

Research Article

A Proxy Outcome Approach for Causal Effect in Observational Studies: A Simulation Study

Wenbin Liang,¹ Yuejen Zhao,² and Andy H. Lee³

¹ National Drug Research Institute, Health Science, Curtin University, G.P.O. Box U 1987, Perth, WA 6845, Australia

² Northern Territory Department of Health, Darwin, NT 0800, Australia

³ School of Public Health, Health Science, Curtin University, Perth, WA 6845, Australia

Correspondence should be addressed to Wenbin Liang; w.liang@curtin.edu.au

Received 1 November 2013; Revised 3 January 2014; Accepted 3 January 2014; Published 18 February 2014

Academic Editor: Abdelwahab Omri

Copyright © 2014 Wenbin Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Known and unknown/unmeasured risk factors are the main sources of confounding effects in observational studies and can lead to false observations of elevated protective or hazardous effects. In this study, we investigate an alternative approach of analysis that is operated on field-specific knowledge rather than pure statistical assumptions. **Method.** The proposed approach introduces a proxy outcome into the estimation system. A proxy outcome possesses the following characteristics: (i) the exposure of interest is not a cause for the proxy outcome; (ii) causes of the proxy outcome and the study outcome are subsets of a collection of correlated variables. Based on these two conditions, the confounding-effect-driven association between the exposure and proxy outcome can then be measured and used as a proxy estimate for the effects of unknown/unmeasured confounders on the outcome of interest. Performance of this approach is tested by a simulation study, whereby 500 different scenarios are generated, with the causal factors of a proxy outcome and a study outcome being partly overlapped under low-to-moderate correlations. **Results.** The simulation results demonstrate that the conventional approach only led to a correct conclusion in 21% of the 500 scenarios, as compared to 72.2% for the alternative approach. **Conclusion.** The proposed method can be applied in observational studies in social science and health research that evaluates the health impact of behaviour and mental health problems.

1. Background

Due to lack of randomization, estimates obtained from observational studies are often affected by uncontrolled or unmeasured confounding effects. Several methods have been proposed to deal with the problem [1–11], but their application relies on assumptions about the distribution of the unknown confounding factor(s) in relation to the outcome, the exposure, and other known covariates.

The present study investigates an alternative approach of analysis that makes use of field-specific knowledge to determine a proxy outcome, which then will be employed to estimate uncontrolled confounding effects. A proxy outcome should satisfy the following conditions: (i) the exposure of interest is not a cause for the proxy outcome; and (ii) causes of the proxy outcome and the main outcome are subsets of a collection of correlated variables. If condition one is satisfied, then it is certain that the observed association between the

exposure and the proxy outcome is completely driven by confounding effects. Nevertheless, condition one may be relaxed to some extent, for example, when it is certain that the confounding effect is by far stronger than possible causal effect. When condition two is satisfied, confounders for the proxy outcome and the outcome of interest are similar or at least correlated. For example, various forms of physical and mental health outcomes can be affected by a cluster of socioeconomic, behavioural, psychological, and genetic factors [12–22]. Researchers can apply their field knowledge and experience to determine the best proxy outcome for their outcome and exposure of interest.

Intuitively, let y and \tilde{y} be the main outcome and the proxy outcome, respectively; r denotes the possible risk of y due to a given level of exposure; t_0 and t_1 the unexposed and exposed person-time at risk; d_0 and d_1 represent the number of cases of y observed in t_0 and t_1 , whereas d'_0 and d'_1 are the number of cases of \tilde{y} observed in t_0 and t_1 , respectively.

