STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:3453–3465
Published online 12 March 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.3246

On effect-measure modification: Relationships among changes in the relative risk, odds ratio, and risk difference

Babette Brumback[1,*,†] and Arthur Berg[2]

[1]*Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL 32610, U.S.A.*
[2]*Department of Statistics, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32610, U.S.A.*

SUMMARY

It is well known that the presence or absence of effect-measure modification depends upon the chosen measure. What is perhaps more disconcerting is that a positive change in one measure may be accompanied by a negative change in another. Therefore, research demonstrating that an effect is 'stronger' in one population when compared with another, but based on only one measure, for example, the odds ratio, may be difficult to interpret for researchers interested in another measure. The present article investigates relationships among changes in the relative risk, odds ratio, and risk difference from one stratum to another. Monte Carlo integration shows that the three measures change in the same direction for 78 or 89 per cent of the volume of the geometric space defined by the four underlying proportions, depending on whether the strata are presumed to share the same direction of effect or not. Analytic results are presented concerning necessary and sufficient conditions for the measures to change in opposite directions. In general, the conditions are seen to be quite complicated, though they do give way to some interesting results. For example, when exposure increases risk but all risks are less than 0.5, it is impossible for the relative risk and risk difference to change in the same direction but opposite to that of the odds ratio. Both data-analytic and hypothetical examples are presented to demonstrate circumstances under which the measures change in opposite directions. Copyright © 2008 John Wiley & Sons, Ltd.

KEY WORDS: case–control studies; clinical trials; cohort studies; common outcomes; cross-sectional studies; heterogeneity

INTRODUCTION

It is well known that the presence of effect-measure modification, also known as heterogeneity of effect, depends upon the chosen measure [1, 2]. Rothman and Greenland [1] present a hypothetical example in which males possess risks of 0.5 and 0.2, whereas the risks for females are 0.1 and 0.04;

*Correspondence to: Babette Brumback, Department of Epidemiology and Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL 32610, U.S.A.
†E-mail: bbrumback@phhp.ufl.edu

Received 5 September 2007
Accepted 22 January 2008

the risk difference is much greater for males, but the risk ratio remains the same. The definition of interaction in statistics is logically equivalent to effect-measure modification [1]. In epidemiology, interaction often refers to biological interaction or synergism, a concept distinct from statistical interaction [1, 3–6]. However, in some contexts, sufficient conditions for the presence of synergism are closely related to effect modification of the risk difference [7, 8].

Rothman and Greenland [1] prove the following three facts about effect-measure modification, assuming that both stratum and exposure have non-zero effects: (a) when there is additivity on the risk scale, there has to be modification of the risk ratio; (b) when there is additivity on the log-risk scale, there must be statistical interaction on the risk scale (modification of the risk difference); and (c) when there is no multiplicative interaction on the odds scale, there must be modification of the risk ratio and risk difference. Morabia *et al.* [9], using simulations, present somewhat general conditions for heterogeneity of the odds ratio without heterogeneity of the relative risks. Furthermore, Rothman and Greenland [1] note that the magnitude of statistical interaction is ‘entirely determined by the scale chosen for measuring additivity of effects.’ However, our literature search has identified almost no discussion pertaining to the relationship between the *direction* of statistical interaction and the scale chosen for measurement. The present article investigates conditions under which the odds ratio, relative risk, and risk difference can change in opposite directions from one pair of risks to the next. This phenomenon is illustrated in a textbook exercise by Agresti [10], who presents data from a cohort study of non-smokers and smokers [11], in which the proportions per year of those who died from lung cancer were 0.0001 and 0.0014 and from coronary heart disease were 0.00413 and 0.00669; the relative risk and odds ratio decrease but the risk difference increases. Although this is not an example of effect-measure modification *per se* because the outcome under consideration changes rather than the subpopulation, there are analogous implications for effect-measure modification. That is, instead of, ‘for which *disease* is the effect of smoking ‘stronger’?’ the question becomes, ‘for which *subpopulation* is the effect of smoking ‘stronger’?’ when set in terms of the same four risks but corresponding to two subpopulations and one disease rather than to two diseases and one population.

Much research having to do with epidemiological effects and thus with effect-measure modification is presented in terms of just one of the three measures. The odds ratio is ‘probably the most widely used measure of 2×2 association in epidemiology’ [12], due largely to the role of logistic regression as the ‘workhorse of contemporary epidemiology’ [13], despite many authors [14–20] having emphasized methods for estimating risk ratios and differences from logistic and other models in studies of common outcomes and in case–control studies. It has been argued that odds ratios are useful estimates of exposure effects only when they serve as estimates of a cumulative-incidence ratio [1, 21] or of an incidence-density ratio (under incidence-density sampling), which does not require rarity of disease [1, 22]. Others [23, 24] have protested the use of prevalence odds ratios in place of prevalence rate ratios for summarizing cross-sectional studies, although not all would agree [25]. The relative risk will be of more interest to investigators interested in the excess or attributable fraction [1], although much caution is required in interpreting this measure causally [26–28]. Similarly, interpretation of the reciprocal of the risk difference as the number needed to treat to benefit (or harm) one patient [29, 30] requires strong assumptions that are seldom plausible [31, 32].

