

# ***Strength* in causality: discerning causal mechanisms in the sufficient cause model**

Etsuji Suzuki<sup>a</sup>, Eiji Yamamoto<sup>b</sup>

<sup>a</sup>Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan. ORCID iD: 0000-0002-1290-5793

<sup>b</sup>Okayama University of Science, 1-1 Ridai-cho, Kita-ku, Okayama 700-0005, Japan.

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

**Acknowledgements:** We thank Edanz Group (<https://en-author-services.edanz.com/>) for editing a draft of this manuscript.

## **Abstract**

The assessment of causality is fundamental to epidemiology and biomedical sciences. One well-known approach to distinguishing causal from noncausal explanations is the nine Bradford Hill viewpoints. A recent article in this journal revisited the viewpoints to incorporate developments in causal thinking, suggesting that the sufficient cause model is useful in elucidating the theoretical underpinning of the first of the nine viewpoints—*strength* of association. In this article, we discuss how to discern the causal mechanisms of interest in the sufficient cause model, which pays closer attention to the relationship between the sufficient cause model and the Bradford Hill viewpoints. To this end, we explicate the link between the sufficient cause model and the potential-outcome model, both of which have become the cornerstone of causal thinking in epidemiology and biomedicine. A clearer understanding of the link between the two models provides significant implications for interpretation of the observed risks in the subpopulations defined by exposure and confounder. We also show that the concept of potential completion times of sufficient causes is useful to fully discerning completed sufficient causes, which leads us to pay closer attention to the fourth of the nine Bradford Hill viewpoints—*temporality*. Decades after its introduction, the sufficient cause model may be vaguely understood and thus implicitly used under unreasonably strict assumptions. To *strengthen* our assessment in the face of multifactorial causality, it is significant to carefully scrutinize the observed associations in a complementary manner, using the sufficient cause model as well as its relevant causal models.

**Keywords:** Bradford Hill; causal inference; causal mechanisms; counterfactual; sufficient cause model; potential-outcome model

## Introduction

The assessment of causality is fundamental to epidemiology and biomedical sciences. For decades, attempts have been made in the field of epidemiology to use a set of facts to distinguish causal from noncausal explanations [1-5]. One of these well-known approaches is the nine Bradford Hill viewpoints, proposed by Sir Austin Bradford Hill in 1965 [6]. Despite the philosophy that a set of sufficient criteria does not exist [7], the nine Bradford Hill viewpoints have become popular among epidemiologists.

In a recent article by Shimonovich et al. [8], the Bradford Hill viewpoints were revisited to incorporate developments in causal thinking, primarily focusing on the following three causal approaches implicitly or explicitly constructed on the counterfactual framework: directed acyclic graphs (DAGs) [9-11], the sufficient cause model [12], and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology [13]. According to Shimonovich et al. [8], the sufficient cause model draws “attention to the impact of disease prevalence and the prevalence of competing causes on the strength of association or effect estimate,” and thus the sufficient cause model is useful for elucidating the theoretical underpinning of the first of the nine viewpoints—*strength* of association. This point was illustrated using a hypothetical example summarized in Table 2 in Shimonovich et al. [8].

In this article, we discuss how to discern the causal mechanisms of interest in the sufficient cause model, which pays closer attention to the relationship between the sufficient cause model and the Bradford Hill viewpoints. To this end, we explicate the link between the sufficient cause model and the potential-outcome model [14-19], which provides significant implications for interpretation of the observed risks. Our discussion, though relatively simple, would be pedagogically useful without loss of generality.

## Notation and settings

We let  $A$  denote a binary exposure of interest (1 = exposed, 0 = unexposed),  $Y$  a binary outcome (1 = outcome occurred, 0 = outcome did not occur), and  $C$  a binary confounder (1 = confounder is present, 0 = confounder is absent). Note: although Shimonovich et al. [8] let  $T$  and  $O$  denote an outcome and a confounder, respectively, we use a more common notation. In the counterfactual framework, we also let  $Y^{ac}$  denote the potential outcomes of  $Y$  if, possibly contrary to fact, there had been interventions to set  $A$  to  $a$  and to set  $C$  to  $c$ . While there is only one exposure of interest, we consider  $Y^{ac}$  rather than  $Y^a$ , which is useful and functions well in the following discussion [20]. For each individual, there would thus be four possible potential outcomes  $Y^{11}$ ,  $Y^{01}$ ,  $Y^{10}$ , and  $Y^{00}$ , resulting in 16 ( $= 2^4$ ) response types for  $Y$  [14]. Throughout the article, we assume that consistency is met [21, 22].

When considering a binary exposure  $A$ , a binary confounder  $C$ , and a binary outcome  $Y$ , we can enumerate nine ( $= 3^2$ ) types of sufficient causes for  $Y$  along with certain background factors in the sufficient-component cause framework, and there are 512 ( $= 2^9$ ) possible risk status types [15, 18]. A

correspondence between the 512 risk status types and 16 response types is shown in eTable in Suzuki et al. [18]. Under an assumption of sufficient cause positive monotonicity of  $A$  and  $C$  on  $Y$  [18, 23], which was not clearly mentioned in Shimonovich et al. [8], only the following four ( $= 2^2$ ) sufficient causes are considered:  $S_1 = U_1$ ,  $S_2 = U_2A$ ,  $S_3 = U_3C$ , and  $S_4 = U_4AC$  (Fig. 1) [18]. Here,  $U_k$  denotes a set of all components or factors, other than the presence of  $A$  and  $C$ , that may be required for a particular mechanism to operate. For simplicity, we denote the presence of these background factors as  $U_k = 1$  and their absence as  $U_k = 0$ . In this case, the number of possible risk status types is reduced from 512 to 16 ( $= 2^4$ ), and the number of possible response types is consequently reduced from 16 to six [18].

As shown in the lower part of Table 1, two hypothetical studies were considered in Shimonovich et al. [8], one in population 1 and the other in population 2 (see their Table 2). In these hypothetical studies, the exposure, the outcome, and the confounder are alcohol consumption, active tuberculosis, and overcrowding, respectively. In each study, a total of 400 subjects are equally divided into four subpopulations with  $(A, C) = (0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$ . Note that, although these data are apparently derived from observational studies,  $C$  and  $A$  are statistically independent in the two studies (i.e.,  $\Pr(C = c | A = a) = \Pr(C = c) = 0.5$  ( $a, c = 0, 1$ )), and there is no confounding in the two studies. This was apparently achieved by cohort matching, which can be described in a causal DAG (Fig. 2) [11, 24].

As implied in the footnote of their Table 2, Shimonovich et al. [8] apparently proceed with the understanding that individuals in the subpopulations with  $(A, C) = (0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$  have the outcome owing to sufficient causes  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_4$ , respectively. While this argument may seem reasonable, it does not hold without some strong assumptions. To clearly understand this point, it is significant to consider the link between the sufficient cause model and the potential-outcome model, which is discussed in the next section.

### **The link between the sufficient cause model and the potential-outcome model**

The sufficient cause model and the potential-outcome model have become the cornerstone of causal thinking in epidemiology and biomedicine [12, 25-28], and the link between the two models has been addressed [14-19].

