CONFOUNDING, HOMOGENEITY AND COLLAPSIBILITY FOR CAUSAL EFFECTS IN EPIDEMIOLOGIC STUDIES

Zhi Geng[1], Jianhua Guo[1&3], Tai Shing Lau[2] and Wing-Kam Fung[4]

[1]*Peking University,* [2]*The Chinese University of Hong Kong,*
[3]*Jilin University and* [4]*The University of Hong Kong*

Abstract: Detection of confounding and confounders is important for observational studies, and especially so for epidemiologic studies. Miettinen and Cook (1981) derived two criteria for detecting confounders. Using a model, Wickramaratne and Holford (1987) proved that the two criteria are necessary but not sufficient conditions for confounders. We take uniform nonconfounding to mean there is no confounding at a coarse-subpopulation-level obtained by pooling any number of subpopulations. We discuss the necessity and sufficiency of the two criteria for uniform nonconfounding. The concepts of homogeneity and collapsibility for causal effects are also defined, and the relation among confounding, homogeneity and collapsibility is discussed. We show that the common causal effect over all fine subpopulations is just the causal effect of the whole population.

Key words and phrases: Causal effect, collapsibility, confounding, effect modification, homogeneity.

1. Introduction

The presence of confounding in an epidemiologic study hinders the estimation of causal effects. Miettinen and Cook (1981) derived two criteria of a confounder:

(a) it must be predictive of risk in the unexposed population; and

(b) it must have different distributions between the exposed and unexposed populations.

Greenland and Robins (1986) discussed a logical connection among identifiability, exchangeability and confounding. Using a model, Wickramaratne and Holford (1987) showed that the above criteria are necessary but not sufficient conditions for confounding. Based on the counterfactual model (Rubin (1974) and Holland (1986)), Holland (1989) discussed causal effects and confounding in epidemiologic studies. Holland and Rubin (1988) defined causal effects for three different levels: unit-level, subpopulation-level and population-level. The population-level causal effect is measured by comparing the proportion of diseased in the exposed population with the proportion of diseased in the same

population had it not been exposed. Subpopulations are defined by the values of covariates, and a coarse subpopulation is defined by pooling several subpopulations together. The concept of coarse-subpopulation introduced in this paper is important for discussion of the necessity and sufficiency of Miettinen and Cook's two criteria. Similarly, we can measure subpopulation-level and coarse-subpopulation-level causal effects. Uniform nonconfounding occurs if there is no confounding in any coarse-subpopulation-level. Otherwise we say that there is occasional confounding. It is shown in Section 2 that (a) and (b) are necessary and sufficient for occasional confounding under the model of Wickramaratne and Holford (1987). In Section 3, the homogeneity of causal effects is said to hold if the causal effects are equal over all fine subpopulations. We discuss relationships among confounding, homogeneity and collapsibility. In the absence of occasional confounding, causal effects for the whole population, subpopulations and coarse subpopulations can be estimated correctly from observed data. However, causal effects at different coarse-subpopulation-levels may not be equal to one another even if there is no confounding. We show that the common value of an association measure (e.g. the common value of risk difference or the common value of risk ratio) over all fine subpopulations is the corresponding causal effect of the whole exposed population if there is no confounding in any fine subpopulation. Causal effects are strongly collapsible if they are equal for all different population-levels. A necessary and sufficient condition for strong collapsibility is found. When causal effects are strongly collapsible, we can estimate the causal effects based on those subpopulation-level results that are free from confounding. Simpson (1951), Bishop (1971), Armitage (1975), Whittemore (1978), Wermuth (1987), Geng (1992) and Guo and Geng (1995) have discussed collapsibility of association measures. Recently, Greenland, Robins and Pearl (1999) provided a historical overview of confounding and confounder in causal inference, and gave the background on causal models that complements the results in the present paper.

2. Causal Effects and Confounding

Following the notation of Holland (1989), we let E be an exposure with values e and $\bar{e}$ representing the presence and absence of the exposure, respectively. D_e and $D_{\bar{e}}$ are responses under the exposure $E = e$ and $E = \bar{e}$. The responses take values 1 and 0 denoting diseased and nondiseased, respectively.

