Journal of Clinical Epidemiology 82 (2017) 87–93

Journal of Clinical Epidemiology

Odds ratios deconstructed: A new way to understand and explain odds ratios as conditional risk ratios

Fred M. Hoppe[a,*], Daniel J. Hoppe[b], Stephen D. Walter[c]

[a]Department of Mathematics and Statistics, McMaster University, 1280 Main St. W., Hamilton, Ontario L8S 4K1, Canada
[b]Department of Orthopaedic Surgery, Stanford University, 450 Broadway St., Redwood City, CA 94063, USA
[c]Department of Clinical Epidemiology and Biostatistics, McMaster University, 1280 Main St. W., Hamilton, Ontario L8S 4K1, Canada

Accepted 9 August 2016; Published online 24 August 2016

Abstract

Objectives: The aim of this analysis was to provide an alternative derivation of the odds ratio (OR) to provide an intuitive meaning, freeing it from any mention of odds, which may make it a more useful concept for clinicians to use when describing treatment effect.

Study Design and Setting: By examining the four possible combinations of treatment/control and corresponding outcomes, we considered the conditional risk ratio (RR, also known as relative risk) of an event with the treatment compared with an event with the control for pairs of patients for whom treatment and control would yield different results. Both matched and unmatched studies are considered.

Results: We found that the OR could be derived as the RR of an outcome with treatment compared with an outcome with control conditional on the treatment and control resulting in different outcomes, thus providing a measure of the net benefit of treatment.

Conclusion: It has been claimed that the OR comparing the effect of treatment vs. control does not have the same clinical interpretability as RR because it involves ratios of odds and so is difficult to explain in terms of patient numbers. This new derivation provides an interpretation of the OR as an RR but conditional on treatment and control resulting in different outcomes. This may help explain the reason ORs cause interpretation difficulties in practice. Moreover, the OR may be a more clinically useful parameter to patients because it deals with only those situations where the outcome differs between the two groups.

Keywords: Conditional probability; Matched pairs; Number needed to treat; Odds; Odds ratio; Relative risk; Risk; Risk difference; Risk ratio; Unmatched pairs

1. Introduction

There are a number of commonly used numerical measures of clinical effect that report the results of clinical trials, whether they be randomized controlled trials, cohort studies, or case–control studies. These measures are estimates of probabilities, namely parameters that represent the likelihood of an event or an outcome under different conditions in a hypothetical population of patients. Sometimes, the term risk is used interchangeably with probability although risk involves conditions such as time duration or a population denominator. A treatment effect is generally measured by the probabilities of the occurrence of events.

For comparing two populations with rates p_1 (treatment) and p_2 (control) for adverse events,[1] a number of commonly used measures (defined in the Appendix) in the literature are risk difference (RD), risk ratio (RR), odds ratio (OR), relative risk reduction, and number needed to treat (NNT). The emphasis in this article is on the OR and its comparison with RR.

These populations are conceptual only. The samples come from the same pool of patients, and it is the randomization to treatment or to control that generates samples. In addition, the manner in which data are collected to estimate parameters does not change their meaning. An OR should

Conflict of interest: None to report.

* Corresponding author. Tel.: +1-905-525-9140x24688; fax: 905 522-0935.

E-mail address: hoppe@mcmaster.ca (F.M. Hoppe).

[1] Instead of adverse events, we may instead consider beneficial events in which risk is replaced by the probability of a beneficial event, but for the purpose of this article, events are taken as adverse. A parallel interpretation for beneficial events extends appropriately, mutatis mutandis, so will not be repeated here.

http://dx.doi.org/10.1016/j.jclinepi.2016.08.007
0895-4356/

What is new?

