

Review Article

Unraveling Causality: Innovations in Epidemiologic Methods

Etsuji Suzuki

Abstract:

For several decades, the counterfactual model and the sufficient cause model have shaped our understanding of causation in biomedical science and, more recently, the link between these two models has enabled us to obtain a deeper understanding of causality. In this article, I provide a brief overview of these fundamental causal models using a simple example. The counterfactual model focuses on one particular cause or intervention and gives an account of the various effects of that cause. By contrast, the sufficient cause model considers sets of actions, events, or states of nature which together inevitably bring about the outcome under consideration. In other words, the counterfactual framework addresses the question “what if?” while the sufficient cause framework addresses the question “why does it happen?” Although these two models are distinct and address different causal questions, they are closely related and used to elucidate the same cause-effect relationships. Importantly, the sufficient cause model makes clear that causation is a multifactorial phenomenon, and it is a “finer” model than the counterfactual model; an individual is of one and only one response type in the counterfactual framework, whereas an individual may be at risk of none, one, or several sufficient causes. Understanding the link between the two causal models can provide greater insight into causality and can facilitate the use of each model in appropriate contexts, highlighting their respective strengths. I will briefly present three topics of interest from our research: the relationship between the concepts of confounding and of covariate balance; distinctions between attributable fractions and etiologic fractions; and the identification of operating mediation and mechanism. It is important to scrutinize observed associations in a complementary manner, using both the counterfactual model and the sufficient cause model, employing both inductive and deductive reasoning. This holistic approach will better help us to unravel causality.

Key Words:

attributable fraction, causal inference, confounding, counterfactual model, covariate balance, etiologic fraction, mediation, sufficient cause model

Seeing the Invisible – How?

Causality is central to clinical and epidemiologic studies. Without an understanding of cause-effect relationships, we cannot answer basic questions such as “Does this treatment harm or help patients?” or “Does exposure to this material have any effects on our health?” These are causal questions, and humans are innately interested in these types of questions. From birth, we continually gain knowledge about cause-effect relationships around us, sometimes through undesirable experiences, and strive to apply this knowledge to improve our lives and health. Simply put, humans are by nature inquisitive about causality.

In the era of data science, there is growing interest in using data to infer causation in biomedical science, and in this context, artificial intelligence is also of increasing interest. Rapid

progress has been made in concepts and methods for causal inference, leading to innovations in epidemiologic methods that address causal questions in biomedical science. However, as expressed by the axiom “association is not causation,” this is not always an easy task. Furthermore, in most situations, causes in biomedical science are neither necessary nor sufficient.

To grasp the nature of causal inference from data, it is important to understand the underlying, though unobservable, structures behind the observed data. How can we see the invisible? Understanding formal causal models is key⁽¹⁾. For several decades, the counterfactual model and the sufficient cause model have shaped our understanding of causation in biomedical science^{(2), (3), (4), (5), (6)}, and more recently, the link between these two models has enabled us to obtain a deeper understanding of causality^{(7), (8), (9), (10), (11), (12)}. In this article, I provide a brief overview of these fundamental causal models using a

Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan

Corresponding author: Etsuji Suzuki, etsuji-s@cc.okayama-u.ac.jp

JMA J. 2025;8(2):323-337

Received: August 27, 2024 / Accepted: January 20, 2025 / Advance Publication: April 21, 2025 / Published: April 28, 2025

Copyright © Japan Medical Association

	Exposed group ($E = 1$)		Unexposed group ($E = 0$)		Total population	
	N	Proportion	N	Proportion	N	Proportion
Outcome occurred ($Y = 1$)	234	0.390	120	0.300	354	0.354
Outcome did not occur ($Y = 0$)	366	0.610	280	0.700	646	0.646
Total	600	1.000	400	1.000	1000	1.000
Risk difference estimate	$\Pr(Y = 1 E = 1) - \Pr(Y = 1 E = 0) = 0.390 - 0.300 = 0.090$					
Risk ratio estimate	$\Pr(Y = 1 E = 1)/\Pr(Y = 1 E = 0) = 0.390/0.300 = 1.300$					

Table 1. Observed Data in a Hypothetical Cohort Study.

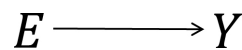


Figure 1. A causal diagram of an exposure E and an outcome Y .

simple example, highlighting how innovations in epidemiologic methods contribute to unraveling causality.

A Hypothetical Example

As an illustration, I use a hypothetical cohort study in **Table 1**. I let E denote a binary exposure of interest (1 = exposed, 0 = unexposed) and Y a binary outcome (1 = outcome occurred, 0 = outcome did not occur). The prevalence of exposure E is 0.6 (i.e., $\Pr(E = 1) = 0.6$). We are interested in the causation between E and Y , which is described as an arrow from E to Y in **Figure 1**. (For the meaning of arrows in causal directed acyclic graphs ^{(13), (14), (15)}, see Section 2-2 of Suzuki et al. ⁽¹⁶⁾) From the observed information in **Table 1**, we can readily calculate the risk difference (RD) estimate as $\Pr(Y = 1|E = 1) - \Pr(Y = 1|E = 0) = 0.390 - 0.300 = 0.090$. Similarly, we can calculate the risk ratio (RR) estimate as $\Pr(Y = 1|E = 1)/\Pr(Y = 1|E = 0) = 0.390/0.300 = 1.300$. To sum up, there is a positive association between E and Y in the population, which may imply the presence of causation between E and Y .

In the following sections, I discuss the underlying data distributions in this cohort study, which, though unobservable in the real world, are important to understand causality. I assume that there is no loss to follow-up and that all variables are measured without errors ⁽¹⁷⁾. For simplicity, I also ignore random errors in this article.

Through the Lens of Counterfactuals

The counterfactual (or potential outcome) model focuses on one particular cause or intervention and gives an account of the various effects of that cause. This framework addresses the question, “What would have occurred if a particular factor were intervened upon and thus set to a different level than it in fact was?” ⁽⁹⁾ Indeed, such “what if?” questions are prevalent

in our lives. The idea of conceptualizing causation in terms of counterfactuals can be traced at least as far back as the 18th-century philosopher, David Hume. In his book *An Enquiry Concerning Human Understanding* ⁽¹⁸⁾, Hume has put forward a counterfactual formulation for causation in the following famous passage: “[W]e may define a cause to be *an object, followed by another, and where all the objects similar to the first are followed by objects similar to the second*. Or in other words *where, if the first object had not been, the second never had existed*.” Since then, counterfactual theories of causation have been further developed, and the counterfactual model has become the dominant framework for contemporary causal thinking in epidemiology ^{(1), (19), (20)}.

In the counterfactual framework, I let Y^e denote the outcome that would have occurred if, possibly contrary to fact, there had been interventions to set E to e . Because E is binary, there are two potential outcomes, Y^1 and Y^0 , corresponding to what would have happened to an individual when the person was exposed and unexposed, respectively. Note that Y^1 and Y^0 are also random variables. Throughout this article, I assume that the consistency assumption is met ^{(21), (22)}, which implies that the observed outcome for an individual is the potential outcome, as a function of intervention, when the intervention is set to the observed exposure (i.e., $Y^e = Y$ if $E = e$ ($e = 0, 1$)). Therefore, for each individual, we only get to observe the outcome corresponding to the actual exposure; one of the potential outcomes is observed. Meanwhile, another potential outcome remains unobserved. This is called “the Fundamental Problem of Causal Inference” ⁽²³⁾ because the causal effect of E on Y for a given individual is defined to be present if $Y^1 \neq Y^0$.

What is the underlying data structure of the hypothetical cohort study through the lens of counterfactuals? Based on the joint potential outcomes (Y^1, Y^0), individuals can be classified into the following four (= 2²) response types ⁽⁴⁾:

Type 1 or “doomed” persons: Exposure is irrelevant be-

Response type	Potential outcomes		Exposed group ($E = 1$)		Unexposed group ($E = 0$)		Total population	
	Y^1	Y^0	N	Proportion	N	Proportion	N	Proportion ^b
1 (doomed)	1	1	84	$p_1 = 0.140$	100	$q_1 = 0.250$	184	$r_1 = 0.184$
2 (causal)	1	0	150	$p_2 = 0.250$	80	$q_2 = 0.200$	230	$r_2 = 0.230$
3 (preventive)	0	1	120	$p_3 = 0.200$	20	$q_3 = 0.050$	140	$r_3 = 0.140$
4 (immune)	0	0	246	$p_4 = 0.410$	200	$q_4 = 0.500$	446	$r_4 = 0.446$
Total			600	1.000	400	1.000	1000	1.000
Target population ^c								
			Exposed group ($E = 1$)		Unexposed group ($E = 0$)		Total population	
Risk when exposed			$p_1 + p_2 = \mathbf{0.390}$		$q_1 + q_2 = 0.450$		$r_1 + r_2 = 0.414$	
Risk when unexposed			$p_1 + p_3 = 0.340$		$q_1 + q_3 = \mathbf{0.300}$		$r_1 + r_3 = 0.324$	
Measures of effect								
causal risk difference			$(p_1 + p_2) - (p_1 + p_3) = 0.050$		$(q_1 + q_2) - (q_1 + q_3) = 0.150$		$(r_1 + r_2) - (r_1 + r_3) = 0.090$	
causal risk ratio			$(p_1 + p_2)/(p_1 + p_3) \approx 1.147$		$(q_1 + q_2)/(q_1 + q_3) = 1.500$		$(r_1 + r_2)/(r_1 + r_3) \approx 1.278$	
Measures of association								
associational risk difference			$(p_1 + p_2) - (q_1 + q_3) = 0.090$					
associational risk ratio			$(p_1 + p_2)/(q_1 + q_3) = 1.300$					

^aI consider a binary exposure E ($1 = \text{exposed}$, $0 = \text{unexposed}$) and a binary outcome Y ($1 = \text{outcome occurred}$, $0 = \text{outcome did not occur}$). I consider two potential outcomes, Y^e , for an individual.

