Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Chiba, J Biom Biostat 2012, 3:6
DOI: 10.4172/2155-6180.1000e113

Editorial Open Access

A Simple Method for Sensitivity Analysis of Unmeasured Confounding

Yasutaka Chiba*

Division of Biostatistics, Clinical Research Center, Kinki University School of Medicine, Japan

Unmeasured confounding is widely recognized as one of the principal problems faced by investigators conducting observational studies. Several sensitivity analysis techniques have been developed to handle unmeasured confounding [1-5]. Recently, VanderWeele and Arah [5] provided a general class of formulas for sensitivity analysis of unmeasured confounding. Their formulas benefit from the fact that they do not presuppose that any particular method is used to yield the initial estimate adjusted only for measured confounders. Three major methods to yield such an initial estimate are reviewed in the appendix. In this editorial, we describe a simple sensitivity analysis method that retains the advantages described above. The method has further advantages that there is only one sensitivity parameter and therefore the results can easily be displayed graphically, and computer programs to yield the initial estimate adjusted only for the measured confounders can be used for the sensitivity analysis without additional programming.

We use the following notation. Let A denote the exposure status of a particular individual. Suppose that A is dichotomous ($A = 1$ if exposed and $A = 0$ if unexposed). Let Y be the observed outcome of that individual. Let X denote a measured confounder or a set of measured confounders, and U denote an unmeasured confounder or a set of unmeasured confounders. We also consider the potential outcomes (or counterfactual) framework [6]. Let Y_a denote the potential outcome of Y for an individual if the exposure A, perhaps contrary to fact, had been set to value a. Using this notation, the causal effects with the total, exposed ($A = 1$), and unexposed ($A = 0$) groups as the target populations are provided by a comparison between $E(Y_1)$ and $E(Y_0)$, between $E(Y_1 \mid A = 1)$ and $E(Y_0 \mid A = 1)$, and between $E(Y_1 \mid A = 0)$ and $E(Y_0 \mid A = 0)$, respectively.

We assume that the potential outcome Y_a for an individual does not depend on the exposure status of other individuals. This assumption is sometimes referred to as the no-interference assumption [7]. Furthermore, we require the consistency assumption $Y_A = Y$, i.e., the value of Y that would have been observed if A had been set to its actual value is equal to the actually observed value of Y. Therefore, the only potential outcome for an individual that we observe is the potential outcome Y_A, i.e., the value of Y that would have been observed if A was set to its actual value. Finally, we suppose that the effect of A on Y is unconfounded given both X and U; in a counterfactual notation, Y_a is independent of A conditional on X and U.

To propose sensitivity analysis formulas of unmeasured confounding for difference measures, we apply the sensitivity parameter introduced by Brumback et al. [2], originally presented in the context of the inverse probability weighting approach [8]. This is represented by the following formula:

$$\delta_a \equiv E(Y_a \mid A = 1, X = x) - E(Y_a \mid A = 0, X = x),$$

where it is assumed that the value of δ_a does not vary between the strata of X. When $\delta_a > 0$, $E(Y_a \mid A = 1, X = x) > E(Y_a \mid A = 0, X = x)$, meaning that the individuals in the exposed group tend to have larger values than those in the unexposed group in the stratum with x. Conversely, when $\delta_a < 0$, $E(Y_a \mid A = 1, X = x) < E(Y_a \mid A = 0, X = x)$, meaning that the individuals in the exposed group tend to have smaller values than those in the unexposed group in the stratum with x. There is no confounding when $\delta_a = 0$.

