Reviewer's report

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

Version: 2 Date: 24 June 2008

Reviewer: Jacobus Lubsen

Reviewer's report:

Response to reviewers:

In point 1, the authors overlook that “the usual effect measure to quantify the impact of a treatment in terms of patient numbers that must be treated to avoid one event (regardless of treatment duration)” is an incomplete definition of an NNT as any NNT based on risk differences must always be defined for a certain duration of follow-up. In this sense, a NNT as the authors wish to define it is similar to a measure of risk. It is meaningless to say that the “risk of death is x%” without specifying the time interval to which the “x%” applies. This is true both for “binary” data and time-to-event data. The between-brackets addition “regardless of treatment duration” is highly confusing in this regard. In point 2, the authors state that “the reviewer equates by mistake follow-up time with treatment duration”. I would like to know where they got this from. They can certainly not have got this from my 2000 Lancet paper that is quoted in the manuscript as reference number 15. In fact, a major point made in that paper is that follow-up time cannot be equated to treatment duration. The example given then goes on to propose a simple and intuitive correction of NNTs based on hazard differences (and hence expressed in person-years of treatment) for the actual use of assigned trial treatment (or similar open treatment after discontinuation of assigned trial treatment) in each of the treatment groups compared. Published data from one of the SOLVD trials are used for this purpose. The authors should (re)read the Lancet paper concerned. Regrettably, the supplemental material to the Lancet paper that explains the calculations in detail is no longer on the Lancet’s website but I am happy to supply this. Point 3. The authors are correct in saying that the reciprocal of hazard differences and the reciprocal of risk differences are different effect measures because these measures have different dimensions (person-time as opposed to a dimensionless quantity). The statement that “only in the case of exponential survival times and rare events the hazard difference can be used to approximate risk difference” is a nonsense however, and is mistakenly based on the fact that for low rates the NNT in person-years and the NNT in persons evaluated after one year of follow-up approach each other numerically. One can always obtain risk differences based on hazard differences and vice versa, if the information needed to do so is available. The authors would be well advised to think these three points through based on simple numerical examples. Here is one:

See attachment for table
The table shows the expected cumulative number of deaths in two imaginary trial cohorts of 1000 patients each, based on exponential survival with a constant hazard. The control and intervention groups are assumed to have constant death rates (hazards) of 6 and 4 deaths per 100 person-years. Note that the NNT taken as the reciprocal of the hazard difference is 50 person-years of follow-up on intervention group care to postpone one death (which is not necessarily the same as 50 years of drug use - see Lancet paper), and that this NNT is numerically close the NNT taken as the reciprocal of the risk difference at one year (52.6 patients). Note also that NNTs based on risk differences depend strongly on the duration of follow-up.

General comments:

In the context of a literature review, the authors should not a priori distinguish between “appropriate” and “inappropriate” methods as is done now already in the background section (i.e. the introduction). Rather, the authors should list the various NNT estimation methods that have been used in the literature in neutral manner, and reserve any comments they may have on their appropriateness for the discussion section of the manuscript. The supplemental table should be part of the manuscript, and be arranged more efficiently. The table should have an additional column that specifies the condition concerned: e.g. “stable coronary artery disease” for EUROPA (first line in the table). The number of patients should be tabulated per group, and the number of endpoints should be given also. In a separate column, the method used to evaluate the NNT given should be identified. When a confidence interval was given, its width and level should be tabulated in addition to the NNT. The authors state in the results that they found 35 papers that reported NNTs for time-to-event outcomes, of which 17 applied “appropriate calculation methods” while the remaining 18 used “naive proportions”. In the supplemental material table a listing is given of the 35 papers concerned but the authors do not state which 17 of 35 reports used “appropriate calculation methods”. This information should be given in the supplemental table.

