Causal Approaches to Disease Progression Analyses

Bronner P. Gonçalves[a] and Etsuji Suzuki[b]

Abstract: Epidemiologic analyses that aim to quantify exposure effects on disease progression are not uncommon. Understanding the implications of these studies, however, is complicated, in part because different causal estimands could, at least in theory, be the target of such analyses. Here, to facilitate interpretation of these studies, we describe different settings in which causal questions related to disease progression can be asked, and consider possible estimands. For clarity, our discussion is structured around settings defined based on two factors: whether the disease occurrence is manipulable or not, and the type of outcome. We describe relevant causal structures and sets of response types, which consist of joint potential outcomes of disease occurrence and disease progression, and argue that settings where interventions to manipulate disease occurrence are not plausible are more common, and that, in this case, principal stratification might be an appropriate framework to conceptualize the analysis. Further, we suggest that the precise definition of the outcome of interest, in particular of what constitutes its permissible levels, might determine whether potential outcomes linked to disease progression are definable in different strata of the population. Our hope is that this paper will encourage additional methodological work on causal analysis of disease progression, as well as serve as a resource for future applied studies.

Keywords: Causal inference; Controlled direct effects; Disease progression; Potential outcomes; Principal stratification

(Epidemiology 2025;36: 732–740)

Submitted November 11, 2024; accepted June 30, 2025

From the [a]Faculty of Health and Medical Sciences, University of Surrey, Guildford, United Kingdom, and [b]Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan.

E.S. is supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP19KK0418 and JP23K09740). The funding sources had no involvement in study design; the collection, analysis, and interpretation of data; the writing of the article; and the decision to submit the article for publication.

Disclosure: The authors report no conflicts of interest.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Bronner P. Gonçalves, Faculty of Health and Medical Sciences, University of Surrey, Surrey Research Park, 30 Priestly Road, Guildford, United Kingdom. E-mail: b.goncalves@surrey.ac.uk; and Etsuji Suzuki, Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan. E-mail: etsuji-s@cc.okayama-u.ac.jp.

Copyright © 2025 Wolters Kluwer Health, Inc. All rights reserved.
ISSN: 1044-3983/25/366-732740
DOI: 10.1097/EDE.0000000000001893

Research on exposures that occur before disease occurrence and that might affect disease progression is important, as it could provide mechanistic information on protection against, or progression to, severe forms of disease. Examples of such studies include analyses on postinfection outcomes[1] as well as on cancer outcomes.[2] Despite the existence of these and other examples, the paucity of methodological discussions that specifically focus on this topic has, most likely, hindered interest in effects on disease progression. Specifically, which causal questions could be addressed in these studies is not always clear, partly due to the challenge of accounting for disease occurrence.

Here, we discuss these challenging aspects from a causal perspective, focusing on two factors: (1) whether the disease occurrence is manipulable and (2) the definition of the outcome of interest—or more precisely, the meaning given to the possible levels of the outcome. The manuscript is structured as follows: in the next section, we introduce the notation and the four possible settings defined based on the two factors (disease manipulability and outcome definition); then, in the following sections, we provide in-depth discussions on two of these settings. The first setting represents the scenario where the disease occurrence is not manipulable and the outcome of interest is undefined in individuals who do not develop the disease (see, for example, the related literature[1–3]); and in the second setting, also discussed in detail because it is theoretically possible and because it differs from the first setting with respect to both factors above, disease manipulability is plausible and the severity outcome is defined even in the absence of disease. Further, in the section *Notes on total effect*, we discuss a different type of analysis on disease severity that does not explicitly account for disease occurrence. The final section provides a summary of key points.

NOTATION AND SETTINGS

We let A denote a binary exposure of interest (1 = exposed, 0 = unexposed), S denote a binary post-treatment variable for the presence of disease (1 = present, 0 = absent), and Y denote a binary outcome of interest (1 = present, 0 = absent). S^a and Y^a correspond to potential outcomes of, respectively, variables S and Y under exposure level a, that is, the values these variables would take had the exposure level been, possibly contrary to the fact, a (see Little and Rubin[4] for a detailed discussion on potential outcomes). Given that the exposure here is binary, the following potential outcomes are

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

defined: S^1 and S^0, and Y^1 and Y^0. As discussed below, the definition of Y may subtly differ between settings. For simplicity, throughout, we assume that the exposure A is randomized, and we let U denote a set of unmeasured confounders between S and Y.

Based on the manipulability of the presence of disease S and the definition of outcome Y, we can consider the following four possible settings: Setting I: epidemiologic studies in which the occurrence of disease is assumed not to be manipulable and the disease progression outcome is only meaningful for participants with the disease; Setting II: analyses of disease progression or severity outcomes that can be defined for individuals who do not develop disease and in which disease occurrence is potentially manipulable; Setting III: studies where disease occurrence is potentially manipulable and the definition of the outcome requires disease occurrence; Setting IV: studies in which the occurrence of disease is assumed not to be manipulable that focus on outcomes that can be defined regardless of occurrence of disease. In Table 1, we present examples for each setting.

Although above we list all possible types of analyses implied by this cross-classification based on disease manipulability and outcome definition, we believe that for many medical conditions, investigators often assume that disease occurrence is not manipulable and hence Settings I and IV might be more relevant in current epidemiologic practice. For this reason, in the next section, we present a detailed discussion on Setting I. However, to contrast and better explain differences between possible estimands, subsequently, Setting II is discussed. For completeness, some remarks on Settings III and IV are presented in the section *Other possible settings*.