Further, suppose that $P(d_0)$ and $P(d_1)$ define the expected probability of one or more sufficient causes that the exposure of interest is not involved, occurring for y in a unit of person-time at risk within t_0 and t_1 , respectively; correspondingly, $P(d'_0)$ and $P(d'_1)$ define the expected probability of one or more sufficient causes occurring for \tilde{y} in a unit of person-time at risk within t_0 and t_1 , respectively. Then, the observed crude risk ratios are

$$\frac{d_1/t_1}{d_0/t_0} = \frac{\{r[1 - P(d_1)] + P(d_1)\}}{P(d_0)} \quad (1)$$

$$\frac{d'_1/t_1}{d'_0/t_0} = \frac{P(d'_1)}{P(d'_0)}.$$

If the sufficient causes of y and \tilde{y} are the same, largely overlapped, or strongly correlated, then

$$\frac{P(d_1)}{P(d_0)} = \frac{P(d'_1)}{P(d'_0)}, \quad (2)$$

$$\frac{d_1/t_1}{d_0/t_0} = \frac{d'_1/t_1}{d'_0/t_0}$$

will be observed if the exposure is not causal for y ($r = 0$), whereas

$$\frac{d_1/t_1}{d_0/t_0} > \frac{d'_1/t_1}{d'_0/t_0} \quad (3)$$

will be observed if the exposure is causal for y ($r > 0$). Given this strict assumption, the method has been successfully applied in time series analysis recently, whereby the proxy outcome was described as a control series [23, 24]. For example, the study by Herttua and colleagues investigated the effect of alcohol price (exposure) on alcohol-related mortality (outcome of interest), while coronary operations were used as the control series (proxy outcome) [23]. Nevertheless, such a method may remain valid when assumptions are relaxed and can be applied to different study designs. Because the details of underlying causal mechanisms for an outcome (i.e., all sufficient causes) are typically unknown, it is best to ascertain the validity of the method through simulations. Therefore, we conduct a simulation study to test its application, focusing on situations when the causes of y and \tilde{y} are only partly overlapped and have only low to moderate strength of association.

2. Method

2.1. Simulation Design. The simulation process follows the sufficient cause model [25]. In the simulation for an event to occur, at least one sufficient cause has to occur, which comprises the occurrence of two matched causal components and absence of any competing event. In addition, a randomly distributed small error term is introduced to ensure that perfect prediction (which interrupts the computing process) will not occur. To account for the fact that only certain real causal factors are known yet some of the noncausal

factors are mistaken as causal factors, a collection of variables are included to encompass exposure, causal factors, and noncausal factors, while a subset from the pool provides the known variables. All simulations are performed within the STATA package release 12.

The simulations contain 500 replicates, with each replicate being generated through an independent process. There are seven simulation steps involved in each replicate.

- (1) Generate a collection of correlated binary variables, $\mathbf{X}_{40 \times 50000} = \{X_{in}\}$, $i = (1, 2, 3, \dots, 40)$, and $n = (1, 2, 3, \dots, 50000)$. For each n , X_{in} is set to 1 if an intermediate process variable $T_{in} \geq 0.75$ and 0 otherwise, where $T_{in} = V_{in}P_i + U_n(1 - P_i)$ is a combination of a variable component (V_i) and a unique component (U) for each i , both being uniform $[0, 1)$ distributed random variables, and P_i is a random proportion drawn from a uniform $[0.3, 0.8)$ distribution. The range $[0.3, 0.8)$ is chosen in order to set a low to moderate level of correlation among X . The mean (standard deviation), 25th, 50th, and 75th percentiles of the correlation coefficients for the matrix \mathbf{X} are 0.29 (0.12), 0.20, 0.26, and 0.34, respectively.
- (2) Determine the real causes for outcomes of interest A , B , and proxy outcome C . Real causes for A , B , and C are subsets of \mathbf{X} in which X_i and X_{i+10} , $i = (1, 2, 3, \dots, 10)$ form ten matched pairs. Let F_{ji} indicate the factual causes, $j = (A, B, C)$ and $i = (1, 2, 3, \dots, 10)$. In this simulation, we force $F_{A,1} = 1$ and $F_{B,1} = F_{C,1} = 0$; that is, X_1 is causal for A , but not causal for B and C . For $i = (2, 3, \dots, 10)$, F_{ji} is a random value drawn from the Bernoulli distribution with probability of success = 0.5, value of success = 1, and value of failure = 0. For example if $F_{A,2} = 1$, then the pair of X_2 and X_{12} is a cause of A .
- (3) Generate competing events for outcomes A , B , and C . Let $E_{j,n}$ denote the competing events for outcomes A , B , and C , $j = (A, B, C)$ and $n = (1, 2, 3, \dots, 50000)$. E is a Bernoulli distributed random variable with a probability of success = 0.1, value of success = 1, and value of failure = 0. E is also independent of X .
- (4) Generate background errors for outcomes A , B , and C . Let $Q_{j,n}$ denote "background" sufficient causes, $j = (A, B, C)$ and $n = (1, 2, 3, \dots, 50000)$. Q is a Bernoulli distributed random variable with a probability of success = 0.05, value of success = 1, and value of failure = 0. Q is independent of E and X . Q services as small random error, and it is introduced to smooth the computing process only.
- (5) Determine the status (occur or not occur) of outcomes A , B , and C . Let $Y_{j,n}$ where $j = (A, B, C)$ and $n = (1, 2, 3, \dots, 50000)$ denote the status of outcomes A , B , and C . Value of each $Y_{j,n}$ is determined as follows. For each observation n , $Y_{j,n} = 1$ (outcome occurred) if $Q_{j,n} = 1$, or for $i = (1, 2, 3, \dots, 10)$, $\sum_i X_{in}X_{(i+10)n}F_{j,i} \geq 1$ when $E_{j,n} = 0$; Otherwise $Y_{j,n} = 0$ (outcome not occurred).

