Received: 21 February 2021 | Revised: 8 April 2022 | Accepted: 9 April 2022

DOI: 10.1002/sim.9418

RESEARCH ARTICLE

Statistics in Medicine WILEY

The number needed to treat adjusted for explanatory variables in regression and survival analysis: Theory and application

Valentin Vancak[1,2] | **Yair Goldberg**[3] | **Stephen Z. Levine**[4]

[1]Department of Statistics, University of Haifa, Haifa, Israel

[2]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[3]Faculty of Industrial Engineering and Management, Technion - Israel institute of Technology, Haifa, Israel

[4]Department of Community Mental Health, University of Haifa, Haifa, Israel

Correspondence
Valentin Vancak, Dept. of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.
Email: valentin.vancak@gmail.com

The number needed to treat (NNT) is an efficacy index commonly used in randomized clinical trials. The NNT is the average number of treated patients for each undesirable patient outcome, for example, death, prevented by the treatment. We introduce a systematic theoretically-based framework to model and estimate the conditional and the harmonic mean NNT in the presence of explanatory variables, in various models with dichotomous and nondichotomous outcomes. The conditional NNT is illustrated in a series of four primary examples; logistic regression, linear regression, Kaplan-Meier estimation, and Cox regression models. Also, we establish and prove mathematically the exact relationship between the conditional and the harmonic mean NNT in the presence of explanatory variables. We introduce four different methods to calculate asymptotically-correct confidence intervals for both indices. Finally, we implemented a simulation study to provide numerical demonstrations of the aforementioned theoretical results and the four examples. Numerical analysis showed that the parametric estimators of the NNT with nonparametric bootstrap-based confidence intervals outperformed other examined combinations in most settings. An R package and a web application have been developed and made available online to calculate the conditional and the harmonic mean NNTs with their corresponding confidence intervals.

KEYWORDS
adjusted NNT, conditional NNT, harmonic NNT, NNT, the cox model

1 | THE NNT

1.1 | Introduction

The number needed to treat (NNT) is an index that is widely used in efficacy analysis and cost-effectiveness analysis in randomized clinical trials, as well as in epidemiology and meta-analysis.[1-6] It is assumed that the outcome Y is dichotomous and may be beneficial or nonbeneficial. The NNT is the average number of patients that have to be treated in

order to observe one less adverse outcome,[7] or alternatively, the NNT can be defined as the average number of patients that are needed to treat in order to attain one more beneficial outcome due to treatment. These are two equivalent definitions since avoiding an adverse effect can be defined as a treatment benefit. Let the probability of treatment benefit be p_s; hence the NNT was initially defined by Laupacis et al[8] as $1/p_s$. Consider now that the probability of a beneficial outcome for treated patients p_t is composed of two additive components: the probability of a beneficial outcome due to treatment p_s (ie, the probability of treatment benefit), plus the probability p_c of a beneficial outcome that is not due to treatment. Therefore, the probability of treatment benefit is defined as $p_s = p_t - p_c$, that is termed the absolute risk difference (ARD). Although the NNT is a commonly used efficacy measure, it is not without limitations. Among the limitations of the NNT are difficulties in its interpretation,[7,9-12] a bi-modal sampling distribution[9] and unbounded disjoint confidence intervals (CIs). There are five main criticisms of the statistical properties of the NNT estimator. First is the singularity at 0 of the inverse of the ARD. Most of the criticism is directed toward this issue and its consequences, which include the difficulty to construct and interpret its CIs.[9,13] In particular, Grieve[9] applied Bayesian analysis to construct CIs with respect to the original definition of the NNT. Rohmel[13] proposed that if the ARD is statistically significant and does not contain zero, it is appropriate to invert the CIs. A comprehensive review of the NNT's statistical limitations can be found in Hutton.[10] Recently Vancak et al[14] resolved the pitfall of singularity at 0 by introducing a modification to the original definition of the NNT. The modified NNT is

$$NNT \equiv g(p_s) = \begin{cases} 1/p_s, & p_s > 0, \\ \infty, & p_s \leq 0. \end{cases} \tag{1}$$