Let $\hat{\Delta}_+$, $\hat{\Delta}_1$, and $\hat{\Delta}_0$ denote the average outcome differences adjusted only for X when the target populations are the total, exposed, and unexposed groups, respectively. Then, using the sensitivity parameter δ_a, the causal effects for the difference measures can be expressed as follows: For the causal effect with the total group as the target population,

$$E(Y_1) - E(Y_0) = \hat{\Delta}_+ - \delta,$$

where $\delta = \delta_1 \Pr(A = 0) + \delta_0 \Pr(A = 1)$ is a weighted mean of δ_0 and δ_1; for the exposed group,

$$E(Y_1 \mid A = 1) - E(Y_0 \mid A = 1) = \hat{\Delta}_1 - \delta_0;$$

and for the unexposed group,

$$E(Y_1 \mid A = 0) - E(Y_0 \mid A = 0) = \hat{\Delta}_0 - \delta_1.$$

Note that δ takes a value between δ_0 and δ_1.

These sensitivity analysis formulas indicate that the causal effects for the difference measures can simply be expressed as the difference between the initial estimate and a sensitivity parameter, and thus we can easily conduct a sensitivity analysis. The sensitivity parameter δ (δ_0 or δ_1) is set by the investigator according to what is thought to be plausible. The parameter can be varied over a range of plausible values to examine how conclusions vary according to different parameter values. To obtain the confidence intervals of the true causal effect for the fixed values of δ (δ_0 or δ_1), δ (δ_0 or δ_1) can be simply subtracted from the upper and lower confidence limits for the average outcome difference. Therefore, we can readily display the results of sensitivity analysis graphically, where the horizontal line represents the sensitivity parameter and the vertical line represents the true causal effect. However, for the total group as the target population, because δ depends on $\Pr(A = a)$, strictly, the variance of this probability should be taken into account to obtain the confidence intervals.

If we are sure that the individuals in the exposed group tend to have larger values than those in the unexposed group in the stratum with x, i.e., $E(Y_a \mid A = 1, X = x) \geq E(Y_a \mid A = 0, X = x)$, it would be assumed that $\delta_a \geq 0$. Conversely, if we are sure that $E(Y_a \mid A = 1, X = x) \leq E(Y_a \mid A$

***Corresponding author:** Yasutaka Chiba, Division of Biostatistics, Clinical Research Center, Kinki University School of Medicine, 377-2, Ohno-higashi, Osakasayama, Osaka 589-8511, Japan, Tel: +81-72-366-0221; Fax: +81-72-368-1193; E-mail: chibay@med.kindai.ac.jp

Received August 09, 2012; **Accepted** August 10, 2012; **Published** August 15, 2012

Citation: Chiba Y (2012) A Simple Method for Sensitivity Analysis of Unmeasured Confounding. J Biom Biostat 3:e113. doi:10.4172/2155-6180.1000e113