- It has been widely asserted that the odds ratio comparing the effect of treatment vs. control does not appear to have the intuitive interpretability that the risk ratio has because it involves ratios of odds and is therefore difficult to explain from a probabilistic perspective that can be understood in terms of patient numbers.
- In this article, we provide a novel derivation of the odds ratio as a ratio of probabilities rather than a ratio of odds, that is, a conditional risk ratio of an event with treatment compared with an event with control among pairs of patients, both matched and unmatched, for whom treatment and control would yield different results, thus providing a measure of the net benefit of treatment.
- This derivation obviates some of the concerns that have been expressed, all of which emanate from the fact that odds are not probabilities.

represent the same quantity regardless of whether it is estimated by independent cohort studies or by matched pairs. It is a quantity that depends only on p_1 and p_2.

To keep the discussion concrete, consider data extracted from a blinded randomized trial in which a tibial shaft fracture was treated with either reamed intramedullary nailing or unreamed intramedullary nailing [1]. Table 1 shows the data for patients having closed fractures. The adverse event involved a primary composite outcome, including nonunion requiring implant exchange or bone grafting and dynamization of the nail.

The estimated treatment risk is shown as $\widehat{p}_1 = \frac{a}{a+b} = \frac{45}{416} = 0.1082$, whereas the control risk is $\widehat{p}_2 = \frac{c}{c+d} = \frac{68}{410} = 0.1659$, giving an RR of treatment to control[2]

$$\widehat{\mathrm{RR}} = \frac{\widehat{p}_1}{\widehat{p}_2} = \frac{0.1082}{0.1659} = 0.65.$$

The statistic used to estimate the OR would be

$$\widehat{\mathrm{OR}} = \frac{\widehat{p}_1/(1-\widehat{p}_1)}{\widehat{p}_2/(1-\widehat{p}_2)} = \frac{\widehat{p}_1(1-\widehat{p}_2)}{\widehat{p}_2(1-\widehat{p}_1)} = \frac{ad}{bc} = 0.61 \qquad (1)$$

Thus, if 1,000 patients were treated with reamed intramedullary nailing, we would expect about $45/416 \times 1000 \approx 108$ to experience an adverse event, whereas of 1,000 treated with unreamed nailing, we would expect about $68/410 \times 1000 \approx 166$ to experience an adverse event. Thus, the proportion of treated patients relative to control who would be expected to experience an adverse event is $108/166 = 0.65$, which is the RR shown in Table 1.

The RR thus has a direct meaning to a physician in terms of patient numbers as a comparison of two risks. What about the OR?

ORs and RRs are different measures and because a larger odds reflect a larger risk, and vice versa, then the ratio of two odds, just like the ratio of two probabilities, clearly provides some measure of the relative effectiveness of treatment vs. control or two treatments. Is there a corresponding meaning in terms of patient numbers?

A number of articles in this journal have recently used the OR in epidemiologic studies but generally with respect to comparison of conclusions with those derived using RR instead [2–6]. In fact, as pointed out in the Cochrane Handbook [7] "The nonequivalence of the risk ratio and odds ratio does not indicate that either is wrong: both are entirely valid ways of describing an intervention effect. Problems may arise, however, if the odds ratio is misinterpreted as a risk ratio."

Yet, among all comparison measures, the OR seems to be the measure that is least understood intuitively [8].

It is our view, that if users are going to refer to the OR, then they should have a careful and rigorous appreciation of what the OR conveys as a parameter in simple language. Unfortunately, denoting an OR as a ratio of odds does not endow it with such an intuitive meaning. We therefore wondered whether ORs have a meaning that could be simply understood in a manner similar to how an RR may be understood.

In this note, we hope to provide a simple novel derivation of the OR that completely bypasses odds and uses only probabilities, from which the OR is then derived as a bona fide ratio of risks.

Precisely, we point out that the OR is the RR of an outcome with the treatment compared with an outcome with the control *conditional* on there being a difference in outcomes between the treatment and the control. This is obvious, once stated, but we have not found such a statement in the literature and believe that this observation is new, useful, and obviates some of the concerns that have been expressed, all of which emanate from the fact that odds are not probabilities.

