Published in final edited form as:

Curr Epidemiol Rep. 2015 September 1; 2(3): 162–171. doi:10.1007/s40471-015-0050-8.

Are all biases missing data problems?

Chanelle J. Howe, PhD MHS MPH[a], **Lauren E. Cain, PhD ScM MHS**[b], and **Joseph W. Hogan, ScD MS**[c]

[a]Department of Epidemiology, Center for Population Health and Clinical Epidemiology, Brown University School of Public Health, 121 South Main Street, Providence, Rhode Island 02912 (Phone: 401-863-7406, Fax: 401-863-3713, chanelle_howe@brown.edu)

[b]Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Kresge, Room 820, Boston, Massachusetts 02115 (Phone: 617-432-6921, Fax: 617-432-1884, lcain@hsph.harvard.edu)

[c]Department of Biostatistics, Center for Statistical Sciences, Brown University School of Public Health, Providence, Rhode Island 02912 (Phone: 401-863-9243, Fax: 401-863-9182, joseph_hogan@brown.edu)

Abstract

Estimating causal effects is a frequent goal of epidemiologic studies. Traditionally, there have been three established systematic threats to consistent estimation of causal effects. These three threats are bias due to confounders, selection, and measurement error. Confounding, selection, and measurement bias have typically been characterized as distinct types of biases. However, each of these biases can also be characterized as missing data problems that can be addressed with missing data solutions. Here we describe how the aforementioned systematic threats arise from missing data as well as review methods and their related assumptions for reducing each bias type. We also link the assumptions made by the reviewed methods to the missing completely at random (MCAR) and missing at random (MAR) assumptions made in the missing data framework that allow for valid inferences to be made based on the observed, incomplete data.

Keywords

Missing data; Confounding bias; Selection bias; Measurement bias

Corresponding author: Chanelle J. Howe, Department of Epidemiology, Center for Population Health and Clinical Epidemiology, Brown University School of Public Health, 121 South Main Street, Providence, RI 02912 (Phone: 401-863-7406, Fax: 401-863-3713, chanelle_howe@brown.edu).

Conflict of Interest

CJ Howe, LE Cain, and JW Hogan all declare no conflicts of interest.

Human and Animal Rights and Informed Consent

All studies by CJ Howe involving animal and/or human subjects were performed after approval by the appropriate institutional review boards. When required, written informed consent was obtained from all participants.

Introduction

A common objective in epidemiologic studies is to estimate the causal effect of an exposure on the occurrence of a specific outcome. Historically, there have been three recognized systematic threats to consistent estimation of causal effects. These three threats are bias due to confounders, selection, and measurement error [1]. Confounding, selection, and measurement bias have typically been characterized as distinct types of biases. However, these biases can also be characterized as missing data problems that arise when incomplete information must be used to estimate quantities of interest (e.g., causal effect) that would be obtained from the complete information if it were available [2, 3].

In studies conducted to estimate causal effects, information is incomplete on the individual-level potential outcomes (also known as the counterfactual outcomes). The individual-level potential outcome is the outcome that would have been observed for a given individual under an intervention to set the exposure for that individual to a specific level [1, 4]. At best researchers can observe the potential outcome for a given individual under their observed exposure while the individual-level potential outcome for their unobserved exposure level(s) is missing. Missing individual-level potential outcomes often shift the focus to estimating aggregate rather than individual causal effects [5]. Estimating aggregate causal effects typically requires calculating the mean or another relevant function (e.g., median) of the individual-level potential outcome distribution.

In epidemiologic studies confounding bias typically occurs when incomplete observed data are used to estimate relevant functions of the individual-level potential outcomes and in turn aggregate causal effects. Selection and measurement error can lead to additional missing individual-level potential outcomes, which can hinder accurate estimation of relevant functions and in turn result in bias. As with any missing data problem, however, these biases can be addressed by accurately completing the missing individual-level potential outcomes before estimating the relevant function [5–15]. Alternatively, the relevant function can often be directly estimated from the observed data without completing the missing individual-level potential outcomes [4]. Accurate estimation of the relevant function will result in the aggregate associational effects that are commonly calculated using the observed data equaling the aggregate causal effects of interest that would have been obtained if the individual-level potential outcomes were known/available [4].

Therefore, the objective of this paper is to first describe how confounding, selection, and measurement bias arise from missing data on individual-level potential outcomes. Second, we will review methods that reduce bias stemming from missing individual-level potential outcomes while emphasizing techniques that directly estimate the relevant function of the potential outcome distribution without completing the missing individual-level potential outcomes given their greater use in the epidemiologic literature. Third, we will detail the assumptions that must be met for each of the described methods to reduce the relevant bias type. To further characterize the aforementioned biases as missing data problems, we will link the assumptions made by the reviewed methods to the missing completely at random (MCAR) and missing at random (MAR) assumptions that are made in the missing data framework and allow for valid inferences to be made based on the observed, incomplete data

[2]. A summary of our review is included in Table 1. To aid in our description, we first define general notation and provide basic definitions that we will build upon in subsequent sections of the paper.

General notation and basic definitions

Let capital letters denote random variables while lower case letters and numbers represent particular values of the random variables. Now suppose in a hypothetical cohort study, A(the exposure) represents a binary indicator of aspirin use at the start of the study (1: yes; 0: no) and Y(the outcome) is an individual-level indicator of dying subsequent to the start of the study (1: yes; 0: no). Further, Y^a denotes the individual-level potential outcome for Y under an intervention to set aspirin use to level a and L is a binary indicator of male gender (1: yes; 0: no).

Next we specify the consistency condition for all subjects that $Y^A = Y$. This condition states that the observed individual-level outcome among study participants is the individual-level potential outcome that would have been observed under an intervention to set the exposure to the observed exposure level if the exposure was measured without error [16–19]. For instance, study participants who used aspirin at study entry and died during follow up would have still died if they were assigned to use aspirin at the start of the study. The consistency condition is required to make appropriate causal inference based on the observed data.