Table 1 shows a correspondence between the 16 risk status types based on the sufficient cause model and the six response types based on the potential-outcome model, under the assumption of sufficient cause positive monotonicity of  $A$  and  $C$  on  $Y$ . Note that there is a surjective but not injective correspondence; for example, an individual of response type 1 corresponds to either one of risk status types 1 to 8. Indeed, the sufficient cause model is a “finer” model than the potential-outcome model. We let  $p_i$ ,  $q_i$ ,  $r_i$ , and  $s_i$  ( $i = 1, \dots, 16$ ) denote the proportions of risk status type  $i$  in the subpopulations with  $(A, C) = (1, 1)$ ,  $(0, 1)$ ,  $(1, 0)$ , and  $(0, 0)$ , respectively. We also let  $t_i$  ( $i = 1, \dots, 16$ ) denote the proportion of risk status type  $i$  in the total population, which can be calculated as  $p_i \times 0.25 + q_i \times 0.25 + r_i \times 0.25 + s_i \times 0.25$ . The cells highlighted in either green, blue, or red in Table 1 represent proportions of those who

have the outcome in each subpopulation. For instance, in both population 1 and population 2, the observed risks are 0.9 in the subpopulation with  $(A, C) = (1, 1)$ , which is equivalent to the sum of  $p_1$  through  $p_{15}$ .

### Causal mechanisms responsible for the outcome occurrence

A clearer understanding of the link between the two models enables us to discern causal mechanisms [29] or sufficient causes responsible for the outcome occurrence. In this section, we discuss a general setting under the sufficient cause positive monotonicity assumption, which is followed in the next section by an explanation of possible assumptions that make the argument in Shimonovich et al. [8] hold.

In the columns of the potential outcomes of  $Y$  in Table 1, sufficient causes responsible for the occurrence of the outcome are shown in parentheses. Note that those who have the outcome are classified into three groups, as illustrated by the three colors. As an example, consider the subpopulation with  $(A, C) = (1, 1)$ . The first group is classified into the cells highlighted in blue. For instance, if their risk status type is 1 (i.e.,  $(U_1, U_2, U_3, U_4) = (1, 1, 1, 1)$ ), the outcome inevitably occurs when any one of the four sufficient causes— $S_1, S_2, S_3$ , or  $S_4$ —completes. Thus, although some may have the outcome owing to sufficient cause  $S_4$ , others may have it owing to one of the other three sufficient causes. Accordingly, we cannot conclude that  $S_4$  is responsible for the outcome occurrence in this group. The same is true for individuals of risk status types 3, 5, 7, 9, 11, and 13. The second group is classified into the cells highlighted in red. For example, let us consider individuals of risk status type 2 (i.e.,  $(U_1, U_2, U_3, U_4) = (1, 1, 1, 0)$ ). Note that individuals of risk status types 1 and 2 cannot be distinguished from the perspective of the counterfactual framework because they are classified in the same response type 1 (i.e.,  $(Y^{11}, Y^{01}, Y^{10}, Y^{00}) = (1, 1, 1, 1)$ ). By definition, however,  $S_4$  never completes among individuals of risk status type 2, and the outcome occurs when one of the other three sufficient causes completes. Thus, we can conclude that  $S_4$  is not responsible for the outcome occurrence among individuals of risk status type 2. The same is true for individuals of risk status types 4, 6, 8, 10, 12, and 14. The third group is classified into the cell highlighted in green. Among individuals of risk status type 15 (i.e.,  $(U_1, U_2, U_3, U_4) = (0, 0, 0, 1)$ ), only sufficient cause  $S_4$  can complete to induce the outcome, and we can conclude that  $S_4$  is responsible for the outcome occurrence. In other words, in the subpopulation with  $(A, C) = (1, 1)$ , the presence of risk status type 15 is a sufficient *but not necessary* condition for the presence of sufficient cause  $S_4$  that completes to induce the outcome. An analogous discussion applies to the subpopulations with  $(A, C) = (0, 1)$  and  $(1, 0)$ , and those who have the outcome can be classified into the three groups. In the subpopulation with  $(A, C) = (0, 0)$ , however, individuals of risk status types 1 through 8 have the outcome, and all owing to sufficient cause  $S_1$ . When  $(A, C) = (0, 0)$ , the other three sufficient causes cannot, by definition, complete.

The above discussion provides significant implications for interpretation of the observed risks.

In the subpopulations with  $(A, C) = (1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$ , the observed risks are not equivalent to the prevalence of the completed sufficient causes  $S_2$ ,  $S_3$ , and  $S_4$ , respectively. Rather, the observed risks are merely upper bounds of their prevalence, and their trivial lower bounds are 0. Therefore, as further discussed below, unlike the explanation by Shimonovich et al. [8], the differing observed risks when  $(A, C) = (0, 1)$  in population 1 (i.e., 0.7) and population 2 (i.e., 0.4) could result from differing prevalence of completed sufficient cause  $S_1$  even if the prevalence of completed sufficient cause  $S_3$  is identical between populations. For instance, let us assume that the following equations hold in the subpopulation with  $(A, C) = (0, 1)$  in population 1:

$$q_1 + q_2 + q_3 + q_4 + q_5 + q_6 + q_7 = 0.0$$

$$q_8 = 0.3$$

$$q_9 + q_{10} + q_{13} + q_{14} = 0.4$$

$$q_{11} + q_{12} + q_{15} + q_{16} = 0.3$$

The observed risk is 0.7. Furthermore, let us assume that the following equations hold in the subpopulation with  $(A, C) = (0, 1)$  in population 2:

$$q_1 + q_2 + q_3 + q_4 + q_5 + q_6 + q_7 + q_8 = 0.0$$

$$q_9 + q_{10} + q_{13} + q_{14} = 0.4$$

$$q_{11} + q_{12} + q_{15} + q_{16} = 0.6$$

The observed risk is 0.4. In this scenario, the prevalence of completed sufficient cause  $S_3$  is 0.4 in both populations. The difference of the observed risks (i.e., 0.3) is entirely attributable to the different proportions of those who have the outcome owing to sufficient cause  $S_1$  (i.e.,  $q_8$ ). Regarding the subpopulation with  $(A, C) = (0, 0)$ , however, the interpretation of observed risk is completely different. We can unequivocally say that the observed risk is equivalent to the prevalence of the completed sufficient cause  $S_1$  because the other three sufficient causes cannot, by definition, complete. Thus, in both populations, the prevalence of completed sufficient cause  $S_1$  is 0.2 in the subpopulation with  $(A, C) = (0, 0)$ .

### Two possible but strict assumptions

On the basis of the above discussion, this section considers two possible assumptions, under both of which the argument of Shimonovich et al. [8] holds.

First, we consider a very strict assumption: individuals in the subpopulation with  $(A, C) = (0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , or  $(1, 1)$  can be at risk of only one sufficient cause,  $S_1$ ,  $S_2$ ,  $S_3$ , or  $S_4$ , respectively. This assumption can be written in terms of the proportions of the risk status types as follows:

Assumption 1:

$$p_i = 0 \ (i = 1, \dots, 14)$$

$$q_i = 0 \ (i = 1, \dots, 13, 15)$$

$$r_i = 0 \ (i = 1, \dots, 11, 13, 14, 15)$$

$$s_i = 0 \ (i = 1, \dots, 7, 9, \dots, 15)$$

Thus, of the 64 ( $= 16 \times 4$ ) cells of the four subpopulations in Table 1, a total of 56 cells are assumed to be 0, and individuals can be classified into the remaining eight cells. Under this assumption, the proportions of risk status types in population 1 and population 2 become

$$p_{15} = 0.9$$

$$p_{16} = 0.1$$

$$q_{14} = 0.7$$

$$q_{16} = 0.3$$

$$r_{12} = 0.6$$

$$r_{16} = 0.4$$

$$s_8 = 0.2$$

$$s_{16} = 0.8$$

and

$$p_{15} = 0.9$$

$$p_{16} = 0.1$$

$$q_{14} = 0.4$$

$$q_{16} = 0.6$$

$$r_{12} = 0.6$$

$$r_{16} = 0.4$$

$$s_8 = 0.2$$

$$s_{16} = 0.8,$$

respectively. Note that the difference of the risks in the subpopulation with  $(A, C) = (0, 1)$  between population 1 (i.e., 0.7) and population 2 (i.e., 0.4) is entirely attributable to the different proportions of those who are at risk of only sufficient cause  $S_3$  (i.e.,  $q_{14}$ ). This is a very strict assumption from two perspectives. First, all individuals can be at risk of, at most, only one sufficient cause. As shown in Table 1, 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 the two models. Second, individuals of a specific risk status type are “pre-determined” to be exposed to a specific situation of exposure and confounder. For instance, those who are at risk of only  $S_4$  (i.e., risk status type 15) are always included in the subpopulation with  $(A, C) = (1, 1)$ . This is a very strong assumption of a deterministic, selective treatment assignment.

