

Case-control matching: effects, misconceptions, and recommendations

Mohammad Ali Mansournia¹  · Nicholas Patrick Jewell^{2,3} · Sander Greenland^{4,5}

Received: 6 February 2017 / Accepted: 7 October 2017
© Springer Science+Business Media B.V. 2017

Abstract Misconceptions about the impact of case-control matching remain common. We discuss several subtle problems associated with matched case-control studies that do not arise or are minor in matched cohort studies: (1) matching, even for non-confounders, can create selection bias; (2) matching distorts dose-response relations between matching variables and the outcome; (3) unbiased estimation requires accounting for the actual matching protocol as well as for any residual confounding effects; (4) for efficiency, identically matched groups should be collapsed; (5) matching may harm precision and power; (6) matched analyses may suffer from sparse-data bias, even when using basic sparse-data methods. These problems support advice to limit case-control matching to a few strong well-measured confounders, which would devolve to no matching if no such confounders are measured. On the positive side, odds ratio modification by matched variables can be assessed in matched case-control studies without further data, and when one knows either the distribution of the matching factors or their relation to the outcome in the

source population, one can estimate and study patterns in absolute rates. Throughout, we emphasize distinctions from the more intuitive impacts of cohort matching.

Keywords Bias · Case-control studies · Confounding · Matching · Odds ratio

Introduction

Matching ensures that the distributions of certain variable(s) are identical (or as close to identical as possible) across exposure groups in cohort studies and outcome groups in case-control studies. While matching on confounders can improve statistical efficiency (i.e., reduce the variance and improve power) for effect estimation, such improvement is not guaranteed, especially in case-control studies [1]. Matching in case-control studies can also have other counterintuitive effects because the matching is across outcome groups rather than exposure groups, and thus does not necessarily result in balancing the matching factors across exposure groups. Unfortunately, misconceptions about the implications of such matching can still be found in expository writings on the topic. We thus review several subtle issues that arise in case-control matching.

There are many types of matching protocols, including individual matching, matching to a given distribution (e.g., frequency matching), partial matching, and marginal matching [2]. There are also many protocols for selecting controls, including cumulative sampling, density sampling, and cohort sampling, for which the sample odds ratios estimate population incidence odds ratios (OR), incidence-rate ratios, or risk ratios, respectively, with little difference among the measures if the outcome is uncommon. Thus,

✉ Mohammad Ali Mansournia
mansournia_ma@yahoo.com

¹ Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, PO Box 14155-6446, Tehran, Iran

² Division of Biostatistics, School of Public Health, University of California, Berkeley, CA, USA

³ Department of Statistics, University of California, Berkeley, CA, USA

⁴ Department of Epidemiology, Fielding School of Public Health, University of California, Los Angeles, CA, USA

⁵ Department of Statistics, College of Letters and Science, University of California, Los Angeles, CA, USA

for simplicity, we will assume a cumulative case–control study of an uncommon disease nested in a closed source population (one with no immigration or emigration) to estimate an incidence OR, with no bias present except those under discussion; we will also focus on balanced matching, in which the same number of matched controls are selected for each case. Except where indicated, however, our points apply to other situations as well [3, 4]. We will also assume all conditional associations are in the same direction across levels of variables (monotonicity).

Bias introduced by case–control matching is an intentional selection bias

Over the past two decades, a consensus has emerged in epidemiology that causal reasoning, with the help of directed acyclic graphs, has improved our understanding of confounding and its control [5–10]. When confounding is defined by characteristic structures among causal relationships in the source population, the definition has proven to be more robust to challenging examples in theory and in practice than earlier definitions based only on associations. For example, when a covariate is affected by exposure or disease, it does not fit the causal definition of a confounder yet its associations in a study might fit one of the obsolete associational definitions of a confounder. Control of such a covariate will usually introduce bias.

A similar consensus about selection bias is growing but much more slowly. In particular, selection bias arising from matching in case–control studies, which has puzzled investigators for almost a century, is still widely misunderstood and often considered a type of confounding (using an associational definition.) For example, p. 237 and 239 of the first edition of *Modern Epidemiology* [11] says “Indeed, for case control studies it would be more accurate to state that matching introduces confounding rather than that it prevents confounding.... In case–control studies, matching on factors associated with exposure builds confounding into the data, whether or not there was confounding in the source population”[5]. Thirty years later, Pearce [12] wrote that matching “can introduce confounding by the matching factors even when it did not exist in the source population... if there is an association between the matching factor and the exposure, then matching will introduce confounding”.

These descriptions are inconsistent with systems that treat confounding as a consequence of causal relationships among confounders and the exposure and disease under study [5–10]. Associations are not sufficient to define confounding because they cannot distinguish between confounding and non-confounding relationships of the putative confounding covariate with exposure and disease

(e.g., a covariate affected by exposure or disease is not a confounder, and its control will usually introduce bias).

