Research Article

Etsuji Suzuki* and Eiji Yamamoto

Attributable fraction and related measures: Conceptual relations in the counterfactual framework

https://doi.org/10.1515/jci-2021-0068
received December 10, 2021; accepted December 14, 2022

Abstract: The attributable fraction (population) has attracted much attention from a theoretical perspective and has been used extensively to assess the impact of potential health interventions. However, despite its extensive use, there is much confusion about its concept and calculation methods. In this article, we discuss the concepts of and calculation methods for the attributable fraction and related measures in the counterfactual framework, both with and without stratification by covariates. Generally, the attributable fraction is useful when the exposure of interest has a causal effect on the outcome. However, it is important to understand that this statement applies to the exposed group. Although the target population of the attributable fraction (population) is the total population, the causal effect should be present not in the total population but in the exposed group. As related measures, we discuss the preventable fraction and prevented fraction, which are generally useful when the exposure of interest has a preventive effect on the outcome, and we further propose a new measure called the attributed fraction. We also discuss the causal and preventive excess fractions, and provide notes on vaccine efficacy. Finally, we discuss the relations between the aforementioned six measures and six possible patterns using a conceptual schema.

Keywords: attributable fraction, counterfactual model, excess fraction, preventable fraction, prevented fraction, vaccine efficacy

MSC 2020: 62D20, 62P10

1 Introduction

Since the paper by Doll in 1951 [1], the attributable fraction (population) has attracted much attention from a theoretical perspective [2–17] and has been used extensively in empirical studies to assess the impact of potential health interventions. Epidemiology textbooks provide several formulae for calculating the attributable fraction (population) [18], including the Levin formula introduced in 1953 [19] and Miettinen formula introduced in 1974 [2]. Despite its extensive use, there is much confusion about the concept of and calculation methods for the attributable fraction (population), presumably because less attention has been paid to its definition in the counterfactual framework.

In this article, we discuss the concepts of and calculation methods for the attributable fraction and related measures in the counterfactual framework, providing their conceptual relations in a comprehensive manner. As related measures, we discuss the preventable fraction and prevented fraction, which are

* **Corresponding author: Etsuji Suzuki,** Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama 700-8558, Japan; Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA, e-mail: esuzuki@hsph.harvard.edu, etsuji-s@cc.okayama-u.ac.jp
Eiji Yamamoto: Okayama University of Science, Okayama 700-0005, Japan

generally useful when the exposure of interest has a preventive effect on the outcome, and we further propose a new measure called the attributed fraction. We also discuss the causal and preventive excess fractions. As a result of this, we show that, as a prerequisite for the calculation of these measures, it is important to have clear definitions of them in the counterfactual framework, which would improve the interpretation and use of these measures.

2 Notation and setting

Let E denote a binary exposure of interest (1 = exposed, 0 = unexposed), Y denote a binary adverse outcome (1 = outcome occurred, 0 = outcome did not occur), and C denote a set of covariates. In the counterfactual framework, also let Y^e denote the potential outcomes of Y if, possibly contrary to the fact, there had been interventions to set E to e. Then, for each individual, there would be two relevant potential outcomes, Y^1 and Y^0, corresponding to what would have happened to that individual when that person was exposed and unexposed, respectively. When no confusion occurs for a binary variable X, we write X and $\bar{X}$ as shorthand for the events of $X = 1$ and $X = 0$, respectively.

Using this notation, the associational risk ratio (aRR) is defined as $\Pr(Y|E)/\Pr(Y|\bar{E})$, whereas the causal risk ratio (cRR) is defined as $\Pr(Y^1)/\Pr(Y^0)$. Similarly, we also define the cRRs in the exposed group (cRR_E) and unexposed group (cRR_U) as $\Pr(Y^1|E)/\Pr(Y^0|E)$ and $\Pr(Y^1|\bar{E})/\Pr(Y^0|\bar{E})$, respectively. Furthermore, we define the standardized mortality (or morbidity) ratio (SMR) as $\Pr(Y|E)/\Pr(Y^0|E)$ in the counterfactual framework [20], which is equal to cRR_E under the assumption of partial consistency under exposure (i.e., $Y^1 = Y$ if $E = 1$). Similarly, we define the standardized risk ratio in the unexposed group (SRR_U) as $\Pr(Y^1|\bar{E})/\Pr(Y|\bar{E})$, which is equal to cRR_U under the assumption of partial consistency under no exposure (i.e., $Y^0 = Y$ if $E = 0$). Note that each of the assumptions of partial consistency is slightly weaker than the assumption of consistency (i.e., $Y^e = Y$ if $E = e$ ($e = 0, 1$)) [21,22], which we assume throughout the article. However, it is worth mentioning that only partial consistency is required in some derivations. When the measures of interest are defined in stratum c, we add the subscript c to the corresponding measures (e.g., $\text{aRR}_c \equiv \Pr(Y|E, c)/\Pr(Y|\bar{E}, c)$). We use $\Pr(c)$ as shorthand for $\Pr(C = c)$. In Table S1, a summary of the measures is provided. The lemmas and their proofs are provided in Supplementary Appendix A.

3 Attributable fraction

In this section, we discuss the concepts of and calculation methods for the attributable fraction in the counterfactual framework, both with and without stratification by covariates. First, we provide an overview of the attributable fraction by presenting its definition, theorems, and corollaries. Then, we discuss a common misuse of the Levin formula (i.e., the "partially adjusted" method) and then discuss the condition under which the Levin formula is valid. Finally, we provide concluding remarks, and highlight the importance of understanding the definition of the attributable fraction to avoid common misunderstandings and confusion about it.

3.1 Overview

Generally, the attributable fraction (population) is used when one is 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; that is, the attributable fraction compares the observed risk with the counterfactual risk under $e = 0$.

Definition 1: When the target population is the total population, the attributable fraction (population) is defined as $\{\Pr(Y) - \Pr(Y^0)\}/\Pr(Y)$, whereas when the target population is the exposed group, the attributable fraction (exposed) is defined as $\{\Pr(Y|E) - \Pr(Y^0|E)\}/\Pr(Y|E)$.

Note that the numerator event is not included in the denominator event, and the attributable fraction ranges from $-\infty$ to 1. The attributable fraction in Definition 1 corresponds to the attributable caseload proposed by Suzuki et al. [13], who distinguished between the attributable caseload and attributable proportion, both of which compose the broader concept of the attributable fraction. The attributable proportion (population) and attributable proportion (exposed) are defined as $\{\Pr(Y) - \Pr(Y^0, Y)\}/\Pr(Y)$ and $\{\Pr(Y|E) - \Pr(Y^0, Y|E)\}/\Pr(Y|E)$, respectively, and both range from 0 to 1 [13]. Under the assumption of positive monotonicity of E on Y (i.e., $Y^0 \leq Y^1$ for all individuals), the attributable caseload and attributable proportion become equivalent. In this article, we use the term attributable fraction to refer to the attributable caseload.

Next we present Theorem 1.1 and Corollary 1.1, which provide the attributable fraction if one does not use stratification by C.

Theorem 1.1: *The attributable fraction* (*population*) *and attributable fraction* (*exposed*) *are obtained as* $\Pr(E|Y)(\mathrm{SMR} - 1)/\mathrm{SMR}$ *and* $(\mathrm{SMR} - 1)/\mathrm{SMR}$, *respectively*.

Corollary 1.1: *The attributable fraction* (*population*) *and attributable fraction* (*exposed*) *are obtained as* $\Pr(E|Y)(\mathrm{aRR} - 1)/\mathrm{aRR}$ *and* $(\mathrm{aRR} - 1)/\mathrm{aRR}$, *respectively, if and only if partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E$).

Note that partial exchangeability under no exposure (i.e., $Y^0 \perp\!\!\!\perp E$) is equivalent to SMR = aRR (Lemma 1), which is obviously expected to hold in an ideal randomized controlled trial. On a related issue, the equation $\Pr(Y^0|E) = \{\Pr(Y^0) - \Pr(Y, \bar{E})\}/\Pr(E)$ holds. Thus, if $\Pr(Y^0)$ is obtained in experimental studies, $\Pr(Y^0|E)$ is reducible to empirically estimable quantities [23]. Regarding the attributable fraction (population) in Corollary 1.1, the following equation holds irrespective of whether $Y^0 \perp\!\!\!\perp E$ holds:

$$\Pr(E|Y)\frac{\mathrm{aRR} - 1}{\mathrm{aRR}} = \frac{\Pr(E)(\mathrm{aRR} - 1)}{\Pr(E)(\mathrm{aRR} - 1) + 1}.$$

The first and second formulae are often called the Miettinen formula and Levin formula, respectively. By contrast, if one uses stratification by C, Theorem 1.2 and Corollary 1.2 provide the attributable fraction.

Theorem 1.2: *If one uses stratification by* C, *the attributable fraction* (*population*) *and attributable fraction* (*exposed*) *are obtained as* $\Pr(E|Y)\sum_c \Pr(c|Y, E)(\mathrm{SMR}_c - 1)/\mathrm{SMR}_c$ *and* $\sum_c \Pr(c|Y, E)(\mathrm{SMR}_c - 1)/\mathrm{SMR}_c$, *respectively*.