When a research team selects just one of the three measures for dissemination of results, whether based on scientific rationale or convenience, other investigators may wish for an interpretation in terms of an alternative measure. For example, suppose a study in one population produced an odds ratio of 2 and another study in a different population produced an odds ratio of 3. When can we conclude that the effect in the second population is stronger when measured by the relative risk or risk difference? Questions like this one motivated us to delineate conditions under which the three measures can and cannot change in opposite directions, the primary objective of this article. A secondary objective is to emphasize the importance of presenting results more thoroughly than is typical, so that the reader can, at least, construct estimates of his or her measure of choice. For example, when presenting results of a logistic regression, it is relatively common for authors to exclude the intercept and exponentiate the remaining coefficients. But without the intercept, readers cannot possibly use the results to construct estimated risks, and hence relative risks (unless the outcome is rare) or risk differences.

THREE MEASURES OF EFFECT MODIFICATION

Let p_1 and p_2 denote the unexposed and exposed risks in stratum 1, and let p_3 and p_4 denote the same in stratum 2. If $p_1=p_2$ or $p_3=p_4$, changes from stratum 1 to stratum 2 in the relative risk, odds ratio, and risk difference are all in the same direction. Furthermore, when $p_2>p_1$ but $p_4<p_3$ (or when $p_2<p_1$ but $p_4>p_3$), all three measures decrease (or increase) together between strata, because within a given stratum, the relative risk and odds ratio will always be on the same side of one and the risk difference will be on the corresponding side of zero. The contexts in which $p_1<p_2$ and $p_3<p_4$ are analogous to those in which $p_2>p_1$ and $p_4>p_3$ (one need only to recode the exposure or outcome). Therefore, except where specifically mentioned, the investigation to follow will presume that exposure increases risk within both strata, so that an increase in a given measure corresponds to a 'stronger' effect. In that context, three useful measures of effect modification are the ratio of relative risks, $k=p_4p_{1/}/(p_3p_2)$, the ratio of odds ratios, $q=k(1-p_3)(1-p_2)/((1-p_4)(1-p_1))$, and the ratio of risk differences, $m=(p_4-p_3)/(p_2-p_1)$; all ratios represent the effect in stratum 2 divided by the effect in stratum 1. In more general contexts, m is ill-defined, and $s=(p_4-p_3)-(p_2-p_1)$ will take its place. To simplify exposition, the presentation to follow will presume that all four risks are strictly between zero and one.

Definition of disagreement

When exposure does not necessarily increase risk within both strata, disagreement of k, q, and s will mean that two of $k-1$, $q-1$, and s are on opposite sides of zero. When exposure does increase risk within both strata, disagreement will mean that two of k, q, and m are all on opposite sides of one (and in this context k or q will disagree with m exactly when either disagrees with s.) Agreement will refer to the opposite of disagreement; note that this includes situations in which one measure remains constant while the other changes. Although the relative risk, odds ratio, and risk difference always agree with one another qualitatively, the examples given previously illustrate that k, q, and s or m, which contrast the three effect measures across strata, need not agree. However, in all contexts, when $p_1=p_3$ or $p_2=p_4$, k, q, and s will agree.

Monte Carlo study

It is an interesting yet impossible exercise to visualize the four-dimensional hypercube, representing the space of possible settings for the four risks, in order to gain intuition about the subspace corresponding to the agreement of k, q, and s. One can, however, use Monte Carlo integration to better understand the *size* of that subspace and of other important components of the hypercube.

DOI: 10.1002/sim

The authors generated 500 000 independent sets of four independent uniform [0,1] proportions using the statistical programming language and environment R [33]. The subspace of agreement occupies 89.0 per cent of the hypercube. When considering only pairs of measures, k and q, as well as k and s, agree on 91.7 per cent, whereas q and s agree on 94.7 per cent.

Twenty-five per cent of the hypercube represents the component on which exposure increases risk within both strata; the subspace of agreement overlaps with 77.9 per cent of this component. The measures k and q agree on 83.2 per cent, k and m on 83.3 per cent, and q and m on 89.4 per cent.