3

As “appropriate methods” the authors consider the two methods based on K-M curves evaluated at a fixed time point as described by Altman et al.. In addition, they consider an NNT based on a hazard difference appropriate “provided that the assumptions mentioned above are met”. The addition “provided...” is inappropriate in this context. Rather, the authors should include in the supplemental table any report that used this method, irrespective whether certain assumptions were met or not. Whether certain assumptions need to be met for interpretability of such an NNT is something that should be taken up in the discussion section. The authors insist (quite inappropriately) that survival times should be exponentially distributed but do not define exactly what is meant. Do you mean exponentially distributed with constant hazard? If you do, you are wrong because that is too rigid an assumption. It would be OK on the other hand to state that the hazard difference should be constant. Another method that the authors need to mention when the different methods are described is the use of naïve proportions, now mentioned only on page 8 of the manuscript as
“inappropriate”. This is important because this is perhaps the most frequently used method, and because the authors can not necessarily claim to know that in a paper included in the count of “using appropriate methods” the naïve proportion method wasn’t used. Consider for example the EUROPA trial listed in the first line of the supplemental material table as having a NNT of 50. The EUROPA report states indeed “We estimate that 50 patients need to be treated with perindopril for a period of 4 years, to prevent one major cardiovascular event” (c.f. last paragraph of discussion section). Other than this statement (which is reproduced in essentially the same words in the summary), the paper concerned does not mention how the authors came to the NNT estimate mentioned. The 10% and 8% of patients respectively who had the primary endpoint mentioned in the abstract and in Table 3 are NOT the event-probabilities at 4 years estimated by the Kaplan-Meier method, but simply the fractions of patients who reached the primary endpoint. Because the EUROPA report does not state how the NNT of 50 was derived and because this is the most frequent method of estimating NNTs for this type of trial, one can be 99.9999% sure that the authors took NNT as 1/[(603/6108) - (488/6110)], which equals 53. This was then rounded to 50 in the report. The authors need to explain how they came to the conclusion that the NNT reported was “appropriately” calculated. That in the case of EUROPA a risk difference at 4 years of about 2% can be guessed from the Figure 2 does not prove that the authors used Kaplan-Meier event-probabilities. It is therefore not possible to classify this paper among the 17 using an appropriate method, or among the remaining 18 of 35 NNT-reporting papers that used “naïve proportions”. Did they ask the EUROPA investigators how the NNT was derived? If yes, this should have been stated in the methods section. In the results section, it should have been stated how often the investigators concerned replied.

Manuscript:

Page 4, line 2 - 3: The definition of NNT is incomplete because no reference to time is included. Page 4, line 5: “highly controversial” is an opinion, not a statement of fact. Use a more neutral adjective please.

Page 4, line 9 a.f.: “risk communication” makes no sense here, do you mean “effect communication”? Page 4, line 17: in this line you explain what you mean by “binary”, i.e. “where the follow-up is the same for all patients”. You have used the term “binary” several times before. Add the explanation when you use the term for the first time. “Where the follow-up is the same for all patients” is not the correct terminology in this context. Change to: “where the occurrence of events is considered as a binary outcome for all patients during a fixed duration of follow-up and the time-to-event is ignored”, or something similar. Whether or not time-to-event data are used in the analysis has nothing to do with the question whether the duration of follow-up is fixed or variable. Page 4, 4th line from bottom a.f.: at this point you introduce “two methods” that have been proposed, and then you use a full page text to explain why you consider the first method as proposed by Altman et al. as the “appropriate” one. First of all, you overlook that Altman et al., apart from using raw data, in fact propose two related but different methods: using treatment group-specific K-M estimates of survival probabilities at specific
points in time, and using the hazard ratio to derive the survival probability at some time point in the intervention group based on the corresponding survival probability in the control group. The first method can be considered as entirely non-parametric with respect to the underlying survival distributions. The second is based on the assumption that the hazard ratio is constant over time. In a survey of this kind, you should distinguish between these two different methods, and report in how many papers the first method, and in how many the second, was used.

Pages 5 and 6. Most of this is opinionated and should be put in the discussion.

On page 7, the authors state “we also assessed whether confidence intervals were provided…”. As mentioned earlier, this is an important point not mentioned in the supplemental table. Whether or not confidence intervals were provided is less interesting than the question whether the information needed to calculate a confidence interval as discussed by Altman et al. could be found in the paper. Page 7 bottom, top page 8. It is inappropriate to classify a study that reported a NNT in patient-years based on hazard differences as “non-NNT-reporting” only because the authors did not describe the NNT concerned as “number of patients”. This is not just a matter of terminology! Table 3 describes a result. This table should therefore be mentioned in the results section. In the methods the authors should describe precisely how they compared published NNTs with “appropriately calculated NNTs”. Table 3 is of interest only if each if the 18 reports can be identified by the reader, and if the reasons for the differences are given.