Suppose our goal is to estimate a causal OR, such as the ratio of disease odds if everyone in the population were exposed as compared to the odds if everyone were unexposed. This targeted effect is said to be *standardized* to the source population, and is often called the *marginal* causal effect for that population [13]. We contrast this measure with the unadjusted (“crude”) OR from the 2×2 table for exposure and disease in the source population, which measures only the association of exposure with disease. With an uncommon outcome, confounding is indicated when this unadjusted OR does not equal the causal OR (i.e., if the unadjusted association does not equal what the population effect of changing exposure in the source population would be).

Table 1 gives an example of such confounding (taken from Table 1 of Pearce [12]), a table of expected population counts for which the unadjusted OR is 0.86. If there were no additional confounding beyond age and no residual confounding within age strata, the disease risk if everyone in the population had been exposed would be

$$\begin{aligned} & (100090(80/80080) + 100300(100/20100)) / \\ & (100090 + 100300) \\ & = 598.99/200390 = 0.00299 \end{aligned}$$

where the two terms refer to the young ($C = 1$) and old ($C = 0$) strata, respectively. The analogous risk if everyone is unexposed is $300.14/200390 = 0.00150$. The resulting causal OR is $(0.00299/(1 - 0.00299))/(0.00150/(1 - 0.00150)) = 1.996 \approx 2.00$. The fact that 0.86 differs from 2.00, together with the assumption that age influences exposure and age also influences the disease (and not the other way around for either), indicates the presence of confounding by age. Since we are working with expected rather than observed frequencies, this conclusion does not depend on any statistical procedure but is instead a consequence of the assumed causal structure alone.

The bias introduced by case–control matching does not follow the causal definition of confounding because such confounding exists in the source population independent of

Table 1 Example of confounding in the source population (row 1 of Pearce Table 1) [12]

D	C = 1		C = 0		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	80	10	100	200	180	210
0	80,000	20,000	20,000	80,000	100,000	100,000
	OR = 2		OR = 2		OR = 0.86	

any case-control aspect of the design or analysis strategy. Instead, matching controls to cases on variables associated with exposure alters the sample association of exposure with disease, thereby resulting in a selection bias in the sample OR [1]. Although this bias has been termed “selection confounding”, [14, 15] bias introduced by case-control matching is a type of control-selection bias, where selection bias is a spurious (non-causal) component of association created by causes of selection rather than causes of disease [3, 16].

If there is confounding in the source population, as in Table 1, case-control matching superimposes this selection bias over the initial confounding in estimating the causal OR. Unlike cohort matching, case-control matching does not and cannot remove confounding, but instead may contribute a selection bias that can itself be removed by adjustment for the matching variables [1, 5, 17]. In particular, *balanced* matching (using a constant case/control ratio across matched sets) may improve test power and estimator precision in an analysis that adjusts appropriately for the matching variables, although this benefit is not guaranteed and is quite context dependent [18–21]; it may improve power and precision for inference on modification across matching factors as well [22, 23]. This matching necessarily makes the matching variables independent of disease, apparently removing confounding. But, assuming there are no other confounders, it is necessary that a factor associated with exposure be independent of disease *conditional on the exposure* to ensure that it does not produce confounding [1, 5, 6].

Note that, using the data from Table 2, the adjusted OR remains at 2.00 whereas the unadjusted OR is now 1.68. This difference is sometimes interpreted as the confounding by age left by the matching (residual confounding), but is instead a mixture of that confounding and the selection bias introduced by matching. Perhaps counterintuitively, the proportion of the original confounding remaining depends on the association of *exposure* with disease given age (i.e., C): The stronger that association, the more the exposure-conditional age-disease association differs from the marginal age-disease association after matching (which

was forced to be null by the age matching), although there will be no residual confounding if there is no association of exposure with disease given age. In a parallel fashion, the amount of selection bias produced by the matching depends on the association of age with exposure given disease status: The stronger that association, the more the age-conditional exposure-disease association differs from the marginal age-disease association after matching; on the other hand, there will be no selection bias if there is no association of age with exposure given disease status.

The causal diagrams in Figs. 1 and 2 illustrate the mix of confounding and selection bias in a matched case-control study of the effect of E on D with C as confounder and matching variable [5–8, 17]. The variable S in Figs. 1 and 2 indicates whether an individual from the original cohort is selected into the matched case-control study (1: yes, 0: no), and the square around $S = 1$ indicates that the analysis is conditional on having been selected ($S = 1$). There are arrows from D and C to S because, by definition of a matched case-control study, both D and C affect S . Figure 1 shows that case-control matching does not break the confounding path $E \leftarrow C \rightarrow D$, but instead introduces the selection-bias path $E \leftarrow C \rightarrow S \leftarrow D$. With no effect of E on D , however (Fig. 2), the net bias is zero: The paths $C \rightarrow S \leftarrow D$ and $C \rightarrow D$ now “unfaithfully” cancel each other exactly, leaving C and D independent both marginally and conditional on E after matching [17].