We adopt this modified version of the NNT. Laupacis et al[8] proposed estimating this unadjusted NNT by replacing the unknown probabilities of a beneficial outcome in each arm with the corresponding proportions of beneficial outcomes. Therefore, the second criticism pointed to the bias of Laupacis' estimator with respect to the true NNT.[15] Third, the NNT is well defined only for dichotomous outcomes. However, in many clinical settings, the outcome Y is nondichotomous. Therefore, a dichotomization of the outcome is required,[16] which has limitations such as loss of information.[17,18] Dichotomization often relies on the definition of the minimal clinically important difference (MCID)[19] that is denoted by τ. Therefore, for nondichotomous outcomes, without loss of generality, we define the beneficial outcome as $I = I\{Y > \tau\}$, where $I = 1$ if the beneficial outcome occurs, and 0 otherwise.[14] Fourth, the NNT does not account for time-dependent outcomes and thus can be misleading.[11,12] Fifth, the interpretation and the definition of the NNT are debatable since different clinical scenarios may result in the same NNT.[20] Grieve,[9] and Hutton[10] challenged the common claim that the NNT is an easily interpretable index by emphasizing the statistical properties of the estimator of the NNT that are commonly neglected. Suissa[21] and Smeeth et al[22] present its miscalculations in various settings, while Kristiansen et al[7] present misinterpretation of the original NNT using an empirical study. We suggest that Vancak et al's modification of the NNT makes it easier to understand and interpret the index, and the this article clarifies the appropriate use of the NNT with time-dependent outcomes.

In Section 2, we introduce the conditional NNT, which is an NNT that is conditioned on a given value of the explanatory variables. Next, we present the harmonic mean NNT (hereafter, harmonic NNT), which is defined by applying the function g as defined in (1) to the marginal probability of treatment benefit. This presentation is followed by the derivation of these NNTs' corresponding asymptotically unbiased and efficient estimators, alongside their asymptotic distributions. In this section, we discuss two primary examples: logistic regression and linear regression. In Section 3, we derive the NNT accommodated to right-censored data conditioned and unconditioned on explanatory variables. Subsequently, we define the harmonic NNT accommodated to right-censored data. Then, we derive their corresponding asymptotically unbiased estimators alongside the asymptotic distributions. In this section, we discuss two main examples that are based on the Kaplan-Meier and Cox model. For all estimators we provide asymptotically correct CIs using four different methods: transformation, delta method, nonparametric and parametric bootstrap. In Section 4, we present a simulation study to illustrate the conditional and the harmonic NNTs in the examples discussed above. Notably, we compare point estimators, the lengths, and the coverage rates of the four CIs methods. Finally, we refer readers to the nntcalc R package[23] and the corresponding web application[24] to calculate the conditional and the harmonic NNTs with their corresponding CIs. The R package and the web application are made available for users online.[*] Detailed proofs appear in the Appendix.

2 | ADJUSTED NNT IN REGRESSION ANALYSIS

2.1 | Adjustment of the NNT

The need for adjustment of the NNT was recognized in parallel with the increased popularity of this index. The first known attempt to adjust the NNT was by Riegelman and Schroth.[25] Ebrahim,[26] and Misselbrook and Armstrong,[27] further acknowledged the need for a specified adjustment and made the distinction between the overall and the conditional NNT. In particular, Ebrahim[26] acknowledged the need to condition the NNT on both time and explanatory variables. There, regression was advocated for that purpose, however, practical tools were not presented. Altman and Andersen[28] conducted initial research on accommodating the NNT to right-censored data based on the Kaplan-Meier estimators. A detailed criticism of the use of NNT in the context of survival analysis and time-to-event data was presented by Kristiansen and Dorte,[7] and further by Snapinn and Jiang.[11] The main points of criticism were that the NNT may vary substantially over time, and hence convey different information as a function of the specific time-point of its calculation. Moreover, Snapinn and Jiang[11] showed examples where the information conveyed by the NNT may be incomplete or even contradictory compared to the traditional statistics of interest in survival analysis. Ola et al[29] adjusted the NNT for a particular explanatory variable in the context of cost-effectiveness analysis and redefined it as the cost needed to treat. More comprehensive research on adjustment of the NNT for explanatory variables was conducted by Bender and Blettner[30] and Austin.[31]

