The International Journal of Biostatistics

Volume 7, *Issue* 1 2011 *Article* 36

Commentary on "Principal Stratification — a Goal or a Tool?" by Judea Pearl

Peter B. Gilbert, *Fred Hutchinson Cancer Research Center & University of Washington*
Michael G. Hudgens, *University of North Carolina at Chapel Hill*
Julian Wolfson, *University of Minnesota, Twin Cities*

Recommended Citation:
Gilbert, Peter B.; Hudgens, Michael G.; and Wolfson, Julian (2011) "Commentary on "Principal Stratification — a Goal or a Tool?" by Judea Pearl," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 36.
DOI: 10.2202/1557-4679.1341

Commentary on "Principal Stratification — a Goal or a Tool?" by Judea Pearl

Peter B. Gilbert, Michael G. Hudgens, and Julian Wolfson

Abstract

This commentary takes up Pearl's welcome challenge to clearly articulate the scientific value of principal stratification estimands that we and colleagues have investigated, in the area of randomized placebo-controlled preventive vaccine efficacy trials, especially trials of HIV vaccines. After briefly arguing that certain principal stratification estimands for studying vaccine effects on post-infection outcomes are of genuine scientific interest, the bulk of our commentary argues that the "causal effect predictiveness" (CEP) principal stratification estimand for evaluating immune biomarkers as surrogate endpoints is not of ultimate scientific interest, because it evaluates surrogacy restricted to the setting of a particular vaccine efficacy trial, but is nevertheless useful for guiding the selection of primary immune biomarker endpoints in Phase I/II vaccine trials and for facilitating assessment of transportability/bridging surrogacy.

KEYWORDS: principal stratification, causal inference, vaccine trial

Author Notes: We thank the reviewer for extraordinarily insightful comments that led to major improvements. This research was supported by NIH NIAID grant 2 R37 AI054165-08 and NIAID HIV Vaccine Trials Network grant 5 U01 AI068635-05.

1 Introduction

In agreement with Pearl, we think it is only worth developing methods for inference on a "principal stratification" (PS) estimand [of the form of equation (3) in Pearl's article, $P(Y_x = y|Z_x = z, Z_{x'} = z')$] if the estimand[1] passes the litmus test question: "If we knew the value of the estimand, could we do something useful with it to advance science?" Our initial consideration of PS estimands was motivated by dissatisfaction with the status quo non-causal estimands used in our applied areas of research, stimulating the search for more useful causal estimands; however, we concur with Pearl's warning that this search should not *a priori* restrict consideration to PS estimands. Rather, the scientific question should drive the search– undertaken with vigorous debate– that may or may not land on a PS estimand. This vigorous search is of primary importance in science, with science better served by more articles with extended discussions of estimand choice, at the expense of relegating more technical details to the supplement, and fewer articles with extended technical discussions, at the expense of a cursory treatment of estimand choice.

The remainder of this commentary takes up Pearl's challenge to clarify the role and scientific value of PS estimands that we and colleagues have investigated, in our case in the area of randomized placebo-controlled preventive vaccine efficacy trials, especially HIV trials, which enroll HIV-negative volunteers and follow them for occurrence of HIV infection and for post-infection outcomes. After briefly describing a PS estimand for studying vaccine effects on post-infection outcomes, we focus on a PS estimand for studying immune biomarkers as surrogate endpoints. We suggest that the post-infection PS estimand is of genuine scientific interest (Pearl's category 3) whereas the surrogate PS estimand is not of ultimate scientific interest (because it evaluates surrogacy restricted to the setting of a particular efficacy trial), but is nevertheless useful for guiding the selection of primary endpoints in subsequent Phase I/II HIV vaccine trials and for facilitating assessment of transportability/bridging surrogacy.

2 Evaluating vaccine effects on post-infection outcomes

Our foray into research involving PS estimands addressed a problem parallel to the "truncation by death" problem classified by Pearl as being of genuine research interest. In HIV vaccine efficacy trials, HIV infection is a primary endpoint, and is also intermediate to co-primary or secondary endpoints measured after HIV infection (e.g., HIV viral load). Such post-infection endpoints are only meaningfully measured in HIV infected individuals, just as quality of life is only meaningfully

[1] i.e., a quantity of interest to be estimated

measured in alive individuals. Unsatisfied with the standard non-causal estimand that compares the post-infection outcome in the infected vaccine group versus the infected placebo group (which could be particularly misleading because a safe vaccine could appear to harmfully increase viral load), our research uses PS estimands that compare the mean or the survival probability of the post-infection outcome in those who would be infected under either treatment assignment (e.g., Hudgens, Hoering, and Self, 2003; Gilbert, Bosch, and Hudgens, 2003; Hudgens and Halloran, 2006; Jemiai et al., 2007; Shepherd, Gilbert, and Lumley, 2007; Gilbert and Jin, 2010; Shepherd, Gilbert, and Dupont, 2011), which are equivalent to the truncation by death PS estimand that focuses on those who would be alive under either treatment assignment (Robins, 1986; Rubin, 2000).