SETTING I

As mentioned above, the first setting we consider, Setting I, corresponds to the situation where the presence of disease S is assumed not to be manipulable (or alternatively, manipulability of disease occurrence is not considered in the context of the study) and the outcome of interest Y is only defined for individuals with disease[1,2]: an example of possible outcome levels would be "severe disease" and "nonsevere (mild) disease" (rather than "absence of severe disease"). In Setting I, a causal approach applied in previous research[2,5,6] is to use the principal stratification framework,[7] which does not require manipulability of the post-treatment variable (disease state). Under this framework, the study and target populations are conceptualized as constituted by principal strata defined by the joint potential outcomes of S, under different exposure levels a; in other words, the joint variable (S^1, S^0) is used in defining the strata in the principal stratification approach. Different from some of the other situations where principal stratification ideas apply (e.g., noncompliance of assigned treatment), in which the post-treatment variable might affect the outcome but cannot be understood as determining its definition, in Setting I, the post-treatment variable S can be viewed as necessary for the outcome Y, for example, clinical severity of established disease, to be interpretable. For example, effects of vaccines on disease severity outcomes conditional on infection acquisition have been interpreted using principal strata-related parameters[1,5]; in this context, if the research question involves a comparison between severe and mild clinical presentations, vaccine effects are definable only for individuals whose potential infection (or disease occurrence) outcomes (the variables S^1 and S^0) correspond to them being infected regardless of exposure status ($S^1 = S^0 = 1$).

In Table 2, which is analogous to Table 1 in the article by Suzuki[8], we show, for Setting I, potential outcomes S^a and Y^a, under the two levels of exposure A, for both the post-treatment variable S and the outcome of interest Y, severity of presentation of the medical condition, respectively. As mentioned above, in potential outcomes notation, the principal strata for S are defined by (S^1, S^0). Note that here and below we made a negative monotonicity assumption of A on S, that is, $S^1 \leq S^0$ for all individuals (in words, we assumed that the exposure does not cause disease occurrence for any individual); monotonicity was not assumed for the effect of A on Y. The potential disease progression outcome Y^a is undefined when $S^a = 0$, and an effect of A on Y (that is, a contrast of potential outcomes under different levels of exposure for a common set of individuals[9]) cannot be defined for two of the three principal strata (i.e., $(S^1, S^0) = (0,1) \bigvee (0,0)$); it is only defined for the stratum

TABLE 1. Examples of Studies in Each of the Four Settings

	Setting I	Setting II	Setting III	Setting IV
Manipulability of the presence of disease S	Not manipulable	Manipulable	Manipulable	Not manipulable
Definition of outcome Y	Severity of disease	Occurrence of severe disease	Severity of disease	Occurrence of severe disease
Examples	Observational study	Infection challenge study	Infection challenge study	Observational study
A	Early-life exposure (e.g., during childhood)	Vaccination	Vaccination	Early-life exposure (e.g., during childhood)
S	Cancer occurrence	Infection occurrence	Infection occurrence	Cancer occurrence
Y	Cancer clinical severity	Occurrence of hospitalization	Pathogen density in blood	Occurrence of hospitalization

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 2. Potential Outcomes in Setting I

		Presence of disease			Severity of disease	
		S^a			Y^a	
SY type	Response type of *S*	S^1	S^0	Response type of *Y*	Y^1	Y^0
1	Doomed	1	1	Doomed	1	1
2	Doomed	1	1	Causal	1	0
3	Doomed	1	1	Preventive	0	1
4	Doomed	1	1	Immune	0	0
5	Preventive	0	1	NA	Undefined	1
6	Preventive	0	1	NA	Undefined	0
7	Immune	0	0	NA	Undefined	Undefined

Throughout, we make the stable unit treatment value assumption.

of individuals who would suffer from the disease under both exposure levels (i.e., $(S^1, \ S^0) = (1, 1)$), which may justify $\mathrm{E}\left[Y^1 - Y^0 | \ (S^1, \ S^0) = (1, 1)\right]$ as a causal estimand of interest in this setting. Table 2, and this definition of the possible outcome states, also implies that S, presence of disease, has no effect on whether severe disease or mild disease develops; S only determines whether Y can be defined and meaningfully interpreted for a particular individual.

In Figure 1A, under the rule that arrows in causal directed acyclic graphs are drawn when we suspect that there is a direct causal effect for at least one individual in the population,[11] the above indicates that for Setting I an arrow from S to Y cannot be drawn. Note that response type variables (that is, the joint potential outcome variables, $S^T = (S^1, \ S^0)$ and $Y^T = (Y^1, \ Y^0)$) are also presented in Figure 1A Suzuki et al.[10] In fact, this setting has some similarities with studies affected by "truncation by death,"[12,13] where survival determines whether an outcome can be measured and defined but does not affect its value.[8] For instance, if the outcome of interest is quality of life, death before the scheduled outcome assessment implies that the outcome is undefined; it is not merely missing as a result of incomplete follow-up, as the death state has no equivalence to a valid outcome value. However, the similarities between disease