2.2 | Conditional and harmonic NNT in regression analysis

Bender and Blettner[30] and Austin[31] adjusted the NNT for explanatory variables in logistic regression. In particular, Bender and Blettner[30] and Bender et al[32] adjusted the NNT in the context of cohort data. In this situation, the covariates may have a different distribution in the exposed and the unexposed arms. Therefore, a distinct NNT was proposed for each arm. Notably, to adjust the NNT to covariate distributions in the control (unexposed) arm, they renamed it the Number Needed to be Exposed (NNE). Furthermore, to account for the effect direction, they divided the NNE into NNEH and NNEB, where H stands for a harmful effect (ie, negative valued NNE), and B for a beneficial effect (ie, positively valued NNE). In these two works, the authors suggested to use the multivariate delta method for the CI construction of the adjusted NNE. In addition to the NNE, they presented the Exposure Impact Number (EIN) to adjust the NNT to the covariates' distribution in the treated (exposed) arm.[30,32] Later on, Bender and Vervolgyi,[33] presented the estimation of harmonic NNT, based on the fit of the logistic regression model in the context of randomized clinical trials. In all these works, the authors used the multivariate delta method to compute the CIs of the index. In the context of time-to-event, Altman and Andersen[28] proposed inversion of the CIs of the absolute risk reduction to construct CIs for the NNT. Austin[34] introduced the harmonic NNT and suggested using the nonparametric bootstrap for construction of its CI.

2.3 | Definition of the conditional NNT(x)

In this subsection we introduce the adjusted NNT for explanatory variables and/or covariates X in the context of regression analysis. Consider the following scenario: At the baseline, for every clinical trial study participant, a vector of background measurements X is taken. Each clinical trial study participant is randomly allocated either to the control c or the treatment arm t. Usually the explanatory variables include sex, age, baseline severity and treatment. Randomization ensures that all baseline covariates of patients in both treatment groups share the same joint parent distribution. The outcome variable Y, defined as in Section 1, is a scalar function of the baseline and the endpoint measurements. Without loss of generality, let the beneficial outcome I be defined as $I \equiv I\{Y > \tau\}$; hence, conditioning on the allocated arm is defined as $I_a \equiv I\{Y > \tau | A = a\}$, where $a \in \{c, t\}$. Define the marginal and the conditional probability of beneficial outcome in the ath arm by $\mathbb{E}[I_a] = p_a$ and $\mathbb{E}[I_a|X = x] = p_a(x)$, respectively. Consequently, we define the conditional probability of treatment benefit given a covariate X as

$$p_s(x) \equiv \mathbb{E}[I_t - I_c|X = x] \equiv p_t(x) - p_c(x), \quad \forall x \in \mathcal{X}, \tag{2}$$

where $\mathcal{X}$ is the support set of the covariate X. Hence, we define the conditional NNT(x) for every x in $\mathcal{X}$ as

$$NNT(x) \equiv g(p_s(x)) = \begin{cases} (p_t(x) - p_c(x))^{-1}, & p_t(x) > p_c(x), \\ \infty, & p_t(x) \leq p_c(x). \end{cases} \tag{3}$$