The PS estimands restrict attention to a subgroup of particular scientific interest, namely those with no vaccine effect on HIV infection. Prime-boost HIV vaccines (e.g., the regimen tested in Thailand, Rerks-Ngarm et al., 2009) generate both antibody and T cell responses and thus are hypothesized to have effects on both infection and post-infection outcomes; focusing on individuals with no causal vaccine effect on infection allows isolation of the vaccine's effect on post-infection outcomes. Separating these two effects is helpful for designing improved vaccines and for predicting the public health impact of a licensed vaccine. In addition, the PS estimand has a simple interpretation from the perspective of the study participant, addressing his/her question: If I am going to become infected regardless of treatment assignment, will the vaccine lower my viral load? We conclude that the PS estimand fits Pearl's third category.

3 Evaluating immune biomarkers as surrogate endpoints

3.1 Introduction

A second area of research is the evaluation of surrogate endpoints, i.e., the evaluation of how well vaccine effects (more generally treatment effects) on a biomarker predict vaccine effects on the true clinical endpoint of interest. For our working example (HIV vaccine efficacy trials), HIV uninfected subjects are randomized to receive vaccine or placebo, the biomarker is an HIV-specific immune response measured after the planned immunizations, and the clinical endpoint is HIV infection. Pearl states that a useful surrogate must robustly predict clinical treatment effects in new settings, a point we agree with but feel needs more discussion. Pearl seems to suggest that it is unimportant to evaluate the value of a surrogate endpoint for

DOI: 10.2202/1557-4679.1341

the same setting as the efficacy trial, because the only purpose of a surrogate is transportability. We agree that ultimately this is indeed the only purpose, because every follow-up study takes place in a new setting even if attempts are made to make the conditions identical to those in the original study. Nevertheless, a few years ago we proposed that both goals of evaluating "bridging/general" surrogates and evaluating "specific" surrogates (restricted to the same setting) are important for vaccine development, and suggested meta-analysis of multiple efficacy trials for the former (e.g., Daniels and Hughes, 1997; Molenberghs et al., 2008) and principal stratification-based and Prentice criteria-based (Prentice, 1989) approaches for the latter (Qin et al., 2007; Gilbert, Qin, and Self, 2009). The new transportability/bridging surrogate approach of Pearl and Bareinboim (2011) should also be evaluated for its utility in vaccine development. Below we describe how, among the candidate estimands measuring surrogacy, the PS specific surrogate estimand is particularly useful for guiding vaccine development, especially for a special class of efficacy trials (chiefly HIV) that has motivated our work.

3.2 Specific surrogate estimand

By specific surrogate value, we mean the accuracy with which causal treatment effects on the biomarker Z predict causal treatment effects on the clinical endpoint Y (measured during a follow-up period after the biomarker is measured) for the same setting as the efficacy trial. This value may be measured with a PS estimand that we named the "causal effect predictiveness (CEP) surface" (Gilbert and Hudgens, 2008). This estimand conditions on not yet experiencing the clinical endpoint under either treatment assignment at the fixed time τ (near baseline) that the biomarker is measured; however, to simplify the discussion we assume all subjects qualify for this group. In this case, for Y a binary outcome, the CEP estimand is defined as

$$\begin{aligned} CEP(z_1,z_0) &\equiv P(Y_0=1|Z_1=z_1,Z_0=z_0)-P(Y_1=1|Z_1=z_1,Z_0=z_0) \\ &= E(Y_0-Y_1|Z_1=z_1,Z_0=z_0) \end{aligned} \quad (1)$$

(or some other contrast), where Z_x (Y_x) is the potential biomarker (infection status after τ) if assigned treatment x, for $x=0$ (placebo) and $x=1$ (vaccine) [Gilbert and Hudgens (2008), building on Frangakis and Rubin (2002) and Follmann (2006)]. This estimand measures clinical efficacy in subgroups defined by certain causal vaccine effects on the biomarker. We refer to the CEP estimand as the "specific surrogate estimand," and in Section 3.5 discuss how it can be used in the assessment of transportability/bridging surrogacy that is ultimately of interest.