Typically in epidemiologic studies the aggregate causal effect of interest is the average causal effect [4]. The causal risk ratio (RR) will be the average causal effect that will be the focus of this paper, acknowledging that the issues discussed in subsequent sections often equally apply to other measures (e.g., causal rate ratio). In the context of the hypothetical cohort study, the causal RR is a function of the mean of Y^a, $E[Y^a]$, and compares aspirin users to non-aspirin users via the following quantity: $E[Y^{a=1}]/ E[Y^{a=0}] = P(Y^{a=1} = 1)/P(Y^{a=0} = 1)$. The numerator of the causal RR is the risk of subsequent death had all individuals in the study population used aspirin at study entry, while the denominator is the risk of subsequent death had no individuals in the study population used aspirin at study entry.

When there is a non-zero probability of observing each exposure level (i.e., positivity) and the exposure groups are equivalent/exchangeable on all factors related to Y (i.e., individual-level potential outcomes are MCAR [2]), the mean of Y conditional on A, $E[Y \mid A]$, can validly be used to calculate the aforementioned $E[Y^a]$ and in turn accurately estimate the causal RR via the associational RR, $E[Y \mid A = 1]/E[Y \mid A = 0] = P(Y = 1 \mid A = 1)/P(Y = 1 \mid A = 0)$. The associational RR is obtained by comparing disjoint subsets of the study population with different exposure levels. Specifically, the risk of death among individuals who used aspirin at study entry is compared to the risk of death among individuals who did not use aspirin at study entry. Bias occurs when the average associational and causal effects are unequal due to a lack of positivity or lack of exchangeability [4]. The lack of positivity or exchangeability in this context has also been referred to as non-ignorability of the treatment assignment mechanism [4, 20].

Bias due to missing individual-level potential outcome data

If all potential outcomes were known at the individual level then the presence of factors that have been traditionally referred to as confounders, the presence of selection, or the presence of measurement error would pose no threat to identifying causal effects (i.e., no bias). Therefore, the following sections describe how confounding, selection, and measurement bias arise from missing individual-level potential outcomes. Table 2 depicts which individual-level potential outcomes are missing for each bias type. Further, the Figure represents each bias type as a causal diagram. Unless otherwise specified, when discussing one type of bias we assume the absence of other bias types.

Confounding bias

Confounding bias occurs because potential outcomes for a given person under study are not seen for unobserved exposure levels. For instance, as shown in Table 2, whether an aspirin user ($A = 1$) who died ($Y = 1$) would have lived had they not used aspirin is not observed ($Y^{a=0} = ?$). Similarly, whether a non-aspirin user ($A = 0$) who lived ($Y = 0$) would have died had they used aspirin is not observed ($Y^{a=1} = ?$). Missing individual-level potential outcomes often shifts the focus from estimating individual to aggregate causal effects (e.g., causal RR) in epidemiologic studies given the greater ease in accurately estimating aggregate rather than individual causal effects [4, 5].

Validly estimating the causal RR requires accurately estimating $E[Y^a]$. When positivity and exchangeability hold, the mean of the observed outcomes conditional on A, $E[Y \mid A]$, can be used to accurately estimate $E[Y^a]$. For instance, the mean outcome among persons observed to take aspirin, $E[Y \mid A = 1]$, can be used to estimate the mean potential outcome of the persons observed to not take aspirin had they taken aspirin, $E[Y^{a=1} \mid A = 0]$, and in turn the mean potential outcome had all participants in our hypothetical cohort study taken aspirin, $E[Y^{a=1}]$. Unfortunately, estimation of $E[Y^a]$ is not always accurate due to lack of positivity or exchangeability.

Exchangeability violations can occur when factors such as L in Diagram (I) of the Figure that are associated with the exposure and outcome exist (e.g., males are more likely than females to take aspirin and die). Such factors have traditionally been referred to as confounders [1, 4]. The existence of potential confounders may result in $E[Y \mid A] \neq E[Y^{a=A}]$ since any observed outcome may be due to the effect of the exposure, the related confounder, or both. The lack of exchangeability on confounders across exposure levels can result in bias; meaning the average associational and causal effect measures are unequal (e.g., $P(Y = 1 \mid A = 1)/P(Y = 1 \mid A = 0) \neq P(Y^{a=1} = 1)/P(Y^{a=0} = 1)$). This bias is depicted in Diagram (I) of the Figure by the open non-causal path from A to Y via L.

Selection bias

Now consider that the population from the hypothetical cohort study was selected from a source population that represents a target population that we would like to make inferences about. Therefore, let S be a binary indicator of selection into the study population from the source population (1:selected; 0:not selected) and $Y^{a,s}$ denote the potential outcome for Y

under an intervention to set aspirin use to a and selection to s. The causal RR with selection will now be represented as $P(Y^{a=1,s=1} = 1)/P(Y^{a=0,s=1} = 1)$. This causal RR compares the same individuals had everyone in the source population been selected and aspirin use been assigned versus not assigned. In contrast, the associational RR with selection, $P(Y = 1 \mid A = 1, S = 1)/P(Y = 1 \mid A = 0, S = 1)$, compares disjoint subsets of the observed study population by their observed aspirin use.

Although not captured by the definition of S in the hypothetical cohort study, selection could also occur when going from the study population to the analytic sample and from the analytic sample at study entry to a given risk set subsequent to study entry [15]. As shown in Table 2, selection can result in missing exposure, observed outcome, and individual-level potential outcome values for persons who were not selected into a given study population, analytic sample, or risk set (A = ?, Y = ?, $Y^{a=1}$ = ?, $Y^{a=0}$ = ?) [3].