As displayed in Table 2, the impact of adjustment is not as dramatic after case-control matching, and now the unadjusted OR is in the same direction as the adjusted OR. This example illustrates a general phenomenon under monotonicity that, after matching, the change in the unadjusted OR relative to the adjusted OR is towards the null (which follows from algebra paralleling that given by

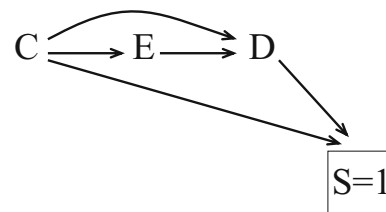


Fig. 1 Case-control matching on a confounder

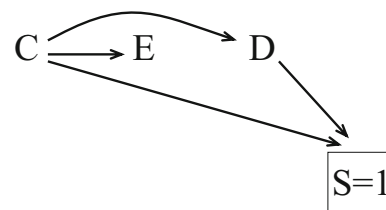


Fig. 2 Case-control matching on a confounder under the causal null hypothesis of no exposure effect

Table 2 Matched case-control study from Table 1 (row 3 of Pearce Table 1) [12]

D	$C = 1$		$C = 0$		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	80	10	100	200	180	210
0	72	18	60	240	132	258
	OR = 2		OR = 2		OR = 1.68	

MH OR 2 (95% CI 1.42, 2.81)

Samuels [20, p. 580]). It also shows why balanced case–control matching does not introduce selection bias if exposure does not affect the disease and why, in this special null case, it counterbalances confounding by the matching factors in the source population: Under the null, a bias toward the null can produce no bias. In purely associational terms, the difference between the adjusted and unadjusted OR from the matched study may also be viewed as an example of *non-collapsibility* of the sample OR, in this case where the unadjusted OR is always closer to the null [20]. We discuss this point further below.

When the matching variables are risk factors, their distribution among the controls will differ from their distribution in the source population. Nonetheless, if the distribution of exposure and matching factors in the study cases equals that distribution among all cases in the source population (i.e., if the cases are “representative”), the standardized ORs in the source population can also be estimated by comparing the numbers of cases expected with and without exposure (either or both of which may be partially or entirely counterfactual) [24, 25 p. 269]. Table 3 provides an example of OR heterogeneity in hypothetical matched case–control data. Here, the ORs standardized to the exposed is $160/(72(10)/18 + 60(240)/260) = 160/95.38 = 1.68$, to the unexposed is $(18(80)/72 + 260(80)/60)/250 = 366.67/250 = 1.47$, and to the total is $(160 + 366.66667)/(95.3846 + 250) = 1.52$. The Mantel–Haenszel OR is 1.53, close to the OR standardized to the total, although slightly biased because its weights are derived by assuming the underlying true stratum-specific ORs are all 1.

Adjustment for matching variables should account for both the actual matching protocol and further confounding effects

With some care in modeling, it is often possible to break the matched sets and instead use the matching variables as regressors in an ordinary logistic regression model. This use

can lead to bias, however, if the variables are not coded to adequately reflect the matching protocol [3, 24]. Furthermore, even if matching is retained, the matching variables may still not be adequately controlled due to coarseness of the matching. Thus, contrary to some assertions [26, p. 182] it is important that the analysis employ sufficient stratification to ensure adequate confounding control.

With matching on a continuous variable, this goal is not achieved by simple regression adjustments because the matching distorts trend relations into a discontinuous form [2, 27]. As noted above, case–control matching requires analytic adjustment for both the selection bias produced by matching and for the confounding effects of the matching variable. It is usually overlooked that distinct adjustments for these two bias sources are needed. In particular, age is often inadequately adjusted in age-matched case–control analyses, in one of two ways:

1. Matched analysis is done but the age matching is too loose to completely adjust for age, leaving unnecessary age confounding in the matched analysis. For example, there can be considerable trends in sarcoma risks within 5-year childhood age categories, and in carcinoma risks within 5-year elderly age categories. These trends can result in non-negligible residual age confounding in matched analyses, despite the age matching.
2. The matches are broken and adjustment for matching is instead done using a regression model [28]. Entering age as a continuous variable then results in an incorrect adjustment for age, because case–control age matching creates a discontinuous “saw-tooth” age trend in disease which is not controlled by continuous age [2, 23].

One model-based solution to these problems is to enter a term for residual age into the regression analysis [27, 29]. An example is the difference between each person’s age in years and the center of their age-matching category (e.g., this residual age variable would equal $-2, -1, 0, 1, 2$ for children age 6, 7, 8, 9, 10 in the 6–10 year category, and also for children age 11, 12, 13, 14, 15 in the 11–15 year category). This term would be entered in the conditional logistic model for the matched data, or the unconditional logistic model when breaking the matches; for the unconditional model, indicators for the matching categories would additionally be needed to control the selection bias produced by age matching [29–31].

