Methods

Implementing GRADE: calculating the risk difference from the baseline risk and the relative risk

10.1136/eb-2013-101340

Robert G Newcombe,1  Ralf Bender2

1Cochrane Institute of Primary Care & Public Health, Cardiff University, Cardiff, Wales, UK; 2Department of Medical Biometry, Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany

Correspondence to: Dr Robert G Newcombe, Cochrane Institute of Primary Care & Public Health, Cardiff University, Cardiff CF14 4YS, Wales, UK; Newcombe@cf.ac.uk

Abstract

A key step in implementing the GRADE (Grading of Recommendations Assessment, Development and Evaluation) system is the estimation of a risk difference based on estimates of the baseline risk and the relative risk estimated from different sources. In this paper we describe a simple and effective method to calculate confidence intervals (CIs) for the risk difference for this situation. Whenever an independent source is available to estimate the baseline risk for the population to which the effect estimates should be applied, this source should be used and CIs for the absolute risk difference should be calculated taking all sources of uncertainty into account.

Availability of high-quality estimates of the absolute difference in effectiveness between alternative treatment options is crucial to the application of evidence-based healthcare to populations of patients and corresponding decisions. One framework for assessing confidence in estimates of the effect of alternative management strategies on patient-relevant outcomes within the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system\(^1\) is summarised by Spencer et al.\(^2\) In this article we single out one of the five domains of the GRADE system, namely imprecision.

Often, the best available evidence for the absolute difference in effectiveness between a treatment under consideration and a standard regime does not come from a single study, but from two totally separate sources. Sometimes, an estimate of the relative risk (RR) of the outcome of interest between the two treatment options is available from a meta-analysis combining evidence from several randomised trials. Owing to the larger sample size available, this will in general have greater precision than an RR derived from a single study. In most contexts, estimates of relative effect of a therapy are more consistent across different baseline risks than absolute effect estimates.\(^3\) Consequently, it is a common practice in systematic reviews to report a pooled estimate of the RR, rather than the absolute risk difference (RD).\(^2\)

To convert an RR into an absolute RD, we also require an estimate of the baseline risk (BR), the rate of occurrence of the event of interest when the standard treatment is used. The absolute RD is then calculated from the BR and RR using the formula RD = BRx(RR−1).

In most applications, the RR is below 1, representing a reduction in risk due to the intervention. The calculated RD is then negative. Sometimes, the RR may be greater than 1, representing an increase in risk due to the intervention. The calculated RD is then positive.

Spencer et al.\(^5\) noted that the calculations currently performed under the GRADE framework take into account the imprecision of the RR estimate, but not that of the baseline risk estimate. They concluded that evaluating uncertainty in baseline risk, and its impact on confidence in absolute estimates of treatment effect, remained an important outstanding issue. The purpose of this article is to describe a simple, effective method that may be used to take both sources of uncertainty into account.

All the quantities we concerned to estimate, such as the BR, the RR or the RD, are derived from series of patients of finite size. A CI is normally used to display the resulting uncertainty of such an estimate. CIs convey information about magnitude and precision of effect simultaneously, keeping these two aspects of measurement closely linked.\(^4\)\(^5\) In the great majority of instances, researchers calculate 95% CIs, as a common metric to quantify sampling imprecision.

Confidence limits for the RD may be calculated from those for the BR and RR by a procedure method of Variance Estimates Recovery (MOVER). This is a general approach that may be used to calculate CIs for sums and differences of two independently estimated quantities. MOVER may be extended to apply to products or ratios, but greater care is required. Neither an approach using logarithms of BR and (RR−1)\(^6\) nor the version of MOVER developed specifically for ratios\(^7\) is guaranteed to yield meaningful results when the CI for the RR can span 1, for reasons explained by Newcombe.\(^8\) An enhanced version described by Newcombe\(^8\) is designed to accommodate datasets with RR and corresponding confidence limits below 1, representing benefit, as well as above 1, representing harm.