Corollary 1.2: *The attributable fraction* (*population*) *and attributable fraction* (*exposed*) *are obtained as* $\Pr(E|Y)\sum_c \Pr(c|Y, E)(\mathrm{aRR}_c - 1)/\mathrm{aRR}_c$ *and* $\sum_c \Pr(c|Y, E)(\mathrm{aRR}_c - 1)/\mathrm{aRR}_c$, *respectively, if conditional partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E|C$).

Note that conditional partial exchangeability under no exposure (i.e., $Y^0 \perp\!\!\!\perp E|C$) is equivalent to $\mathrm{SMR}_c = \mathrm{aRR}_c$ (Lemma 1), which is obviously expected to hold in an observational study with no unmeasured confounding. In Table 1, we present Theorems 1.1 and 1.2, and Corollaries 1.1 and 1.2 in more detail, and we provide their proofs in Supplementary Appendix B. Note that we obtain each formula for the attributable fraction (exposed) in Table 1 by dividing the corresponding formula for the attributable fraction (population) by $\Pr(E|Y)$, which is the exposure prevalence among cases. Although this relation is obvious in the Miettinen formula (i.e., the first formula in Corollary 1.1), it is unclear in the Levin formula (i.e., the second formula in Corollary 1.1).

Generally, the attributable fraction is useful when the exposure of interest has a causal effect on the outcome. However, it is important to understand that this statement applies to the exposed group; that is, the

Table 1: Attributable fractions under a binary exposure and a binary outcome[a]

	Attributable fraction (population)[b]	**Attributable fraction (exposed)**[c]
Definition 1	$\frac{\Pr(Y)-\Pr(Y^0)}{\Pr(Y)}$	$\frac{\Pr(Y\|E)-\Pr(Y^0\|E)}{\Pr(Y\|E)}$
Without stratification by C		
Theorem 1.1	$\Pr(E\|Y)\frac{\mathrm{SMR}-1}{\mathrm{SMR}} = \frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)\left(\mathrm{SMR}-\frac{\mathrm{SMR}}{\mathrm{aRR}}\right)+\frac{\mathrm{SMR}}{\mathrm{aRR}}}$	$\frac{\mathrm{SMR}-1}{\mathrm{SMR}}$
Corollary 1.1 (if and only if $Y^0 \perp\!\!\!\perp E$)[d]	$\Pr(E\|Y)\frac{\mathrm{aRR}-1}{\mathrm{aRR}} = \frac{\Pr(E)(\mathrm{aRR}-1)}{\Pr(E)(\mathrm{aRR}-1)+1}$	$\frac{\mathrm{aRR}-1}{\mathrm{aRR}}$
With stratification by C		
Theorem 1.2	$\Pr(E\|Y)\sum_c \Pr(c\|Y,\ E)\frac{\mathrm{SMR}_c-1}{\mathrm{SMR}_c} = \sum_c \Pr(c\|Y)\Pr(E\|Y,\ c)\frac{\mathrm{SMR}_c-1}{\mathrm{SMR}_c} = \sum_c \Pr(c\|Y)\frac{\Pr(E\|c)(\mathrm{SMR}_c-1)}{\Pr(E\|c)\left(\mathrm{SMR}_c-\frac{\mathrm{SMR}_c}{\mathrm{aRR}_c}\right)+\frac{\mathrm{SMR}_c}{\mathrm{aRR}_c}}$	$\sum_c \Pr(c\|Y,\ E)\frac{\mathrm{SMR}_c-1}{\mathrm{SMR}_c}$
Corollary 1.2 (if $Y^0 \perp\!\!\!\perp E\|C$)[e]	$\Pr(E\|Y)\sum_c \Pr(c\|Y,\ E)\frac{\mathrm{aRR}_c-1}{\mathrm{aRR}_c} = \sum_c \Pr(c\|Y)\Pr(E\|Y,\ c)\frac{\mathrm{aRR}_c-1}{\mathrm{aRR}_c} = \sum_c \Pr(c\|Y)\frac{\Pr(E\|c)(\mathrm{aRR}_c-1)}{\Pr(E\|c)(\mathrm{aRR}_c-1)+1}$	$\sum_c \Pr(c\|Y,\ E)\frac{\mathrm{aRR}_c-1}{\mathrm{aRR}_c}$

Abbreviations: aRR, associational risk ratio; SMR, standardized mortality ratio.

[a]We let E, Y, and C denote a binary exposure, a binary outcome, and a set of covariates, respectively. See text and Table S1 for details. When the measures of interest are defined in stratum c, we add the subscript c to the corresponding measures. Although the attributable fraction (unexposed) may be algebraically defined, it becomes 0 using partial consistency under no exposure (i.e., $Y^0 = Y$ if $E = 0$).

[b]The definition provided corresponds to the attributable caseload (population) proposed by Suzuki et al. [13].

[c]The definition provided corresponds to the attributable caseload (exposed) proposed by Suzuki et al. [13]. Note that we obtain each formula by dividing the corresponding formula for the attributable fraction (population) by $\Pr(E|Y)$.

[d]The assumption of $Y^0 \perp\!\!\!\perp E$ (or equivalently, SMR = aRR) is a necessary and sufficient condition for Corollary 1.1 to yield the attributable fraction. The first and second formulae for the attributable fraction (population) are often called the Miettinen formula and Levin formula, respectively, which are equivalent irrespective of whether $Y^0 \perp\!\!\!\perp E$ holds. Although the conditions may differ, the formula in Corollary 1.1 for the attributable fraction (exposed) is identical to the formula in Corollary 4.1 for the attributed fraction (unexposed) in Table S4, as well as the formulae in Corollary 5.1 for the causal excess fractions in Table S5.

[e]The assumption of $Y^0 \perp\!\!\!\perp E|C$ (or equivalently, $\mathrm{SMR}_c = \mathrm{aRR}_c$) is a sufficient but not necessary condition for Corollary 1.2 to yield the attributable fraction. If the homogeneity of aRR_c across strata of C additionally holds, the homogeneous aRR_c is equal to SMR, and the formulae in Corollary 1.2 are reduced to the formulae in Theorem 1.1. Because the homogeneous aRR_c is not necessarily equal to aRR, the formulae in Corollary 1.2 are not reduced to the formulae in Corollary 1.1, unless $Y^0 \perp\!\!\!\perp E$ holds. Under the same condition, the formula in Corollary 1.2 for the attributable fraction (exposed) is identical to the formula in Corollary 5.2 for the causal excess fraction (exposed) in Table S5.

attributable fraction is useful or straightforwardly interpretable when SMR > 1, in which case the range of the measure becomes (0, 1). Conversely, the attributable fraction is less interpretable when it is a negative value. This is especially relevant for improving the understanding of the attributable fraction (population). Although the target population of this measure is the total population, the causal effect should be present not in the total population but in the exposed group. Thus, for example, even if exposure has a causal effect in the total population (i.e., cRR > 1), the attributable fraction (population) is less useful if exposure has a preventive effect in the exposed group (i.e., SMR < 1). We address this point further in Section 9.

3.2 Notes on the "partially adjusted" method of the Levin formula

When using the Levin formula, calculating the attributable fraction (population) with adjusted RR (or SMR) may be the most common error [6]. This was previously called the "partially adjusted" method [10,14]:

$$\frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)(\mathrm{SMR}-1)+1} \quad \text{["Partially adjusted" Levin formula].}$$

If conditional partial exchangeability under no exposure holds (i.e., $Y^0 \perp\!\!\!\perp E|C$), SMR can be calculated as $\Pr(Y|E)/\sum_c \Pr(Y|\bar{E}, c)\Pr(c|E)$ (refer the proof of Lemma 2). However, as has been emphasized [4], the "partially adjusted" Levin formula does not yield the correct attributable fraction (population), whereas the Miettinen formula does not suffer from this problem. This point can be clearly understood by comparing the formulae for the attributable fraction (population) in Theorem 1.1 and Corollary 1.1 in Table 1. Although the first formula in Theorem 1.1 can be readily obtained by substituting SMR with aRR in the Miettinen formula, the second formula in Theorem 1.1 cannot be obtained by substituting SMR with aRR in the Levin formula. On a related issue, it is important to note that $Y^0 \perp\!\!\!\perp E$ (or equivalently, SMR = aRR) is a necessary and sufficient condition for both the Miettinen formula and Levin formula to yield the attributable fraction (population), whereas $Y^0 \perp\!\!\!\perp E$ is a sufficient *but not necessary* condition for the "partially adjusted" Levin formula to yield the attributable fraction (population). Note also that $Y^0 \perp\!\!\!\perp E|C$ is neither a necessary nor a sufficient condition for the "partially adjusted" Levin formula to yield the correct attributable fraction (population). See the proof of Theorem 1.1 in Supplementary Appendix B for details. A possible reason for the misuse of the Levin formula is that information about $\Pr(E|Y)$ in the target population is relatively less available than information about $\Pr(E)$. Without information about $\Pr(E|Y)$, the first formula in Theorem 1.1 cannot be used, and some may consider using the "partially adjusted" Levin formula because of the lack of a better alternative [4]. However, even if information about $\Pr(E|Y)$ is not available, the second formula in Theorem 1.1, which is similar to the Levin formula, can be used to obtain a true attributable fraction (population) once information about $\Pr(E)$, aRR, and SMR is available.

