



Challenges in epidemiological and statistical evaluations of effect modifiers and confounders

Su Yon Jung*

Translational Sciences Section, School of Nursing, University of California Los Angeles, Los Angeles, CA, USA

*Correspondence: sjung@sonnet.ucla.edu

Edited by:

Jimmy Thomas Efird, Brody School of Medicine, USA

Reviewed by:

Eugenia M. Bastos, Bastos Consulting, USA

Wenlong Gao, Lanzhou University, China

Keywords: interactions, effect modifiers, confounders, mediators, effect sizes

In multiple adjusted regression models, researchers sometimes do not know when the interaction occurs and how to interpret the exposure effect estimate while adjusting for the interaction term, resulting in a misinterpretation of the results; this issue has been raised in previous epidemiologic studies. In addition, when the positions of exposure and outcome are switched in the multiple regression, interpreting covariates is challenging.

Here, we present the epidemiological and statistical challenges in evaluating the effect modifier and confounding factor.

WHEN DOES THE INTERACTION OCCUR?

As an example to illustrate aspects of the evaluation of interaction, we describe below. In our example, we want to determine whether radon exposure (the third factor) is an effect modifier in the relationship between smoking (exposure) and lung cancer (outcome). For a dichotomous potential modifier [radon exposure (yes/no)], the interaction occurs when the effect of the exposure (smoking) on the outcome (lung cancer) is not homogenous in strata formed by a third variable (radon exposure) (1–3). The effect can be measured either by the attributable risk (in the additive model) or by a relative risk (in the multiplicative model); both models share the same conceptual basis for evaluating the interaction (1, 2).

To measure the effect of the interaction between smoking and lung cancer, we set the simple regression model and added a

third variable, radon exposure, along with an interaction term of radon exposure with smoking (i.e., radon*smoking).

The regression model is shown below.

$$\begin{aligned} Y \text{ (lung cancer)} \\ &= B_s X_{\text{smoking}} + B_r X_{\text{radon exposure}} \\ &\quad + B_{sr} X_{\text{smoking*radon exposure}} \end{aligned}$$

where B_s is the effect of X_{smoking} , B_r is the effect of $X_{\text{radon exposure}}$, and B_{sr} is the effect of $X_{\text{smoking*radon exposure}}$.

The most important issue in this model for determining whether or not radon exposure is an effect modifier is to interpret the effect size (i.e., B_{sr}); that is, rather than focusing on only the “ p -value” of the effect size, researchers should concentrate on the magnitude of the interaction term. In addition, when outcome variables are continuous and the levels of a continuous variable are small (e.g., 0.01, 0.02, and 0.03), the effect size of the interaction can be small resulting in a small slope in a graph composed of an x -axis for exposure and a y -axis for outcome. This occurs as a result of the tiny interval between outcome variable numerals; it does not suggest that there is no interaction. In this case, we can graphically test the interaction by plotting the means of the outcome variables for each category of exposure according to the strata defined by the effect modifier. Non-parallel lines suggest the presence of interaction. In both statistical and graphical evaluations, we always focus on the effect size of the interaction term rather than only its p -value.

IN THE REGRESSION MODEL THAT INCLUDES EFFECT MODIFICATION, WHY DOES THE EFFECT OF EXPOSURE VARIABLE (B_s) NOT REPRESENT THE RELATIONSHIP BETWEEN EXPOSURE AND OUTCOME?

When the effect modifier (radon exposure) is determined to evaluate its interaction with exposure (smoking) and outcome (lung cancer), it should never be treated as a confounder in the analysis; the interaction effect (B_{sr}) should be considered. In addition, the effect of smoking (B_s) in a regression analysis in which the effect modification (B_{sr}) is adjusted does not solely indicate the association between exposure and outcome, which has interacted with the effect modifier (4). The regression model shown below accounts for the effect modification:

$$\begin{aligned} Y \text{ (lung cancer)} \\ &= B_s X_{\text{smoking}} + B_r X_{\text{radon exposure}} \\ &\quad + B_{sr} X_{\text{smoking*radon exposure}} \end{aligned}$$

When the radon exposure status is no, the effect of smoking is B_s .

When the radon exposure status is yes, the effect of smoking is $B_s + B_{sr}$.

