Prev Sci (2013) 14:570–580
DOI 10.1007/s11121-012-0339-5

An Introduction to Sensitivity Analysis for Unobserved Confounding in Nonexperimental Prevention Research

Weiwei Liu · S. Janet Kuramoto · Elizabeth A. Stuart

Published online: 14 February 2013

Abstract Despite the fact that randomization is the gold standard for estimating causal relationships, many questions in prevention science are often left to be answered through nonexperimental studies because randomization is either infeasible or unethical. While methods such as propensity score matching can adjust for observed confounding, unobserved confounding is the Achilles heel of most nonexperimental studies. This paper describes and illustrates seven sensitivity analysis techniques that assess the sensitivity of study results to an unobserved confounder. These methods were categorized into two groups to reflect differences in their conceptualization of sensitivity analysis, as well as their targets of interest. As a motivating example, we examine the sensitivity of the association between maternal suicide and offspring's risk for suicide attempt hospitalization. While inferences differed slightly depending on the type of sensitivity analysis conducted, overall, the association between maternal suicide and offspring's hospitalization for suicide attempt was found to be relatively robust to an unobserved confounder. The ease of implementation and the insight these analyses provide underscores sensitivity analysis techniques as an important tool for nonexperimental studies. The implementation of sensitivity analysis can help increase confidence in results from nonexperimental studies and better inform prevention researchers and policy makers regarding potential intervention targets.

Keywords Sensitivity analysis · Causal inference · Unobserved confounding · Suicide prevention

Electronic supplementary material The online version of this article (doi:10.1007/s11121-012-0339-5) contains supplementary material, which is available to authorized users.

W. Liu (✉)
NORC at the University of Chicago,
Bethesda, MD, USA
e-mail: liu-weiwei@norc.org

S. J. Kuramoto
American Psychiatric Institute for Research and Education,
Arlington, VA, USA

E. A. Stuart
Department of Mental Health and Department of Biostatistics,
Bloomberg School of Public Health, Johns Hopkins University,
Baltimore, MD, USA

This paper describes and illustrates a set of tools, broadly known as "sensitivity analysis," to help understand the robustness of nonexperimental findings to a potential unobserved confounder. The goal of this paper is to discuss the relevance of common sensitivity analysis methods to prevention research and to provide an easy-to-understand guide for interested prevention researchers. We do not intend to discuss the technical nuances of sensitivity analysis or to provide a comprehensive listing of all of the methods used to assess sensitivity. However, interested readers can refer to works cited throughout the paper for details and description of additional methods.

Many studies in prevention research aim to investigate causal effects—either the effects of early risk factors or experiences on later outcomes, such as the effects of adolescent drug use on outcomes such as unemployment and drug use during adulthood (Stuart and Green 2008) or the effects of particular interventions, such as the Good Behavior Game (Kellam et al. 2008). Understanding such causal effects is crucial for determining what risk or protective factors should be targeted to improve outcomes or whether a preventive intervention is effective. In some situations, researchers can randomize individuals to receive the intervention or the control condition, a method considered the gold standard for estimating causal effects (Flay et al. 2005). However, randomization is often infeasible or unethical, sometimes due to the nature of the factor under study (such as drug abuse or childhood maltreatment). In those cases, questions of causal inference are left to be answered using nonexperimental methods.

The Achilles heel of nonexperimental studies is that the exposed and unexposed (or treatment and control) groups may not be comparable, a phenomenon formally known as confounding (Rothman et al. 2008). While observed confounding can be addressed with methods, such as propensity score matching (Stuart 2010), methods to assess the consequences of unobserved confounding are less readily available, and researchers tend to shy away from the issue. Being able to determine the robustness of study findings to potential unobserved confounding is crucial for theory testing and development. In addition, the knowledge of to what extent the conclusions drawn from those studies are robust to potential unobserved confounding can help with policy decision making.

The origin of sensitivity analysis to unobserved confounding has been attributed to a study by Cornfield et al. (1959), which quantified the role of unobserved confounding in the observed relationship between smoking and lung cancer. Specifically, Cornfield et al. (1959) showed that, an unobserved confounder, such as a genetic factor, would need to lead to a ninefold increase in the odds of smoking in order to explain away the association between smoking and lung cancer, and it was asserted that such a strong unobserved confounder was very unlikely to exist. Since then, more recent examples demonstrate the use of sensitivity analysis to assess unobserved confounding in various fields, such as sociology, criminology, psychology, and prevention science (e.g., DiPrete and Gangl 2004; Harding 2003; Haviland et al. 2007; Kitahata et al. 2009; Liu 2012), though such examples are still relatively rare. For example, Haviland et al. (2007) found that the estimated effect of gang involvement on subsequent violence was robust to a fairly weak unobserved confounder but may be sensitive to an unobserved confounder moderately associated with gang involvement and subsequent violence. Harding (2003) found that the estimated effect of neighborhood choice on high school dropout was robust to an unobserved confounder if high and low poverty neighborhoods were compared. However, the effect was sensitive to an unobserved confounder if high and moderate poverty neighborhoods were compared or if moderate and low poverty neighborhoods were compared. As a final example, Liu (2012) found that an observed relationship between high school dropout and subsequent adult criminal offending estimated using propensity score matching remained significant even after taking into account an unobserved confounder strongly associated with both high school dropout and criminal offending.

Conceptually, sensitivity analysis can be understood from both statistical and epidemiological perspectives (Luiz and Cabral 2010). Although the two perspectives differ in their approaches to conceptualize sensitivity analysis, they both assess how strong the effects of the unobserved covariate on the exposure and/or the outcome would have to be to change the study inference. From a statistical perspective, Rosenbaum (2002, 2010) emphasizes the difference between randomized trials and nonexperimental studies; that is, in a nonexperimental study, exposed and unexposed groups may differ on an unobserved characteristic even after matching on observed characteristics. In other words, individuals with the same observed covariates may have different probabilities of being exposed if they have different unobserved covariates. A sensitivity parameter is used to quantify the difference in the odds of exposure for two individuals with the same observed covariates (or the same propensity score) but diverge on unobserved covariates. The goal is to determine the smallest value of this parameter that will change the p value of the "true" outcome–exposure association to a nonsignificant level.

