + \pi^{00}_{\cdot1\cdot1} + \pi^{10}_{0\cdot\cdot1} + \pi^{10}_{1111}$	0	0	1	·	0	·	·	·	1
Placebo/no Abs/success	ϕ_{pns}	$\pi^{00}_{0000} + \pi^{10}_{0000}$	0	0	0	·	0	·	·	·	0
Passive Imm/failure	ϕ_{if}	$\pi^{00}_{0011} + \pi^{00}_{0111} + \pi^{00}_{1111}+$ $\pi^{10}_{0011} + \pi^{10}_{0111} + \pi^{10}_{1111}$	2	·	1	·	·	·	·	1	·
Passive Imm/success	ϕ_{is}	$\pi^{00}_{0000} + \pi^{00}_{0001} + \pi^{00}_{0101}+$ $\pi^{10}_{0000} + \pi^{10}_{0001} + \pi^{10}_{0101}$	2	·	0	·	·	·	·	0	·

The last six columns represent the observed values (A, M_A, Y_{AM_A}) put in the appropriate column, with "·" representing unobserved potential outcomes.

several different experimental designs, all of which allow three types of intervention, vaccine, placebo, and passive immunization.

5.1 Three-arm trial

Consider a randomized vaccine trial with three arms: placebo, vaccine, and passive immunization. We denote the three arms as p, v, and i, with the associated values of A equal to 0, 1 and 2, and similarly for antibody potential responses (M_0, M_1, and M_2) and potential outcomes (Y_{0m}, Y_{1m}, and Y_{2m}).

We start by reviewing some results in Gilbert et al. [20], who studied controlled VE (CVE) from a placebo-controlled randomized vaccine trial, and also considered the case when $\Pr[M_0 = 0^*] = 1$. Gilbert et al. [20] defined the CVE, which in our notation (i.e., when Y_{am} is binary) is $\mathrm{CVE}(m_1, m_0) = 1 - \frac{E[Y_{1m_1}]}{E[Y_{0m_0}]}$. When $\Pr[M_0 = 0^*] = 1$, we write

$$\mathrm{CVE}(m, 0^*) \equiv \mathrm{CVE}(m) = 1 - \frac{E[Y_{1m}]}{E[Y_{00}]} \equiv 1 - \theta_C(m),$$

where Y_{00} is the placebo response when $m = 0^*$. The $\mathrm{CVE}(m)$ used $E(Y_{1m})$, the expected value of the potential response Y_{am}, where A is set to a and M is set of m. To identify $E(Y_{1m})$ from a random sample where individuals are vaccinated and each individual has their natural mediator value, we need to make some strong assumptions. Let X be baseline covariates such that the sequential ignorability assumptions, SI_1 and SI_2, hold, and assume the following positivity assumption holds:

PosM1(SM1): $\Pr[M_1 = m|A = 1, X = x] > 0$, for all x and $m \in \mathcal{S}_{M_1}$.

Then,

$$1 - \theta_C(m) = \mathrm{CVE}(m) = 1 - \frac{\sum_{(m,x)} E[Y_{1m}|X = x, M_1 = m]\Pr[M_1 = m \text{ and } X = x]}{E[Y_{00}]}, \tag{12}$$

where the summation is over the support of X and M_1 combined, and we assume discrete support for ease of exposition. Then, $\theta_C(m)$ is identifiable for each $m \in \mathcal{S}_{M_1}$. See Gilbert et al. [20] for the details of identification results, inferential methods, and sensitivity analyses.

We can define an analogous controlled protective efficacy (CPE) by comparing the passive immunization arm to placebo,

$$\mathrm{CPE}(m) = 1 - \frac{E[Y_{2m}]}{E[Y_{00}]} \equiv 1 - \theta_{\mathrm{I_a}}(m), \tag{13}$$

where we use the notation $\theta_{\mathrm{I_a}}(m) = 1 - \mathrm{CPE}(m)$ because here Y_{2m} (potential outcome from passive immunization where M is set to m) is acting like Y_{0m} (potential outcome from placebo recipient where M is set to m). For the passive immunization arm, we can randomly assign individuals to M_2 with a known distribution with support $\mathcal{S}_{M_1}$. In other words, we ensure by design that $\Pr[M_2 = m|A = 2] > 0$ for each $m \in \mathcal{S}_{M_1}$ and $M_2 \perp\!\!\!\perp Y_{\mathrm{am}}$ instead of assuming sequential ignorability and positivity.