Target population	Average potential outcome	Approach		
		Model-based standardization	Inverse probability weighting	Doubly robust estimation
Total	$E(Y_1)$	$\frac{1}{n}\sum_{i=1}^{n} m_i(1)$	$\frac{1}{n}\sum_{i=1}^{n} \frac{y_i a_i}{p_i}$	$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i a_i}{p_i} - \frac{(a_i - p_i)m_i(1)}{p_i}\right] = \frac{1}{n}\sum_{i=1}^{n}\left[m_i(1) + \frac{a_i\{y_i - m_i(1)\}}{p_i}\right]$
	$E(Y_0)$	$\frac{1}{n}\sum_{i=1}^{n} m_i(0)$	$\frac{1}{n}\sum_{i=1}^{n} \frac{y_i(1-a_i)}{1-p_i}$	$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i(1-a_i)}{1-p_i} + \frac{(a_i - p_i)m_i(0)}{1-p_i}\right] = \frac{1}{n}\sum_{i=1}^{n}\left[m_i(0) + \frac{(1-a_i)\{y_i - m_i(0)\}}{1-p_i}\right]$
Exposed	$E(Y_1 \mid A=1)$	$\frac{1}{n_1}\sum_{i=1}^{n} a_i y_i$	$\frac{1}{n_1}\sum_{i=1}^{n} a_i y_i$	$\frac{1}{n_1}\sum_{i=1}^{n} a_i y_i$
	$E(Y_0 \mid A=1)$	$\frac{1}{n_1}\sum_{i=1}^{n} a_i m_i(0)$	$\frac{1}{n_1}\sum_{i=1}^{n} \frac{p_i y_i(1-a_i)}{1-p_i}$	$\frac{1}{n_1}\sum_{i=1}^{n}\left[\frac{p_i y_i(1-a_i)}{1-p_i} + \frac{(a_i - p_i)m_i(0)}{1-p_i}\right] = \frac{1}{n_1}\sum_{i=1}^{n}\left[a_i m_i(0) + \frac{p_i(1-a_i)\{y_i - m_i(0)\}}{1-p_i}\right]$
Unexposed	$E(Y_1 \mid A=0)$	$\frac{1}{n_0}\sum_{i=1}^{n}(1-a_i)m_i(1)$	$\frac{1}{n_0}\sum_{i=1}^{n}\frac{(1-p_i)y_i a_i}{p_i}$	$\frac{1}{n_0}\sum_{i=1}^{n}\left[\frac{(1-p_i)y_i a_i}{p_i} - \frac{(a_i - p_i)m_i(1)}{p_i}\right] = \frac{1}{n_0}\sum_{i=1}^{n}\left[(1-a_i)m_i(1) + \frac{(1-p_i)a_i\{y_i - m_i(1)\}}{p_i}\right]$
	$E(Y_0 \mid A=0)$	$\frac{1}{n_0}\sum_{i=1}^{n}(1-a_i)y_i$	$\frac{1}{n_0}\sum_{i=1}^{n}(1-a_i)y_i$	$\frac{1}{n_0}\sum_{i=1}^{n}(1-a_i)y_i$

Table 1: The estimators from the model-based standardization approach, inverse probability weighting approach, and doubly robust estimation.

$= 0, X = x$), it would be assumed that $\delta_a \leq 0$. Note that $\mathrm{E}(Y_a \mid A = 1, X = x) \geq \mathrm{E}(Y_a \mid A = 0, X = x)$ holds when both relationships between the unmeasured confounder-outcome and the unmeasured confounder-exposure are positive or negative in the stratum with x [9,10]. The reverse results are obtained when one of the relationships is positive and the other is negative. The other assumptions to derive a range of the sensitivity parameter are found elsewhere [11].

The above sensitivity analysis formulas for difference measures can be straightforwardly extended to ratio measures. Here, we assume that the outcome is binary; i.e., $\mathrm{E}(Y \mid A = a) = \Pr(Y = 1 \mid A = \mathrm{a})$ and $\mathrm{E}(Y_a) = \Pr(Y_a = 1)$. To propose the sensitivity analysis formulas for ratio measures, we introduce the following sensitivity parameter [4] instead of δ_a,:

$$\gamma_a \equiv E(Y_a \mid A = 1, X = x) / E(Y_a \mid A = 0, X = x),$$

where it is assumed that the value of γ_a, similar to δ_a, does not vary between the strata of X. Whether the value of γ_a is greater or less than 1 is interpreted in a similar manner to whether, the value of δ_a is larger or smaller than 0.

Let $\hat{\Gamma}_+$, $\hat{\Gamma}_1$, and $\hat{\Gamma}_0$ denote the average outcome ratios adjusted only for X when the target populations are the total, exposed, and unexposed groups, respectively. Then, using the sensitivity parameter γ_a, the causal effects for the ratio measures can be expressed as the ratio between the initial estimate and a sensitivity parameter. The formulas are follows: For the causal effect with the total group as the target population,