Epidemiologic studies look for causal effects in an exposed population. Let $P(D_e = 1 \mid E = e)$ and $P(D_{\bar{e}} = 1 \mid E = \bar{e})$ be the proportions of diseased in the exposed population and the unexposed population respectively, and let $P(D_{\bar{e}} = 1 \mid E = e)$ be the hypothetical proportion of exposed individuals who would have developed the disease had they not been exposed. Causal effects are

measured by comparing $P(D_e = 1 \mid E = e)$ with $P(D_{\bar{e}} = 1 \mid E = e)$, for example, $P(D_e = 1 \mid E = e) - P(D_{\bar{e}} = 1 \mid E = e)$ or $P(D_e = 1 \mid E = e)/P(D_{\bar{e}} = 1 \mid E = e)$ (Rosenbaum and Rubin (1983)). According to Rothman and Greenland (1998), we call $P(D_e = 1 \mid E = e) - P(D_{\bar{e}} = 1 \mid E = e)$ the causal risk difference (CRD) for the whole exposed population, and call $P(D_e = 1 \mid E = e)/P(D_{\bar{e}} = 1 \mid E = e)$ the causal risk ratio (CRR) for the whole exposed population. Since $P(D_{\bar{e}} = 1 \mid E = e)$ is the hypothetical proportion, the model is called a counterfactual model (Rubin (1974), Holland (1986), Holland and Rubin (1988)). In the following we consider the causal risk difference first. The risk difference (RD) usually used to measure associations in the whole exposed population is $RD = P(D_e = 1 \mid E = e) - P(D_{\bar{e}} = 1 \mid E = \bar{e})$, which certainly is estimable from the data in a cohort study. If $P(D_{\bar{e}} = 1 \mid E = e) = P(D_{\bar{e}} = 1 \mid E = \bar{e})$, then CRD is equal to RD. One says there is nonconfounding in the whole population or, simply, one has population nonconfounding (Miettinen and Cook (1981), Wickramaratne and Holford (1987), Greenland and Robins (1986)).

Let C be an extraneous factor with values $1, \ldots, K$, where each value defines a fine subpopulation; C is a covariate in the sense that it is not affected by the exposure or it is not in the causal sequence from exposure to disease. Variables measured prior to the exposure are always covariates. Let $P(D_e = 1 \mid E = e, C = k)$ be the proportion of diseased in the exposed subpopulation $C = k$. Similarly, $P(D_{\bar{e}} = 1 \mid E = e, C = k)$ is the hypothetical proportion of individuals in the exposed subpopulation $C = k$ who would have developed the disease had they not been exposed. The subpopulation-level causal risk difference for the exposed subpopulation with $C = k$ is $CRD_k = P(D_e = 1 \mid E = e, C = k) - P(D_{\bar{e}} = 1 \mid E = e, C = k)$. The risk difference for association in the exposed subpopulation with $C = k$ is given by $RD_k = P(D_e = 1 \mid E = e, C = k) - P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k)$. Let $A \perp\!\!\!\perp B \mid C$ denote conditional independence between A and B given C (Dawid (1979)). If $P(D_{\bar{e}} = 1 \mid E = e, C = k) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k)$ for all k (denoted by $D_{\bar{e}} \perp\!\!\!\perp E \mid C$), then $CRD_k = RD_k$ for all k, and we say that there is no confounding in the fine subpopulations or, simply, there is subpopulation nonconfounding (Wickramaratne and Holford (1987)).

There are two important assumptions for causal inference in observational studies. One is strongly ignorable treatment assignment: $(D_e, D_{\bar{e}}) \perp\!\!\!\perp E \mid C$ (Rosenbaum and Rubin (1983), Rosenbaum (1984)). That is, the responses and the treatment assignment are conditionally independent given C. The other is weakly ignorable treatment assignment: $D_e \perp\!\!\!\perp E \mid C$ and $D_{\bar{e}} \perp\!\!\!\perp E \mid C$ (Rubin (1978), Stone (1993)). We can see that subpopulation nonconfounding (i.e., $D_{\bar{e}} \perp\!\!\!\perp E \mid C$) is weaker than both strong ignorability and weak ignorability. Subpopulation nonconfounding can be satisfied by selecting a control group which is comparable with the treatment group had the individuals in this treatment

group not been treated. In addition, weak ignorability requires that the control group is also comparable with the treatment group had the individuals in the control group been treated. In epidemiologic studies, it is customary to restrict the causal effects to individuals in the exposed population. Thus subpopulation nonconfounding is a more essential assumption than ignorability for causal inference in epidemiologic studies.