Next, we slightly weaken assumption 1: although individuals can be at risk of two or more sufficient causes, only sufficient causes  $S_1, S_2, S_3$ , and  $S_4$  may *potentially* complete to induce the outcome in the subpopulations with  $(A, C) = (0, 0), (1, 0), (0, 1)$ , and  $(1, 1)$ , respectively. This assumption can be written in terms of the proportions of the risk status types as follows:

Assumption 2:

$$p_i = 0 \ (i = 1, \dots, 14)$$

$$q_i = 0 \ (i = 1, \dots, 8)$$

$$r_i = 0 \ (i = 1, \dots, 8)$$

Note that, of the 64 ( $= 16 \times 4$ ) cells of the four subpopulations in Table 1, a total of 30 cells highlighted in either blue or red are assumed to be 0, and individuals can be classified into the remaining 34 cells, which are either highlighted in green or not highlighted. Under this assumption, the proportions of risk status types in population 1 and population 2 become

$$p_{15} = 0.9$$

$$p_{16} = 0.1$$

$$q_9 + q_{10} + q_{13} + q_{14} = 0.7$$

$$q_{11} + q_{12} + q_{15} + q_{16} = 0.3$$

$$r_9 + r_{10} + r_{11} + r_{12} = 0.6$$

$$r_{13} + r_{14} + r_{15} + r_{16} = 0.4$$

$$S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + S_7 + S_8 = 0.2$$

$$S_9 + S_{10} + S_{11} + S_{12} + S_{13} + S_{14} + S_{15} + S_{16} = 0.8$$

and

$$p_{15} = 0.9$$

$$p_{16} = 0.1$$

$$q_9 + q_{10} + q_{13} + q_{14} = 0.4$$

$$q_{11} + q_{12} + q_{15} + q_{16} = 0.6$$

$$r_9 + r_{10} + r_{11} + r_{12} = 0.6$$

$$r_{13} + r_{14} + r_{15} + r_{16} = 0.4$$

$$s_1 + s_2 + s_3 + s_4 + s_5 + s_6 + s_7 + s_8 = 0.2$$

$$s_9 + s_{10} + s_{11} + s_{12} + s_{13} + s_{14} + s_{15} + s_{16} = 0.8,$$

respectively. Note that the difference of the risks in the subpopulation with  $(A, C) = (0, 1)$  between population 1 (i.e., 0.7) and population 2 (i.e., 0.4) is entirely attributable to the different proportions of those in which only sufficient cause  $S_3$  can *potentially* complete (i.e.,  $q_9 + q_{10} + q_{13} + q_{14}$ ). Assumption 2 is indeed less strict than assumption 1 from two perspectives. First, unlike assumption 1, an individual may be at risk of none, one, or several sufficient causes. Second, individuals of the same risk status type may be exposed to other situations of exposure and confounder. For instance, those who are at risk of only  $S_4$  (i.e., risk status type 15) can be classified in any of the four subpopulations with  $(A, C) = (1, 1)$ ,  $(0, 1)$ ,  $(1, 0)$ , and  $(0, 0)$ , although individuals who are at risk of sufficient cause  $S_1$  (i.e., risk status types 1 to 8) are always included in the subpopulation with  $(A, C) = (0, 0)$ .

Two points are further worth mentioning. First, assumption 1 implies that there are no individuals of risk status types 1 to 7, 9, 10, 11, and 13 in the total population (i.e.,  $t_i = 0$  ( $i = 1, \dots, 7, 9, 10, 11, 13$ )), which excludes individuals of response type 2 (or risk status types 9 and 10). Under the sufficient cause positive monotonicity, this is the only response type of which individuals are at risk of two or more sufficient causes (Table 1). Therefore, if one assumes that individuals can be at risk of, at most, only one sufficient cause, this response type is inevitably excluded. Note that response type 2 has been termed “competing antagonism” [30] or “agonism” [31-34] in the counterfactual framework. By contrast, under assumption 2, individuals of all the 16 risk status types may exist in the total population. Second, unlike assumption 1, assumption 2 can be also written in terms of the proportions of response types; there are no individuals of response types 1 to 4 in the subpopulation with  $(A, C) = (1, 1)$ , and there are no individuals of response type 1 in the subpopulations with  $(A, C) = (0, 1)$  and  $(1, 0)$ . Thus, assumption 2 is technically based on the potential-outcome model rather than the sufficient cause model.

Though less strict than assumption 1, assumption 2 still imposes some unnecessary restrictions because the proportions in the cells highlighted in blue in Table 1 are assumed to be all 0. Recall that these cells include individuals who have the outcome owing to  $S_2$ ,  $S_3$ , and  $S_4$  in the subpopulations with  $(A, C) = (1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$ , respectively. Then, how can one discern the prevalence of the completed sufficient causes in these subpopulations? In this regard, it is helpful to use the concept of potential completion times of sufficient causes [35], which is discussed in the next section.

### Discerning completed sufficient causes by the potential completion times of sufficient causes

Let  $d_1$ ,  $d_2$ ,  $d_3$ , and  $d_4$  denote potential completion times of sufficient causes  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_4$  at which an outcome would occur in an individual, respectively [35]. We assume that each potential completion

time is different. We also let  $d^{11}$ ,  $d^{01}$ ,  $d^{10}$ , and  $d^{00}$  denote the potential outcome occurrence time of an individual when  $(A, C) = (1, 1)$ ,  $(0, 1)$ ,  $(1, 0)$ , and  $(0, 0)$ , respectively, which can be defined as:

$$d^{11} = \min\{d_1, d_2, d_3, d_4\}$$

$$d^{01} = \min\{d_1, d_3\}$$

$$d^{10} = \min\{d_1, d_2\}$$

$$d^{00} = d_1$$

Let  $h$  denote a maximum follow-up time for an individual. Then, the potential outcomes of  $Y$  can be described as:

$$Y^{11} = I(d^{11} \leq h)$$

$$Y^{01} = I(d^{01} \leq h)$$

$$Y^{10} = I(d^{10} \leq h)$$

$$Y^{00} = I(d^{00} \leq h)$$

Because we consider four types of sufficient causes, a total of 120  $(= (4 + 1)!)^4$  sequence types can be enumerated [35], which is a finer classification than the 16 risk status types.

In Table 2, we show risk in each subpopulation in terms of potential completion times of sufficient causes. Note that, in each of the subpopulations with  $(A, C) = (1, 1)$ ,  $(0, 1)$ , and  $(1, 0)$ , the risk shown in the first row is decomposed into the risks shown in the second and the third rows. By using the concept of potential completion times of sufficient causes, we can clearly define the prevalence of each of the completed sufficient causes.