As for the post-infection PS estimand, our research for the CEP estimand was initially motivated from dissatisfaction with the standard estimand traditionally used for evaluating an "immune correlate" of vaccine protection, which simply measures the association between Z_1 and Y_1 using data from the vaccine group, and, where a strong association is found, an inference is ventured that the correlate of infection may be used to reliably predict vaccine efficacy (Qin et al., 2007). However, as extensively discussed in the surrogate endpoint evaluation literature, association between Z_1 and Y_1 does not imply association between $Z_1 - Z_0$ and $Y_1 - Y_0$, and yet, for our goal of predicting vaccine efficacy, it is the latter association that is truly of interest. To illustrate how the former association may not predict the latter, suppose $Z_0 = 0$, $Z_1 = U + \varepsilon_{z1}$, $Y_0 = U + \varepsilon_{y0}$, and $Y_1 = U + \varepsilon_{y1}$, where $U, \varepsilon_{z1}, \varepsilon_{y0}, \varepsilon_{y1}$ are all independent with $U \sim N(0, \sigma = 10)$, $\varepsilon_{z1} \sim N(1,1)$, $\varepsilon_{y0} \sim N(0,1)$, and $\varepsilon_{y1} \sim N(1,1)$. Then $cor(Z_1, Y_1) = 100/101$ and yet $cor(Z_1 - Z_0, Y_1 - Y_0) = 0$, i.e., Z_1 is an excellent predictor of Y_1 but $Z_1 - Z_0$ is a worthless predictor of the vaccine effect $Y_1 - Y_0$. The vaccine literature was void of estimands that quantify the correlation between treatment effects on biomarker and outcome, and we proposed the CEP estimand for this purpose. Evaluating $CEP(z_1, z_0)$ at different fixed values (z_1, z_0) amounts to a series of subgroup analyses, equivalent to the common secondary objective in clinical trials to assess how the treatment effect (e.g., vaccine efficacy) varies with baseline covariates. Here again, principal stratification is useful in defining subgroups of particular interest, such as those who do and those who do not experience a causal vaccine effect on the biomarker in question.

Paraphrasing a question from Ross Prentice, "Why not instead focus research on discovering actual baseline covariates that predict vaccine efficacy?" The first part of the answer is that there often exist immune biomarkers measured after the immunizations that are much stronger efficacy predictors than any baseline covariates, simply because the post-immunization biomarkers are selected to putatively measure the functional immune response that kills the pathogen of interest before it can establish infection (see Plotkin, 2010, for a review). But, the objection continues, the PS estimand is less useful than an actual baseline covariate because (Z_1, Z_0) can never be measured simultaneously on the same individual. While true for many vaccine efficacy trials, this is false for the large special class of trials that only enroll subjects without previous infection with the pathogen. For such trials, Z_0 is known to be zero/negative for all subjects, because the laboratory instrument used to measure Z is designed to detect only a pathogen-specific immune response. For this class the estimand simplifies to

$$CEP(z_1) \equiv P(Y_0 = 1 | Z_1 = z_1) - P(Y_1 = 1 | Z_1 = z_1) = E(Y_0 - Y_1 | Z_1 = z_1). \qquad (2)$$

This simplification implies (Z_1, Z_0) are observed simultaneously for subjects assigned vaccine, greatly aiding identifiability (addressed briefly in Section 4). More-

DOI: 10.2202/1557-4679.1341

over, this simplification is appealing for the parsimony of studying how clinical efficacy varies with a univariate (or low-dimensional Z_1) biomarker, and was an important motivator for us to use the CEP estimand in our HIV vaccine efficacy trials research.

In addition, due to the lack of common support of the vaccine and placebo group biomarker distributions, treatment effects are undefined within all subgroups with observed biomarker $Z = z$ for $z > 0$, such that the Prentice (1989) approach (Chan et al., 2002; Gilbert and Hudgens, 2008) does not apply. The utility of the natural direct/indirect effect approach in this setting is also not clear (see Section 3.6). Therefore, in our motivating application principal stratification appears to yield the only well-defined estimand for assessing surrogate value. In other settings with Z_0 variable, the estimand (2) may still be useful, as the ability to predict $Y_1 - Y_0$ from Z_1 alone regardless of Z_0 would be useful for vaccine development (Wolfson and Gilbert, 2010); generally what is sought is accurate prediction of $Y_1 - Y_0$ based on any baseline covariate information (i.e., actual baseline covariates and/or Z_1 and Z_0, which are treated as baseline covariates).

3.3 Utility of the specific surrogate estimand for selecting the primary biomarker endpoints in follow-up Phase I/II vaccine trials

For a given field of researchers working to develop a vaccine against a certain pathogen, a handful of pivotal vaccine efficacy trials are conducted over a period of decades, and the generated data are used to make decisions on the immune biomarkers to use as primary study endpoints in subsequent Phase I/II trials that evaluate and compare a number of refined candidate vaccine regimens. Typically no direct data on clinical efficacy in new settings are available for informing these decisions, such that the decisions are traditionally based on the primary efficacy data together with the observed associations between the immune biomarkers and the clinical outcome within the efficacy trials (i.e., Z_1 and Y_1), informally combined with theories/models of mechanisms of protection. In particular, for many pathogens the first efficacy trial demonstrates partial vaccine efficacy that is too low to warrant licensure, which makes it a top priority to assess immune correlates for guiding the selection of immune biomarker endpoints in follow-up trials (e.g., such immune correlates assessment is now occurring for the first trial to show low-level efficacy of an HIV vaccine, Rerks-Ngarm et al., 2009).