2. Common OR as a conditional RR: parallel groups

Whichever of RR or OR is used for comparison of two treatments or comparison of treatment with control, the choice is to provide a comparison between two probabilities (risks) p_1 and p_2. Applying such comparison to an individual patient, a researcher may then envisage the possibility of giving the treatment and control simultaneously to the same sample from the target population of individuals and under the same conditions. Of course, this is not possible. Instead, one might have a parallel group

[2] For clarity of exposition, we distinguish between parameters such as RR and their estimates, correspondingly denoted as $\widehat{RR}$, for instance, even if the meaning is clear from the context.

Table 1. Reamed and unreamed intramedullary nailing

Choice	Adverse	Not adverse	Totals	Risk	Odds	Risk ratio	Odds ratio
Treatment (reamed)	$a = 45$	$b = 371$	416	0.1082	0.1213		
						0.65	0.61
Control (unreamed)	$c = 68$	$d = 342$	410	0.1659	0.1988		
Totals	113	713	826				

design, as in the aforementioned example in Table 1, in which n_1 patients are randomized to the treatment and n_2 patients are randomized to the control. If a, c are the numbers of patients experiencing adverse events observed in the two groups, whereas b, d are the corresponding numbers not experiencing adverse events, then the sample proportions $\widehat{p}_1 = \frac{a}{n_1}, \widehat{p}_2 = \frac{c}{n_2}$ are estimates of p_1 and p_2 and may be substituted for them to obtain corresponding estimates of RR and OR.

Consider comparing treatment and control, but now collect data pairwise in unmatched pairs in which n pairs of patients are randomized within each pair, one to treatment, one to control, but instead of merely calculating event rates in each group separately, we now have four possible combinations of outcomes, depending on whether there was an adverse event or not in either pair. This is shown in Table 3 in the Appendix.

Because p_1 and p_2 are the individual (marginal) probabilities of occurrence of events with treatment and control, respectively, then because the pairs are unmatched, the outcomes in each pair are independent, meaning that the joint probabilities of all four possible combinations of outcomes are given by the products:

1. Event with both treatment and control: probability $= p_1 p_2$.
2. No event with both treatment and control: probability $= (1-p_1)(1-p_2)$.
3. Event with treatment, no event with control: probability $= p_1(1-p_2)$.
4. Event with control, no event with treatment: probability $= (1-p_1)p_2$.

It is easier to use integers to interpret these probabilities [9,10] by considering the expected breakdown for 1,000 pairs of randomized patients, according to the results of treatment (reamed) or control (unreamed) data in Table 1.

1. In $45/416 \times 68/410 \times 1000 \approx 18$ pairs, both treatment and control lead to an event.
2. In $371/416 \times 342/410 \times 1000 \approx 744$, neither treatment nor control leads to an event.
3. In $45/416 \times 342/410 \times 1000 \approx 90$ pairs, the treatment yields an event but not the control.
4. In $371/416 \times 68/410 \times 1000 \approx 148$ pairs, the control yields an event but not the treatment.

Table 2 summarizes these computations.

Table 2. Expected results for 1,000 pairs: reamed vs. unreamed nailing

Both adverse	Neither adverse	Only treatment adverse	Only control adverse
18	744	90	148

Thus, of 1,000 pairs of randomized patients, we expect about $18 + 744 = 762$ to show no difference in outcome between reamed or unreamed nailing. Of the remaining 238 patients, a fraction $90/238 = 0.3782$ would experience an adverse event that occurs only with the treatment, whereas in a fraction $148/238 = 0.6218$, the adverse event would occur only with the control. The corresponding RR of treatment, conditional on a difference in outcomes between treatment and control, is $0.3782/0.6218 = 0.61$ or equivalently, the RR that an event occurs with treatment among those cases where treatment and control produce different outcomes is 0.61, which is just the OR.

Thus, the OR is like an RR, but only in cases where the outcomes would be different for treatment and control. It has an interpretation in terms of patient numbers, like the RR, but gives different information concerning the possible benefits of a treatment.

So, in the aforementioned example, in situations where there is a difference between treatment and control, then 61% of the time, the treatment will be better.