Sensitivity analysis can also be understood from an epidemiological perspective (e.g., Harding 2003) as a tool to assess the extent to which a significant association found between observed variables could be due to unobserved confounding. Sensitivity parameters are used to quantify the strengths of the associations between a hypothetical unobserved confounder and the exposure and outcome. The goal is to arrive at a "true" association between the exposure and the outcome, adjusting for the hypothetical unobserved confounder with various values of the sensitivity parameters.

We selected seven methods to present in this paper. In selecting these methods, we attempted to keep a balance between introducing a broad range of methods that accommodate different interests and data availability and selecting the methods that are relatively straightforward to understand and easy to apply to a number of different settings. For example, the selected methods capture the two different perspectives previously mentioned to conceptualize sensitivity analysis. Additionally, these seven methods have different targets of interest. Specifically, the first group, i.e., Rosenbaum's three approaches (Gastwirth et al. 1998) focus on the statistical significance of the "true" outcome–exposure association, while the second group, i.e., the approaches of Greenland (1996), Harding (2003), Lin et al. (1998), VanderWeele and Arah (2011), and Arah et al. (2008) obtain the point estimate of the "true" outcome–exposure association with a 95 % confidence interval (CI). Other differences between these approaches (such as study design and outcome distribution) are described later in the manuscript and summarized in Table 1. Importantly, the seven selected methods are all relatively straightforward to understand and can be computed by hand or with standard statistical software.

Motivating Study

As a demonstration, we apply the seven sensitivity analysis methods to investigate potential unobserved confounding in

Table 1 Summary of sensitivity analysis

	Rosenbaum			$OR_{yx \times cu}$ with CI			
	Primal	Dual	Simultaneous	Greenland	Harding	Lin et al.	VanderWeele and Arah
Target of interest	Nonsignificance of test statistic			$OR_{yx \times cu}$ with CI			
Study design	1:1 matched pairs			Any study design			
Outcome distribution	Any			Binary	Binary	Binary, continuous, or censored	Any
Parameters obtained from the data	Number of discordant pairs			2×2 joint cells from x and y	2×2 joint cells from x and y	$OR_{yx \times c}$	$OR_{yx \times c}$
Parameters set by the method	OR_{yu}=infinity	OR_{xu}=infinity	None	None	None	None	None
Parameters explicitly set by the user	OR_{xu}	OR_{yu}	OR_{yu} OR_{xu}	OR_{yu} OR_{xu} $p(u\|x=0)$	OR_{yu} OR_{xu}	$OR(yu\|x=1)$ $OR(yu\|x=0)$ $p(u\|x=1)$ $p(u\|x=0)$	OR_{yu} $p(u\|x=1)$ $p(u\|x=0)$
Pros	Reflects uncertainty given sample size Easy to implement Does not require the specification of the prevalence of u Gives a conservative result		Can vary both OR_{yu} and OR_{xu}	Can vary both OR_{yu} and OR_{xu} Yields CI Can accommodate any study design Do not need the raw data to conduct sensitivity analysis Easier to implement (e.g., compared to Harding's method)	Does not require the specification of the prevalence of u	Easy to implement Can accommodate more general settings Allows three way interaction between x, y, and u	Easy to implement Can accommodate more general settings Can accommodate the violation of assumptions
Cons	CI for the estimate not easily available Implementation limited to matching design Requires the number of discordant pairs, which may not always be readily available			Need to specify $p(u\|x=1)$ or $p(u\|x=0)$	More involved computation	Need to specify $p(u\|x=1)$ or $p(u\|x=0)$	Need to specify $p(u\|x=1)$ or $p(u\|x=0)$
Software for implementation	"rbounds" in Stata or R Love's Excel spreadsheet	Hand computation (or with standard statistical software)	Hand computation (or with standard statistical software)	Hand computation (or with standard statistical software) Regression analysis	System equation solver (or with standard statistical software) Regression analysis	Hand computation (or with standard statistical software)	Hand computation (or with standard statistical software)

a recent study by Kuramoto et al. (2010), which examined the association between maternal suicide and their offspring's hospitalization for suicide attempt. Parental suicide has been examined as a risk factor for adolescent's suicide attempt (Wilcox et al. 2010). However, the association between parental suicide and offspring's risk has been somewhat equivocal, with most of the research in this area coming from short-term prospective or cross-sectional studies of referred samples. Additionally, most previous studies compared offspring of suicide decedents with offspring of living parents, which cannot clarify our understanding on the impact of parental suicide over and beyond the impact of sudden parental death. Understanding the causal mechanisms, and not just associations, is particularly crucial to suicide prevention to identify populations in which prevention efforts can be targeted. For example, if parental suicide was found to indeed cause offspring's suicide attempt, then prevention efforts should target individuals whose parents died from suicide. If it was found that such a relationship can be easily explained away by some other factors, it is important to search for those factors that may better explain hospitalization for suicide attempt. However, we cannot randomly assign parental suicide, leaving the investigation of such a relationship relying solely on nonexperimental studies. With such a goal, Kuramoto et al. (2010) compared 5,600 offspring who lost a mother to suicide before age 18 (exposed group) with 2,872 offspring who lost a mother to an accident (unexposed group). This comparison group allowed a clearer distinction of risk, over and beyond the stress and disruption associated with sudden parental death. Propensity score matching was used to make the exposed and unexposed groups as similar as possible on observed characteristics, such as deceased parent and surviving parent's psychiatric hospitalization. One to one nearest neighbor propensity score matching with replacement (Stuart 2010) was used. The authors concluded that maternal suicide was associated with a 1.86-fold increased risk (95 % CI=1.49, 2.32)[1] of their offspring being hospitalized for suicide attempt, as compared to matched offspring who lost their mother to an accident, a result in congruent with the increasing body of literature on the impact of parental suicide on offspring's risk for suicidal behavior (Wilcox et al. 2010; Niederkrotenthaler et al. 2012).