While it is simple to calculate the RD from the BR and RR, the formulae to derive confidence limits for the RD from those of the BR and RR are quite complicated.\(^2\)\(^8\) However, there is no need for the user to perform any part of the calculations. An Excel spreadsheet ‘RD from BR and RR.xls’ to perform all the calculations described in this article is freely downloadable from: http://medicine.cf.ac.uk/primary-care-public-health/resources/

The calculations in the spreadsheet start with estimates of the BR and RR and the corresponding CIs. The RD together with its CI is then derived using the 95% confidence limits for the RD may be calculated from those for the BR and RR, the formulae to derive confidence limits for the RD from those of the BR and RR are quite complicated.\(^2\)\(^8\) However, there is no need for the user to perform any part of the calculations. An Excel spreadsheet ‘RD from BR and RR.xls’ to perform all the calculations described in this article is freely downloadable from: http://medicine.cf.ac.uk/primary-care-public-health/resources/

An example where the intervention is beneficial

Spencer et al.\(^2\) reported calculations taken from Bates et al.\(^9\) relating to the use of low-dose, low-molecular-weight heparin (LMWH) to prevent venous thromboembolic events in women undergoing assisted reproduction who develop severe ovarian hyperstimulation syndrome. The RR used here is 0.36 (95% CI 0.20 to 0.67), taken from a meta-analysis. The baseline risk is 2/49=0.041, with 95% CI 0.011 to 0.137 calculated with the Wilson score method\(^10\) taken from a small Czech study.\(^11\) Bates et al.\(^9\) combined the RR and BR here to give RD=0.041x(0.36−1)=−0.026, indicating that use of low-dose LMWH is estimated to prevent 26 venous thromboembolic events per 1000 patients treated. Bates et al.\(^9\) reported 95% confidence limits of 13 to 32 events prevented. However, these limits take account of the uncertainty of the RR only. Spencer et al.\(^2\) observed that the uncertainty of the BR should also be taken into account here, and reported a CI of 4 to 110 for the number of events prevented per 1000 women. These figures are obtained by directly combining the 95% confidence limits for the RR and BR, which leads to an unnecessarily wide interval. The correct 95% CI calculated by MOVER-R is −0.089 to −0.006, indicating prevention of between 6 and 89 events per 1000 women.

We could equally well construct a CI representing the uncertainty of the BR only, −0.088 to −0.007 here. Figure 1 displays all four 95% CIs for the RD, expressed as a risk reduction. In this example, the correct
Mover-R interval.
ing the uncertainty of the RR only that approximates closely the
is the dominant source of uncertainty, hence it is the interval represent-
pared with the use of the conventional dose, with 95% CI 15% to 55%.
need for analgesic supplementation in an additional 30% of women com-
0.546. These results indicate that using low-dose bupivacaine leads to the
12.5%. The resulting RD is 0.109×(3.76-
line risk of needing analgesic supplementation during surgery was taken
receiving the low dose compared with the conventional dose. The base-
thetic ef
hypotension resulting from spinal anaesthesia, it may compromise anaes-
ara low dose instead of the conventional dose (>8 mg) may help prevent
facy. In a meta-analysis, the need for analgesic supplementation
during surgery was higher (RR=3.76, 95% CI 2.38 to 5.92) in women
0.368, with 95% CI15 0.209 to 0.649. The RR representing the
cereas of patients experiencing the event of interest in the intervention group
minus the corresponding proportion for the control group. A CI for this RD
hospital re-evaluating the use of low-dose bupivacaine (≤8 mg) in spinal anae.
Arzola and Wieczorek 12 evaluated the use of low-dose bupivacaine
An example where the intervention increases risk
Arzola and Wieczorek12 evaluated the use of low-dose bupivacaine
(≤8 mg) in spinal anaesthesia for elective caesarean section. While use of a
low dose instead of the conventional dose (>8 mg) may help prevent
hypotension resulting from spinal anaesthesia, it may compromise anaes-
thetic efficacy. In a meta-analysis, the need for analgesic supplementation
during surgery was taken as 10.9% based on Garry and Davies. 13 This is derived from 175 occur-
ences among 1610 women, leading to a 95% CI for BR of 9.4% to
12.5%. The resulting RD is 0.109×(3.76–1)=0.301, with 95% CI 0.149 to 0.546. These results indicate that using low-dose bupivacaine leads to the
need for analgesic supplementation in an additional 30% of women com-
pared with the use of the conventional dose, with 95% CI 15% to 55%.