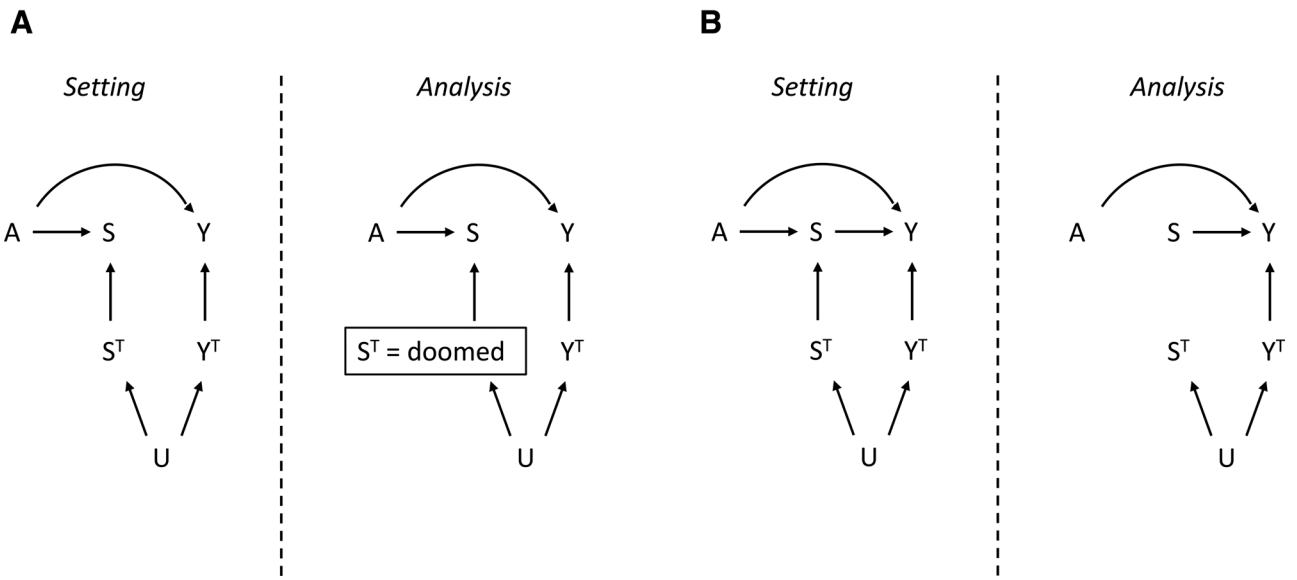

FIGURE 1. Directed acyclic graphs corresponding to causal assumptions in Setting I (Panel A) and Setting II (Panel B). As in the text, A, S, and Y correspond, respectively, to the exposure of interest, a post-treatment variable ("presence of disease"), and the outcome ("severity of disease" in Panel A, and "occurrence of severe disease" in Panel B); A was assumed randomized. S^{T} and Y^{T} correspond to response type variables, as described in the related literature[8,10]; the variable U represents unmeasured factors that affect both S^{T} and Y^{T}. Each panel consists of two subpanels, corresponding to setting-specific causal structures before analysis and for the analysis. The box around S^{T} in Panel A (analysis subpanel) represents the restriction of the principal stratification analysis to the doomed principal stratum. In Panel B (analysis subpanel), the arrows from A to S and from S^{T} to S, that would be present in an observational scenario, were removed to indicate an intervention to set the value of the variable S to s.

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

progression studies and the "truncation by death" setting are only partial, as while death can be considered a competing event, the absence of disease development cannot be considered a competing event. Further, for the binary disease progression outcome Y considered here, although replacing the undefined potential outcomes with the numerical quantity 0 would effectively change the interpretation of the outcome of interest, this modified outcome might still be meaningful. The fact that the "severity of disease" outcome is undefined for individuals without the condition becomes more obvious when the focus is on quantitative Y variables that reflect disease severity, such as the size of lesions, viral load or number of days from diagnosis to recovery—in these situations, the use of 0 for the variable Y in nondiseased individuals would be less meaningful. Consistent with this, in discussing their analysis on viral load endpoints in HIV vaccine trials, Hudgens and colleagues[3] mentioned that assigning zero viral load to uninfected individuals would imply an interest in the combination of effects on susceptibility and disease progression.

SETTING II

The second setting we discuss is relevant when it is possible to manipulate disease occurrence, that is, when it is possible to conceive an intervention that would solely change disease status, and when the outcome of interest Y is defined regardless of disease status. In this Setting II, a possible outcome Y would be "occurrence of severe disease" (which may be defined even among those who do not develop the disease), rather than "severity of disease" (which is only defined in diseased individuals); the implication is that $Y = 1$ and $Y = 0$ correspond then to "presence of severe disease" and "absence of severe disease," respectively; the latter category includes "presence of nonsevere (mild) disease," as in Setting I, as well as "absence of disease." A research scenario that might correspond to Setting II is that of infection challenge studies. These are studies where participants are exposed to a particular infectious agent in a controlled setting as part of the study protocol; there are multiple examples of challenge studies in the literature that focused on different pathogens (see, for example, the related literature[14–16] and the recent review by Abo et al[17]). A hypothetical example that is relevant for Setting II would thus be an infection challenge study in which vaccination (which would correspond to the exposure, A) is randomly assigned, and all participants receive an inoculum dose of the pathogen that ensures infection occurrence (S), and the outcome of interest is occurrence of hospitalization (Y); it is worth noting that in challenge studies, infection status is manipulated indirectly via exposure to the pathogen. Here, a possible inferential goal would be the quantification of controlled direct effects as a causal estimand,[18,19] where disease occurrence S would represent a controlled mediator in a directed path from exposure A to outcome Y. Potential outcomes of the form Y^{as} would thus be needed to represent the outcome that an individual would develop had she or he, possibly contrary to the fact, been exposed to level a, and had the disease status been set to s (Table 3); thus, potential outcomes of this form, Y^{as}, imply interventions on both the exposure and the disease occurrence. Note that in this setting, because of the definition of Y, none of the potential outcomes Y^{as} is undefined; rather, even among individuals with $S^a = 0$, all of the four potential outcomes of Y can be defined. Different from the principal effect