As an improvement to the traditional approach that selects biomarkers with the strongest (Z_1, Y_1) associations, we suggest utilizing the $CEP(z_1)$ curve to select biomarkers with the strongest $(Z_1, Y_1 - Y_0)$ associations. Biomarkers with $CEP(z_1)$

large for some range of z_1 and $CEP(z_1)$ small for z_1 equal to or near zero may be prioritized as primary endpoints. Given selection of the best biomarker, the follow-up Phase I/II trials would rank the vaccine regimens by the proportion of vaccine recipients with immune response z_1 in the estimated high-protection range, forming the basis for advancing the most promising vaccine regimen to the next efficacy trial. Because accurate prediction of vaccine efficacy internal to an efficacy trial does not imply accurate prediction to a new setting where the clinical outcome is not measured, it is important to address if and how the CEP estimand may be useful for this purpose; we begin this discussion in Section 3.5.

3.4 Remarks on evaluating bridging surrogates

Based on the above discussion, theories of mechanisms of protection must be combined with an empirically supported specific surrogate to make a bridging prediction, and the accuracy of the prediction depends on the veracity of the theory. Building credible theories has often been more achievable in the preventive vaccine setting than for many chronic disease settings, because the biological pathways of treatment effects on Z and Y are often better understood and these pathways can be studied more readily in the lab, due to the specificity of the biomarker and of the infection endpoint. For example, commonly theories have proposed that functional antibodies (e.g., neutralizing) directed to certain pathogen epitopes are protective against infection, and manipulation experiments are conducted (e.g., antibody infusion challenge experiments in animals or humans) to provide evidence that the functional antibodies actually kill the pathogen before it can establish infection (Plotkin, 2008). The nature of the bridge is fundamental to the needed theory. If the only change from the conditions of the efficacy trial is adding a fourth pathogen strain to the existing 3-strain vaccine (e.g., for influenza), then it may be relatively easy to develop a compelling biological theory justifying accurate bridging, whereas if a new vaccine formulation is tested in a new population against new circulating virus types, then the needed biological theory will be more elaborate, raising the bar for credibility.

Pearl and Bareinboim (2011) suggest it is useful to mathematically formalize the process by which evaluating a predictive biomarker in an experimental setting is combined with theory/assumptions to yield accurate bridging, a point we agree with. As noted above vaccine development over the past 60 years has proceeded by informally combining the two elements, which, while not ideal, has worked reasonably well, as judged by the fact that the identification and assessment of efficacy-predictive immune biomarkers has ubiquitously played a central role in the development and deployment of vaccines, many of which were confirmed over

DOI: 10.2202/1557-4679.1341

the course of decades to confer high levels of vaccine efficacy in many populations (Falk and Ball, 2001; Plotkin, 2010). However, the use of a formal mathematical framework for bridging may have allowed for even greater success, and future research in this area seems merited. Similarly, a formal framework is needed for understanding when the CEP estimand provides reliable guidance for bridging. We sketch a start to this problem in the next section.

3.5 Toward criteria for reliable bridging based on the CEP estimand

Consider a new setting different from that in the efficacy trial, which may entail a new vaccine regimen, a new study population, or both. First consider the case of a new vaccine and the same study population. To illustrate the bridging problem (which is realistic for HIV vaccine efficacy trials), suppose the initial efficacy trial demonstrates partial vaccine efficacy that is promising but too low to warrant licensure, and also identifies a promising biomarker, with $\widehat{CEP}(0)$ near zero (i.e., supporting average causal necessity of a vaccine-induced immune response for protection, Gilbert and Hudgens, 2008) and $\widehat{CEP}(z_1)$ increasing monotonically in z_1. These results stimulate research on various refined candidate vaccines, leading to the advancement of a promising new vaccine to a follow-up Phase II trial in the identical population that was studied in the efficacy trial (identical inclusion and exclusion criteria), which shows that the distribution of the immune biomarker is substantially shifted upwards compared to that for the previous vaccine. The field of vaccine researchers hopes that the new vaccine improves the overall vaccine efficacy.