Consider then the perspective of a patient deciding between treatment or control based on Table 3. The first column can be summed and divided by the common sample size n to arrive at $\widehat{p}_1$, estimating the probability of an adverse treatment event, and likewise summing the first row and dividing by n gives an estimated probability $\widehat{p}_2$ of an adverse control event. Division then yields an estimated risk ratio $\widehat{\mathrm{RR}} = \frac{\widehat{p}_1}{\widehat{p}_2}$.

Alternatively, it might be reasoned using Table 2 as follows. For cases (1) and (2), neither treatment nor control yields an excess adverse event.[3] In case (3), the treatment results in an excess adverse event, and in case (4), the control results in an excess adverse event. Because the first two cases yield the same outcome of interest and so have the same effect (ignoring any other attributable to the treatment), they might reasonably be excluded from comparison. In other words, it could be considered how likely case (3) is relative to (4), which is the probability of case (3) divided by the probability of case (4), that is,

$$\mathrm{OR} = \frac{p_1(1-p_2)}{p_2(1-p_1)} = \textit{Conditional}\,\mathrm{RR} \tag{2}$$

(and would be estimated by Eq. (1)) the familiar OR now arrived at without introducing odds, then taking the ratio of two odds and simplifying, but instead obtained directly

[3] We use the term excess to refer to a situation where either control or treatment results in an event, but not both, or equivalently, when the outcomes are different.

as an RR comparing the probabilities of *excess* adverse events under treatment and control.

We point out that a patient would not know in advance if she would be in a discordant group. If not, then the expected effect of treatment or control would be the same. But if she is, the OR then reflects the (conditional) RR of an event occurring with treatment, as indicated previously.

We have thus freed ORs from any mention of odds themselves. The fact that the OR has traditionally been expressed as a ratio of odds may obscure its meaning as the RR of an excess adverse event and so having a direct interpretation as a ratio of probabilities. In other words, the OR is just an RR but conditional on treatment and control resulting in different outcomes.

3. General OR as a conditional RR: matched pairs

Benefit of pairing accrues when the pairs are matched to reduce variation, so next consider how the aforementioned ideas extend to matched pairs (for instance, pairs of twin patients who are randomized one to treatment and one to control or case–control studies). The only difference is that we can no longer assume independence of outcomes within each pair nor with case–control can we even estimate p_1 and p_2.

The four possible combinations of outcomes and probabilities are now the following:

1. Event with both treatment and control: probability = p_{11}.
2. No event with both treatment and control: probability = p_{00}.
3. Event with treatment and no event with control: probability = p_{10}.
4. Event with control and no event with treatment: probability = p_{01}.

Here, the first subscript represents treatment whereas the second subscript represents control, and 1 represents event while 0 represents no event. A table of expected values like Table 2 of pairs of combinations can be constructed, and likewise, a patient using such a table would argue exactly as in the unmatched case.

The natural definition for the OR parameter would no longer be Eq. (2), as for independent samples, but instead it should be defined more generally as

$$\mathrm{OR} = \frac{p_{10}}{p_{01}} \tag{3}$$

subsuming Eq. (2), and this is the general definition that we are proposing for the OR.

Following a reviewer's suggestion, we can express Eq. (3) as a ratio of expected number of discordant events in patients responsive to treatment relative to the expected number of discordant events in patients who are responsive to control because we can multiply both numerator and denominator in Eq. (3) by the number of pairs, in which case, both the numerator and denominator become expected number of events, as is done in Table 2.

The issue of estimation of OR then becomes secondary and dependent on the sampling method. An estimate might be given by Eq. (1), assuming independence within each pair, or by

$$\widehat{\mathrm{OR}} = \frac{X_{10}}{X_{01}} \tag{4}$$

if the pairs were dependent.

However, regardless of the sampling design, the OR, considered as a parameter, rather than an estimate, would still be given by Eq. (3), consistent with our interpretation as a conditional RR.