TABLE 3. Potential Outcomes in Setting II

		Presence of disease		Occurrence of severe disease					
		S^a		Y^{as}				Y^{aS^a}	
***SY* type**	**Response type of S**	S^1	S^0	Y^{11}	Y^{01}	Y^{10}	Y^{00}	Y^{1S^1}	Y^{0S^0}
1	Doomed	1	1	1	1	(0)	(0)	1	1
2	Doomed	1	1	1	0	(0)	(0)	1	0
3	Doomed	1	1	0	1	(0)	(0)	0	1
4	Doomed	1	1	0	0	(0)	(0)	0	0
5	Preventive	0	1	(1)	1	0	(0)	0	1
6	Preventive	0	1	(1)	0	0	(0)	0	0
7	Preventive	0	1	(0)	1	0	(0)	0	1
8	Preventive	0	1	(0)	0	0	(0)	0	0
9	Immune	0	0	(1)	(1)	0	0	0	0
10	Immune	0	0	(1)	(0)	0	0	0	0
11	Immune	0	0	(0)	(1)	0	0	0	0
12	Immune	0	0	(0)	(0)	0	0	0	0

Parentheses indicate potential outcomes that are not observable in observational scenarios. In addition to potential outcomes of the form Y^{as}, used to define controlled direct effects, and to the compound potential outcomes Y^{aS^a}, that can be used to define total exposure effects (see the section *Notes on total effect*), we also present the potential outcomes S^a, and the corresponding response types, to explicate that the 12 *SY* types in the table correspond to possible response types based on the joint potential outcomes Y^{a1}, i.e. Y^{11} and Y^{01}, for each of the three response types defined by (S^1, S^0).

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

defined above (i.e., $\mathrm{E}\left[Y^1 - Y^0 | \left(S^1,\ S^0\right) = (1,1)\right]$), whose target population is a subpopulation with $\left(S^1,\ S^0\right) = (1,1)$, controlled direct effects are generally defined for the total population as the target, and under an intervention that sets the value of the mediator S to 1, a controlled direct effect [CDE] $(s = 1)$ corresponds to a contrast between Y^{11} and Y^{01} (i.e., $\mathrm{E}\left[Y^{11} - Y^{01}\right]$). Note that Y^{a0} is, by definition, 0 in all individuals, although it may be unobservable in observational settings. This implies that, in the context discussed here, under an intervention that sets the value of the mediator S to 0, a CDE $(s = 0)$ is 0 (i.e., $\mathrm{E}\left[Y^{10} - Y^{00}\right] = 0$). Relatedly, note that the effect of S on Y, which is defined by comparing Y^{a1} and Y^{a0}, may be present in all individuals, except for SY types 4, 8, and 12 (Table 3), which may justify drawing an arrow from S to Y in Figure 1B.

COMPARING SETTINGS I AND II

In Figure 2, we present an illustration of the different definitions of outcomes; this graphical representation is analogous to a statement by Zhang and Rubin[13] that the outcome for individuals with "truncation by death" would only be defined in an extended sample space (here, {Severity variable range, $*$}, where $*$ stands for the outcome with truncation). It is useful to mention that graphs similar to those in Figure 2 could be conceived for an example described in Rubin[12] on the effect of a job training program on wages, rather than income (where the proportion of those unemployed and the distribution of wages of those employed, and the fraction with high wage [or income], would then be represented). Although the observed data would be the same in both cases, the outcome definition would change, focusing on wages implies relevance of a modified Panel B, while focusing on income, relevance of a modified Panel C. In other words, trivially, wage can be defined only among those who are employed, while income is 0, rather than undefined, among those who are unemployed. Similarly, in Setting I, the outcome Y can be conceived as a dichotomized version of a continuous severity scale that is only defined for diseased individuals; the focus in Setting II is on the presence or absence of severe disease, rather than the relative frequencies of mild and severe presentations. Consistent with this, in their study on the effect of finasteride on the severity of prostate cancer, Shepherd and colleagues[2] dichotomized an ordinal severity score used to grade prostate cancer. Related arguments have

