

Using and Interpreting Adjusted NNT Measures in Biomedical Research

Ralf Bender^{1,2,*}

¹Department of Medical Biometry, Institute for Quality and Efficiency in Health Care (IQWiG), Cologne, Germany

²Faculty of Medicine, University of Cologne, Germany

Abstract: The number needed to treat (NNT) is a popular effect measure to present study results in biomedical research. NNTs were originally proposed to describe the absolute effect of a new treatment compared with a standard treatment or placebo in randomized controlled trials (RCTs) with binary outcome. The concept of the NNT measure has been applied to a number of other research areas involving the development of related measures and more sophisticated techniques to calculate and interpret NNT measures in biomedical research. In epidemiology and public health research an adequate adjustment for covariates is usually required leading to the application of adjusted NNT measures. An overview of the recent developments regarding adjustment of NNT measures is given. The use and interpretation of adjusted NNT measures is illustrated by means of examples from dentistry research.

Key Words: Number needed to treat, evidence-based medicine, confounding, adjustment for covariates, regression analysis.

INTRODUCTION

The number needed to treat (NNT) is a popular measure to describe the absolute effect of a new treatment compared with a standard treatment or placebo in randomized controlled trials (RCTs) with binary outcome [1,2]. The use of NNTs has been advocated in general medical journals in the last 20 years [3-7] as well as in a periodontal journal [8]. Additionally, the explanatory document of the Consolidated Standards of Reporting Trials (CONSORT) statement [9] exposes that NNTs are helpful for expressing the results of studies with binary and survival time outcomes. In spite of its widespread use, NNTs are frequently misused, incorrectly calculated, incompletely or misleadingly presented, and incorrectly interpreted [2,10,11]. There are various reasons for misleading applications of NNTs. Two major reasons are firstly, that the basic features of NNTs frequently are insufficiently understood and secondly, that simple standard methods for NNT calculation are applied in complex data situations in which more sophisticated methods are required. In this paper the basic issues required for adequate application of NNTs to present research findings are summarized. An overview of recent developments to estimate adjusted NNT measures in epidemiological and clinical trials is given. Examples from dentistry research are presented to illustrate the use and interpretation of adjusted NNTs. Additionally, other NNT-related effect measures, the so called impact numbers, are discussed.

BASIC ISSUES

The effect measure NNT is defined by the inverse of the difference between the risk r_c of an adverse outcome in the control group (CG) and the corresponding risk r_i of the inter-

vention group (IG), i.e. $NNT = 1/(r_c - r_i)$. NNT describes the expected number of patients that must be treated to prevent an event in one patient within a specific period of time. For example, the effect of an oral health program (OHP) on caries in children can be described in terms of NNT as follows. If a randomized controlled trial (RCT) is performed in which the OHP is applied to the IG, whereas the CG receives a conventional program and after, say, 5 years 50 of 500 children have caries in the IG ($r_i=0.1$) and 125 of 500 children have caries in the CG ($r_c=0.25$), then the NNT is given by $NNT = 1/(0.25 - 0.1) = 6.7$ (95% confidence interval 5.1 to 9.7).

How can this statistical result be presented in words? An incomplete, potentially misleading presentation is given by a statement such as "The number needed to treat was 7 for the OHP group". The basic information required to allow an adequate interpretation of NNT values is given by the alternative treatment to which the considered intervention is compared, the follow-up period, the outcome, the direction of the effect, and an appropriate confidence interval (CI). Rough rounding up to the next integer – although frequently recommended and used in the medical literature – should be avoided for low NNTs. Stang *et al.* [11] proposed that NNTs from 1 to 100 should be reported to at least one decimal place.

An adequate presentation of the result described above is given as follows. "On average, 6 to 7 children must receive the OHP to avoid one case of caries within 5 years compared to the conventional program. Due to estimation uncertainty the NNT may also lie between 5 and 10 children receiving OHP to prevent caries within 5 years in one additional child compared to the conventional program."