- (6) Determine the known/suspected (not necessary the fact) “causal” factors (except X_1) for outcomes A and B . The known/suspected “causal” factors for outcomes A , B , and C are determined for X_i for $i = (1, 2, 3, \dots, 40)$. Let $K_{j,i}$ denote the researcher’s knowledge (not necessary the fact) of causes for outcomes A and B . Let $j = (A, B)$, and $K_{j,i}$ indicates a known “causal” factor. Because X_1 is the exposure of interest, so we force each $K_{j,i} = 1$ when $i = 1$. For $i = (2, 3, \dots, 20)$, $K_{j,i}$ is a random value drawn from a Bernoulli distribution with a probability of success = 0.5. For $i = (21, 22, 23, \dots, 40)$, $K_{j,i}$ is a random value drawn from the Bernoulli distribution with a probability of success = 0.15, value of success = 1, and value of failure = 0. The difference in the success rates between the two groups indicates that a real causal factor is more likely to be acknowledged than a noncausal factor.
- (7) Estimate the effects of X_1 on outcomes A and B based on the fact model, and compare the conventional approach with the proposed approach. Let $G_{j,n}$, $j = (A, B)$, $n = (1, 2, 3, \dots, 50000)$, be the presence of causes (except X_1 and X_{11}) for outcomes A and B for each observation. For each n , $G_{j,n} = 1$ for $i = (2, 3 \dots, 10)$, if $\sum_i X_{in} X_{(i+10)n} F_{j,i} \geq 1$; $G_{j,n} = 0$ otherwise.

The true effects of X_1 on outcomes A and B based on the fact model are estimated by

$$P(Y_{j,n} = 1 | X_{1n}, G_{j,n}) = \frac{\exp(\beta_{j,1} X_{1n} + \beta_j G_{j,n})}{1 + \exp(\beta_{j,1} X_{1n} + \beta_j G_{j,n})} \tag{4}$$

where $\beta_{j,i}$ is the estimated real effect of X_i on outcome j for $j = (A, B)$. To estimate the effects of X_1 on outcomes A and B based on known/suspected confounders and applying standard multivariate logistic regression as the adjustment method, we have

$$P(Y_{j,n} = 1 | X_{in}, K_{j,i}) = \frac{\exp(\beta'_{j,1} X_{1n} + \sum_{i=2}^{40} \beta'_{j,i} X_{in} K_{j,i})}{1 + \exp(\beta'_{j,1} X_{1n} + \sum_{i=2}^{40} \beta'_{j,i} X_{in} K_{j,i})} \tag{5}$$