4. Generalizing RD and number needed to treat

Referring again to Sections 2 and 3, observe that

$$p_2(1-p_1) - p_1(1-p_2) = p_2 - p_1 = \mathrm{RD}.$$

In n unmatched pairs of patients, randomized one to treatment and one to control, $np_1(1-p_2)$ are expected to experience an adverse event with treatment and not control, whereas $np_2(1-p_1)$ are expected to experience an adverse event with control but not treatment. The difference $n(p_2(1-p_1)-p_1(1-p_2))$ is the net number of excess adverse events with control and not treatment. So on average, it takes n pairs to obtain $n(p_2(1-p_1)-p_1(1-p_2))$ additional net excess adverse events with the control, and on average, it takes $n/n(p_2(1-p_1)-p_1(1-p_2)) = 1/p_2-p_1$ pairs to have one additional adverse event. This recovers NNT on the basis of net adverse events in an unmatched paired experiment. Note that in this context, we are viewing NNT as a parameter rather than an estimate based on data. However, the same conclusions hold for the latter because algebraically $\widehat{p}_2(1-\widehat{p}_1) - \widehat{p}_1(1-\widehat{p}_2) = \widehat{p}_2 - \widehat{p}_1$.

This suggests definitions of RD and NNT applicable to matched pairs by replacing $p_2(1-p_1)$ with p_{01} and replacing $p_1(1-p_2)$ with p_{10}. This leads to the following proposed definitions of RD and NNT in such settings

$$\mathrm{RD} = p_{01} - p_{10}$$

$$\mathrm{NNT} = \frac{1}{p_{01} - p_{10}}.$$

These definitions follow from $p_1 = p_{10} + p_{11}, p_2 = p_{01} + p_{11}$ implying $p_2 - p_1 = p_{01} - p_{10}$ because the common p_{11} cancels. These parameters can be estimated from data available from matched pair and case–control studies.

5. Discussion

A number of authors have observed that the OR does not have the same clinical interpretability as the RR. Sackett

et al. [11] present five properties of ORs that interfere with their clinical application; Bland and Altman [12] state that ORs can cause difficulties in interpretation; Katz [13] regards ORs as intuitively difficult to explain, in contrast to RR; Fleiss [14] states that ORs are intuitively less understandable than RRs; Knol et al. [15] warn of potential misinterpretation of ORs, in particular when the incidence rate is large, whereas Sinclair and Bracken [8] say that "In our view, this nonuse is quite justified not only is there a lack of clinical meaning but also there may be several grounds for confusion when the treatment effect is expressed as an odds ratio." Sometimes an OR has even been misinterpreted as an RR, as pointed out in Refs. [15,16].

Walter [17] notes that often the OR is compared with an apriori standard, the RR, and is not considered as useful because the two measures differ. He states that it is not a sufficient reason to reject use of OR in favor of RR merely because of having different values. He stresses the importance of separating the statistical evaluation of data with communication of risk to physicians and patients in comparing the different parameters RD, RR, OR and provides numerical and graphical examples examining the behavior of a treatment group risk as the control group risk varies while each parameter simultaneously remains constant.

We believe that the difficulties and criticisms expressed previously lie not so much with odds themselves, but rather in a lack of understanding of *ratios* of odds. Probabilities and odds are easier to understand than an OR, which is a ratio of ratios.

The risk p of an event can be understood in terms of patient numbers. In a random sample of n patients, the number of events is given by a binomial random variable. If independent random samples of size n_1, n_2 are taken from populations with risks p_1, p_2, respectively, then the numbers of corresponding events are binomial random variables on n_1, n_2 trials and success probabilities p_1, p_2, respectively. As a result, the risks can be understood in terms of patient numbers. This accounts for the success of NNT in translating an RD to more concrete terms, despite NNT and RD merely being inverses of each other.

In the same spirit, the RR can be understood in terms of a ratio of numbers. For instance, if RR = 0.5, then one would expect roughly half as many events with treatment vs. control for the same sample sizes. As a result, both RD and RR have a concrete clinical interpretability.