Much of epidemiology is focused on rare diseases for which all four risks are less than 0.10 [14]. To study this setting, 500 000 sets of uniformly distributed proportions were generated as before but restricted to [0,0.10]. This subcube represents 0.01 per cent of the original hypercube, of which 25 per cent corresponds to increased risk of exposure within both strata. Of that 0.0025 per cent, the subspace of agreement overlaps with 83.3 per cent. The measures k and q agree on 99.5 per cent, k and m (s) on 83.3 per cent, and the odds ratio and risk difference on 83.7 per cent. When disease is rare but exposure does not necessarily increase risk within both strata, the subspace of agreement overlaps with 91.7 per cent; the measures k and q agree on 99.8 per cent, k and m (s) on 91.7 per cent; and the odds ratio and risk difference on 91.9 per cent.

The above results are reassuring in that disagreement occurs for only 22 per cent, or fewer, of the possible combinations of risks, depending on the component of interest within the hypercube and on the pairing or tripling of measures of effect modification. The next focus of the investigation is on conditions necessary and/or sufficient for disagreement or agreement, assuming that exposure increases risk within both strata. Each pair of measures is first considered separately.

RELATIVE RISK AND ODDS RATIO

Greenland [21] showed that for a rare disease (all risks less than 0.10), if the odds never exceed x in a stratum, the odds ratio will equal the relative risk multiplied by a bias factor falling between $1-x$ and $1+x$. The following presentation focuses on common outcomes as well as rare diseases. Let $p_2 = rp_1$ and $p_3 = up_1$. Then $p_4 = krup_1$. Algebra leads to the relation

$$k(1+rup_1^2-(r+u)p_1)=q(1+krup_1^2-(kru+1)p_1) \tag{1}$$

Suppose $k>1$; the goal is then to find conditions under which $q<1$. From equation (1), $q<1$ exactly when

$$p_1>(k-1)/(rk+k(1-r)u-1) \quad \text{and} \quad rk+k(1-r)u-1>0 \tag{2}$$

These conditions, together with constraints on r, u, and k needed for the four risks to be between zero and one, are necessary and sufficient for the relative risk to increase but the odds ratio to decrease. It is not necessary to consider the opposite scenario in which the relative risk decreases while the odds ratio increases, because one can just relabel the two strata.

A necessary condition for disagreement

The conditions in equation (2) are quite complicated, but they imply that a necessary condition for $q<1$ is $p_1>(k-1)/rk$ (because $0<rk+k(1-r)u-1<rk$). Because $p_2 = rp_1$, an equivalent

condition is $p_2 > (k-1)/k$. Therefore, *a necessary condition for* $k > 1$ *and* $q < 1$ *is that one of the four risks must exceed* $(k-1)/k$. For rare diseases, all risks are less than 0.10, which means that if the odds ratio decreases ($q < 1$), the relative risk cannot increase by a factor greater than $k = \frac{1}{0.9} = 1.11$. For more common outcomes, when the odds ratio decreases, the relative risk cannot double unless one of the four risks exceeds 0.5.

RELATIVE RISK AND RISK DIFFERENCE

Algebra shows that $u = m(r-1)/(kr-1)$ and, thus, $p_3 = p_1 m(r-1)/(kr-1)$ and $p_4 = p_1 m(r-1) kr/(kr-1)$. Suppose $k > 1$; the goal is to find conditions under which $m < 1$. Necessary and sufficient conditions are obtained directly from the constraints on the four risks to be between zero and one. These constraints are $p_1 r < 1$ and $p_1 m(r-1)kr/(kr-1) < 1$ (recall that $r > 1$ because the exposure is presumed to increase risk within both strata). Several values of $k > 1$ and $m < 1$ can be found which satisfy these constraints; hence, one need only to choose p_1 small enough and practically any combination of $k > 1$ and $m < 1$ can occur. Furthermore, when m is small, p_1 need not be.

Some necessary conditions for disagreement

Again, the necessary and sufficient conditions are quite complicated, but simpler necessary conditions emerge. *Supposing* $k > 1$, *then both* $p_3 < mp_1$ and $p_4 < mp_2$ *are necessary regardless of* $m < 1$ *or* $m > 1$ (because $(r-1)/(kr-1) < 1$ and $(r-1)kr/(kr-1) < r$). To put this in context, suppose that the relative risk increases; then in order for the risk difference to halve, p_3 must be less than half of p_1 and p_4 must be less than half of p_2.

On the other hand, *suppose* $m < 1$, *then both* $p_3 < p_1/k$ *and* $p_4 < p_2$ *are necessary for* $k > 1$. Thus, if the risk difference decreases, then in order for the relative risk to double, p_3 must be less than half of p_1 and p_4 must be less than p_2.