If $Y^0 \perp\!\!\!\perp E$ does not hold (i.e., SMR ≠ aRR) and SMR > 1, the following inequalities hold between the Levin formula, second formula in Theorem 1.1 in Table 1, and "partially adjusted" Levin formula (refer Supplementary Appendix C for the proof):

$$\frac{\Pr(E)(\mathrm{aRR}-1)}{\Pr(E)(\mathrm{aRR}-1)+1} > \frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)\left(\mathrm{SMR}-\frac{\mathrm{SMR}}{\mathrm{aRR}}\right)+\frac{\mathrm{SMR}}{\mathrm{aRR}}} > \frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)(\mathrm{SMR}-1)+1} \text{ if } (\mathrm{SMR}>1)\wedge(\mathrm{SMR}<\mathrm{aRR})$$

and

$$\frac{\Pr(E)(\mathrm{aRR}-1)}{\Pr(E)(\mathrm{aRR}-1)+1} < \frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)\left(\mathrm{SMR}-\frac{\mathrm{SMR}}{\mathrm{aRR}}\right)+\frac{\mathrm{SMR}}{\mathrm{aRR}}} < \frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)(\mathrm{SMR}-1)+1} \text{ if } (\mathrm{SMR}>1)\wedge(\mathrm{SMR}>\mathrm{aRR}).$$

Recall that, in this scenario, only the second formula in Theorem 1.1 yields the correct attributable fraction (population). Thus, the "partially adjusted" formula underestimates the true attributable fraction (population) if SMR < aRR, which is equivalent to $\Pr(Y^0|E) > \Pr(Y|\bar{E}) = \Pr(Y^0|\bar{E})$ using partial consistency under no exposure. Conversely, the "partially adjusted" formula overestimates the true attributable fraction (population) if SMR > aRR, which is equivalent to $\Pr(Y^0|E) < \Pr(Y|\bar{E}) = \Pr(Y^0|\bar{E})$. Thus, the direction of the inherent "bias" in the "partially adjusted" formula is determined by the sum of proportions of the "doomed"

response type (i.e., $(Y^1, Y^0) = (1, 1)$) and "preventive" response type (i.e., $(Y^1, Y^0) = (0, 1)$) in the exposed and unexposed groups. In the two inequalities above, the correct formula lies between the Levin formula and "partially adjusted" formula. Thus, compared with the correct formula, the "partially adjusted" formula over-adjusts the Levin formula.

3.3 Further remarks on the Levin formula – when is the formula valid?

Regarding the calculation of the attributable fraction (population), in some previous studies, researchers have argued that the Levin formula is valid only in the absence of confounding or effect modification [11,12,15,18]. Our study shows that this argument requires clarification.

First, we consider a scenario in which partial exchangeability under no exposure holds (i.e., $Y^0 \perp\!\!\!\perp E$). In this case, as shown in Corollary 1.1, the Levin formula yields the correct attributable fraction (population). Next we consider a scenario in which conditional partial exchangeability under no exposure holds (i.e., $Y^0 \perp\!\!\!\perp E|C$). As mentioned above, the Levin formula cannot yield the correct attributable fraction (population) in this case, and one may use the formulae in Corollary 1.2 in Table 1 instead, the third of which is a stratification method of the Levin formula. In this regard, it is worth mentioning that, if the homogeneity of aRR_c across strata of C additionally holds in Corollary 1.2, the homogeneous aRR_c is equal to SMR (Lemma 2). If this is the case, the formulae in Corollary 1.2 are reduced to the formulae in Theorem 1.1. Because the homogeneous aRR_c is not necessarily equal to aRR, the formulae in Corollary 1.2 are not reduced to the formulae in Corollary 1.1, unless $Y^0 \perp\!\!\!\perp E$ holds. Note that $Y^0 \perp\!\!\!\perp E|C$ is neither stronger nor weaker than $Y^0 \perp\!\!\!\perp E$.

To summarize, even if both $Y^0 \perp\!\!\!\perp E|C$ and the homogeneity of aRR_c across strata of C hold, the Levin formula cannot be used to obtain the correct attributable fraction (population). As shown in Corollary 1.1, the Levin formula is valid if and only if $Y^0 \perp\!\!\!\perp E$ holds.

3.4 Importance of the definition

In previous studies, researchers have focused on how to calculate attributable fractions, particularly paying attention to the Levin formula and Miettinen formula. However, as shown in Table 1, it is important to understand the definition of the attributable fraction first, and then derive relevant theorems and corollaries accordingly. Some confusion about attributable fractions may originate from the lack of an overview, such as that shown in Table 1. As another source of confusion, some researchers have used the terms "attributable fraction" and "excess fraction" interchangeably [20,24]. However, as has been emphasized [13], it is important to clearly distinguish them in the counterfactual framework, which we discuss further later.

Finally, it cannot be emphasized too much that the attributable fraction is distinct from the etiologic fraction [5,13,25], which becomes clearer by considering the link between the potential outcome model and the sufficient cause model [13]. The etiologic fraction has been broadly defined as the fraction of cases that were "caused" by exposure, and a lower bound can be calculated by the attributable fraction if it yields a positive value. See previous studies for further discussions on the link between the two models [26–33].

In Section 4, we outline hypothetical examples to illustrate some formulae shown in Table 1 and then discuss other related measures.

4 Hypothetical examples

Following Flegal [14], we use hypothetical examples of being overweight and mortality. In these examples, we let E and Y denote overweight (1 = overweight, 0 = normal weight) and mortality (1 = deceased, 0 = not

Table 2: Hypothetical examples of being overweight and mortality[a]

	Total population		Nonsmokers ($C = 0$)		Smokers ($C = 1$)	
	Overweight ($E = 1$)	**Normal weight ($E = 0$)**	**Overweight ($E = 1$)**	**Normal weight ($E = 0$)**	**Overweight ($E = 1$)**	**Normal weight ($E = 0$)**
Study 1[b]						
Death ($Y = 1$)	72	90	54	30	18	60
Survival ($Y = 0$)	428	410	396	270	32	140
Total	500	500	450	300	50	200
$\Pr(Y = 1 \mid E = e, C = c)$	0.144	0.18	0.12	0.1	0.36	0.3
aRR	0.8		1.2		1.2	
$\Pr(Y^1) = \sum_c \Pr(Y \mid E, c)\Pr(c) = 0.18$						
$\Pr(Y^0) = \sum_c \Pr(Y \mid \bar{E}, c)\Pr(c) = 0.15$						
$\text{cRR} = \Pr(Y^1)/\Pr(Y^0) = 1.2$						
$\text{SMR} = \Pr(Y \mid E)/\sum_c \Pr(Y \mid \bar{E}, c)\Pr(c \mid E) = 1.2$						
Attributable fraction (population) $= \{\Pr(Y) - \Pr(Y^0)\}/\Pr(Y) = (0.162 - 0.15)/0.162 \approx 0.074$						
Study 2[c]						
Death ($Y = 1$)	81	90	54	30	27	60
Survival ($Y = 0$)	419	410	396	270	23	140
Total	500	500	450	300	50	200
$\Pr(Y = 1 \mid E = e, C = c)$	0.162	0.18	0.12	0.1	0.54	0.3
aRR	0.9		1.2		1.8	
$\Pr(Y^1) = \sum_c \Pr(Y \mid E, c)\Pr(c) = 0.225$						
$\Pr(Y^0) = \sum_c \Pr(Y \mid \bar{E}, c)\Pr(c) = 0.15$						
$\text{cRR} = \Pr(Y^1)/\Pr(Y^0) = 1.5$						
$\text{SMR} = \Pr(Y \mid E)/\sum_c \Pr(Y \mid \bar{E}, c)\Pr(c \mid E) = 1.35$						
Attributable fraction (population) $= \{\Pr(Y) - \Pr(Y^0)\}/\Pr(Y) = (0.171 - 0.15)/0.171 \approx 0.12$						

Abbreviations: aRR, associational risk ratio; cRR, causal risk ratio; SMR, standardized mortality ratio.

[a]We assume $Y^e \perp\!\!\!\perp E|C$ ($e = 0, 1$) in these two studies.

[b]This hypothetical study is derived from Table 1 in the paper by Flegal [14], in which the values of aRR_c are homogeneous across strata of C. Note that $(Y^0 \perp\!\!\!\perp E|C) \wedge (\text{aRR}_c = k \text{ for } \forall c)$ is a sufficient but not necessary condition for SMR $= k$. The attributed fraction (population) and causal excess fraction (population) are $(0.18 - 0.162)/0.18 = 0.1$ and $(0.18 - 0.15)/0.18 \approx 0.17$, respectively, which, by definition, differ from the attributable fraction (population).