When estimating the causal RR, an analysis that disregards the aforementioned selection results in the mean of the observed outcomes among the selected persons (e.g., $P(Y = 1 \mid A, S = 1)$) being used to estimate the mean potential outcome of the not selected persons (e.g., $P(Y^a = 1 \mid A, S = 0)$). This estimation is inaccurate when the probability of being selected is not greater than zero (i.e., non-positivity) or the mean potential outcomes among persons who were selected (e.g., $P(Y^a = 1 \mid A, S = 1)$) are not exchangeable with the missing mean potential outcomes among those who were not selected (e.g., $P(Y^a = 1 \mid A, S = 0)$).Here, exchangeability is defined as equivalence between those who were and were not selected on factors related to the outcome and in turn their potential outcomes conditional on the exposure [21]. Exchangeability occurs when the potential outcomes among those who were not selected are MCAR or MAR conditional on the exposure [2]. Positivity plus exchangeability is equivalent to ignorability of the selection mechanism [4, 20].

The lack of exchangeability between those who were and were not selected can occur if factors such as L in Diagram (II) of the Figure that are associated with selection and the outcome exist such that those who were selected (e.g., mostly men) are more or less likely to develop the outcome (e.g., die) than those who were not selected. The lack of exchangeability between those who were and were not selected may result in a lack of exchangeability across different exposure levels and in turn bias (e.g., $P(Y = 1 \mid A = 1, S = 1)/P(Y = 1 \mid A = 0, S = 1) \neq P(Y^{a=1,s=1} = 1)/P(Y^{a=0,\, s=1} = 1)$) [21, 22]. This bias in Diagram (II) of the Figure is represented by the open non-causal path from A to Y via L and S.

Measurement bias

Now consider that in the hypothetical cohort the exposure, outcome, or confounders/covariates may have been measured with error. Measurement error is a type of missing data because the true value for the exposure, outcome, or confounder/covariate is not known. Therefore, let A^* be the measured version of A via self-report, U_A be the measurement error for A, Y^* be the measured version of Y obtained from medical record abstraction, U_Y be the measurement error for Y, L^* be the measured version of L obtained from medical record abstraction, and U_L be the measurement error for L. The associational effect with measurement error is now represented as $P(Y^* = 1 \mid A^* = 1, S = 1)/P(Y* = 1 \mid A* = 0, S = 1)$

and compares disjoint subsets of the observed study population by their reported rather than actual/true aspirin use.

In the presence of U_Y the individual-level potential outcomes that correspond to the observed exposure levels are missing since the observed outcomes are missing. Even when the outcome is measured perfectly (no U_Y) like in Table 2, if the exposure is measured with error ($A \neq A^*$), the individual-level potential outcomes that correspond to the observed exposure levels are still missing ($Y^{a=A^*} = ?$) due to the incorrect labeling of persons who are actually exposed as unexposed and vice versa [3].

Therefore, like in the case of Diagram (III) in the Figure where confounding and selection bias do not exist, errors in measurement of the exposure (U_A) and/or outcome (U_Y) can result in bias (e.g., $P(Y^* = 1 \mid A^* = 1, S = 1)/P(Y^* = 1 \mid A^* = 0, S = 1) \neq P(Y^{a=1,s=1} = 1)/P(Y^{a=0,s=1} = 1)$) that would not exist had study participants been compared based on their true outcome level and aspirin use (e.g., $P(Y = 1 \mid A = 1, S = 1)/P(Y = 1 \mid A = 0, S = 1) = P(Y^{a=1,s=1} = 1)/P(Y^{a=0,s=1} = 1)$). Even when the exposure and outcome are measured perfectly, confounder/covariate measurement error ($L \neq L^*$) is problematic when confounding or selection bias exist. Specifically, errors in the measurement of factors that contribute to the confounding or selection bias hinders the accurate estimation of $E[Y^a]$ based on information about the aforementioned factors using methods outlined in the next section [23].

Methods and assumptions necessary to reduce bias due to missing individual-level potential outcomes

Confounding bias

Approaches that have been more frequently used in the epidemiologic literature to address confounding bias due to missing individual-level potential outcomes include randomization, standardization, restriction, matching, stratification, standard regression adjustment, propensity scores, and inverse probability weighting (IPW) [1, 4, 24–28]. In the case of the binary time-fixed indicator of aspirin use, randomization, stratification, restriction, matching, standardization, standard regression adjustment, and propensity scores would use the mean observed outcome among the non-aspirin users to estimate the mean potential outcome among aspirin users and in turn the entire study population had they not used aspirin. Likewise, the mean observed outcome among the aspirin users would be used to estimate the mean potential outcome of non-aspirin users and in turn the entire study population had they used aspirin.

Randomization helps ensure that the estimate is accurate by assigning the exposure randomly such that the distribution of measured and unmeasured confounders is expected to be equivalent across different levels of the exposure. Standardization and matching (e.g., individual, frequency, propensity score) reduce differences in the confounder distribution across different exposure groups. Stratification, restriction, and standard regression adjustment ensures the estimation is accurate by only performing the estimation within strata of measured potential confounders where greater balance on potential confounders is expected. Propensity score subclassification analogously performs the estimation within strata of scores that are a function of the potential confounders where again greater balance on potential confounders is expected.

The greater balance on the confounder distribution across different exposure levels expected to be achieved by each of the above described methods blocks the open non-causal pathway from *A* to *Y* via *L* in Diagram (I) of the Figure by effectively either removing the arrow from *L* to *A* or by conditioning on *L*. This blocking in turn reduces the occurrence of differences in the outcome across exposure levels that occur for reasons beyond the exposure that hinder accurate estimation. Despite the aforementioned expectation, balance is only achieved when necessary assumptions and conditions are met.

