Debate & Analysis

Against all odds?

Improving the understanding of risk reporting

GPs are increasingly called upon to make or guide patients with choices about medical interventions, but there is gathering evidence that clinicians’ understanding of risk is poor and, correspondingly, that their ability to communicate risk is deficient.1 In this article we aim to improve health professionals’ understanding of risk reporting and clarify common misunderstandings in interpreting risk, odds, relative risk (RR), and odds ratios (ORs).

RISK

The risk of an event (or disease) happening is simply the number of times it occurs divided by the number of occasions on which it could potentially occur. It is usually expressed as a proportion or as a percentage. For example, to give the risk of breast cancer is to say that for every eight women, one will develop breast cancer, which is a lifetime risk of 1/8, or 0.13.2 We might also express this by saying 12% of women develop breast cancer.

This lifetime risk is not the same as the prevalence of breast cancer, which is the total number of cases in a given population at a specific time. Prevalence is dependent on factors including survivability and associated comorbidity, and is closer to 2% in the UK.2

Using simple frequency statements — such as ‘one in eight women will develop breast cancer’ — avoids a potential source of miscommunication with patients. He had intended to breast cancer ‘— avoids a potential source of miscommunication with patients.

Gigerenzer’s work highlights this issue in the common misunderstandings in interpreting risk, odds, relative risk (RR), and odds ratios (ORs).

In health, odds are the number of people who experience an event, divided by the number of those who do not; the odds of a woman developing breast cancer are 1/7, which is 0.14. However, saying that 0.14 women develop breast cancer for every one that does not, makes little intuitive sense. Indeed, odds are much easier to understand if they are >1 than if they are <1; for example, for every one woman who develops breast cancer, seven will not — the ‘odds against’ developing breast cancer are seven to one. Unfortunately, as we will see below, ORs are often based on odds of <1.

Whether risk or odds are easier to interpret is subjective, but many authors consider the former to be more intuitive.3 Health researchers often imply that having an understanding of odds is dependent on having an insight into the betting industry;4 however, the analogy between health statistics and gambling is a perilous one. Odds in health are an accurate representation of frequencies in a specific setting whereas, in sports betting, the bookmakers’ odds are a statement of the return on your stake; this is an indication of how much you will win.7

THE RELATIONSHIP BETWEEN RISK AND ODDS

When a condition is rare, risk and odds are approximately equal. For example, in our breast cancer example, the lifetime risk is 1/8 (or 0.13) and the lifetime odds 1/7 (or 0.14). As conditions become rarer, the approximation becomes more exact. The lifetime risk of developing pancreatic cancer is 1/86 (or 0.0116) and its lifetime odds 1/85 (or 0.0117).2 However, for more common conditions, the approximation is poor: the lifetime risk of developing any cancer is estimated to be 1/3 (or 0.33), but the lifetime odds are 1/2 (or 0.50).2

Table 1 not only illustrates how, at low values, the odds and the risk of an event are equivalent — a proportion of 0.01 is the equivalent of odds of 0.01 — but also demonstrates that, as the proportion rises, the odds diverge such that, above values of 20%, the odds start to bear no relation to the proportions. In addition, risk ranges from zero to one, whereas odds range from zero (the event will never happen) to infinity (the event will always happen).

COMPARING OUTCOMES IN PROSPECTIVE STUDIES USING RELATIVE RISKS AND ODDS RATIOS

In prospective studies, such as clinical trials and observational studies, groups of participants with different characteristics are followed up to determine whether one group is more prone to a particular outcome than another. The difference in outcomes between an ‘exposed group’ and an ‘unexposed group’ (in a clinical trial, the intervention and control groups) can be measured on the risk or the odds scales.

The RR is calculated as the risk in the exposed group divided by the risk in the unexposed group. It tells us how much risk has increased or decreased from an initial level; most authors consider it to be readily understood.5,8 The OR is calculated as the odds in the exposed group divided by the odds in the unexposed group. As many authors have recognised, the OR is more difficult to visualise.5,9,10 How would we interpret an OR of, say, 0.5 or 3? One suspects that many readers will have no intuitive feel for the size of the difference when expressed in this way.