[c]This hypothetical study is slightly modified from Study 1 so that the values of aRR_c are non-homogeneous across strata of C. The attributed fraction (population) and causal excess fraction (population) are $(0.225 - 0.171)/0.225 = 0.24$ and $(0.225 - 0.15)/0.225 \approx 0.33$, respectively, which, by definition, differ from the attributable fraction (population).

deceased), respectively. We also consider smoking status C (1 = smoker, 0 = nonsmoker) as a confounder of the effect of E on Y and assume that conditional exchangeability holds (i.e., $Y^e \perp\!\!\!\perp E|C$ ($e = 0, 1$)).

In the upper part of Table 2, we show data for Study 1, which we derived from Example 1 in the paper by Flegal [14]. Although being overweight has a causal effect on mortality (cRR = 1.2), a negative association is observed (aRR = 0.8) because of the presence of confounding by the smoking status. Note that the aRR_c values are homogeneous across the smoking status ($\text{aRR}_{c=1} = \text{aRR}_{c=0} = 1.2$), which is equivalent to SMR (Lemma 2). The true attributable fraction (population) in Study 1 is approximately 0.074. When aRR in the Miettinen formula is substituted with SMR, it becomes the first formula in Theorem 1.1, which, by definition, correctly yields the attributable fraction (population) as

$$\Pr(E|Y)\frac{\text{SMR} - 1}{\text{SMR}} = \frac{72}{162} \times \frac{1.2 - 1}{1.2} \approx 0.074. \quad (1)$$

Particularly when information about $\Pr(E|Y)$ is not available, some may attempt to use the "partially adjusted" Levin formula as follows:

$$\frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)(\mathrm{SMR}-1)+1}=\frac{(500/1000)\times(1.2-1)}{(500/1000)\times(1.2-1)+1}\approx 0.091, \tag{2}$$

which is slightly higher than the true attributable fraction (population). This is expected because SMR (i.e., 1.2) is higher than aRR (i.e., 0.8). However, even without information about $\Pr(E|Y)$, the true attributable fraction (population) can be obtained using the second formula in Theorem 1.1 as follows:

$$\frac{\Pr(E)(\mathrm{SMR}-1)}{\Pr(E)\left(\mathrm{SMR}-\frac{\mathrm{SMR}}{\mathrm{aRR}}\right)+\frac{\mathrm{SMR}}{\mathrm{aRR}}}=\frac{(500/1000)\times(1.2-1)}{(500/1000)\times(1.2-1.2/0.8)+1.2/0.8}\approx 0.074. \tag{3}$$

When using equations (1)–(3), information about SMR is necessary. However, because $Y^0 \perp\!\!\!\perp E|C$ holds, the formulae in Corollary 1.2 can also be used to directly obtain the true attributable fraction (population) as follows:

$$\Pr(E|Y)\sum_c \Pr(c|Y,E)\frac{\mathrm{aRR}_c-1}{\mathrm{aRR}_c}=\frac{72}{162}\times\left(\frac{54}{72}\times\frac{1.2-1}{1.2}+\frac{18}{72}\times\frac{1.2-1}{1.2}\right)\approx 0.074. \tag{4}$$

For comparison, when we use the Levin formula, we obtain

$$\frac{\Pr(E)(\mathrm{aRR}-1)}{\Pr(E)(\mathrm{aRR}-1)+1}=\frac{(500/1000)\times(0.8-1)}{(500/1000)\times(0.8-1)+1}\approx -0.11. \tag{5}$$

Similarly, when we use the Miettinen formula, we obtain

$$\Pr(E|Y)\frac{\mathrm{aRR}-1}{\mathrm{aRR}}=\frac{72}{162}\times\frac{0.8-1}{0.8}\approx -0.11, \tag{6}$$

which is equivalent to the Levin formula. Note that the relation shown in the second inequality in Section 3.2 holds in Study 1 because SMR = 1.2 and aRR = 0.8.

In the lower part of Table 2, we slightly modify Study 1 to make aRR_c non-homogeneous across the smoking status. The true attributable fraction (population) in Study 2 is approximately 0.12. Because the values of aRR_c are non-homogeneous, we need to calculate SMR (i.e., 1.35) before using equations (1)–(3). Equations (1) and (3) both yield the true attributable fraction (population), but equation (2) yields approximately 0.15, which is, as expected, higher than the true attributable fraction (population). However, equation (4) directly yields the true attributable fraction (population) without calculating SMR, as follows:

$$\Pr(E|Y)\sum_c \Pr(c|Y,E)\frac{\mathrm{aRR}_c-1}{\mathrm{aRR}_c}=\frac{81}{171}\times\left(\frac{54}{81}\times\frac{1.2-1}{1.2}+\frac{27}{81}\times\frac{1.8-1}{1.8}\right)\approx 0.12.$$

For comparison, when we use the Levin formula (equation (5)) or Miettinen formula (equation (6)), we obtain approximately −0.053. Thus, the relation shown in the second inequality in Section 3.2 again holds in Study 2 because SMR = 1.35 and aRR = 0.9.

To summarize, when the values of aRR_c are homogeneous across strata of C as in Study 1, one does not have to calculate SMR, which is equal to the homogeneous aRR_c, and may simply proceed to use the formulae in Theorem 1.1 to calculate the attributable fraction. However, even when the homogeneity of aRR_c is not met as in Study 2, the approach is essentially the same, and one may choose the most appropriate formula on each occasion, partly depending on the availability of data.

5 Preventable fraction and prevented fraction

When the exposure of interest has a preventive effect on the outcome, the following two measures are used in the literature: preventable fraction and prevented fraction [34,35]. In this section, we briefly discuss these measures in the counterfactual framework.

The preventable fraction (population) is used when one is interested in the reduction in incidence that would be achieved if the population had been entirely exposed compared with its "current" or observed exposure pattern; that is, the preventable fraction compares the observed risk with the counterfactual risk under $e = 1$.

Definition 2: When the target population is the total population, the preventable fraction (population) is defined as $\{\Pr(Y) - \Pr(Y^1)\}/\Pr(Y)$, whereas when the target population is the unexposed group, the preventable fraction (unexposed) is defined as $\{\Pr(Y|\bar{E}) - \Pr(Y^1|\bar{E})\}/\Pr(Y|\bar{E})$.

Note that the preventable fraction is equivalent to the attributable fraction when the coding of exposure is reversed. Thus, a discussion analogous to that concerning the attributable fraction applies to the preventable fraction. If one does not use stratification by C, Theorem 2.1 and Corollary 2.1 provide the preventable fraction.

Theorem 2.1: *The preventable fraction* (*population*) *and preventable fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y)(1 - \mathrm{SRR_U})$ *and* $1 - \mathrm{SRR_U}$, *respectively*.

Corollary 2.1: *The preventable fraction* (*population*) *and preventable fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y)(1 - \mathrm{aRR})$ *and* $1 - \mathrm{aRR}$, *respectively, if and only if partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E$).

By contrast, if one uses stratification by C, Theorem 2.2 and Corollary 2.2 provide the preventable fraction.

Theorem 2.2: *If one uses stratification by C, the preventable fraction* (*population*) *and preventable fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y)\sum_c \Pr(c|Y, \bar{E})(1 - \mathrm{SRR_{Uc}})$ *and* $\sum_c \Pr(c|Y, \bar{E})(1 - \mathrm{SRR_{Uc}})$, respectively.

Corollary 2.2: *The preventable fraction* (*population*) *and preventable fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y)\sum_c \Pr(c|Y, \bar{E})(1 - \mathrm{aRR}_c)$ and $\sum_c \Pr(c|Y, \bar{E})(1 - \mathrm{aRR}_c)$, *respectively, if conditional partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E|C$).

In Table S2, we present Theorems 2.1 and 2.2, and Corollaries 2.1 and 2.2 in more detail, and we provide their proofs in Supplementary Appendix B. Note that we obtain each formula for the preventable fraction (unexposed) in Table S2 by dividing the corresponding formula for the preventable fraction (population) by $\Pr(\bar{E}|Y)$.

Moreover, the prevented fraction (population) is used when one is interested in the reduction in incidence that has been achieved with the "current" or observed exposure compared with the scenario in which the population had been entirely unexposed; that is, the prevented fraction compares the counterfactual risk under $e = 0$ with the observed risk, where the reference is the counterfactual risk.

Definition 3: When the target population is the total population, the prevented fraction (population) is defined as $\{\Pr(Y^0) - \Pr(Y)\}/\Pr(Y^0)$, whereas when the target population is the exposed group, the prevented fraction (exposed) is defined as $\{\Pr(Y^0|E) - \Pr(Y|E)\}/\Pr(Y^0|E)$.

Compared with the attributable fraction, the roles of $\Pr(Y)$ (or $\Pr(Y|E)$) and $\Pr(Y^0)$ (or $\Pr(Y^0|E)$) are reversed in the prevented fraction. If one does not use stratification by C, Theorem 3.1 and Corollary 3.1 provide the prevented fraction.

Theorem 3.1: *The prevented fraction* (*population*) *and prevented fraction* (*exposed*) *are obtained as* $\Pr(E|Y^0)(1 - \mathrm{SMR})$ *and* $1 - \mathrm{SMR}$, *respectively*.