ODDS RATIO AND RISK DIFFERENCE

Let $t = p_2(1-p_1)/(p_1(1-p_2))$ be the odds ratio in stratum 1. Then algebra establishes that $p_2 - p_1 = (p_1 - p_1^2)(t-1)/(1+(t-1)p_1)$ and $p_4 - p_3 = (p_3 - p_3^2)(tq-1)/(1+(tq-1)p_3)$. Suppose $q > 1$. Then necessary and sufficient conditions for $m < 1$ are obtained by substituting the preceding expressions (in terms of t, q, p_1, and p_3) into $p_2 - p_1 < p_4 - p_3$, together with the conditions on t and q that they be small enough for all four risks to be less than one.

Some necessary conditions for disagreement

Because $a, b, c, d > 0$; $c > a$; and $a/(1+b) > c/(1+d)$ imply that $ad > bc$, algebra establishes that the complicated necessary and sufficient conditions for $q > 1$ and $m < 1$ cannot hold whenever both $p_1 > p_3$ and $p_3 - p_3^2 > p_1 - p_1^2$. Therefore, *when* $q > 1$, *a necessary condition for* $m < 1$ *is that either* $p_3 > p_1$ *or* $p_1 - p_1^2 > p_3 - p_3^2$. It also follows that $0.5 < p_3 < p_1$ renders $m < 1$ and $q > 1$ impossible (but $m > 1$ and $q < 1$ remain possible), whereas similarly $0.5 < p_1 < p_3$ renders $m > 1$ and $q < 1$ impossible (but $m < 1$ and $q > 1$ remain possible). The breadth of these conditions may help to explain one

DOI: 10.1002/sim

result of the simulation study, which showed that the odds ratio and risk difference tend to change together more often than the other two pairs.

A sufficient condition for agreement

Supposing again that $q>1$, it follows by algebra (by rewriting $p_i - p_i^2$ as $p_i(1-p_i)$, and noting that both $p_1>p_3$ implies $1-p_1<1-p_3$ and $0<a<b$ implies $a/(1+a)<b/(1+b)$) from the necessary and sufficient conditions for $m>1$ (simply reverse the above necessary and sufficient conditions for $m<1$) that $p_3(tq-1)/(t-1)\geqslant p_1>p_3$ implies $m>1$. *Thus, when* $q>1$, *a sufficient condition for* $m>1$ *is that* $p_3(tq-1)/(t-1)\geqslant p_1>p_3$. To put this in context, if the odds ratio doubles ($q=2$) and the odds ratio in stratum 1 is $t=1+\varepsilon$, then $p_3\ (2\varepsilon+1)/\varepsilon\geqslant p_1>p_3$ ensures that the risk difference increases, i.e. if $\varepsilon=0.5$, then $4p_3\geqslant p_1>p_3$ ensures agreement.

ALL THREE MEASURES

Necessary and sufficient conditions for agreement of all three measures are so complicated that they are not reported here. However, the following results are intriguing. Suppose $q>1$. It was shown that $0.5<p_3<p_1$ renders $m<1$ impossible; thus, when all four risks are greater than 0.5, $m<1$ requires $p_3>p_1$. Furthermore, when $m<1$, $k>1$ was shown to imply $p_3<p_1/k$, and thus $p_3<p_1$. Hence, it is impossible that $q>1$, $m<1$, and $k>1$. Reversing strata shows that it is impossible that $q<1$, $m>1$, and $k<1$. Although the above proofs were given for the case when exposure increases risk, relabeling yields the same result when exposure decreases risk. *Thus, when all risks are greater than 0.5, it is impossible for the relative risk and odds ratio to change in the same direction but in a direction opposite to that of the risk difference*.

Appendix A furthermore shows that *when all risks are less than 0.5, it is impossible for the relative risk and risk difference to change in the same direction but in a direction opposite to that of the odds ratio*. In fact, as shown in Appendix A, all that is really needed for this result is $p_1+p_4\leqslant 1$ or $p_2+p_3\leqslant 1$.

HYPOTHETICAL EXAMPLES OF DISAGREEMENT

Table I illustrates several examples of disagreement by presenting various settings of the four risks together with k (ratio of relative risks), q (ratio of odds ratios), and m (ratio of risk differences). In order for $k>1$ and $q<1$ for a rare disease, the risks need to be displayed with more precision.

Table I. Hypothetical examples of disagreement.

p_1	p_2	p_3	p_4	k	q	m
0.6	0.8	0.33	0.5	1.14	0.76	0.85
0.6	0.8	0.79	0.95	0.90	1.89	0.80
0.0157	0.0963	0.0020	0.0128	1.04	0.97	0.13
0.06	0.19	0.01	0.11	3.47	3.33	0.77
0.03	0.26	0.28	0.77	0.32	0.76	2.13

DOI: 10.1002/sim

A DATA-ANALYTIC EXAMPLE OF DISAGREEMENT

Ford *et al.* [34] used logistic regression to analyze data from the 1996 Behavioral Risk Factor Surveillance System (BRFSS) survey in order to examine the relationship between physical health-related quality of life (QOL) and body mass index (BMI). A tangible example of disagreement of k, q, and m is obtained by analyzing data from the 2005 BRFSS in an analogous manner. Details of BRFSS survey methodology are presented in [35, 36] and can also be found at http://www.cdc.gov/brfss/. Records were available on 356 112 United States residents, of which 265 703 contained complete information on key variables, including income.