With respect to odds and ORs, such interpretations are not as apparent. An odds is the ratio of the probability that an event occurs divided by the probability that it does not occur. It is however difficult to relate odds to patient numbers. Rather, odds are commonly used in games of chance because they relate directly to payoffs, which are expressed as payoff odds vs. the true odds and so have a meaning to a gambler by directly indicating what the payoff is compared with what the payoff should be if a game were fair. In fact, it is stated by Sinclair and Bracken [8] that risk is familiar to physicians, whereas odds are familiar to gamblers, yet an OR is a stranger to both physicians and gamblers. But in clinical situations, there is no game of chance with a numerical payoff.

Just as RR can be understood in terms of a ratio of numbers, now so can OR. If OR = 0.5, then one would expect roughly half as many events with treatment vs. control for the same sample sizes in those cases where treatment and control would give different outcomes.

This interpretation also provides a *nonmathematical explanation* for the known fact that RR $<$ 1 implies OR $<$ RR. Recall that OR considers only those situations where treatment pairs have the expected effect (treatment does not result in an adverse outcome, whereas control does) vs. where the treatment pairs have the nonexpected effect (treatment results in an adverse outcome, whereas control does not) and is the RR for such pairs. When RR $<$ 1, it is more likely, when the outcomes are different, for the control to have yielded an excess adverse event, thus making the corresponding conditional RR, in other words, the OR, smaller than it would be if the concordant events were also included as in the RR. In other words, RR $<$ 1 implies OR $<$ RR. Of course, this inequality is obvious from algebra, and our explanation merely serves to provide some insight.

Note that if RR $>$ 1, then a parallel argument gives the opposite inequality OR $>$ RR.

6. Conclusions

We believe that if health researchers use statistical terms in communicating research results, there should be a clear understanding of the clinical meaning of these terms. In this article, we examined the OR, whose use has been subject to criticism and have provided a new derivation that omits mention of odds and which can be extended to a more general definition of the OR. Whereas RR describes how a treatment will compare with control with a long sequence of patients by giving the ratio of events our derivation gives the OR a clinical interpretation as a RR of treatment compared with control in a large sequence of patients for whom treatment and control yield *different* results, thus providing an interpretation of another measure of the net benefit of treatment. Our purpose was to provide an understanding of the OR that can be described in words, and we have shown how the interpretation applies equally to parallel groups, to matched pairs, and to unmatched pairs.

We hope that our perspective dispositions the criticisms leveled at the OR as lacking clinical value while at the same time providing an easily understood intuitive meaning.

Acknowledgments

The authors thank Gordon Guyatt for helpful comments on an early draft of this article. F.M.H. and S.D.W. were supported by the Natural Sciences and Engineering Research Council of Canada.

Appendix

Suppose that $p_1 < p_2$ represent event rates of adverse events in a treatment group and control group, respectively. Among the measures of clinical benefit, most commonly used are:

$$\text{risk difference} = RD = \text{control risk} - \text{treatment risk} = p_2 - p_1$$

$$\text{risk ratio} = RR = \frac{\text{treatment risk}}{\text{control risk}} = \frac{p_1}{p_2}$$

$$\text{odds} = \frac{\text{risk}}{1-\text{risk}} = \frac{p}{1-p}$$

$$\text{odds ratio} = OR = \frac{\text{treatment odds}}{\text{control odds}} = \frac{p_1/(1-p_1)}{p_2/(1-p_2)} \quad (5)$$

$$\text{relative risk reduction} = RRR = \frac{RD}{p_2}$$

$$\text{number needed to treat} = NNT = \frac{1}{p_2 - p_1}$$

In the definition of odds, the risk p may be either p_2 or p_1. The relative risk reduction = 1 − risk ratio and so provides no additional information than the latter, while the number needed to treat (NNT assumes $p_1 < p_2$.) is the inverse of the risk difference and likewise provides no additional information than the risk difference except that it is a more intuitive measure to grasp in terms of patient numbers needed to achieve some outcome on average.

Unmatched pairs

Suppose data are collected with unmatched pairs of patients who are randomized, one to treatment and one to control. The results are summarized in Table 3 with the same convention on subscripts as in Section 3. For instance, X_{10} is the number of pairs with a treatment event but no control event.