where $\beta'_{j,i}$ is the estimated effect of X_i on outcome j for $j = (A, B)$. To estimate the confounding effects on X_1 on outcomes A and B using the proxy outcome C , the logistic model for adjustment becomes

$$P(Y_{C,n} = 1 | X_{in}, K_{j,i}) = \frac{\exp(\tilde{\beta}'_{j,1} X_{1n} + \sum_{i=2}^{40} \tilde{\beta}'_{j,i} X_{in} K_{j,i})}{1 + \exp(\tilde{\beta}'_{j,1} X_{1n} + \sum_{i=2}^{40} \tilde{\beta}'_{j,i} X_{in} K_{j,i})} \tag{6}$$

where $\tilde{\beta}'_{j,i}$ is the estimated effect of X_i on outcome C (proxy outcome) for $j = (A, B)$.

The adjusted effects of X_1 on A and B are $\hat{\beta}_{A,1} = \beta'_{A,1} - \tilde{\beta}'_{A,1}$ and $\hat{\beta}_{B,1} = \beta'_{B,1} - \tilde{\beta}'_{B,1}$, respectively.

2.2. *Classification of Effect of X_1 on A and B .* Based on the fact model, X_1 increases risk of outcome A if $\beta_{A,1} > 0.05$ and P value for $\beta_{A,1} < 0.05$; otherwise X_1 has no effect on A . Also, X_1 increases risk of outcome B if $\beta_{B,1} > 0.05$ and P value for $\beta_{B,1} < 0.05$; otherwise X_1 has no effect on B . The same effect patterns hold analogously for the conventional model and the alternative approach by replacing the regression coefficient $\beta_{j,1}$ with $\beta'_{j,1}$ and $\tilde{\beta}'_{j,1}$, respectively. Classifications of the effect of X_1 based on the fact model are then used as the gold standard to compare with the classifications based on the conventional approach and the alternative approach.

2.3. *Empirical Application.* A simple example is provided to clarify the methodology. For additional illustration of the proxy outcome method to adjust for residual confounding effects, interested readers are referred to the first author’s recently published research [26]. Briefly, when investigating the effect of alcohol use (exposure) on general health status (outcome), both measured and unmeasured confounding factors are involved. Many of these confounding factors are clustered within the family such as socioeconomic determinants, environmental factors, lifestyle, and genetic susceptibility. Although current alcohol use by adults does not produce any physiological effect on their children’s current health, observed effect of current alcohol use (exposure) on their children’s health status (proxy outcome) can be used as an approximation of confounding effects.

This example used the data from the 2010 National Health Interview Survey. A first logistic regression model was fitted to compare the likelihood of having undesirable (poor or fair) health status (outcome) between lifetime abstainers and current light drinkers (exposure). A second logistic regression model was then applied to compare the likelihood of having undesirable health status in the children (proxy outcome) in relation to the drinking status of their parents. To adjust for confounding effects, natural logarithm of the odds ratios from the second model was introduced as an offset variable into the first model.

3. Results

Estimates based on the knowledge and conventional model from one replicate are shown in Tables 1 and 2 as an example. Both outcomes A and B are treated as the outcomes of interest, while outcome C is used as the proxy outcome. In this replicate, between outcomes A and C , there are four common causal factors. These account for 33% and 40% of all causal factors for A and C , respectively. Except for the exposure of interest (X_1), 54% of causal factors for A are known. The true effect of X_1 on outcome A ($\beta_{A,1}$) based on the fact model is 1.27 (P value < 0.001) indicating that X_1 is a real causal factor to A . The estimated effect based on the conventional approach ($\beta'_{A,1}$) is 0.79 (P value < 0.001). The estimated effect based on the alternative approach ($\hat{\beta}_{A,1}$)

TABLE 1: Data example of a replicate/scenario, estimated effects (coefficients from logistic models) of exposure (X_i), and known “causal”/confounding factors of A on A and proxy outcome C .