Let ω be a nonempty subset of the range of C, that is, $\omega \subseteq \{1, \ldots, K\}$ and $\omega \neq \emptyset$. The causal risk difference for the coarse exposed subpopulation of $C \in \omega$ is given by $CRD_\omega = P(D_e = 1 \mid E = e, C \in \omega) - P(D_{\bar{e}} = 1 \mid E = e, C \in \omega)$. The word "coarse" means that a "coarse" subpopulation consists of several fine subpopulations. Similarly we denote the risk difference for association as $RD_\omega = P(D_e = 1 \mid E = e, C \in \omega) - P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega)$. If $P(D_{\bar{e}} = 1 \mid E = e, C \in \omega) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega)$ for all ω, then there is no confounding in any coarse subpopulation, and thus $CRD_\omega = RD_\omega$ for all ω. We call it uniform nonconfounding. This means no confounding occurs no matter how the fine subpopulations are pooled into coarse subpopulations. In particular, the case $\omega = \{1, \ldots, K\}$ implies that uniform nonconfounding is stronger than population nonconfounding. In many studies, we are not only interested in the causal effect for the whole exposed population but also interested in those for coarse subpopulations. For example, epidemiologists routinely consider coarse populations formed by different age groups, groupings of household income levels, groupings of occupations, and so on. Confounding may arises in some coarse populations but not in others. Thus, recognizing uniform nonconfounding is also important in epidemiologic studies.

Example 1. Consider an example with $K = 3$. From the probabilities given in Table 1, we see that the criteria of Miettinen and Cook are satisfied but there is no confounder.

Table 1. C is not a confounder even if it satisfies Miettinen and Cook's criteria.

	$P(D_s = 1 \mid E, C)$ for $s = e$ or $\bar{e}$		$E = \bar{e}$	$P(C \mid E)$	
	$E = e$		$E = \bar{e}$	E	
C	$D_e = 1$	$D_{\bar{e}} = 1$	$D_{\bar{e}} = 1$	e	$\bar{e}$
1	0.5	0.2	0.2	0.2	0.4
2	0.7	0.4	0.4	0.2	0.4
3	0.6	0.3	0.3	0.6	0.2

(1) There is no confounding in the fine subpopulations since $P(D_{\bar{e}} = 1 \mid E = e, C = k) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k)$ for $k = 1, 2, 3$. Thus $CRD_k = RD_k = 0.3$ for all k.

(2) There is no confounding in the coarse subpopulation of $C \in \omega_1$ where $\omega_1 = \{1, 2\}$, since

$$P(D_{\bar{e}} = 1 \mid E = e, C \in \omega_1)$$
$$= \frac{\sum_{k \in \omega_1} P(D_{\bar{e}} = 1 \mid E = e, C = k)P(C = k \mid E = e)}{\sum_{k \in \omega_1} P(C = k \mid E = e)} = 0.3,$$

which is equal to $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega_1)$. Also $CRD_{\omega_1} = RD_{\omega_1} = P(D_e = 1 \mid E = e, C \in \omega_1) - P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega_1) = 0.6 - 0.3 = 0.3$.

(3) There is confounding in the coarse subpopulation of $C \in \omega_2 = \{2, 3\}$. Here $P(D_{\bar{e}} = 1 \mid E = e, C \in \omega_2) = 0.325$, but $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega_2) = 0.367$. Thus we cannot claim uniform nonconfounding since $RD_{\omega_2} = 0.625 - 0.367 = 0.258$ and $CRD_{\omega_2} = 0.625 - 0.325 = 0.3$.

(4) There is no confounding in the whole population since $P(D_{\bar{e}} = 1 \mid E = e) = P(D_{\bar{e}} = 1 \mid E = \bar{e}) = 0.30$. Thus C is not a confounder even if it does satisfy Miettinen and Cook's criteria: it is predictive of risk in the unexposed population and has different distributions between the exposed and unexposed populations. It can be seen that Miettinen and Cook's criteria are not sufficient conditions for confounders. It can be shown that $CRD = RD = 0.60 - 0.30 = 0.30$.

We show that Miettinen and Cook's criteria are necessary and sufficient for uniform nonconfounding under the assumption of subpopulation nonconfounding.

Theorem 1. *Assuming subpopulation nonconfounding, the necessary and sufficient condition for uniform nonconfounding is*
(i) $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k) = P(D_{\bar{e}} = 1 \mid E = \bar{e})$ *for* $k = 1, \ldots, K$, *or*
(ii) $P(C = k \mid E = \bar{e}) = P(C = k \mid E = e)$ *for* $k = 1, \ldots, K$.

See the Appendix for a proof.

Condition (i) means that the conditional independence of $D_{\bar{e}}$ and C given $E = \bar{e}$ is the negation of condition (a) for a confounder. Condition (ii) is the marginal independence of C and E and is the negation of condition (b). Under the same condition of subpopulation nonconfounding, Wickramaratne and Holford (1987) showed only that conditions (i) and (ii) are sufficient (but not necessary) conditions for population nonconfounding.