Ford *et al.* [34] reported on the relationship between QOL and BMI in terms of odds ratios, which were furthermore adjusted to control for confounding. Separate odds ratios were presented for males and females because effect-measure modification was uncovered. Judging by the adjusted odds ratios, the effect of high BMI on poor QOL is 'stronger' for women than for men, all else being equal. The comparison might lead one to conclude that women are more in need of an intervention. But what if the relative risk or risk difference were to indicate that the effect of high BMI on poor QOL was in fact 'stronger' for men? For the following discussion we will pretend that we can interpret the effect measures causally; after all, intervention trials are often justified using results of observational studies. If one had to choose between intervening on a relatively homogeneous (e.g. each member sharing approximately the same risk) group of women or of men, the goal of reducing each participant's risk of poor QOL could lead to a choice based on the relative risk; e.g. a relative risk of 2 for women but of 1.5 for men would indicate that the risk for women would be halved by the intervention, whereas that for men would be reduced only by a third. However, the goal of maximizing the net benefit by either intervening on 100 women or 100 men could lead to a choice based on the risk difference. A risk difference of 0.05 for women but 0.1 for men corresponds to a net benefit of 10 men *versus* 5 women helped by the intervention. This particular conflict of goals and choices arises exactly when the risks for men are equal to $p_1 = 0.2$ and $p_2 = 0.3$ and the risks for women are equal to $p_3 = 0.05$ and $p_4 = 0.1$.

Because the results of Ford *et al.* [34] are presented strictly in terms of odds ratios, we will use our analysis of the 2005 BRFSS data to study the problem further. We will let risks p_1 and p_2 represent the probability of reporting more than five physically unhealthy days during the past 30, for not obese and obese (BMI$\geqslant$30) people in a relatively homogeneous group of men (subpopulation one). The risks p_3 and p_4 represent the same for a relatively homogeneous group of women (subpopulation two). We will estimate these risks using a logistic regression model together with the coefficient estimates presented in Table II. SAS Proc Surveylogistic was used for the computation of these results, based on the 2005 BRFSS data. A product term for female and not obese was included as in Ford *et al.* [34].

From Table II, we see that the product term for female and not obese is statistically significantly negative, indicating that the odds ratio representing the association between obesity and poor physical QOL is higher for women than for men, all else being equal (all other covariates being the same). One can use the results in Table II to calculate the two risks for a given subpopulation. For example, let subpopulation one represent all white, college-educated males over the age of 65 who currently smoke everyday. Then the risk for the non-obese members of this group is estimated as $\text{expit}(-1.71-0.36+0.53)=0.177$, where $\text{expit}(x)=\exp(x)/(1+\exp(x))$ is the inverse logit transform. The risk for the obese members of this group is obtained similarly as 0.236.

By making these calculations for both women and men for each setting of all possible combinations of the remaining covariates, and then computing the relative risks and risk differences,

DOI: 10.1002/sim

Table II. Estimating the risk of more than five physically unhealthy days during the past 30 as a function of obesity and confounders, using logistic regression and data from the 2005 United States BRFSS survey.

Coefficient	Estimate (SE)*
Intercept	−1.71 (0.04)
Age	
18–24	−0.99 (0.06)
25–34	−0.99 (0.04)
45–54	−0.71 (0.03)
55–64	−0.44 (0.03)
65+	
Female	0.51 (0.04)
Male	
Not obese	−0.36 (0.04)
Obese	
Female and not obese	−0.19 (0.05)
Black non-hispanic	0.06 (0.04)
Hispanic	0.16 (0.05)
Multiracial non-hispanic	0.48 (0.08)
Other race non-hispanic	0.13 (0.07)
White non-hispanic	
Did not graduate high school	0.85 (0.04)
Graduated high school	0.45 (0.03)
Attended college	0.36 (0.03)
Graduated college	
Current smoker every day	0.53 (0.03)
Current smoker some days	0.37 (0.05)
Former smoker	0.25 (0.03)
Never smoked	

*Unexponentiated coefficients and standard errors are presented so that the reader can calculate the estimated risks.

one finds that both the relative risk and the risk difference are always higher for women than for men, all else being equal. Thus, it may seem safe to conclude that women need an intervention more than men. However, when subpopulation one is defined as above and subpopulation two represents all white, college-educated females between the ages of 18 and 24 who never smoked (so that the covariate set differs in terms of not just gender but also age and smoking status), we find an example of disagreement. Table III presents the results. Standard errors were obtained by simulation, based on the idea presented by Greenland [37]. Specifically, in R we simulated 10 000 multivariate normal coefficient estimates with covariance matrix returned by Proc Surveylogistic, and for each set of coefficients, we calculated the quantities in Table III (except for m, because it could possibly be negative).