These results are obtained by applying simple standard methods which are appropriate in RCTs with individual randomization, two parallel groups, fixed follow-up time, binary outcome and sufficient sample size. In other situations in

*Address correspondence to this author at the Department of Medical Biometry, Institute for Quality and Efficiency in Health Care (IQWiG), Dillenburger Str. 27, D-51105 Cologne, Germany; Tel: +49 221 35685-451; Fax: +49 221 35685-891; E-mail: Ralf.Bender@iqwig.de

which clustered data, time-to-event outcomes or confounding play a role, more complex methods are required to estimate NNTs appropriately. In the following, we focus attention on application of adjusted NNTs which allow the consideration of important confounders in epidemiology as well as accounting for balanced covariates and covariate treatment interactions in RCTs.

NNT WITH ADJUSTMENT FOR COVARIATES

Methods to Adjust for Covariates

Besides randomized controlled trials the number needed to treat is also used in epidemiology and public health research. As the term "number needed to treat" makes no sense if the explanatory factor is an exposure rather than a treatment, the terms number needed to be exposed (NNE) [12,13] and exposure impact number (EIN) [13,14] have been proposed to apply the NNT concept in epidemiological studies. Regardless of terminology, in the simplest case NNT measures (NNT, NNE, EIN) are calculated by taking the reciprocal of the difference of two risks given by a 2x2 table. The use of simple 2x2 tables may be appropriate in RCTs. However, in observational studies covariates usually have to be taken into account to minimize bias due to confounding.

Within the framework of logistic regression a method was recently derived to perform point and interval estimation of NNT measures with adjustment for confounding by using the so called average risk difference (ARD) approach [13]. The main principle of this approach is given by averaging of the risk differences of all individuals of an appropriate (sub-) population taking the distribution of the confounders into account. Adjusted NNT measures are obtained by inverting the corresponding ARD. Technical details including methods to calculate confidence intervals for ARDs and NNTs can be found elsewhere [13]. The ARD approach to perform point and interval estimates of NNTs with adjustment for covariates can also be applied within the framework of the Cox regression model to analyze time-to-event data [15,16].

Application and Interpretation

The choice of the appropriate population over which the averaging of risk differences is performed depends on the research question and the study design [17,18]. In the context of cohort studies investigating the effect of exposures, averaging is performed separately over the unexposed or the exposed person leading to two different NNT measures [13,17]. In the first case the effect of allocating the exposure to unexposed persons (NNE) and in the second the effect of removing the exposure from exposed persons (EIN) is described. In the case of equal distributions of the covariates NNE and EIN are identical. However, usually the distributions of the covariates are different between the unexposed and exposed persons in the context of cohort studies leading to different values for NNE and EIN.

In the context of clinical trials it makes sense to average risk differences over the whole sample which leads to one unique adjusted NNT. This adjusted NNT describes the average effect of moving all patients from untreated to treated [18]. As in epidemiological studies, this concept allows the adjustment for potential confounding also in non-randomized clinical trials. In randomized controlled trials with adequate

randomization, in which the covariates are balanced, the application of the ARD approach leads to a gain in estimation precision concerning adjusted risk differences and NNTs so that the corresponding confidence intervals are shorter [19].

In summary, depending on the research question and the study design, different adjusted NNT measures should be applied. In the context of cohort studies NNE describes the average effect of allocating an exposure to unexposed persons, whereas EIN describes the average effect of removing the exposure from exposed persons. In the context of clinical trials (randomized or non-randomized) NNT describes the average treatment effect in the whole population of patients.

EXAMPLES

In order to illustrate the use and interpretation of adjusted NNTs two examples from dentistry research are considered. The first example was chosen to show the drawbacks of a naive and incorrect use of NNTs. In the second example it was possible to reconstruct the original individual data from the information given in the article so that own calculations could be performed to illustrate how adjusted NNTs can be used to describe absolute treatment effects appropriately in a complex data situation.