Both RRs and ORs are >1 when the outcome of interest is more common in the exposed group (an example being a prospective trial showing harm); both are <1 when the outcome is more common in the unexposed group (typically, a prospective trial showing benefit). Confidence intervals (CIs) around each ratio, as well as P-values, can be used to assess whether differences in the

<table>
<thead>
<tr>
<th>Risk as percentage (%)</th>
<th>Risk as decimal</th>
<th>Odds as decimal</th>
<th>Odds as fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.01</td>
<td>1/99</td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>0.05</td>
<td>1/19</td>
</tr>
<tr>
<td>10</td>
<td>0.10</td>
<td>0.11</td>
<td>1/9</td>
</tr>
<tr>
<td>20</td>
<td>0.20</td>
<td>0.25</td>
<td>1/4</td>
</tr>
<tr>
<td>30</td>
<td>0.30</td>
<td>0.5</td>
<td>1/2</td>
</tr>
<tr>
<td>50</td>
<td>0.50</td>
<td>1</td>
<td>1/1</td>
</tr>
<tr>
<td>60</td>
<td>0.60</td>
<td>1.5</td>
<td>6/4</td>
</tr>
<tr>
<td>70</td>
<td>0.70</td>
<td>2.3</td>
<td>7/3</td>
</tr>
<tr>
<td>80</td>
<td>0.80</td>
<td>4</td>
<td>8/2</td>
</tr>
<tr>
<td>90</td>
<td>0.90</td>
<td>9</td>
<td>9/1</td>
</tr>
<tr>
<td>100</td>
<td>1.00</td>
<td>Infinity</td>
<td>10/0</td>
</tr>
</tbody>
</table>

Table 1. Comparison of risk and odds
Box 1. Calculation of relative risk and odds ratio to illustrate the 'rare disease assumption'

<table>
<thead>
<tr>
<th></th>
<th>Mercury sphygmomanometers</th>
<th>Aneroid sphygmomanometers</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHS standard not achieved</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>BHS standard achieved</td>
<td>71</td>
<td>150</td>
</tr>
<tr>
<td>Total</td>
<td>75</td>
<td>191</td>
</tr>
</tbody>
</table>

Relative risk = (4/75)/(41/191) = 0.25; odds ratio = (4/71)/(41/150) = 0.21

Because the outcome of interest (devices failing) is ‘rare’, the risk of failure (number of devices failing divided by number of all devices) is close in value to the odds of failure (number of devices failing divided by number of devices not failing); the RR and OR will, therefore, be close in value. In epidemiology, this is called the ‘rare disease assumption’.

BHS = British Hypertension Society

should be described using RR or ORs. The debate stems from the premise that, if the RR is more intuitively understandable than an OR, then, even when baseline risks are high, readers may interpret ORs as risk ratios and, hence, reach an incorrect conclusion.13

How commonly the two are confused is illustrated by a study of papers published in two journals reporting that 26% of the articles that used an OR interpreted it as a risk ratio.9 It has even been argued that the effect may be exploited by healthcare researchers hoping to make their findings more dramatic and publishable.14 For example in a study comparing National Psoriasis Foundation members with non-members, an OR of 24 for having heard of calcipotriene was reported as ‘members (being) more than 20-fold more likely to have heard of the treatment, although the actual RR was 3.5.15

When should odds ratio not be interpreted as relative risk?