IPW procedures can be used to re-weight the observed data to generate a pseudo-population with the corresponding outcomes that would have been observed had everyone in the study population been exposed and unexposed (e.g., used aspirin and did not use aspirin). This re-weighting is usually done as a function of the measured potential confounders and similar to randomization removes the arrow from *L* to *A*. The re-weighted data therefore yields a pseudo-population where the aforementioned measured potential confounders are not associated with the exposure and the estimation of $E[Y^a]$ is therefore accurate when necessary assumptions and conditions are met. Stabilized versions of the aforementioned weights can be estimated that preserve the original sample size of the study population and enhance the precision of estimates. When a measured time-varying confounder that is affected by prior exposure exists, IPW procedures may be less biased, but more imprecise, than more traditional approaches including standard regression adjustment [27, 28].

Less commonly used methods in the epidemiologic literature that address confounding bias when necessary assumptions and conditions are met include instrumental variable approaches [29–34], g-estimation [8–12], the g-computation formula [35–37], and Bayesian techniques [5, 7]. Despite their potential to circumvent bias due to measured and unmeasured potential confounders, instrumental variable approaches have been less frequently employed in observational epidemiologic studies in part due to the limited number of suitable instruments in this setting [31, 32]. Although g-estimation and the g-computation formula may also be less biased than traditional approaches in the setting of a measured time-varying confounder that is affected by prior exposure, these g-methods are also infrequently used by epidemiologists along with Bayesian techniques [5, 7] likely due to their greater complexity compared to other methods. However, more recent applications of these g-methods and Bayesian approaches that provide code should facilitate greater consideration of these valuable techniques [5, 8, 9, 37, 38].

Another set of complex and therefore also less utilized methods include doubly robust estimators. Doubly robust estimators represent more flexible strategies for confounder control, which in certain settings can yield more valid and precise effect estimates compared to the aforementioned techniques [39, 40]. Thus, despite their greater complexity, doubly robust estimators should be more readily considered for use by applied researchers as well especially because code is now also widely available to implement these methods [41].

Each of the previously described methods requires the consistency condition and assumes exchangeability (potential outcomes are MCAR or MAR conditional on measured potential confounders), positivity, and correct model specification (when semi-parametric and fully parametric techniques are employed) [4, 42]. Here positivity requires a non-zero probability of each instrument/exposure level marginally or within every observed combination of potential confounders. The exclusion restriction, which requires the instrument to only affect the outcome through the exposure, is also necessary for the instrumental variable approach.

Some methods are more sensitive to assumption violations and may be less efficient than others. Therefore, when selecting which method(s) to use to estimate a given causal effect, careful consideration should be given to which method(s) is most feasible and appropriate in a given research setting. The results from multiple methods can also be compared. Further, sensitivity analysis techniques [43–50] should be readily employed concurrently with the selected technique(s) to assess the robustness of inferences in the presence of potential assumption violations.

Selection bias

The two approaches that have been most commonly used in the epidemiologic literature to address selection bias due to missing individual-level potential outcomes include standard regression adjustment and IPW [4, 15, 51, 52]. Standard regression adjustment ensures that the estimation of a relevant function of the individual-level potential outcomes (e.g., $P(Y^a = 1 \mid A, S = 0)$) is accurate by only performing the estimation within strata of measured covariates that are associated with selection and the outcome of interest such as L in Diagram (II) of the Figure. Differences in the distribution of covariates like L between selected and not selected persons is the source of differences in the relevant function between selected and not selected persons and in turn the selection bias. Thus estimating the relevant function within strata of L should be accurate and reduce selection bias when necessary assumptions and conditions are met. Further, estimating the relevant function within strata of L is equivalent to conditioning on L in Diagram (II) of the Figure and blocking the open non-causal pathway from A to Y via S and L.

IPW can be used in a broader number of selection bias scenarios than standard regression adjustment to facilitate the accurate estimation of relevant functions and in turn reduce selection bias [15]. Specifically, IPW procedures re-weight the observed data to generate a pseudo-population that includes the missing individual-level potential outcomes of those individuals who were not selected. This re-weighting is performed as a function of the measured covariates that are associated with selection and the outcome of interest (e.g., L). The re-weighted data therefore yields a pseudo-population where the covariates used to estimate the weights are no longer associated with selection (e.g., the arrow from L to S in Diagram (II) of the Figure is removed) and the resulting estimated relevant function of the individual-level potential outcomes is accurate when necessary assumptions and conditions are met.

Stabilized versions of the aforementioned selection weights can be estimated that preserve the number of observed outcomes and enhance the precision of estimates. When competing risks are a source of the potential selection bias (e.g., dying before the outcome occurs or is assessed), IPW [53, 54] as well as other methods [55–58] have been used to address the potential selection bias. However, there remains considerable debate regarding whether estimating relevant functions of the individual-level potential outcomes using methods such as IPW is appropriate when the potential outcome is undefined, like in the case where the competing risk is death [59, 60].

Standard regression adjustment, IPW, as well as other techniques including more flexible doubly robust estimators [13, 14, 35, 41, 54–57, 61] for reducing selection bias require the consistency condition and assume exchangeability, positivity, and correct model specification (when semi-parametric and fully parametric techniques are employed) [4, 21, 42]. Here exchangeability requires no unmeasured covariates that contribute to the selection bias (i.e., potential outcomes are MAR conditional on exposure and measured covariates) while positivity requires a non-zero probability of being selected within every exposure level and observed combination of the exposure and the covariates that contribute to the selection bias. Given that unmeasured covariates likely exist, more recently employed instrumental variable approaches [62–64] to address selection bias related to measured and unmeasured covariates are appealing. However, suitable instruments that satisfy the exclusion restriction (instrument only associated with the outcome through selection) are likely limited. Therefore, after selecting the technique(s) most appropriate for the particular research setting, sensitivity analysis procedures [50, 65–72] should be employed concurrently with the selected technique(s) to assess the robustness of inferences in the presence of potential assumption violations.