Identically matched sets should be collapsed together

Although case–control matching usually dictates adjustment for matching variables, matched sets with *identical* matching values are best combined into a single matched

Table 3 Matched case–control study with OR heterogeneity

D	C = 1		C = 0		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	80	10	80	240	160	250
0	72	18	60	260	132	278
	OR = 2.00		OR = 1.44		OR = 1.35	

*MH OR 1.53; OR standardized to exposed: $160/(40 + 55.4) = 1.68$; to unexposed: $(20 + 346.7)/250 = 1.47$; to total: $(160 + 20 + 346.7)/(40 + 55.4 + 250) = 1.52$

set for the analysis. For example, if pair matching is not based on variables that tend to be unique to very few individuals (e.g., sibling status, residence location at block level, etc.), but rather over variables shared by multiple study subjects (e.g., sex, age categories), the matched design does not require analysis at the pair level so long as the matching strata are retained [12, 30]. The latter kind of matched data is often referred to as *stratum matched* or *frequency matched* [1, 31], although frequency matching often refers instead to a selection protocol based on the frequency distribution among cases.

With stratum matching, retaining pair matched data is superfluous since cases and controls cannot be distinguished from other pairs in the same stratum based on the matching factors alone. Therefore, analysis at the pair-level involves unnecessary stratification, which increases variability without reducing bias [32, 33]. Combining pairs with identical values for the matching factors into a single stratum thus improves accuracy over an analysis keeping the pairs separate. Another advantage of combining is that it can eliminate “double loss” of subjects: that is, when one member of a pair has missing data, its corresponding match will also be ignored by the analysis at the pair level [30].

Case-control matching on a non-confounder associated with disease may lead to selection bias

It is often stated that case-control matching on a variable not associated with exposure does not introduce selection bias [1, 12, 26, p. 180–181]. This is practical advice to the extent that case-control studies are ordinarily recommended for situations involving uncommon diseases. When however the disease is common these statements can become technically incorrect due to the effect of disease on control selection. In particular, for the usual case-control designs, bias can arise from ignoring a matched disease predictor if the disease is common, even if the predictor is unassociated with exposure in the source population and thus not a confounder [17].

In cumulative case-control studies (as in Table 1), controls are sampled from those who do not develop disease by the end of follow-up. Similarly, in incidence density case-control studies, controls are sampled based on person-time at risk. In order for the matching-factor adjusted and unadjusted OR to be equal in a cumulative case-control study, independence between matching factors and exposure should be among those available as controls; in an incidence density study, the independence is among the total person-time at risk [17, 21]. In either case, however, this independence will be broken if the disease predictor is initially independent of exposure, before events affecting availability as a control occur (such as the study

disease, a competing risk, or loss from the source population). In that case, matching on the predictor introduces a selection bias if exposure indeed affects disease or otherwise affects the chance of becoming a control, a bias that is removed by adjustment for the predictor.

As an example, if both the matching variable and exposure affect disease risk, they will both reduce the number of non-cases available as controls, and so will (apart from artificial exceptions) become associated among the controls. This association arises from the joint effect of the variables on events that remove one from availability as a control. Graphically, this phenomenon is an example of collider bias [34]: a bias arising from a collision of causal arrows [17, 35]. Fortunately, the bias is negligible when exposure effects are small or the disease is uncommon over the study period (which typifies the usual setting in which a case-control design is recommended). In that setting, these control-source populations will differ only negligibly from the starting population, and in particular initial independence will be little altered by subsequent events. The bias also does not occur in case-cohort studies, for in the latter the controls are sampled from the total cohort and thus independent of disease, which leaves matching factors and exposure independent in the controls (apart from sampling error).

Matching may lead to overadjustment, thus harming precision or creating uncorrectable bias

Table 1 shows extreme confounding, to the point that without adjustment exposure appears to be protective, but appears harmful within age strata. Yet the precision improvement from case-control matching is small: The unmatched confidence interval for the age-specific OR (assumed constant) shrinks only slightly from (1.38, 2.89) in the unmatched scenario to (1.42, 2.81) in the matched scenario (Tables 2, 4). When an exposure is strongly associated with both the confounder and disease, but the association between confounder and disease is weak, matching can even lead to a loss in efficiency [21].