The concept of potential completion times of sufficient causes enables us to consider a further less strict assumption, which is consistent with the argument of Shimonovich et al. [8]:

Assumption 3:

$$\Pr((d^{11} \leq h) \wedge (d_4 > \min\{d_1, d_2, d_3\}) | (A, C) = (1, 1)) = 0$$

$$\Pr((d^{01} \leq h) \wedge (d_3 > d_1) | (A, C) = (0, 1)) = 0$$

$$\Pr((d^{10} \leq h) \wedge (d_2 > d_1) | (A, C) = (1, 0)) = 0$$

Note that, unlike assumption 2, while a total of 15 cells highlighted in red in Table 1 are assumed to be 0, individuals may be classified into the 15 cells highlighted in blue. For instance, when the first equation in assumption 3 is met, there is no one who has the outcome owing to either one of sufficient causes  $S_1$ ,  $S_2$ , and  $S_3$  in the subpopulation with  $(A, C) = (1, 1)$ , and all of them may have the outcome owing to sufficient cause  $S_4$  (see footnote e of Table 2). Assumption 3 is indeed less strict than assumption 2 from

two perspectives. First, unlike assumption 2, two or more sufficient causes may *potentially* complete. Although only one sufficient cause with the shortest potential completion time may actually complete, depending on the situations of  $(A, C)$  [35], this certainly allows more types of individuals to exist. Second, more individuals of the same risk status type may be exposed to different situations of  $(A, C)$ . For example, recall that, under assumption 2, individuals who are at risk of sufficient cause  $S_1$  (i.e., risk status types 1 to 8) are always included in the subpopulation with  $(A, C) = (0, 0)$ . Under assumption 3, however, individuals of risk status types 1 to 7 may find themselves in the subpopulations with conditions other than  $(A, C) = (0, 0)$ , although individuals of risk status type 8 always find themselves in the subpopulation with  $(A, C) = (0, 0)$ .

In the Appendix, we show the relationships between assumptions 1 to 3 in terms of the potential completion times of sufficient causes.

### **Discussion: strengthening our assessment of causality**

In this article, we discuss a conceptual framework to discern causal mechanisms of interest in the sufficient cause model. Under the assumption of sufficient cause positive monotonicity [18, 23], the observed risks in the subpopulations with  $(A, C) = (1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$  do not necessarily correspond to the prevalence of causal mechanisms responsible for the outcome or to completed sufficient causes  $S_2$ ,  $S_3$ , and  $S_4$ , respectively. Rather, the observed risks generally provide their upper bounds. Regarding the subpopulation with  $(A, C) = (0, 0)$ , however, the observed risk is equivalent to the prevalence of the completed sufficient cause  $S_1$ . On a related issue, VanderWeele [36] provided methods for obtaining lower bounds on the proportion of the outcome due to sufficient cause  $S_4$  in the population, both with and without counterfactual monotonicity. Indeed, the *strength* of effect depends on the distributions of relevant background factors in the population, which may well vary across time [12, 35, 37]. The concept of potential completion times of sufficient causes is useful to fully discerning completed sufficient causes, which provides a clearer mechanistic understanding of the amount of time until the occurrence of the outcome. This leads us to pay closer attention to the fourth of the nine Bradford Hill viewpoints—*temporality* [38].

Decades after its introduction, the sufficient cause model may be vaguely understood and thus implicitly used under unreasonably strict assumptions. Importantly, our discussions are based on the link between the sufficient cause model and the potential-outcome model [14-19], which contributes to a clearer understanding of the sufficient cause model. The potential-outcome model provides a formal theory of causal assessment, and has been extensively applied in a recent methodological development of epidemiology. However, no one should overlook other closely related but distinct causal models [39, 40]. Indeed, the sufficient cause model is essential to addressing important causal concepts in epidemiology, including “biologic” interaction [17, 41, 42], mediation [43-45], and etiologic fraction [35, 36, 46], which would be useful to consider the *plausibility* of association—the sixth of the nine

Bradford Hill viewpoints. The sufficient cause model can also be used to distinguish and relate concepts of confounding and of covariate balance [37, 47]. Furthermore, the use of DAGs is now widespread among epidemiologists as a tool for describing what is already known about relevant causal structures [9-11]. Some developments have been made in incorporating sufficient causes into the DAG framework [37, 44, 45, 48, 49].

Within the philosophical literature, the sufficient cause framework is most closely associated with the work of Mackie [50], 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 a shorthand for this statement [50]. Following Rothman [12], we refer to each component of a sufficient cause as a “component cause” or simply as a “cause”, which is consistent with the notion of an INUS condition. Thus, our discussion based on the link between the sufficient cause model and the potential-outcome model is essentially compatible with Mackie’s notion of an INUS condition. Note that, by borrowing an expression from Mackie [50], the causal mechanism of interest can be described as “a condition which is itself unnecessary but sufficient for the result.”

Despite their popularity, the nine Bradford Hill viewpoints do not provide a set of sufficient criteria to distinguish causal from noncausal explanations, and temporality is a *sine qua non* for causal explanations of observed associations [5]. Although Bradford Hill viewpoints are still helpful in assessing causality from an inductive standpoint, it is important to have a deductive form of logic in the formal causal models to draw causal conclusions, by clarifying the premises to be evaluated [51]. To *strengthen* our assessment in the face of multifactorial causality, it is significant to carefully scrutinize the observed associations in a complementary manner, using the sufficient cause model as well as its relevant causal models, with both inductive and deductive reasoning in mind.

## Appendix

In terms of the potential completion times of sufficient causes, assumptions 1 and 2 can be algebraically rewritten as follows:

Assumption 1:

$$\Pr(\min\{d_1, d_2, d_3\} \leq h | (A, C) = (1, 1)) = 0$$

$$\Pr(\min\{d_1, d_2, d_4\} \leq h | (A, C) = (0, 1)) = 0$$

$$\Pr(\min\{d_1, d_3, d_4\} \leq h | (A, C) = (1, 0)) = 0$$

$$\Pr(\min\{d_2, d_3, d_4\} \leq h | (A, C) = (0, 0)) = 0$$

Assumption 2:

$$\Pr(\min\{d_1, d_2, d_3\} \leq h | (A, C) = (1, 1)) = 0$$

$$\Pr(d_1 \leq h | (A, C) = (0, 1)) = 0$$

$$\Pr(d_1 \leq h | (A, C) = (1, 0)) = 0$$

These clearly show that assumption 2 is less strict than assumption 1. As mentioned in the text, assumption 3 is less strict than assumption 2, which can be more readily seen by rewriting assumption 3 as follows:

Assumption 3:

$$\Pr(\min\{d_1, d_2, d_3\} \leq \min\{d_4, h\} | (A, C) = (1, 1)) = 0$$

$$\Pr(d_1 \leq \min\{d_3, h\} | (A, C) = (0, 1)) = 0$$

$$\Pr(d_1 \leq \min\{d_2, h\} | (A, C) = (1, 0)) = 0$$

### Figure captions

**Fig. 1** Four types of sufficient causes for  $Y$

We consider a binary exposure  $A$ , a binary outcome  $Y$ , and a binary confounder  $C$ , under the assumption of sufficient cause positive monotonicity of  $A$  and  $C$  on  $Y$ .

**Fig. 2** Causal directed acyclic graph illustrating relationships between exposure  $A$ , outcome  $Y$ , and confounder  $C$

Unlike Fig. 1 in Shimonovich et al. [8], cohort matching on a confounder  $C$  is described. The variable  $S$  indicates whether an individual in the source population is selected for the matched study (1 = selected, 0 = not selected). The square around  $S = 1$  indicates that the analysis is conditional on having been selected. Although the variables  $C$  and  $A$  are d-connected via two paths,  $C \rightarrow A$  and  $C \rightarrow \boxed{S = 1} \leftarrow A$ , these variables are independent by design (i.e.,  $C \perp\!\!\!\perp A | S = 1$ ). In other words, the two paths “unfaithfully” cancel each other exactly.