Measurement bias

Bias analysis techniques [1, 50, 73–75] can be used to obtain more accurate estimates of relevant functions of the individual-level potential outcomes when measurement error is present and necessary assumptions and conditions are met. Specifically, simple bias analysis [1, 50] uses validity measures (e.g., sensitivity and specificity or positive predictive value and negative predictive value) obtained from validation data, expert opinions, or the published literature to correctly classify participants by their misclassified exposure, outcome, or covariate. When the validity measures are accurate this reclassification aids in the valid estimation of the functions of the individual-level potential outcomes of interest by effectively removing the arrows from U_A to A^* and U_Y to Y^* in Diagram (III) of the Figure or from U_L to L^* (not shown).

Recently Funk and Landi [75] provided a nice review of the aforementioned simple bias analysis as well as other methods for addressing measurement bias. Other discussed methods include probabilistic bias analysis [1, 50], Bayesian bias analysis [1, 73, 74], regression calibration [76, 77], modified maximum likelihood [78–80], multiple imputation [77, 81], and propensity score calibration [82–84] which all also require that appropriate assumptions are met. The review also articulates the research settings when each of these approaches is most appropriate.

Conclusions

In epidemiologic studies, the three main threats to obtaining consistent estimates of causal effects can be characterized as missing data problems that can be addressed using a myriad of methods so that associational effects equal the desired causal effects of interest. Each of these methods makes assumptions. Many of these assumptions cannot be tested empirically. Therefore, the application of these methods should be based on the research setting and combined with sensitivity analyses to examine how robust inferences are to potential violations in relevant assumptions.

Acknowledgements

LE Cain was supported by the National Institutes of Health [grant number R01-AI102634]. JW Hogan was supported by grants P30-AI42853, P01-AA019072 and R01-AI-108441 from the National Institutes of Health.

References

Papers of particular interest, published recently, have been highlighted as:

•• Of major importance

• Of importance

1. Rothman, KJ.; Greenland, S.; Lash, TL. Modern Epidemiology. 3rd ed.. Philadelphia, PA: Lippincott Williams & Wilkins; 2008.

2. Rubin DB. Inference and missing data. Biometrika. 1976; 63:581–592.

3. 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; 28 Consistent with the present review paper the authors use a simple example to describe causal inference as a problem of missing potential outcomes particularly focusing on the case of estimating a causal effect in the presence of potential bias due to measurement error.

4. Hernán MA, Robins J. Causal Inference Book. http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/. The authors provide a cohesive introductory text to concepts and methods for causal inference.

5. Hill J. Bayesian Nonparametric Modeling for Causal Inference. Journal of Computational and Graphical Statistics. 2011; 20(1):217–240.

6. Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. New York: John Wiley & Sons; 1987.

7. Rubin DB. Bayesian inference for causal effects: the role of randomization. Ann Stat. 1978; 6(1): 34–58.

8. Naimi AI, Cole SR, Hudgens MG, Richardson DB. Estimating the effect of cumulative occupational asbestos exposure on time to lung cancer mortality: using structural nested failure-time models to account for healthy-worker survivor bias. Epidemiology. 2014; 25(2):246–254. [PubMed: 24487207] The authors use g-estimation to estimate the cumulative effect of occupational asbestos exposure on time to lung cancer mortality with annotated SAS code provided in an earlier commentary [9].

9. Naimi AI, Richardson DB, Cole SR. Causal inference in occupational epidemiology: accounting for the healthy worker effect by using structural nested models. Am J Epidemiol. 2013; 178(12):1681–1686. Epub 2013 Sep 27. [PubMed: 24077092]

10. Robins JM. Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. Biometrika. 1992; 79:321–334.

11. Robins, JM. Causal inference from complex longitudinal data. In: Berkane, M., editor. Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics 120. New York: Springer-Verlag; 1997. p. 69-117.

12. Robins JM, Blevins D, Ritter G, Wulfsohn M. G-estimation of the effect of prophylaxis therapy for Pneumocystis carinii pneumonia on the survival of AIDS patients. Epidemiology. 1992; 3(4):319–336. [PubMed: 1637895]

13. Hsu CH, Taylor JM, Murray S, Commenges D. Survival analysis using auxiliary variables via non-parametric multiple imputation. Stat Med. 2006; 25(20):3503–3517. [PubMed: 16345047]

14. Malani HM. A Modification of the Redistribution to the Right Algorithm Using Disease Markers. Biometrika. 1995; 82(3):515–526.

15. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004; 15(5):615–625. [PubMed: 15308962]

16. VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology. 2009; 20(6):880–883. [PubMed: 19829187]

17. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009; 20(1):3–5. [PubMed: 19234395]

18. Pearl J. On the consistency rule in causal inference: axiom, definition, assumption, or theorem? Epidemiology. 2010; 21(6):872–875. [PubMed: 20864888]

19. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. Epidemiology. 2011; 22(3):368–377. [PubMed: 21399502]

20. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.