While the exposed and unexposed groups were matched on a large number of characteristics, potential unobserved confounders remain a concern. For example, the observed association may be partly explained by genetic predisposition to suicidal behavior, which has been suggested to be associated with both suicide and offspring's suicide attempt (Brent and Mann 2005; Lieb et al. 2005). In order to accommodate some methods (e.g., primal sensitivity analysis) that are most easily applied to a 1:1 matched setting and ease the comparison of these methods with other methods, the 1:1 matching with replacement was modified to resemble a 1:1 match without replacement. This resulted in 5,600 matched pairs, with a total sample size of 11,200. In these pairs, 233 offspring of suicide decedents and 128 offspring of accident decedents were hospitalized for suicide attempt. This modification did not change the inferences about the association between maternal suicide and offspring's suicide attempt.

[1] The original study estimated hazard ratio of 1.80 with a 95 % confidence interval of 1.19, 2.74.

Setting, Assumptions, and Notation

In order to demonstrate and motivate the use of sensitivity analysis, we focus on a relatively simple setting with a binary exposure, a binary outcome, and a binary unobserved confounder. The focus on binary exposures and outcomes is common in the causal inference literature (Stuart 2010), helps simplify the sensitivity analysis techniques, and is a common assumption in that setting (e.g., VanderWeele and Arah 2011; Harding 2003). The binary unobserved confounder can also be thought of as a combination of a number of unobserved confounders (Lin et al. 1998). Some of the methods discussed in this paper can be generalized to accommodate continuous normally distributed unobserved confounders, although the computation is more involved. The basic ideas remain the same, except that, when the confounder is continuous, the relationships between the unobserved confounder and the exposure and the outcome are expressed as mean differences rather than odds ratios (ORs) (a brief discussion is provided later in this paper). In addition, some of these methods can be generalized to accommodate continuous or censored outcomes (e.g., the approach of Lin et al.; see Table 1).

Two common assumptions are made (though not necessarily required by every method, as discussed later in the paper) to apply the seven methods to the motivating example. First, we assume that the relationships between the unobserved confounder and the exposure and the outcome do not vary as a function of the observed covariates. As VanderWeele and Arah (2011) discussed when presenting their more general approach, it becomes virtually impossible to allow the specified parameters to differ across levels of the observed covariates when multiple observed covariates are involved. Second, we assume no three-way interaction between the exposure, the outcome, and the unobserved confounder, an assumption made by most studies using sensitivity analysis (e.g., Harding 2003). While of course these two assumptions are not always met in reality, demonstrating methods to accommodate violation of these assumptions is beyond the scope of this paper. A brief discussion is provided later in this paper; for more detailed

discussions of those approaches, see VanderWeele and Arah (2011) and Lin et al. (1998).

The following notation will be used throughout the paper:

x	Binary treatment status/exposure
y	Binary outcome
u	Unobserved binary confounder
c	Observed confounders
$p(x)$	Prevalence of the exposure
$p(u)$	Prevalence of the unobserved confounder
$p(u\|x=1)$	Prevalence of the unobserved confounder among the exposed group
$p(u\|x=0)$	Prevalence of the unobserved confounder among the unexposed group
OR_{yu}	OR of the relationship between the outcome and unobserved confounder
OR_{xu}	OR of the relationship between the exposure and unobserved confounder
$OR_{yx\times c}$	Observed OR of the relationship between the outcome and the exposure from the data, adjusted for c (but not for u)
$OR_{yx\times cu}$	True (bias-free/bias-adjusted) OR of the relationship between the outcome and the exposure, adjusted for both c and u

Generally, when performing sensitivity analyses, researchers should specify a range of parameter values that are suggested by the literature or based on the relationships between observed confounders and the exposure and outcome (e.g., Harding 2003) to examine the sensitivity of study inferences under different specifications. This approach is particularly useful when some parameters require outside knowledge that is not easily obtained. However, for demonstration purposes and to make the results from different methods comparable, we specify a single set of parameters to be used in the motivating example. Since the relationship between the hypothetical unobserved confounder (u; such as genetic predisposition) and the exposure (OR_{xu}) and the outcome (OR_{yu}) (net of all the covariates matched on) were not readily available in the literature, they were obtained by examining the relationships of the observed confounders (c) available in the motivating study with the exposure (having a mother die of suicide; OR_{xc}) and the outcome (offspring's hospitalization for suicide attempt; OR_{yc}). The ORs between the observed confounders and the exposure ranged from 0.98 to 7.39, with the strongest factor being the deceased parent's history of psychiatric hospitalization (OR_{xc}=7.39). The ORs between the observed confounder and the outcome (y) ranged from 0.95 to 1.84, with the strongest factor being the psychiatric hospitalization of the surviving parent prior to the death of the parent (OR_{yc}= 1.84). To err on the conservative side, we fixed the values of OR_{xu} and OR_{yu} at these two highest values of the observed ORs (i.e., OR_{xu}=7.39; OR_{yu}=1.84). Some approaches require the specification of $p(u|x=1)$and $p(u|x=0)$ instead of OR_{xu}. The prevalence of an unobserved confounder such as genetic predisposition to suicidal behavior in the general population is not available; hence, we specified a range of $p(u|x=0)$ from 1 to 25 % with the fixed OR_{xu}=7.39 to obtain varying $p(u|x=1)$. In the remainder of the paper, we introduce each of the seven methods and apply these methods to the motivating example. Details of the computations, sample R code, and links to relevant Excel spreadsheets and web equation solvers can be found in the eAppendix.