Table 4 Unmatched case-control study from Table 1 (row 2 of Pearce Table 1) [12]

D	C = 1		C = 0		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	80	10	100	200	180	210
0	156	39	39	156	195	195
	OR = 2		OR = 2		OR = 0.86	

* MH OR 2 (95% CI 1.38, 2.89)

Tables 5, 6, 7 gives an example in which the stratum-specific association of E and D is stronger than in Table 1, but the E-specific association between C and D is weaker, and case-control matching slightly harms precision. Matching can also harm precision by creating unnecessary concordance (correlation) between case and control exposures [1], e.g., while matching on sibling can control genetic factors, dietary variables may also vary too little among siblings to allow precise estimation of diet and nutrient effects. Such examples suggest avoiding case-control matching on covariates only weakly related to the disease, even if they are strongly related to the exposure (exceptions occur although they appear difficult to identify in advance [36]).

Matching can also produce irremediable selection bias, especially if matching on a variable affected by exposure.

Table 5 An example of confounding in the source population based on a modification of Table 1 with the age stratum-specific E-D association strengthened, and the E-specific C-D association weakened

D	$C = 1$		$C = 0$		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	160	10	80	80	240	90
0	80,000	20,000	20,000	80,000	100,000	100,000
	OR = 4		OR = 4		OR = 2.67	

Table 6 Unmatched case-control study using Table 5

D	$C = 1$		$C = 0$		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	160	10	80	80	240	90
0	132	33	33	132	165	165
	OR = 4		OR = 4		OR = 2.67	

* MH OR 4 (95% CI 2.65, 6.03)

Table 7 Matched case-control study using Table 5

D	$C = 1$		$C = 0$		Collapsed	
	E		E		E	
	1	0	1	0	1	0
1	160	10	80	80	240	90
0	136	34	32	128	168	162
	OR = 4		OR = 4		OR = 2.67	

* MH OR 4 (95% CI 2.65, 6.04)

Consider next surviving birth at the same hospital as a matched control in a study of prenatal death. The choice of hospital may be dictated by certain pregnancy conditions including the study exposure, and hospital itself will ordinarily affect death risk, making hospital a mediator. Because matching results in adjustment for hospital, the study OR will be biased for the total effect of any exposure that affects hospital choice (e.g., neonatal care). If one had total births by date and place of delivery, one could correct for this bias by reweighting the observations with weights inversely proportional to the date- and hospital-specific sampling fractions in cases and controls to “undo” the matching [37, 38], but such corrective data are often unreliable or unavailable.

Subtleties in assessing modification of the OR in matched case-control studies

It is well known that the effect of a matching variable usually cannot be examined without further data to “undo” the matching effect, but that modification of the OR for exposure effect by a matching factor can be examined by stratifying on the factor [31, 39]. As with unmatched factors, there is almost never an empirical basis for assuming homogeneity (no modification) across matching factors, and so checking for important violations is good practice. Nonetheless, tests for modification generally have little power to detect the direction let alone degree of modification [40], and the multiple tests involved in looking at many factors creates a high risk of reporting exaggerated or spurious “false positive” findings.

A sophisticated answer to this problem is to conduct an analysis that accounts for the multiplicity, using for example “shrinkage” methods (such as hierarchical, empirical, or partial-Bayes methods, or penalized-likelihood) applied to the product terms that represent log-OR modification in a conditional-logistic model [29]. A simpler heuristic approach (which is the implicit default in many studies) is to not explore modification, instead treating the (fictional) constant OR estimated using a main-effects only (“no-interaction”) model as an estimate of the marginal (standardized) OR in the underlying source population. This can be a fair approximation in an unmatched study of an uncommon disease, but can break down if matching is done on a strong modifier of the exposure-effect OR, or if the exposure or matching factor have strong effects on the probability of being selected as a control: In either case, matching alters the weighting across matching strata implicit in the regression analysis, leading to discrepancies between the OR estimated under homogeneity and the population OR standardized to the total (which is the marginal causal OR if there is no further

confounding) [41, 42]. This discrepancy is often small however, as in Table 3.

Another question is how to estimate modification of the OR across a continuous variable. Using unordered categories may severely harm power of an already low-power test, which can be mitigated by using continuous versions of the variable in product terms. The crucial point for matched studies is that, unlike for main effects, the variable as entered into product terms need *not* follow the form used for matching. For example, age is often (and wisely) matched in categories so small as to preclude analysis of the categories separately. Nonetheless, exposure can be multiplied by a continuous version of age and entered into the regression model. What then should be the scale of age in the product (“interaction”) term in the model? We recommend a simple form close to a well-informed prior expectation regarding effect modification. In particular, given the exponential dependence of the OR on a continuous variable in a logistic model, we advise transforming the variable to dampen extreme projections, e.g., by using $\log(\text{age})$ rather than untransformed age [43].

Further differences between the effects of cohort and case-control matching

Matching has a more favorable cost-benefit analysis in cohort settings: Cohort matching balances confounder distributions across exposure groups, and hence can prevent confounding by the matched variables (although this balance can be disrupted by further adjustments [44]). And although (as with case-control matching) cohort matching can sometimes lead to efficiency loss, this problem appears less severe than in case-control studies [45]. Nonetheless, non-random losses from the cohort or adjustments for unmatched variables may lead to bias from failure to control matching variables [1, 17, 44].