$$\mathrm{E}(Y_1)/\mathrm{E}(Y_0) = \hat{\Gamma}_+ / \gamma,$$

where

$$\gamma = \frac{\Pr(A=0)\mathrm{E}(Y \mid A=0) + \gamma_0 \Pr(A=1)\sum_x \mathrm{E}(Y \mid A=0, X=x)\Pr(X=x \mid A=1)}{\Pr(A=0)\mathrm{E}(Y \mid A=0) + \Pr(A=1)\sum_x \mathrm{E}(Y \mid A=0, X=x)\Pr(X=x \mid A=1)} \Big/ \frac{\frac{\Pr(A=0)}{\gamma_1}\sum_x \mathrm{E}(Y \mid A=1, X=x)\Pr(X=x \mid A=0) + \Pr(A=1)\mathrm{E}(Y \mid A=1)}{\Pr(A=0)\sum_x \mathrm{E}(Y \mid A=1, X=x)\Pr(X=x \mid A=0) + \Pr(A=1)\mathrm{E}(Y \mid A=1)}; \quad (1)$$

for the exposed group,

$$\mathrm{E}(Y_1 \mid A=1)/\mathrm{E}(Y_0 \mid A=1) = \hat{\Gamma}_1/\gamma_0;$$

and for the unexposed group,

$$\mathrm{E}(Y_1 \mid A=0)/\mathrm{E}(Y_0 \mid A=0) = \hat{\Gamma}_0/\gamma_1.$$

While the sensitivity analysis formulas for the exposed and unexposed groups are simple, the formula for the total group is complex and causes problems for the interpretation. However, when both γ_0 and γ_1 are greater (less) than 1, γ is also greater (less) than 1 and takes a value between the values of γ_0 and γ_1.

A sensitivity analysis for ratio measures can also be conducted easily, and the procedure is identical to that for difference measures. Using the log scale to obtain the confidence intervals of the true causal effect, a sensitivity parameter can be simply subtracted from the upper and lower confidence limits for the initial estimate. However, for the total group as the target population, strictly, the variances of estimators in (1) should be taken into account and it is troublesome to obtain the confidence intervals.

In this editorial, we have described a simple method for sensitivity analysis of unmeasured confounding in three target populations: total, exposed, and unexposed groups. The method can also be applied to the attributable fraction [12]. While the method described here has advantages mentioned at the beginning, it has a disadvantage that we need a strong assumption that the values of γ_a and γ_a do not vary between the strata of X. This assumption may not be reasonable in many actual studies. However, it is troublesome to set each value of the sensitivity parameters within each stratum of X, and further somewhat complex programming will be required to conduct a sensitivity analysis. In addition, the result of a sensitivity analysis with this assumption may not be largely different from that without the assumption, although the former always has the narrower confidence intervals than the latter.

Sensitivity analysis will aid in exploration of the potential impact of unmeasured confounding. We recommend performing a sensitivity analysis to evaluate the influence of unmeasured confounders on study results.

Appendix: Adjustment for Measured Confounding

We here introduce three approaches to adjust for the measured confounders: the model-based standardization approach, the Inverse Probability Weighting (IPW) approach, and the Doubly Robust (DR) estimation. The estimators from these three approaches are summarized in Table 1. In this table, $i = 1, \ldots, n$ denotes an individual and n is the total number; n_1 and n_0 are the number of individuals in the exposed and unexposed groups, respectively. $\mathrm{E}(Y \mid A = 1)$ and $\mathrm{E}(Y \mid A = 0)$ are the average outcomes for individuals in the exposed and unexposed groups, respectively.

The model-based standardization approach specifies a single model in which we simultaneously estimate the exposure-outcome association and the confounder-outcome association as follows:

$$\mathrm{E}(Y \mid A, X) = \alpha_0 + \alpha_1 A + \alpha_2 X_1 + \ldots + \alpha_{k+1} X_k ,$$

where $(\alpha_0, \ldots, \alpha_{k+1})$ is a set of regression parameters and can be estimated using standard software. Using this regression model, the expectations of the potential outcomes can be estimated as shown in Table 1, where $m_i(1) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 x_{1,i} + \ldots + \hat{\alpha}_{k+1} x_{k,i}$ is the predicted outcome given $A = 1$ for an individual i and $m_i(0) = \hat{\alpha}_0 + \hat{\alpha}_2 x_{1,i} + \ldots + \hat{\alpha}_{k+1} x_{k,i}$ is the predicted outcome given $A = 0$ for the individual i.