Similar to Peal and Bareinboim (2011), our formulation for evaluating bridging envisages two experiments, the original efficacy trial and the follow-up Phase II trial, and considers conditions for transportability. The Phase II trial randomizes subjects to the new vaccine or new placebo ($X = 1'$ or $X = 0'$), and uses an identical procedure for measuring the same biomarker as was used in the efficacy trial, yielding information on Z. However, information on the outcome Y is not collected and therefore interest focuses on predicting $CE^{new} \equiv P(Y_{0'} = 1) - P(Y_{1'} = 1)$, i.e., the overall effect on Y of the new vaccine. The overall effect on Y of the original vaccine can be expressed as $CE \equiv P(Y_0 = 1) - P(Y_1 = 1) = \int CEP(z_1)dF(z_1)$ and the overall effect on Y of the new vaccine can be expressed as $CE^{new} = \int CEP^{new}(z_1)dF'(z_1)$, where F is the cdf of Z_1, F' is the cdf of $Z_{1'}$, and $CEP^{new}(z_1) \equiv P(Y_{0'} = 1|Z_{1'} = z_1) - P(Y_{1'} = 1|Z_{1'} = z_1)$. Here for simplicity we focus on the common special case that Z_0 and $Z_{0'}$ are constant. The field of vaccine researchers receives reliable guidance about bridging efficacy if CE^{new} can

be accurately predicted, suggesting a general criterion for Z to be a useful bridging surrogate:

[**Bridging Surrogate Criterion.**] Z is a useful bridging surrogate if CE^{new} can be accurately predicted based on $CEP(z_1)$ and F from the efficacy trial and F' from the follow-up trial.

In particular, the field of vaccine researchers hopes for accurate prediction in the following way: if a new vaccine is selected for efficacy testing based on the criterion that Phase I/II trials demonstrate increases in the percentage of vaccine recipients with z_1 values in regions where $\widehat{CEP}(z_1)$ from the previous efficacy trial is high, then the selected vaccine is accurately predicted to have $CE^{new} > CE$. That is, successfully modifying the vaccine based on the biomarker reliably leads to a more efficacious vaccine.

Following Gilbert and Hudgens (2008), if Z_1 and $Z_{1'}$ have the same support, then the prediction of CE^{new} may be based on

$$CE^{new} = \int \psi(z_1) CEP(z_1) dF'(z_1), \tag{3}$$

where $\psi(z_1) \equiv CEP^{new}(z_1)/CEP(z_1)$ (with convention 0/0 = 1). This equation re-weights the original CEP curve by two factors: the relationship between $CEP^{new}(z_1)$ and $CEP(z_1)$ for each value of z_1 and the distribution of $Z_{1'}$ for the new vaccine. A numerical prediction is obtained by substituting estimates for $CEP(\cdot)$ and $F'(\cdot)$ into (3) and by assuming a fully specified form for $\psi(\cdot)$; therefore the prediction combines empirical evidence with a bridging assumption. A perfectly accurate prediction is obtained if $CEP^{new}(\cdot) = CEP(\cdot)$, i.e., $\psi(\cdot) = 1$. If this perfect bridging assumption holds, then CE^{new} can be accurately predicted by

$$\widehat{CE}^{new} = \int \widehat{CEP}(z_1) d\widehat{F}'(z_1). \tag{4}$$

Expression (4) is similar in spirit to the "transport formula" of Pearl and Barenboim [2011, equation (5)], except in (4) we are integrating over principal strata rather than observed biomarker levels. In words, the perfect bridging assumption $\psi(\cdot) = 1$ states that given a vaccine induces an immune response z, the protective effect (on Y) will be the same regardless of whether it was the new or original vaccine that induced the immune response, and regardless of any differences in the placebos used in the two studies. Note that the perfect bridging assumption $\psi(\cdot) = 1$ implies transportability of the average causal necessity condition from the efficacy trial to the new trial: $CEP(0) = 0$ implies $CEP^{new}(0) = 0$.

On a case-by-case basis, vaccine researchers must deliberate the plausibility of the perfect bridging equality. One way it would fail is if the additional group of

DOI: 10.2202/1557-4679.1341

individuals achieving an immune response in the high-protection range with the new vaccine differs in a critical way from the subgroup that achieved the high-protection range with the original vaccine in the efficacy trial; for example, the originally protected subgroup may have all possessed a critical (unmeasured) host genotype that is absent in the additional group.

If perfect bridging ($\psi(\cdot) = 1$) fails, imperfect but useful bridging may still be achieved, depending on the nature of the departure of $\psi(\cdot)$ from unity. Even if $\psi(\cdot)$ does not equal 1, (4) should provide a reasonable estimate of CE^{new} provided $CEP(z_1) \approx CEP^{new}(z_1)$ for z_1 where $dF'(z_1)$ is large. While $\psi(\cdot)$ is not identifiable without evaluating the new vaccine in an efficacy trial, a sensitivity analysis may be conducted where one considers how $\widehat{CE}^{new}$ changes with different assumed forms for $\psi(\cdot)$. For example, if the cautious assumption is made that $\psi(\cdot) = 1/2$, is $\widehat{CE}^{new}$ still sufficiently large to justify moving forward with a new efficacy trial?