Rosenbaum's Approaches

Rosenbaum's approaches, in general, are interested in finding the thresholds of the association(s) between the unobserved confounder and the exposure (OR_{xu}) and/or between the unobserved confounder and the outcome (OR_{yu}) that would render the test statistics of the study inference ($OR_{yx\times cu}$) insignificant. This method is most frequently used when the observed confounders have been dealt with using matching methods (such as propensity score matching) that form matched pairs of exposed and unexposed individuals who are similar on the observed covariates. These approaches are further broken down into primal, dual, and simultaneous analyses (Gastwirth et al. 1998), which differ in their specified parameters. Primal sensitivity analysis varies the association between the unobserved confounder and the exposure (OR_{xu}, with an upper bound denoted Γ), while setting OR_{yu} at infinity. In contrast, dual sensitivity analysis varies the association between the unobserved confounder and the outcome (OR_{yu}, with an upper bound denoted Δ), while setting OR_{xu} at infinity. Simultaneous sensitivity analysis varies both OR_{xu} and OR_{yu}. The primal and dual sensitivity analyses are, therefore, special cases of the simultaneous sensitivity analysis. We focus our further discussion on primal and simultaneous sensitivity analyses, as the steps involved in dual sensitivity analysis are similar to primal sensitivity analysis, except that the parameter that is varied is OR_{yu} instead of OR_{xu}.

Primal Sensitivity Analysis

OR_{xu} estimates are bounded by $\Gamma : \frac{1}{\Gamma} \leq OR_{xu} \leq \Gamma$, where $\Gamma \geq 1$.[2] The upper ($p+$) and lower ($p-$) bounds on the probability of being exposed, accounting for u, can then be calculated. In particular, $p+$ is of most interest and can be expressed as $\frac{\Gamma}{1+\Gamma}$.

A modified McNemar's exact test (McNemar 1947) is then used to examine the association between x and y, accounting for u (i.e., by computing an upper-bound p value using $p+$ instead

[2] Note that this is not a loss of generality; if the unobserved confounder is negatively associated with exposure status, we could simply redefine the unobserved confounder to meet this scenario.

of the observed probability of exposure of 0.5 in the matched pairs). In Eq. 1, T denotes the total number of discordant pairs (those where the outcomes differ within the pair) and a denotes the number of discordant pairs, in which the exposed had an outcome and the unexposed did not. This is repeated with different values of Γ to find the value of Γ at which the upper-bound p value becomes nonsignificant (e.g., $p>0.05$). A higher value of Γ required to render the upper-bound p value nonsignificant is preferred, as it indicates that $OR_{yx\times cu}$ is more robust to unobserved bias, i.e., a stronger association between the unobserved confounder and the exposure is necessary for the $OR_{yx\times cu}$ to become nonsignificant.

$$\text{upper-bound p value} = \sum_{a}^{T}\binom{T}{a}(p^{+})^{a}(1-p^{+})^{T-a} \quad (1)$$

The primal sensitivity analysis for a 1:1 matched design with a binary outcome can be implemented by hand, using the "rbounds" package in R (Keele 2010) or using an available Excel spreadsheet (Love 2008; for details, see the eAppendix). These tools can also be used for continuous outcomes; for example, by using a function written for Stata (Gangl 2004). Although generalization of this technique to study designs beyond 1:1 matches is possible (e.g., Rosenbaum 2002; Keele 2010), it is not as easily implemented.

Simultaneous Sensitivity Analysis

The simultaneous sensitivity analysis allows researchers to vary not only OR_{xu} (with an upper-bound Γ) but also OR_{yu} (with an upper-bound Δ). The goal is to find the combinations of Γ and Δ at which $OR_{yx\times cu}$ becomes statistically nonsignificant. The steps are similar to primal sensitivity analysis. One first specifies values for Γ and Δ, which can be used to calculate the upper and lower bounds of the probability of being exposed given the unobserved confounder ($p+$ and $p-$). In particular, $p+$ can be calculated using Eq. 2, where $p(\theta)=\frac{\Delta}{(1+\Delta)}$ and $p(\pi)=\frac{\Gamma}{(1+\Gamma)}$. We then use $p+$ and the numbers of discordant pairs to calculate the upper-bound p value of $OR_{yx\times cu}$ using McNemar's exact test.

$$p+ = p(\pi)\times p(\theta) + (1-p(\pi))\times(1-p(\theta)) \quad (2)$$

A combination of values of Δ and Γ for which the test statistic becomes nonsignificant is a point at which the result is sensitive to an unobserved confounder. Although there is no specific package available, this analysis can be computed by hand or easily programmed using Excel or R for a 1:1 matched pair design (for details, see the eAppendix).

Application of Rosenbaum's Approaches

Primal Sensitivity Analysis

Using the counts of discordant pairs, i.e., $a=226$ and $T=347$, and different values of Γ, four methods (hand computation, "rbounds" package in R, computation using R codes, and Love's spreadsheet) were used for the analysis. The results suggest that when $\Gamma\geq 1.55$, the association between maternal death by suicide and offspring's hospitalization for suicide attempt would no longer be significant (with a p value of 0.054).

Simultaneous Sensitivity Analysis

Using the counts of discordant pairs, as well as the specified parameters ($\Gamma=7.39$, $\Delta=1.84$), we computed $p+$ and the upper-bound p value by hand and by using R codes. Results suggest moderate sensitivity of the study inference to an unobserved confounder, as when Γ is 7.39 and Δ is 1.84, $OR_{yx\times cu}$ is no longer significant (with a p value of 0.08).

Methods to Obtain $OR_{yx\times cu}$ with Confidence Interval

While Rosenbaum's approaches were primarily concerned with finding the point at which effects became nonsignificant, the following methods quantify the unobserved confounder under certain specifications and then arrive at an estimate of the target of interest, $OR_{yx\times cu}$ (i.e., the true relationship between x and y), and an associated CI, adjusting for the unobserved confounder. Two groups of methods are presented. The first group (Greenland's and Harding's approaches) utilizes the association between x, y, and u to create the actual data as if u was observed, which is then used to estimate the $OR_{yx\times cu}$. The second group (the approaches of Lin et al. and VanderWeele and Arah) utilizes the association between x, y, and u to compute an adjustment or bias factor that is then used to obtain the $OR_{yx\times cu}$.