^bNote that r_j can be calculated as: $p_j \times \Pr(E = 1) + q_j \times \Pr(E = 0)$, where $\Pr(E = e)$ represents the prevalence of $E = e$ in the total population.

^cIn this cohort study, only the numbers shown in bold can be observed under the assumption of consistency.

Table 2. Underlying Hypothetical Data of the Hypothetical Cohort Study in Terms of Response Types^a.

cause outcome occurs with or without exposure (i.e., $(Y^1, Y^0) = (1, 1)$)

Type 2 or “causal” persons: Outcome occurs if and only if they are exposed (i.e., $(Y^1, Y^0) = (1, 0)$)

Type 3 or “preventive” persons: Outcome occurs if and only if they are unexposed (i.e., $(Y^1, Y^0) = (0, 1)$)

Type 4 or “immune” persons: Exposure is irrelevant because outcome does not occur with or without exposure (i.e., $(Y^1, Y^0) = (0, 0)$)

In **Table 2**, I show underlying hypothetical data in terms of response types. I let p_j , q_j , and r_j , $j = 1-4$, be the proportions of response type j in the exposed group, the unexposed group, and the total population, respectively. As described below, the concept of response type is fundamental to causal inference because the causal effect of exposure on disease frequency in a population is determined by the distribution of response types of individuals in that population, and not necessarily by the population distribution of covariates⁽⁴⁾. This point, however, has been relatively underappreciated because, despite its importance, the response type of each individual is unobservable⁽²⁴⁾.

Recall that, when calculating the RD and RR estimates in **Table 1**, two conditional probabilities, $\Pr(Y = 1|E = 1)$ and $\Pr(Y = 1|E = 0)$, were used. Among the exposed group, only type 1 and type 2 persons develop the outcome Y , and the risk, or incidence proportion, of the outcome in the exposed group

is given as $\Pr(Y = 1|E = 1) = \Pr(Y^1 = 1|E = 1) = p_1 + p_2 = 0.140 + 0.250 = 0.390$. Similarly, among the unexposed group, only type 1 and type 3 persons develop the outcome Y , and the corresponding risk is given as $\Pr(Y = 1|E = 0) = \Pr(Y^0 = 1|E = 0) = q_1 + q_3 = 0.250 + 0.050 = 0.300$. Therefore, the associational RD is obtained using the proportions of response types as $(p_1 + p_2) - (q_1 + q_3) = 0.390 - 0.300 = 0.090$, which is equivalent to the RD estimate. Similarly, the associational RR is obtained as $(p_1 + p_2)/(q_1 + q_3) = 0.390/0.300 = 1.300$, which is equivalent to the RR estimate. These are *measures of association* that can be observed in the real world.

By contrast, an average causal effect of E on Y in the population is defined to be present if $E[Y^1] = \Pr(Y^1 = 1) \neq \Pr(Y^0 = 1) = E[Y^0]$ in the population of interest (**Figure 2**). Note that these are not conditional but marginal probabilities. Importantly, the concept of the target population plays a key role in causal inference^{(25), (26), (27), (28), (29)}. Target parameters for causal inference, or *measures of effect*, cannot be defined unless the target population is clearly defined. For example, when the total population is used as the target population, the counterfactual risk when everyone in the total population is exposed is given as $\Pr(Y^1 = 1) = r_1 + r_2$. In a similar manner, the counterfactual risk when everyone in the total population is unexposed is given as $\Pr(Y^0 = 1) = r_1 + r_3$. Accordingly, the causal RD in the total population is calculated as $(r_1 + r_2) - (r_1 + r_3) = 0.414 - 0.324 = 0.090$, whereas the causal RR is calculated as $(r_1 +$

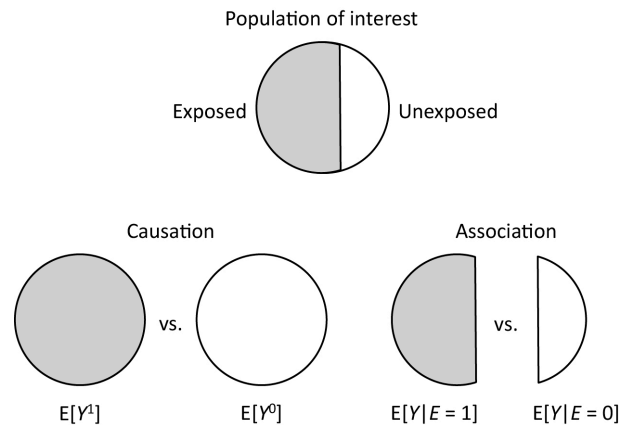


Figure 2. Causation and association in the population of interest. Causation is defined by a contrast of risks in the entire population under two potential exposure values, whereas association is defined by a contrast of risks in subsets of the population determined by the subjects' actual exposure value. Adapted from Hernán and Robins⁽¹⁹⁾ with permission from the authors.

$$r_2)/(r_1 + r_3) = 0.414/0.324 \approx 1.278.$$

Once information about these target parameters, or *causal estimands*, is available for the target population, the concept of confounding—or more precisely, *confounding in measure*—becomes clearer by comparing the causal estimands with the corresponding measures of association^{(27), (28), (29), (30), (31), (32), (33), (34)}. If we use RD as a measure of interest when the target population is the total population, the associational RD (i.e., 0.090) is identical to the causal RD (i.e., 0.090), and we say there is no confounding (in measure for RD). However, when either the exposed or unexposed group is used as the target population, there is confounding (in measure for RD); the associational RD is higher than the causal RD in the exposed group (i.e., 0.050), while it is lower than the causal RD in the unexposed group (i.e., 0.150). This exemplifies that the concept of the target population is important when explaining confounding because this depends on the population selected as the target of inference.

However, confounding depends not only on the target population but also on the notions of *confounding in distribution* and *confounding in measure*^{(27), (28), (29), (30), (31), (32), (33), (34)}. As summarized in **Table 3**^{(19), (35), (36), (37)}, when the total population is used as the target population, no confounding in distribution is stronger than no confounding in measure. Thus, as shown in our hypothetical cohort study, even when confounding in distribution is present, confounding in measure is not always present. Furthermore, while confounding in distribution is scale-independent, confounding in measure is scale-dependent. In this cohort study, there is no confounding in measure for RD, while there is confounding in measure for RR. When either the exposed or unexposed group is used as the target population, distinguishing between the two notions is subtle because the necessary and sufficient conditions for no confounding become equivalent. Moreover, to clarify a fur-

ther distinction between the notions of *confounding in expectation* and *realized confounding*, it is useful to examine the distribution of exposure status in the target population⁽²⁷⁾. To grasp the explicit distinction between these two notions, we need to understand the “mechanism” that generates exposure events, rather than the “product of that mechanism”^{(27), (31), (38)}. Despite its importance, the differing notions of confounding have not been fully appreciated in the literature, leading to confusion in epidemiology regarding causal concepts. See Suzuki et al.⁽²⁷⁾ for further discussion.

Through the Lens of Sufficient Causes

Mechanisms are particularly important in understanding causal processes in biology⁽³⁹⁾. Although the meaning of mechanisms may substantially vary across disciplines, careful scrutiny of the concept of the mechanism provides deeper insight into why uncertainty or randomness is so prevalent in biomedical and social sciences⁽⁴⁰⁾. The sufficient cause model was introduced in the field of epidemiology by Rothman⁽³⁾, and it has shaped our understanding of causation in biomedical science. Similar models arise in philosophy⁽⁴¹⁾, law⁽⁴²⁾, and psychology⁽⁴³⁾.

