

The (mis)use of overlap of confidence intervals to assess effect modification

Mirjam J. Knol · Wiebe R. Pestman ·
Diederick E. Grobbee

Received: 19 August 2010 / Accepted: 4 March 2011 / Published online: 19 March 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

In randomized controlled trials as well as in observational studies, researchers are often interested in effects of treatment or exposure in different subgroups, i.e. effect modification [1, 2]. There are several methods to assess effect modification and the debate on which method is best is still ongoing [2–5]. In this article we focus on an invalid method to assess effect modification, which is often used in articles in health sciences journals [6], namely concluding that there is no effect modification if the confidence intervals of the subgroups are overlapping [7–9].

When assessing effect modification by looking at overlap of the 95% confidence intervals in subgroups, a type 1 error probability of 0.05 is often mistakenly assumed. In other words, if the confidence intervals are overlapping, the difference in effect estimates between the two subgroups is judged to be statistically insignificant. By using mathematical derivation, we calculated that the chance of finding non-overlapping 95% confidence intervals under the null hypothesis is 0.0056 if the variance of both effect estimates is equal and the effect estimates are independent (see Supplementary material for derivation of this probability). If the variance of the effect estimates is not equal, the chance of finding non-overlapping 95% confidence intervals can be calculated by taking into account ρ , i.e. the ratio between the standard deviations in the subgroups, σ_2/σ_1 (Supplementary material, formula (3)). Figure. 1 shows the

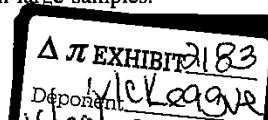
relation between ρ and the type 1 error probability if the effect estimates are independent. If the effect estimates are not independent, the correlation coefficient between the effect estimates can also be accounted for (Supplementary material, formula (3)).

To arrive at a type 1 error probability of 0.05, 83.4% confidence intervals should be calculated around the effect estimates in subgroups if the variance is equal and the effect estimates are independent (see Supplementary material for derivation of this percentage). If the variance is not equal, ρ should be taken into account (Supplementary material, formula (11)). Figure. 2 shows the relation between ρ and the level of the confidence interval. If the effect estimates are not independent, the correlation coefficient should be taken into account (Supplementary material, formula (11)). Adapting the level of the confidence interval can be especially useful for graphical presentations, for example in meta-analyses [10]. However, it is necessary to explicitly and clearly state which percentage confidence interval is calculated and its meaning should be thoroughly explained to the reader. Many readers will still interpret this ‘new’ confidence interval as if it were a 95% confidence interval, because this percentage is so commonly used. To prevent such confusion, other methods to assess effect modification could be used, such as calculating a 95% confidence interval around the difference in effect estimates [8].

The assumption used in the formulas presented in the appendices is that the effect estimators in the subgroups are normally distributed. Assuming that epidemiologic effect measures, such as the odds ratio, risk ratio, hazard ratio and risk difference, follow a normal distribution, the methods presented can also be used for these epidemiologic measures. Note that the assumption for normality is generally unreasonable in small samples, but a satisfactory approximation in large samples.

Electronic supplementary material The online version of this article (doi:10.1007/s10654-011-9563-8) contains supplementary material, which is available to authorized users.

M. J. Knol (✉) · W. R. Pestman · D. E. Grobbee
Julius Center for Health Sciences and Primary Care, University
Medical Center Utrecht, PO Box 85500, 3508 GA Utrecht,
The Netherlands
e-mail: m.j.knol@umcutrecht.nl



 Springer

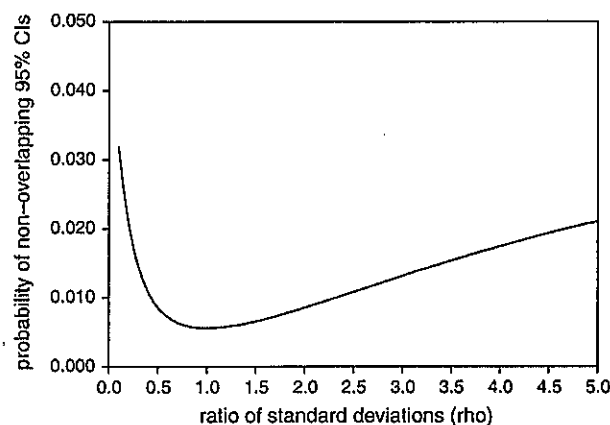


Fig. 1 Relation between ρ , which is the ratio of σ_2 and σ_1 , and the probability of non-overlapping confidence intervals under the null hypothesis (type I error)