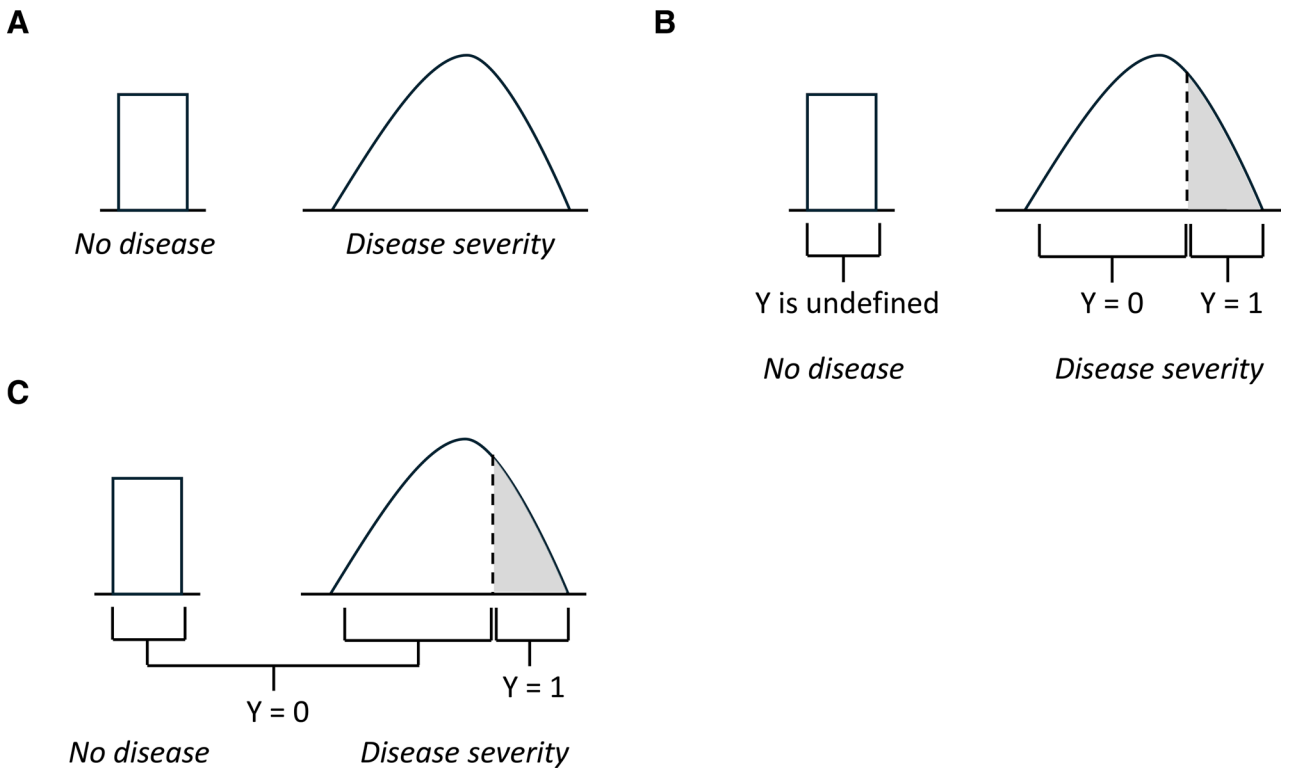

FIGURE 2. Illustration of outcomes of interest for the different settings discussed here. In Panels A–C, y-axes, which are not shown, represent numbers of individuals in a hypothetical study population. Panel A presents the distribution of disease severity, as a continuous variable (x-axis), among cases, and also shows that this quantity corresponding to disease severity is undefined for individuals who do not have the disease (separate x-axis section). In Panel B, the different levels of the outcome variable (Y) for Setting I and Setting III are shown. The outcome levels defined in Panel C imply that interest lies in the occurrence of severe disease (Setting II and Setting IV), rather than the relative frequencies of different levels of disease severity. Note that Panel B is more informative than Panel C, as the former includes information on three outcome-related groups, rather than two, as in the latter panel.

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

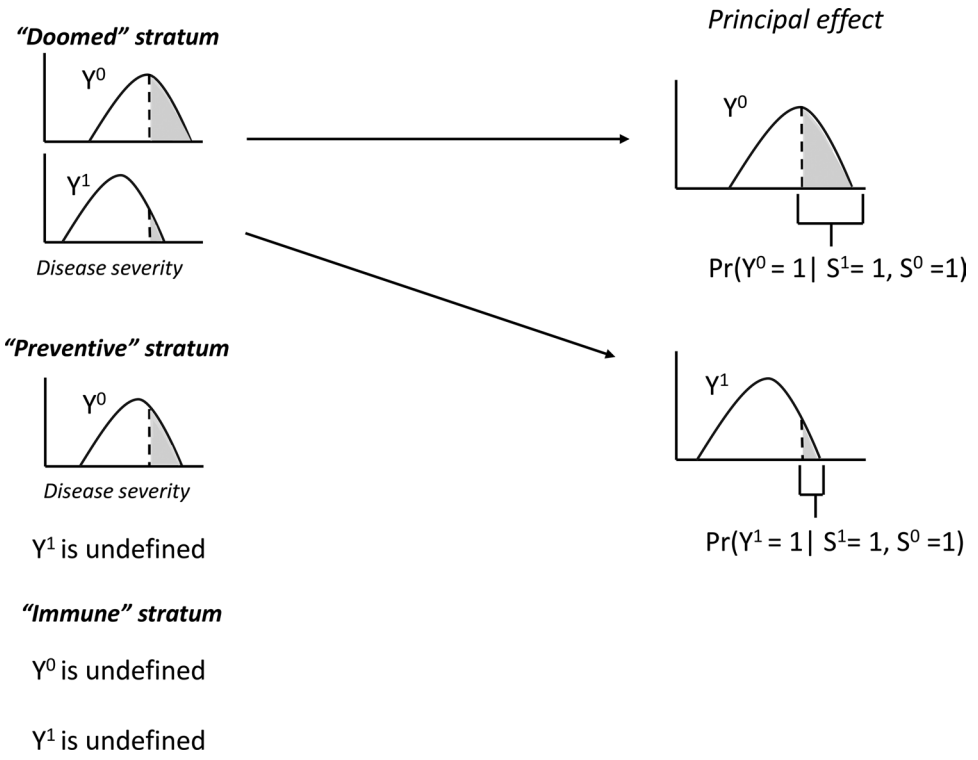

FIGURE 3. Illustration of Setting I. Distributions of potential values for the outcome Y are presented for each principal stratum and treatment level. As mentioned in the text and in **Table 2**, for the "preventive" stratum the potential outcome Y is undefined under treatment level 1; for the "immune" stratum, the potential values for Y are undefined under both treatment levels, 0 and 1. The right side of the figure shows components of the target comparison for the principal effect estimand; here, this is a contrast between the proportions of the distributions above a certain threshold of disease severity.