Greenland's and Harding's Approaches

Both approaches break down the observed 2×2 table of x and y into eight combinations of x, y, and u, imagining the data that would be observed if u was observed. Table 2 presents the observed 2×2 cross-tabulation of x and y. Table 3 presents the 2×2 cross-tabulation of x and y stratified by u, which we would see if u was observed. The goal is to estimate the cell counts a–h by specifying aspects of u so that Table 3 can be recreated.

Greenland's and Harding's approaches slightly differ in the necessary parameter inputs and in the manner in which

Table 2 Observed data: 2×2 table for x and y

	$y=0$	$y=1$
$x=0$	$a+e$	$b+f$
$x=1$	$c+g$	$d+h$

Table 3 Underlying true data: 2×2 table for x and y controlling for u

		y=0	y=1
u=0	x=0	a	b
	x=1	c	d
u=1	x=0	e	f
	x=1	g	h

these a–h cell counts are obtained. In his paper, Harding (2003) set u to be evenly distributed in the population (i.e., $p(u)$=0.5), but this can be easily modified. Harding's approach requires the specification of OR_{yu} and OR_{xu}. Given those values and the observed data ($(a+e)$, $(e+f)$, $(c+g)$, $(d+h)$), these eight cell counts (a–h) can be solved using Eq. 3.

$$af/be = \mathrm{OR_{yu}} \quad ch/dg = \mathrm{OR_{yu}} \quad ag/ce = \mathrm{OR_{xu}} \quad bh/fd = \mathrm{OR_{xu}}$$
$$a+b+c+d = e+f+g+h \tag{3}$$

Instead of specifying OR_{yu} and OR_{xu}, Greenland (1996) specifies the prevalence of the unobserved confounder in the unexposed individuals ($p(u|x=0)$) and the exposed individuals ($p(u|x=1)$).

Greenland's approach then finds the cell counts e–h using the values specified by the user ($p(u|x=1)$, $p(u|x=0)$, and OR_{yu}) through Eq. 4.

$$g = p(u|x=1)\times(c+g) \quad e = p(u|x=0)\times(a+e)$$
$$h = \frac{\mathrm{OR_{yu}}\times(\mathrm{d+h})\times\mathrm{g}}{\mathrm{OR_{yu}}\times\mathrm{g}+(\mathrm{c+g})-\mathrm{g}} \quad f = \frac{\mathrm{OR_{yu}}\times(\mathrm{b+f})\times\mathrm{e}}{\mathrm{OR_{yu}}\times\mathrm{e}+(\mathrm{a+e})-\mathrm{e}} \tag{4}$$

Instead of specifying $p(u|x=1)$ and $p(u|x=0)$, an alternative approach is to specify OR_{xu} and $p(u|x=0)$ and then calculate the implied $p(u|x=1)$ using Eq. 5.

$$p(u|x=1) = \frac{\mathrm{OR_{xu}}\times\mathrm{p(u|x=0)}}{1-p(u|x=0)+\mathrm{OR_{xu}}\times\mathrm{p(u|x=0)}} \tag{5}$$

Once we know four of the eight cell counts, the values of a–d can be easily obtained from the observed data and the values of e–h using simple algebra.

For both Harding's and Greenland's approaches, $OR_{yx\times cu}$ is then estimated by using these cell counts as frequency weights to recreate a dataset that contains information on the unobserved confounder. From this recreated data, a weighted logistic regression can be performed to obtain $OR_{yx\times cu}$, including its CI (Harding 2003). It is also possible to supplement the original data by explicitly creating the number of observations that reflect each cell count and then performing regular logistic regression.

The Approaches of Lin et al. and VanderWeele and Arah

Although the approaches of Lin et al. (1998), VanderWeele and Arah (2011), and Arah et al. (2008) also aim to arrive at the $OR_{yx\times cu}$ with a CI, they use a slightly different approach than Greenland's and Harding's approaches previously discussed. It is important to note that the approaches of Lin et al. and VanderWeele and Arah both allow relaxing the no-three-way-interaction assumption. In addition, VanderWeele and Arah's approach allows the relationships between the unobserved confounder and the exposure and outcome to vary as a function of observed covariates. To ease the comparison between these methods to the other methods previously presented, we maintain these assumptions in the motivating example.

In the approach of Lin et al., the information about the relationship between y and u is first summarized by an equation that estimates an adjustment factor (AF) derived from the following parameters specified by the researcher: $p(u|x=1)$, $p(u|x=0)$, $OR(yu|x=1)$, and $OR(yu|x=0)$ using Eq. 6.

$$\mathrm{AF} = \frac{\mathrm{OR(yu|x=1)}\times\mathrm{p(u|x=1)}+(1-\mathrm{p(u|x=1)})}{\mathrm{OR(yu|x=0)}\times\mathrm{p(u|x=0)}+(1-\mathrm{p(u|x=0)})} \tag{6}$$

When assuming no three-way interaction, i.e., $OR(yu|x=1)=OR(yu|x=0)$, the equation is simplified to Eq. 7.

$$\mathrm{AF} = \frac{\mathrm{OR_{yu}}\times\mathrm{p(u|x=1)}+(1-\mathrm{p(u|x=1)})}{\mathrm{OR_{yu}}\times\mathrm{p(u|x=0)}+(1-\mathrm{p(u|x=0)})} \tag{7}$$

$OR_{yx\times cu}$ can then be calculated by dividing $OR_{yx\times c}$ by AF. The CI for $OR_{yx\times cu}$ can be obtained using the same AF and dividing the upper and lower bounds of the CI for $OR_{yx\times c}$ by the AF.