Table III shows that the risk difference decreases by a factor of 0.69 (from 0.059 to 0.04), whereas the relative risk and odds ratio increase by factors of 1.25 ($1.33 \rightarrow 1.65$) and 1.22 ($1.44 \rightarrow 1.73$), respectively. Results of the simulation confirmed that the disagreement was statistically significant: 99.5 per cent of the simulations found $k>1, q>1$, and $s<0$.

What if these results had been presented in terms of odds ratios or relative risks only? Then one might be led to conclude that, even when the women are younger and have never smoked while the men are older smokers, the women are still more in need of an intervention. If only the

DOI: 10.1002/sim

Table III. Estimated risks and measures of effect modification, with standard errors.

Parameter	Estimate (SE)
p_1	0.177 (0.006)
p_2	0.236 (0.009)
p_3	0.061 (0.003)
p_4	0.101 (0.006)
k	1.252* (0.049)
q	1.215† (0.057)
s	−0.018‡ (0.007)
m	0.687‡

*Relative risk increases from 1.33 to 1.65.
†Odds ratio increases from 1.44 to 1.73.
‡Risk difference decreases from 0.06 to 0.04.

odds ratios had been presented, the results of this article would allow us to conclude that the risk difference would increase provided that 1.66 $p_3 \geqslant p_1 \geqslant p_3$. The reader would likely believe that the baseline risk for the men would be greater than that for the women ($p_1 \geqslant p_3$), because the men are older and smoke. Thus, if the reader also believed that $p_3 \geqslant p_1/1.66$, he or she would be able to conclude that the risk difference increases. In fact, it turns out that $p_3 = 0.061$, whereas $p_1 = 0.177$ and, thus, $p_3 < p_1/1.66$; fittingly, the risk difference does not increase.

If only the relative risks had been presented, we would know that the odds ratio increases provided that all four of the risks were less than $\frac{0.25}{1.25} = 0.2$. Or, if we believe that the risk difference increases, we would conclude that the odds ratio also increases. We would know that in order for the risk difference to decrease, p_3 needs to be less than p_1 and p_4 needs to be less than p_2—two things that are likely given that the men are older and smoke.

This data analysis highlights the importance of presenting coefficient estimates rather than odds ratio summaries, so that the reader can compute risks and whatever effect measure is sought. However, much literature reports only on a chosen measure, and the tools developed in this article can then guide the reader in trying to interpret results in terms of an alternative measure. It is also important to note that, although in this example, all three measures increased together from males to females when the other model covariates were fixed, this need not have been the case. Simply changing the coefficient of female from 0.51 to −0.8 leads to disagreement for some of the instances in which the other covariates are fixed. A final word of caution: effect-measure modification can, of course, be present but not causal. In the data-analytic example, we pretended that we wanted to use the results of a cross-sectional survey to plan an intervention study. As is often the case when planning an intervention, estimates based on observational studies are prone to multiple biases.

DISCUSSION

This article presented conditions under which the relative risk, odds ratio, and risk difference can and cannot change in opposite directions. Although necessary and sufficient conditions were seen to be quite complicated, necessary conditions were uncovered for each pair of measures. Whenever the relevant necessary conditions are implausible, an analysis reporting that one measure (e.g. the

DOI: 10.1002/sim

odds ratio) increases from one stratum to the next would also imply that another (e.g. the risk difference) does not decrease. Perhaps the most intriguing findings pertained to all three measures, assuming that exposure increases risk; if all four risks are less than 0.5, it cannot happen that k (ratio of relative risks) and m (ratio of risk differences) are on the same side of one but opposite of q (ratio of odds ratios), and if all four risks are greater than 0.5, it cannot happen that k and q are on the same side of one but opposite of m. Potentially useful results also emerged for each pair of measures. For example, when all risks are less than 0.10, if the odds ratio decreases from stratum 1 to stratum 2 ($q<1$), the relative risk cannot increase from stratum 1 to stratum 2 by a factor greater than $k=1.11$; for more common outcomes, when the odds ratio decreases across the two strata, the relative risk cannot double unless one of the four risks exceeds 0.5.

The simulation study showed that, in general, the odds ratio and risk difference change together more readily than the other two pairs of measures. So far, although the preceding investigation pointed to possible reasons for this, the true reasons remain unclear; further research might reveal the answer. For studies of rare diseases, the simulation study confirmed the obvious, which is that the relative risk and odds ratio will change together almost all the time, and certainly more often than the other two pairs.