If a second efficacy trial is conducted with the new vaccine, then transportability is supported by a numerical prediction $\widehat{CE}^{new}$ [obtained from (4)] near the estimate of CE^{new} obtained in the primary analysis of the new efficacy trial which ignores the biomarker data. Gilbert and Hudgens (2008) note that even in the absence of a second efficacy trial, a partial check of transportability (or 'projective validity') can be conducted by cross-validation of data from the first efficacy trial. In particular, individuals in the trial can be partitioned into two subgroups. Then the CEP curve can be estimated from subgroup 1 data and CE can be predicted for subgroup 2 based on the observed distribution of Z_1 in subgroup 2 and the estimated CEP curve from subgroup 1. Transportability across subgroups is supported if the predicted CE in subgroup 2 is similar to the estimate of CE in that subgroup which ignores data on Z_1.

Next we suppose the follow-up Phase I/II trial is done with a new vaccine in a new population. In this case the bridging criterion described above carries over under a slight modification that accounts for different distributions of baseline covariates W in the two settings. Specifically, the CEP estimand now conditions on W, $CEP(z_1, w) \equiv P(Y_0 = 1|Z_1 = z_1, W = w) - P(Y_1 = 1|Z_1 = z_1, W = w)$, and the integrations in (3) and (4) are replaced with integrations over the joint distribution of Z_1 and W, now requiring common support of this joint distribution for the old and new settings. A challenge with the bridging criterion is that the numerical prediction $\widehat{CEP}^{new}$ may be inaccurate if the baseline covariates are inadequately informative about disease risk to fully adjust for differences in risk between the two settings. In addition, the bridging criterion relies on a particular functional contrast between the conditional disease risks under the two treatment assignments specified by the CEP estimand; we have focused on a difference on the additive scale. These challenges may be especially problematic if the placebo group disease incidence

differs substantially between the two settings. For scenarios where the support of the biomarker and/or the other covariates differs between the old and new settings, additional research is needed to delineate if and how the *CEP* estimand may be useful for assessing bridging surrogate utility.

3.6 Natural versus principal strata direct/indirect effects

Pearl suggests that the PS direct effect (PSDE) estimand (VanderWeele, 2008) is inadequate for measuring mediation of a treatment effect and is generally less interesting scientifically (especially for identifying and explaining causal mechanisms) than the natural direct effect (NDE) estimand of Robins and Greenland (1992) and Pearl (2001). In a sense, this comment does not apply to our specific surrogate estimand as we use it in vaccine efficacy trials– not for measuring anything about the causal biological mechanism of protection, but merely for measuring a biomarker's predictiveness of vaccine efficacy. Therefore on the one hand we do not view the PS estimand *CEP* [given either by (1) or (2)] as a competitor with other causal estimands trying to identify and explain causal mechanisms; it simply has a different purpose, prediction.

On the other hand, the *CEP* estimand at $Z_1 = Z_0$ is the PSDE, which raises the question as to whether this estimand or the alternative NDE estimand is of greater value for the surrogate endpoint problem in vaccine efficacy trials. To address this, we first review the definition of the NDE estimand, which considers the potential clinical endpoint $Y_{x,z}$ under assignment to both $X = x$ and $Z = z$, thus requiring that the biomarker Z is manipulable. By consistency $Y_{x,Z_x} = Y_x$, i.e., the potential outcome when X is set to x and Z is set to Z_x is the same as when X is set to x and Z is not manipulated and therefore (naturally) takes on the value Z_x. The average NDE (Pearl 2001, equation 6) estimand can be defined by

$$E(Y_{1,Z_1} - Y_{0,Z_1}) = E(Y_1 - Y_{0,Z_1}), \tag{5}$$

i.e., the average effect of treatment when setting the intermediate Z to the value it would have been with treatment (i.e., when $X = 1$). This estimand is entirely symmetric such that a second average NDE estimand is defined as

$$E(Y_{1,Z_0} - Y_{0,Z_0}) = E(Y_{1,Z_0} - Y_0), \tag{6}$$

i.e., the average effect of treatment when setting Z to the value it would have been without treatment (i.e., when $X = 0$). Thus in considering the NDE estimand for the vaccine setting, one needs to conceive of either placebo recipients having their immune responses set to Z_1, as in (5), or vaccine recipients having their immune responses set to Z_0, as in (6).