### Declarations

**Funding:** ES is supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP20K10471, JP18K10104, and JP20K10499). Funding sources had no involvement in study design, the collection, analysis and interpretation of data, the writing of the report, and the decision to submit the article for publication.

**Conflicts of interest:** The authors have no relevant financial or non-financial interests to disclose.

**Availability of data and material:** Not applicable.

**Code availability:** Not applicable.

**Ethics approval:** This is a theoretical study, and no ethical approval is required.

**Consent to participate:** Not applicable.

**Consent to publication:** Not applicable.

## References

- 1 Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95 Suppl 1:S144-50.
- 2 Lucas RM, McMichael AJ. Association or causation: evaluating links between "environment and disease". *Bull World Health Organ*. 2005;83:792-5.
- 3 Morabia A. Hume, Mill, Hill, and the sui generis epidemiologic approach to causal inference. *Am J Epidemiol*. 2013;178:1526-32.
- 4 Olsen J, Jensen UJ. Causal criteria: time has come for a revision. *Eur J Epidemiol*. 2019;34:537-41.
- 5 VanderWeele TJ, Lash TL, Rothman KJ. Causal inference and scientific reasoning. In: Lash TL, VanderWeele TJ, Haneuse S, et al., editors. *Modern Epidemiology*. 4th ed. Philadelphia, PA: Wolters Kluwer, 2021. pp. 17-31.
- 6 Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58:295-300.
- 7 Rothman KJ, Lash TL, Haneuse S, et al. The scope of epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, et al., editors. *Modern Epidemiology*. 4th ed. Philadelphia, PA: Wolters Kluwer, 2021. pp. 3-15.
- 8 Shimonovich M, Pearce A, Thomson H, et al. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. *Eur J Epidemiol*. 2021.
- 9 Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*. 1999;10:37-48.
- 10 Glymour MM. Using causal diagrams to understand common problems in social epidemiology. In: Oakes JM, Kaufman JS, editors. *Methods in Social Epidemiology*. 2nd ed. San Francisco, CA: Jossey-Bass, 2017. pp. 458-92.
- 11 Suzuki E, Shinozaki T, Yamamoto E. Causal diagrams: pitfalls and tips. *J Epidemiol*. 2020;30:153-62.
- 12 Rothman KJ. Causes. *Am J Epidemiol*. 1976;104:587-92.
- 13 Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924-6.
- 14 Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scand J Work Environ Health*. 1988;14:125-9.
- 15 Flanders WD. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. *Eur J Epidemiol*. 2006;21:847-53.
- 16 VanderWeele TJ, Hernán MA. From counterfactuals to sufficient component causes and vice versa. *Eur J Epidemiol*. 2006;21:855-8.
- 17 VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika*. 2008;95:49-61.

- 18 Suzuki E, Yamamoto E, Tsuda T. On the link between sufficient-cause model and potential-outcome model. *Epidemiology*. 2011;22:131-2.
- 19 VanderWeele TJ, Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. *Ann Stat*. 2012;40:2128-61.
- 20 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.
- 21 Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*. 2009;20:3-5.
- 22 VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology*. 2009;20:880-3.
- 23 VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. *Epidemiology*. 2007;18:329-39.
- 24 Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *Int J Epidemiol*. 2013;42:860-9.
- 25 Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol*. 1974;66:688-701.
- 26 Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15:413-9.
- 27 Kaufman JS, Poole C. Looking back on "causal thinking in the health sciences". *Annu Rev Public Health*. 2000;21:101-19.
- 28 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.
- 29 Suzuki E, VanderWeele TJ. Mechanisms and uncertainty in randomized controlled trials: a commentary on Deaton and Cartwright. *Soc Sci Med*. 2018;210:83-5.
- 30 VanderWeele TJ, Knol MJ. Remarks on antagonism. *Am J Epidemiol*. 2011;173:1140-7.
- 31 Lin SH, Huang YT, Yang HI. On identification of agonistic interaction: Hepatitis B and C interaction on hepatocellular carcinoma. *Stat Med*. 2019;38:2467-76.
- 32 Lin JH, Lin KI, Lee WC, et al. Stochastic approach for mechanistic interaction under longitudinal studies with noninformative right censoring. *Stat Med*. 2020;39:114-28.
- 33 Huang YT, Tai AS, Chou MY, et al. Six-way decomposition of causal effects: Unifying mediation and mechanistic interaction. *Stat Med*. 2020;39:4051-68.
- 34 Tai AS, Huang YT, Lee WC, et al. Conceptualization of agonistic interaction in a marginal sufficient component cause model: an alternative interpretation for subadditive interaction. *Journal of the Chinese Statistical Association*. 2020;58:168-98.
- 35 Suzuki E, Yamamoto E, Tsuda T. On the relations between excess fraction, attributable fraction, and etiologic fraction. *Am J Epidemiol*. 2012;175:567-75.

- 36 VanderWeele TJ. Attributable fractions for sufficient cause interactions. *Int J Biostat.* 2010;6:5.
- 37 Suzuki E, Tsuda T, Yamamoto E. Covariate balance for no confounding in the sufficient-cause model. *Ann Epidemiol.* 2018;28:48-53.
- 38 Suzuki E. Time changes, so do people. *Soc Sci Med.* 2012;75:452-6.
- 39 Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol.* 2002;31:1030-7.
- 40 Suzuki E. Generalized causal measure: the beauty lies in its generality. *Epidemiology.* 2015;26:490-5.
- 41 VanderWeele TJ. Sufficient cause interactions and statistical interactions. *Epidemiology.* 2009;20:6-13.
- 42 VanderWeele TJ. A word and that to which it once referred: assessing "biologic" interaction. *Epidemiology.* 2011;22:612-3.
- 43 Hafeman DM. A sufficient cause based approach to the assessment of mediation. *Eur J Epidemiol.* 2008;23:711-21.
- 44 VanderWeele TJ. Mediation and mechanism. *Eur J Epidemiol.* 2009;24:217-24.
- 45 Suzuki E, Yamamoto E, Tsuda T. Identification of operating mediation and mechanism in the sufficient-component cause framework. *Eur J Epidemiol.* 2011;26:347-57.
- 46 Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol.* 1988;128:1185-97.
- 47 Suzuki E, Yamamoto E. Re: A graphical description of partial exchangeability. *Epidemiology.* 2021;32:e7-e9.
- 48 VanderWeele TJ, Robins JM. Directed acyclic graphs, sufficient causes, and the properties of conditioning on a common effect. *Am J Epidemiol.* 2007;166:1096-104.
- 49 VanderWeele TJ, Robins JM. Minimal sufficient causation and directed acyclic graphs. *Ann Stat.* 2009;37:1437-65.
- 50 Mackie JL. Causes and conditions. *Am Philos Q.* 1965;2:245-64.
- 51 VanderWeele TJ. Hill's causal considerations and the potential outcomes framework. *Observ Stud.* 2020;6:47-54.