Because the pairs are unmatched and independently randomized within each pair, the random variables in Table 3 are marginally binomial. (In fact, more generally the random vector of the four possible combinations of outcomes has a multinomial distribution with the above probabilities of success.)

Table 3. Pairing

	Treatment	
Control	**Event**	**No event**
Event	X_{11}	X_{01}
No Event	X_{10}	X_{00}

$$\begin{aligned} X_{11} &\sim \text{Binomial}(n, p_1 p_2) \\ X_{00} &\sim \text{Binomial}(n, (1-p_1)(1-p_2)) \\ X_{10} &\sim \text{Binomial}(n, p_1(1-p_2)) \\ X_{01} &\sim \text{Binomial}(n, (1-p_1)p_2) \end{aligned}$$

The estimated probability of an adverse event following treatment is $\widehat{p}_1 = \frac{X_{11}+X_{10}}{n}$, while with a control, it is $\widehat{p}_2 = \frac{X_{11}+X_{01}}{n}$, yielding an estimated risk ratio $\widehat{RR} = \frac{\widehat{p}_1}{\widehat{p}_2}$ and an estimated odds ratio $\widehat{OR}$, as if there were no matching, which is given by Eq. (1).

Our intention in this paper has been on interpreting the odds ratio, for which existing estimators are well known, for instance as developed in [18]. When the data representing treatment and control groups arise independently, as described by Table 1, the appropriate estimated odds ratio $\widehat{OR}$ is given by Eq. (1) whose variance is estimated by

$$\widehat{OR}^2 \left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \right)$$

The same estimator is used in a matched pairs situation, when the pairs are independently randomized to treatment or control or where the matching is on variables not associated with the outcome of interest, leading to independence within pairs and then $a = X_{11} + X_{10}, b = X_{01} + X_{00}$, $c = X_{11} + X_{10}, d = X_{10} + X_{00}$ with the notation of Table 3.

Matched pairs

Paired data arise more generally when the pairs are matched on similar characteristics. Such data can also be represented by a table like Table 3, but we cannot assume independence of outcomes within each pair. The random variables in Table 3 are still marginally binomial but with occurrence probabilities that are not products, rather

$$\begin{aligned} X_{11} &\sim \text{Binomial}(n, p_{11}) \\ X_{00} &\sim \text{Binomial}(n, p_{00}) \\ X_{10} &\sim \text{Binomial}(n, p_{10}) \\ X_{01} &\sim \text{Binomial}(n, p_{01}) \end{aligned}$$

A patient using the data in Table 3 would argue exactly as in the unmatched case and come up with an estimated odds ratio

$$\widehat{OR} = \frac{X_{10}}{X_{01}}$$

as given by Eq. (4) because $\frac{X_{10}}{n}$ estimates p_{10} and $\frac{X_{01}}{n}$ estimates p_{01}. Now, the variance of the estimated odds ratio is given by

$$\widehat{OR}^2\left(\frac{1}{X_{10}}+\frac{1}{X_{01}}\right)$$

Eq. (4) is known as the Mantel–Haenszel estimate of the odds ratio from matched pairs case–control data [19, 20]. The Mantel–Haenszel estimate is given by

$$\widehat{OR}^{MH}=\frac{\sum_{i=1}^{n}a_i d_i}{\sum_{i=1}^{n}b_i c_i} \tag{6}$$

where $\{a_i, b_i, c_i, d_i\}$ play the same role as $\{a, b, c, d\}$ in Eq. (1) but with respect only to the data from pair $1 \le i \le n$. The possible values of $\{a_i, b_i, c_i, d_i\}$ are in $\{0, 1\}$ and Eq. (6) then reduces to Eq. (4).