VanderWeele and Arah (2011) and Arah et al. (2008) proposed a more general framework to assess sensitivity to an unobserved confounder, which can accommodate continuous or categorical outcomes, exposures and observed and unobserved confounders, while also relaxing the two assumptions discussed in the beginning of this paper. Briefly, they suggest that the bias for any outcome distribution can be estimated as long as one can specify two conditions for each level of the observed confounders, c: (1) the relationship between u and y across different levels of x (u' denotes a chosen reference value for u), e.g., $\{E(y|x=x_1,u=u,c)-E(y|x=x_1,u=u',c)\}$ and $\{E(y|x=x_2,u=u,c)-E(y|x=x_2,u=u',c)\}$, and (2) the comparison between the prevalence of u when x is at different levels with the overall prevalence of u set, e.g., $p(u|x=x_1,c)-p(u|c)$ and $p(u|x=x_2,c)-p(u|c)$. For formulas to apply this method, interested readers can refer to VanderWeele and Arah (2011). While attractive in its generality, VanderWeele and Arah (2011) also discuss the challenges in using such a general setting as it requires a large number of parameters to be specified. They recommended simplifying the approach for specific settings. Applying this general method to our specific setting (i.e., x, y, and u are all binary) and assumptions (i.e., (1) the

relationships between y and u and between x and u are the same across different levels of c; (2) no three-way interaction between x, y, and u), we can estimate the bias in the $OR_{yx\times c}$ using the following simple Eq. 8, which is essentially equivalent to Eq. 7 in the approach of Lin et al.:

$$\text{bias} = \frac{1 + (OR_{yu} - 1)p(u|x = 1)}{1 + (OR_{yu} - 1)p(u|x = 0)} \quad (8)$$

Similar to the approach of Lin et al., we can then divide the $OR_{yx\times c}$ by this bias term to obtain $OR_{yx\times cu}$, as well as its CIs.

Application of Methods to Obtain $OR_{yx\times cu}$ with Confidence Interval

As a reminder, to ease the comparison between methods, we fixed the values of OR_{xu} and OR_{yu} at the two highest observed values (OR_{xu}=7.39; OR_{yu}=1.84). For methods that require the prevalence of the unobserved confounder, we specified a range of $p(u|x=0)$ from 1 to 25 % and then obtained $p(u|x=1)$ using $p(u|x=0)$ and OR_{xu}=7.39 (for detailed computation and sample R code, see the eAppendix). Results are summarized in Table 4 for the range of $p(u|x=0)$ specified. Similar results were observed across different methods, with $OR_{yx\times cu}$ estimates decreasing with increasing $p(u|x=0)$. However, the CI never included one, suggesting that the study inference is not sensitive to an unobserved binary confounder that is associated with sevenfold increased odds of having a mother die from suicide as compared to an accident and approximately twofold increased odds of the offspring being hospitalized for suicide attempt.

Comparisons and Synthesis of Methods

The seven approaches described differed in their targets of interest, implementable study designs, as well as the necessary specifications of parameters. Rosenbaum's approaches focus on obtaining the value Γ (upper bound of OR_{xu}) and/or Δ (upper bound of OR_{yu}) at which $OR_{yx\times cu}$ becomes nonsignificant. Given that it uses information on the actual number of pairs in the study, this method reflects the uncertainty of the analysis associated with sample size. The results of sensitivity analyses may then change as sample size changes: the values of Γ and Δ tend to be slightly larger as sample size increases. In other words, the study conclusion tested may appear to be more robust when the sample size is large. The other class of methods we described, which we labeled the $OR_{yx\times cu}$ with CI approaches, obtains both the point estimate of the treatment effect and its CI. This group of methods does not use the actual sample from the original study, thus sensitivity results do not change as a function of the sample size. Additionally, these seven methods also differ in the study designs in which they can be implemented. The software to implement Rosenbaum's approach can generally only be applied to 1:1 matching designs. Compared to Rosenbaum's approaches, the $OR_{yx\times cu}$ with CI approaches are more flexible in that they accommodate any study design.

Each of these techniques requires the specification of a different set of parameters related to the unobserved confounder. The differences in the parameters specified across methods suggest that researchers should choose a method based on the parameters that they feel comfortable specifying or on which there is existing literature. The advantage of Rosenbaum's approach is that users are able to directly specify the values of both OR_{yu} and OR_{xu}, which may be more readily obtained from the literature. However, unlike the other approaches described in the paper, Rosenbaum's approach does not allow users to conduct sensitivity analysis on results from published studies, since it requires the number of discordant pairs, which is not always readily available. The approaches of Greenland, Lin et al., and VanderWeele and Arah require the specification of $p(u|x=0)$ and/or $p(u|x=1)$. While Harding's approach does not require the specification of these parameters, the computation is more involved.

As a motivating example, we applied these seven methods to assess the sensitivity of an observed relationship between maternal suicide and offspring's hospitalization for suicide attempt to an unobserved binary confounder, such as genetic predisposition to suicidal behavior. Rosenbaum's primal

Table 4 Estimated $OR_{yx\times cu}$ (95 % CI) using sensitivity analysis methods that obtain $OR_{yx\times cu}$ and CI

Approach	$p(u\|x=0)$					
	1 %	5 %	10 %	15 %	20 %	25 %
Greenland	1.77 (1.42, 2.22)	1.57 (1.24, 1.98)	1.46 (1.15, 1.86)	1.42 (1.12, 1.82)	1.41 (1.10, 1.80)	1.40 (1.10, 1.79)
Harding[a]	1.77 (1.42, 2.22)	1.57 (1.24, 1.98)	1.46 (1.15, 1.86)	1.42 (1.11, 1.82)	1.40 (1.10, 1.80)	1.41 (1.10, 1.80)
Lin et al./VanderWeele and Arah	1.77 (1.42, 2.21)	1.57 (1.26, 1.96)	1.46 (1.17, 1.82)	1.42 (1.14, 1.77)	1.41 (1.13, 1.75)	1.41 (1.13, 1.76)

OR_{yu}=1.84, OR_{xu}=7.39, $OR_{yx\times c}$=1.86

[a] $p(u)$ was obtained from the specified $p(u|x=0)$, $p(x)$=50 %, and OR_{xu}=7.39

approach suggests that the study inference will be no longer significant when the upper bound of OR_{xu} is >1.55. While this would imply sensitivity to unobserved confounding, it is important to keep in mind that this method assumes a situation in which the unobserved confounder perfectly predicts the outcome of interest, e.g., genetic predisposition to suicidal behavior perfectly predicts offspring's hospitalization for suicide attempt. As a result, this method may overstate the study sensitivity.

