

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) system¹ 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 $RD = BR \times (RR - 1)$.

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.*² 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 called 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 ratios⁷ 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 Newcombe⁸ 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.⁷⁻⁸ 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 from these figures. If 95% CIs are used for the BR and RR, the resulting CI for the RD is also a 95% CI. The spreadsheet is designed to be a highly user-friendly resource, though needless to say, great care is needed with the negative numbers used to represent benefits.

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 method¹⁰ taken from a small Czech study.¹¹ Bates *et al.*⁹ combined the RR and BR here to give $RD = 0.041 \times (0.36 - 1) = -0.026$, 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

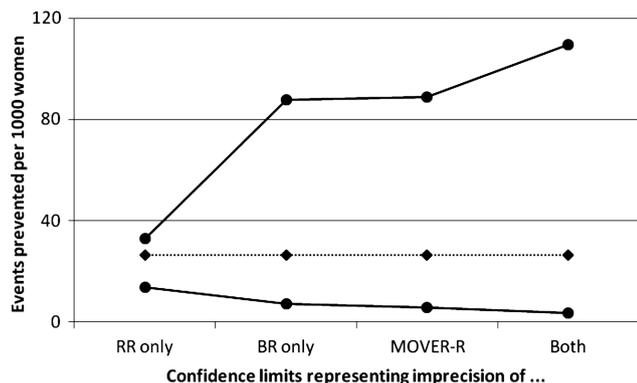


Figure 1 Absolute risk reduction (expressed as events prevented per 1000 women) for effect of low-dose, low-molecular-weight heparin on venous thrombotic events, from Bates *et al.*⁹ Four 95% CIs are shown, (A) representing imprecision of relative risk (RR) only, (B) imprecision of baseline risk (BR) 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 higher (RR=3.76, 95% CI 2.38 to 5.92) in women receiving the low dose compared with the conventional dose. The baseline risk of needing 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 $0.109 \times (3.76 - 1) = 0.301$, 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 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 representing the uncertainty of the RR only that approximates closely the MOVER-R interval.

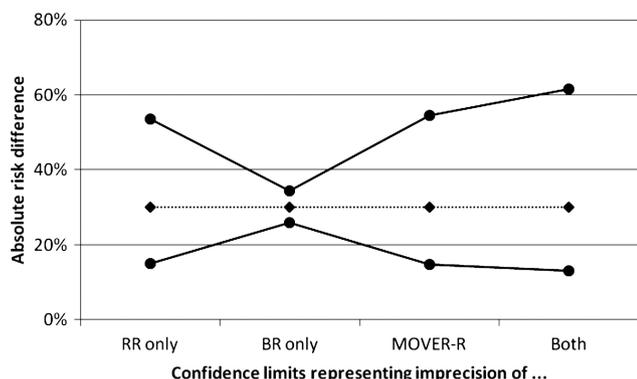


Figure 2 Absolute risk difference (in %) for effect of low-dose bupivacaine on spinal anaesthesia efficacy, from Arzola and Wieczorek.¹² Four 95% CIs are shown in figure 1.

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 parameter 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 *CIPROPORTION.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 al.*¹⁴ 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 levodopa) 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 narrower 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 independently 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 uncertainty of the estimated BR the greater is the need to include this uncertainty 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 available. This method improves the currently used methods within the GRADE system, because both sources of uncertainty, namely the estimation 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>).

Evid Based Med: first published as 10.1136/eb-2013-101340 on 22 August 2013. Downloaded from <http://ebm.bmj.com/> on April 26, 2024 by guest. Protected by copyright.