21. Howe CJ, Cole SR, Chmiel JS, Munoz A. Limitation of inverse probability-of-censoring weights in estimating survival in the presence of strong selection bias. Am J Epidemiol. 2011; 173(5):569–577. Epub 2011 Feb 2. [PubMed: 21289029]

22. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. Stat Methods Med Res. 2012; 21(3):243–256. Epub 2011 Mar 9. [PubMed: 21389091]

23. Hernán MA, Cole SR. Invited Commentary: Causal diagrams and measurement bias. Am J Epidemiol. 2009; 170(8):959–962. discussion 63–4. Epub 2009 Sep 15. [PubMed: 19755635]

24. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med. 1997; 127(8 Pt 2):757–763. [PubMed: 9382394]

25. Slade EP, Stuart EA, Salkever DS, Karakus M, Green KM, Ialongo N. Impacts of age of onset of substance use disorders on risk of adult incarceration among disadvantaged urban youth: a propensity score matching approach. Drug Alcohol Depend. 2008; 95(1–2):1–13. Epub 08 Jan 31. [PubMed: 18242006]

26. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010; 25(1):1–21. [PubMed: 20871802]

27. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000; 11(5):561–570. [PubMed: 10955409]

28. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11:550–560. [PubMed: 10955408]

29. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol. 2000; 29(4):722–729. [PubMed: 10922351]

30. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology. 2006; 17(3): 268–275. [PubMed: 16617275]

31. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology. 2006; 17(4):360–372. [PubMed: 16755261]

32. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology. 2006; 17(3):260–267. [PubMed: 16617274]

33. Davies NM, Smith GD, Windmeijer F, Martin RM. COX-2 selective nonsteroidal anti-inflammatory drugs and risk of gastrointestinal tract complications and myocardial infarction: an instrumental variable analysis. Epidemiology. 2013; 24(3):352–362. [PubMed: 23532054]

34. Swanson SA, Hernan MA. Commentary: how to report instrumental variable analyses (suggestions welcome). Epidemiology. 2013; 24(3):370–374. [PubMed: 23549180] The authors offer guidelines for how to report instrumental variable analyses to address confounding bias using the Davies et al. [33] paper as an example.

35. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period–application to control of the healthy worker survivor effect. Math Model. 1986; 7:1393–1512.

36. Taubman SL, Robins JM, Mittleman MA, Hernan MA. Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. Int J Epidemiol. 2009; 38(6):1599–1611. Epub 2009 Apr 23. [PubMed: 19389875]

37. Keil AP, Edwards JK, Richardson DB, Naimi AI, Cole SR. The parametric g-formula for time-to-event data: intuition and a worked example. Epidemiology. 2014; 25(6):889–897. [PubMed: 25140837] The authors provide a simple introduction to the parametric g-formula with annotated SAS code for implementing the method and demonstrate its use when examining the effect of a hypothetical treatment to prevent graft-versus-host disease on mortality among bone marrow transplant patients.

38. HSPH Program on Causal Inference Software. http://www.hsph.harvard.edu/causal/software/.

39. Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. Biometrics. 2005; 61(4):962–973. [PubMed: 16401269]

40. van der Laan MJ, Gruber S. Targeted minimum loss based estimation of causal effects of multiple time point interventions. Int J Biostat. 2012; 8(1)

41. Neugebauer R, Schmittdiel JA, van der Laan MJ. Targeted learning in real-world comparative effectiveness research with time-varying interventions. Stat Med. 2014; 33(14):2480–2520. Epub 2014 Feb 17. [PubMed: 24535915] The authors use doubly robust targeted minimum loss-based estimation with super learning to address confounding and selection bias while examining the effect of various glucose-lowering strategies on albuminuria among adults with Type-2 diabetes and provide annotated R code for implementation.

42. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008; 168(6):656–664. [PubMed: 18682488]

43. Vanderweele TJ, Arah OA. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology. 2011; 22(1):42–52. [PubMed: 21052008]

44. Brumback BA, Hernan MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat Med. 2004; 23(5): 749–767. [PubMed: 14981673]

45. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. Int J Biostat. 2007:3–14.

46. Small DS. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. J Am Statist Assoc. 2007; 102:1049–1058.

47. Small DS, Rosenbaum P. War and wages: the strength of instrumental variables and their sensitivity to unobserved biases. J Am Statist Assoc. 2008; 103:924–933.

48. Baiocchi M, Small DS, Lorch S, Rosenbaum P. Building a stronger instrument in an observational study of perinatal care for premature infants.. 2010;105. J Am Statist Assoc. 2010; 105:1285–1296.

49. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological challenges in mendelian randomization. Epidemiology. 2014; 25(3):427–435. [PubMed: 24681576]

50. Lash, TL.; Fox, MP.; Fink, AK. Statistics for Biology and Health. New York, NY: Springer Science+Business Media, LLC; 2009. Applying Quantitative Bias Analysis to Epidemiologic Data.

51. Hernán MA, McAdams M, McGrath N, Lanoy E, Costagliola D. Observation plans in longitudinal studies with time-varying treatments. Stat Methods Med Res. 2009; 18(1):27–52. [PubMed: 19036915]

52. Gottesman RF, Rawlings AM, Sharrett AR, Albert M, Alonso A, Bandeen-Roche K, et al. Impact of differential attrition on the association of education with cognitive change over 20 years of follow-up: the ARIC neurocognitive study. Am J Epidemiol. 2014; 179(8):956–966. Epub 2014 Mar 13. [PubMed: 24627572]

53. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. Epidemiology. 2012; 23(1):119–128. [PubMed: 21989136]

54. Shardell M, Hicks GE, Ferrucci L. Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. Biostatistics. 2015; 16(1):155–168. Epub 2014 Jul 4. [PubMed: 24997309] The authors use doubly robust augmented inverse probability weighted estimation to address selection bias due to death and lost to follow up when examining the effect of Vitamin D use on physical functioning among older adults.

55. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am J Epidemiol. 2009; 170(2):244–256. Epub 2009 Jun 3. [PubMed: 19494242]

56. Lau B, Cole SR, Gange SJ. Parametric mixture models to evaluate and summarize hazard ratios in the presence of competing risks with time-dependent hazards and delayed entry. Stat Med. 2011; 30(6):654–665. Epub 2010 Nov 30. [PubMed: 21337360]

57. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics. 2002; 58(1):21–29. [PubMed: 11890317]

58. Vanderweele TJ. Principal stratification--uses and limitations. Int J Biostat. 2011; 7(1) (pii):Article 28. Epub 2011 Jul 11.