The sufficient cause model considers sets of actions, events, or states of nature which together bring about the outcome under consideration. In contrast to the counterfactual model, the sufficient cause model gives an account of the causes of a particular effect, addressing the question, “Given a particular effect, what are the various events which might have been its cause?”⁽⁹⁾ Note that the sufficient cause framework and the counterfactual framework address different questions. The counterfactual framework addresses the question “what if?” By contrast, the sufficient cause framework addresses the question “why does it happen?,” which is also a frequently oc-

Target population; Notion of confounding	Necessary and sufficient conditions for no confounding ^a	True/false in Table 2
Exposed group		
No confounding in distribution ^b	$(p_1 + p_3) = (q_1 + q_3)$	False
No confounding in measure for RD	$(p_1 + p_2) - (p_1 + p_3) = (p_1 + p_2) - (q_1 + q_3)$ $\Leftrightarrow (p_1 + p_3) = (q_1 + q_3)$	False
No confounding in measure for RR	$(p_1 + p_2)/(p_1 + p_3) = (p_1 + p_2)/(q_1 + q_3)$ $\Leftrightarrow (p_1 + p_3) = (q_1 + q_3)$	False
Unexposed group		
No confounding in distribution ^c	$(p_1 + p_2) = (q_1 + q_2)$	False
No confounding in measure for RD	$(q_1 + q_2) - (q_1 + q_3) = (p_1 + p_2) - (q_1 + q_3)$ $\Leftrightarrow (p_1 + p_2) = (q_1 + q_2)$	False
No confounding in measure for RR	$(q_1 + q_2)/(q_1 + q_3) = (p_1 + p_2)/(q_1 + q_3)$ $\Leftrightarrow (p_1 + p_2) = (q_1 + q_2)$	False
Total population		
No confounding in distribution ^d	$\{(r_1 + r_2) = (p_1 + p_2)\} \wedge \{(r_1 + r_3) = (q_1 + q_3)\}$ $\Leftrightarrow \{(p_1 + p_2) = (q_1 + q_2)\} \wedge \{(p_1 + p_3) = (q_1 + q_3)\}$	False
No confounding in measure for RD	$(r_1 + r_2) - (r_1 + r_3) = (p_1 + p_2) - (q_1 + q_3)$ $\Leftrightarrow (p_1 + p_3) \times \Pr(E = 1) + (p_1 + p_2) \times \Pr(E = 0)$ $= (q_1 + q_3) \times \Pr(E = 1) + (q_1 + q_2) \times \Pr(E = 0)$	True
No confounding in measure for RR	$(r_1 + r_2)/(r_1 + r_3) = (p_1 + p_2)/(q_1 + q_3)$ $\Leftrightarrow (p_1 + p_2)\{(p_1 + p_3) - (q_1 + q_3)\} \times \Pr(E = 1)$ $= (q_1 + q_3)\{(q_1 + q_2) - (p_1 + p_2)\} \times \Pr(E = 0)$	False

Abbreviations: RD: risk difference; RR: risk ratio.

^aSee Table 2 for notations.

^bA necessary and sufficient condition for no confounding in distribution is equivalent to $Y^0 \perp\!\!\!\perp E$, which is termed a “partial exchangeability under no treatment” condition in the counterfactual model^{(19),(35),(36),(37)}. When the exposed group is used as the target population, distinguishing between the two notions is a subtle issue because the necessary and sufficient conditions for no confounding become equivalent.

^cA necessary and sufficient condition for no confounding in distribution is equivalent to $Y^1 \perp\!\!\!\perp E$, which is termed a “partial exchangeability under treatment” condition in the counterfactual model^{(19),(35),(36),(37)}. When the unexposed group is used as the target population, distinguishing between the two notions is a subtle issue because the necessary and sufficient conditions for no confounding become equivalent.

^dA necessary and sufficient condition for no confounding in distribution is equivalent to $Y^e \perp\!\!\!\perp E$ ($e = 0, 1$), which is termed an “(marginal) exchangeability” condition in the counterfactual model^{(19),(35)}. When the total population is used as the target population, no confounding in distribution is stronger than no confounding in measure.

Table 3. Necessary and Sufficient Conditions for No Confounding.

curing question about causality.

In the sufficient cause model, causation is conceptualized as a collection of different causal mechanisms, each sufficient to produce the outcome. These causal mechanisms are called “sufficient causes,” with each consisting of a minimal set of conditions or “component causes.” Whenever all the component causes of a particular causal mechanism are present, the mechanism operates and the outcome inevitably occurs. Within the philosophical literature, the sufficient cause framework is most closely associated with the work of Mackie⁽⁴¹⁾, who proposed that when we refer to something as a “cause,” it is then generally known to be an “*insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for the result.” From the initial letters of the words italicized above, the term “INUS condition” is used as shorthand for this statement. Following Rothman⁽³⁾, I refer to each component of a sufficient cause as a “component cause” or simply a “cause,” in

alignment with the concept of an INUS condition. Drawing on Mackie’s terminology⁽⁴¹⁾, the causal mechanism of interest can be described as “a condition which is itself unnecessary but sufficient for the result.”

Within the sufficient cause framework, we usually include one component cause to represent unspecified events, conditions, and characteristics that must be present or must have occurred at the instance of the outcome of interest. The necessity of the unknown or unmeasured component cause underscores the pervasive uncertainty observed in the biomedical and social sciences⁽⁴⁰⁾. Then, how many types of sufficient causes can we enumerate in the hypothetical cohort study? In the sufficient cause framework, each sufficient cause for the outcome Y might require the presence of E , the presence of $\bar{E}$, or may not require either, where I let $\bar{E}$ denote the complement of E in the terminology of events. We could thus enumerate three different types of sufficient causes for Y along

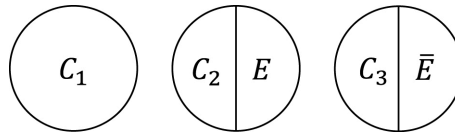


Figure 3. Three types of sufficient causes for Y . I consider a binary exposure E and a binary outcome Y .

Risk status type	Background factors			Exposed group ($E = 1$)		Unexposed group ($E = 0$)		Total population	
	C_1	C_2	C_3	N	Proportion	N	Proportion	N	Proportion ^b
1	1	1	1	30	$s_1 = 0.050$	28	$t_1 = 0.070$	58	$u_1 = 0.058$
2	1	1	0	18	$s_2 = 0.030$	20	$t_2 = 0.050$	38	$u_2 = 0.038$
3	1	0	1	12	$s_3 = 0.020$	16	$t_3 = 0.040$	28	$u_3 = 0.028$
4	1	0	0	18	$s_4 = 0.030$	24	$t_4 = 0.060$	42	$u_4 = 0.042$
5	0	1	1	6	$s_5 = 0.010$	12	$t_5 = 0.030$	18	$u_5 = 0.018$
6	0	1	0	150	$s_6 = 0.250$	80	$t_6 = 0.200$	230	$u_6 = 0.230$
7	0	0	1	120	$s_7 = 0.200$	20	$t_7 = 0.050$	140	$u_7 = 0.140$
8	0	0	0	246	$s_8 = 0.410$	200	$t_8 = 0.500$	446	$u_8 = 0.446$
Total				600	1.000	400	1.000	1000	1.000

^aI consider a binary exposure E ($1 = \text{exposed}$, $0 = \text{unexposed}$) and a binary outcome Y ($1 = \text{outcome occurred}$, $0 = \text{outcome did not occur}$). I consider three different types of sufficient causes for outcome Y along with certain binary background factors as follows: C_1 , C_2E , and $C_3\bar{E}$, where I let $\bar{E}$ denote the complement of E .

^bNote that u_j can be calculated as: $s_j \times \Pr(E = 1) + t_j \times \Pr(E = 0)$, where $\Pr(E = e)$ represents the prevalence of $E = e$ in the total population.

Table 4. Underlying Hypothetical Data of the Hypothetical Cohort Study in Terms of Risk Status Types^a.

with certain background factors C_k : C_1 , C_2E , and $C_3\bar{E}$ (**Figure 3**). Here, C_k denotes a set of all components or factors, other than the presence of E and $\bar{E}$, that may be required for a particular mechanism to operate. Thus, these background factors may comprise several combinations of variables, each of which is part of the sufficient causes. For simplicity, I denote the presence of these background factors as $C_k = 1$ and their absence as $C_k = 0$. An individual is at risk of, or susceptible to, sufficient cause k if C_k is present for that person.

What is the underlying data structure of the hypothetical cohort study through the lens of sufficient causes? Importantly, an individual may be at risk of none, one, or several sufficient causes. Therefore, based on the patterns of joint background factors (C_1 , C_2 , C_3), we can enumerate eight ($= 2^3$) risk status types as shown in **Table 4**⁽⁴⁴⁾. For example, individuals of risk status type 1 are at risk of sufficient causes 1, 2, and 3 (i.e., $(C_1, C_2, C_3) = (1, 1, 1)$). **Table 4** also shows underlying hypothetical data in terms of risk status types. I let s_j , t_j , and u_j , $j = 1-8$, be the proportions of risk status type j in the exposed group, the unexposed group, and the total population, respectively. Like the response type, the risk status type of each individual is unobservable.