In clinical trials, randomized treatment assignment ensures that the treatment assignment is independent of all other factors, which implies $(D_e, D_{\bar{e}}, C) \perp\!\!\!\perp E$ (Rubin (1978)). Thus randomized treatment assignment can ensure that both condition (ii) and the assumption of subpopulation nonconfounding hold. Stratified-randomized treatment assignment in each subpopulation can ensure that the treatment assignment is independent of other factors in each subpopulation,

which implies $D_{\bar{e}} \perp\!\!\!\perp E \mid C$. In cohort studies, there are three designs— frequency-matching (Weinberg (1985)), pair-matching and stratification (Kleinbaum, Kupper and Morgenstern (1982)), which can ensure that C has same distributions between the exposed and unexposed populations. For stratified cohort studies, condition (i) should be rewritten as $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k')$ for $k \neq k'$.

Since the uniform nonconfounding implies subpopulation nonconfounding, we can rewrite Theorem 1 as follows.

Theorem 1′. *The necessary and sufficient condition for uniform nonconfounding is* (i) $D_{\bar{e}} \perp\!\!\!\perp C \mid E = \bar{e}$ *or* $C \perp\!\!\!\perp E$ *and* (ii) $D_{\bar{e}} \perp\!\!\!\perp E \mid C$.

We say that there is occasional confounding if there is a coarse subpopulation with confounding. For Table 1, there is occasional confounding in the coarse subpopulation of $\omega_2 = \{2, 3\}$ since $P(D_{\bar{e}} = 1 \mid E = e, C \in \omega_2) = 0.325 \neq P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega_2) = 0.367$.

Corollary 1. *There is occasional confounding if and only if* (i) $D_{\bar{e}} \not\perp\!\!\!\perp C \mid E = \bar{e}$ *and* $C \not\perp\!\!\!\perp E$, *or* (ii) $D_{\bar{e}} \not\perp\!\!\!\perp E \mid C$.

Condition (i) in Corollary 1 is the criteria of Miettinen and Cook (1981). It is sufficient for occasional confounding and can be verified by observed data. Consider again Table 1. We find $D_{\bar{e}} \not\perp\!\!\!\perp C \mid E = \bar{e}$ and $C \not\perp\!\!\!\perp E$. Thus there must be occasional confounding, and it is found in the subpopulation ω_2.

3. Homogeneity and Collapsibility for Causal Effects

Uniform nonconfounding means that causal effects for the whole exposed population, subpopulations and coarse-subpopulations can be estimated from observed data. However uniform nonconfounding does not imply causal effects at different population-levels are equal. In this section we discuss collapsibility of causal effects and give a necessary and sufficient condition for collapsibility. We discuss the relationships among confounding, homogeneity and collapsibility.

If causal risk differences for all fine exposed subpopulations are the same, we say that causal risk differences are homogeneous across C, and we denote the common causal risk difference by $_aCRD$. Thus homogeneity means that there is no effect modification for CRDs. Moreover, if the causal risk difference in the whole exposed population is equal to those for each fine exposed subpopulations, that is, we say that the causal risk differences are collapsible over C. Further, if the causal risk difference in the whole exposed population is equal to those for each coarse exposed populations ($CRD = CRD_\omega$ for all ω), we say that the causal risk differences are strongly collapsible over C. Thus strong collapsibility implies collapsibility, which in turn implies homogeneity. The collapsibility condition

is introduced to avoid dramatic change of causal effects. When the causal risk differences are strongly collapsible, we may obtain the CRD from some coarse subpopulations even if $CRD \neq RD$. More explicitly, if there is no confounding in the subpopulation $C = k$, we have $CRD = RD_k$ for all k because of collapsibility.

To illustrate homogeneity and collapsibility of causal effects, consider again Table 1.

(1) The causal risk differences are homogeneous across C since $CRD_k = 0.3$ for $k = 1$, 2 and 3, so $_aCRD = 0.3$.
(2) The causal risk differences are collapsible over C since $CRD = {}_aCRD = 0.3$.
(3) $CRD = CRD_\omega = 0.3$ for all nonempty $\omega \subseteq \{1, 2, 3\}$. Thus the causal risk differences are strongly collapsible over C.

It seems that strong collapsibility is stronger than collapsibility which in turn is stronger than homogeneity. Theorem 2 shows these conditions are equivalent for causal risk differences.

Theorem 2. *Causal risk differences are strongly collapsible over C if and only if causal risk differences are homogeneous across C.*

See the Appendix for a proof.

