Errors in causal inference: an organizational schema for systematic error and random error

Etsuji Suzuki MD, PhD[a], Toshihide Tsuda MD, PhD[b], Toshiharu Mitsuhashi MD, PhD[c], Mohammad Ali Mansournia MD, MPH, PhD[d], Eiji Yamamoto PhD[e]

[a] Department of Epidemiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan. Email: etsuji-s@cc.okayama-u.ac.jp

[b] Department of Human Ecology, Graduate School of Environmental and Life Science, Okayama University, Okayama, 3-1-1 Tsushima-naka, Kita-ku, Okayama 700-8530, Japan. Email: tsudatos@md.okayama-u.ac.jp

[c] Center for Innovative Clinical Medicine, Okayama University Hospital, Okayama University, Okayama, 2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan. Email: sankyoh@gmail.com

[d] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, P.O. Box 14155-6446, Tehran, Iran. Email: mansournia_ma@yahoo.com

[e] Department of Information Science, Faculty of Informatics, Okayama University of Science, Okayama, 1-1 Ridai-cho, Kita-ku, Okayama 700-0005, Japan. Email: eiji@mis.ous.ac.jp

Corresponding author: Etsuji Suzuki
2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, Japan
Tel.: +81-86-235-7174, Fax: +81-86-235-7178, E-mail: etsuji-s@cc.okayama-u.ac.jp

ABSTRACT

Purpose: To provide an organizational schema for systematic error and random error in estimating causal measures, aimed at clarifying the concept of errors from the perspective of causal inference.

Methods: We propose to divide systematic error into *structural error* and *analytic error*. With regard to random error, our schema shows its four major sources: non-deterministic counterfactuals, sampling variability, a mechanism that generates exposure events, and measurement variability.

Results: Structural error is defined from the perspective of counterfactual reasoning, and divided into *non-exchangeability bias* (which comprises *confounding bias* and *selection bias*) and *measurement bias*. Directed acyclic graphs are useful to illustrate this kind of error. Non-exchangeability bias implies a lack of "exchangeability" between the selected exposed and unexposed groups. A lack of exchangeability is not a primary concern of measurement bias, justifying its separation from confounding bias and selection bias. Many forms of analytic errors result from the small sample properties of the estimator used and vanish asymptotically. Analytic error also results from wrong (misspecified) statistical models and inappropriate statistical methods.

Conclusions: Our organizational schema is helpful for understanding the relationship between systematic error and random error from a previously less investigated aspect, enabling us to better understand the relationship between accuracy, validity, and precision.

Keywords: bias, causality, epidemiologic methods

List of abbreviations: DAG (directed acyclic graph)

INTRODUCTION

Error is common in science, contrary to some popular belief. No matter how diligently one tries to prevent errors, they do persist in some biomedical research. Epidemiology is no exception. Much epidemiologic research is aimed at obtaining an accurate estimate of the causal effect of exposure on the health outcome in the source population of the study [1], and the quest for it can be viewed as a struggle against errors.

Two broad kinds of errors can occur in studies in the health, life, and social sciences: systematic error and random error [2]. In introductory epidemiology textbooks [3, 4], the distinction between them is usually explained using a schematic illustration of target shooting. Suppose the parameter is the bull's-eye of a target, the estimator is the process of shooting at the target, and the individual bullet holes are estimates. Bias, or systematic error or invalidity, in this example is described as the distance between the average position of the bullet holes and the bull's-eye. This definition of bias has been often explained in statistical dictionaries as: $E(\hat{\theta}) - \theta$, where θ is the parameter of interest and $E(\hat{\theta})$ is the expected value of an estimator $\hat{\theta}$ of the parameter θ [5, 6]. This bias is also referred to as *exact bias* [7]. Epidemiology has borrowed the term "bias" from the statistical literature, using it as a synonym for systematic (non-random) error in the design and conduct of the study [8-10]. Variance, or random error, is described as the degree of dispersion of the bullet holes.

In this research, we provide an organizational schema for errors in estimating causal measures (e.g., causal risk difference, causal risk ratio) in the counterfactual framework [11, 12]. In so doing, we discuss both systematic error and random error from the perspective of causal reasoning. Understanding these errors is of great significance to epidemiologists because it enables them to better understand the relationship between accuracy, validity, and precision when examining cause–effect relationships. In this paper we focus on errors in estimating causal measures, and do not discuss errors relating to descriptions of a particular measurement.

NOTATION

We let A denote an exposure of interest and Y an outcome of interest for an individual. Then, we let Y_a denote the potential outcomes for an individual if, possibly contrary to fact, there had been interventions to set $A = a$. For simplicity, we will generally assume a binary exposure variable (1 = exposed, 0 = unexposed) and a binary outcome variable (1 = outcome occurred, 0 = outcome did not occur), but the remarks here are applicable more generally. Throughout the manuscript we will assume that the consistency assumption is met [13, 14], which implies that the observed outcome for an individual is the potential outcome, as a function of intervention, when the

intervention is set to the actual exposure (i.e., $Y = AY_1 + (1-A)Y_0$). We also assume that there is no interference [15], so that one person's outcome does not depend on the exposure of others. We denote parameters or measures of interest by $\mu(\phi_1, \phi_0)$, which is a function of the parameters ϕ_1 and ϕ_0. When defining causal effects, ϕ_a is a counterfactual parameter for the distribution of potential outcomes Y_a if A had been set to a for all in the target population [16]. In epidemiology, ϕ is often risk when the outcome of interest is binary [17]. Then, causal measures in the target population (e.g., risk difference, risk ratio) can be described as:

$$\mu\left(E(Y_1), E(Y_0)\right), \quad (1)$$

which is the bull's-eye (i.e., θ) in the schematic illustration of target shooting. In the causal inference literature, causal measures have been referred to as *effect measures* or *measures of effect* [18, 19]. Note that these are defined in terms of potential outcomes; they compare what would happen to one population under two possible but distinct life courses or conditions, of which, at most, only one can occur [19].