When cohort-matching factors are strong effect-measure modifiers (strong heterogeneity is present) and, as usual, the unexposed are matched to the exposed, the matched marginal effect measure can differ dramatically from the marginal effect in the original unmatched source population from which the cohort was drawn [46]. This concern does not apply to matched randomized trials with no loss, because the marginal distribution of matching variables is not altered by the randomization process [45]. For case-control matching however the problem is worse in this sense: Matching alters the distribution of the factors to follow the distribution in the cases rather than in the source population, thus making the usual matched OR estimators (which assume a constant OR) biased for any marginal causal OR in the source population if the factors are

important modifiers of the OR. Again, if the modification can be captured using stratified methods or product terms in a logistic model, the data can still be used to obtain population standardized OR estimates as in Table 3, although nonstandard estimators and variance formulas are needed [24, 25 p. 269].

Another divergence arises because the OR suffers from non-collapsibility: Even in the absence of confounding, factor-specific ORs can differ from the unadjusted OR [6, 47–49]. In a cohort, baseline independence between risk factors and exposure (e.g., as induced by matching on the factors) *does* imply no confounding by the factors if no further adjustments are made [17, 44]. The degree of OR non-collapsibility that remains depends directly on disease frequency and the strength of covariate and exposure effects (with only slight non-collapsibility if the disease is uncommon or either the covariate or exposure effects are weak) [50]. Unfortunately, these observations do not carry over to case-control studies with matching on the same factors, since (again) the matching does not change the source population and thus does not prevent confounding. This means that the difference between the matching-adjusted and unadjusted OR reflects not only non-collapsibility but also includes the matching-selection bias in the unadjusted OR, as discussed above.

Discussion

Although matched case-control studies date at least back to 1920s [51, 52] and the purpose of matching in case-control and cohort studies was clarified long ago [19, 21, 45], there is still much confusion among researchers and data analysts. Some confusion may have arisen from false analogies with matched experiments, although these misconceptions had been recognized by the 1980s [18–21].

We often encounter misconceptions to the effect that any covariate that temporally precedes and predicts exposure *or* disease (and so *might* be a confounder) should be controlled. These notions appear to be an incorrect generalization from theoretical results that control of all such variables will prevent confounding by those variables [53]. Another incorrect notion is that variables whose distribution differs between cases and controls should be matched. Such naïve recommendations ignore the variance inflation and biases due to modeling error or sparse data that unnecessary covariate adjustment can produce. We have reviewed how they can also lead to matching that harms study precision or validity. A further, practical problem is that matching may decrease cost efficiency if finding matches requires effort (as in studies collecting original data). The effort in finding closely matched subjects can instead go toward selecting a greater number of less

matched or unmatched subjects, thus increasing precision beyond that of a matched design [1].

Matching on strong confounders should however remain a core design option. Modeling variables with strong effects can disrupt typical fitting methods by creating very sparse data in certain categories, and can also increase sensitivity to model misspecification; control of such variables can be improved by at least partial matching, rather than by modeling alone [54].

These points are seen most clearly for age: The majority of cancers have huge (power-law) relations to age, as do common dementias and vascular disorders. To fail to match for age in a case–control study of such an outcome would guarantee little overlap between the cases and controls in the relevant age ranges. The result would be that control of age confounding would have to rely mostly on correct specification of the age–incidence relation to extrapolate age adjustments between the mostly younger controls and the mostly older cases. With age matching and its proper control, that extrapolation dependence would be reduced to the residual relation within age-matching categories. Similar comments apply when other very strong measured confounders are present.

Matching can also enable adjustment for confounding variables that are difficult to measure. For example, use of siblings as controls partially adjusts for genetic and childhood environmental factors; use of neighborhood controls partially adjusts for social class, ambient air and water, and other local geographic factors [31, sec. 16.7]. Matching on such factors also provides a convenient sampling frame for controls (e.g., the control for each prenatal death was the next surviving birth at the same hospital). A minor complication arising from such matching is that it will usually result in very small numbers of cases and controls in each matched set, and thus sparse-data methods will be required to analyze the data without breaking the matching. A major drawback however is that the matched sets so produced may be too concordant in exposure to provide accurate estimates.

Although common sparse-data methods (such as conditional logistic regression and Mantel–Haenszel techniques) were initially developed as a remedy for sparse-data bias in conventional unconditional logistic regression analysis, they too can suffer from considerable sparse-data bias when certain types of discordant matched sets are infrequent or when the model contains too many parameters [54, 55]. As an example, matched-pair estimates develop bias away from the null value if the number of discordant pairs is low [56]. Penalization and related “shrinkage” methods can be applied to matched samples both to reduce sparse-data bias while achieving finer confounding adjustments [54, 55, 57–59].