It is not by chance that the causal risk differences in Table 1 are both homogeneous and collapsible. If there is no confounding in any subpopulation, then the homogeneity of causal risk differences becomes the homogeneity of RD_k for all k. For Table 1, we see that both RD_k and CRD_k are homogeneous across C since there is no confounding in any subpopulation. Denote the common value by $_aRD$. The test of homogeneity of risk differences is suggested by Gart and Nam (1990) and Lipsitz, Dear, Laird and Molenberghs (1998). In a stratified cohort study, causal risk difference for the whole exposed population and coarse subpopulations cannot be obtained since there is no information on the distribution of C. But they can be obtained from $_aRD$ if RDs are homogeneous and there is no confounding in any subpopulation.

Finally, we consider the causal risk ratio. The risk ratio (RR) usually used for association in the whole exposed population is $RR = P(D_e = 1 \mid E = e)/P(D_{\bar{e}} = 1 \mid E = \bar{e})$. Similar to CRD, population nonconfounding occurs if $P(D_{\bar{e}} = 1 \mid E = e) = P(D_{\bar{e}} = 1 \mid E = \bar{e})$ or, equivalently, $CRR = RR$. Similarly, we define CRR_k, $_aCRR$, CRR_ω, RR_k and $_aRR$. While the homogeneity and collapsibility of causal measure depends on which measure is considered, the existence of confounding does not. This can be seen from Table 1, where the CRDs are both homogeneous and collapsible, but the CRRs are neither homogeneous nor collapsible. The following result for the CRR can be shown in a similar way for the CRD.

Theorem 3. *The CRRs are strongly collapsible over C if and only if the CRRs are homogeneous across C.*

From Theorem 3 we can see the equivalence of homogeneity, collapsibility and strong collapsibility for the causal risk ratios. Also, the homogeneity of CRRs becomes the homogeneity of the RRs, and this can be tested when there is no confounding in any subpopulation (Gart (1985)).

There are a lot of methods for estimating a common association measure over subpopulations defined by a covariate C. Suppose the condition of subpopulation nonconfounding holds, for example, a stratified-randomized experiment, or in a cohort study where confounding in subpopulation can be removed by stratifying the population into subpopulations defined by a covariate C. Then homogeneity of CRDs and CRRs is equivalent to the homogeneity of RDs and RRs, respectively. According to Theorems 2 and 3, if CRDs (or CRRs) are homogeneous, then $_aCRD$ (or $_aCRR$) is equal to all CRDs (or CRRs) for different population-levels, and equal to CRD (or CRR) for the whole exposed population. Even if there exists confounding in the whole population so that $CRD \neq RD$ (or $CRR \neq RR$), we still have $CRD = {}_aRD$ (or $CRR = {}_aRR$). If $_aRD \neq RD$ (or $_aRR \neq RR$), then there must exist confounding in the whole population, and then $_aRD$ (or $_aRR$) can be considered as a non-confounded estimate of CRD (or CRR).

There are two broad criteria for confounding (Greenland and Robins (1986)). One is "comparability-based", that is, the type of criteria at (a) and (b). The other is "collapsibility-based", that is, the collapsibility of association measures over a covariate C. We are going to show the relation between the two.

Corollary 2. *Assuming subpopulation nonconfounding, the collapsibility of RDs (or RRs) is equivalent to* (i) *population nonconfounding and* (ii) *RDs (or RRs) are homogeneous over C.*

Proof. We only need to prove the result for RDs since the proof for RRs is the same. The subpopulation nonconfounding means that $CRD_k = RD_k$ for all k. Thus the homogeneity of RDs is equivalent to the homogeneity of CRDs. By Theorem 2 and subpopulation nonconfounding, the homogeneity of RDs is also equivalent to the collapsibility of CRDs, which together with population nonconfounding is equivalent to the collapsibility of RDs.

For Table 1, we have shown both population nonconfounding and the homogeneity of RDs, which is equivalent to RDs being collapsible over C. RRs are not homogeneous and thus not collapsible. Corollary 2 shows the relation between collapsibility and population nonconfounding. For randomized trials, Gail, Wieand and Piantadosi (1984) and Gail (1986) discussed consistent estimates of

treatment effect in nonlinear models without interaction between the treatment assignment and the covariate C. Randomization implies subpopulation nonconfounding, and no interaction and consistency correspond to homogeneity and collapsibility in our discussion. Thus it can be seen that, under the assumption of randomization and no interaction, statistical inferences on causal risk differences and causal risk ratios are not affected by omitting a covariate. The following result shows the relation between the strong collapsibility and uniform nonconfounding.