In Figure 1, we show an organizational schema for errors in estimating the causal measure (i.e., Formula 1). In the following sections, we explain our schema step by step.

--- Place Figure 1 about here ---

SYSTEMATIC ERROR

Many epidemiologists have been interested in obtaining unbiased estimates of the effect of an exposure on the health outcome, and there is a large volume of literature on systematic error or bias [20-25]. Recall that the amount of exact bias is defined by the difference between the expected value of an estimator and the causal measure (i.e., $E(\hat{\theta}) - \theta$), which can be algebraically described as:

$$E\left[\hat{\mu}\left(E(Y_1), E(Y_0)\right)\right] - \mu\left(E(Y_1), E(Y_0)\right). \quad (2)$$

Note that $\mu\left(\hat{E}(Y_1), \hat{E}(Y_0)\right)$ is commonly used as $\hat{\mu}\left(E(Y_1), E(Y_0)\right)$, and an issue of estimators

is addressed in the Analytic error section. In this section, we show how Formula 2 can be decomposed from the perspective of causal reasoning, explaining the upper part of Figure 1. In so doing, we use a traditional tripartite categorization of biases, i.e., confounding, selection, and measurement/information [7, 10, 26]. Recently published literature has reported that these three major sources of bias can be graphically described using directed acyclic graphs (DAGs) [27-34]. As shown in Figure 2, we let *L* denote a common cause of *A* and *Y*. Typically, *L* is called a confounder of the effect of *A* on *Y*. We also let *S* denote a selection variable (1 = selected, 0 = not selected), which is a common effect of *A* (or a cause of *A*) and *Y* (or a cause of *Y*). Conditioning on *S*, shown by square around *S*, yields a spurious association between *A* and *Y* (in at least one strata of *S*, say *S* = 1), which is called selection bias [30]. We also consider a situation in which there is measurement bias due to misclassification of exposure and outcome, and we use asterisks to denote misclassified exposure and outcome (i.e., *A** and *Y**, respectively).

In our schema (Figure 1), systematic error is divided into *analytic error* and *structural error*. To simplify the explanation, we first explain structural error followed by analytic error.

--- Place Figure 2 about here ---

Structural error

Obviously, we are only able to observe one of the two potential outcomes (i.e., Y_1 and Y_0) for an individual, and the measure of effect is not available in actual studies. Recall that a measure of effect compares what would happen to one population under two possible but distinct life courses or conditions. Thus, in actual studies, we use measures that compare what happens in two distinct populations, although the two distinct populations may correspond to one population in different time periods [19]. In the causal inference literature, these have been referred to as *associational measures* or *measures of association* [18, 19]. Using the notation above, they can be algebraically described as:

$$\mu\left(E\left(Y^* \mid A^*=1, S=1\right), E\left(Y^* \mid A^*=0, S=1\right)\right), \quad (3)$$

where $E\left(Y^* \mid A^*=a^*, S=1\right)$ equals $\sum_l E\left(Y^* \mid A^*=a^*, S=1, L=l\right) P\left(L=l \mid A^*=a^*, S=1\right)$.

We here propose to define structural error by comparing a measure of association with a measure of effect as:

$$\mu\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right). \quad (4)$$

Unlike Formula 2, structural error is defined by comparing the two measures, each of which is a function of two population parameters. Note that the newly proposed structural error corresponds to previously known "bias" in epidemiology, including confounding bias, selection bias, and measurement bias. Because the error described in Formula 4 is structural, DAGs are useful to illustrate this kind of error (Figure 2).

We here propose to decompose structural error into *measurement bias* and *non-exchangeability bias* from the perspective of causal reasoning. We first explain non-exchangeability bias followed by measurement bias.

Non-exchangeability bias

The assumption of exchangeability often receives most attention in discussions about causal inference [11, 35-37], and it is commonly explained as statistical independence between the observed exposure A and the potential outcome Y_a (i.e., $A \coprod Y_a$ for $\forall a$) [18]. It is fairly well known that both confounding bias and selection bias imply a lack of exchangeability between the exposed and unexposed groups [30, 36-38]. Here, we propose to define non-exchangeability bias by comparing a measure of association without measurement error and a measure of effect as:

$$\mu\left(E\left(Y\mid A=1,S=1\right),E\left(Y\mid A=0,S=1\right)\right)-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right). \quad (5)$$

The presence of the newly proposed non-exchangeability bias implies a lack of "exchangeability" between the selected exposed and unexposed groups. Note that, even without exchangeability between these two groups, Formula 5 may become 0, depending on the measure used. Thus, it should be emphasized that the presence of non-exchangeability bias is not a necessary but a sufficient condition for a lack of exchangeability [39].

Graphically, non-exchangeability bias implies the presence of an undirected open path (i.e., a biasing path) between the exposure A and the outcome Y, and can be divided into *confounding bias* and *selection bias* based on the causal structures [30]. As mentioned above, confounding bias occurs because of common causes of exposure and outcome, whereas selection bias occurs because of conditioning on common effects of exposure (or a cause of exposure) and outcome (or a cause of outcome) (Figure 2) [30]. Although these two types of bias imply the same phenomenon, i.e., a lack of exchangeability between the exposed and unexposed groups in the study (selected) population, the distinction based on the causal structure is important because

it frequently guides the choice of analytic methods and study design to reduce or avoid the bias [30]. Algebraically, confounding bias and selection bias can be described as:

$$\mu\left(E\left(Y \mid A=1\right), E\left(Y \mid A=0\right)\right) - \mu\left(E\left(Y_1\right), E\left(Y_0\right)\right), \quad (6)$$

and

$$\mu\left(E\left(Y \mid A=1, S=1\right), E\left(Y \mid A=0, S=1\right)\right) - \mu\left(E\left(Y \mid A=1\right), E\left(Y \mid A=0\right)\right), \quad (7)$$