Both data-analytic and hypothetical examples were presented for which the three measures changed in opposite directions. Finding a hypothetical example in which the relative risk and odds ratio change in opposite directions for a rare outcome was challenging. Finding a real example may be next to impossible. The data-analytic example showed that the relative risk and odds ratio increase and the risk difference decreases from subpopulation one to two. Further research will uncover still more data-analytic examples of each pair of measures changing in opposite directions. Additionally, while the present article focused on effect-measure modification from one stratum to the next, further research could investigate the much more complex situation that occurs when more than two strata are considered.

For researchers without a clear preference for one measure *versus* another, testing for effect modification is an ambiguous task. A possible remedy would be to formulate the alternative hypothesis in terms of all three measures; i.e. all three measures increase or decrease together. The multivariate delta method [10] would give the asymptotic joint distribution of (k, q, s), and a level α test could be obtained by rejecting whenever a $(1-\alpha)\times 100$ per cent approximate simultaneous confidence region was strictly contained within either of the subspaces $\{k<1, q<1, s<0\}$ or $\{k>1, q>1, s>0\}$. One could construct a one-sided test in an analogous fashion. It is likely that better tests could be constructed to increase power, but this remains a topic for future research. For case–control study designs allowing only for estimation of q, a Bayesian approach to hypothesis testing that incorporated priors on the four risks would perhaps be feasible.

APPENDIX A

Here we prove the following result: Suppose p_1, p_2, p_3, and p_4 are strictly between zero and one and satisfy $p_1+p_4\leqslant 1$ or $p_2+p_3\leqslant 1$. Then

$$p_2/p_1<p_4/p_3 \tag{A1}$$

and

$$p_2-p_1<p_4-p_3 \tag{A2}$$

DOI: 10.1002/sim

imply

$$\frac{p_2/(1-p_2)}{p_1/(1-p_1)} < \frac{p_4/(1-p_4)}{p_3/(1-p_3)} \tag{A3}$$

Furthermore, (A1) and (A2) imply (A3) when $<$ is replaced with $>$ in each of the inequalities.

Proof
Note that (A3) is equivalent to the inequality

$$p_2 p_3 + p_1 p_4 (p_2 + p_3) < p_1 p_4 + p_2 p_3 (p_1 + p_4) \tag{A4}$$

At this point, we state a lemma. □

Lemma 1
Suppose $a<b$, $c<d$, and either $c<1$ or $d<1$, then $a+bc<b+ad$.

Proof
Since $a+bc<b+ad$ is equivalent to $a(1-d)-b(1-c)<0$, we shall show the latter. Noting $1-d<1-c$, then if $c<1$ we have

$$a(1-d)-b(1-c)<a(1-c)-b(1-c)=(a-b)(1-c)<0$$

Again using $1-d<1-c$, then if $d<1$ we have

$$a(1-d)-b(1-c)<a(1-d)-b(1-d)=(a-b)(1-d)<0$$

Therefore, if either $c<1$ or $d<1$, then $a(1-d)-b(1-c)<0$, which proves the lemma. □

Assuming $p_2+p_3<1$ or $p_1+p_4<1$, we see that the lemma applies to (A4) with $a=p_2 p_3$, $b=p_1 p_4$, $c=p_2+p_3$, and $d=p_1+p_4$, where the assumptions are satisfied since (A1) implies $p_2 p_3<p_1 p_4$ and (A2) implies $p_2+p_3<p_1+p_4$. If $p_2+p_3=1$ or $p_1+p_4=1$ but not both, then (A4) is trivially true. If $p_2+p_3=p_1+p_4=1$, then (A2) is not satisfied.

Owing to the symmetry of the inequalities, if all $<$ symbols are replaced with $>$ symbols except on the inequalities $p_1+p_4 \leqslant 1$ and $p_2+p_3 \leqslant 1$, we obtain a proof to the second part of the problem involving the inequalities reversed.

ACKNOWLEDGEMENTS

The authors thank Dr Sander Greenland for his helpful comments, Dr Huamei Dong for her help with Appendix A, and Dr Elena Andresen for her encouragement to write this article.

REFERENCES

1. Rothman KJ, Greenland S. *Modern Epidemiology* (2nd edn). Lippincott-Raven: Philadelphia, PA, 1998.
2. Szklo M, Nieto FJ. *Epidemiology*: *Beyond the Basics* (2nd edn). Jones and Bartlett: Sudbury, MA, 2007.
3. Blot WJ, Day NE. Synergism and interaction: are they equivalent? *American Journal of Epidemiology* 1979; **110**:99–199.
4. Rothman KJ, Greenland S, Walker AM. Concepts of interaction. *American Journal of Epidemiology* 1980; **112**:467–470.