Corollary 3.1: *The prevented fraction* (*population*) *and prevented fraction* (*exposed*) *are obtained as* $\Pr(E|Y^0)(1 - \text{aRR})$ *and* $1 - \text{aRR}$, *respectively, if and only if partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E$).

Note that if $Y^0 \perp\!\!\!\perp E$ holds, the prevented fraction (population) can be written as $\Pr(E)(1 - \text{aRR})$, but not *vice versa*, which is identified from observed data. By contrast, if one uses stratification by C, Theorem 3.2 and Corollary 3.2 provide the prevented fraction.

Theorem 3.2: *If one uses stratification by C, the prevented fraction* (*population*) *and prevented fraction* (*exposed*) *are obtained as* $\Pr(E|Y^0)\sum_c \Pr(c|Y^0, E)(1 - \text{SMR}_c)$ *and* $\sum_c \Pr(c|Y^0, E)(1 - \text{SMR}_c)$, *respectively*.

Corollary 3.2: *The prevented fraction* (*population*) *and prevented fraction* (*exposed*) *are obtained as* $\Pr(E|Y^0)\sum_c \Pr(c|Y^0, E)(1 - \text{aRR}_c)$ *and* $\sum_c \Pr(c|Y^0, E)(1 - \text{aRR}_c)$, *respectively, if conditional partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E|C$).

In Table S3, we present Theorems 3.1 and 3.2, and Corollaries 3.1 and 3.2 in more detail, including identifiable estimands, and we provide their proofs in Supplementary Appendix B. Note that we obtain each formula for the prevented fraction (exposed) in Table S3 by dividing the corresponding formula for the prevented fraction (population) by $\Pr(E|Y^0)$. Although the formula in Corollary 2.1 for the preventable fraction (unexposed) is identical to the formula in Corollary 3.1 for the prevented fraction (exposed), it is important to understand the difference between their definitions, as well as their differing conditions.

To summarize, both the preventable and prevented fractions are useful when the exposure of interest has a preventive effect on the outcome. However, the preventable fraction is useful when the preventive effect is present in the unexposed group (i.e., $\text{SRR}_\text{U} < 1$), whereas the prevented fraction is useful when the preventive effect is present in the exposed group (i.e., $\text{SMR} < 1$). In these cases, the range of both measures becomes (0, 1]. On a related issue, in line with the approach to distinguish between the attributable caseload and attributable proportion [13], it is also possible to define the preventable and prevented proportions in the counterfactual framework [36], which are both always within the range [0, 1]. For example, the preventable proportion (population) is defined as $\{\Pr(Y) - \Pr(Y^1, Y)\}/\Pr(Y)$, whereas the prevented proportion (population) is defined as $\{\Pr(Y^0) - \Pr(Y, Y^0)\}/\Pr(Y^0)$ [36]. Under the assumption of negative monotonicity of E on Y (i.e., $Y^0 \geq Y^1$ for all individuals), these measures are identifiable [36], and they become equivalent to the preventable fraction (population) and prevented fraction (population) in this section, respectively. To provide the conceptual relations of these measures, we propose a new measure in Section 6.

6 Attributed fraction

The three measures discussed previously all compare the observed risk under the "current" or observed exposure pattern with the counterfactual risk under either $e = 1$ or $e = 0$. Accordingly, a new measure may be defined that compares the observed risk and the counterfactual risk under $e = 1$, where the reference is the counterfactual risk under $e = 1$. Following the relations between the preventable fraction and prevented fraction, we call this newly proposed measure the "attributed fraction."

Definition 4: When the target population is the total population, the attributed fraction (population) is defined as $\{\Pr(Y^1) - \Pr(Y)\}/\Pr(Y^1)$, whereas when the target population is the unexposed group, the attributed fraction (unexposed) is defined as $\{\Pr(Y^1|\bar{E}) - \Pr(Y|\bar{E})\}/\Pr(Y^1|\bar{E})$.

Note that the attributed fraction is equivalent to the prevented fraction when the coding of exposure is reversed. Furthermore, compared with the preventable fraction, the roles of $\Pr(Y)$ (or $\Pr(Y|\bar{E})$) and $\Pr(Y^1)$

(or $\Pr(Y^1|\bar{E})$) are reversed in the attributed fraction. If one does not use stratification by C, Theorem 4.1 and Corollary 4.1 provide the attributed fraction.

Theorem 4.1: *The attributed fraction* (*population*) *and attributed fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y^1)(\mathrm{SRR_U}-1)/\mathrm{SRR_U}$ *and* $(\mathrm{SRR_U}-1)/\mathrm{SRR_U}$, *respectively*.

Corollary 4.1: *The attributed fraction* (*population*) *and attributed fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y^1)(\mathrm{aRR}-1)/\mathrm{aRR}$ *and* $(\mathrm{aRR}-1)/\mathrm{aRR}$, *respectively, if and only if partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E$).

Note that, if $Y^1 \perp\!\!\!\perp E$ holds, the attributed fraction (population) can be written as $\Pr(\bar{E})(\mathrm{aRR}-1)/\mathrm{aRR}$, but not *vice versa*, which is identified from observed data. By contrast, if one uses stratification by C, Theorem 4.2 and Corollary 4.2 provide the attributed fraction.

Theorem 4.2: *If one uses stratification by C, the attributed fraction* (*population*) *and attributed fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y^1)\sum_c \Pr(c|Y^1,\bar{E})(\mathrm{SRR_{U}}_c-1)/\mathrm{SRR_{U}}_c$ *and* $\sum_c \Pr(c|Y^1,\bar{E})(\mathrm{SRR_{U}}_c-1)/\mathrm{SRR_{U}}_c$, *respectively*.

Corollary 4.2: *The attributed fraction* (*population*) *and attributed fraction* (*unexposed*) *are obtained as* $\Pr(\bar{E}|Y^1)\sum_c \Pr(c|Y^1,\bar{E})(\mathrm{aRR}_c-1)/\mathrm{aRR}_c$ *and* $\sum_c \Pr(c|Y^1,\bar{E})(\mathrm{aRR}_c-1)/\mathrm{aRR}_c$, *respectively, if conditional partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E|C$).

In Table S4, we present Theorems 4.1 and 4.2, and Corollaries 4.1 and 4.2 in more detail, including identifiable estimands, and we provide their proofs in Supplementary Appendix B. Note that we obtain each formula for the attributed fraction (unexposed) in Table S4 by dividing the corresponding formula for the attributed fraction (population) by $\Pr(\bar{E}|Y^1)$. Although the formula in Corollary 1.1 for the attributable fraction (exposed) is identical to the formula in Corollary 4.1 for the attributed fraction (unexposed), it is important to understand the difference between their definitions, as well as their differing conditions.

As an illustration, recall the hypothetical examples of being overweight and mortality (Table 2). One may be interested in the reduction in mortality that has been achieved with the "current" or observed distribution of being overweight compared with the scenario in which the entire population had been overweight. In this case, one should estimate the attributed fraction (population) instead of the attributable fraction (population). In Study 1, the attributed fraction (population) is $(0.18-0.162)/0.18 = 0.1$, which, by definition, differs from the attributable fraction (population) (i.e., 0.074). Similarly, the attributed fraction (population) in Study 2 is $(0.225-0.171)/0.225 = 0.24$, which, by definition, differs from the attributable fraction (population) (i.e., 0.12).

Like the attributable fraction, the attributed fraction is useful when the exposure of interest has a causal effect on the outcome. However, unlike the attributable fraction, the attributed fraction is useful when the causal effect is present in the unexposed group (i.e., $\mathrm{SRR_U} > 1$), in which case, the range of the measure becomes (0, 1).

7 Excess fraction: causal and preventive

As mentioned above, there has been some confusion regarding the terms "attributable fraction" and "excess fraction," and they are sometimes used interchangeably [20,24]. To avoid any confusion, these measures should be clearly distinguished in the counterfactual framework [13]. The four measures in Definitions 1–4 compare the observed risk under the "current" or observed exposure pattern with the counterfactual risk under either $e = 1$ or $e = 0$. By contrast, we define the excess fraction as a measure that compares the counterfactual risk under $e = 1$ and the counterfactual risk under $e = 0$. In this study, we

propose distinguishing between the causal excess fraction and preventive excess fraction. First, we define the former.

Definition 5: When the target population is the total population, the causal excess fraction (population) is defined as $\{\Pr(Y^1) - \Pr(Y^0)\}/\Pr(Y^1)$. Similarly, when the target population is the exposed group, the causal excess fraction (exposed) is defined as $\{\Pr(Y^1|E) - \Pr(Y^0|E)\}/\Pr(Y^1|E)$, whereas when the target population is the unexposed group, the causal excess fraction (unexposed) is defined as $\{\Pr(Y^1|\bar{E}) - \Pr(Y^0|\bar{E})\}/\Pr(Y^1|\bar{E})$.

If one does not use stratification by C, Theorem 5.1 and Corollary 5.1 provide the causal excess fraction.