respectively. Note that the definition of confounding bias essentially corresponds to that of confounding *in measure* in the original population [40]. It is also notable that the target parameter in Formula 7 refers to the difference in association measure between the original (unselected) and the study (selected) populations, and thus is not a causal measure. Although it may seem inconsistent with the conventional definition of bias, however, when there is no confounding bias, the second term of Formula 7 is equal to $\mu\left(E\left(Y_1\right), E\left(Y_0\right)\right)$. The decomposition of non-exchangeability bias in Formula 5 to confounding bias in Formula 6 and selection bias in Formula 7 is consistent with the fact that confounding usually precedes selection bias (e.g., case-control sampling, censoring in cohort) in observational studies. In Supplementary Appendix A, we discuss alternative decomposition of non-exchangeability bias. Finally, we note that measurement error in confounders (for simplicity, not assumed in Figure 2) or misspecification of the functional forms of confounders in the model (e.g., error in scale of the continuous confounders) results in residual confounding bias in adjusted estimates. Similar argument applies to selection bias adjustment.

Measurement bias

Measurement error in both the exposure and the outcome is a common problem in epidemiology studies. However, bias due to measurement error has been relatively less addressed in the causal inference literature. Recently, Edwards et al. [41] demonstrated that bias due to measurement error can be incorporated into the counterfactual framework. We here propose to define measurement bias as the difference between structural error in Formula 4 and non-exchangeability bias in Formula 5, which is algebraically described as:

$$\mu\left(E\left(Y^{*} \mid A^{*}=1, S=1\right), E\left(Y^{*} \mid A^{*}=0, S=1\right)\right)-\mu\left(E\left(Y \mid A=1, S=1\right), E\left(Y \mid A=0, S=1\right)\right). \quad (8)$$

Note that, unlike non-exchangeability bias, the definition of measurement bias is based on the comparison of two associational measures; the first term of Formula 8 is calculated using the misclassified exposure and the misclassified outcome among the selected individuals, whereas the second term is calculated using the true exposure and the true outcome among the selected individuals. In other words, the target parameter is not a true causal parameter but a "true" associational measure among the selected individuals. Thus, the newly defined measurement bias may seem inconsistent with the conventional definition of bias. When there is neither confounding bias nor selection bias, however, the second term of Formula 8 is equal to $\mu\left(E\left(Y_1\right), E\left(Y_0\right)\right)$.

The rationale to separate measurement bias from the other two major types of bias (i.e., confounding and selection) may be readily understood by describing measurement bias in DAGs [31-34]. Hernán and Cole [31] showed that DAGs can be used to represent four types of measurement error: independent non-differential, dependent non-differential, independent differential, and dependent differential. To simplify our discussion, Figure 2d shows a situation in which there is an independent non-differential measurement error of the exposure and the outcome, in addition to confounding bias and selection bias. Note that, conditional on *S*, there are three open backdoor paths between *A** and *Y** (i.e., *A**←*A*→*Y*→*Y**, *A**←*A*←*L*→*Y*→*Y**, and *A**←*A*→[*S*]←*Y*→*Y**), and bias due to the causal structure of the third path has been specifically referred to as M-bias [42]. Thus, unlike confounding bias and selection bias, measurement bias is not described as yielding an undirected open path between the true exposure *A* and the true outcome *Y*; rather, measurement bias is explained by the presence of backdoor path(s) between the misclassified exposure *A** and the misclassified outcome *Y**. Therefore, a lack of exchangeability between the truly exposed and truly unexposed groups is not a primary concern of measurement bias, which would justify distinguishing measurement bias from confounding bias and selection bias.

Finally, under the assumption of no confounding bias and no selection bias, independent non-differential measurement error of the exposure and outcome does not result in bias under the causal null hypothesis of no effect of the exposure on the outcome [31]. Although this is a notable exception, measurement error will generally result in bias. Even when there is no effect of *A* on *Y* in Figure 2d, the two open backdoor paths between *A** and *Y** (i.e., *A**←*A*←*L*→*Y*→*Y** and *A**←*A*→[*S*]←*Y*→*Y**) remain and the independent non-differential measurement bias persists.

Analytic error

Comparison of the systematic error (Formula 2) and the newly defined structural error (Formula 4) yields the identification of analytic error. Using Formula 3, we may estimate $\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$ by $\hat{\mu}\left(E\left(Y_1\right),E\left(Y_0\right)\right)=\hat{\mu}\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)$ so that the bias in Formula 2 becomes $E\left[\hat{\mu}\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)\right]-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$. Then, analytic error is algebraically described as:

$$E\left[\hat{\mu}\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)\right]$$
$$-\mu\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right). \quad (9)$$

Note that the target of analytic error is the measure of association (Formula 3), whereas the target of conventional bias is the measure of effect (Formula 1). When there is no structural error, however, the second term of Formula 9 is equal to $\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$. In short, analytic error is interpreted as a component of systematic error that cannot be explained by structural error.

Many forms of analytic errors result from the small sample properties of the estimator used and vanish asymptotically. For example, ratio measures estimates suffer from a positive small-sample bias (of order 1/n); their exact bias based on Formula 9 is infinite, because there is a nonzero probability of having a zero denominator. This bias can occur even with large sample sizes if the data lack adequate numbers of observations for some combination of risk factor and outcome levels, which may arise even if the total sample size appears large and thus the bias is better termed sparse-data bias [43]. Stratification on a vector of confounders thins out cells much faster than risk factor/outcome levels alone. Regression modeling can mitigate the sparse-data bias inherent in stratification by imposing modeling assumptions (e.g., linearity, no interaction between confounders) at the expense of introducing model-misspecification bias. Sparse-data bias can be reduced using penalization methods [43-46].

There are other important and common sources of analytic error such as bias due to wrong (misspecified) statistical models and bias due to inappropriate statistical methods, which can generally occur even if the size of population is large. For example, when the target parameter is the marginal causal effect, standardization and inverse-probability weighted estimates better estimate the marginal causal effect than covariate-conditional odds ratios derived from direct outcome regression even if both odds ratios are unconfounded. This is because the

target parameter of the latter method is not a marginal but a common causal effect [47, 48].