School-Based Education and Oral Cleanliness

The short-term effect of a school-based educational program on oral cleanliness was evaluated by means of a cluster randomized trial and described in terms of NNT [20]. In short, 15 year old students at public schools in Teheran, Iran, were cluster-randomized to the control group (n=130) or one of two oral health intervention groups, a leaflet group (n=148) and a videotape group (n=139) and outcomes were evaluated after 12 weeks. For illustration, we consider only the control and videotape groups and the outcome improvement in oral cleanliness (IOC). For statistical analyses paired and unpaired t-tests and the chi-square test were used; NNT was calculated as inverse of the absolute risk reduction. The result concerning IOC was presented as "... improvement of oral cleanliness occurred ... in 37% ($p < 0.001$) in the videotape group, and in 10% in the control group ... NNT was ... three in the videotape group." Additionally, the proportions and NNTs (both rounded to integers) were given in a table separately for boys (NNT = 2) and girls (NNT = 10).

Unfortunately, this is one example where NNTs are incorrectly calculated and presented for the following reasons. 1) Incorrect statistical tests were applied without accounting for the cluster randomization which may lead to spurious positive findings [21]. 2) There are obvious counting errors. For example, the proportions with IOC for girls are given as 18% in the videotape and 14% in the control group, which lead to $NNT \approx 25$, not to $NNT = 10$ as reported. 3) No confidence intervals for the estimated NNTs were presented. 4) Proportions and NNTs were rounded much too roughly. 5) Due to the different gender distributions in the groups and the different effect estimates for boys and girls the estimation of an adjusted NNT accounting for gender seems to be preferable to describe the overall average treatment effect. 6) To test whether the treatment effect is significantly different for boys and girls, an appropriate interaction test is required. 7) Only if the interaction test is statistically significant, the conclusion that "Boys in the videotape group showed more

Table 1. Results for the Main Outcome Dental Caries at 5 Years of Age in MS Positive Children (Hypothetical Data Based Upon the Finish Oral Health Program Study [22])

		White Collar			Blue Collar			Total		
		Dental Caries		Total	Dental Caries		Total	Dental Caries		Total
		Yes	No		Yes	No		Yes	No	
OHP	Yes	12	75	87	75	105	180	87	180	267
	No	48	63	111	57	96	153	105	159	264
Total		60	138	198	132	201	333	339	192	531

improvement ... than girls" is valid. In this case, the presentation of different effect estimates for boys and girls is adequate. The best method for data analysis is given by multiple logistic regression with appropriate interaction term and application of the ARD approach. 8) An appropriate explanation of the estimated NNTs is useful but was lacking in the considered example.

Consequently, the presented NNTs of the school-based intervention are misleading in this example because the estimates – at least in part – are too low, i.e., the corresponding reported absolute treatment effects are erroneously too large, and because no information is given about the estimation uncertainty. Additionally, the reported *p*-values are too low due to the application of an invalid statistical method.

Oral Health Program and Preschool Dental Caries

The preventive effect of a risk-based oral health program (OHP) in comparison with a traditional program on occurrence of dental caries was evaluated in a prospective controlled study of Finnish children followed from 18 months to 5 years of age [22]. The study reported a protective effect of OHP in white-collar families and NNTs were applied to present study results. The data set contains an interesting covariate treatment effect. Unfortunately, the study power was too low to show a significant overall average treatment effect. For illustrative purposes the original data were tripled to increase study power. Although based upon real data, the following results are hypothetical because the amount of data was artificially increased.

The intervention was targeted to mutans streptococci (MS) positive children. Only MS positive children are considered in the following for simplification. A complete analysis of all data would require the consideration of additional covariates. After triplication, $n=531$ MS positive "children" were obtained, $n_{IG}=267$ in the intervention (OHP) and $n_{CG}=264$ in the control group (traditional program). An important covariate in this example is given by the occupation of caretakers (blue collar vs. white collar). The results for the main outcome dental caries with stratification for occupation are given in Table 1.