DOI: 10.2202/1557-4679.1341

With that background, we are sympathetic to Pearl's statement, "...it is hard to accept the PSDE restriction that nature's pathways should depend on whether we have the technology to manipulate one variable or another," but only for a certain category of manipulations. In particular, we distinguish between manipulations that may not be possible now but conceivably can be developed, versus manipulations that can never conceivably be developed. An example of the former is a controlled direct effect estimand (Pearl, 2001) that sets all subjects to be fully compliant to the assigned inoculations; while this manipulation may be unachievable in an efficacy trial, once an excellent vaccine is licensed, many individuals will receive it, conceivably even those who would have been non-compliant in the efficacy trial (Robins and Greenland, 1996). We suggest that for HIV vaccine trials (for which $Z_0 = 0$ for all subjects) both NDE estimands (5) and (6) are examples of the latter. First, estimand (5) requires that placebo recipients can conceivably have their HIV-specific immune response Z set to exceed 0; however Z is measured using an immunoassay that mixes certain HIV peptides/isolates with the individual's blood sample, and, by the nature of the adaptive immune system, HIV antigenic exposure (created by HIV vaccination or natural exposure) is the only thing that could stimulate a positive HIV-specific response. Similarly, estimand (6) requires that all vaccine recipients can be manipulated to have $Z = 0$, which is also difficult to conceive given the nature of the assay for measuring Z. Many others have suggested that causal estimands requiring inconceivable manipulations are of dubious scientific value (e.g., Holland, 1986; Angrist, Imbens, and Rubin, 1996). VanderWeele (2008) wrote, "Whether it is reasonable to consider counterfactual variables of the form Y_{xz} will depend on whether an intervention on the intermediate variable is conceivable," and "Principal strata direct and indirect effects have the advantage that the concepts are defined irrespective of whether an intervention on the intermediate variable is conceivable." Specifically addressing the surrogate endpoint problem, Gallop et al. (2009) and Joffe and Greene (2009) made the same point.

However, it is not easy to definitively answer the question as to whether a conceivable manipulation exists; a negative answer produced by a feeble imagination could be reversed by a fertile one. For the vaccine example and NDE estimand (5), we can imagine manipulations to set $Z > 0$ in placebo recipients. For instance, in passive immunization experiments, antibodies or T cells from another individual or stimulated in vitro may be transferred to an unvaccinated individual. However, such a manipulation poses another difficulty to using the NDE estimand: the consistency assumption (Cole and Frangakis, 2009) becomes dubious. In particular, consistency implies that the outcome for an individual observed to have $Z_1 = z_1 > 0$ when vaccinated ($X = 1$) would be the same as if we set $Z = z_1$ through passive immunization.

Thus use of the natural direct/indirect approach may require strong assumptions about manipulation and consistency, a problem not faced by the PS estimand. We conclude there are unsolved challenges posed to use of the NDE estimand for our motivating class of efficacy trials with Z_0 constant.

4 Concluding remarks

Following Pearl, we have largely ignored identifiability in our comments so as to focus attention on the value/interpretability of the PS estimands. While this is appropriate because an estimand must be valuable to make a discussion of identifiability important, where multiple estimands of similar value are being compared, identifiability is a relevant criterion for preferring certain estimands. If the assumptions needed to identify estimand 1 are weaker/more realistic than those needed to identify estimand 2, then that is something to consider in choosing the estimand to attempt to make inference about. The two PS estimands we have considered are not identified from the observed data plus standard assumptions in randomized trials; and hence, extra identifiability assumptions and sensitivity analysis are needed for inference. Augmented study designs (Follmann, 2006) can aid such analyses.

In conclusion, we have suggested the value of principal stratification estimands for providing insight into vaccine effects on post-infection outcomes and for evaluating specific surrogate biomarkers in vaccine efficacy trials. For the former setting the PS estimand delineates a scientifically meaningful subgroup within which vaccine effects are of interest, while in the latter, the CEP estimand facilitates discovery and characterization of efficacy-predictive biomarkers. The CEP estimand provides guidance for selecting the immune response endpoints to use in follow-up Phase I/II vaccine trials before adequate data are available on bridging surrogacy, and is particularly appealing in efficacy trials that enroll participants naive to the pathogen (such that $Z_0 = 0$ for all subjects), both because the estimand is well-defined while alternative estimands are not, and identifiability is achieved with fewer and weaker assumptions. At this point in our surrogate endpoint evaluation research for vaccine trials, we conclude that the CEP estimand is superior for selecting immune biomarkers as primary endpoints in Phase I/II trials compared to traditionally used estimands, and that additional research is needed to understand the utility of the CEP estimand for evaluating bridging surrogates.

DOI: 10.2202/1557-4679.1341

References

Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of causal effects using instrumental variables (with comments)." *Journal of the American Statistical Association*, 91, 444–472.

Cole, S. R. and C. E. Frangakis (2009): "The consistency statement in causal inference: a definition or an assumption?" *Epidemiology*, 20, 3–5.

Daniels, M. and M. Hughes (1997): "Meta-analysis for the evaluation of potential surrogate markers." *Statistics in Medicine*, 16, 1965–1982.

Falk, L. and L. Ball (2001): "Current status and future trends in vaccine regulation." *Vaccine*, 19, 1567–1572.

Follmann, D. (2006): "Augmented designs to assess immune response in vaccine trials." *Biometrics*, 62, 1161–1169.

Frangakis, C. and D. Rubin (2002): "Principal stratification in causal inference." *Biometrics*, 58, 21–29.