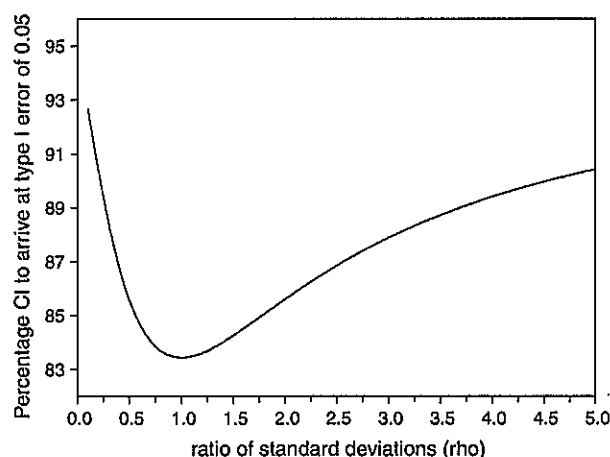


Fig. 2 Relation between ρ , which is the ratio of σ_2 and σ_1 , and the percentage confidence intervals to be calculated to arrive at a type I error probability of 0.05

Example

As an example, imagine a large randomized controlled trial that investigates the effect of some intervention on mortality and that includes 10,000 men and 5,000 women. Besides the main effect of treatment, the researchers are interested in assessing whether the treatment effect is different for men and women. Suppose that the risk ratio in men is 0.67 (95% CI: 0.59-0.75) and in women is 0.83 (95% CI: 0.71-0.98). The confidence intervals are partly overlapping, which the researchers may wrongly interpret as no effect modification by sex. Filling in formula (3)

(Supplementary material) results in a probability of non-overlapping 95% confidence intervals under the null hypothesis of 0.006. A confidence level of 83.8% could have been calculated to arrive at a type I error probability of 0.05, resulting in a confidence interval of 0.61-0.73 for men and 0.74-0.93 for women. Now, the confidence intervals do not overlap, so the p-value is at least smaller than 0.05, indicating statistically significant effect modification. Calculating the difference in risk ratios with a 95% confidence interval results in a ratio of risk ratios of 0.80 with a 95% confidence interval of 0.66-0.98, corresponding to a p-value of 0.028. This confirms our earlier observation of statistically significant effect modification.

Acknowledgement This study was performed in the context of the Escher project (T6-202), a project of the Dutch Top Institute Pharma.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Knol MJ, Egger M, Scott P, Geerlings MI, Vandembroucke JP. When one depends on the other: reporting of interaction in case-control and cohort studies. *Epidemiology*. 2009;20:161-6.
2. Wang R, Lagakos SW, Ware JH, Hunter DJ, Drazen JM. Statistics in medicine—reporting of subgroup analyses in clinical trials. *N Engl J Med*. 2007;357:2189-94.
3. Greenland S. Interactions in epidemiology: relevance, identification and estimation. *Epidemiology*. 2009;20:14-7.
4. Pocock SJ, Collier TJ, Dandreo KJ. Issues in the reporting of epidemiological studies: a survey of recent practice. *BMJ*. 2004;329:883.
5. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet*. 2000;355:1064-9.
6. Schenker N, Gentleman JF. On judging the significance of differences by examining the overlap between confidence intervals. *The Am Stat*. 2001;55:182-6.
7. Ryan GW, Leadbetter SD. On the misuse of confidence intervals for two means in testing for the significance of the difference between the means. *J Mod Appl Stat Methods*. 2002;1:473-8.
8. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ*. 2003;326:219.
9. Austin PC, Hux JE. A brief note on overlapping confidence intervals. *J Vasc Surg*. 2002;36:194-5.
10. Goldstein H, Healy MJR. The graphical presentation of a collection of means. *J R Stat Soc*. 1995;158:175-7.