**Table 1** Correspondence between sufficient causes and potential outcomes under the assumption of sufficient cause positive monotonicity<sup>a</sup>

| Sufficient cause model |                                         |                |                |                | Potential-outcome model <sup>b</sup> |                                                                           |                                       |                                       |                     | Proportions of risk status types in population with (A, C) = c, d, e |                 |                 |                 |                    |
|------------------------|-----------------------------------------|----------------|----------------|----------------|--------------------------------------|---------------------------------------------------------------------------|---------------------------------------|---------------------------------------|---------------------|----------------------------------------------------------------------|-----------------|-----------------|-----------------|--------------------|
| Risk status type       | Background factors of sufficient causes |                |                |                | Response type                        | Potential outcomes of Y                                                   |                                       |                                       |                     | (1, 1)                                                               | (0, 1)          | (1, 0)          | (0, 0)          | Total <sup>f</sup> |
|                        | U <sub>1</sub>                          | U <sub>2</sub> | U <sub>3</sub> | U <sub>4</sub> |                                      | γ <sup>11</sup>                                                           | γ <sup>01</sup>                       | γ <sup>10</sup>                       | γ <sup>00</sup>     | (N = 100)                                                            | (N = 100)       | (N = 100)       | (N = 100)       | (N = 400)          |
| 1                      | 1                                       | 1              | 1              | 1              | 1                                    | 1 (S <sub>1</sub> or S <sub>2</sub> or S <sub>3</sub> or S <sub>4</sub> ) | 1 (S <sub>1</sub> or S <sub>3</sub> ) | 1 (S <sub>1</sub> or S <sub>2</sub> ) | 1 (S <sub>1</sub> ) | p <sub>1</sub>                                                       | q <sub>1</sub>  | r <sub>1</sub>  | s <sub>1</sub>  | t <sub>1</sub>     |
| 2                      | 1                                       | 1              | 1              | 0              | 1                                    | 1 (S <sub>1</sub> or S <sub>2</sub> or S <sub>3</sub> )                   | 1 (S <sub>1</sub> or S <sub>3</sub> ) | 1 (S <sub>1</sub> or S <sub>2</sub> ) | 1 (S <sub>1</sub> ) | p <sub>2</sub>                                                       | q <sub>2</sub>  | r <sub>2</sub>  | s <sub>2</sub>  | t <sub>2</sub>     |
| 3                      | 1                                       | 1              | 0              | 1              | 1                                    | 1 (S <sub>1</sub> or S <sub>2</sub> or S <sub>4</sub> )                   | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> or S <sub>2</sub> ) | 1 (S <sub>1</sub> ) | p <sub>3</sub>                                                       | q <sub>3</sub>  | r <sub>3</sub>  | s <sub>3</sub>  | t <sub>3</sub>     |
| 4                      | 1                                       | 1              | 0              | 0              | 1                                    | 1 (S <sub>1</sub> or S <sub>2</sub> )                                     | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> or S <sub>2</sub> ) | 1 (S <sub>1</sub> ) | p <sub>4</sub>                                                       | q <sub>4</sub>  | r <sub>4</sub>  | s <sub>4</sub>  | t <sub>4</sub>     |
| 5                      | 1                                       | 0              | 1              | 1              | 1                                    | 1 (S <sub>1</sub> or S <sub>3</sub> or S <sub>4</sub> )                   | 1 (S <sub>1</sub> or S <sub>3</sub> ) | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> ) | p <sub>5</sub>                                                       | q <sub>5</sub>  | r <sub>5</sub>  | s <sub>5</sub>  | t <sub>5</sub>     |
| 6                      | 1                                       | 0              | 1              | 0              | 1                                    | 1 (S <sub>1</sub> or S <sub>3</sub> )                                     | 1 (S <sub>1</sub> or S <sub>3</sub> ) | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> ) | p <sub>6</sub>                                                       | q <sub>6</sub>  | r <sub>6</sub>  | s <sub>6</sub>  | t <sub>6</sub>     |
| 7                      | 1                                       | 0              | 0              | 1              | 1                                    | 1 (S <sub>1</sub> or S <sub>4</sub> )                                     | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> ) | p <sub>7</sub>                                                       | q <sub>7</sub>  | r <sub>7</sub>  | s <sub>7</sub>  | t <sub>7</sub>     |
| 8                      | 1                                       | 0              | 0              | 0              | 1                                    | 1 (S <sub>1</sub> )                                                       | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> )                   | 1 (S <sub>1</sub> ) | p <sub>8</sub>                                                       | q <sub>8</sub>  | r <sub>8</sub>  | s <sub>8</sub>  | t <sub>8</sub>     |
| 9                      | 0                                       | 1              | 1              | 1              | 2                                    | 1 (S <sub>2</sub> or S <sub>3</sub> or S <sub>4</sub> )                   | 1 (S <sub>3</sub> )                   | 1 (S <sub>2</sub> )                   | 0                   | p <sub>9</sub>                                                       | q <sub>9</sub>  | r <sub>9</sub>  | s <sub>9</sub>  | t <sub>9</sub>     |
| 10                     | 0                                       | 1              | 1              | 0              | 2                                    | 1 (S <sub>2</sub> or S <sub>3</sub> )                                     | 1 (S <sub>3</sub> )                   | 1 (S <sub>2</sub> )                   | 0                   | p <sub>10</sub>                                                      | q <sub>10</sub> | r <sub>10</sub> | s <sub>10</sub> | t <sub>10</sub>    |
| 11                     | 0                                       | 1              | 0              | 1              | 3                                    | 1 (S <sub>2</sub> or S <sub>4</sub> )                                     | 0                                     | 1 (S <sub>2</sub> )                   | 0                   | p <sub>11</sub>                                                      | q <sub>11</sub> | r <sub>11</sub> | s <sub>11</sub> | t <sub>11</sub>    |
| 12                     | 0                                       | 1              | 0              | 0              | 3                                    | 1 (S <sub>2</sub> )                                                       | 0                                     | 1 (S <sub>2</sub> )                   | 0                   | p <sub>12</sub>                                                      | q <sub>12</sub> | r <sub>12</sub> | s <sub>12</sub> | t <sub>12</sub>    |
| 13                     | 0                                       | 0              | 1              | 1              | 4                                    | 1 (S <sub>3</sub> or S <sub>4</sub> )                                     | 1 (S <sub>3</sub> )                   | 0                                     | 0                   | p <sub>13</sub>                                                      | q <sub>13</sub> | r <sub>13</sub> | s <sub>13</sub> | t <sub>13</sub>    |
| 14                     | 0                                       | 0              | 1              | 0              | 4                                    | 1 (S <sub>3</sub> )                                                       | 1 (S <sub>3</sub> )                   | 0                                     | 0                   | p <sub>14</sub>                                                      | q <sub>14</sub> | r <sub>14</sub> | s <sub>14</sub> | t <sub>14</sub>    |
| 15                     | 0                                       | 0              | 0              | 1              | 5                                    | 1 (S <sub>4</sub> )                                                       | 0                                     | 0                                     | 0                   | p <sub>15</sub>                                                      | q <sub>15</sub> | r <sub>15</sub> | s <sub>15</sub> | t <sub>15</sub>    |
| 16                     | 0                                       | 0              | 0              | 0              | 6                                    | 0                                                                         | 0                                     | 0                                     | 0                   | p <sub>16</sub>                                                      | q <sub>16</sub> | r <sub>16</sub> | s <sub>16</sub> | t <sub>16</sub>    |

*Population 1<sup>g</sup>*

|                              |      |           |     |           |     |
|------------------------------|------|-----------|-----|-----------|-----|
| Observed risk                | 0.9  | 0.7       | 0.6 | 0.2       | 0.6 |
| Risk ratio                   | 4.5  | 3.5       | 3.0 | 1.0 (ref) |     |
| Risk ratio (stratified by C) | 1.29 | 1.0 (ref) | 3.0 | 1.0 (ref) |     |
| Risk difference              | 0.7  | 0.5       | 0.4 | 0.0 (ref) |     |

|                                   |      |           |     |           |       |
|-----------------------------------|------|-----------|-----|-----------|-------|
| Risk difference (stratified by C) | 0.2  | 0.0 (ref) | 0.4 | 0.0 (ref) |       |
| <i>Population 2<sup>g</sup></i>   |      |           |     |           |       |
| Observed risk                     | 0.9  | 0.4       | 0.6 | 0.2       | 0.525 |
| Risk ratio                        | 4.5  | 2.0       | 3.0 | 1.0 (ref) |       |
| Risk ratio (stratified by C)      | 2.25 | 1.0 (ref) | 3.0 | 1.0 (ref) |       |
| Risk difference                   | 0.7  | 0.2       | 0.4 | 0.0 (ref) |       |
| Risk difference (stratified by C) | 0.5  | 0.0 (ref) | 0.4 | 0.0 (ref) |       |

Abbreviation: ref, reference.