By means of multiple logistic regression containing treatment, occupation and the corresponding interaction term it can be shown that there is a statistically significant interaction between treatment and occupation ($p < 0.001$). Therefore, the effect of OHP is different between children from white and blue collar families. Nevertheless, the overall av-

erage treatment effect is of interest accounting for the distribution of the caretakers' occupation.

The natural relative effect measure in logistic regression is given by the odds ratio (OR). In the case of no interaction the odds ratio is simply given by $exp(b)$, where *b* is the estimated regression coefficient. Methods to calculate odds ratios in the case of interactions are described for example by Hosmer and Lemeshow [23]. The OHP effects on dental caries at 5 years in terms of odds ratios in this example are given by $OR = 1.20$ (95% CI 0.77 to 1.87) for blue collar families and $OR = 0.21$ (95% CI 0.10 to 0.43) for white collar families demonstrating clearly different relative treatment effects in dependence on occupation.

In applications of NNTs in biomedical research a frequent problem is given by the fact that – although adjusted ORs are estimated and presented – crude naive NNTs based upon simple standard methods are calculated [13]. If we neglect the occupation of caretakers and estimate the NNT based upon the 2x2 table of white and blue collar caretakers together by means of standard methods the result $NNT = 13.9$ is obtained. However, the chi-square test yields a not significant result ($p = 0.085$) and the confidence region for NNT includes infinity (i.e. the zero effect). This result can be presented as $NNTB = 13.9$ (95% CI: $NNTB 6.5$ to ∞ to $NNTH 103.8$), where *NNTB* and *NNTH* mean number needed to treat for one patient to benefit or to be harmed, respectively, to indicate the direction of the effect [24,25]. However, the crude NNT estimation is inefficient and potentially biased because the covariate occupation is not taken into account.

We now consider how adjusted NNTs can be used to describe the absolute treatment effect of OHP adequately. If we are interested in the overall average effect of OHP in the population of MS positive children taking the distribution of the caretakers' occupation into account an adequate approach is given by an adaptation of the ARD approach allowing a covariate treatment interaction, here the interaction between occupation and OHP. This approach yields the result $NNTB = 12.2$ (95% CI: 6.2 to 331.2, $p = 0.042$), i.e. a statistically significant overall beneficial treatment effect. This result means that, on average, 12 to 13 MS positive children from a population with a distribution of occupation as in the considered sample are needed to receive the OHP to have one case of dental caries at age 5 years less compared to the traditional program. Due to estimation uncertainty NNT may also lie between 6 and 331 MS positive children receiving

OHP to prevent dental caries at age 5 years in one additional child compared to the traditional program.

Due to the interaction between occupation and OHP it is natural to also estimate the treatment effect separately for white and blue collar caretakers. These analyses make use of the two 2x2 tables shown in Table 1. By means of standard methods the following results are obtained: white collar caretakers $NNTB = 3.4$ (95% CI: 2.4 to 5.6, $p < 0.001$), blue collar caretakers $NNTB = 22.7$ (95% CI: $NNTB = 6.7$ to ∞ to $NNTB = 16.4$, $p = 0.412$). These results mean firstly that, on average, 3 to 4 MS positive children from white collar families are needed to receive the OHP to have one case of dental caries at age 5 years less compared to the traditional program. Due to estimation uncertainty NNT may also lie between 2 and 6 MS positive children from white collar families receiving OHP to prevent dental caries at age 5 years in one additional child compared to the traditional program. Secondly, a significant effect of OHP in MS positive children from blue collar families could not be found; the estimation uncertainty is quite large demonstrating that the effect of OHP in this group could not be reliably estimated and neither benefit, nor harm, nor a zero effect can be excluded. The large heterogeneity in the group of children from blue collar families indicates the existence of other unmeasured covariates which should be taken into account to yield more reliable estimates. In summary, a large and statistically significant preventive effect of OHP was found in MS positive children from white collar families, whereas the effect of OHP in MS positive children from blue collar families is still unknown.