In Table 1 we summarize algebraic descriptions of the decomposition of systematic error.

Table 1. Decomposition of systematic error

Error	**Algebraic description**
Systematic error (2)	$E\left[\hat{\mu}\left(E\left(Y_1\right),E\left(Y_0\right)\right)\right]-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$
Analytic error [a] (9)	$E\left[\hat{\mu}\left(E\left(Y_1\right),E\left(Y_0\right)\right)\right]-\mu\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)$ $=E\left[\hat{\mu}\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)\right]-\mu\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)$
Structural error (4)	$\mu\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$
Measurement bias (8)	$\mu\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)-\mu\left(E\left(Y\mid A=1,S=1\right),E\left(Y\mid A=0,S=1\right)\right)$
Non-exchangeability bias (5)	$\mu\left(E\left(Y\mid A=1,S=1\right),E\left(Y\mid A=0,S=1\right)\right)-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$
Selection bias (7)	$\mu\left(E\left(Y\mid A=1,S=1\right),E\left(Y\mid A=0,S=1\right)\right)-\mu\left(E\left(Y\mid A=1\right),E\left(Y\mid A=0\right)\right)$
Confounding bias (6)	$\mu\left(E\left(Y\mid A=1\right),E\left(Y\mid A=0\right)\right)-\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$

We let *A* denote a binary exposure of interest (1 = exposed, 0 = unexposed) and *Y* denote an outcome of interest. We also let *S* denote a selection variable (1 = selected, 0 = not selected), which is a common effect of *A* and *Y*. We use asterisks to denote misclassified exposure and outcome (i.e., *A** and *Y**, respectively). We denote measures of interest by $\mu\left(\phi_1,\phi_0\right)$, which is a function of the parameters. See text for details.

[a] We show an algebraic definition of analytic error when estimating $\mu\left(E\left(Y_1\right),E\left(Y_0\right)\right)$ by $\hat{\mu}\left(E\left(Y_1\right),E\left(Y_0\right)\right)=\hat{\mu}\left(E\left(Y^*\mid A^*=1,S=1\right),E\left(Y^*\mid A^*=0,S=1\right)\right)$.

RANDOM ERROR

Much epidemiology literature has simply explained random error as an error other than systematic error, and few have provided systematic discussion of random error. Indeed, there is no entry for "random error" in *A Dictionary of Epidemiology* [2], although there is an entry for "systematic error". Hernán [18] discussed stochastic (non-deterministic) counterfactuals and sampling variability as two sources of random error. Note that the former error is at the individual level and the latter error is at the sample level. A non-deterministic definition of counterfactual outcome attaches a statistical distribution of Y_a to each subject. See Hernán [18] and Hernán and Robins [7] for further discussion.

We here propose to add "mechanism that generates exposure events" and "measurement variability" as sources of random error (see the lower part of Figure 1). The former error is at the exposure assignment level in a particular sample, while the latter error occurs at measurement, which is related to issues of inter-rater/intra-rater variability. Note that these errors can occur even when there is no sampling involved. For example, even if one has accounted for measurement bias in a particular sample, there is a remaining variability in measurement, which is referred to as "error due to measurement variability". Our schema does not show type 1 error or type 2 error, which are defined based on statistical decision-making in the face of random error.

To illustrate the relationship between systematic error and random error in general, we explain the relationship between "confounding bias" and "error due to a mechanism that generates exposure events", addressing a distinction between the notions of confounding "in expectation" and "realized" confounding [36, 40, 49]. In an ideal randomized controlled trial, the randomized groups will be comparable in their potential outcomes on average over repeated experiments. For any given experiment, however, the particular randomization may result in imbalances by chance because of the particular allocation or exposure assignment [50]. Such a scenario would be one in which there is no confounding in expectation but there is realized confounding for the particular experiment [40, 49]. (This phenomenon has been also referred to as "random confounding" [51]. Under stochastic potential outcome models for continuously distributed risks, there is always some random confounding in finite samples.) To grasp the profound distinction between these notions of confounding, we need to understand the *mechanism* that generates exposure events, not the product of that mechanism.

In Figure 1, we refer to "confounding bias" by employing the notion of confounding in expectation. See Supplementary Appendix B for its strict algebraic description. When there is confounding in expectation (a form of systematic error), we also expect to observe realized confounding in a study. In this case, realized confounding is a result from "confounding bias" as well as "error due to a mechanism that generates exposure events". However, even if there is no

bias or systematic error, there is a possibility of realized confounding because of "error due to a mechanism that generates exposure events", particularly when the size of the population is small. In other words, irrespective of whether there is "confounding bias" or confounding in expectation, "error due to a mechanism that generates exposure events" is a determinant of the presence of realized confounding. Since error due to a mechanism that generates exposure events is a type of random error, it declines as the size of the population increases, and realized confounding converges to confounding in expectation. On a related issue, it is notable that, in such a situation in which there is no confounding in expectation, one describes the estimator as being an unbiased estimator and the realized value is referred to as an unbiased estimate even if there is realized confounding [2, 6]. This unbiasedness, however, is only "unconditional" (or pre-allocation state) and not helpful for analysis of a given experiment and there would be "conditional" (or post-allocation state) bias in the unadjusted estimator [11, 36, 50, 51].

Analogous discussions apply to the relationship between "selection bias" and "error due to sampling variability" as well as the relationship between "measurement bias" and "error due to measurement variability" in Figure 1.