been made in the "truncation by death" setting: Young and Stensrud[20] suggested that while for a continuous outcome such as birth weight, failure to conceive or miscarriage means that the outcome is undefined, for other outcomes, for example neonatal intensive care unit admission, failure to conceive or miscarriage might imply that the outcome is zero (when using 0/1 coding), rather than undefined; see also the Appendix D of the work by Young and colleagues.[21] However, as mentioned above, using a binary variable for disease progression outcome Y means that the distinction between outcome definitions under these two settings is a subtle one. In Figure 3 and eFigure 1; https://links.lww.com/EDE/C258 we illustrate distributions of potential values for the outcome of interest Y by S-defined principal strata and how they determine the causal contrasts of Settings I and II, respectively.

Should studies on disease progression and severity aim to estimate a principal stratum effect or a controlled direct effect? In Table 4, we present comparisons between the two settings discussed above. Specifically, as mentioned above, one factor to consider is the manipulability of S, which could be plausible for some conditions but not others. In fact, in the disease progression setting, where the mediator is a complex process (disease development), for many noninfectious conditions, it might be difficult to conceive a valid intervention on S. Having said that, when a key step of the pathogenesis has been identified to be triggerable by intervention, manipulability of disease occurrence is plausible; examples of procedures that can induce a medical condition include methacholine challenge, a test that triggers bronchoconstriction in asthma patients,[22,23] and allergen exposure chambers for allergic patients.[24] Further, for infectious diseases, in the context of challenge studies, it could be argued that controlling infection status is plausible for some pathogens even though in these studies what is controlled is the variable "exposure to pathogen"; see also work by Stensrud and Smith,[25] that discusses vaccine effects and their relation to exposure to infectious agents, and Chiu et al[26] for a discussion on the interpretation of controlled direct effects in a structurally similar situation. Moreover, although controlled direct effects can provide insights into the underlying causal structures, if the goal is to guide public health decision-making, even when valid hypothetical interventions on disease occurrence are, or become, plausible, it is unclear, for example, that the inclusion of a stratum corresponding to the immune response type (i.e., $(S^1, \ S^0) = (0, 0)$) in the estimation of a CDE $(s = 1)$ that has as target the total population would be desired, as in real-world conditions individuals in this stratum would not develop the disease of interest. Estimation of the principal effect in the doomed stratum (i.e., $(S^1, \ S^0) = (1, 1)$) would answer the question of

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

TABLE 4. Comparative Summary of Setting I and Setting II

	Setting I (Table 2 and Figure 1A)	Setting II[a] (Table 3 and Figure 1B)
	Setting-defining factors	
Manipulability of the presence of disease S	Not manipulable	Manipulable
Definition of outcome Y	Severity of disease	Occurrence of severe disease
Examples of outcome Y (B, binary; Q, quantitative)	Hospitalization in individuals with disease (B), disease-related death in individuals with disease (B), pathogen levels (Q), duration of disease-related hospitalization (Q)	Hospitalization (B), death (B)
Analytical considerations		
Example of research questions answerable in the setting	What is the preventive effect of A on severity of disease Y among those who suffer from disease S regardless of whether they receive A or not?	What is the preventive effect of A on the occurrence of severe disease Y among the total population, when all the individuals are set to suffer from the disease S?
Target causal estimand	$\mathrm{E}\left[Y^1 - Y^0 \middle\vert \left(S^1,\ S^0\right) = (1,1)\right]$	$\mathrm{E}\left[Y^{11} - Y^{01}\right]$ $\left(\mathrm{E}\left[Y^{10} - Y^{00}\right] = 0\right)$
Target of (potentially hypothetical) intervention	A	A and S
Target population	"Doomed" individuals with $\left(S^1,\ S^0\right) = (1,1)$	Total population

[a]Note that for analyses that focus on the different research question of estimating the total effect of an exposure, that is, on $\mathrm{E}\left[Y^1 - Y^0\right]$, the definition of the outcome might be similar to that of Setting II, and explicit consideration of S, in particular with regard to its manipulability, is not needed.

whether exposure affects severity of disease presentation in the only group of individuals for whom the question is valid; for example, it could help to understand whether vaccines reduce clinical severity, or duration of symptoms, even when infection itself is not preventable by the vaccine or whether a particular exposure affects oncologic outcomes (e.g. survival) in individuals who would have cancer regardless of exposure history. The general rationale for using this estimand in studies on disease progression is thus related to quantifying, or at least defining a plausible range of, effects that are directly relevant to real-world settings where control of the disease occurrence is not, and will not be, in place and that are not explained by the relevant exposure's effect on disease occurrence. This approach, however, suffers from nonidentifiability under standard assumptions (note that sensitivity analysis methods have been described that could be applied in this context; for example using additive scale, see Chiba and VanderWeele[27]) and, relatedly, the fact that this stratum is not discernible.