CONCLUSIONS AND FINAL REMARKS

In biomedical research, simple standard methods for NNT calculation are frequently applied also in complex data situations in which more sophisticated methods are required [11]. This is hard to understand, especially if adequate sophisticated methods are used to estimate relative effect measures, but simple invalid standard methods based upon 2x2 tables are applied for NNT estimation. In the case of clustered data, the corresponding correlations have to be taken into account, e.g., by application of generalized estimating equations (GEEs) or mixed models [8]. For the analysis of time-to-event outcomes the use of survival time methods is required. However, a recent systematic literature review of RCTs with parallel group design and individual randomization published in 4 major medical journals in the period 2003-2005 found that in the case of time-to-event outcomes inadequate methods to estimate NNTs have been used in 50% of articles presenting NNTs (17 of 34 articles) [10]. In studies in which the effects of covariates are taken into account to estimate relative effect measures such as adjusted odds ratios, these covariates should also be used to estimate adjusted NNTs [13]. In this paper, the use and interpretation of adjusted NNT measures was described and illustrated by means of examples.

Besides the number needed to treat several other absolute effect measures, the so called impact numbers have been proposed, especially for use in public health research [14, 26, 27]. The NNT, which represents a special case of the impact numbers, is given by the inverse of a risk difference. The effect measures population impact number (PIN), case im-

pact number (CIN) and exposed case impact number (ECIN) represent the reciprocals of the population risk difference (i.e. the difference of the risk in the whole population and the risk of the unexposed or untreated persons), the population attributable risk (PAR), and the attributable fraction among the exposed (AF_e), respectively. PIN, CIN, and ECIN describe the number of persons of the whole population (PIN), the number of cases (CIN), or the number of exposed cases (ECIN) among which one case is attributable to the exposure or treatment. In the example of the Finish OHP study, the application of these effect measures would allow to describe the effect of the complete program including screening step in the whole population, and not only the preventive effect of OHP in the MS positive children. Methods are available to calculate confidence intervals for impact numbers [28, 29]. However, methods to perform point and interval estimation of adjusted impact numbers accounting for covariates are currently not completely developed.

In summary, in data situations where the effects of covariates play an important role, the application of NNTs with adjustment for covariates is required to present study results in terms of NNTs. To describe treatment effects with a population perspective the new impact numbers can be used, however, methods to estimate adjusted impact numbers have to be developed.

ACKNOWLEDGEMENT

I thank Consuela Jakobi-Yniguez for editorial support.

REFERENCES

- [1] Bender R. Number needed to treat (NNT). In: Armitage P, Colton T, Eds. *Encyclopedia of Biostatistics*. Chichester: Wiley, 2005; 3752-61.
- [2] McAlister FA. The "number needed to treat" turns 20 – and continues to be used and misused. *Can Med Assoc J* 2008; 179: 549-53.
- [3] Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318: 1728-33.
- [4] Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. *J Clin Epidemiol* 1994; 47: 881-9.
- [5] Cook RJ, Sackett DL. The number needed to treat: A clinically useful measure of treatment effect. *BMJ* 1995; 310: 452-4.
- [6] Sackett DL. On some clinically useful measures of the effects of treatment. *Evid Based Med* 1996; 1: 37-8.
- [7] Nuovo J, Melnikow J, Chang D. Reporting number needed to treat and absolute risk reduction in randomized controlled trials. *JAMA* 2002; 287: 2813-4.
- [8] Greenstein G, Nunn ME. A method to enhance determining the clinical relevance of periodontal research data: Number needed to treat (NNT). *J Periodontol* 2004; 75: 620-4.
- [9] Altman DG, Schulz KF, Moher D, et al. For the CONSORT Group. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Ann Intern Med* 2001; 134: 663-94.
- [10] Hildebrandt M, Vervölgyi E, Bender R. Calculation of NNTs in RCTs with time-to-event outcomes: A literature review. *BMC Med Res Methodol* 2009; 9: 21.
- [11] Stang A, Poole C, Bender R. Common problems related to the use of number needed to treat. *J Clin Epidemiol* 2010; 63: in press (DOI: 10.1016/j.jclinepi.2009.08.006).
- [12] Bender R, Blettner M. Calculating the "number needed to be exposed" with adjustment for confounding variables in epidemiological studies. *J Clin Epidemiol* 2002; 55: 525-30.
- [13] Bender R, Kuss O, Hildebrandt M, Gehrman U. Estimating adjusted NNT measures in logistic regression analysis. *Stat Med* 2007; 26: 5586-95.