Corollary 3. *Assuming subpopulation nonconfounding, the strong collapsibility of RDs (or RRs) is equivalent to* (i) *uniform nonconfounding and* (ii) *RDs (or RRs) are homogeneous over C.*

Proof. We give the proof for RDs. Similar to the proof of Corollary 2 we obtain, under the assumption of subpopulation nonconfounding, that the homogeneity of RDs is equivalent to strong collapsibility of CRDs by Theorem 3, but not to the strong collapsibility of RDs. The uniform nonconfounding means that $CRD_\omega = RD_\omega$ for all ω. Thus, under the assumption of subpopulation nonconfounding, the homogeneity of RDs together with the uniform nonconfounding is equivalent to the strong collapsibility of RDs.

Condition (ii) in Corollaries 2 and 3 also means that there is no effect modification. According to Corollaries 2 and 3, we should identify confounding using the collapsibility-based criteria after confirming homogeneity. For Table 1, we have shown that RDs are collapsible, that is, $RD = RD_k$ for all $k = 1$, 2 and 3. According to Corollary 2, we can assess population nonconfounding and thus $CRD = RD$. But RDs are not strongly collapsible over C since $RD_{\omega_2} = 0.258 \neq RD = 0.3$. By Corollary 3, there must exist occasional nonconfounding, and thus there is an ω such that $CRD_\omega \neq RD_\omega$. In fact, even without the collapsibility or strong collapsibility of RDs (or RRs), we still can estimate CRD (or CRR) using estimate of $_aRD$ (or $_aRR$). The comparability-based criteria can be used to identify confounding even if measures are not homogeneous. Under the assumption of subpopulation nonconfounding and homogeneity (i.e., no effect modification), comparability-based and collapsibility-based criteria are equivalent.

Related work on collapsibility of the risk ratio and risk difference is in papers by Wermuth (1987), Geng (1992), Guo and Geng (1995), among others. Miettinen and Cook (1981) and Greenland (1996) showed by counterexample that nonconfounding is neither necessary nor sufficient for collapsibility of odds ratios or rate ratios. Example 3 of Miettinen and Cook (1981) can be used to illustrate that odds ratios do not admit results analogous to Corollary 2 or 3.

4. Examples Related to the Two Approaches to Confounding

Example 2. This example illustrates the equivalency of the two criteria, comparability-based and collapsibility-based, for confounding under the homogeneity of association measures and subpopulation nonconfounding. Suppose that

$$P(D_e = 1 \mid E = e, C = 1) = 2/3, \quad P(D_e = 1 \mid E = e, C = 2) = 1/3,$$

$$P(D_{\bar{e}} = 1 \mid E = e, C = 1) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = 1) = 1/2,$$

$$P(D_{\bar{e}} = 1 \mid E = e, C = 2) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = 2) = 1/6,$$

$$P(C = 1 \mid E = e) = 1/4, \quad P(C = 1 \mid E = \bar{e}) = 3/4.$$

It can be seen that the RDs are homogeneous over C: $RD_1 = P(D_e = 1 \mid E = e, C = 1) - P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = 1) = 1/6 = RD_2$. Since there is no confounding in the subpopulations $C = 1$ and $C = 2$, we have that $CRD_k = RD_k = 1/6$ for $k = 1, 2$. From homogeneity, we have $CRD = CRD_1 = CRD_2$. So we find $CRD = CRD_k = RD_k = 1/6$ for any k. The RDs are not simply collapsible over C since $RD = 0 \neq RD_k = 1/6$. According to the collapsibility-based criteria, there must be confounding in the whole population since RDs are homogeneous but not collapsible. It can also be seen by the comparability-based criteria that there is confounding in the whole population, since $CRD \neq RD$. In this case, even if there is confounding in the whole population, we can still obtain CRD from the $_aRD$.

In this example, since RRs are not homogeneous, the collapsibility of RRs cannot be used as a criterion for identifying population confounding. That is, there exists effect modification for CRRs.

Example 3. This example illustrates that the comparability-based criteria can be used to identify confounding even if RDs and RRs are non-homogeneous. Consider the same supposition as in Example 2 except that $P(D_{\bar{e}} = 1 \mid E = e, C = 1) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = 1) = 1/6$. It can be seen that there is no confounding in subpopulations. Since $RD_1 = 1/2$, $RD_2 = 1/6$ and $RR_1 = 4$, $RR_2 = 2$, we cannot use collapsibility to identify confounding. According to Theorem 1', we can assess uniform nonconfounding because $D_{\bar{e}} \perp\!\!\!\perp C \mid E = \bar{e}$. Thus we get $CRD = RD = 1/4$.