<sup>a</sup>We consider a binary exposure  $A$ , a binary outcome  $Y$ , and a binary confounder  $C$ . We let  $Y^{ac}$  denote a potential outcome of  $Y$ . Under an assumption of sufficient cause positive monotonicity of  $A$  and  $C$  on  $Y$ , we consider four types of sufficient causes for  $Y$ :  $S_1 = U_1$ ,  $S_2 = U_2A$ ,  $S_3 = U_3C$ , and  $S_4 = U_4AC$ . See text for details.

<sup>b</sup>Sufficient causes responsible for the occurrence of outcome  $Y$  are shown in parentheses.

<sup>c</sup>In the cells highlighted in green (or light gray in the print version), only one sufficient cause ( $S_1$ ,  $S_2$ ,  $S_3$ , or  $S_4$ ) completes by the end of the follow-up time in the subpopulation with  $(A, C) = (0, 0)$ ,  $(1, 0)$ ,  $(0, 1)$ , or  $(1, 1)$ , respectively. Under assumption 1 in the main text, 13 of these 17 cells are assumed to be 0.

<sup>d</sup>In the cells highlighted in blue (or medium gray in the print version), two or more sufficient causes *potentially* complete by the end of the follow-up time, and the outcome of interest inevitably occurs. However, we cannot conclude whether sufficient causes  $S_2$ ,  $S_3$ , and  $S_4$  are responsible for the outcome occurrence in the subpopulations with  $(A, C) = (1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$ , respectively. Under assumptions 1 and 2 in the main text, these 15 cells are assumed to be 0.

<sup>e</sup>In the cells highlighted in red (or dark gray in the print version), at least one sufficient cause *potentially* completes by the end of the follow-up time, and the outcome of interest inevitably occurs. However, sufficient causes  $S_2$ ,  $S_3$ , and  $S_4$  never complete in the subpopulations with  $(A, C) = (1, 0)$ ,  $(0, 1)$ , and  $(1, 1)$ , respectively. Under assumptions 1 to 3 in the main text, these 15 cells are assumed to be 0.

<sup>f</sup>Note that  $t_i$  can be calculated as  $p_i \times 0.25 + q_i \times 0.25 + r_i \times 0.25 + s_i \times 0.25$ .

<sup>g</sup>These hypothetical studies are derived from Table 2 in Shimonovich et al. [8]. In population 1, the observed risks of the outcome in the exposed group and the unexposed group are  $\Pr(Y = 1 | A = 1) = \Pr(Y = 1 | A = 1, C = 1) \times \Pr(C = 1 | A = 1) + \Pr(Y = 1 | A = 1, C = 0) \times \Pr(C = 0 | A = 1) = 0.9 \times 0.5 + 0.6 \times 0.5 = 0.75$  and  $\Pr(Y = 1 | A = 0) = \Pr(Y = 1 | A = 0, C = 1) \times \Pr(C = 1 | A = 0) + \Pr(Y = 1 | A = 0, C = 0) \times \Pr(C = 0 | A = 0) = 0.7 \times 0.5 + 0.2 \times 0.5 = 0.45$ , respectively. Because  $C$  and  $A$  are independent, these risks are equivalent to the counterfactual risks of the outcome when everyone in the population is exposed and unexposed, respectively. Thus, causal risk ratio and causal risk difference of  $A$  on  $Y$  in population 1 are obtained as  $\Pr(Y = 1 | A = 1) / \Pr(Y = 1 | A = 0) = 0.75 / 0.45 \approx 1.67$  and  $\Pr(Y = 1 | A = 1) - \Pr(Y = 1 | A = 0) = 0.75 - 0.45 = 0.3$ , respectively. In a similar manner, causal risk ratio and causal risk difference of  $A$  on  $Y$  in population 2 are obtained as  $\Pr(Y = 1 | A = 1) / \Pr(Y = 1 | A = 0) = 0.75 / 0.3 = 2.5$  and  $\Pr(Y = 1 | A = 1) - \Pr(Y = 1 | A = 0) = 0.75 - 0.3 = 0.45$ , respectively. From Theorem 2 of VanderWeele [36], lower bounds on the proportions of the outcome due to sufficient cause  $S_4$  in population 1 and population 2 can be calculated as  $\max\{0, (0.9 - 0.7 - 0.6 + 0.2) \times 0.25\} / 0.6 = 0$  and  $\max\{0, (0.9 - 0.4 - 0.6 + 0.2) \times 0.25\} / 0.525 \approx 0.048$ , respectively, if the effects of  $A$  and  $C$  on  $Y$  are unconfounded.

**Table 2** Risk in each subpopulation in terms of potential completion times of sufficient causes

| Risks <sup>a</sup>                                                                                                                 | Sufficient causes responsible for $Y = 1$ | Number of sequence types <sup>b</sup> |
|------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------|---------------------------------------|
| Subpopulation with $(A, C) = (1, 1)$                                                                                               |                                           |                                       |
| $\Pr(Y = 1   (A, C) = (1, 1)) = \Pr(d^{11} \leq h   (A, C) = (1, 1)) = \Pr(\min\{d_1, d_2, d_3, d_4\} \leq h   (A, C) = (1, 1))^c$ | $S_1$ or $S_2$ or $S_3$ or $S_4$          | 96                                    |
| $\Pr((d^{11} \leq h) \wedge (d_4 < \min\{d_1, d_2, d_3\})   (A, C) = (1, 1))^d$                                                    | $S_4$                                     | 24                                    |
| $\Pr((d^{11} \leq h) \wedge (d_4 > \min\{d_1, d_2, d_3\})   (A, C) = (1, 1))^{e, f}$                                               | $S_1$ or $S_2$ or $S_3$                   | 72                                    |
| Subpopulation with $(A, C) = (0, 1)$                                                                                               |                                           |                                       |
| $\Pr(Y = 1   (A, C) = (0, 1)) = \Pr(d^{01} \leq h   (A, C) = (0, 1)) = \Pr(\min\{d_1, d_3\} \leq h   (A, C) = (0, 1))^c$           | $S_1$ or $S_3$                            | 80                                    |
| $\Pr((d^{01} \leq h) \wedge (d_3 < d_1)   (A, C) = (0, 1))^d$                                                                      | $S_3$                                     | 40                                    |
| $\Pr((d^{01} \leq h) \wedge (d_3 > d_1)   (A, C) = (0, 1))^e$                                                                      | $S_1$                                     | 40                                    |
| Subpopulation with $(A, C) = (1, 0)$                                                                                               |                                           |                                       |
| $\Pr(Y = 1   (A, C) = (1, 0)) = \Pr(d^{10} \leq h   (A, C) = (1, 0)) = \Pr(\min\{d_1, d_2\} \leq h   (A, C) = (1, 0))^c$           | $S_1$ or $S_2$                            | 80                                    |
| $\Pr((d^{10} \leq h) \wedge (d_2 < d_1)   (A, C) = (1, 0))^d$                                                                      | $S_2$                                     | 40                                    |
| $\Pr((d^{10} \leq h) \wedge (d_2 > d_1)   (A, C) = (1, 0))^e$                                                                      | $S_1$                                     | 40                                    |
| Subpopulation with $(A, C) = (0, 0)$                                                                                               |                                           |                                       |
| $\Pr(Y = 1   (A, C) = (0, 0)) = \Pr(d^{00} \leq h   (A, C) = (0, 0)) = \Pr(d_1 \leq h   (A, C) = (0, 0))^c$                        | $S_1$                                     | 60                                    |

<sup>a</sup>We consider a binary exposure  $A$ , a binary outcome  $Y$ , and a binary confounder  $C$  under the assumption of sufficient cause positive monotonicity. See text for details.