- [14] Heller RF, Dobson AJ, Attia J, Page J. Impact numbers: Measures of risk factor impact on the whole population from case-control and cohort studies. *J Epidemiol Commun Health* 2002; 56: 606-10.
- [15] Laubender RP, Bender R. Estimating adjusted risk difference (RD) and number needed to treat (NNT) measures in the Cox regression model. *Stat Med* 2010; 29: in press.
- [16] Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. *J Clin Epidemiol* 2010; 63: 46-55.
- [17] Bender R, Kuss O. Methods to calculate relative risks, risk differences, and numbers needed to treat from logistic regression. *J Clin Epidemiol* 2010; 63: 7-8.
- [18] Austin PC. Different measures of treatment effect for different research questions. *J Clin Epidemiol* 2010; 63: 9-10.
- [19] Bender R. Estimating the number needed to treat with adjustment for balanced covariates in clinical trials. 30th Annual Conference of the International Society for Clinical Biostatistics, 23-27 August 2009, Prague, Czech Republic, Programme & Abstract Book: 100.
- [20] Yazdani R, Vehkalahti MM, Nouri M, Murtomaa H. School-based education to improve oral cleanliness and gingival health in adolescents in Tehran, Iran. *Int J Paediatr Dent* 2009; 19: 274-81.
- [21] Bland JM, Kerry SM. Trials randomised in clusters. *BMJ* 1997; 315: 600.
- [22] Meurman P, Pienihakkinen K, Eriksson AL, Alanen P. Oral health programme for preschool children: A prospective, controlled study. *Int J Paediatr Dent* 2009; 19: 263-73.
- [23] Hosmer DW Jr, Lemeshow S. *Applied Logistic Regression*; 2nd ed. New York: Wiley 2000.
- [24] Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998; 317: 1309-12.
- [25] Bender R. Calculating confidence intervals for the number needed to treat. *Control Clin Trials* 2001; 22: 102-10.
- [26] Heller RF, Dobson AJ. Disease impact number and population impact number: Population perspectives to measures of risk and benefit. *BMJ* 2000; 321: 950-2.
- [27] Attia J, Page J, Heller RF, Dobson AJ. Impact numbers in health policy decisions. *J Epidemiol Commun Health* 2002; 56: 600-5.
- [28] Hildebrandt M, Bender R, Gehrmann U, Blettner M. Calculating confidence intervals for impact numbers. *BMC Med Res Methodol* 2006; 6: 32.
- [29] Bender R, Grouven U. Berechnung von Konfidenzintervallen für die Population Impact Number (PIN). In: Hilgers R-D, Heussen N, Herff W, Ortseifen C, Eds. *KSFE 2008 – Proceedings der 12. Konferenz der SAS[®]-Anwender in Forschung und Entwicklung (KSFE)*. Aachen, Shaker 2008: 41-7.

Received: September 25, 2009

Revised: October 14, 2009

Accepted: October 14, 2009

© Ralf Bender; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.