	Effects on A		Effects on C		Indicators for real causal factors (1 = Yes, 0 = No)	
	Coefficient	P value	Coefficient	P value	Causal to A	Causal to C
X_1	0.79	0.000	0.09	0.011	1	0
X_2	0.19	0.000	0.92	0.000	0	1
X_3	0.69	0.000	0.03	0.360	1	0
X_4	0.46	0.000	0.50	0.000	0	0
X_5	*		*		1	1
X_6	*		*		1	1
X_7	0.87	0.000	0.29	0.000	1	0
X_8	0.20	0.000	0.84	0.000	0	1
X_9	0.04	0.293	0.80	0.000	0	1
X_{10}	0.77	0.000	0.13	0.000	1	0
X_{11}	*		*		1	0
X_{12}	0.15	0.000	0.79	0.000	0	1
X_{13}	*		*		1	0
X_{14}	*		*		0	0
X_{15}	0.85	0.000	0.91	0.000	1	1
X_{16}	*		*		1	1
X_{17}	0.73	0.000	0.10	0.009	1	0
X_{18}	0.15	0.000	0.89	0.000	0	1
X_{19}	*		*		0	1
X_{20}	0.95	0.000	0.35	0.000	1	0
X_{21}	*		*		0	0
X_{22}	0.27	0.000	0.16	0.000	0	0
X_{23}	0.25	0.000	0.19	0.000	0	0
X_{24}	0.20	0.000	0.18	0.000	0	0
X_{25}	*		*		0	0
X_{26}	*		*		0	0
X_{27}	*		*		0	0
X_{28}	*		*		0	0
X_{29}	0.50	0.000	0.47	0.000	0	0
X_{30}	*		*		0	0
X_{31}	*		*		0	0
X_{32}	*		*		0	0
X_{33}	*		*		0	0
X_{34}	*		*		0	0
X_{35}	*		*		0	0
X_{36}	*		*		0	0
X_{37}	*		*		0	0
X_{38}	0.13	0.001	0.17	0.000	0	0
X_{39}	*		*		0	0
X_{40}	0.16	0.000	0.17	0.000	0	0

* indicates variable is not known as a “causal” factor for A , therefore is not included in the models.

TABLE 2: Data example of a replicate/scenario, estimated effects (coefficients from logistic models) of exposure (X_i), and known “causal”/confounding factors of B on B and proxy outcome C .

	Effects on B		Effects on C		Indicators for real causal factors (1 = Yes, 0 = No)	
	Coefficient	P value	Coefficient	P value	Causal to B	Causal to C
X_1	0.12	0.001	0.16	0.000	0	0
X_2	*		*		1	1
X_3	*		*		1	0
X_4	0.39	0.000	0.62	0.000	0	0
X_5	*		*		0	1
X_6	*		*		0	1
X_7	0.24	0.000	0.43	0.000	0	0
X_8	*		*		1	1
X_9	*		*		0	1
X_{10}	*		*		0	0
X_{11}	*		*		0	0
X_{12}	1.22	0.000	0.88	0.000	1	1
X_{13}	1.15	0.000	0.17	0.000	1	0
X_{14}	0.23	0.000	0.27	0.000	0	0
X_{15}	0.16	0.000	1.04	0.000	0	1
X_{16}	0.31	0.000	1.56	0.000	0	1
X_{17}	0.13	0.000	0.19	0.000	0	0
X_{18}	*		*		1	1
X_{19}	*		*		0	1
X_{20}	0.28	0.000	0.43	0.000	0	0
X_{21}	*		*		0	0
X_{22}	0.17	0.000	0.24	0.000	0	0
X_{23}	*		*		0	0
X_{24}	*		*		0	0
X_{25}	*		*		0	0
X_{26}	*		*		0	0
X_{27}	*		*		0	0
X_{28}	0.23	0.000	0.26	0.000	0	0
X_{29}	0.31	0.000	0.58	0.000	0	0
X_{30}	*		*		0	0
X_{31}	*		*		0	0
X_{32}	*		*		0	0
X_{33}	*		*		0	0
X_{34}	*		*		0	0
X_{35}	*		*		0	0
X_{36}	*		*		0	0
X_{37}	*		*		0	0
X_{38}	*		*		0	0
X_{39}	*		*		0	0
X_{40}	*		*		0	0

* indicates variable is not known as a “causal” factor for B , therefore is not included in the models.