5. Saracci R. Interaction and synergism. *American Journal of Epidemiology* 1980; **112**:465–466.
6. Greenland S. Basic problems in interaction assessment. *Environmental Health Perspectives* 1993; **101**(Suppl 4): 59–66.
7. VanderWeele TJ, Robins JM. The identification of synergism in the sufficient-component-cause framework. *Epidemiology* 2007; **18**:329–339.
8. Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scandinavian Journal of Work, Environment and Health* 1988; **14**:125–129.
9. Morabia A, Ten Have T, Landis RJ. Interaction fallacy. *Journal of Clinical Epidemiology* 1997; **50**:809–812.
10. Agresti A. *Categorical Data Analysis* (2nd edn). Wiley: Hoboken, NJ, 2002.
11. Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. *British Medical Journal* 1976; **2**:1525–1536.
12. Kraemer HC. Reconsidering the odds ratio as a measure of 2×2 association in a population. *Statistics in Medicine* 2004; **23**:257–270.
13. Skrondal A. Interaction as departure from additivity in case–control studies: a cautionary note. *American Journal of Epidemiology* 2003; **158**:251–258.
14. McNutt L, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. *American Journal of Epidemiology* 2003; **157**:940–943.
15. Cummings P. Re: 'Estimating the relative risk in cohort studies and clinical trials of common outcomes' (Letter). *American Journal of Epidemiology* 2004; **159**:213–215.
16. Zou G. A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* 2004; **159**:702–706.
17. Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case–control studies. *American Journal of Epidemiology* 2004; **160**:301–305.
18. Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *American Journal of Epidemiology* 2005; **162**:199–200.
19. Petersen MR, Deddens JA. Re: 'Easy SAS calculations for risk or prevalence ratios and differences' (Letter). *American Journal of Epidemiology* 2006; **163**:1157–1163.
20. Kleinbaum DG, Klein M. *Logistic Regression*: *A Self-learning Text* (2nd edn). Springer: New York, NY, 2002.
21. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *American Journal of Epidemiology* 1987; **125**:761–768.
22. Greenland S, Thomas DC. On the need for the rare disease assumption in case–control studies. *American Journal of Epidemiology* 1982; **116**:547–553.
23. Lee J. Odds ratio or relative risk for cross-sectional data? *International Journal of Epidemiology* 1994; **23**:201–203.
24. Hughes K. Odds ratios in cross-sectional studies (Letter). *International Journal of Epidemiology* 1995; **24**: 463–464.
25. Osborn J, Cattaruzza MS. Odds ratio and relative risk for cross-sectional data (Letter). *International Journal of Epidemiology* 1995; **24**:464–465.
26. Greenland S, Robins J. Conceptual problems in the definition and interpretation of attributable fractions. *American Journal of Epidemiology* 1988; **128**:1185–1197.
27. Robins J, Greenland S. The probability of causation under a stochastic model for individual risk. *Biometrics* 1989; **45**:1125–1138.
28. Greenland S. Relation of probability of causation to relative risk and doubling dose: a methodologic error that has become a social problem. *American Journal of Public Health* 1999; **89**:1166–1169.
29. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *The New England Journal of Medicine* 1988; **318**:1728–1733.
30. McQuay HJ, Moore R. Using numerical results from systematic reviews in clinical practice. *Annals of Internal Medicine* 1997; **126**:712–720.
31. Pearl J. *Causality*: *Models*, *Reasoning*, *and Inference*. Cambridge University Press: Cambridge, U.K., 2000.
32. Poole C. Understanding the number needed to treat (abstract). *American Journal of Epidemiology* 2007; **165**:S143.
33. The R Development Core Team. *R 2.5.0—A Language and Environment*. The R Foundation for Statistical Computing, 2007. Available from: http://www.r-project.org/.
34. Ford ES, Moriarty DG, Zack MM, Mokdad AH, Chapman DP. Self-reported body mass index and health-related quality of life: findings from the behavioral risk factor surveillance system. *Obesity Research* 2001; **9**: 21–31.

DOI: 10.1002/sim

35. Gentry EM, Kalsbeek WD, Hogelin GC, Jones J, Gaines K, Forman M, Marks J, Trowbridge F. The behavioral risk factor surveys: design, methods, and estimates from combined state data. *American Journal of Preventive Medicine* 1985; **1**:9–14.
36. Remington PL, Smith MY, Williamson DF, Anda RF, Gentry EM, Hogelin GC. Design, characteristics, and usefulness of state-based behavioral risk factor surveillance: 1981–1987. *Public Health Report* 1988; **103**:366–375.
37. Greenland S. Interval estimation by simulation as an alternative to and extension of confidence intervals. *International Journal of Epidemiology* 2004; **33**:1389–1397.

DOI: 10.1002/sim