DISCUSSION

To our knowledge, there is little in the literature providing comprehensive schema of both systematic error and random error. Here we provide an organizational schema for systematic error and random error in estimating causal measures, aimed at clarifying the concept of errors from the perspective of causal inference. We propose to divide systematic error into structural error and analytic error. Structural error comprises three main sources of bias: confounding bias, selection bias, and measurement bias. Although some researchers have provided very extensive lists of specific biases [52, 53], there is concern that these lists are not necessarily comprehensive, and can be confusing because of the inconsistent terminology [22]. Also many, if not most, of these types of biases are structural and can be easily mapped to epidemiologic confounding, measurement bias, and selection bias [54]. Thus, we aim to provide an overall relationship of "bias" by employing the well-known tripartite taxonomy. Although analytic error has not been fully addressed in epidemiology literature, it provides a unifying perspective of epidemiology and statistics. We also detail the four major sources of random error in our schema. This clear explication is useful for understanding the relationship between systematic error and random error from a previously less investigated aspect.

We should note that our schema is not truly complete. For example, we assumed that there is no measurement error of confounders, and we did not consider dependent measurement error between covariates, etc. Also exact mean-unbiasedness considered in this paper is just one of several useful definitions of bias. In practice, validity is sometimes taken to mean a strong

combination of uniform consistency, asymptotic unbiasedness, and normality to ensure that approximate Wald intervals are asymptotically valid. Although these are important issues in causal inference, our schema is helpful for understanding the overall relationship of errors.

The target of inference is important in evaluating the bias in effect estimates. The target can be the total population, or the exposed, marginal or conditional on the covariates. The effects in the total population can be different from the effects in the exposed if there is modification of effect measures across levels of the confounders. Similarly, the marginal causal effects can be different from conditional causal effects simply due to non-collapsibility of effect measures [55]. For example, in a matched cohort study with matching variable as the sole confounder, the crude and matching variable conditional odds ratios are different even though both have causal interpretations for their target (the target of the crude odds ratio is the exposed population if as usual unexposed subjects are matched to the exposed subjects) [39, 56, 57]. This example highlights the point that non-collapsibility should not be confused with bias.

Although the best endeavors are usually made to ensure the accuracy of scientific research, unfortunately there are potential errors in all scientific papers. However, in public health practice, we must take appropriate action to ensure the safety and health of the public before it is too late, while still being aware of potential errors. A balanced approach should be applied to interpretation, and the aim of the research should not be overlooked, such as disease control and prevention. In so doing, it is also important to bear in mind that quantitative bias analyses are now recommended to obtain more valid study results [58]. A systematic understanding of errors is key in the pursuit of causal inference.

Competing interests

None declared.

Funding

This work was supported by Japan Society for the Promotion of Science [KAKENHI grant number JP26870383]. The funder had no role in the writing of the report or the decision to submit the paper for publication.

REFERENCES

1. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health 2001;22:189-212.
2. Porta MS, editor. A Dictionary of Epidemiology. 6th ed. New York, NY: Oxford University Press; 2014.
3. Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring III JR. Medical Epidemiology.

4th ed. Lange Medical Books/McGraw-Hill: New York, NY; 2005.
4. Jekel JF, Katz DF, Elmore JG, Wild DMG. Epidemiology, Biostatistics, and Preventive Medicine. 3rd ed. Saunders/Elsevier: Philadelphia, PA; 2007.
5. Everitt B, Skrondal A. The Cambridge Dictionary of Statistics. 4th ed. Cambridge University Press: Cambridge, UK; 2010.
6. Upton G, Cook I. A Dictionary of Statistics. 2nd ed. Oxford University Press: New York, NY; 2008.
7. Hernán MA, Robins JM. Causal Inference. Chapman & Hall/CRC: Boca Raton, FL; 2016, forthcoming.
8. Newman SC. Biostatistical Methods in Epidemiology. John Wiley & Sons: New York, NY; 2001.
9. Szklo M, Nieto FJ. Epidemiology: Beyond the Basics. 3rd ed. Jones & Bartlett Learning: Burlington, MA; 2012.
10. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Modern Epidemiology, Rothman KJ, Greenland S, Lash TL, editors. Lippincott Williams & Wilkins: Philadelphia, PA, 2008, p. 128-47.
11. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. Int J Epidemiol 1986;15(3):413-9.
12. 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.
13. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology 2009;20(1):3-5.
14. VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology 2009;20(6):880-3.
15. Tchetgen Tchetgen EJ, VanderWeele TJ. On causal inference in the presence of interference. Stat Methods Med Res 2012;21(1):55-75.
16. Flanders WD, Klein M. A general, multivariate definition of causal effects in epidemiology. Epidemiology 2015;26(4):481-9.
17. Cole SR, Hudgens MG, Brookhart MA, Westreich D. Risk. Am J Epidemiol 2015;181(4):246-50.
18. Hernán MA. A definition of causal effect for epidemiological research. J Epidemiol Community Health 2004;58(4):265-71.
19. Greenland S, Rothman KJ, Lash TL. Measures of effect and measures of association. In: Modern Epidemiology, Rothman KJ, Greenland S, Lash TL, editors. Lippincott Williams & Wilkins: Philadelphia, PA, 2008, p. 51-70.
20. Steineck G, Ahlbom A. A definition of bias founded on the concept of the study base.