59. Chaix B, Evans D, Merlo J, Suzuki E. Commentary: Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. Epidemiology. 2012; 23(1):129–131. discussion 32–7. [PubMed: 22157307]

60. Tchetgen Tchetgen EJ, Glymour M, Shpitser I, Weuve J. To weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. Epidemiology. 2012; 23(4):644–646. [PubMed: 22659551]

61. Murray S, Tsiatis AA. Nonparametric survival estimation using prognostic longitudinal covariates. Biometrics. 1996; 52(1):137–151. [PubMed: 8934589]

62. Barnighausen T, Bor J, Wandira-Kazibwe S, Canning D. Correcting HIV prevalence estimates for survey nonparticipation using Heckman-type selection models. Epidemiology. 2011; 22(1):27–35. [PubMed: 21150352]

63. Hogan DR, Salomon JA, Canning D, Hammitt JK, Zaslavsky AM, Barnighausen T. National HIV prevalence estimates for sub-Saharan Africa: controlling selection bias with Heckman-type selection models. Sex Transm Infect. 2012; 88(Suppl 2):i17–i23. [PubMed: 23172342]

64. McGovern ME, Barnighausen T, Salomon JA, Canning D. Using interviewer random effects to remove selection bias from HIV prevalence estimates. BMC Med Res Methodol. 2015; 15(1):8. [PubMed: 25656226] The authors use an instrumental variable approach to correct for selection bias when estimating the prevalence of HIV among men in Ghana and Zambia.

65. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for non-ignorable drop-out using semi-parametric non-response models. J Am Statist Assoc. 1999; 94:1096–1120.

66. Scharfstein DO, Rotnitzky A, Robins JM. Adjusting for non-ignorable drop-out using semi-parametric non-response models [Comments and Rejoinder]. J Am Statist Assoc. 1999

67. Robins, JM.; Rotnitzky, A.; Scharfstein, DO. Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In: Halloran, MEB.; D, editors. Statistical Models in Epidemiology: The Environment and Clinical Trials. IMA. Vol. 116. New York: Springer-Verlag; 1999. p. 1-92.

68. Scharfstein D, Robins JM, Eddings W, Rotnitzky A. Inference in randomized studies with informative censoring and discrete time-to-event endpoints. Biometrics. 2001; 57(2):404–413. [PubMed: 11414563]

69. Scharfstein DO, Robins JM. Estimation of the failure time distribution in the presence of informative censoring. Biometrika. 2002; 89(3):617–634.

70. Robins J, Rotnitzky A, Vansteelandt S, Frangakis CE, Rubin DB, An M, MacKenzie E. Principal stratification designs to estimate input data missing due to death. Biometrics. 2007; 63(3):650–653. In discussion of: [PubMed: 17824996]

71. Long DM, Hudgens MG. Comparing competing risk outcomes within principal strata, with application to studies of mother-to-child transmission of HIV. Stat Med. 2012; 31(27):3406–3418. Epub 2012 Aug 28. [PubMed: 22927321]

72. Geneletti S, Mason A, Best N. Adjusting for selection effects in epidemiologic studies: why sensitivity analysis is the only "solution". Epidemiology. 2011; 22(1):36–39. [PubMed: 21150353]

73. Chu H, Wang Z, Cole SR, Greenland S. Sensitivity analysis of misclassification: a graphical and a Bayesian approach. Ann Epidemiol. 2006; 16(11):834–841. Epub 2006 Jul 13. [PubMed: 16843678]

74. MacLehose RF, Olshan AF, Herring AH, Honein MA, Shaw GM, Romitti PA. Bayesian methods for correcting misclassification: an example from birth defects epidemiology. Epidemiology. 2009; 20(1):27–35. [PubMed: 19234399]

75. Funk MJ, Landi SN. Misclassification in Administrative Claims Data: Quantifying the Impact on Treatment Effect Estimates. Curr Epidemiol Rep. 2014; 1:175–185. [PubMed: 26085977] The authors review the strengths and limitations including assumptions of various methods to reduce bias due to measurement error when estimating causal effects using administrative claims data.

76. Spiegelman D, McDermott A, Rosner B. Regression calibration method for correcting measurement-error bias in nutritional epidemiology. Am J Clin Nutr. 1997; 65(4 Suppl):1179S–1186S. [PubMed: 9094918]

77. Bang H, Chiu YL, Kaufman JS, Patel MD, Heiss G, Rose KM. Bias Correction Methods for Misclassified Covariates in the Cox Model: comparison offive correction methods by simulation and data analysis. J Stat Theory Pract. 2013; 7(2):381–400. [PubMed: 24072991]

78. Magder LS, Hughes JP. Logistic regression when the outcome is measured with uncertainty. Am J Epidemiol. 1997; 146(2):195–203. [PubMed: 9230782]

79. Neuhaus J. Bias and efficiency loss due to misclassified responses in binary regression. Biometrika. 1999; 86(4):843–855.

80. Lyles RH, Tang L, Superak HM, King CC, Celentano DD, Lo Y, et al. Validation data-based adjustments for outcome misclassification in logistic regression: an illustration. Epidemiology. 2011; 22(4):589–597. [PubMed: 21487295]

81. Cole SR, Chu H, Greenland S. Multiple-imputation for measurement-error correction. Int J Epidemiol. 2006; 35(4):1074–1081. Epub 2006 May 18. [PubMed: 16709616]

82. Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. Am J Epidemiol. 2005; 162(3):279–289. Epub 2005 Jun 29. [PubMed: 15987725]

83. Sturmer T, Glynn RJ, Rothman KJ, Avorn J, Schneeweiss S. Adjustments for unmeasured confounders in pharmacoepidemiologic database studies using external information. Med Care. 2007; 45 Supl 2(10):S158–S165. [PubMed: 17909375]