The simultaneous approach yielded an inference closer to the approaches that estimate $OR_{yx\times cu}$ with its CI, which suggests that the association between maternal suicide and offspring's hospitalization for suicide attempt is relatively robust to an unobserved confounder. However, the simultaneous approach still suggested that the study of Kuramoto et al. (2010) is somewhat sensitive to an unobserved confounder. This evidence of this heightened sensitivity may relate to the fact that Rosenbaum's approaches depend on the sample size in the original study and, in particular, the number of discordant pairs, which is relatively small in the study of Kuramoto et al. since the outcome is rare.

Other Sensitivity Analysis Methods

Although we focused on a few particular sensitivity analysis techniques because of their relative ease of implementation and fairly intuitive explanations, other sensitivity analysis methods are available. For example, Schneeweiss (2006) described methods that quantify the unobserved confounder under certain specifications and then arrive at an estimate of $OR_{yx\times cu}$, but do not provide an estimate of the CI. Harding (2009) used a sensitivity analysis based on omitted variable bias calculations for ordinary least squares regression (an approach relatively common in economics). Ridgeway (2006) and McCaffrey et al. (2004) discussed a method that can be applied to studies that utilize propensity score weights and is available in the "twang" package for R. Methods are also available that yield nonparametric bounds on the treatment effects, without characterizing the relationship between the unobserved confounder with the exposure and the outcome (Manski et al. 1992). A method using Monte Carlo methods and Bayesian analysis techniques to allow sensitivity parameters to come from specified distributions is described in McCandless et al. (2007) and Steenland and Greenland (2004), with *R* and WinBUGS codes provided. One of the advantages of simulation methods is that the CIs obtained may be more accurate than those provided by the simpler approaches. For a practical example of sensitivity analysis using Monte Carlo simulation, see Arah et al. (2008).

Discussion

While observed confounding in nonexperimental studies can be addressed by methods such as propensity score matching, researchers have not had good tools to handle potential unobserved confounding. This paper discussed different methods of sensitivity analysis that quantify the sensitivity of study results to unobserved confounding. When applying these methods to Kuramoto et al. (2010), they yielded a fairly consistent result, suggesting that the relationship between maternal suicide and offspring's hospitalization for suicide attempt is relatively robust to an unobserved confounder strongly associated with maternal suicide and moderately associated with offspring's hospitalization for suicide attempt. The sensitivity analysis gives us the confidence to conclude that the observed association between parent's suicide and offspring's suicide attempt is likely to be causal. Our study conclusion suggests that prevention strategies need to particularly focus on individuals whose parents died from suicide, for example, by ensuring access to counseling services. Future studies should replicate the study findings to establish true causality, and of course, most studies, this one included, also have other limitations, such as measurement error and other threats to validity. In addition, future studies should further investigate the mechanism of such a relationship, such as behavioral changes that might mediate the relationship between parent's suicide and offspring's hospitalization for suicide attempt.

Most of the methods described in this paper can be implemented relatively easily with available software packages or by hand. The preferred sensitivity analysis for a particular study may be driven by the target of interest, the study design, the details of a potential unobserved confounder that can be easily specified using external sources, and its ease of implementation. However, as the methods discussed rely on slightly different sets of assumptions and come from different perspectives, it may also be helpful to explore several sensitivity analysis tools. In addition, it is important to note that the second set of methods ($OR_{yx\times cu}$ with CI) do not require the original data and can be implemented using just the results in a published study. This allows researchers to conduct sensitivity analyses for published results, enabling them to determine how much confidence should be placed in those results.

Although more general and more complex approaches are available, the availability of these simple approaches leaves researchers little excuse for not performing sensitivity analyses when conducting nonexperimental studies. Although we are not aware of their use in that context, these methods may also be useful for nonexperimental comparisons conducted within the context of randomized trials, such as to handle noncompliance (e.g., Jo and Stuart

2009). We hope that this paper will raise awareness and the use of these important methods. By giving researchers insight into how sensitive studies may be to unobserved confounding, these methods can serve a critical role in nonrandomized studies aiming to establish causal relationships. Importantly, these methods can inform prevention researchers and policy makers when drawing conclusions from nonexperimental studies.

Acknowledgments The authors wish to acknowledge the NARSAD Young Investigator Award to Dr. Holly C. Wilcox, and Dr. Holly C. Wilcox for allowing us to use the motivating example. We also thank the National Institute of Drug Abuse for the training support for S. Janet Kuramoto (1F31DA0263182), the National Institute of Mental Health (NIMH) Prevention Research T32 Training Grant for the training support for Weiwei Liu (T32 MH18834), and NIMH for the support of Elizabeth Stuart's time (K25 MH083846). This work was performed while Weiwei Liu was a postdoctoral fellow and S. Janet Kuramoto was a student at Johns Hopkins Bloomberg School of Public Health.

References

Arah, O. A., Chiba, Y., & Greeland, S. (2008). Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders. *Annals of Epidemiology, 18*, 637–646. doi:10.1016/j.annepidem.2008.04.003.

Brent, D. A., & Mann, J. J. (2005). Family genetic studies, suicide, and suicidal behavior. *American Journal of Medical Genetics, 133*, 13–24. doi:10.1002/ajmg.c.30042.

Cornfield, J., Haenszel, W., Hammon, E., Lilienfeld, A., Shimkin, M., & Wynder, E. (1959). Smoking and lung cancer: Recent evidence and a discussion of some questions. *Journal of the National Cancer Institute, 22*, 173–203. Available at http://ije.oxfordjournals.org/content/38/5/1175.full.

DiPrete, T. A., & Gangl, M. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. *Sociological Methodology, 34*, 271–310. doi:10.1111/j.0081-1750.2004.00154.x.

Flay, B., Biglan, A., Boruch, R., Castro, F., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151–175. doi:10.1007/s11121-005-5553-y.