Since $\theta_C(m)$ and $\theta_{\mathrm{I_a}}(m)$ are acting like a total effect and an adding indirect effect at m, we can define the proportion of the controlled ratio effect at m due to the controlled indirect ratio effect at m as

$$\lambda_{\mathrm{a}}(m) = \frac{\log\{\theta_{\mathrm{I_a}}(m)\}}{\log\{\theta_C(m)\}}. \tag{14}$$

These are useful estimands themselves, but they can also be used to identify θ_T and $\theta_{\mathrm{I_a}}$. Because the distribution of M_1 is identifiable from the vaccine arm, and because the denominator of the ratios in equations (12) and (13) do not depend on m, we obtain

$$\theta_{\mathrm{T}} = \sum_m \theta_C(m)\Pr[M_1 = m] \tag{15}$$

and

$$\theta_{\mathrm{I_a}} = \frac{E[Y_{0M_1}]}{E[Y_{0M_0}]} = \sum_m \theta_{\mathrm{I_a}}(m)\Pr[M_1 = m], \tag{16}$$

and $\lambda_{\mathrm{a}} = \log(\theta_{\mathrm{I_a}})/\log(\theta_T)$, where the summations are over $\mathcal{S}_{M_1}$.

For this section, we have identified λ_{a} using the CVE and its analog for passive immunization, CPE. Since this is a randomized trial, both SI_1 and *PosA* are met for both CVE and CPE, but CVE additionally requires the SI_2 and PosM1($\mathcal{S}_{M_1}$) as untestable assumptions, while the CPE can meet those later assumptions by design.

5.2 Three-arm trial with closeout vaccination

Follmann [21] considered closeout vaccination, where individuals in the placebo arm are vaccinated at the end of the response follow-up period, to identify M_1 in the placebo arm without using baseline variables. We modify that idea for the three-arm trial.

To start, as in the previous section, we assume $E(Y_{0M_1}) = E(Y_{2M_1})$. We rewrite $E(Y_{2M_1})$ as

$$\begin{aligned} E[Y_{2M_1}] &= \sum_m \Pr(Y_{2m} = 1 | M_1 = m) \times \Pr(M_1 = m) \\ &= \sum_m \frac{\Pr(M_1 = m | Y_{2m} = 1)\Pr(Y_{2m} = 1)}{\Pr(M_1 = m)} \times \Pr(M_1 = m) \\ &= \sum_m \Pr(M_1 = m | Y_{2m} = 1)\Pr(Y_{2m} = 1) \\ &= \sum_m \{\Pr[M_1 = m] - \Pr(M_1 = m | Y_{2m} = 0)\Pr(Y_{2m} = 0)\}, \end{aligned} \tag{17}$$

where the second step uses the Bayes theorem and the last step uses

$$P(M_1 = m) = P(M_1 = m | Y_{2m} = 1)P(Y_{2m} = 1) + P(M_1 = m | Y_{2m} = 0)P(Y_{2m} = 0).$$

As in the previous section, we independently draw M_2 from a known distribution with support $\mathcal{S}_{M_1}$, such that M_2 is independent of all potential mediators and outcomes, and $\Pr[M_2 = m | A = 2] > 0$ for each $m \in \mathcal{S}_{M_1}$. Then, by independence and consistency,

$$\Pr[Y_{2m} = 0] = \Pr[Y_{2m} = 0 | M_2 = m] = \Pr[Y | A = 2, M_2 = m],$$

and $\Pr[Y_{2m} = 0]$ is identifiable. Also, $\Pr[M_1 = m]$ is identifiable from the vaccine arm. The only remaining piece to identify in equation (17) is $\Pr(M_1 = m | Y_{2m} = 0)$, which we identify using closeout vaccination in the passive immunization arm.