Theorem 5.1: *The causal excess fraction* (*population*) *is obtained as* $(\mathrm{cRR} - 1)/\mathrm{cRR}$. *Similarly, the causal excess fraction* (*exposed*) *and causal excess fraction* (*unexposed*) *are obtained as* $(\mathrm{cRR_E} - 1)/\mathrm{cRR_E}$ *and* $(\mathrm{cRR_U} - 1)/\mathrm{cRR_U}$, *respectively*.

Corollary 5.1: *The causal excess fraction* (*population*) *is obtained as* $(\mathrm{aRR} - 1)/\mathrm{aRR}$ *if exchangeability holds* (i.e., $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$). *The causal excess fraction* (*exposed*) *is obtained as* $(\mathrm{aRR} - 1)/\mathrm{aRR}$ *if and only if partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E$). *The causal excess fraction* (*unexposed*) *is obtained as* $(\mathrm{aRR} - 1)/\mathrm{aRR}$ *if and only if partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E$).

By contrast, if one uses stratification by C, Theorem 5.2 and Corollary 5.2 provide the causal excess fraction.

Theorem 5.2: *If one uses stratification by C, the causal excess fraction* (*population*) *is obtained as* $\sum_c \Pr(c|Y^1)(\mathrm{cRR}_c - 1)/\mathrm{cRR}_c$. *Similarly, the causal excess fraction* (*exposed*) *and causal excess fraction* (*unexposed*) *are obtained as* $\sum_c \Pr(c|Y^1, E)(\mathrm{cRR_{E}}_c - 1)/\mathrm{cRR_{E}}_c$ *and* $\sum_c \Pr(c|Y^1, \bar{E})(\mathrm{cRR_{U}}_c - 1)/\mathrm{cRR_{U}}_c$, *respectively*.

Corollary 5.2: *The causal excess fraction* (*population*) *is obtained as* $\sum_c \Pr(c|Y^1)(\mathrm{aRR}_c - 1)/\mathrm{aRR}_c$ *if conditional exchangeability holds* (i.e., $Y^e \perp\!\!\!\perp E|C$ $(e = 0, 1)$). *The causal excess fraction* (*exposed*) *is obtained as* $\sum_c \Pr(c|Y, E)(\mathrm{aRR}_c - 1)/\mathrm{aRR}_c$ *if conditional partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E|C$). *The causal excess fraction* (*unexposed*) *is obtained as* $\sum_c \Pr(c|Y^1, \bar{E})(\mathrm{aRR}_c - 1)/\mathrm{aRR}_c$ *if conditional partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E|C$).

In Table S5, we present Theorems 5.1 and 5.2, and Corollaries 5.1 and 5.2 in more detail, including identifiable estimands, and we provide their proofs in Supplementary Appendix B. The causal excess fraction (population), causal excess fraction (exposed), and causal excess fraction (unexposed) are generally useful when the causal effect is present in the total population, exposed group, and unexposed group, respectively (i.e., $\mathrm{cRR} > 1$, $\mathrm{cRR_E} > 1$, and $\mathrm{cRR_U} > 1$, respectively). Using partial consistency under exposure (i.e., $Y^1 = Y$ if $E = 1$), the causal excess fraction (exposed) becomes equivalent to the attributable fraction (exposed) in Definition 1, which may partly explain why some researchers have used the terms "attributable fraction" and "excess fraction" interchangeably [20,24]. Similarly, using partial consistency under no exposure (i.e., $Y^0 = Y$ if $E = 0$), the causal excess fraction (unexposed) becomes equivalent to the attributed fraction (unexposed) in Definition 4. However, the causal excess fraction (population) is distinct from both the attributable fraction (population) and attributed fraction (population), even under some assumptions. Regarding this, note that the causal excess fraction (population) is useful when the causal effect is present in the total population (i.e., $\mathrm{cRR} > 1$), whereas the attributable fraction (population) and attributed fraction (population) are useful when the causal effect is present in the exposed group (i.e., $\mathrm{SMR} > 1$) and unexposed group (i.e., $\mathrm{SRR_U} > 1$), respectively; that is, although the target population of these three measures is the total population, the "target" of causation differs between them. We address this point further in Section 9.

As an illustration, the causal excess fraction (population) in Study 1 is $(0.18 - 0.15)/0.18 \approx 0.17$, which, by definition, differs from the attributable fraction (population) (i.e., 0.074) and attributed fraction (population) (i.e., 0.1). Similarly, the causal excess fraction (population) in Study 2 is $(0.225 - 0.15)/0.225 \approx 0.33$,

which, by definition, differs from the attributable fraction (population) (i.e., 0.12) and attributed fraction (population) (i.e., 0.24).

When the exposure of interest has a preventive effect on the outcome, the preventive excess fraction, in which the coding of exposure is reversed, is useful.

Definition 6: When the target population is the total population, the preventive excess fraction (population) is defined as $\{\Pr(Y^0) - \Pr(Y^1)\}/\Pr(Y^0)$. Similarly, when the target population is the exposed group, the preventive excess fraction (exposed) is defined as $\{\Pr(Y^0|E) - \Pr(Y^1|E)\}/\Pr(Y^0|E)$, whereas when the target population is the unexposed group, the preventive excess fraction (unexposed) is defined as $\{\Pr(Y^0|\bar{E}) - \Pr(Y^1|\bar{E})\}/\Pr(Y^0|\bar{E})$.

If one does not use stratification by C, Theorem 6.1 and Corollary 6.1 provide the preventive excess fraction.

Theorem 6.1: *The preventive excess fraction* (*population*) *is obtained as* $1 - \mathrm{cRR}$. *Similarly, the preventive excess fraction* (*exposed*) *and preventive excess fraction* (*unexposed*) *are obtained as* $1 - \mathrm{cRR_E}$ *and* $1 - \mathrm{cRR_U}$, *respectively*.

Corollary 6.1: *The preventive excess fraction* (*population*) *is obtained as* $1 - \mathrm{aRR}$ *if exchangeability holds* (i.e., $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$). *The preventive excess fraction* (*exposed*) *is obtained as* $1 - \mathrm{aRR}$ *if and only if partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E$). *The preventive excess fraction* (*unexposed*) *is obtained as* $1 - \mathrm{aRR}$ *if and only if partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E$).

By contrast, if one uses stratification by C, Theorem 6.2 and Corollary 6.2 provide the preventive excess fraction.

Theorem 6.2: *If one uses stratification by* C, *the preventive excess fraction* (*population*) *is obtained as* $\sum_c \Pr(c|Y^0)(1 - \mathrm{cRR}_c)$. *Similarly, the preventive excess fraction* (*exposed*) *and preventive excess fraction* (*unexposed*) *are obtained as* $\sum_c \Pr(c|Y^0, E)(1 - \mathrm{cRR}_{Ec})$ *and* $\sum_c \Pr(c|Y^0, \bar{E})(1 - \mathrm{cRR}_{Uc})$, *respectively*.

Corollary 6.2: *The preventive excess fraction* (*population*) *is obtained as* $\sum_c \Pr(c|Y^0)(1 - \mathrm{aRR}_c)$ *if conditional exchangeability holds* (i.e., $Y^e \perp\!\!\!\perp E|C$ $(e = 0, 1)$). *The preventive excess fraction* (*exposed*) *is obtained as* $\sum_c \Pr(c|Y^0, E)(1 - \mathrm{aRR}_c)$ *if conditional partial exchangeability under no exposure holds* (i.e., $Y^0 \perp\!\!\!\perp E|C$). *The preventive excess fraction* (*unexposed*) *is obtained as* $\sum_c \Pr(c|Y, \bar{E})(1 - \mathrm{aRR}_c)$ *if conditional partial exchangeability under exposure holds* (i.e., $Y^1 \perp\!\!\!\perp E|C$).

In Table S6, we present Theorems 6.1 and 6.2, and Corollaries 6.1 and 6.2 in more detail, including identifiable estimands, and we provide their proofs in Supplementary Appendix B. The preventive excess fraction (population), preventive excess fraction (exposed), and preventive excess fraction (unexposed) are generally useful when the preventive effect is present in the total population, exposed group, and unexposed group, respectively (i.e., $\mathrm{cRR} < 1$, $\mathrm{cRR_E} < 1$, and $\mathrm{cRR_U} < 1$, respectively). Using partial consistency under exposure (i.e., $Y^1 = Y$ if $E = 1$), the preventive excess fraction (exposed) becomes equivalent to the prevented fraction (exposed) in Definition 3. Similarly, using partial consistency under no exposure (i.e., $Y^0 = Y$ if $E = 0$), the preventive excess fraction (unexposed) becomes equivalent to the preventable fraction (unexposed) in Definition 2. However, the preventive excess fraction (population) is distinct from both the preventable fraction (population) and prevented fraction (population), even under some assumptions.

Finally, it is worth mentioning that, if $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$ holds (e.g., in ideal randomized controlled trials), the differing target populations in either the causal or preventive excess fraction become irrelevant; that is, the formulae in Corollary 5.1 all become identical and so do those in Corollary 6.1. This is in sharp contrast to the measures in Definitions 1–4; even if $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$ holds, the concept of the target

population is important in these measures. Conversely, even if $Y^e \perp\!\!\!\perp E|C$ $(e = 0, 1)$ holds, the differing target populations are relevant in all the measures in Definitions 1–6, unless $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$ holds.