Importantly, the sufficient cause model makes clear that causation is a multifactorial phenomenon. It is the combina-

tion of various conditions that leads to the health or social outcomes that we seek to study. Therefore, in most population health settings, it is not meaningful to try to identify “the” cause; rather we examine “causes”⁽⁴⁵⁾. Furthermore, when many causes or conditions are required for a particular causal mechanism to be operative, elimination of any of them suffices, by definition, to render that sufficient cause inoperative. If multiple conditions are required for a sufficient cause, then they effectively interact within the sufficient cause framework; each needs the other to activate a particular mechanism. We can thus sometimes study and identify the multiple causes that, if eliminated, may be sufficient to substantially reduce the occurrence of a disease. This point is closely related to what I will discuss in the next section.

The Link between the Counterfactual Model and the Sufficient Cause Model

As highlighted above, the counterfactual model gives an account of the various effects or outcomes of a particular cause or intervention, whereas the sufficient cause model gives an account of the causes of a particular effect (**Figure 4**). Indeed, they are distinct models addressing different causal questions. Importantly, however, these two models are closely related

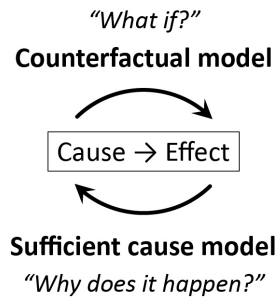


Figure 4. Logical flows of the counterfactual model and the sufficient cause model.

Response types						Risk status types						
Type	Potential outcomes		Proportion of types in ^b			Type	Background factors			Proportion of types in ^b		
	γ^1	γ^0	Exposed	Unexposed	Total population		C_1	C_2	C_3	Exposed	Unexposed	Total population
1	1	1	p_1	q_1	r_1	1 ^d	1	1	1	s_1	t_1	u_1
						2	1	1	0	s_2	t_2	u_2
						3 ^d	1	0	1	s_3	t_3	u_3
						4	1	0	0	s_4	t_4	u_4
						5 ^d	0	1	1	s_5	t_5	u_5
2	1	0	p_2	q_2	r_2	6	0	1	0	s_6	t_6	u_6
3 ^c	0	1	p_3	q_3	r_3	7 ^d	0	0	1	s_7	t_7	u_7
4	0	0	p_4	q_4	r_4	8	0	0	0	s_8	t_8	u_8

^aI consider a binary exposure E (1 = exposed, 0 = unexposed) and a binary outcome Y (1 = outcome occurred, 0 = outcome did not occur). I consider two potential outcomes, Y^e , for an individual. I consider three different types of sufficient causes for outcome Y along with certain binary background factors as follows: C_1 , C_2E , and $C_3\bar{E}$, where I let $\bar{E}$ denote the complement of E .

^bNote that r_j can be calculated as: $p_j \times \Pr(E = 1) + q_j \times \Pr(E = 0)$, where $\Pr(E = e)$ represents the prevalence of $E = e$ in the total population. Similarly, u_j can be calculated as: $s_j \times \Pr(E = 1) + t_j \times \Pr(E = 0)$.

^cUnder the assumption of counterfactual positive monotonicity (i.e., $Y^0 \leq Y^1$ for all individuals), this response type is excluded^{(11),(47)}.

^dUnder the assumption of sufficient cause positive monotonicity⁽⁴⁸⁾, or no preventive action⁽⁴⁷⁾ (i.e., $C_3 = 0$ for all individuals), these risk status types are excluded^{(11),(44)}. Note that sufficient cause positive monotonicity implies counterfactual positive monotonicity.

Table 5. Correspondence between Response Types and Risk Status Types under a Binary Exposure and a Binary Outcome^a.

and used to elucidate the same underlying cause-effect relationships⁽⁴⁶⁾, and the link between them has been addressed^{(7), (8), (9), (10), (11), (12)}. In **Table 5**^{(11), (44), (47), (48)}, I show a correspondence between the four response types and the eight risk status types, combining the underlying data structure of the hypothetical cohort study through the lens of both counterfactuals and sufficient causes⁽⁴⁴⁾. Linking sufficient causes to the counterfactual model is challenging because of its non-bijective (or more precisely, surjective but not injective) correspondence. Indeed, the sufficient cause model is a “finer” model than the counterfactual model. That is, an individual is of one and only one response type in the counterfactual framework, whereas an individual may be at risk of none, one, or several sufficient causes. This is a fundamental issue when

considering the link between these two models. Note that the potential outcomes of Y can be described using the background factors C_k as: $Y^1 = \max\{C_1, C_2\}$ and $Y^0 = \max\{C_1, C_3\}$. For example, $\max\{C_1, C_2\}$ is a binary random variable with a value of 1 unless both $C_1 = 0$ and $C_2 = 0$.

Understanding the link between the two causal models can provide greater insight and facilitate the use of each model in appropriate contexts, highlighting their respective strengths⁽⁴⁹⁾. Here, I briefly present three topics of interest from our research. First, the notion of covariate balance was formalized and related to the concept of confounding^{(35), (36)}. We proposed to use each of the background factors in sufficient causes to represent a set of covariates of interest and to examine the presence of covariate balance by comparing joint

Target population	Counterfactual model	Sufficient cause model	
	Exchangeability ^b	Exchangeability in terms of background factors	Covariate balance ^c
Exposed group	$Y^0 \perp\!\!\!\perp E$	$\Leftrightarrow \max\{C_1, C_3\} \perp\!\!\!\perp E$	$\Leftrightarrow (C_1, C_3) \perp\!\!\!\perp E$
Unexposed group	$Y^1 \perp\!\!\!\perp E$	$\Leftrightarrow \max\{C_1, C_2\} \perp\!\!\!\perp E$	$\Leftrightarrow (C_1, C_2) \perp\!\!\!\perp E$
Total population	$Y^e \perp\!\!\!\perp E (e = 0,1)$	$\Leftrightarrow \max\{C_1, C_k\} \perp\!\!\!\perp E (k = 2,3)$	$\Leftrightarrow (C_1, C_k) \perp\!\!\!\perp E (k = 2,3)$
	$\uparrow$ $(Y^0, Y^1) \perp\!\!\!\perp E$	$\uparrow$ $\Leftrightarrow (\max\{C_1, C_3\}, \max\{C_1, C_2\}) \perp\!\!\!\perp E$	$\uparrow$ $\Leftrightarrow (C_1, C_2, C_3) \perp\!\!\!\perp E$

^aSee Table 5 for notations. In this table, I primarily use the notion of confounding in distribution. For details, see Suzuki et al.⁽³⁵⁾

^bAs explained in the footnotes of Table 3, $Y^0 \perp\!\!\!\perp E$ and $Y^1 \perp\!\!\!\perp E$ are termed partial exchangeability, whereas $Y^e \perp\!\!\!\perp E (e = 0,1)$ is termed (marginal) exchangeability. Note that $(Y^0, Y^1) \perp\!\!\!\perp E$ is referred to as full exchangeability⁽¹⁹⁾, which is equivalent to complete comparability of response types between the exposed and the unexposed groups (i.e., $(p_1, p_2, p_3, p_4) = (q_1, q_2, q_3, q_4)$).

^cSuzuki et al.⁽³⁵⁾ referred to $(C_1, C_3) \perp\!\!\!\perp E$ and $(C_1, C_2) \perp\!\!\!\perp E$ as partial covariate balance. In addition, Suzuki et al.⁽³⁵⁾ referred to $(C_1, C_2, C_3) \perp\!\!\!\perp E$ as full covariate balance, which is equivalent to complete comparability of risk status types between the exposed and the unexposed groups (i.e., $(s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8) = (t_1, t_2, t_3, t_4, t_5, t_6, t_7, t_8)$).

Table 6. The Relationship between Exchangeability and Covariate Balance for No Confounding^a.

distributions of the relevant background factors between the exposed and the unexposed groups⁽³⁵⁾. Note that a set of covariates can be generally divided into covariates that have no causal coaction with an exposure (i.e., C_1) and those that have causal coaction with an exposure (i.e., C_2 and C_3). Given that sufficient cause 2 (i.e., C_2E) contains exposure as a component, it can never complete when the individual is unexposed. Therefore, when the exposed group is the target population, we do not have to consider the comparability of C_2 between the exposed and the unexposed groups. In other words, C_2 represents an irrelevant set of covariates for no confounding, and we need to consider the comparability of only C_1 and C_3 between the exposed and the unexposed groups. Conversely, when the unexposed group is the target population, C_3 represents an irrelevant set of covariates for no confounding because sufficient cause 3 (i.e., $C_3\bar{E}$) can never complete when the individual is exposed. Thus, we need to consider the comparability of joint distributions of C_1 and C_2 between the exposed and the unexposed groups. Consequently, irrespective of the target population, covariate balance is a sufficient, but not a necessary, condition for no confounding (**Table 6**)^{(19), (35)}. In other words, confounding implies the presence of covariate imbalance, but not vice versa.