TABLE 3: Effect of X_1 on A and B based on the fact model, conventional approach, and the alternative approach.

Fact model	Conventional approach		Alternative approach (%)	
	X_1 causal to A but not B (%)	X_1 causal to A and B (%)	X_1 causal to A but not B (%)	X_1 causal to A and B (%)
X_1 causal to A but not B	93 (18.6%)	394 (78.8%)	347 (69.4%)	140 (28.0%)
X_1 causal to A and B	1 (0.2%)	12 (2.4%)	4 (0.8%)	9 (1.8%)

TABLE 4: Empirical application, predicting adults' health status (outcome of interest) and their children's health status (proxy outcome) by alcohol use (proxy outcome).

	Have undesirable health status in adults		Have undesirable health status in children	
	Odds Ratio	95% confidence interval	Odds ratio	95% confidence interval
Parents' drinking behaviour				
Lifetime abstainers	1.00	Reference	1.00	Reference
Current light drinkers	0.54	0.42 - 0.69	0.60	0.50 - 0.71

is $0.79 - 0.09 = 0.70$. Both the conventional approach and alternative approach lead to the same correct conclusion that X_1 is a causal factor to A .

Between outcomes B and C there are also four common factors. These account for 67% and 40% of all causal factors for B and C , respectively. Except the exposure of interest (X_1), 33% of causal factors for B are known. The true effect of X_1 on outcome B ($\beta_{B,1}$) based on the fact model is -0.024 (P value = 0.668) indicating that X_1 is not a causal factor to B . The estimated effect based on the conventional approach ($\beta'_{B,1}$) is 0.12 (P value < 0.001). The estimated effect based on the alternative approach ($\hat{\beta}_{B,1}$) is $0.12 - 0.16 = -0.04$. Therefore, based on estimation from the conventional approach one would mistakenly draw the conclusion that " X_1 is a causal factor to B ." However, given that $\hat{\beta}_{B,1} < 0.05$, the alternative approach has led to the correct conclusion that X_1 is not a causal factor to B .

Table 3 summarises findings from the 500 replicates. Based on the fact model, the exposure of interest (X_1) is classified as a causal factor for outcome A in all replicates. This is in perfect agreement with the simulation process that X_1 is set to be a causal factor for A . In the simulation process, X_1 is set to be a noncausal factor for B ; in 97.4% of the 500 replicates, the fact model concludes that X_1 is not a causal factor for B , but in 2.6% of the replicates the fact model concludes that X_1 is a causal factor for B . The disagreement between the simulation process and the fact model in these 2.6% replicates is a result of type I error (setting the two-sided confidence interval to 95%). Nevertheless, in all replicates, both the conventional and alternative approaches have classified X_1 as a causal factor for A . When comparing the estimates between the conventional approach and the fact model, the two models have led to the same conclusion in only 21% of the replicates, while in 72.2% of the replicates, the alternative approach has led to the same conclusion as the fact model. When the simulation process is used as the gold standard for classification (i.e., X_1 is causal for A , but not causal for B), the sensitivity of the new approach is 100%, and the specificity is 70.2%.

Table 4 presents results of the example. Alcohol use by adults has similar effects on their health status and their

children's health status, whereas the effect on children's health status is mediated by confounding factors. To account for the uncontrolled confounding effects when estimating the effect of adults' alcohol use on their health status, an offset variable, which takes on the value of the natural logarithm of 1 for lifetime abstainers and 0.60 for current light drinkers, is added to the model. The adjustment changes the odds ratio from 0.54 ($P < 0.001$) to 0.90 ($P = 0.38$).