At the end of the response follow-up period, we implement a closeout vaccination on individuals in the passive immunization arm with $Y_{2M_2} = 0$. We do not vaccinate individuals with $Y_{2M_2} = 1$ because the antibody response after acquiring the disease and vaccination cannot substitute for M_1. We assume that on the closeout vaccinated individuals the antibody response after closeout vaccination would equal M_1, the value they would have gotten if vaccinated at the start of the trial. This may be a tenuous assumption if the response is disease and there are asymptomatic infections that have $Y_{2M_2} = 0$, since asymptomatic infections may affect a subsequent immune response to vaccination. We do not assume that $M_2 = M_1$. Instead, we use the independence and positivity assumptions on M_2 previously mentioned to obtain

$$\Pr[M_1 = m | Y_{2m} = 0] = \Pr[M_1 = m | Y_{2m} = 0, M_2 = M_1],$$

so that $\Pr[M_1 = m | Y_{2m} = 0]$ is identifiable. Thus, by equation (17), we can identify $E[Y_{2M_1}]$, and hence, also identify $\theta_{\mathrm{I_a}}$.

In summary, a three-arm trial with closeout vaccination of the passive immunization arm allows identification of $\theta_{\mathrm{I_a}}$ and λ_{a} without the use of any baseline covariates, without assuming SI$_2$, and allowing $\Pr[M_0 = 0^*] = 1$.

5.3 Two randomized trials

In this section, we consider combining two randomized trials, a vaccine vs placebo (VP) trial and a passive immunization vs placebo (IP) trial. This requires more care because there will be differences between the trials, such as distributions for exposure, baseline covariates, potential mediator responses, and potential outcomes. To emphasize these differences, we use superscripts VP and IP to differentiate the two trials in the variables.

Suppose that using the VP trial, we can identify baseline predictors of M_1, such that for each $m \in \mathcal{S}_{M_1}$, there is a predictor set of baseline variables, $\mathcal{S}_x(m) = \{x: \text{ if } X = x \text{ then } M_1 = m\}$. Furthermore, assume that $\mathcal{S}_x(m)$ can identify participants in the IP trial with $M_1 = m$. Then, we design the IP trial using a quota sampling approach so that we chose the participants for the IP trial from the pool of available volunteers such that the baseline distribution of X is similar between the two trials. For all participants in the IP trial with $X \in S_x(m)$, we set $M_2 = m$. If the sampling is done well enough, then the distribution of M_1 will match the distribution of M_2 and

$$\theta_{\mathrm{I_a}} = \frac{E[Y^{VP}_{0M_1}]}{E[Y^{VP}_{0M_0}]} \approx \frac{E[Y^{IP}_{2M_2}]}{E[Y^{IP}_{0M_0}]}. \tag{18}$$

In expression (18), the approximation may be reasonable because even though we will have different exposure distributions in the two trials, the exposure effects should cancel out as in equation (1).

A second approach is similar to that of Section 5.1. Rewrite equation (12) from the VP trial using the new notation,

$$\mathrm{CVE}(m) = 1 - \frac{\sum_x E[Y^{VP}_{1m} | X^{VP} = x] \Pr[X^{VP} = x]}{E[Y^{VP}_{00}]} = 1 - \theta_C(m), \tag{19}$$

and make the same sequential ignorability and positivity assumptions as in Section 5.1. For the IP trial, we assume that within levels of X, the study populations between the trials are comparable and capture all the differences between the trials except the exposure effects, which we assume cancel out by equation (1). Then, we modify equation (13) to standardize using the distribution of X from the VP trial,

$$\mathrm{CPE}(m) = 1 - \frac{\sum_x E[Y^{IP}_{2m} | X^{IP} = x] \Pr[X^{VP} = x]}{E[Y^{IP}_{00} | X^{IP}] \Pr[X^{VP} = x]} = 1 - \theta_{\mathrm{I_a}}(m), \tag{20}$$

where now we must standardize in both the numerator and denominator. As in Section 5.1, we can impose the independence of the m and potential outcomes in the $A = 2$ arm by randomization in the design, not by assumption. In summary, we are controlling for differences in X between the trials by standardization, and

Table 5: Example for combining two trials, VP trial and IP trial

m (Ab level)	CVE (m)	CPE (m)	θ_C (m)	$\theta_{\mathrm{I_a}}$ (m)	λ_{a} (m)	P ($M_1 = m$)
0	62.0	0.0	38.0	100.0	0.0	10
1	92.0	60.0	8.0	40.0	36.3	40
2	96.0	85.0	4.0	15.0	58.9	50
Overall	91.0	66.5	9.0	33.5	45.4	100

All values except Ab level are in percentage.

controlling for differences in exposure by equation (1). We again define $\lambda_a(m)$ as in equation (14), and because the denominators still do not depend on m, we can obtain θ_T and θ_{I_a} by equations (15) and (16), respectively.