84. Lunt M, Glynn RJ, Rothman KJ, Avorn J, Sturmer T. Propensity score calibration in the absence of surrogacy. Am J Epidemiol. 2012; 175(12):1294–1302. Epub 2012 Apr 24. [PubMed: 22688682]

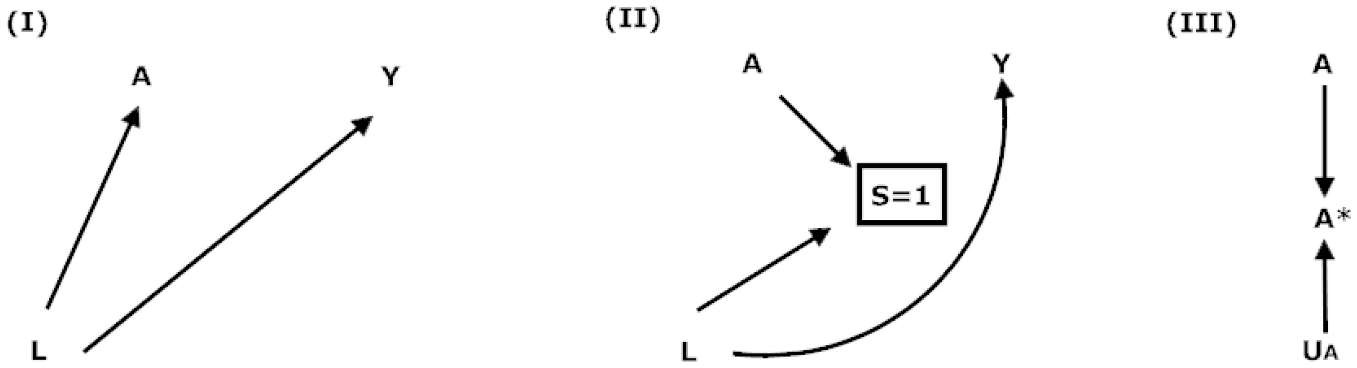

Figure.

Causal diagram depicting confounding, selection, and measurement bias in the absence of a true causal effect between an exposure (*A*) and outcome (*Y*)

Boxes represent restriction due to selection

Table 1

Source of and methods to reduce bias due to missing data on the individual-level potential outcomes

Bias type and source	Method	Applications of less commonly used methods published in the literature in the last 3 years
Confounding - missing individual-level potential outcomes for unobserved exposure levels	Randomization [1, 4], stratification [1, 4], restriction [1, 4], matching [1, 4], standardization [1, 4], standard regression adjustment [1, 4], propensity scores [24–26], inverse probability weighting [27, 28], instrumental variables [29–32], g-estimation [10–12], g-computation formula [35, 36], bayesian approaches [5, 7], and doubly robust estimators [39, 40]	Davies et al. [33] use an instrumental variable approach to estimate the effect of COX-2 selective nonsteroidal antiinflammtory drugs on the incidence of upper gastrointestinal complications and myocardial infarction. Swanson et al. [34] offer guidelines for how to report instrumental variable analyses using the Davies et al. [33] paper as an example. Naimi et al. [8] use g-estimation to estimate the cumulative effect of occupational asbestos exposure on time to lung cancer mortality with annotated SAS code provided in an earlier commentary [9]. Keil et al. [37] provide a simple introduction to the parametric g-formula with annotated SAS code for implementing the method and demonstrate its use when examining the effect of a hypothetical treatment to prevent graft-versus-host disease on mortality among bone marrow transplant patients. Neugebauer et al. [41] use doubly robust targeted minimum loss-based estimation with super learning to address confounding bias while examining the effect of various glucose-lowering strategies on albuminuria among adults with Type-2 diabetes and provide annotated R code for implementation.
Selection - missing individual-level potential outcomes among persons not selected	Standard regression adjustment [4, 15, 51], inverse probability weighting [4, 15, 51], redistribute-to-the-right algorithm [14], standardization [61], g-computation formula [35], multiple imputation [13], principal stratification [57], doubly robust estimators [41, 54], and instrumental variables [62–64]	Gottesman et al. [52] use inverse probability weighting as well as imputation to address potential selection bias due to death and loss to follow up when examining the effect of education on cognitive change. Neugebauer et al. [41] use doubly robust targeted minimum loss-based estimation with super learning to address selection bias while examining the effect of various glucose-lowering strategies on albuminuria among adults with Type-2 diabetes and provide annotated R code for implementation. Shardell et al. [54] use doubly robust augmented inverse probability weighted estimation to address selection bias due to death and lost to follow up when examining the effect of Vitamin D use on physical functioning among older adults. McGovern et al. [64] use an instrumental variable approach to correct for selection bias when estimating the prevalence of HIV among men in Ghana and Zambia.
Measurement - missing individual-level potential outcomes when exposure, outcome, or covariates are measured with error	Bias analysis [1, 50, 73–75], regression calibration [76, 77], modified maximum likelihood [78–80], multiple imputation [77, 81], and propensity score calibration [82–84]	See Funk and Landi [75] for recent published applications

Table 2

Missing individual-level potential outcomes by bias type given the consistency condition

Bias type	***S***	***A***	***A******	***Y***	$Y^{a=0}$	$Y^{a=1}$
Confounding - missing individual-level potential outcomes for unobserved exposure levels	1	1	1	1	?	1
	1	0	0	0	0	?
Selection - missing individual-level potential outcomes among persons not selected	0	?	?	?	?	?
	0	?	?	?	?	?
Measurement - missing individual-level potential outcomes when exposure, outcome, or covariates are measured with error	1	1	0	1	?	1
	1	0	1	0	0	?

S (binary indicator of selection); *A* (binary exposure); A^* (measured version of *A*); *Y* (outcome);

Y^a (potential outcome for *Y*; note for simplicity the superscripted *s* has been suppressed)