Second, the link between the two models also provides important distinctions between attributable fractions and etiologic fractions^{(44), (48), (50)}. Since the work published by Doll in 1951⁽⁵¹⁾, the attributable fraction has been widely used to assess the potential impact of health interventions in epidemiology. For example, the attributable fraction (population) is generally used when we are interested in the reduction in incidence that would be achieved if the population had been entirely unexposed compared with its current or observed exposure pattern^{(48), (52), (53), (54)}. This is defined in the counterfactual framework as^{(19), (44), (55)}:

$$\begin{aligned} & \frac{\Pr(Y=1) - \Pr(Y^0=1)}{\Pr(Y=1)} \\ &= \frac{\sum_e \Pr(Y=1|E=e)\Pr(E=e) - \sum_e \Pr(Y^0=1|E=e)\Pr(E=e)}{\sum_e \Pr(Y=1|E=e)\Pr(E=e)} \\ &= \frac{\{(p_1 + p_2)\Pr(E=1) + (q_1 + q_3)\Pr(E=0)\} - \{(p_1 + p_3)\Pr(E=1) + (q_1 + q_3)\Pr(E=0)\}}{(p_1 + p_2)\Pr(E=1) + (q_1 + q_3)\Pr(E=0)} \\ &= \frac{(p_2 - p_3)\Pr(E=1)}{(p_1 + p_2)\Pr(E=1) + (q_1 + q_3)\Pr(E=0)}, \end{aligned}$$

which is approximately calculated as 0.085 in the hypothetical cohort study. Note that this corresponds to the attributable caseload (population) proposed by Suzuki et al.⁽⁴⁴⁾, who distinguished between the attributable caseload and attributable proportion, both of which compose the broader concept of attributable fraction. The attributable proportion (population) is defined as $\{\Pr(Y=1) - \Pr(Y^0=1, Y=1)\}/\Pr(Y=1)$, which ranges from 0 to 1⁽⁴⁴⁾, and is approximately calculated as 0.424 in the hypothetical cohort study. By contrast, the etiologic fraction has been broadly defined as the fraction of cases that were “caused” by exposure^{(50), (52), (53), (56), (57), (58), (59), (60)}. By taking into account the potential completion time of each sufficient cause, we showed that, when considering a binary exposure and a binary outcome, individuals can be classified into 24 (= 4!) sequence types as shown in **Table 7**^{(11), (44), (47), (48)}, which helps to clarify the two types of etiologic fraction: the accelerating etiologic proportion and the total etiologic proportion⁽⁴⁴⁾. The former refers to the proportion of the diseased for whom the exposure “sped up” the time at which the outcome occurred, whereas the latter refers to the proportion of the diseased for whom the exposure is the “actual cause of the outcome.” The accelerating etiologic proportion (population) and the total etiologic proportion (population) are defined as⁽⁴⁴⁾:

Response types						Risk status types						Sequence types							
Type	Potential outcomes		Proportion of types in ^b			Type	Background factors			Proportion of types in ^b			Type	Sequence of potential completion time	Potential outcome occurrence time		Proportion of types in ^b		
	y^1	y^0	Exposed	Unexposed	Total population		C_1	C_2	C_3	Exposed	Unexposed	Total population			d_1	d_0	Exposed	Unexposed	Total population
1	1	1	p_1	q_1	r_1	1 ^d	1	1	1	s_1	t_1	u_1	1	$d_e < d_e < d_e \leq h$	d_e	d_e	v_1	w_1	x_1
													2	$d_e < d_e < d_e \leq h$	d_e	d_e	v_2	w_2	x_2
													3	$d_e < d_e < d_e \leq h$	d_e	d_e	v_3	w_3	x_3
													4	$d_e < d_e < d_e \leq h$	d_e	d_e	v_4	w_4	x_4
													5 ^e	$d_e < d_e < d_e \leq h$	d_e	d_e	v_5	w_5	x_5
													6 ^e	$d_e < d_e < d_e \leq h$	d_e	d_e	v_6	w_6	x_6
						2	1	1	0	s_2	t_2	u_2	7	$d_e < d_e \leq h < d_e$	d_e	d_e	v_7	w_7	x_7
													8	$d_e < d_e \leq h < d_e$	d_e	d_e	v_8	w_8	x_8
						3 ^d	1	0	1	s_3	t_3	u_3	9	$d_e < d_e \leq h < d_e$	d_e	d_e	v_9	w_9	x_9
													10 ^e	$d_e < d_e \leq h < d_e$	d_e	d_e	v_{10}	w_{10}	x_{10}
						4	1	0	0	s_4	t_4	u_4	11	$d_e \leq h < d_e < d_e$	d_e	d_e	v_{11}	w_{11}	x_{11}
													12	$d_e \leq h < d_e < d_e$	d_e	d_e	v_{12}	w_{12}	x_{12}
						5 ^d	0	1	1	s_5	t_5	u_5	13	$d_e < d_e \leq h < d_e$	d_e	d_e	v_{13}	w_{13}	x_{13}
													14 ^e	$d_e < d_e \leq h < d_e$	d_e	d_e	v_{14}	w_{14}	x_{14}
2	1	0	p_2	q_2	r_2	6	0	1	0	s_6	t_6	u_6	15	$d_e \leq h < d_e < d_e$	d_e	d_e	v_{15}	w_{15}	x_{15}
													16	$d_e \leq h < d_e < d_e$	d_e	d_e	v_{16}	w_{16}	x_{16}
3 ^e	0	1	p_3	q_3	r_3	7 ^d	0	0	1	s_7	t_7	u_7	17 ^e	$d_e \leq h < d_e < d_e$	d_e	d_e	v_{17}	w_{17}	x_{17}
													18 ^e	$d_e \leq h < d_e < d_e$	d_e	d_e	v_{18}	w_{18}	x_{18}
4	0	0	p_4	q_4	r_4	8	0	0	0	s_8	t_8	u_8	19	$h < d_e < d_e < d_e$	d_e	d_e	v_{19}	w_{19}	x_{19}
													20	$h < d_e < d_e < d_e$	d_e	d_e	v_{20}	w_{20}	x_{20}
													21	$h < d_e < d_e < d_e$	d_e	d_e	v_{21}	w_{21}	x_{21}
													22	$h < d_e < d_e < d_e$	d_e	d_e	v_{22}	w_{22}	x_{22}
													23 ^e	$h < d_e < d_e < d_e$	d_e	d_e	v_{23}	w_{23}	x_{23}
													24 ^e	$h < d_e < d_e < d_e$	d_e	d_e	v_{24}	w_{24}	x_{24}

^aI consider a binary exposure E ($1 = \text{exposed}$, $0 = \text{unexposed}$) and a binary outcome Y ($1 = \text{outcome occurred}$, $0 = \text{outcome did not occur}$). I consider two potential outcomes, Y^e , for an individual. I consider three different types of sufficient causes for outcome Y along with certain binary background factors as follows: C_1 , C_2E , and C_3E , where I let $\bar{E}$ denote the complement of E . I let d_e , d_e , and d_e denote the potential completion times of sufficient causes C_1 , C_2E , and C_3E at which the outcome would occur in an individual, respectively. I also let d_1 and d_0 denote the potential outcome occurrence time of an individual when exposed and unexposed, respectively. In other words, I denote $d_1 = \min\{d_e, d_e\}$ and $d_0 = \min\{d_e, d_e\}$. Furthermore, h denotes a maximum follow-up time of an individual. The structure of this table is slightly modified from Table 1 of Suzuki et al. (44).

^bNote that r_j can be calculated as: $p_j \times \Pr(E = 1) + q_j \times \Pr(E = 0)$, where $\Pr(E = e)$ represents the prevalence of $E = e$ in the total population. Similarly, u_j can be calculated as: $s_j \times \Pr(E = 1) + t_j \times \Pr(E = 0)$ and x_j can be calculated as: $v_j \times \Pr(E = 1) + w_j \times \Pr(E = 0)$.

^cUnder the assumption of counterfactual positive monotonicity (i.e., $Y^0 \leq Y^1$ for all individuals), this response type is excluded^{(41),(47)}.

^dUnder the assumption of sufficient cause positive monotonicity⁽⁴⁸⁾, or no preventive action⁽⁴⁷⁾ (i.e., $C_3 = 0$ for all individuals), these risk status types are excluded^{(11),(44)}.

^eUnder the assumption of no preventive sequence (i.e., $d_1 \leq d_0$ for all individuals), these sequence types are excluded. Note that no preventive sequence is neither a necessary nor a sufficient condition for sufficient cause positive monotonicity, and vice versa. However, both no preventive sequence and sufficient cause positive monotonicity imply counterfactual positive monotonicity.