8 Notes on "vaccine efficacy"

In *Modern Epidemiology*, the preventive excess fraction (population) in Definition 6 is called the preventable fraction, which is further explained as follows: "[in] vaccine studies, this measure is also known as the *vaccine efficacy*" [24, p. 83]. By contrast, in *A Dictionary of Epidemiology*, vaccine efficacy is defined as "[the] proportion of persons in the placebo group of a vaccine trial who under ideal conditions would not have become ill if they had received the vaccine" [34, p. 287], which may well correspond to the preventable fraction (unexposed) in Definition 2. This possible confusion may be caused by the fact that, if $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$ holds, both the preventable fraction (unexposed) and preventive excess fraction (population) become equivalent to 1 – aRR (Corollaries 2.1 and 6.1, respectively). Vaccine efficacy is typically measured when a study is conducted under ideal conditions (e.g., clinical trials). However, notably, numerous observational study designs are used to evaluate vaccine efficacy, including cohort studies and test-negative studies [37].

Conventionally, vaccine efficacy is calculated using the formula 1 – aRR, which is equivalent to $\{\Pr(E) - \Pr(E|Y)\}/\{\Pr(E)(1 - \Pr(E|Y))\}$ [38–40]. This so-called screening technique has been used as a method for rapidly estimating vaccine efficacy [40]. However, as elaborated in this article, this calculation method provides vaccine efficacy only when certain conditions hold. When vaccine efficacy is defined as the preventable fraction (unexposed), the calculation method is valid if and only if $Y^1 \perp\!\!\!\perp E$ holds (Corollary 2.1). By contrast, when vaccine efficacy is defined as the preventive excess fraction (population), the calculation method is valid if $Y^e \perp\!\!\!\perp E$ $(e = 0, 1)$ holds (Corollary 6.1). Thus, when vaccine efficacy is estimated in ideal randomized controlled trials, the inconsistent definitions may not make a substantive difference. However, when these conditions are violated, the calculation method is invalid irrespective of the definitions. Once vaccine efficacy is clearly defined, one may proceed to estimate it using the appropriate formulae under certain conditions. Note that the preventable fraction (unexposed) and preventive excess fraction (population) do not become equivalent, even if $Y^e \perp\!\!\!\perp E|C$ $(e = 0, 1)$ holds (Corollaries 2.2 and 6.2, respectively). Given the growing need to assess vaccine efficacy, it is important to have a clearer understanding of the definitions of the relevant measures to avoid confusion.

9 Conclusion

In this study, we discussed a total of six measures, including a newly proposed measure – the attributed fraction – in the counterfactual framework. All the measures could also be considered to be conditional on strata of C. Figure 1 shows a conceptual schema of these six measures when the target population is the total population. In the figure, we consider six (= 3!) possible patterns based on $\Pr(Y)$, $\Pr(Y^0)$, and $\Pr(Y^1)$, assuming that $\Pr(Y)$, $\Pr(Y^0)$, and $\Pr(Y^1)$ all differ. The six arrows show the contrasts of interest in the corresponding six measures. The arrow tails represent the reference points, whereas the arrow heads represent the points to be compared. The height of the gray areas represents the risk of the outcome for each pattern. The difference between $\Pr(Y)$ and $\Pr(Y^0)$ represents the presence of either causal or preventive causation in the exposed group (highlighted in light red), whereas the difference between $\Pr(Y)$ and $\Pr(Y^1)$ represents the presence of either causal or preventive causation in the unexposed group (highlighted in light blue). Note that, although $\Pr(Y^0)$ and $\Pr(Y^1)$ are uniquely determined in the target population, $\Pr(Y)$ may vary according to the exposure distribution patterns, even if the proportion of exposure remains constant.

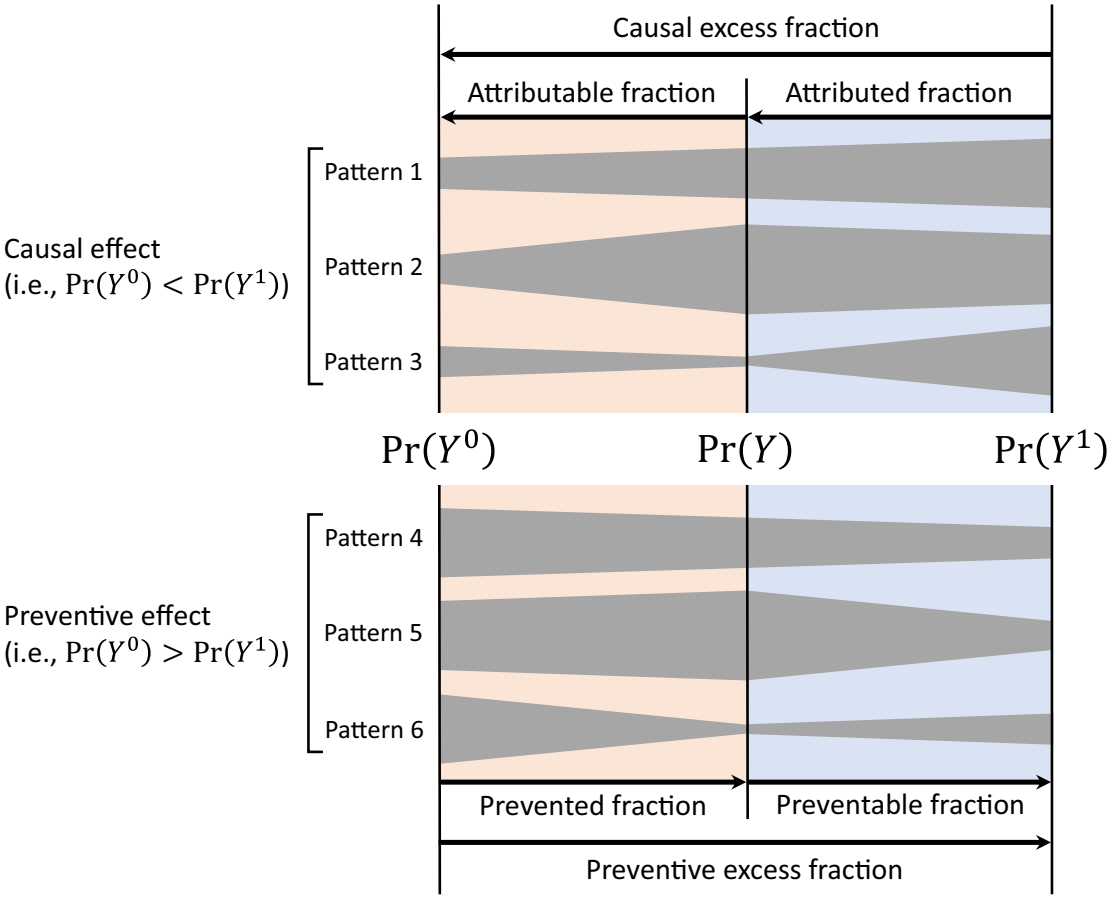

Figure 1: Conceptual schema of the six measures and six possible patterns. We show a schema when the target population is the total population. The six arrows show the contrasts of interest in the corresponding six measures. The arrow tails represent the reference points, whereas the arrow heads represent the points to be compared. The height of the gray areas represents the risk of the outcome for each pattern. We assume that $\Pr(Y)$, $\Pr(Y^0)$, and $\Pr(Y^1)$ all differ. Note that we conveniently describe their relationships in a linear manner. The difference between $\Pr(Y)$ and $\Pr(Y^0)$ represents the presence of either causal or preventive causation in the exposed group (highlighted in light red), whereas the difference between $\Pr(Y)$ and $\Pr(Y^1)$ represents the presence of either causal or preventive causation in the unexposed group (highlighted in light blue). For example, the attributable fraction (population) is a positive value in Patterns 1, 2, and 5 because $\Pr(Y)$ (i.e., the reference point) is higher than $\Pr(Y^0)$ (i.e., the point to be compared). In these patterns, the causal effect is present in the exposed group and SMR is higher than 1. However, in Pattern 5, there is a preventive effect in the total population. An analogous discussion applies to the attributed fraction, preventable fraction, and prevented fraction.