Gallop, R., D. Small, J. Lin, M. Elliott, M. Joffe, and T. Ten Have (2009): "Mediation analysis with principal stratification." *Statistics in Medicine*, 28, 1108–1130.

Gilbert, P., R. Bosch, and M. Hudgens (2003): "Sensitivity analysis for the assessment of vaccine effects on viral load in HIV vaccine trials." *Biometrics*, 59, 531–541.

Gilbert, P. and M. Hudgens (2008): "Evaluating candidate principal surrogate endpoints." *Biometrics*, 64, 1146–1154.

Gilbert, P. and Y. Jin (2010): "Semiparametric estimation of the average causal effect of treatment on an outcome measured after a post-randomization event, with missing outcome data." *Biostatistics*, 11, 34–47.

Gilbert, P., L. Qin, and S. Self (2009): "Response to Andrew Dunning's comment on "Evaluating a surrogate endpoint at three levels, with application to vaccine development"," *Statistics in Medicine*, 28, 716–719.

Holland, P. (1986): "Statistics and causal inference." *Journal of the American Statistical Association*, 81, 945–961.

Hudgens, M. and M. Halloran (2006): "Causal vaccine effects on binary postinfection outcomes." *Journal of the American Statistical Association*, 101, 51–64.

Hudgens, M., A. Hoering, and S. Self (2003): "On the analysis of viral load endpoints in HIV vaccine trials." *Statistics in Medicine*, 22, 2281–2298.

Jemiai, Y., A. Rotnitzky, B. Shepherd, and P. Gilbert (2007): "Semiparametric estimation of treatment effects given base-line covariates on an outcome measured after a post-randomization event occurs." *Journal of the Royal Statistical Society, Series B*, 69, 879–902.

Joffe, M. and T. Greene (2009): "Related causal frameworks for surrogate outcomes." *Biometrics*, 65, 530–538.

Molenberghs, G., T. Burzykowski, A. Alonso, P. Assam, A. Tilahum, and M. Buyse (2008): "The meta-analytic framework for the evaluation of surrogate endpoints in clinical trials." *Journal of Statistical Planning and Inference*, 138, 432–449.

Pearl, J. (2001): "Direct and indirect effects." *Proceedings of of the 17th Conference in Uncertainty in Artificial Intelligence*, 411–420.

Pearl, J. (2011): "Principal stratification– a goal or a tool?" *The International Journal of Biostatistics*, 7, Article 20.

Pearl, J. and E. Bareinboim (2011): "Transportability across studies: A formal approach." *Technical Report*, 1–33.

Plotkin, S. A. (2008): "Vaccines: Correlates of vaccine-induced immunity," *Clinical Infectious Diseases*, 47, 401–409, URL `http://dx.doi.org/10.1086/589862`.

Plotkin, S. A. (2010): "Correlates of protection induced by vaccination." *Clinical Vaccine Immunology*, 17, 1055–1065.

Prentice, R. (1989): "Surrogate endpoints in clinical trials: definition and operational criteria." *Statistics in Medicine*, 8, 431–440.

Qin, L., P. Gilbert, L. Corey, J. McElrath, and S. Self (2007): "A framework for assessing an immunological correlate of protection in vaccine trials." *The Journal of Infectious Diseases*, 196, 1304–1312.

Rerks-Ngarm, S., P. Pitisuttithum, S. Nitayaphan, and et al. (2009): "Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in thailand." *New England Journal of Medicine*, 361, 2209–2220.

Robins, J. (1986): "A new approach to causal inference in mortality studies with sustained exposure periods - application to control of the healthy worker survivor effect." *Mathematical Modeling*, 7, 1393–1512.

Robins, J. and S. Greenland (1992): "Identifiability and exchangeability of direct and indirect effects." *Epidemiology*, 3, 143–155.

Robins, J. and S. Greenland (1996): "Comment on, "Identification of causal effects using instrumental variables (with comments)"." *Journal of the American Statistical Association*, 91, 456–458.

Rubin, D. (2000): "Comment on "Causal inference without counterfactuals," by A.P. Dawid." *Journal of the American Statistical Association*, 95, 435–437.

Shepherd, B., P. Gilbert, and C. Dupont (2011): "Sensitivity analyses for comparing time-to-event outcomes only existing in a subset selected postrandomization and relaxing monotonicity." *Biometrics*, 67, 1100–1110.

DOI: 10.2202/1557-4679.1341

Shepherd, B., P. Gilbert, and T. Lumley (2007): "Sensitivity analyses comparing time-to-event outcomes only existing in a subset selected post-randomization, conditional on covariates, with application to HIV vaccine trials." *Journal of the American Statistical Association*, 102, 573–582.

VanderWeele, T. (2008): "Simple relations between principal stratification and direct and indirect effects." *Statistics and Probability Letters*, 78, 2957–2962.