Epidemiology 1992;3(6):477-82.
21. Maclure M, Schneeweiss S. Causation of bias: the episcope. Epidemiology 2001;12(1):114-22.
22. Weisberg HI. Bias and Causation: Models and Judgment for Valid Comparisons. Wiley: Hoboken, NJ; 2010.
23. Gatto NM, Campbell UB, Schwartz S. An organizational schema for epidemiologic causal effects. Epidemiology 2014;25(1):88-97.
24. Schwartz S, Campbell UB, Gatto NM, Gordon K. Toward a clarification of the taxonomy of "bias" in epidemiology textbooks. Epidemiology 2015;26(2):216-22.
25. Howe CJ, Cain LE, Hogan JW. Are all biases missing data problems? Curr Epidemiol Rep 2015;2(3):162-71.
26. Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol 1981;114(4):593-603.
27. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10(1):37-48.
28. Robins JM. Data, design, and background knowledge in etiologic inference. Epidemiology 2001;12(3):313-20.
29. Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. Am J Epidemiol 2002;155(2):176-84.
30. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology 2004;15(5):615-25.
31. Hernán MA, Cole SR. Invited commentary: causal diagrams and measurement bias. Am J Epidemiol 2009;170(8):959-62.
32. Shahar E. Causal diagrams for encoding and evaluation of information bias. J Eval Clin Pract 2009;15(3):436-40.
33. Shahar E, Shahar DJ. On the causal structure of information bias and confounding bias in randomized trials. J Eval Clin Pract 2009;15(6):1214-6.
34. VanderWeele TJ, Hernán MA. Results on differential and dependent measurement error of the exposure and the outcome using signed directed acyclic graphs. Am J Epidemiol 2012;175(12):1303-10.
35. Hernán MA. Beyond exchangeability: the other conditions for causal inference in medical research. Stat Methods Med Res 2012;21(1):3-5.
36. Greenland S, Robins JM. Identifiability, exchangeability and confounding revisited. Epidemiol Perspect Innov 2009;6:4. doi:10.1186/1742-5573-6-4.
37. Flanders WD, Eldridge RC. Summary of relationships between exchangeability, biasing

paths and bias. Eur J Epidemiol 2015;30(10):1089-99.
38. Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A counterfactual approach to bias and effect modification in terms of response types. BMC Med Res Methodol 2013;13:101. doi:10.1186/1471-2288-13-101.
39. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. Stat Sci 1999;14(1):29-46.
40. VanderWeele TJ. Confounding and effect modification: distribution and measure. Epidemiol Method 2012;1(1):55-82. doi:10.1515/2161-962X.1004.
41. Edwards JK, Cole SR, Westreich D. All your data are always missing: incorporating bias due to measurement error into the potential outcomes framework. Int J Epidemiol 2015;44(4):1452-9.
42. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003;14(3):300-6.
43. Cole SR, Chu H, Greenland S. Maximum likelihood, profile likelihood, and penalized likelihood: a primer. Am J Epidemiol 2014;179(2):252-60.
44. Agresti A. Categorical Data Analysis. John Wiley & Sons: New York, NY; 1990.
45. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Stat Med 2015;34(23):3133-43.
46. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. BMJ 2016;352:i1981.
47. Mansournia MA, Altman DG. Inverse probability weighting. BMJ 2016;352:i189.
48. Mansournia MA, Danaei G, Forouzanfar MH, Mahmoodi M, Jamali M, Mansournia N, et al. Effect of physical activity on functional performance and knee pain in patients with osteoarthritis : analysis with marginal structural models. Epidemiology 2012;23(4):631-40.
49. Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A typology of four notions of confounding in epidemiology. J Epidemiol: in press.
50. Greenland S. Randomization, statistics, and causal inference. Epidemiology 1990;1(6):421-9.
51. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. Eur J Epidemiol 2015;30(10):1101-10.
52. Sackett DL. Bias in analytic research. J Chronic Dis 1979;32(1-2):51-63.
53. Delgado-Rodríguez M, Llorca J. Bias. J Epidemiol Community Health 2004;58(8):635-41.
54. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. Epidemiology: in press.
55. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and

propensity-based weighting under conditions of nonuniform effect. Am J Epidemiol 2006;163(3):262-70.
56. Mansournia MA, Greenland S. The relation of collapsibility and confounding to faithfulness and stability. Epidemiology 2015;26(4):466-72.
57. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. Int J Epidemiol 2013;42(3):860-9.
58. Lash TL, Fox MP, Fink AK. Applying Quantitatvie Bias Analysis to Epidemiologic Data. Springer: New York, NY; 2009.

Figure captions

Fig. 1 An organizational schema for errors in estimating causal measures

The presence of non-exchangeability bias implies a lack of exchangeability between the selected exposed and unexposed groups. Non-exchangeability bias is divided into the two types of bias based on their causal structures; confounding bias occurs because of common causes whereas selection bias occurs because of conditioning on common effects. Note that random errors due to stochastic (non-deterministic) counterfactuals, sampling variability, a mechanism that generates exposure events, and measurement variability are at the individual level, at the sample level, at the exposure assignment level, and at the measurement level, respectively.

Fig. 2 Directed acyclic graphs illustrating structural error

A, *Y*, *L*, and *S* denote exposure, outcome, confounder, and selection variable, respectively, and we use asterisks to denote misclassified variables. The square around *S* in Figure 2c and d indicates that the analysis is restricted to those who do not drop out (i.e., $S = 1$). Figure 2a shows a situation in which one can readily identify a causal relationship between the exposure *A* and the outcome *Y*. Figure 2b illustrates a situation in which there is confounding bias, and Figure 2c illustrates a situation in which there are both confounding bias and selection bias. Figure 2d illustrates a situation in which there is an independent non-differential measurement error of the exposure and the outcome, in addition to confounding bias and selection bias. Hernán and Cole [31] represent on their diagrams factors responsible for measurement error, U_A and U_Y, other than *A* and *Y*. Following a recent approach suggested by VanderWeele and Hernán [34], we do not represent the factors in Figure 2d.

