Author's response to reviews

Title: Calculation of NNTs in RCTs with time-to-event outcomes: A literature review

Authors:

Mandy Hildebrandt (Mandy.Hildebrandt@iqwig.de)
Elke Vervoelgyi (Elke.Vervoelgyi@iqwig.de)
Ralf Bender (Ralf.Bender@iqwig.de)

Version: 4 Date: 3 December 2008

Author's response to reviews: see over
Response to the Advisor

03.12.2008

1) We added two examples (based upon studies 4 and 10 of Table 3) to demonstrate our calculations explicitly. The two outcomes in study 5 are uncorrelated (because they refer to different groups), whereas the two outcomes in study 10 are correlated (cardiovascular death or nonfatal MI and nonfatal MI alone measured in the same patients). However, there was no problem due to this correlation because we considered the two outcomes separately as it was done in the paper.

2) We think that our statements in the Methods section (page 7) are a sufficient explanation how we classified the NNT calculations as appropriate or inappropriate:

"… we considered a method as appropriate if the NNT was calculated either from survival probabilities estimated by means of the Kaplan-Meier method or the Cox regression model [14] or if it was calculated as the inverse of the hazard difference and both assumptions mentioned above are met (constant hazard difference and low event rates) [15,16,18]. When the method to calculate NNTs was not described in the article, we tried to verify the reported NNTs by recalculation from the presented data. When it was clear that an inappropriate method was used, the method was classified as "inappropriate", otherwise as "appropriate".

As explained in the Discussion section (page 10) this methods leads to an underestimation of the proportion of articles with inappropriate NNT calculation:

"Our finding that about 50% of the NNT-reporting articles with survival time data used inadequate calculation methods underlines the requirement to point out that special methods based on survival time techniques have to be used to calculate NNTs in this situation. This observed proportion probably underestimates the true proportion because we classified the method to calculate NNTs as "appropriate" if the method used was unclear and the reported NNT equalled the recalculated NNT from survival probabilities. It could be that in fact naive proportions have been used (i.e. an inappropriate method) but the result haphazardly equalled the correct result based upon survival probabilities. Thus, the true proportion of NNT-reporting articles with survival time data and inadequate calculation methods may be even higher than the observed proportion of about 50%.

We hope that our article has some impact on the peer review in major medical journals that more attention is given on the adequacy of NNT calculations.

We do not know exactly why only 34% of the NNT-reporting articles provide confidence intervals. One possible reason is given in the Discussion (page 12):

"Due to the unusual scale of NNTs their confidence intervals are difficult to describe if the effect is not significant [22]. This may be one reason why confidence intervals for the number needed to treat were given in one third of the investigated articles only (time-to-event and binary data)."
3) We have not extracted systematically all effect measures used in the non-NNT-reporting articles, because this is not the scope of our paper. Frequently used effect measures were the risk ratio and the odds ratio for binary data, the hazard ratio for survival time data and mean differences for continuous data.

An RCT “without NNT” is an RCT that reports the result without using NNT. We changed Figure 1 accordingly to make this clear.

4) We added one more disadvantage of the NNT, namely that NNT is undefined if ARR=0. A full description of the advantages and disadvantages of NNTs is beyond the scope of our paper. Detailed explanations and discussions can be found in the cited papers, which were in part included due to a former suggestion of one of the reviewers.