Acknowledgement

Appendix

Proof of Theorem 1. For sufficiency, we have

$$P(D_{\bar{e}}=1 \mid E=\bar{e}, C \in \omega) = \sum_{k \in \omega} P(D_{\bar{e}} = 1 \mid E=\bar{e}, C=k)P(C=k \mid E=\bar{e}, C \in \omega),$$
(1)

$$P(D_{\bar{e}}=1 \mid E=e, C \in \omega) = \sum_{k \in \omega} P(D_{\bar{e}}=1 \mid E=e, C=k)P(C=k \mid E=e, C \in \omega).$$
(2)

If condition (ii) in Theorem 1 holds, $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega) = \sum_{k \in \omega} P(D_{\bar{e}} = 1 \mid E = e, C = k)P(C = k \mid E = e, C \in \omega) = P(D_{\bar{e}} = 1 \mid E = e, C \in \omega)$.

If the condition (i) holds, from (1) we get $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega) = P(D_{\bar{e}} = 1 \mid E = \bar{e}) \sum_{k \in \omega} P(C = k \mid E = \bar{e}, C \in \omega) = P(D_{\bar{e}} = 1 \mid E = \bar{e})$.

Similarly, from (2) we have $P(D_{\bar{e}} = 1 \mid E = e, C \in \omega) = P(D_{\bar{e}} = 1 \mid E = \bar{e})$. Thus $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega) = P(D_{\bar{e}} = 1 \mid E = e, C \in \omega)$.

For necessity, we first consider the case of $\omega = \{i, j\}$ for any two levels $i \neq j$. From (1) and (2) we get $0 = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C \in \omega) - P(D_{\bar{e}} = 1 \mid E = e, C \in \omega) = [P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = i) - P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = j)][P(C = i \mid E = \bar{e}, C \in \{i, j\}) - P(C = i \mid E = e, C \in \{i, j\})]$. Thus one of (3) or (4) holds where

$$P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = i) = P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = j),$$
(3)

$$P(C = i \mid E = \bar{e}, C \in \{i, j\}) = P(C = i \mid E = e, C \in \{i, j\}).$$
(4)

Since $P(C = i \mid E = \bar{e}) = P(C = i \mid E = \bar{e}, C \in \{i, j\})P(C \in \{i, j\} \mid E = \bar{e})$, and $P(C = j \mid E = \bar{e}, C \in \{i, j\}) = 1 - P(C = i \mid E = \bar{e}, C \in \{i, j\})$, (4) is equivalent to

$$\frac{P(C = i \mid E = \bar{e})}{P(C = i \mid E = e)} = \frac{P(C \in \{i, j\} \mid E = \bar{e})}{P(C \in \{i, j\} \mid E = e)} = \frac{P(C = j \mid E = \bar{e})}{P(C = j \mid E = e)}.$$
(5)

Now we show that (3) holds simultaneously for all pairs ω or that (5) holds simultaneously for all pairs ω. Suppose there exists a pair $\{i, j\}$ with $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = i) \neq P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = j)$. Then (5) must hold simultaneously for all pairs. In fact, for any level k, either $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k) \neq P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = i)$ or $P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = k) \neq P(D_{\bar{e}} = 1 \mid E = \bar{e}, C = j)$ from the supposition. Thus from (5), for any k,

$$\frac{P(C = k \mid E = \bar{e})}{P(C = k \mid E = e)} = \frac{P(C = i \mid E = \bar{e})}{P(C = i \mid E = e)} = \frac{P(C = j \mid E = \bar{e})}{P(C = j \mid E = e)}.$$

So

$$\frac{P(C = k \mid E = \bar{e})}{P(C = k \mid E = e)} = \frac{\sum_m P(C = m \mid E = \bar{e})}{\sum_m P(C = k \mid E = e)} = 1$$

for all k, that is (ii).

Proof of Theorem 2. Necessity is obvious since strong collapsibility implies homogeneity. For sufficiency, since $P(D_e = 1 \mid E = e, C \in \omega) = \sum_{k \in \omega} P(D_e = 1 \mid E = e, C = k)P(C = k \mid E = e, C \in \omega)$ and $P(D_{\bar{e}} = 1 \mid E = e, C \in \omega) = \sum_{k \in \omega} P(D_{\bar{e}} = 1 \mid E = e, C = k)P(C = k \mid E = e, C \in \omega)$, we obtain $P(D_e = 1 \mid E = e, C \in \omega) - P(D_{\bar{e}} = 1 \mid E = e, C \in \omega) = \sum_{k \in \omega}[P(D_e = 1 \mid E = e, C = k) - P(D_{\bar{e}} = 1 \mid E = e, C = k)]P(C = k \mid E = e, C \in \omega)$.

From the homogeneity of causal effects for all subpopulations we have, for any k, $P(D_e = 1 \mid E = e, C \in \omega) - P(D_{\bar{e}} = 1 \mid E = e, C \in \omega) = P(D_e = 1 \mid E = e, C = k) - P(D_{\bar{e}} = 1 \mid E = e, C = k)$.