4. Discussion

In this study, we introduce a new analysis approach for causal effect. Although the new approach is only applicable to measurements of relative effects (i.e., risk ratios, odds ratios), it does not require any distributional assumption for the confounding variables in relation to the outcome, the exposure, and other known confounding variables. Instead, the approach merely assumes that the causes for the outcome of interest and proxy outcome are partly overlapped and correlated. The choice of an optimal proxy outcome is achievable by directly applying field expertise without advanced knowledge in statistics. The simulation results show that the alternative approach is far more accurate than the conventional approach in classifying causal associations, even under conditions of low to moderate correlation between the causes for the outcome and causes for the proxy outcome. The proposed approach appears to be suitable for observational studies in social science and health research that evaluate the health impact of behaviour and mental health problems, especially where clusters of causes for various outcomes are strongly correlated and overlapped in these fields [12, 27, 28].

It should be remarked that the analysis can only be performed when effects are measured by relative risk difference such as risk ratio or odds ratio. Another limitation is that false classification remains possible, even though the proposed method appears to have an advantage over the conventional approach. In this study, we demonstrate a new simulation process that incorporates the component causes, competing events, difference between the fact and the knowledge, to model realistic scenarios in observational studies. This simulation process could be further developed

and used to determine how knowledge that deviates from the fact can introduce bias in estimates.

5. Conclusion

In conclusion, the proposed proxy outcome approach can be applied in observational studies in social science and health research that evaluate the health impact of behaviour and mental health problems.

Conflict of Interests

The authors declare that they have no conflict of interests.