Propensity-score methods are sometimes promoted to address the concerns we have discussed. Even in cohort studies, however, propensity-score matching may lead to overadjustment and variance inflation, or poor control of strong confounders [60–62], and can also generate spurious results in case–control studies [63]. Thus propensity scoring does not address the need to consider causal structure, associational strength, and potential artefacts when formulating a matching protocol.

In conclusion, we concur with advice that matching should be used with great caution, especially in case–control studies [1]. Variables expected to be strong confounders (like age and sex) are good candidates for direct matching, whereas weak confounders may be better addressed via subsequent model-based adjustments, while matching or adjustment for variables unrelated to disease is best avoided. Nonetheless, practical considerations may dictate use of conveniently matched controls such as relatives, neighbors, or friends, despite risks of efficiency overmatching and overlap bias [64]. In particular, we think it highly misguided if not destructive to ignore the practical difficulties of locating and recruiting valid population control groups while attempting to avoid theoretical biases that are likely to be minor.

The most practical option may often be to match only on age and sex, and perhaps one more important nominal-scale confounder, especially those with a large number of possible values (e.g., neighborhood, occupation) for which model-based adjustment is difficult [1]. Regardless, one should account for matching variables in the analysis, paying special attention to the matching protocol and the distortions produced by case–control matching, as well as sparse-data bias [54–59].

Acknowledgements The authors are grateful to David Clayton and the referees for helpful comments on earlier drafts of this paper.

References

1. Rothman KJ, Greenland S, Lash TL. Design strategies to improve study accuracy. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2008. p. 168–82.
2. Greenland S. Partial and marginal matching in case-control studies. In: Moolgavkar SH, Prentice RL, editors. *Modern statistical methods in chronic disease epidemiology*. New York: Wiley; 1986. p. 35–49.
3. Rothman KJ, Greenland S, Lash TL. Case-control studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2008. p. 111–27.
4. Jewell NP. *Statistics for epidemiology*, chapter 5. Boca Raton: Chapman & Hall/CRC; 2004.
5. Glymour MM, Greenland S. Causal diagrams. In: Rothman KJ, Greenland S, Lash T, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 183–209.