References

Armitage, P. (1975). The use of the cross-ratio in aetiological surveys. In *Perspectives in Probability and Statistics: Papers in Honour of M. S. Bartlett on the Occasion of His Sixty-Fifth Birthday* (Edited by J. Gani), 349-355. Wiley, New York.

Bishop, Y. M. M. (1971). Effects of collapsing multidimensional contingency tables. *Biometrics* **27**, 545-562.

Dawid, A. P. (1979). Conditional independence in statistical theory. *J. Roy. Statist. Soc. Ser. B* **41**, 1-31.

Gail, M. H. (1986). Adjusting for covariates that have the same distribution in exposed and unexposed cohorts. In *Modern Statistical Methods in Chronic Disease Epidemiology* (Edited by S. H. Moolgavkar and R. L. Prentice), 3-18. Wiley, New York.

Gail, M. H., Wieand, S. and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71**, 431-444.

Gart, J. J. (1985). Approximate tests and interval estimation of the common relative risk in the combination of 2×2 tables. *Biometrika* **72**, 673-677.

Gart, J. J. and Nam, J. (1990). Approximate interval estimation of the difference in binomial parameters: correction for skewness and extension to multiple tables. *Biometrics* **46**, 637-643.

Geng, Z. (1992). Collapsibility of relative risks in contingency tables with a response variable. *J. Roy. Statist. Soc. Ser. B* **54**, 585-593.

Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* **7**, 498-501.

Greenland, S. and Robins, J. M. (1986). Identifiability, exchangeability, and epidemiologic confounding. *Internat. J. Epidemiology* **15**, 413-419.

Greenland, S., Robins, J. M. and Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statist. Sci.* **14**, 29-46.

Guo, J. H. and Geng, Z. (1995). Collapsibility of logistic regression coefficients. *J. Roy. Statist. Soc. Ser. B* **57**, 263-267.

Holland, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81**, 945-970.

Holland, P. W. (1989). Reader reactions: confounding in epidemiologic studies. *Biometrics* **45**, 1310-1316.

Holland, P. W. and Rubin, D. B. (1988). Causal inference in retrospective studies. *Evaluation Review* **12**, 203-231.

Kleinbaum, D. G., Kupper, L. L. and Morgenstern, H. (1982). *Epidemiologic Research: Principles and Quantitative Methods*. Van Nostrand Reinhold, New York.

Lipsitz, S. R., Dear, K. B. G., Laird, N. M. and Molenberghs, G. (1998). Tests for homogeneity of the risk difference when data are sparse. *Biometrics* **54**, 148-160.

Miettinen, O. S. and Cook, E. F. (1981). Confounding: essence and detection. *Amer. J. Epidemiology* **114**, 593-603.

Rosenbaum, P. R. (1984). From association to causation in observational studies: the role of tests of strongly ignorable treatment assignment. *J. Amer. Statist. Assoc.* **79**, 41-48.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41-55.

Rothman, K. J. and Greenland, S. (1998). *Modern Epidemiology*. 2nd edition. Lippintcott-Raven, USA.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educational Psychology* **66**, 688-701.

Rubin, D. B. (1978). Bayesian inference for causal effects: the role of randomization. *Ann. Statist.* **6**, 34-58.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc. Ser. B* **13**, 238-241.

Stone, R. (1993). The assumptions on which causal inferences rest. *J. Roy. Statist. Soc. Ser. B* **55**, 455-466.

Weinberg, C. R. (1985). On pooling across strata when frequency matching has been followed in a cohort study. *Biometrics* **41**, 117-127.

Wermuth, N. (1987). Parametric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. Roy. Statist. Soc. Ser. B* **49**, 353-364.

Whittemore, A. S. (1978). Collapsibility of multidimensional contingency tables. *J. Roy. Statist. Soc. Ser. B* **40**, 328-340.

Wickramaratne, P. J. and Holford, T. R. (1987). Confounding in epidemiologic studies: the adequacy of the control groups as a measure of confounding. *Biometrics* **43**, 751-765.

Department of Probability and Statistics, Peking University, Beijing 100871, China.

E-mail: zgeng@math.pku.edu.cn

E-mail: jhguo@statms.stat.pku.edu.cn

Department of Statistics, The Chinese University of Hong Kong, Shatin, Hong Kong.

E-mail: tslau@sparc2.sta.cuhk.edu.hk

Institute of Mathematics, Jilin University, Changchun 130023, Jilin Province, China.

Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong.

E-mail: hrntfwk@hku.hk