Figure 2 displays four 95% CIs for the RD, as in figure 1. Here, the RR
is the dominant source of uncertainty, hence it is the interval represent-
ing the uncertainty of the RR only that approximates closely the
Mover-R interval.

Figure 1 Absolute risk reduction (expressed as events prevented per 1000 women) for effect of low-dose, low-molecular-weight heparin on venous thromboembolic events, from Bates et al.9 Four 95% CIs are shown, (A) representing imprecision of relative risk (RR) only, (B) imprecision of baseline risk (BR) only, (C) a correct interval using Mover-R and (d) calculated directly from upper and lower limits for both BR and RR.

Mover-R interval is very similar to the interval derived from that for the
BR. Here, the BR is the dominant source of imprecision.

An example where the intervention increases risk
Arzola and Wieczorek12 evaluated the use of low-dose bupivacaine (≤8 mg) in spinal anaesthesia for elective caesarean section. While use of a low dose instead of the conventional dose (>8 mg) may help prevent hypotension resulting from spinal anaesthesia, it may compromise anaesthetic efficacy. In a meta-analysis, the need for analgesic supplementation during surgery was taken as 10.9% based on Garry and Davies.13 This is derived from 175 occurrences among 1610 women, leading to a 95% CI for BR of 9.4% to 12.5%. The resulting RD is 0.109×(3.76–1)=0.301, with 95% CI 0.149 to 0.546. These results indicate that using low-dose bupivacaine leads to the need for analgesic supplementation in an additional 30% of women compared with the use of the conventional dose, with 95% CI 15% to 55%.

Figure 2 Absolute risk difference (in %) for effect of low-dose bupivacaine on spinal anaesthesia efficacy, from Arzola and Wieczorek.12 Four 95% CIs are shown in figure 1.

Discussion
Whenever the BR and RR are derived from separate studies and thus are
estimated independently, the calculations described here, based on
Mover-R, lead to an appropriate CI for the RD which correctly allows
for the degree of imprecision of both the BR and RR. As in figures 1 and
2, more simplistic approaches either reflect the imprecision of one param-
eter only, or else produce an interval that is, unnecessarily wide. In both
instances, the interval derived from the limits for both the BR and RR is
substantially wider than the correct Mover-R interval. Thus in the
LMWH example, it is implausible that the BR would be at its upper 95%
limit, 0.0137 and RR at its lower limit, 0.20, which is what Spencer’s upper limit of 110 represents. Versions of Mover that are less refined than the algorithm used here fail to give useful results when the RR and its confidence limits can be either side of 1.

In many applications, the RR is taken from a meta-analysis. However, the
method described here must not be used when the BR and RR are
derived from exactly the same series of individuals, because the assumption
that they are statistically independent is violated. In the situation of a single
study, the RD should be calculated directly from the data, as the proportion
of patients experiencing the event of interest in the intervention group
minus the corresponding proportion for the control group. A CI for this RD
is calculated using the second block of the spreadsheet CIPROPORTIONS.xls
available from the same website. An indirect procedure in which intervals
are first calculated for the BR and RR and then combined may lead to a very
different interval and is simply incorrect in this scenario.