Consider making inferences using equations (19) and (20). In Table 5 and Figure 2, we give a made-up example assuming the covariate adjustments of equations (19) and (20) have already been made. In this example, there are three levels of the Ab mediator ($m = 0, 1, 2$). At $m = 0$, the value of $\hat{\theta}_{I_a}(0) = 100\%$ so that $\hat{\lambda}_a(0) = 0$. The overall effects are estimated by weighted averages, so that using equation (15), we obtain $\hat{\theta}_T$ and using equation (16), we obtain $\hat{\theta}_{I_a}$. The overall λ_a is then estimated by $\hat{\lambda}_a = \log(\hat{\theta}_{I_a})/\log(\hat{\theta}_T)$. Compare $\hat{\theta}_{I_a} = 45.4\%$ and $\hat{\lambda}_a = 45.4\%$ from Table 5 to $\hat{\theta}_{I_s} = 0.09/0.38 = 23.7\%$ and $\hat{\lambda}_s = \log(0.38)/\log(0.09) = 40.2\%$ calculated using assumption SI$_2$ and equation (6), which gives $\hat{E}[Y_{1M_0}] = 38\%$. So if SI$_2$ is true and the sample size is large enough that we can ignore variability of the estimators, then $\theta_{I_s} < \theta_{I_a}$ and there is an interaction (since $\theta_{I_s} \neq \theta_{I_a}$).

6 Discussion

We have explored the assumptions related to identifying the proportion of the total ratio effect mediated through the antibodies in a VP trial. We have shown that because the placebo arm will likely not have positive