OTHER POSSIBLE SETTINGS

Above, we focused our discussion only on two settings. Setting I is likely more relevant in current practice, because disease occurrence is often assumed to be nonmanipulable, or at least its manipulability is not explicitly discussed by investigators, and because of the broader-in-scope outcome definition (e.g., binary outcomes for which imputation of zero would result in meaningful, although different, outcomes and quantitative outcomes, for which this imputation in nondiseased individuals might not be interpretable). Setting II is also discussed because it represents, at least theoretically, a possible scenario where disease occurrence is manipulable, and because, as it differs from Setting I also in terms of type of outcome definition, it provides an opportunity to discuss this latter issue in some detail and contrast estimands in these settings. However, two other settings are conceivable, and in the eAppendix we present tables (eTables 1–3; https://links.lww.com/EDE/C258), and corresponding causal diagrams (eFigure 2; https://links.lww.com/EDE/C258) for these settings. Briefly, Setting III corresponds to the situation where disease is manipulable but the outcome is "severity of disease," being only defined for patients with the disease; here investigators could either estimate a principal effect in the doomed stratum, as in Setting I, or a $\mathrm{CDE}\,(s = 1)$, as in Setting II, depending on their research question. Setting IV on the other hand represents the scenario of a nonmanipulable condition where the focus is on "occurrence of severe disease" (see eTable 3; https://links.lww.com/EDE/C258 for examples of binary outcomes in this setting); in this case, an analysis on disease progression would involve principal effects and the target population is the doomed principal stratum. In fact, Setting IV could be viewed as the relevant setting whenever it is possible to meaningfully impute zero for the nondiseased individuals in Setting I; note that for many quantitative outcomes (e.g. pathogen density in the blood or number of skin lesions), an imputed value of zero would change the interpretation of the outcome definition, and for quantitative outcomes such as these, that are more meaningful for diseased individuals, Setting I is more relevant. Note also that although in discussing the appropriate causal estimands for Setting III and Setting IV (eTable 3; https://links.lww.com/EDE/C258) we do not introduce estimands different from those for Setting I and Setting II (Table 4), the corresponding tables (eTables 1 and 2; https://links.lww.com/

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

EDE/C258 and Tables 2 and 3) of implied potential outcomes are different. Finally, it is important to clarify that although we discussed Setting II in more detail than Setting IV, Setting IV might be more relevant for future research on conditions for which manipulability of occurrence is not envisaged.

NOTES ON TOTAL EFFECT

While our focus here is on analyses that aim to quantify effects on disease progression (i.e., on analyses that account for the [potential] disease occurrence when assessing the impact of an exposure on disease severity), the meaning of the outcome levels in Setting II ("presence of severe disease" and "absence of severe disease") is also the definition used when studying the effect on severe disease occurrence that involves an intervention only on the exposure, rather than on the exposure and the mediator. This effect of the exposure on severe disease occurrence corresponds to the contrast between the potential outcomes Y^a under the two exposure levels, that is, Y^1 and Y^0, both of which are, by the outcome definition above, theoretically observable among all individuals. Although under the composition assumption,[28,29] which states that $Y^a = Y^{aS^a}$, the total effect of A on Y could be defined using the compound potential outcomes presented in the rightmost columns of Table 3 that imply interventions on both A and S, Y^{1S^1} and Y^{0S^0}, here we refer to total effect as a contrast of potential outcomes of the form Y^a, which does not require the composition assumption. Notice that although the outcome definition in Setting II and Setting IV is similar to that of analyses on total effects, the latter do not require consideration of manipulability of S, and in this case, the four settings do not apply.

It is worth noting that for the purpose of improving public health strategies, where mechanistic insights might not be a priority, this total effect, which does not involve consideration of the variable S and hence does not correspond to a principal effect or a controlled direct effect, would, in many cases, provide the information needed for policy decision-making. Indeed, in addition to effects on outcomes that are not defined with respect to disease progression, for example, infection occurrence, many recent observational studies on vaccines and COVID-19 also report effects on severe outcomes (e.g., hospitalization) that are unconditional on disease or infection occurrence and on the corresponding potential outcome variables.[30–32]

CONCLUSIONS

Our focus here was on the description and comparison of disease progression settings; however, it is important to mention that there is a rich literature on the identification and estimation of the estimands discussed here; in addition to established results for direct effects,[19,29,33] of particular relevance for principal effects are the work by Hudgens and Halloran on postinfection outcomes[1] and the statistical literature on the related problem of "truncation by death."[27,34–36]

In sum, we argue that studies that aim to quantify effects of factors on disease progression might imply different types of outcome definitions and involve either (1) the nondiscernibility of the population strata for which a causal estimand can be defined (under the principal stratification framework) or (2) the conceivability of disease occurrence manipulation (which is required for controlled direct effects). In describing the complexity of studying disease progression, our discussion supports the view that these studies are fundamentally different from those examining total effect as a causal estimand. Therefore, careful scrutiny is required to address the causal estimands of interest in disease progression studies so that this type of epidemiologic research is appropriately used when guiding decision-making processes and, perhaps more relevantly, for gaining insights into mechanisms of disease progression and protection.