Not only is the odds ratio, as we have derived it, a conditional risk ratio, but, curiously, it is also a conditional odds (not ratio) being the ratio of probabilities of two events that are conditionally mutually exclusive, as we can see by writing

$$OR=\frac{p_1(1-p_2)}{p_1(1-p_2)+p_2(1-p_1)}\Bigg/\left(1-\frac{p_1(1-p_2)}{p_1(1-p_2)+p_2(1-p_1)}\right)=\text{Conditional Odds}$$

We can then define a conditional risk given that there is a difference in outcomes as follows. In the usual manner in which an odds can be converted into a probability, a conditional odds ratio of *OR* considered, as we have just shown, to be a conditional odds, is equivalent to a conditional risk (not ratio)

$$\textit{Conditional}\ \text{Risk}=\frac{OR}{1+OR} \tag{7}$$

which can be interpreted as a conditional probability that the treatment does not prevent an adverse event in those cases where treatment and control have different outcomes.

Interestingly, an equation identical to Eq. (7) appears as Equation (5.2) in [20] in the context of case–control studies although in an earlier Eq. (2.13) of [20] the odds ratio is still defined as a ratio of odds as in Eq. (5) unlike our more general Eq. (3) that does not involve odds.

References

[1] SPRINT Investigators, Bhandari M, Guyatt G, Tornetta P 3rd, Schemitsch EH, Swiontkowski M, Sanders D, et al. Randomized trial of reamed and unreamed intramedullary nailing of tibial shaft fractures. J Bone Joint Surg Am 2008;90:2567–78.

[2] Furuya-Kanamori L, Suhail AR, Doi SAR. The outcome with higher baseline risk should be selected for relative risk in clinical studies: a proposal for change to practice. J Clin Epidemiol 2014; 67:364–7.

[3] Mirzazadeh A, Malekinejad M, Kahn JG. Relative risk reduction is useful metric to standardize effect size for public heath interventions for translational research. J Clin Epidemiol 2015;68:317–23.

[4] Shrier I, Pang M. Confounding, effect modification, and the odds ratio: common misinterpretations. J Clin Epidemiol 2015;68:470–4.

[5] Papageorgiou SN, Tsiranidoud E, Antonoglou GN, Deschner J, Jäger A. Choice of effect measure for meta-analyses of dichotomous outcomes influenced the identified heterogeneity and direction of small-study effects. J Clin Epidemiol 2015;68:534–41.

[6] Meister R, von Wolff A, Kriston L. Odds ratios of treatment response were well approximated from continuous rating scale scores for meta-analysis. J Clin Epidemiol 2015;68:740–51.

[7] Cochrane Handbook for Systematic Reviews of Interventions. Effect measures for dichotomous outcomes, Chapter 9, Section 2. Available at: http://handbook.cochrane.org/chapter_9. Accessed June 10, 2016.

[8] Sinclair JC, Bracken MG. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol 1994;47:881–9.

[9] Kahneman D, Slovic P, Tversky A. Judgement under uncertainty: heuristics and biases. Cambridge, UK: Cambridge University Press; 1982.

[10] Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: frequency formats. Psych Rev 1995;102: 684–704.

[11] Sackett DL, Deeks JJ, Altman DG. Down with odds ratios!. Evid Based Med 1996;1:164.

[12] Bland JM, Altman DG. Statistics notes: the odds ratio. Br Med J 2000;320:1468.

[13] Katz KA. The (relative) risk of using odds ratios. Arch Dermatol 2006;142:761–4.

[14] Fleiss JL. The statistical basis of meta-analysis. Stat Methods Med Res 1993;2:121–45.

[15] Knol MJ, Le Cessie S, Algra A, Vandenbroucke JP, Groenwold RH. Overestimation of risk ratios by odds ratios in trials and cohort studies: alternatives to logistic regression. Can Med Assoc J 2012; 184:895–9.

[16] Davies HTO, Crombie IK, Tavakoli M. When can odds ratios mislead? Br Med J 1998;316:989–91.

[17] Walter SD. Choice of effect measure for epidemiological data. J Clin Epidemiol 2000;53:931–9.

[18] Fleiss JL. Statistical methods for rates and proportions. 3rd ed. Hoboken, NJ: John Wiley & Sons; 2003.

[19] Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. J Natl Cancer Inst 1959;22: 719–48.

[20] Breslow NE, Day NE. Statistical methods in cancer research. Volume 1—The analysis of case control studies. Lyon, France: International Agency for Research on Cancer; 1980.