Table 7. Correspondence between Response Types, Risk Status Types, and Sequence Types under a Binary Exposure and a Binary Outcome^a.

$$\begin{aligned}
 & \frac{\Pr(\{d_e \leq t\}, \{d_e < d_0\} | E = 1) \Pr(E = 1)}{\Pr(Y = 1)} \\
 &= \frac{\Pr(\{d_e \leq t\}, \{d_e < d_\phi\}, \{d_e < d_e\} | E = 1) \Pr(E = 1)}{\sum_e \Pr(Y = 1 | E = e) \Pr(E = e)} \\
 &= \frac{(v_3 + v_4 + v_8 + v_{13} + v_{15} + v_{16}) \Pr(E = 1)}{(p_1 + p_2) \Pr(E = 1) + (q_1 + q_3) \Pr(E = 0)} \\
 & \text{and} \\
 & \frac{\Pr(\{d_e \leq t\}, \{d_e < d_\phi\} | E = 1) \Pr(E = 1)}{\Pr(Y = 1)} \\
 &= \frac{\Pr(\{d_e \leq t\}, \{d_e < d_\phi\} | E = 1) \Pr(E = 1)}{\sum_e \Pr(Y = 1 | E = e) \Pr(E = e)} \\
 &= \frac{(v_3 + v_4 + v_6 + v_8 + v_{13} + v_{14} + v_{15} + v_{16}) \Pr(E = 1)}{(p_1 + p_2) \Pr(E = 1) + (q_1 + q_3) \Pr(E = 0)},
 \end{aligned}$$

respectively. The latter includes two more sequence types (types 6 and 14) in the numerator, in which d_e is longer than d_e and thus equal to or larger than the former. These measures cannot generally be inferred from epidemiologic data, and a lower bound can be calculated by the attributable case-load (population) if it yields a positive value⁽⁴⁴⁾. Although the differences between the two types of etiologic fraction might be subtle, they are closely related to the definition of causality. Therefore, it is important to clarify which measures are used

on each occasion. Our recent articles provide further discussion of the attributable fraction and related measures^{(55), (61), (62), (63)}.

Third, the sufficient cause model has been extended to and used to elucidate the phenomenon of mediation^{(64), (65), (66), (67)}. The assessment of mediation is one way to explore cause-effect relationships in more depth, providing a stronger test for and explanation of the observed associations^{(68), (69), (70), (71), (72)}. In the context of a binary exposure E , a binary mediator M , and a binary outcome Y (**Figure 5**), I let M^e denote the potential outcomes of M if, possibly contrary to fact, there had been interventions to set E to e . In addition, I let Y^{em} denote the potential outcomes of Y if, possibly contrary to fact, there had been interventions to set E to e and to set M to m . By using the compound potential outcomes (or nested counterfactuals), the total effect of E on Y in the population of interest is given as $E[Y^{LM}] - E[Y^{0LM}]$ on the RD scale, or equivalently $E[Y^1] - E[Y^0]$, by making a composition assumption (i.e., $Y^e = Y^{eM}$)^{(73), (74)}. As has been well appreciated in the causal mediation literature^{(68), (73), (74), (75)}, the total effect can be decomposed into the pure direct effect (PDE) (i.e., $E[Y^{LM}] - E[Y^{0LM}]$) and the total indirect effect (TIE) (i.e., $E[Y^{LM}] -$

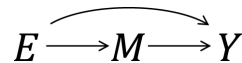


Figure 5. A causal diagram of an exposure E , a mediator M , and an outcome Y .

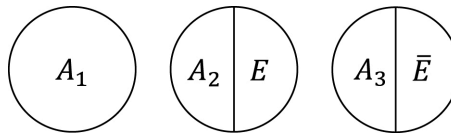


Figure 6. Three types of sufficient causes for M . I consider a binary exposure E and a binary mediator M . Under the assumption of sufficient cause positive monotonicity of E , I consider only A_1 and A_2E .

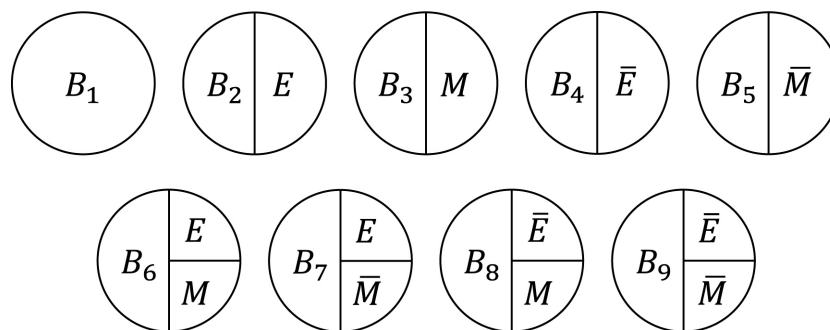


Figure 7. Nine types of sufficient causes for Y . I consider a binary exposure E , a binary mediator M , and a binary outcome Y . Under the assumption of sufficient cause positive monotonicity of E and M , I consider only B_1 , B_2E , B_3M , and B_6EM .

$E[Y^{1M}]$), or alternatively into the total direct effect (TDE) (i.e., $E[Y^{1M}] - E[Y^{0M}]$) and the pure indirect effect (PIE) (i.e., $E[Y^{0M}] - E[Y^{00}]$). Under the sufficient cause positive monotonicity of E and M (**Figure 6** and **7**), we described how we can identify mediation as well as mechanism by showing their correspondence with direct and indirect effects in the counterfactual framework (**Table 8**)⁽⁶⁶⁾. We defined both mediation and mechanism by explicating their presence and operation in the sufficient cause framework. For example, we define that mediating paths (i.e., $E \rightarrow A_2E \rightarrow M \rightarrow B_3M \rightarrow Y$ and $E \rightarrow A_2E \rightarrow M \rightarrow B_6EM \rightarrow Y$ in **Figure 8**) are present in an individual if the individual is at risk of a sufficient cause(s) involving E in the M -stage and simultaneously is at risk of a sufficient cause(s) involving M in the Y -stage. Similarly, we define that mediation operates in an individual if the mediating path from E to Y via M is actually activated to make Y occur. We demonstrated that, although the PIE implies the presence of mediating pathways, it does not necessarily imply their operation because a non- M -mediating path, $E \rightarrow B_2E \rightarrow Y$, may operate to induce Y . However, this is not the case for TIE, and

non-zero TIE implies the operation, not simply the presence, of mediation. Thus, when assessing mediation, our findings give priority to the TIE over the PIE, exerting influence on practice, even indirectly, by shaping the methods that are available and routinely employed. We also demonstrated how a researcher can decompose the total effect into the effect due to mediated paths and the effect due to non-mediated paths in terms of the probabilities of background factors of sufficient causes⁽⁶⁶⁾. These findings contribute to facilitating looking inside the “black box” between exposure and outcome in epidemiologic studies⁽⁷⁶⁾.

Further Remarks

The counterfactual model and the sufficient cause model have become cornerstones for causal thinking in epidemiology^{(2), (3), (4), (5), (6)}. Meanwhile, for decades, attempts have been made in the field of epidemiology to use a set of facts to distinguish causal from noncausal explanations^{(77), (78), (79), (80), (81)}. One well-known approach is the nine viewpoints proposed by Sir

Counterfactual model										Sufficient cause model					
Potential outcomes of M		Potential outcomes of Y^b				Compound potential outcomes ^b				Background factors of sufficient causes for M		Background factors of sufficient causes for Y			
M^e		Y^{em}				$Y^{eM^{e'}}$									
M^1	M^0	Y^{11}	Y^{01}	Y^{10}	Y^{00}	Y^{1M^1}	Y^{1M^0}	Y^{0M^1}	Y^{0M^0}	A_1	A_2	B_1	B_2	B_3	B_6
1	1	1	1	(1)	(1)	1	(1)	(1)	1	1	0/1	1	0/1	0/1	0/1
1	0	1	(1)	(1)	1	1	(1)	(1)	1	0	1	1	0/1	0/1	0/1
0	0	(1)	(1)	1	1	1	(1)	(1)	1	0	0	1	0/1	0/1	0/1
1	1	1	1	(1)	(0)	1	(1)	(1)	1	1	0/1	0	1	1	0/1
1	0	1	(1)	(1)	0	1	(1)	(1)	0	0	1	0	1	1	0/1
0	0	(1)	(1)	1	0	1	(1)	(0)	0	0	0	0	1	1	0/1
1	1	1	1	(0)	(0)	1	(1)	(1)	1	1	0/1	0	0	1	0/1
1	0	1	(1)	(0)	0	1	(0)	(1)	0	0	1	0	0	1	0/1
0	0	(1)	(1)	0	0	0	(0)	(0)	0	0	0	0	0	1	0/1
1	1	1	0	(1)	(0)	1	(1)	(0)	0	1	0/1	0	1	0	0/1
1	0	1	(0)	(1)	0	1	(1)	(0)	0	0	1	0	1	0	0/1
0	0	(1)	(0)	1	0	1	(1)	(0)	0	0	0	0	1	0	0/1
1	1	1	0	(0)	(0)	1	(1)	(0)	0	1	0/1	0	0	0	1
1	0	1	(0)	(0)	0	1	(0)	(0)	0	0	1	0	0	0	1
0	0	(1)	(0)	0	0	0	(0)	(0)	0	0	0	0	0	0	1
1	1	0	0	(0)	(0)	0	(0)	(0)	0	1	0/1	0	0	0	0
1	0	0	(0)	(0)	0	0	(0)	(0)	0	0	1	0	0	0	0
0	0	(0)	(0)	0	0	0	(0)	(0)	0	0	0	0	0	0	0

^aI consider a binary exposure E , a binary mediator M , and a binary outcome Y . I consider two different types of sufficient causes for M along with a certain binary background factors as follows: A_1 and A_2E . Similarly, I consider four different types of sufficient causes for Y as follows: B_1 , B_2E , B_3M , and B_6EM . See Figures 6 and 7. The structure of this table is slightly modified from Table 1 of Suzuki et al.⁽⁶⁶⁾

^bParentheses indicate that this particular outcome will never be observed.