Gangl, M. (2004). Rbounds: Stata module to perform Rosenbaum sensitivity analysis for average treatment effects on the treated. Available at http://EconPapers.repec.org/RePEc:boc:bocode:s438301.

Gastwirth, J., Krieger, A., & Rosenbaum, P. (1998). Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika, 85*, 907–920. doi:10.1093/biomet/85.4.907.

Greenland, S. (1996). Basic methods for sensitivity analysis of biases. *International Journal of Epidemiology, 25*, 1107–1116. doi:10.1093/ije/25.6.1107.

Harding, D. (2003). Counterfactual models of neighborhood effects: The effect of neighborhood poverty on dropping out and teenage pregnancy. *The American Journal of Sociology, 109*, 676–719. doi:10.1086/379217.

Harding, D. J. (2009). Collateral consequences of violence in disadvantaged neighborhoods. *Social Forces, 88*, 757–784. doi:10.1353/sof.0.0281.

Haviland, A., Nagin, D., & Rosenbaum, P. (2007). Combining propensity score matching and group-based trajectory analysis in an observational study. *Psychological Methods, 12*, 247–267. doi:10.1037/1082-989X.12.3.247.

Jo, B., & Stuart, E. A. (2009). On the use of propensity scores in principal causal effect estimation. *Statistics in Medicine, 28*, 2857–2875. PMCID PMC 2757143.

Keele, L. (2010). Rbounds: An R package for sensitivity analysis with matched data. R package. Available at http://www.polisci.ohio-state.edu/faculty/lkeele/rbounds.html.

Kellam, S. G., Brown, C. H., Poduska, J. M., Ialongo, N. S., Wang, W., Toyinbo, P., et al. (2008). Effects of a universal classroom behavior management program in first and second grades on young adult behavioral, psychiatric, and social outcomes. *Drug and Alcohol Dependence, 95*, S5–S28. doi:10.1016/j.drugalcdep.2008.01.004.

Kitahata, M. M., Gange, S. J., Abraham, A. G., Merriman, B., Saag, M. S., Justice, A. C., et al. (2009). Effect of early versus deferred antiretroviral therapy for HIV on survival. *The New England Journal of Medicine, 360*, 1815–1826. Available at http://www.nejm.org/doi/full/10.1056/NEJMoa0807252#t=articleTop.

Kuramoto, S. J., Stuart, E. A., Runeson, B., Lichtenstein, P., Langstrom, N., & Wilcox, H. C. (2010). Maternal or paternal suicide and offspring's psychiatric and suicide-attempt hospitalization risk. *Pediatrics, 126*, e1026–e1032. doi:10.1542/peds.2010-0974.

Lieb, R., Bronisch, T., Hofler, M., Schreier, A., & Wittchen, H.-U. (2005). Maternal suicidality and risk of suicidality in offspring: Findings from a community study. *The American Journal of Psychiatry, 162*, 1665–1671. doi:10.1176/appi.ajp.162.9.1665.

Lin, D. Y., Psaty, B. M., & Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics, 54*, 948–963. Available at http://www.jstor.org/pss/2533848.

Liu, W. (2012). *The adult offending and school dropout nexus: A life course analysis*. El Paso: LFB Scholarly Publishing LLC (in press).

Love, T. (2008). Spreadsheet-based sensitivity analysis calculations for matched samples. Center for Health Care Research & Policy, Case Western Reserve University, Cleveland. Available at http://www.chrp.org/propensity/.

Luiz, R., & Cabral, M. (2010). Sensitivity analysis for an unmeasured confounder: A review of two independent methods. *Revista Brasileira de Epidemiologia, 13*, 188–198. Available at http://www.scielosp.org/pdf/rbepid/v13n2/02.pdf.

Manski, C. F., Sandefur, G. D., McLanahan, S., & Power, D. (1992). Alternative estimates of the effect of family structure during adolescence on high school graduation. *Journal of the American Statistical Association, 87*, 25–37. Available at http://www.jstor.org/pss/2290448.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods, 9*, 403–425. doi:10.1037/1082-989X.9.4.403.

McCandless, L. C., Gustafson, P., & Levy, A. (2007). Bayesian sensitivity analysis for unmeasured confounding in observational studies. *Statistics in Medicine, 26*, 2331–2347. doi:10.1002/sim.

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika, 12*, 153–157. doi:10.1007/BF02295996.

Niederkrotenthaler, T., Floderus, B., Alexanderson, K., Rasmussen, F., & Mittendorfer-Rutz, E. (2012). Exposure to parental mortality and markers of morbidity, and the risks of attempted and completed suicide in offspring: An analysis of sensitive life periods. *Journal of Epidemiology and Community Health, 66*, 233–239. doi:10.1136/jech.2010.109595.

Ridgeway, G. (2006). Assessing the effect of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology, 22*, 1–29. doi:10.1007/s10940-005-9000-9.

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.

Rosenbaum, P. R. (2010). *Design of observational studies*. New York: Springer.

Rothman, K. J., Greenland, S., & Lash, T. L. (2008). *Modern epidemiology*. Philadelphia: Lippincott Williams & Wilkins.

Schneeweiss, S. (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and Drug Safety, 15*, 291–303. doi:10.1002/pds.1200.

Steenland, K., & Greenland, S. (2004). Monte Carlo sensitivity analysis and Bayesian analysis of smoking as an unmeasured confounder in a study of silica and lung cancer. *American Journal of Epidemiology, 160*, 384–392. doi:10.1093/aje/kwh211.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1–21. doi:10.1214/09-STS313.

Stuart, E. A., & Green, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology, 44*, 395–406.

VanderWeele, T. J., & Arah, O. A. (2011). Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders. *Epidemiology, 22*, 42–52. doi:10.1097/EDE.0b013e3181f74493.

Wilcox, H. C., Kuramoto, S. J., Lichtenstein, P., Långström, N., Brent, D. A., & Runeson, B. (2010). Psychiatric morbidity, violent crime and suicide among children and adolescents exposed to parental death. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*, 514–523. doi:10.1016/j.jaac.2010.01.020.