The above two cases of the dichotomous effect modifier indicate that the effect of smoking is not B_s only and that it depends on the effect of the interaction (B_{sr}). Thus, when there is an interaction between smoking and radon exposure in lung cancer (i.e., B_{sr} has a noticeable effect), the effect of smoking must be measured in separate analyses, stratified by radon exposure status, to produce the effect of

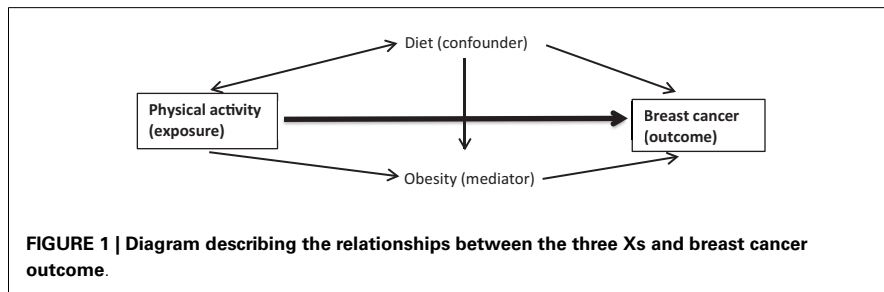


FIGURE 1 | Diagram describing the relationships between the three Xs and breast cancer outcome.

smoking within the radon exposure groups ($B_s; B_s + B_{sr}$).

INTERPRETING COVARIATES (CONFOUNDERS OR MEDIATORS) IN THE REGRESSION MODEL WHERE THE POSITIONS BETWEEN EXPOSURE (X) AND OUTCOME (Y) ARE SWITCHED IS CHALLENGING

Researchers sometimes face the situation where the positions between exposure and outcome are switched. For example, in a multiple regression model (see the below formula and **Figure 1**) of the relationship between physical activity (exposure) and breast cancer (outcome), we evaluate the effect of physical activity on breast cancer risk after accounting for the third factors such as diet (as a confounder) and obesity (as a mediator), as shown below:

$$\begin{aligned}
 Y \text{ (breast cancer: outcome)} &= B_p X_{\text{physical activity (exposure)}} \\
 &+ B_d X_{\text{diet (confounder)}} \\
 &+ B_o X_{\text{obesity (mediator)}}.
 \end{aligned}$$

The diagram in **Figure 1** depicts the relationships between the three Xs and breast cancer outcome.

If researchers want to switch the positions between exposure and outcome, the following formula appears:

$$\begin{aligned}
 Y \text{ (physical activity: outcome)} &= B_b X_{\text{breast cancer (exposure)}}
 \end{aligned}$$

$$\begin{aligned}
 &+ B_d X_{\text{diet}} \text{ (1}^{\text{st}} \text{ of the third variables)} \\
 &+ B_o X_{\text{obesity}} \text{ (2}^{\text{nd}} \text{ of the third variables)}.
 \end{aligned}$$

The new model does not appear to be mathematically problematic. However, the diet variable (X_{diet}) is no longer a confounder because X_{diet} has no causal direction to physical activity outcome. Additionally, the causal relation between exposure and outcome is reversed. For the same reason, the obesity variable (X_{obesity}) is not a mediator (i.e., obesity \rightarrow physical activity; breast cancer \rightarrow physical activity). Therefore, in this new formula, B_b does not reflect the effect of breast cancer adjusted for the confounder and mediator. Furthermore, when researchers set a multiple regression model, they should first understand whether the third variables are included as confounders, mediators, or effect modifiers.

Overall, when using the formula for switching between exposure and outcome and adding third variables without considering their technical roles in a given multiple regression analysis, the results cannot be interpreted correctly to determine the effect of the exposure after adjusting for third variables.

In conclusion, previous epidemiological studies to evaluate effect modification using the multiple regression model may have overlooked an important issue in interpreting the effect modifier estimate by

focusing on only the p -value rather than the effect size. Additionally, they may have misinterpreted the exposure effect in the regression analysis adjusted for the effect modification. Moreover, when the positions are switched between exposure and outcome in the multiple regression, the third variables in the formula do not have the same epidemiological roles as those in the formula prior to the switch; they need to be considered carefully to determine their roles as confounders, mediators, or effect modifiers.

REFERENCES

1. Sukon Kanchanaraks. *Interaction*. The Johns Hopkins University (2008). Available from: <http://ocw.jhsph.edu/courses/fundepiii/PDFs/Lecture20.pdf>
2. Szklo M, Nieto FJ. *Epidemiology: Beyond the Basics*. 3rd ed. Burlington, MA: Jones & Bartlett Learning (2014). p. 185–94.
3. *The Part A Examination of the Faculty of Public Health*. (2014). Available from: www.edmundjessop.org.uk
4. Westreich D, Greenland S. The table 2 fallacy: presenting and interpreting confounder and modifier coefficients. *Am J Epidemiol* (2013) 177(4):292–8. doi:10.1093/aje/kws412

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 July 2014; accepted: 25 November 2014; published online: 10 December 2014.

Citation: Jung SY (2014) Challenges in epidemiological and statistical evaluations of effect modifiers and confounders. *Front. Public Health* 2:277. doi: 10.3389/fpubh.2014.00277

This article was submitted to *Epidemiology*, a section of the journal *Frontiers in Public Health*.

Copyright © 2014 Jung. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.