Table 8. Correspondence between Potential Outcomes and Sufficient Causes under the Assumption of Sufficient Cause Positive Monotonicity in the Context of Mediation^a.

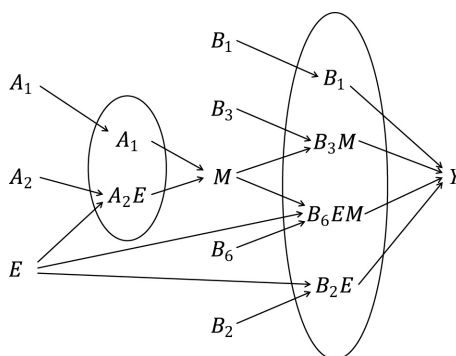


Figure 8. Mediation and mechanism in a causal diagram with a sufficient causation structure. I consider a binary exposure E , a binary mediator M , and a binary outcome Y , under the assumption of sufficient cause positive monotonicity of E and M . See Figures 6 and 7.

Austin Bradford Hill in his President's Address to the Royal Society of Medicine in 1965 (**Table 9**)^{(52), (82), (83)}. Unfortunately,

the list is often erroneously referred to as the "Bradford Hill criteria" or "causal criteria", although he warned that there are

Viewpoints	Descriptions ^a	Problems ^b
1. Strength	The stronger the size of the risk measured by appropriate statistical estimates is, the more likely to be causal.	Strength depends on the prevalence of other causes; it is not a biologic characteristic and can be confounded.
2. Consistency	The results are replicated in studies in different settings using different methods.	Causal relations have exceptions that are understood best with hindsight.
3. Specificity	A putative cause produces a specific effect, as hypothesized or predicted by background theory.	A cause can have many effects.
4. Temporality	Exposure always precedes the outcome.	It may be difficult to establish the temporal sequence between cause and effect.
5. Biological gradient	An increasing level of exposure (in amount and/or time) increases risk.	It can be confounded; threshold phenomena would not show a progressive relation.
6. Plausibility	The association is coherent with firmly established knowledge on pathobiological processes.	Too subjective.
7. Coherence	The association is compatible with existing theory and knowledge.	How does it differ from consistency or plausibility?
8. Experiment	An experiment shows that the caused condition can be altered by changing exposure to the putative cause.	Not always available.
9. Analogy	Similar relations have previously been established.	Analogies abound.

^aSee the entry “Hill’s considerations for causation” in *A Dictionary of Epidemiology*⁽⁵²⁾.

^bSee Table 3-2 in *Epidemiology: An Introduction* by Rothman⁽⁸²⁾.

Table 9. The Nine Bradford Hill Viewpoints.

no “hard-and-fast rules of evidence” for causation ⁽⁸³⁾. Indeed, despite their popularity, the nine Bradford Hill viewpoints do not provide a set of sufficient criteria to distinguish causal from noncausal explanations ⁽⁸⁴⁾, and temporality is a *sine qua non* for causal explanations of observed associations ^{(81), (85)}.

Although the Bradford Hill viewpoints should not be used as a checklist for evaluating whether an observed association can be interpreted as causal, they are still helpful in assessing causality from an inductive standpoint. However, as emphasized in this article, it is also important to use a deductive form of logic in the formal causal models to draw causal conclusions, by clarifying the premises to be evaluated ⁽⁸⁶⁾. Epidemiologic perspectives have been used to shed light on the underlying mechanisms at the genetic, molecular, and cellular levels ^{(87), (88), (89), (90), (91)}. To strengthen our assessment in the face of multifactorial causality across multiple levels and dimensions ^{(85), (92)}, with time-varying exposures ⁽⁹³⁾, it is important to carefully scrutinize observed associations in a complementary manner, using both the counterfactual model and the sufficient cause model, employing both inductive and deductive reasoning. This holistic approach will better help us to unravel causality.

Article Information

This article is based on the study, which received the Medical Research Encouragement Prize of The Japan Medical Association in 2023.

Conflicts of Interest

None

Sources of Funding

ES is supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP19KK0418 and JP23K09740).

Acknowledgement

The author thanks Eiji Yamamoto for helpful discussions. The author also thanks Edanz Group (<https://jp.edanz.com/ac>) for editing a draft of this manuscript.

Author Contributions

ES conceptualized the study and wrote the manuscript.

Approval by Institutional Review Board (IRB)

Not applicable.

References

1. VanderWeele TJ, Rothman KJ. Formal causal models. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. *Modern epidemiology*. 4th ed. Philadelphia (PA): Wolters Kluwer; 2021. p. 33-51.
2. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66(5):688-701.
3. Rothman KJ. Causes. *Am J Epidemiol*. 1976;104(6):587-92.
4. Greenland S, Robins JM. Identifiability, exchangeability, and

- epidemiological confounding. *Int J Epidemiol*. 1986;15(3):413-9.
5. Kaufman JS, Poole C. Looking back on “causal thinking in the health sciences”. *Annu Rev Public Health*. 2000;21:101-19.
 6. Little RJ, Rubin DB. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annu Rev Public Health*. 2000;21:121-45.
 7. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scand J Work Environ Health*. 1988;14(2):125-9.
 8. Flanders WD. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *Eur J Epidemiol*. 2006;21(12):847-53.
 9. VanderWeele TJ, Hernán MA. From counterfactuals to sufficient component causes and vice versa. *Eur J Epidemiol*. 2006;21(12):855-8.
 10. VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*. 2008;95(1):49-61.
 11. Suzuki E, Yamamoto E, Tsuda T. On the link between sufficient-cause model and potential-outcome model. *Epidemiology*. 2011;22(1):131-2.
 12. VanderWeele TJ, Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. *Ann Stat*. 2012;40(4):2128-61.
 13. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10(1):37-48.
 14. Suzuki E, Komatsu H, Yorifuji T, et al. Causal inference in medicine part II. Directed acyclic graphs - a useful method for confounder selection, categorization of potential biases, and hypothesis specification. *Nihon Eiseigaku Zasshi*. 2009;64(4):796-805.
 15. Suzuki E, Komatsu H, Yorifuji T, et al. Causal interpretation based on DAGs. *Epidemiology*. 2008;19(2):361;author reply 361-2 [author reply].
 16. Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams: pitfalls and tips. *J Epidemiol*. 2020;30(4):153-62.
 17. Suzuki E, Tsuda T, Mitsuhashi T, et al. Errors in causal inference: an organizational schema for systematic error and random error. *Ann Epidemiol*. 2016;26(11):788-93.e1.
 18. Hume D, Selby-Bigge LA, Nidditch PH. *Enquiries concerning human understanding and concerning the principles of morals*. 3rd ed. Oxford: Clarendon Press; 1975.
 19. Hernán MA, Robins JM. *Causal inference: what if*. Boca Raton (FL): Chapman & Hall/CRC; 2020.
 20. Suzuki E, Komatsu H, Yorifuji T, et al. Causal inference in medicine part I. Counterfactual models - an approach to clarifying discussions in research and applied public health. *Nihon Eiseigaku Zasshi*. 2009;64(4):786-95.
 21. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20(1):3-5.
 22. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20(6):880-3.
 23. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396):945-60.
 24. Suzuki E, Mitsuhashi T, Tsuda T, et al. A counterfactual approach to bias and effect modification in terms of response types. *BMC Med Res Methodol*. 2013;13:101.
 25. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health*. 2001;22:189-212.
 26. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31(2):422-9.
 27. Suzuki E, Mitsuhashi T, Tsuda T, et al. A typology of four notions of confounding in epidemiology. *J Epidemiol*. 2017;27(2):49-55.
 28. Suzuki E. Generalized causal measure: the beauty lies in its generality. *Epidemiology*. 2015;26(4):490-5.
 29. Suzuki E, Yamamoto M, Yamamoto E. A general explanation of the counterfactual definition of confounding. *J Clin Epidemiol*. 2022;148:189-92.
 30. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci*. 1999;14(1):29-46.
 31. VanderWeele TJ. Confounding and effect modification: distribution and measure. *Epidemiol Methods*. 2012;1(1):55-82.
 32. Suzuki E, Yamamoto E. Further refinements to the organizational schema for causal effects. *Epidemiology*. 2014;25(4):618-9.
 33. Suzuki E, Mitsuhashi T, Tsuda T, et al. A simple example as a pedagogical device? *Ann Epidemiol*. 2014;24(7):560-1.
 34. Suzuki E, Yamamoto M, Yamamoto E. Exchangeability of measures of association before and after exposure status is flipped: its relationship with confounding in the counterfactual model. *J Epidemiol*. 2023;33(8):385-9.
 35. Suzuki E, Tsuda T, Yamamoto E. Covariate balance for no confounding in the sufficient-cause model. *Ann Epidemiol*. 2018;28(1):48-53.e2.
 36. Suzuki E, Yamamoto E. Re: a graphical description of partial exchangeability. *Epidemiology*. 2021;32(2):e7-9.
 37. Sarvet AL, Wanis KN, Stensrud MJ, et al. A graphical description of partial exchangeability. *Epidemiology*. 2020;31(3):365-8.
 38. Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. *Epidemiol Perspect Innov*. 2009;6:4.
 39. Suzuki E, Yamamoto E. *Strength* in causality: discerning causal mechanisms in the sufficient cause model. *Eur J Epidemiol*. 2021;36(9):899-908.
 40. Suzuki E, VanderWeele TJ. Mechanisms and uncertainty in randomized controlled trials: a commentary on Deaton and Cartwright. *Soc Sci Med*. 2018;210:83-5.
 41. Mackie JL. Causes and conditions. *Am Philos Q*. 1965;2(4):245-64.
 42. Wright RW. Causation, responsibility, risk, probability, naked statistics, and proof: pruning the bramble bush by clarifying the concepts. *Iowa Law Rev*. 1988;73(5):1001-77.
 43. Novick LR, Cheng PW. Assessing interactive causal influence.