Rather than adjusting for the association between confounders and the outcome, we can control for confounding using the propensity score, which is defined as the conditional probability of exposure given confounders [13]. The propensity score is typically estimated from the observed data with a model such as the following:

$$\Pr(A = 1 \mid X) = \mathrm{expit}(\beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k),$$

where $\mathrm{expit}(s) = \exp(s)/\{1 + \exp(s)\}$, and where $(\beta_0, \ldots, \beta_k)$ is a set of regression parameters that can be estimated by standard software. Using this regression model, the IPW approach estimates the expectations of the potential outcomes as in Table 1, where $p_i = \mathrm{expit}(\hat{\beta} + \hat{\beta}_1 x_{1,i} + \ldots + \hat{\beta}_k x_{k,i})$ [8]. Note that Sato and Matsuyama [14] exemplified an SAS code to yield the IPW estimate using the marginal structural model.

The DR estimation requires specification of two regression models for the outcome and exposure as a function of confounders. Having estimated $m_i(a)$ and p_i, we combine these values as in Table 1 to calculate the DR estimates. These expressions suggest an intuitive explanation of the properties of DR. The formulas of the left hand side indicate that the DR estimators are equivalent to the unbiased estimators from the IPW approach if the exposure regression model is correctly specified. Likewise, the formulas of the right hand side indicate that the DR estimators are equivalent to the unbiased estimators from the model-based standardization approach if the outcome regression model is correctly specified. Note that Funk et al. [15] presented an SAS macro to yield the DR estimate with the total group as the target population.

Acknowledgements

This work was supported partially by Grant-in-Aid for Scientific Research (No. 23700344) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. Greenland S (2003) The impact of prior distributions for uncontrolled confounding and response bias: a case study of the relation of wire codes and magnetic fields to childhood leukemia. J Am Stat Assoc 98: 47-54.
2. Brumback BA, Hernán MA, Haneuse SJ, Robins JM (2004) Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat Med 23: 749-767.
3. McCandless LC, Gustafson P, Levy A (2007) Bayesian sensitivity analysis for unmeasured confounding in observational studies. Stat Med 26: 2331-2347.
4. Chiba Y (2010) Sensitivity analysis of unmeasured confounding for the causal risk ratio by applying marginal structural models. Commun Stat Theory Methods 39: 65-76.
5. Vanderweele TJ, Arah OA (2011) Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. Epidemiology 22: 42-52.
6. Little RJ, Rubin DB (2000) Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu Rev Public Health 21: 121-145.
7. Cole SR, Hernán MA (2008) Constructing inverse probability weights for marginal structural models. Am J Epidemiol 168: 656-664.
8. Robins JM, Hernán MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. Epidemiology 11: 550-560.
9. Vanderweele TJ (2008) The sign of the bias of unmeasured confounding. Biometrics 64: 702-706.
10. Chiba Y (2009) The sign of the unmeasured confounding bias under various standard populations. Biom J 51: 670-676.
11. Chiba Y, Sato T, Greenland S (2007) Bounds on potential risks and causal risk differences under assumptions about confounding parameters. Stat Med 26: 5125-5135.
12. Chiba Y (2012) Sensitivity analysis for unmeasured confounding of attributable fraction. Epidemiology 23: 175-176.
13. Joffe MM, Rosenbaum PR (1999) Invited Commentary: Propensity scores. Am J Epidemiol 150: 327-333.
14. Sato T, Matsuyama Y (2003) Marginal structural models as a tool for standardization. Epidemiology 14: 680-686.
15. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, et al. (2011) Doubly robust estimation of causal effects. Am J Epidemiol 173: 761-767.