References

- [1] D. Y. Lin, B. M. Psaty, and R. A. Kronmal, "Assessing the sensitivity of regression results to unmeasured confounders in observational studies," *Biometrics*, vol. 54, no. 3, pp. 948–963, 1998.
- [2] S. Greenland, "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," *Journal of the American Statistical Association*, vol. 98, no. 461, pp. 47–54, 2003.
- [3] S. Greenland, J. Copas, D. R. Jones et al., "Multiple-bias modelling for analysis of observational data," *Journal of the Royal Statistical Society A*, vol. 168, no. 2, pp. 267–306, 2005.
- [4] L. C. McCandless, P. Gustafson, and A. Levy, "Bayesian sensitivity analysis for unmeasured confounding in observational studies," *Statistics in Medicine*, vol. 26, no. 11, pp. 2331–2347, 2007.
- [5] O. A. Arah, Y. Chiba, and S. Greenland, "Bias formulas for external adjustment and sensitivity analysis of unmeasured confounders," *Annals of Epidemiology*, vol. 18, no. 8, pp. 637–646, 2008.
- [6] T. M. Palmer, J. R. Thompson, M. D. Tobin, N. A. Sheehan, and P. R. Burton, "Adjusting for bias and unmeasured confounding in Mendelian randomization studies with binary responses," *International Journal of Epidemiology*, vol. 37, no. 5, pp. 1161–1168, 2008.
- [7] T. J. VanderWeele, M. A. Hernán, and J. M. Robins, "Causal directed acyclic graphs and the direction of unmeasured confounding bias," *Epidemiology*, vol. 19, no. 5, pp. 720–728, 2008.
- [8] Y. Chiba, "The sign of the unmeasured confounding bias under various standard populations," *Biometrical Journal*, vol. 51, no. 4, pp. 670–676, 2009.
- [9] W.-C. Lee, "Bounding the bias of unmeasured factors with confounding and effect-modifying potentials," *Statistics in Medicine*, vol. 30, no. 9, pp. 1007–1017, 2011.
- [10] T. J. Vanderweele and O. A. Arah, "Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments, and confounders," *Epidemiology*, vol. 22, no. 1, pp. 42–52, 2011.
- [11] L. C. McCandless, P. Gustafson, A. R. Levy, and S. Richardson, "Hierarchical priors for bias parameters in Bayesian sensitivity analysis for unmeasured confounding," *Statistics in Medicine*, vol. 31, no. 4, pp. 383–396, 2012.
- [12] F. Jacobi, H.-U. Wittchen, C. Hölting et al., "Prevalence, comorbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS)," *Psychological Medicine*, vol. 34, no. 4, pp. 597–611, 2004.
- [13] Y.-F. Chan, M. L. Dennis, and R. R. Funk, "Prevalence and comorbidity of major internalizing and externalizing problems among adolescents and adults presenting to substance abuse treatment," *Journal of Substance Abuse Treatment*, vol. 34, no. 1, pp. 14–24, 2008.
- [14] M. Teesson, W. Hall, T. Slade et al., "Prevalence and correlates of DSM-IV alcohol abuse and dependence in Australia: findings of the 2007 National Survey of Mental Health and Wellbeing," *Addiction*, vol. 105, no. 12, pp. 2085–2094, 2010.
- [15] W. Liang and T. Chikritzhs, "Does light alcohol consumption during pregnancy improve offspring's cognitive development?" *Medical Hypotheses*, vol. 78, no. 1, pp. 69–70, 2012.
- [16] W. Liang and T. Chikritzhs, "Alcohol consumption and health status of family members: health impacts without ingestion," *Internal Medicine Journal*, vol. 43, no. 9, pp. 1012–1016, 2013.
- [17] E. Petridou, X. Zavitsanos, N. Dessypris et al., "Adolescents in high-risk trajectory: clustering of risky behavior and the origins of socioeconomic health differentials," *Preventive Medicine*, vol. 26, no. 2, pp. 215–219, 1997.
- [18] J. W. Lynch, G. A. Kaplan, and J. T. Salonen, "Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse," *Social Science and Medicine*, vol. 44, no. 6, pp. 809–819, 1997.
- [19] A. Rozanski, J. A. Blumenthal, K. W. Davidson, P. G. Saab, and L. Kubzansky, "The epidemiology, pathophysiology, and management of psychosocial risk factors in cardiac practice: the emerging field of behavioral cardiology," *Journal of the American College of Cardiology*, vol. 45, no. 5, pp. 637–651, 2005.
- [20] K.-J. Suzuki, S. Nakaji, S. Tokunaga, T. Shimoyama, T. Umeda, and K. Sugawara, "Confounding by dietary factors in case-control studies on the efficacy of cancer screening in Japan," *European Journal of Epidemiology*, vol. 20, no. 1, pp. 73–78, 2005.
- [21] R. H. H. Groenwold, A. W. Hoes, and E. Hak, "Confounding in publications of observational intervention studies," *European Journal of Epidemiology*, vol. 22, no. 7, pp. 413–415, 2007.
- [22] P. Lagiou, H.-O. Adami, and D. Trichopoulos, "Causality in cancer epidemiology," *European Journal of Epidemiology*, vol. 20, no. 7, pp. 565–574, 2005.
- [23] K. Herttua, P. Mäkelä, and P. Martikainen, "An evaluation of the impact of a large reduction in alcohol prices on alcohol-related and all-cause mortality: time series analysis of a population-based natural experiment," *International Journal of Epidemiology*, vol. 40, no. 2, Article ID dyp336, pp. 441–454, 2011.
- [24] E. Poirier, L. Watier, E. Espie, F.-X. Weill, H. de Valk, and J.-C. Desenclos, "Evaluation of the impact on human salmonellosis of control measures targeted to *Salmonella* Enteritidis and Typhimurium in poultry breeding using time-series analysis and intervention models in France," *Epidemiology and Infection*, vol. 136, no. 9, pp. 1217–1224, 2008.
- [25] K. J. Rothman, S. Greenland, C. Poole, and T. L. Lash, "Causation and causal inference," *Modern Epidemiology*, vol. 2, pp. 7–28, 1998.
- [26] W. Liang and T. Chikritzhs, "Observational research on alcohol use and chronic disease outcome: new approaches to counter biases," *The Scientific World Journal*, vol. 2013, Article ID 860915, 14 pages, 2013.

- [27] M. Zuckerman and D. M. Kuhlman, "Personality and risk-taking: common biosocial factors," *Journal of Personality*, vol. 68, no. 6, pp. 999–1029, 2000.
- [28] R. C. Kessler, C. B. Nelson, K. A. McGonagle, M. J. Edlund, R. G. Frank, and P. J. Leaf, "The epidemiology of co-occurring addictive and mental disorders: implications for prevention and service utilization," *American Journal of Orthopsychiatry*, vol. 66, no. 1, pp. 17–31, 1996.