For example, Rascol et al14 compared the incidence of dyskinesia
after ropinirole (17/179, 0.095) and levodopa (23/89, 0.258) in a prospective
study in early Parkinson’s disease. Here, the baseline risk (on levodo-
paa) is 0.258, with 95% CI10 0.179 to 0.358. The RR representing the
reduction of risk using ropinirole is 0.368, with 95% CI15 0.209 to 0.649.
The absolute reduction in risk is 0.163, calculated either directly or from the
BR and RR. However, the 95% CI for the risk reduction calculated incorrectly from the BR and RR is 0.082–0.242. This is substantially nar-
rower than the correct interval calculated directly,16 0.068–0.269. Conversely, for other datasets, the interval calculated indirectly by
Mover-R can be too wide.

In the context of a meta-analysis in which it makes sense to use RD as the effect measure, the RD should be estimated in each study and then
pooled using meta-analysis methods. One meta-analysis situation in
which no clear solution has yet been established is where the RD should
not be used as an effect measure due to heterogeneity, but the relative
effect measure, the RR can be pooled adequately and the BR is taken to
be the median or some other summary measure derived from the
observed absolute risks of the control group across the
same studies included in the meta-analysis. If there is little variation in the BR across
the studies this is the recommended procedure of the Cochrane
Collaboration.17 In this situation BR and RR are not estimated independ-
ently and Mover-R cannot be used. The solution used in Cochrane
reviews is to consider the BR (called assumed control risk (ACR) in
Cochrane reviews) as a fixed constant. However, the larger the uncer-
tainty of the estimated BR the greater is the need to include this uncer-
tainty in the CI for the RD.

In summary, a simple and effective method to calculate CIs for the
RD from independent estimates of the baseline risk and the RR is avail-
able. This method improves the currently used methods within the
GRADE system, because both sources of uncertainty, namely the estima-
tion of the RR as well as that of the BR are taken into account.

Competing interests None.

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/eb-2013-101340).

Methods
Methods

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/3.0/

References
**Spreadsheet MOVER-R.**

**Confidence interval for the ratio of two independently estimated quantities.**

This spreadsheet starts with confidence intervals for two quantities theta1 and theta2. It calculates a confidence interval for the ratio \( \theta_1/\theta_2 \) using the MOVER-R algorithm. Reference: Newcombe, Statistical Methods in Medical Research, 2013. The algorithm assumes that the quantities \( \theta_1 \) and \( \theta_2 \) are estimated independently. This will be true if the two estimates are derived from separate studies. Or if they come from two independent groups of individuals within the same study. For example, treatment groups A and B, or males and females.

The algorithm assumes that the same confidence level applies to the intervals for \( \theta_1 \) and \( \theta_2 \). This confidence level then applies to the ratio \( \theta_1/\theta_2 \).

The algorithm requires that \( \theta_2 \) and its confidence limits must be strictly positive. \( \theta_1 \) and its confidence limits may be positive or negative. \( \theta_1 \) and \( \theta_2 \) may represent continuous or binary data.

E.g. \( \theta_1 \) and \( \theta_2 \) could be means of the same strictly positive continuous variable in 2 different groups.

The algorithm may also be used to get an interval for the product of two quantities, \( \theta_1 \times \theta_3 \). This is done by letting \( \theta_2 = 1/\theta_3 \).

Let \( L_3 \) and \( U_3 \) denote the lower and upper limits for \( \theta_3 \). \( L_2 = 1/U_3 \) and \( U_2 = 1/L_3 \) are then used as the lower and upper limits for \( \theta_2 \). It is essential to input these limits in the correct order.

As an example derived from binary data with \( \theta_1 \) not necessarily positive:

Let \( \theta_1 \) denote the relative risk minus 1. If \( RR>1 \), \( \theta_1>0 \); if \( RR<1 \), \( \theta_1<0 \). Let \( \theta_2 \) the reciprocal of the baseline risk, from a separate study. Then \( \theta_1/\theta_2 \) represents the absolute risk difference.