6. Jewell NP. Statistics for epidemiology, chapter 8. Boca Raton: Chapman & Hall/CRC; 2004.
7. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *Eur J Epidemiol*. 2015;30:1101–10.
8. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in randomized trials: a conversation between trialists and epidemiologists. *Epidemiology*. 2017;28:54–9.
9. Suzuki E, Tsuda T, Mitsuhashi T, Mansournia MA, Yamamoto E. Errors in causal inference: an organizational schema for systematic error and random error. *Ann Epidemiol*. 2016;26:788–93.
10. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ* 2017;359:j4587.
11. Rothman KJ. Modern epidemiology, chapter 13. Boston: Little, Brown; 1986.
12. Pearce N. Analysis of matched case-control studies. *BMJ*. 2016;352:i969.
13. Gharibzadeh S, Mohammad K, Rahimiforoushani A, Amouzegar A, Mansournia MA. Standardization as a tool for causal inference in medical research. *Arch Iran Med*. 2016;19:666–70.
14. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2008. p. 345–80.
15. Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–25.
16. Gail MH. Selection bias. In: Armitage P, Colton T, editors. *Encyclopedia of biostatistics*. 2nd ed. Hoboken: John Wiley & Sons; 2005. p. 4869–70.
17. Mansournia MA, Hernán MA, Greenland S. Matched designs and causal diagrams. *Int J Epidemiol*. 2013;42:860–9.
18. Smith PG, Day NE. Matching and confounding in the design and analysis of epidemiological case-control studies. In: Blithell JF, Coppi R, editors. *Perspectives in medical statistics*. New York: Academic Press; 1981.
19. Kupper LL, Karon JM, Kleinbaum DG, Morgenstern H, Lewis DK. Matching in epidemiologic studies: validity and efficiency considerations. *Biometrics*. 1981;37:271–92.
20. Samuels ML. Matching and design efficiency in epidemiological studies. *Biometrika*. 1981;68:577–88.
21. Thomas DC, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. *J Chronic Dis*. 1983;36:685–97.
22. Smith PG, Day NE. The design of case-control studies: the influence of confounding and interaction effects. *Int J Epidemiol*. 1984;13:356–65.
23. Thomas DC, Greenland S. The efficiency of matching in case-control studies of risk-factor interactions. *J Chronic Dis*. 1985;38:569–74.
24. Greenland S. Estimating variances of standardized estimators in case-control studies and sparse data. *J Chronic Dis*. 1986;39:473–7.
25. Greenland S, Rothman KJ. Introduction to stratified analysis. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2008. p. 258–82.
26. Clayton D, Hills M. Statistical models in epidemiology, chapter 18. New York: Oxford University Press; 1993.
27. Greenland S. Re: Estimating relative risk functions in case-control studies using a nonparametric logistic regression. *Am J Epidemiol*. 1997;146:883–4.
28. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. *J Am Stat Assoc*. 1983;78:1–12.
29. Greenland S. Introduction to regression modeling. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2008. p. 418–55.
30. Greenland S. Applications of stratified analysis methods. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins; 2008. p. 283–302.
31. Jewell NP. Statistics for epidemiology, chapter 16. Boca Raton: Chapman & Hall/CRC; 2004.
32. Robinson LD, Jewell NP. Some surprising results about covariate adjustment in logistic regression. *Int Stat Rev*. 1991;59:227–40.
33. Brookmeyer R, Liang KY, Linet M. Matched case-control designs and overmatched analyses. *Am J Epidemiol*. 1986;124:693–701.
34. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–6.
35. Didelez V, Kreiner S, Keiding N. On the use of graphical models for inference under outcome dependent sampling. *Stat Sci*. 2010;25:368–87.
36. Kalish LA. Matching on a non-risk factor in the design of case-control studies does not always result in an efficiency loss. *Am J Epidemiol*. 1986;123:551–4.
37. Mansournia MA, Altman DG. Inverse probability weighting. *BMJ*. 2016;15(352):i189.
38. Mansournia MA, Danaei G, Forouzanfar MH, Mahmoudi M, Jamali M, Mansournia N, Mohammad K. Effect of physical activity on functional performance and knee pain in patients with osteoarthritis: analysis with marginal structural models. *Epidemiology*. 2012;23:631–40.
39. Szklo M, Nieto F. *Epidemiology: beyond the basics*, chapter 6. 3rd ed. Sudbury: Jones and Bartlett Publishers; 2014.
40. Greenland S. Tests for interaction in epidemiologic studies: a review and a study of power. *Stat Med*. 1983;2:243–51.
41. Greenland S, Maldonado G. The interpretation of multiplicative model parameters as standardized parameters. *Stat Med*. 1994;13:989–99.
42. Mohammad K, Hashemi Nazari SS, Mansournia N, Mansournia MA. Marginal versus conditional causal effects. *J Biostat Epidemiol*. 2015;1:121–8.
43. Greenland S. Dose-response and trend analysis: alternatives to category-indicator regression. *Epidemiology*. 1995;6:356–65.
44. Sjölander A, Greenland S. Ignoring the matching variables in cohort studies: when is it valid and why? *Stat Med*. 2013;32:4696–708.
45. Greenland S, Morgenstern H. Matching and efficiency in cohort studies. *Am J Epidemiol*. 1990;131:151–9.
46. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*. 2006;163:262–70.
47. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci*. 1999;14:29–46.
48. Mansournia MA, Greenland S. The relation of collapsibility and confounding to faithfulness and stability. *Epidemiology*. 2015;26:466–72.
49. Greenland S, Pearl J. Adjustments and their consequences: collapsibility analysis using graphical models. *Int Stat Rev*. 2011;79:401–26.
50. Pang M, Kaufman JS, Platt RW. Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models. *Stat Methods Med Res*. 2016;25:1925–37.
51. Lombard HL, Doering CR. Cancer studies in Massachusetts. 2. Habits, characteristics and environment of individuals with and without cancer. *N Engl J Med*. 1928;198:481–7.

52. Lane-Claypon JE. A further report on cancer of the breast. London: Her Majesty's Stationery Office; 1926.
53. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. *Biometrics*. 2011;67:1406–13.
54. Greenland S, Schwartzbaum JA, Finkle WD. Problems from small samples and sparse data in conditional logistic regression analysis. *Am J Epidemiol*. 2000;151:531–9.
55. Greenland S. Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics*. 2000;1:113–22.
56. Jewell NP. Small-sample bias of point estimators of the odds ratio from matched sets. *Biometrics*. 1984;40:421–35.
57. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ*. 2016;27(352):i1981.
58. Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. *Stat Med*. 2015;34:3133–43.
59. Mansournia MA, Geroldinger A, Greenland S, Heinze G. Separation in logistic regression—causes, consequences, and control. *Am J Epidemiol*. 2017. doi:10.1093/aje/kwx299.
60. Shrier I. Re: the design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2008;27:2740–1.
61. Pearl J. Remarks on the method of propensity score. *Stat Med*. 2009;28:1415–6.
62. King G, Nielsen R. Why propensity scores should not be used for matching. Vers. 2 Feb. 2016 downloaded from <http://j.mp/1FQhySn>.
63. Mansson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case-control and case-cohort studies. *Am J Epidemiol*. 2007;166:332–9.
64. Austin H, Flanders WD, Rothman KJ. Bias arising in case-control studies from selection of controls from overlapping groups. *Int J Epidemiol*. 1989;18:713–6.