Altman and colleagues advise that the OR should not be interpreted as an approximate RR unless the events are rare in both groups, ‘say, less than 20–30%’.16 Misinterpretation of the OR as an approximation to the RR is particularly ill-advised when the event rate is high in only one group.16 A more conservative benchmark of 10% has been suggested as the threshold above which, ‘authors, reviewers, and editors should consider presenting and interpreting odds ratios with caution.3

The impact of an intervention [effect size] also affects the discrepancy between RR and OR [Figure 1]. Taking both factors into account, an article in Bandolier concluded that, ‘as both the prevalence [initial risk] and the OR increase, the error in the approximation quickly becomes unacceptable.’17 As ORs are often reported without the corresponding RR, various authors have offered guidance on when to expect a problematic degree of divergence. Davies asserted that:

‘So long as the event rate in both the intervention and control groups is less than 30%, and the effect size no more than moderate [say, a halving or doubling of risk], then interpreting an OR as a RR will overestimate the size of effect by less than one-fifth.’16,18

Holcomb et al suggested a rule of five:

‘If the OR is greater than 1 but no greater than 5, and the occurrence of the outcome in the unexposed group is no greater than 5%, then the OR exceeds the risk ratio by less than 20%.’

Does it matter?

In defence of ORs, Davies has also argued that misinterpretation of them does not materially affect qualitative judgements.19 To illustrate this point, it was argued that reporting specialist stroke units as reducing adverse outcomes by an OR of 0.66 (instead of by an RR of 0.81), at worst, suggested a reduction in poor outcomes of one-third

![Figure 1. Divergence between RR and OR with increase in baseline risk or effect size (risk ratio).](image-url)
Box 2. How odds ratios are the ideal statistical tool for case–control studies

The following data from a cross-sectional study show the prevalence of hay fever and eczema in 11-year-old children.22 Should we address the risk (or odds) of hay fever in children with eczema, or the risk (or odds) of eczema in children with hay fever?

<table>
<thead>
<tr>
<th>Hay fever</th>
<th>No hay fever</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eczema</td>
<td>141</td>
<td>420</td>
</tr>
<tr>
<td>No eczema</td>
<td>928</td>
<td>13 525</td>
</tr>
<tr>
<td>Total</td>
<td>1069</td>
<td>13 945</td>
</tr>
</tbody>
</table>

What is probability that a child with eczema will also have hay fever?

For 11-year olds with eczema, risk of hay fever is 141/14 526 (25%). For 11-year olds without eczema, risk of hay fever is 928/14 433 (6.4%).

RR = 25.1/6.4 = 3.9, so a child with eczema is 3.9 times as likely to have hay fever.

OR = 141/620/928/13 525 = 4.9, demonstrating 4.9 greater odds of hay fever if a child has eczema.

What is probability that a child with hay fever will also have eczema?

For 11-year olds with hay fever, risk of eczema is 141/1069 (13.2%). For 11-year olds without hay fever, risk of eczema is 420/13 945 (3%).

RR = 4.28, so a child with hay fever is 4.36 times as likely to have eczema.

OR = 141/928/420/13 525 = 4.9, demonstrating 4.9 greater odds of eczema if a child has hay fever.

Note that the RR depends on the choice of analysis, but the OR is always 4.9.

For more detail see Bland and Altman.² OR = odds ratio. RR = relative risk.

Can use of odds ratios in prospective studies be defended?

Some authors defend the OR over the RR.¹²,¹³ Sometimes, it is an arbitrary decision whether or not to treat an outcome, or its converse, as the ‘event’. For instance, trials can choose to report the RR of dying, or of surviving. Any such choice affects the RR and OR very differently. For example, in Schulman and colleagues’ study of cardiac catheterisation rates and ethnic group, the OR for referral in black versus white ethnic groups is 0.6, and the OR for non-referral is 1.7 (the reciprocal of 0.6).₂⁶ Had the authors reported the RR, they could have quoted an RR of 0.9 for referral or 1.6 for non-referral, which could be translated into a 10% lower referral rate, or 60% higher non-referral rate. Each is correct, yet gives a different impression of the magnitude of any association between ethnic group and referral. ORs circumvent this effect by accounting for ‘success’ or ‘failure’ symmetrically (demonstrated by the derivation in Box 2).²¹ However, we are not persuaded that this pleasing mathematical property is worth the risk of ORs being misinterpreted as an RR; in particular, in this case, the use of ORs did not prevent misreporting in the popular press.²²

WHEN ARE ODDS THE STATISTIC OF CHOICE?