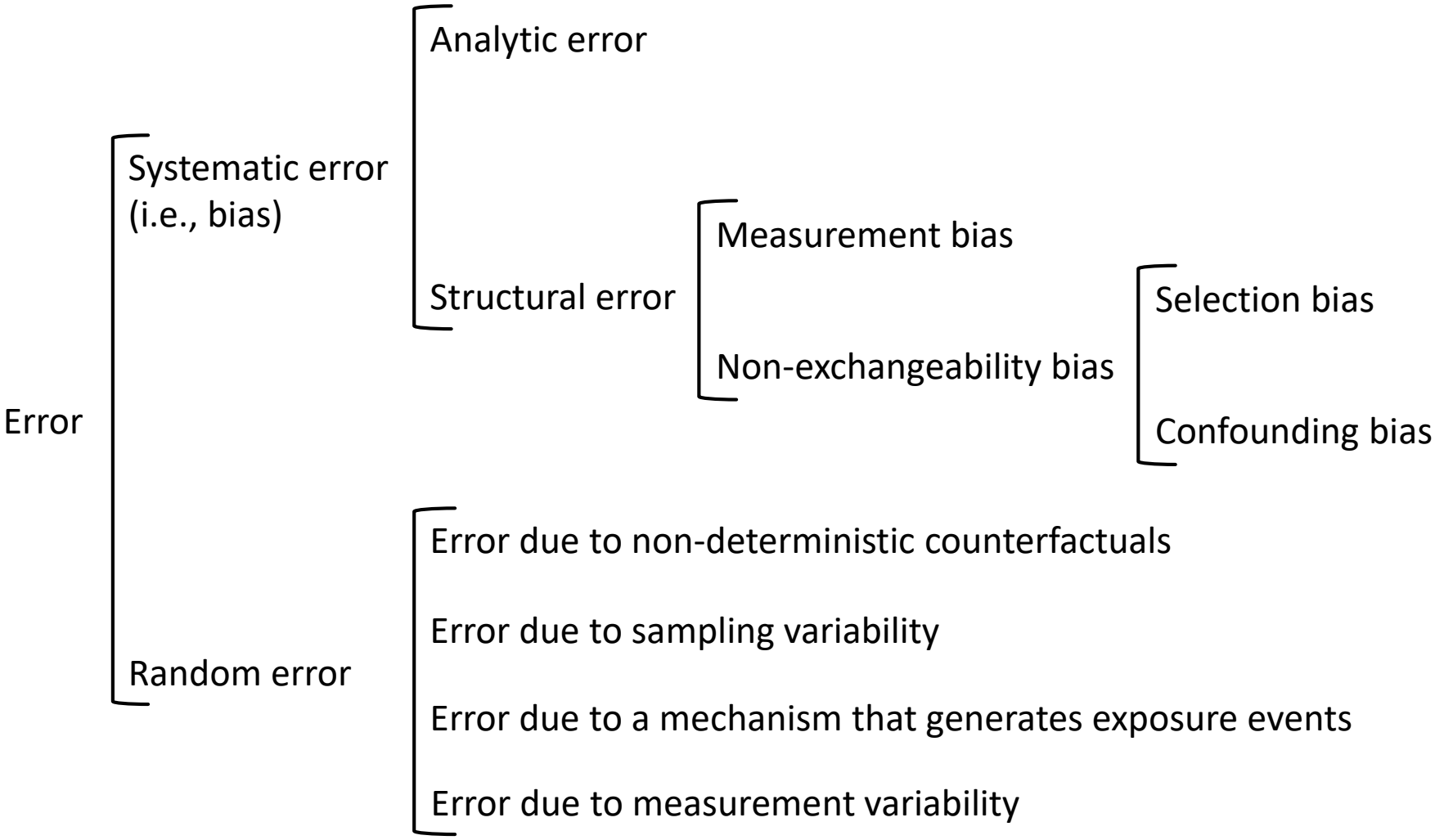
Error
Systematic error (i.e., bias)
Analytic error
Structural error
Measurement bias
Non-exchangeability bias
Selection bias
Confounding bias
Random error
Error due to non-deterministic counterfactuals
Error due to sampling variability
Error due to a mechanism that generates exposure events
Error due to measurement variability

Fig. 1

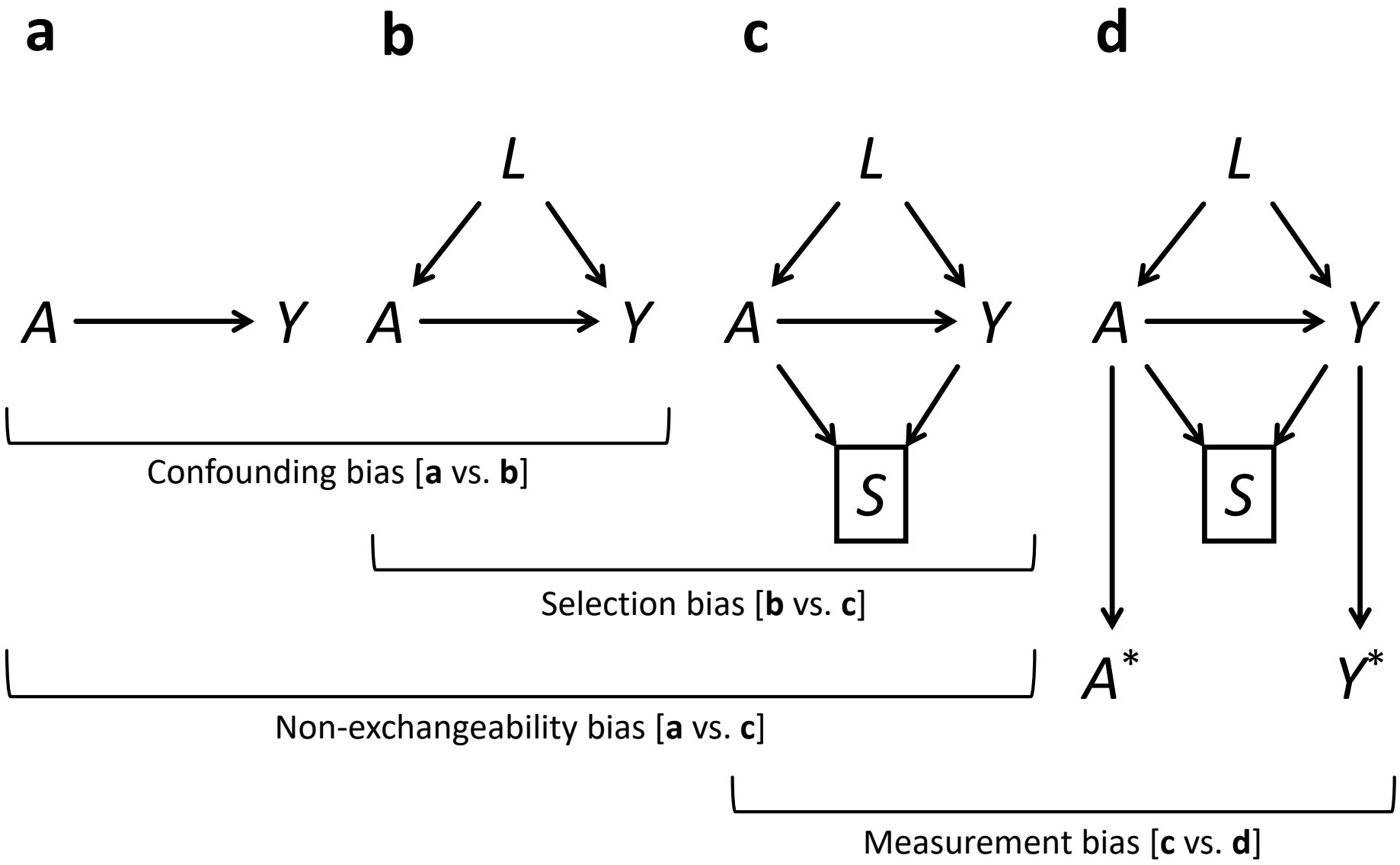
a
b
c
d
L
A
Y
S
A*
Y*
Confounding bias [a vs. b]
Selection bias [b vs. c]
Non-exchangeability bias [a vs. c]
Measurement bias [c vs. d]