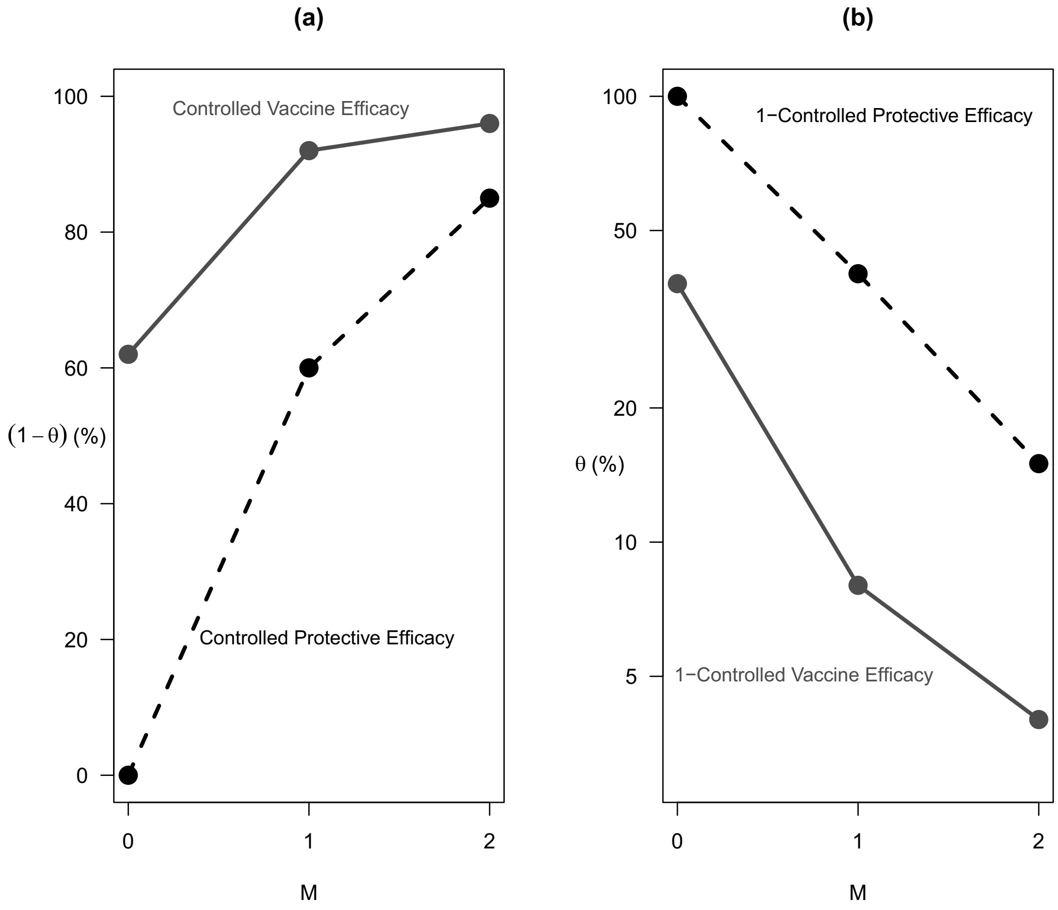

Figure 2: Example for combining two trials, VP trial and a IP trial. In panel (a), we plot the CVE ($\mathrm{CVE}(m) = 1\ \ \theta_C(m)$) for the VP trial and the CPE ($\mathrm{CPE}(m) = 1\ \ \theta_{I_a}(m)$) for the IP trial, while in panel (b), we plot the same data in terms of ratio effects ($\theta_C(m)$ for the VP trial, and $\theta_{I_a}(m)$ for the IP trial) but on a log scale. The efficacies (i.e., vertical axes) are in percentage.

antibody values, from the VP trial, we can only identify one of the two ways to define that proportion (λ_s), and that requires an independence assumption between potential mediators and potential outcomes. We discussed supplementing the VP trial with extra experimental information: either adding to the VP trial a third arm with a passive immunization, or combining the VP trial with a IP trial. These newly proposed experimentally supplemented trials allow us to identify the proportion of the total ratio effect due to adding antibodies to the placebo arm, λ_a.

We note several differences between the traditional (i.e., VP trial alone) and non-traditional (i.e., VP trial with supplemental experimental passive immunization) mediation analyses. First, the traditional mediation analysis is estimating λ_s, while the non-traditional one is estimating λ_a. If we can make the simplifying assumption that the size of the effect of adding antibodies to participants in the placebo arm is the same as that of subtracting antibodies from participants in the vaccine arm, then $\lambda = \lambda_s = \lambda_a$, and the proposed non-traditional supplemental experiments provide a way to estimate λ in a different way. For the passive immunization arm, we do not need to assume the antibodies are independent of the potential outcomes because we can impose that independence by actively randomizing participants to their antibody values. In practice, we can do both mediation analyses and can compare estimates of both λ_s and λ_a. If they are similar, then it may be reasonable to assume no interaction such that $\lambda_s = \lambda_a$ and they are two estimators of the same parameter, λ.

Even if we cannot assume $\lambda_s = \lambda_a$, an advantage of estimating λ_a using the non-traditional mediation analyses is that the cross-world parameter $E(Y_{0M_1})$ is estimated experimentally (see, e.g., Section 5.2). In contrast, the cross-world parameter $E(Y_{1M_0})$ used in traditional mediation analyses for estimating λ_s requires making necessary and often difficult independence assumptions, which ideally should be accompanied with sensitivity analyses (see similar sensitivity analyses for CVE in [20]). In practice, some thought is required to accept the consistency assumptions with respect to the vaccine-induced antibodies having an equal effect as monoclonal antibodies. It could be that the monoclonal antibodies are made from a different virus strain that the vaccine-induced ones and the monoclonal antibodies may have different immunological characteristics. Furthermore, the decay pharmacokinetics are likely to be different between the two types of antibodies so timing of antibody measurements is important for these types of studies.

In this study, we have focused on vaccine studies where very few or no placebo participants would produce antibodies. We have also focused almost exclusively on examining assumptions and identifiability. We explored different study designs to identify these mediation effects. One design in particular (a three-arm randomized trial with closeout vaccination, see Section 5.2) is feasible and does not require as severe assumptions as the others. There is much room for future work in spelling out the details of that design and the other proposed study designs, including examining how the designs may need to be modified for feasibility reasons. Future work could explore these issues when some placebo participants have existing antibodies due to natural exposure. Other work is needed to explore details of estimators (e.g., [4]), and this article could be complementary to some of that work, since the estimator work typically does not have the space to fully discuss the implications of all their assumptions.