Three aspects are worth mentioning. First, when the coding of exposure is reversed, the attributable fraction, attributed fraction, and causal excess fraction in the upper panel become equivalent to the preventable fraction, prevented fraction, and preventive excess fraction in the lower panel, respectively. Second, the contrasts of interest in the attributable fraction, attributed fraction, and causal excess fraction in the upper panel are the same as those in the prevented fraction, preventable fraction, and preventive excess fraction in the lower panel, respectively, whereas their reference points differ. Third, and related to the second aspect, we should clearly understand the differing "targets" of causation in the six measures. In both the causal and preventive excess fractions, the "target" of causation is the total population, which is the same as the target population. However, the "targets" of causation of the other four measures are not the total population; they are the exposed group in the attributable fraction and prevented fraction, whereas they are the unexposed group in the attributed fraction and preventable fraction. This point may be less relevant in relatively simple scenarios, such as Patterns 1 and 4. Under the assumption of positive monotonicity of E on Y, one may observe only Pattern 1. Similarly, under the assumption of negative monotonicity of E on Y, one may observe only Pattern 4. However, in more complex scenarios,

this has important ramifications. For example, as mentioned above, even if exposure has a causal effect in the total population (i.e., cRR > 1), the attributable fraction (population) is a negative value and is less useful if exposure has a preventive effect in the exposed group (i.e., SMR < 1). This scenario corresponds to Pattern 3, and the prevented fraction would be useful in this case. Furthermore, note that, because $\Pr(Y)$ is lower than $\Pr(Y^0)$ or $\Pr(Y^1)$ in Pattern 3, interventions to set either $E = 0$ or $E = 1$ increase the risk of the adverse outcome Y in the total population. In some specific situations (i.e., $\Pr(Y) = \Pr(Y^1, Y^0)$), one can minimize the risk of the adverse outcome Y by preserving the status quo. Moreover, one may consider implementing interventions on variables other than E with the aim to further lower the risk of Y. It would be important to have such an overview when considering possible interventions. To understand these aspects, we believe that it is important to understand the conceptual relations between the six measures in Figure 1.

Regarding the interpretations of these measures, it is also worth mentioning that information about temporality is not included in their definitions. Some may consider that the attributable and preventable fractions reflect the potential impact in the *future*, whereas the attributed and prevented fractions reflect the potential impact in the *past* [35]. However, we note that ad hoc information about temporality is unwittingly incorporated into this interpretation. For example, in Figure 1, the attributable fraction (population) is generally useful in Patterns 1, 2, and 5 because $\Pr(Y^0)$ is lower than the observed risk $\Pr(Y)$. In this scenario, one may implicitly consider that $\Pr(Y^0)$ represents the risk in the future. By contrast, the attributed fraction (population) is generally useful in Patterns 1, 3, and 6 because $\Pr(Y^1)$ is higher than the observed risk $\Pr(Y)$, and one may implicitly consider that $\Pr(Y^1)$ represents the risk in the past. Note that, in both cases, one innately considers that time flows from the arrow tail to the arrow head in the measure of interest. This conception may seem reasonable because, in the real world, interventions are typically implemented only when they are expected to reduce adverse outcomes in the population. However, the counterfactual risks in the definitions of the measures do not incorporate information about time; rather they are merely counterfactual or contrary to the fact (in the current scenario). Thus, for the interpretation above to hold, the counterfactual risks should be invariant in the future or past. This implicit assumption should be explicitly addressed for the appropriate use of these measures.

To summarize, the six measures are, if correctly used, valuable for health impact assessment and decision-making. However, despite their importance and potential usefulness, the definitions of these measures remain inconsistent in the literature, which has led to some confusion among researchers and health practitioners. It is important to have a clearer understanding of their definitions in the counterfactual framework, which is essential for their appropriate interpretation in the real world.

Acknowledgements: We thank Edanz (https://jp.edanz.com/ac) for editing a draft of this manuscript.

Funding information: E.S. is supported by the Japan Society for the Promotion of Science (JSPS KAKENHI Grant Numbers JP20K10471, JP19KK0418, and JP20K10499). The funding sources had no involvement in the study design; collection, analysis, and interpretation of data; writing of the article; and decision to submit the article for publication.

Author contributions: E.S. conceptualized the authors' views and drafted the manuscript. E.Y. critically revised the manuscript for intellectual content.

Conflict of interest: Authors state no conflict of interest.

Data availability statement: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

[1] Doll R. On the aetiology of cancer of the lung. Acta Unio Int Contra Cancrum. 1951;7(1 Spec. No.):39–50.
[2] Miettinen OS. Proportion of disease caused or prevented by a given exposure, trait or intervention. Am J Epidemiol. 1974;99(5):325–32.
[3] Walter SD. The estimation and interpretation of attributable risk in health research. Biometrics. 1976;32(4):829–49.
[4] Greenland S. Bias in methods for deriving standardized morbidity ratio and attributable fraction estimates. Stat Med. 1984;3(2):131–41.
[5] Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. Am J Epidemiol. 1988;128(6):1185–97.
[6] Rockhill B, Newman B, Weinberg C. Use and misuse of population attributable fractions. Am J Public Health. 1998;88(1):15–9.
[7] Bénichou J. A review of adjusted estimators of attributable risk. Stat Methods Med Res. 2001;10(3):195–216.
[8] Greenland S. Attributable fractions: bias from broad definition of exposure. Epidemiology. 2001;12(5):518–20.
[9] Hanley JA. A heuristic approach to the formulas for population attributable fraction. J Epidemiol Community Health. 2001;55(7):508–14.
[10] Flegal KM, Graubard BI, Williamson DF. Methods of calculating deaths attributable to obesity. Am J Epidemiol. 2004;160(4):331–8.
[11] 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.
[12] Darrow LA, Steenland NK. Confounding and bias in the attributable fraction. Epidemiology. 2011;22(1):53–8.
[13] 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.
[14] Flegal KM. Bias in calculation of attributable fractions using relative risks from nonsmokers only. Epidemiology. 2014;25(6):913–6.
[15] Darrow LA. Commentary: Errors in estimating adjusted attributable fractions. Epidemiology. 2014;25(6):917–8.
[16] Poole C. A history of the population attributable fraction and related measures. Ann Epidemiol. 2015;25(3):147–54.
[17] Greenland S. Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities. Ann Epidemiol. 2015;25(3):155–61.
[18] Khosravi A, Nazemipour M, Shinozaki T, Mansournia MA. Population attributable fraction in textbooks: time to revise. Glob Epidemiol. 2021;3:100062. https://doi.org/10.1016/j.gloepi.2021.100062.
[19] Levin ML. The occurrence of lung cancer in man. Acta Unio Int Contra Cancrum. 1953;9(3):531–41.
[20] Hernán MA, Robins JM. Causal Inference: What If. Boca Raton, FL: Chapman & Hall/CRC, 2020.
[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] Pearl J. Causality: Models, Reasoning, and Inference. (2nd ed.). New York, NY: Cambridge University Press, 2009.
[24] Rothman KJ, VanderWeele TJ, Lash TL. Measures of effect and measures of association. In: Lash TL, VanderWeele TJ, Haneuse S, et al. (Eds.). Modern Epidemiology. (4th ed.). Philadelphia, PA: Wolters Kluwer, 2021. p. 79–103.
[25] VanderWeele TJ. Attributable fractions for sufficient cause interactions. Int J Biostat. 2010;6(2):Article 5. https://doi.org/10.2202/1557-4679.1202.
[26] Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. Scand J Work Environ Health. 1988;14(2):125–9.
[27] Flanders WD. On the relationship of sufficient component cause models with potential outcome (counterfactual) models. Eur J Epidemiol. 2006;21(12):847–53.
[28] VanderWeele TJ, Hernán MA. From counterfactuals to sufficient component causes and vice versa. Eur J Epidemiol. 2006;21(12):855–8.
[29] VanderWeele TJ, Robins JM. Empirical and counterfactual conditions for sufficient cause interactions. Biometrika. 2008;95(1):49–61.
[30] Suzuki E, Yamamoto E, Tsuda T. On the link between sufficient-cause model and potential-outcome model. Epidemiology. 2011;22(1):131–2.
[31] VanderWeele TJ, Richardson TS. General theory for interactions in sufficient cause models with dichotomous exposures. Ann Stat. 2012;40(4):2128–61.
[32] Suzuki E, Yamamoto E. Marginal sufficient component cause model: an emerging causal model with merits? Epidemiology. 2021;32(6):838–45.
[33] Suzuki E, Yamamoto E. *Strength* in causality: discerning causal mechanisms in the sufficient cause model. Eur J Epidemiol. 2021;36(9):899–908.
[34] Porta MS. (Ed.) A Dictionary of Epidemiology. (6th ed.). New York, NY: Oxford University Press, 2014.

[35] Morgenstern H. Attributable fractions. In: Boslaugh S (Ed.). Encyclopedia of Epidemiology. Thousand Oaks, CA: Sage Publications, 2008. p. 55–63.
[36] Yamada K, Kuroki M. Counterfactual-based prevented and preventable proportions. J Causal Inference. 2017;5(2):20160020. https://doi.org/10.1515/jci-2016-0020.
[37] Fox MP, Gower EW. Infectious disease epidemiology. In: Lash TL, VanderWeele TJ, Haneuse S, et al. (Eds.). Modern Epidemiology. (4th ed.). Philadelphia, PA: Wolters Kluwer, 2021. p. 805–43.
[38] Orenstein WA, Bernier RH, Dondero TJ, Hinman AR, Marks JS, Bart KJ, et al. Field evaluation of vaccine efficacy. Bull World Health Organ. 1985;63(6):1055–68.
[39] Orenstein WA, Bernier RH, Hinman AR. Assessing vaccine efficacy in the field: further observations. Epidemiol Rev. 1988;10(1):212–41.
[40] Hatton P. The use of the screening technique as a method of rapidly estimating vaccine efficacy. Public Health. 1990;104(1):21–5.