There are two settings when it makes sense to use ORs: case–control studies and logistic regression.

In a cross-sectional study, it may be arbitrary which variable is the outcome and which is the exposure. The same OR is obtained either way, for reasons explained by Bland and Altman (Box 2);²³ however, this does not apply to the RR.

A special type of cross-sectional study is the case–control study, in which the percentage of cases does not reflect the risk (baseline prevalence) in the population, but the design chosen by the investigators. Simply varying the number of cases and controls included in the study would vary the value obtained for the calculated risks in the exposed and unexposed groups; these are, therefore, meaningless, as would be the RR. The odds in the exposed and unexposed groups are also meaningless; however, the OR is invariant to the number of subjects recruited, as well as to how we order the categories in the rows and columns.²⁴ It has also been argued for case–control studies that, as the outcome of interest is usually rare, the OR also gives a good approximation to the RR.²⁵ The advantage is instead of the true reduction of one-fifth, and that this was unlikely to affect the conclusion that stroke units improve outcomes and should be provided. Interestingly, this argument looks less secure now, in current cash-constrained times when cost efficacy and opportunity costs are seemingly at the forefront of most commissioning decisions. When the baseline event rate is high, interpretation of the OR as the RR results in a systematic and important underestimate of the number needed to treat.³⁰ Moreover, for clinicians to balance the good and bad outcomes of an intervention requires full understanding of the probability of each.³¹ Many authorities seem to agree that authors reporting on prospective studies should use the RR or risk reduction in preference to ORs.⁵,⁸,¹⁰,¹³

ADDRESS FOR CORRESPONDENCE

Christine A’Court, Department of Primary Care Health Sciences, University of Oxford, 23–38 Hythe Bridge Street, Oxford, OX1 2JF. E-mail: christine.acourt@btinternet.com

that, in this situation, the researcher benefits from the intuitive understanding offered by RR, but gets a value that is not unduly swayed by the study design. For instance, a GP with prior Guillain-Barre syndrome (GBS), trying to decide whether to be vaccinated against H1N1, identified a case–control study reporting an OR of 18.6 (95% CI = 7.5 to 66.4) for the development of GBS in the 2 months following influenza-like illness, and the contrasting finding of an OR of GBS developing after H1N1 vaccination of 0.16 (95% CI = 0.02 to 1.29).²³ An approximate 19-fold risk increase after native infection, compared with no increase or possibly reduced risk after vaccination, enabled a decision to be made.

The second reason for the widespread use of ORs is that, unlike risks, odds can take any value between zero and infinity. This mathematical property is not of interest in itself, but it makes possible logistic regression analyses. These allow researchers to examine the effect of multiple exposure variables on a dichotomous outcome, or the effect of one exposure variable while adjusting for potential confounders.¹³ ORs are the natural output of logistic regression software and are, therefore, widely quoted in the scientific literature.

For these reasons the OR, as well as the RR, will continue to be used in clinical research, and users of the medical literature will continue to need to be familiar with both.

Christine A’Court, Clinical Researcher, Department of Primary Care Health Sciences, University of Oxford, Oxford.

Richard Stevens, University Research Lecturer in Statistics, Department of Primary Care Health Sciences, University of Oxford, Oxford.

Carl Heneghan, Reader in Evidence Based Medicine, Department of Primary Care Health Sciences, University of Oxford, Oxford.

Provenance: Commissioned; externally peer reviewed.

©British Journal of General Practice

This is the full-length article (published online 27 Feb 2012) of an abridged version published in print. Cite this article as: Br J Gen Pract 2012; DOI: 10.3399/bjgp12X630223.
ratios should be used only in case-control studies and logistic regression analyses [letter]. BMJ 1998; 317(7166): 1155–1157.