Fig. 2

Supplementary Appendix A: Alternative decomposition of non-exchangeability bias

In the main text, we show the decomposition of non-exchangeability bias (Formula 5) into confounding bias (Formula 6) and selection bias (Formula 7). We here show that non-exchangeability bias can be alternatively decomposed into:

$$\mu\big(E(Y \mid A=1, S=1), E(Y \mid A=0, S=1)\big) - \mu\big(E(Y_1 \mid S=1), E(Y_0 \mid S=1)\big), \quad \text{(A.1)}$$

and

$$\mu\big(E(Y_1 \mid S=1), E(Y_0 \mid S=1)\big) - \mu\big(E(Y_1), E(Y_0)\big). \quad \text{(A.2)}$$

Note that Formula A.1 is the definition of confounding bias (Formula 6) among the selected individuals (i.e., S = 1). In other words, the difference between Formulas 6 and A.1 can be understood as a difference in the target populations; the target of the former is the total population whereas that of the latter is the subpopulation of S = 1. Formula A.2 is the difference between the 2 distinct causal measures, and may arise as an error due to selection of comparison groups. By contrast, recall that Formula 7 is the difference between the 2 distinct associational measures, and this bias is induced by selecting individuals into the study or the analysis [1].

In conclusion, Formula A.1 is related to internal validity of the subpopulation in randomized controlled trials, whereas Formula A.2 is related to external validity or generalizability in randomized controlled trials. In perfect randomized controlled trials, we expect that Formula A.1 becomes zero. Although non-exchangeability bias can be alternatively decomposed into Formulas A.1 and A.2, we note that this decomposition uses 2 distinct target populations. It would thus be natural to decompose non-exchangeability bias into Formulas 6 and 7 by consistently using a total population as the target population.

Supplementary Appendix B: Strict algebraic description of confounding in expectation

As mentioned in the main text, we refer to "confounding bias" in our schema by employing the notion of confounding in expectation [2, 3]. In a strict sense, however, the algebraic description of confounding bias in Formula 6 is based on the notion of realized confounding. Here we provide a strict algebraic description of confounding bias in the notion of confounding in expectation. To show a clear distinction between the notions of realized confounding and confounding in expectation, we need to introduce new notations, clarifying the *mechanism* that generates exposure events. We let J_m denote a scenario of exposure allocation among the target population, which is generated by the mechanism m. We also let A_j denote a binary exposure (1 = exposed, 0 = unexposed) under the scenario j. Then, realized confounding under scenario j can be

algebraically described as:

$$\mu\left(E(Y \mid A_j = 1), E(Y \mid A_j = 0)\right) - \mu\left(E\left(Y_1\right), E\left(Y_0\right)\right), \quad \text{(B.1)}$$

which essentially corresponds to Formula 6. On a related issue, Greenland and Mansournia [4] formalized "random confounding" using the potential-outcome (counterfactual) model and the notion of "sample exchangeability".

Confounding in expectation under mechanism *m* can be algebraically described as:

$$\mu\left(E_{J_m}E_Y(Y \mid A_{J_m} = 1), E_{J_m}E_Y(Y \mid A_{J_m} = 0)\right) - \mu\left(E\left(Y_1\right), E\left(Y_0\right)\right), \quad \text{(B.2)}$$

where $E_{J_m}E_Y\left(Y \mid A_{J_m} = a\right)$ can be calculated as $\sum_j E_Y\left(Y \mid A_j = a\right)P\left(J_m = j\right)$. Note that confounding in expectation (Formula B.2) is dependent on each mechanism that generates exposure events, whereas realized confounding (Formula B.1) is dependent on each product of that mechanism (or scenario). To simplify our discussion, we do not use Formula B.2 in the main text. Analogous discussions apply to selection bias and measurement bias.

References

1. Porta MS, editor. A Dictionary of Epidemiology. 6th ed. New York, NY: Oxford University Press; 2014.
2. VanderWeele TJ. Confounding and effect modification: distribution and measure. Epidemiol Method 2012;1(1):55-82. doi:10.1515/2161-962X.1004.
3. Suzuki E, Mitsuhashi T, Tsuda T, Yamamoto E. A typology of four notions of confounding in epidemiology. J Epidemiol: in press.
4. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. Eur J Epidemiol 2015;30(10):1101-10.