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/>



OPEN ACCESS



References

- Guyatt GH, Oxman AD, Schünemann HJ, *et al.* GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *J Clin Epidemiol* 2011;64:380–2.
- Spencer FA, Iorio A, You J, *et al.* Uncertainties in baseline risk estimates and confidence in treatment effects. *Br Med J* 2012;345:e7401.
- Deeks JJ. Issues in selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575–600.
- Gardner MJ, Altman DG. Confidence intervals rather than P values: estimating rather than hypothesis testing. *Br Med J* 1986;292:746–50.
- Rothman KJ, Greenland S. *Modern epidemiology*. 2nd edn. Philadelphia: Lippincott-Raven, 1998.
- Zou GY, Donner A. Construction of confidence limits about effect measures: a general approach. *Stat Med* 2008;27:1693–702.
- Donner A, Zou GY. Closed-form confidence intervals for functions of the normal mean and standard deviation. *Stat Methods Med Res* 2012;21: 347–59.
- Newcombe RG. MOVER-R confidence intervals for ratios and products of two independently estimated quantities. *Stat Methods Med Res* 2013. In press.
- Bates SM, Greer IA, Veenstra D, *et al.* Venous thromboembolism, thrombophilia, antithrombotic therapy, and pregnancy. ACCP Evidence-based Clinical Practice Guidelines, 9th Edition. *Chest* 2012;141(2_Suppl):e691S–736S.
- Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927;22:209–12.
- Mara M, Koryntova D, Rezabek K, *et al.* [Thromboembolic complications in patients undergoing in vitro fertilization: retrospective clinical study.] *Ceska Gynekol* 2004;69:312–16.
- Arzola C, Wiecek PM. Efficacy of low-dose bupivacaine in spinal anaesthesia for caesarean delivery: systematic review and meta-analysis. *Br J Anaesth* 2011;107:308–18.
- Garry M, Davies S. Failure of regional blockade for caesarean section. *Int J Obstet Anaesth* 2002;11:9–12.
- Rascol O, Brooks D, Korczyn AD, *et al.* A five year study of the incidence of dyskinesia in patients with early Parkinson's disease who were treated with ropinirole or levodopa. *N Engl J Med* 2000;342:1484–91.
- Miettinen O, Nurminen M. Comparative analysis of two rates. *Stat Med* 1985;4:213–26.
- Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat Med* 1998;17:873–90.
- Schünemann HJ, Oxman AD, Higgins JPT, *et al.* Chapter 11: presenting results and 'summary of findings' tables. In: Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions version 5.1.0 (updated March 2011)*. The Cochrane Collaboration, 2011. <http://www.cochrane-handbook.org>

Spreadsheet MOVER-R.

Confidence interval for the ratio of two independently estimated quantities.

This spreadsheet starts with confidence intervals for two quantities θ_1 and θ_2 . It calculates a confidence interval for the ratio θ_1/θ_2 using the MOVER-R algorithm. Reference: Newcombe, Statistical Methods in Medical Research, 2013. The algorithm assumes that the quantities θ_1 and θ_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 θ_1 and θ_2 . This confidence level then applies to the ratio θ_1/θ_2 .

The algorithm requires that θ_2 and its confidence limits must be strictly positive. θ_1 and its confidence limits may be positive or negative. θ_1 and θ_2 may represent continuous or binary data.

E.g. θ_1 and θ_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, $\theta_1 * \theta_3$. This is done by letting $\theta_2 = 1/\theta_3$. Let L_3 and U_3 denote the lower and upper limits for θ_3 . $L_2 = 1/U_3$ and $U_2 = 1/L_3$ are then used as the lower and upper limits for θ_2 . It is essential to input these limits in the correct order. As an example derived from binary data with θ_1 not necessarily positive: Let θ_1 denote the relative risk minus 1. If $RR > 1$, $\theta_1 > 0$; if $RR < 1$, $\theta_1 < 0$. Let θ_2 the reciprocal of the baseline risk, from a separate study. Then θ_1/θ_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 θ_2 , all input values must be strictly positive. For θ_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 θ_1 and θ_2 (b) the calculated point estimate and confidence limits for θ_1/θ_2 lie within the relevant ranges for validity.

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

Input data:	Estimate	Lower limit	Upper limit
Theta1	0.3	0.04	0.625
Theta2	25	10	62.5

Input data validity check	TRUE
---------------------------	------

Results:			
Theta1/Theta2	0.012000	0.001392	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 \leq estimate \leq 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 preferred. 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.

Input data:	Estimate	Lower limit	Upper limit
Baseline risk (BR)	0.041	0.011	0.137
Relative risk (RR)	0.36	0.2	0.67

Input data validity check	TRUE
---------------------------	------

Results:	Estimate	Lower limit	Upper limit
Risk difference (RD)	-0.0262	-0.0888	-0.0058

RR).

vention.

preferable.
limits.