REFERENCES

1. Hudgens MG, Halloran ME. Causal vaccine effects on binary postinfection outcomes. *J Am Stat Assoc*. 2006;101:51–64.
2. Shepherd BE, Redman MW, Ankerst DP. Does finasteride affect the severity of prostate cancer? A causal sensitivity analysis. *J Am Stat Assoc*. 2008;103:1392–1404.
3. Hudgens MG, Hoering A, Self SG. On the analysis of viral load endpoints in HIV vaccine trials. *Stat Med*. 2003;22:2281–2298.
4. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000;21:121–145.
5. Halloran ME, Longini IM, Struchiner CJ. Design and Analysis of Vaccine Studies. Statistics for Biology and Health. Springer. 2010.
6. Gonçalves BP, Suzuki E. Preventable fraction in the context of disease progression. *Epidemiology*. 2024;35:801–804.
7. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics*. 2002;58:21–29.
8. Suzuki E. Generalized causal measure: the beauty lies in its generality. *Epidemiology*. 2015;26:490–495.
9. Rubin DB. Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc*. 2005;100:322–331.
10. Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A counterfactual approach to bias and effect modification in terms of response types. *BMC Med Res Methodol*. 2013;13:101.
11. Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams: pitfalls and tips. *J Epidemiol*. 2020;30:153–162.
12. Rubin DB. Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. *Stat Sci*. 2006;21:299–309.
13. Zhang JL, Rubin DB. Estimation of causal effects via principal stratification when some outcomes are truncated by "death". *J Educ Behav Stat*. 2003;28:353–368.
14. Talaat KR, Alaimo C, Martin P, et al. Human challenge study with a Shigella bioconjugate vaccine: analyses of clinical efficacy and correlate of protection. *EBioMedicine*. 2021;66:103310.
15. Minassian AM, Silk SE, Barrett JR, et al. Reduced blood-stage malaria growth and immune correlates in humans following RH5 vaccination. *Med*. 2021;2:701–719.e19.
16. Rapeport G, Smith E, Gilbert A, Catchpole A, McShane H, Chiu C. SARS-CoV-2 human challenge studies - establishing the model during an evolving pandemic. *N Engl J Med*. 2021;385:961–964.
17. Abo YN, Jamrozik E, McCarthy JS, Roestenberg M, Steer AC, Osowicki J. Strategic and scientific contributions of human challenge trials for vaccine development: facts versus fantasy. *Lancet Infect Dis*. 2023;23:e533–e546.

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.

18. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3:143–155.
19. Pearl J. Direct and Indirect Effects. In: Proceedings of the Seventeenth Conference on Uncertainy in Articial Intelligence. San Francisco, CA. 2001.
20. Young JG, Stensrud MJ. Identified versus interesting causal effects in fertility trials and other settings with competing or truncation events. *Epidemiology*. 2021;32:569–572.
21. Young JG, Stensrud MJ, Tchetgen Tchetgen EJ, Hernán MA. A causal framework for classical statistical estimands in failure-time settings with competing events. *Stat Med*. 2020;39:1199–1236.
22. Mauger EA, Mauger DT, Fish JE, Chinchilli VM, Israel E; Asthma Clinical Trials Network. Summarizing methacholine challenges in clinical research. *Control Clin Trials*. 2001;22:244S–251S.
23. Cockcroft DW, Davis BE, Roh Y, Lourens J-A. Effect of ingested H(1) antihistamines on methacholine challenge. *J Allergy Clin Immunol*. 2015;135:579–580.
24. Pfaar O, Zieglmayer P. Allergen exposure chambers: implementation in clinical trials in allergen immunotherapy. *Clin Transl Allergy*. 2020;10:33.
25. Stensrud MJ, Smith L. Identification of vaccine effects when exposure status is unknown. *Epidemiology*. 2023;34:216–224.
26. Chiu YH, Stensrud MJ, Dahabreh IJ, et al. The effect of prenatal treatments on offspring events in the presence of competing events: an application to a randomized trial of fertility therapies. *Epidemiology*. 2020;31:636–643.
27. Chiba Y, VanderWeele TJ. A simple method for principal strata effects when the outcome has been truncated due to death. *Am J Epidemiol*. 2011;173:745–751.
28. Suzuki E, Yamamoto E, Tsuda T. Identification of operating mediation and mechanism in the sufficient-component cause framework. *Eur J Epidemiol*. 2011;26:347–357.
29. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Its Interface*. 2009;2:457–468.
30. Barda N, Dagan N, Cohen C, et al. Effectiveness of a third dose of the BNT162b2 mRNA COVID-19 vaccine for preventing severe outcomes in Israel: an observational study. *Lancet*. 2021;398:2093–2100.
31. Dagan N, Barda N, Biron-Shental T, et al. Effectiveness of the BNT162b2 mRNA COVID-19 vaccine in pregnancy. *Nat Med*. 2021;27:1693–1695.
32. Dickerman BA, Gerlovin H, Madenci AL, et al. Comparative effectiveness of third doses of mRNA-based COVID-19 vaccines in US veterans. *Nat Microbiol*. 2023;8:55–63.
33. VanderWeele TJ. Controlled direct and mediated effects: definition, identification and bounds. *Scand Stat Theory Appl*. 2011;38:551–563.
34. Tchetgen Tchetgen EJ. Identification and estimation of survivor average causal effects. *Stat Med*. 2014;33:3601–3628.
35. Wang L, Zhou XH, Richardson TS. Identification and estimation of causal effects with outcomes truncated by death. *Biometrika*. 2017;104:597–612.
36. Chiba Y, Taguri M, Uemura Y. On the identification of the survivor average causal effect. *J Biomet Biostat*. 2011;02:1000e104.

© 2025 Wolters Kluwer Health, Inc. All rights reserved.

Copyright © 2025 Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.