The exemplar dataset represents this situation with BR 0.04 (0.016 to 0.1) and RR 1.3 (1.04-1.625). Intervals for variables derived from binary data should be calculated by boundary-respecting methods.

The algorithm checks data validity as follows. For \( \theta_2 \), all input values must be strictly positive. For \( \theta_1 \), they may be either positive or negative. Both variables must satisfy the conditions lower limit \( \leq \) estimate \( \leq \) upper limit. Error codes are returned for the confidence limits if these conditions are violated.

Depending on the context of application, the user should also check whether

(a) the input values - point estimates and confidence limits for \( \theta_1 \) and \( \theta_2 \)
(b) the calculated point estimate and confidence limits for \( \theta_1/\theta_2 \)

lie within the relevant ranges for validity.

To perform these calculations, replace values in **bold** as appropriate.

<table>
<thead>
<tr>
<th>Input data:</th>
<th>Estimate</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theta1</td>
<td>0.3</td>
<td>0.04</td>
<td>0.625</td>
</tr>
<tr>
<td>Theta2</td>
<td>25</td>
<td>10</td>
<td>62.5</td>
</tr>
</tbody>
</table>

Input data validity check: **TRUE**

Results:

\[
\theta_1/\theta_2 = 0.012000 \quad 0.001392 \quad 0.038514
\]
Spreadsheet RD from BR and RR.

This spreadsheet calculates the risk difference (RD) from the baseline risk (BR) and relative risk (RR). The BR and RR are assumed to be estimated independently. This will be true if the BR and RR are derived from separate studies. In many applications, the RR comes from a meta-analysis.

Confidence limits for the RD are then calculated from confidence limits for the BR and RR. The calculation uses the MOVER-R algorithm. Reference: Newcombe, Statistical Methods in Medical Research, 2013.

It is assumed that the same confidence level applies to the intervals for BR and RR. This confidence level then applies to the RD.

The relative risk will usually be below 1, representing a reduction in risk due to the intervention. The calculated risk difference is then negative. Sometimes, the relative risk may be greater than 1, representing an increase in risk due to the intervention. The calculated risk difference is then positive.

The spreadsheet checks data validity as follows. The BR must lie between 0 and 1. The RR and its lower and upper confidence limits may be below or above 1. Both the BR and RR must satisfy the conditions lower limit <= estimate <= upper limit. Error codes are displayed for the confidence limits if these conditions are violated. Entered confidence limits for BR below 0 or above 1 are truncated to 0 or 1.

Also, when BR and RR are large, the calculated RD and its confidence limits may be >=1. If this is a possibility, the RR based model may be inappropriate and an OR based model may be more appropriate. If the calculated RD >= 1, a warning is displayed and error codes are displayed for both confidence limits. Otherwise, a calculated upper limit >=1 is truncated and a warning displayed.

To perform these calculations, replace values in **bold** as appropriate.

<table>
<thead>
<tr>
<th>Input data:</th>
<th>Estimate</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline risk (BR)</td>
<td>0.041</td>
<td>0.011</td>
<td>0.137</td>
</tr>
<tr>
<td>Relative risk (RR)</td>
<td>0.36</td>
<td>0.2</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Input data validity check: **TRUE**

<table>
<thead>
<tr>
<th>Results:</th>
<th>Estimate</th>
<th>Lower limit</th>
<th>Upper limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk difference (RD)</td>
<td>-0.0262</td>
<td>-0.0888</td>
<td>-0.0058</td>
</tr>
</tbody>
</table>
This spreadsheet calculates the risk difference (RD) from the baseline risk (BR) and relative risk (RR). Sometimes, the relative risk may be greater than 1, representing an increase in risk due to the intervention. If this is a possibility, the RR based model may be inappropriate and an OR based model may be preferable. If the calculated RD >= 1, a warning is displayed and error codes are displayed for both confidence limits.