- Psychol Rev. 2004;111(2):455-85.
44. Suzuki E, Yamamoto E, Tsuda T. On the relations between excess fraction, attributable fraction, and etiologic fraction. *Am J Epidemiol*. 2012;175(6):567-75.
 45. VanderWeele TJ. Invited commentary: the continuing need for the sufficient cause model today. *Am J Epidemiol*. 2017;185(11):1041-3.
 46. Suzuki E. Toward a further understanding of causality in medicine: a contribution to "seizon and life sciences". *J Seizon Life Sci*. 2016;27(1):97-106.
 47. Greenland S, Lash TL, Rothman KJ. Concepts of interaction. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia (PA): Lippincott Williams & Wilkins; 2008. p. 71-83.
 48. VanderWeele TJ. Attributable fractions for sufficient cause interactions. *Int J Biostat*. 2010;6(2):Article 5.
 49. Suzuki E, Yamamoto E. Marginal sufficient component cause model: an emerging causal model with merits? *Epidemiology*. 2021;32(6):838-45.
 50. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol*. 1988;128(6):1185-97.
 51. Doll R. On the aetiology of cancer of the lung. *Acta Unio Int Contra Cancrum*. 1951;7(1 Spec No.):39-50.
 52. Porta MS, editor. *A dictionary of epidemiology*. 6th ed. New York (NY): Oxford University Press; 2014.
 53. Rothman KJ, VanderWeele TJ, Lash TL. Measures of effect and measures of association. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. *Modern epidemiology*. 4th ed. Philadelphia (PA): Wolters Kluwer; 2021. p. 79-103.
 54. Steenland K, Armstrong B. An overview of methods for calculating the burden of disease due to specific risk factors. *Epidemiology*. 2006;17(5):512-9.
 55. Suzuki E, Yamamoto E. Attributable fraction and related measures: conceptual relations in the counterfactual framework. *J Causal Inference*. 2023;11(1):20210068.
 56. Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. *Stat Med*. 1989;8(7):845-59.
 57. Robins JM, Greenland S. The probability of causation under a stochastic model for individual risk. *Biometrics*. 1989;45(4):1125-38.
 58. Greenland S. Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem. *Am J Public Health*. 1999;89(8):1166-9.
 59. Beyea J, Greenland S. The importance of specifying the underlying biologic model in estimating the probability of causation. *Health Phys*. 1999;76(3):269-74.
 60. Boslaugh S, editor. *Encyclopedia of epidemiology*. Thousand Oaks (CA): Sage Publications; 2008.
 61. Suzuki E, Yamamoto E. Errors in the calculation of the population attributable fraction. *Epidemiology*. 2024;35(4):469-72.
 62. Gonçalves BP, Suzuki E. Preventable fraction in the context of disease progression. *Epidemiology*. 2024;35(6):801-4.
 63. Suzuki E, Yamamoto E. Re: bias in calculation of attributable fractions using relative risks from nonsmokers only. *Epidemiology*. 2024;35(6):e21-2.
 64. Hafeman DM. A sufficient cause based approach to the assessment of mediation. *Eur J Epidemiol*. 2008;23(11):711-21.
 65. VanderWeele TJ. Mediation and mechanism. *Eur J Epidemiol*. 2009;24(5):217-24.
 66. Suzuki E, Yamamoto E, Tsuda T. Identification of operating mediation and mechanism in the sufficient-component cause framework. *Eur J Epidemiol*. 2011;26(5):347-57.
 67. VanderWeele TJ. Subtleties of explanatory language: what is meant by "mediation"? *Eur J Epidemiol*. 2011;26(5):343-6.
 68. VanderWeele TJ. *Explanation in causal inference: methods for mediation and interaction*. New York (NY): Oxford University Press; 2015.
 69. Chiba Y, Suzuki E. Causal inference with intermediates: simple methods for principal strata effects and natural direct effects. In: Rodriguez-Morales AJ, editor. *Current topics in public health*. London: InTech; 2013. p. 37-60.
 70. Suzuki E, Mitsuhashi T, Tsuda T, et al. Alternative definitions of "proportion eliminated". *Epidemiology*. 2014;25(2):308-9.
 71. Suzuki E, Evans D, Chaix B, et al. On the "proportion eliminated" for risk differences versus excess relative risks. *Epidemiology*. 2014;25(2):309-10.
 72. Shrier I, Suzuki E. The primary importance of the research question: implications for understanding natural versus controlled direct effects. *Int J Epidemiol*. 2022;51(4):1041-6.
 73. Pearl J. *Causality: models, reasoning, and inference*. 2nd ed. New York (NY): Cambridge University Press; 2009.
 74. VanderWeele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. *Stat Interface*. 2009;2(4):457-68.
 75. Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3(2):143-55.
 76. Hafeman DM, Schwartz S. Opening the Black Box: a motivation for the assessment of mediation. *Int J Epidemiol*. 2009;38(3):838-45.
 77. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95(suppl 1):S144-50.
 78. Lucas RM, McMichael AJ. Association or causation: evaluating links between "environment and disease". *Bull World Health Organ*. 2005;83(10):792-5.
 79. Morabia A. Hume, Mill, Hill, and the sui generis epidemiologic approach to causal inference. *Am J Epidemiol*. 2013;178(10):1526-32.
 80. Olsen J, Jensen UJ. Causal criteria: time has come for a revision. *Eur J Epidemiol*. 2019;34(6):537-41.
 81. VanderWeele TJ, Lash TL, Rothman KJ. Causal inference and scientific reasoning. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. *Modern epidemiology*. 4th ed. Philadelphia (PA): Wolters Kluwer; 2021. p. 17-31.
 82. Rothman KJ. *Epidemiology: an introduction*. 2nd ed. New York (NY): Oxford University Press; 2012.
 83. Hill AB. *The environment and disease: association or*

- causation? *Proc R Soc Med.* 1965;58(5):295-300.
84. Rothman KJ, Lash TL, Haneuse S, et al. The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ, editors. *Modern epidemiology*. 4th ed. Philadelphia (PA): Wolters Kluwer; 2021. p. 3-15.
 85. Suzuki E. Time changes, so do people. *Soc Sci Med.* 2012;75(3):452-6.
 86. VanderWeele TJ. Hill's causal considerations and the potential outcomes framework. *Obs Stud.* 2020;6(2):47-54.
 87. VanderWeele TJ, Asomaning K, Tchetgen Tchetgen EJ, et al. Genetic variants on 15q25.1, smoking, and lung cancer: an assessment of mediation and interaction. *Am J Epidemiol.* 2012;175(10):1013-20.
 88. VanderWeele TJ. Epistatic interactions. *Stat Appl Genet Mol Biol.* 2010;9(1):Article 1.
 89. VanderWeele TJ. Empirical tests for compositional epistasis. *Nat Rev Genet.* 2010;11(2):166.
 90. VanderWeele TJ, Laird NM. Tests for compositional epistasis under single interaction-parameter models. *Ann Hum Genet.* 2011;75(1):146-56.
 91. Suzuki E, VanderWeele TJ. Compositional epistasis: an epidemiologic perspective. In: Moore JH, Williams SM, editors. *Epistasis: methods and protocols*. New York (NY): Springer; 2015. p. 197-216.
 92. Suzuki E, Yamamoto E, Takao S, et al. Clarifying the use of aggregated exposures in multilevel models: self-included vs. self-excluded measures. *PLoS One.* 2012;7(12):e51717.
 93. Shinozaki T, Suzuki E. Understanding marginal structural models for time-varying exposures: pitfalls and tips. *J Epidemiol.* 2020;30(9):377-89.

JMA Journal is an Open Access journal distributed under the Creative Commons Attribution 4.0 International License. To view the details of this license, please visit (<https://creativecommons.org/licenses/by/4.0/>).