Supplementary material: See supplementary material for proofs and extra mathematical details.

Acknowledgements: We thank Peter Gilbert and three anonymous reviewers for insightful comments that appreciably improved the article.

Funding information: None declared.

Author contributions: Both authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors state no conflict of interest.

Data availability statement: All data generated or analyzed for this article are included in this published article.

References

[1] Halloran ME, Longini IM, Struchiner CJ, Longini IM. Design and analysis of vaccine studies. New York: Springer; 2010.

[2] VanderWeele T. Explanation in causal inference: methods for mediation and interaction. Oxford University Press; 2015.

[3] Gilbert PB, Montefiori DC, McDermott AB, Fong Y, Benkeser D, Deng W, et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. Science. 2022;375:43–50.

[4] Benkeser D, Diaz I, Ran J. Inference for natural mediation effects under case-cohort sampling with applications in identifying COVID-19 vaccine correlates of protection. 2021. arXiv:210302643v1.

[5] Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. Epidemiology. 1992;3:143–55.

[6] Seigrist CA, Lambert PH. Chapter 2: how vaccines work. In: Bloom B, Lambert P, editors. The vaccine book. New York: Elsevier; 2016. p. 33–42.

[7] Rapaka RR, Hammershaimb EA, Neuzil KM. Are some COVID-19 vaccines better than others? Interpreting and comparing estimates of efficacy in vaccine trials. Clin Infect Diseases. 2022;74(2):352–8.

[8] Cowling BJ, Lim WW, Perera RA, Fang VJ, Leung GM, Peiris JM, et al. Influenza hemagglutination-inhibition antibody titer as a mediator of vaccine-induced protection for influenza B. Clin Infect Diseases. 2019;68(10):1713–7.

[9] Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen Tchetgen EJ. Practical guidance for conducting mediation analysis with multiple mediators using inverse odds ratio weighting. American J Epidemiol. 2015;181(5):349–56.

[10] Hudgens MG, Gilbert PB, Self SG. Endpoints in vaccine trials. Stat Methods Med Res. 2004;13(2):89–114.

[11] Senn S. The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy. Pharmaceutical Stat. 2022;21(4):790–807.

[12] Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009;20(1):3–5.

[13] Hafeman DM, Schwartz S. Opening the Black Box: a motivation for the assessment of mediation. Int J Epidemiol. 2009;38(3):838–45.

[14] Pearl J. Direct and Indirect Effects. In: Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann; 2001. p. 411–20.

[15] Judea P. The mediation formula: a guide to the assessment of causal pathways in nonlinear models. In: Berzuini C, Dawid P, Bernardinelli L, editors. Causality: statistical perspectives and applications. Chichester, West Sussex, UK: Wiley; 2012. p. 151–79.

[16] Imai K, Keele L, Yamamoto T. Identification, inference and sensitivity analysis for causal mediation effects. Stat Sci. 2010;25(1):51–71.

[17] Goel RR, Apostolidis SA, Painter MM, Mathew D, Pattekar A, Kuthuru O, et al. Distinct antibody and memory B cell responses in SARS-CoV-2 naiiiiive and recovered individuals after mRNA vaccination. Sci Immunol. 2021;6(58):eabi6950.

[18] Hafeman DM, VanderWeele TJ. Alternative assumptions for the identification of direct and indirect effects. Epidemiology. 2011;22:753–64.

[19] Shukla R, Ramasamy V, Shanmugam RK, Ahuja R, Khanna N. Antibody-dependent enhancement: a challenge for developing a safe dengue vaccine. Front Cellular Infect Microbiol. 2020;10:Article 572681.

[20] Gilbert PB, Fong Y, Kenny A, Carone M. A controlled effects approach to assessing immune correlates of protection. Biostatistics. 2023;24(4):850–65.

[21] Follmann D. Augmented designs to assess immune response in vaccine trials. Biometrics. 2006;62(4):1161–9.