<sup>b</sup>Because we consider four types of sufficient causes, a total of 120 ( $= (4 + 1)!$ ) sequence types can be enumerated.

<sup>c</sup>These risks correspond to a sum of the cells highlighted in either green, blue, or red in each subpopulation in Table 1.

<sup>d</sup>An upper bound of these risks is a sum of the cells highlighted in either green or blue in each subpopulation in Table 1, while a lower bound is a sum of the cells highlighted in green in each subpopulation.

<sup>e</sup>An upper bound of these risks is a sum of the cells highlighted in either red or blue in each subpopulation in Table 1, while a lower bound is a sum of the cells highlighted in red in each subpopulation. Under the assumption that these risks are 0 (i.e., assumption 3), the argument of Shimonovich et al. [8] holds.

<sup>f</sup>This risk can be further decomposed into three risks according to the completed sufficient causes responsible for the occurrence of outcome  $Y$ .

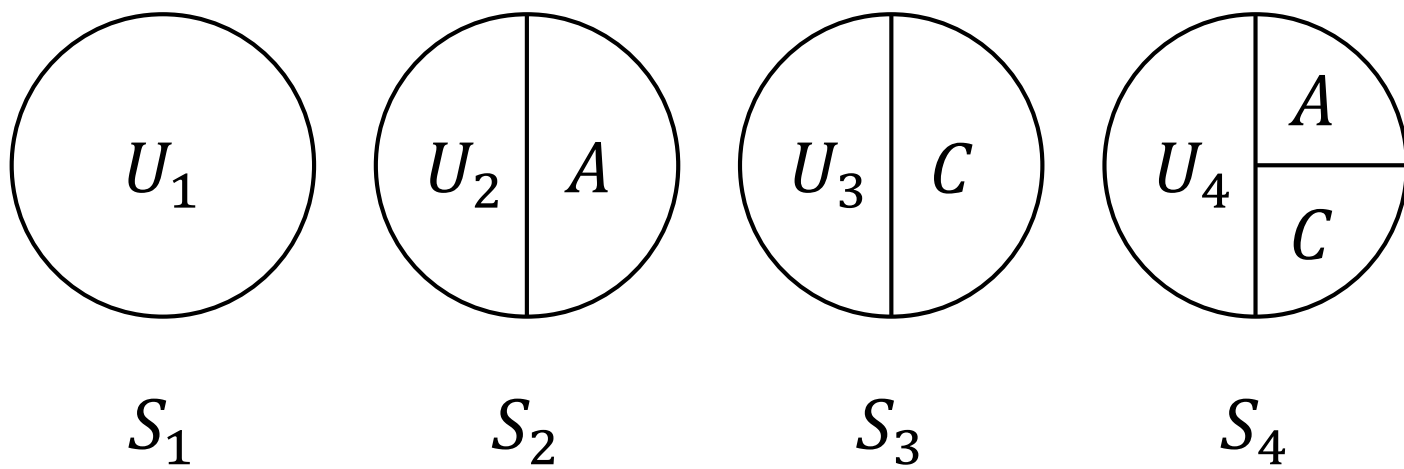


Fig. 1

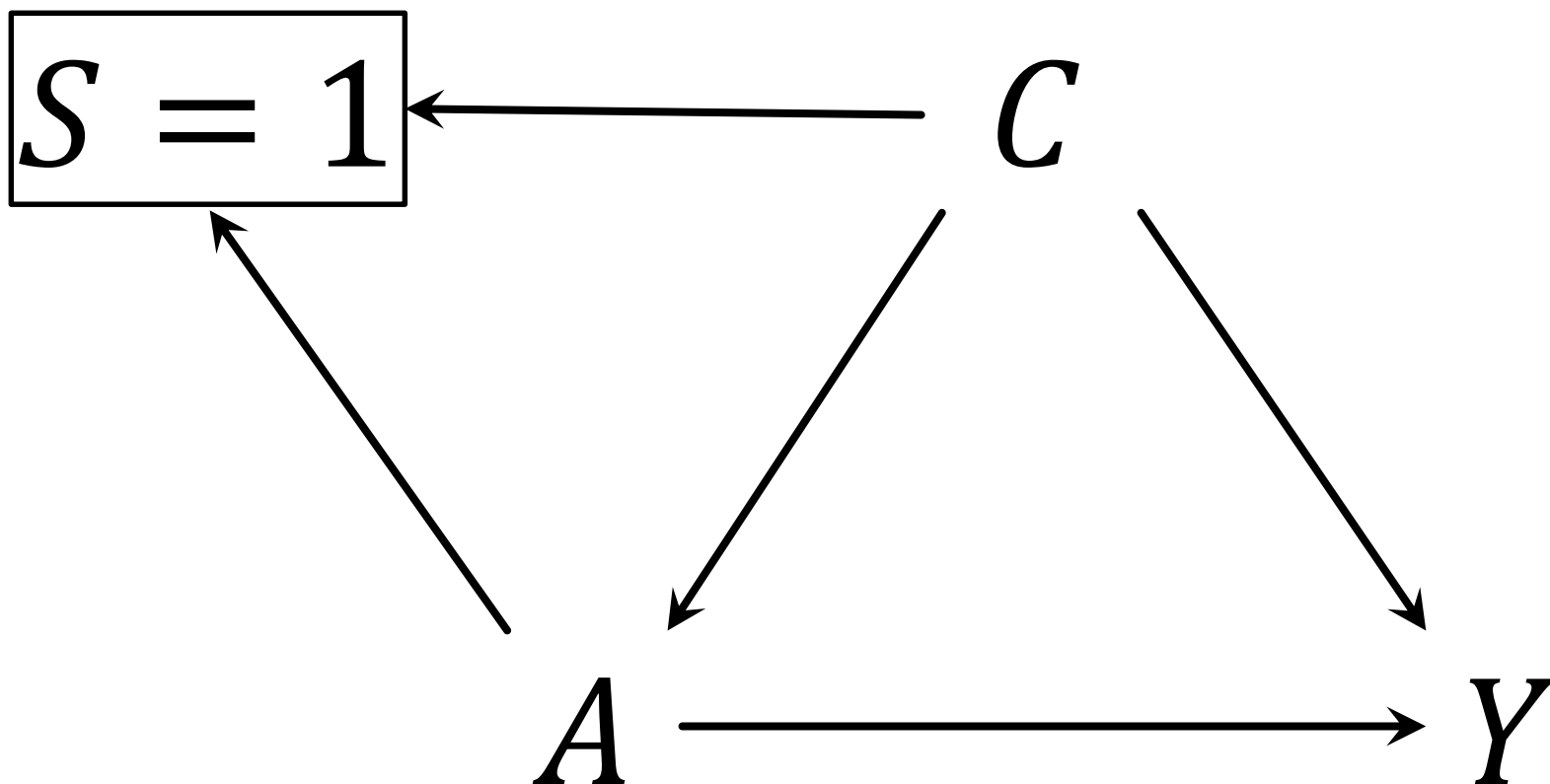


Fig. 2