The conditional NNT(x) allows us to calculate the NNT for every possible value x of the covariate X. Equation 4 below establishes the connection between the conditional NNT(x) and the harmonic NNT. Let the NNT be as defined in (1) and the NNT(x) as defined in (3). Then,

$$g(p_s) = g\left(\mathbb{E}[p_s(X)]\right). \tag{4}$$

Equation (4) states that, since an expectation of the conditional probability of treatment benefit $p_s(X)$ w.r.t. X results in the average (marginal) probability of treatment benefit p_s, therefore, the harmonic NNT can be calculated by applying g as defined in (1) to $\mathbb{E}[p_s(X)]$. Moreover, as g is a convex function for all x in $\mathcal{X}$, by applying Jensen's inequality[35] it is evident that

$$\mathbb{E}[g(p_s(X))] \geq g(\mathbb{E}[p_s(X)]). \tag{5}$$

Averaging the NNT(x) instead of averaging the $p_s(x)$ will result in biased NNT. Moreover, it may lead to a wrong conclusion since the distribution of $g(p_s(X))$ can be dominated by its extreme values. Consider, for example, a scenario where there is a possible realization x_0 of the covariates X such that $p_s(x_0) = 0$, while for every other realization x, $x \neq x_0$, $p_s(x) > 0$. Therefore, NNT(x_0) $\equiv$ $g(p_s(x_0)) = \infty$: thus the NNT computed as a mean on the NNT scale equals infinity, that is, $\mathbb{E}[g(p_s(X))] = \infty$. However, if $\mathbb{E}[p_s(X)] > 0$, consequently, the NNT computed as a mean on the risk scale is finite, that is, NNT $\equiv$ $g(\mathbb{E}[p_s(X)]) < \infty$.

2.4 | Estimation of NNT(x)

This section introduces the parametric approach to conditional NNT(x) estimation. Assume that the conditional probability of treatment benefit $p_s(x)$ can be described via a vector of unknown parameters $\theta_0^T = (\theta_c^T, \theta_t^T)$. Assume that the probability of beneficial outcome in the ath arm is $p_a(x;\theta) = p(x;\theta_a)$, for $a \in \{c, t\}$. Consequently, $p_s(x;\theta) = p(x;\theta_t) - p(x;\theta_c)$, and by (3), NNT($x$) $\equiv$NNT($x;\theta$) $\equiv g(p_s(x;\theta))$. Notably, $p_s(x;\theta)$ is the average treatment effect that is commonly a nonlinear function of the covariates X. The fundamental problem of causal inference states that it is impossible to observe both I_t and I_c within the same clinical trial study participant,[36] as the participant is allocated to either the treatment or the control arm. Therefore, we observe only one of the two possibilities, while the missing one can be estimated using a model. Thus, we estimate NNT($x;\theta$) by replacing the unknown parameters θ with their corresponding point estimators.

For the estimation of the harmonic NNT, the marginal distribution of X is required. Usually such a distribution is unavailable; therefore we propose to estimate $\mathbb{E}[p_s(X;\theta)]$ using the corresponding sample average and then applying g as defined in (1) to the result, that is,

$$\widehat{NNT} = g\left(\frac{1}{n}\sum_{a,i_a} p_s\left(x_{i_a};\hat{\theta}_c,\hat{\theta}_t\right)\right). \tag{6}$$

Namely, for the harmonic NNT, we use a parametric model to estimate the conditional risk difference $p_s(x;\theta)$ given a set of covariates X, then average it over the empirical distribution of the covariates, and finally take its inverse. Next we present two examples to compute the conditional and the harmonic NNTs: logistic regression and linear regression.

Example 1: NNT(x) in logistic regression

Let $X = (X_1, \ldots, X_p)^T$ be the vector of explanatory variables with a support set $\mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{R}^p$. Assume that the response variable Y is dichotomous, that is, $Y \in \{0, 1\}$. Without loss of generality, let the beneficial outcome be defined as $Y = 1$. Assume that given realization x, and an allocated arm a, Y follows a Bernoulli distribution with probability

$p_a(x) \equiv p(x;\beta_a)$, where $p(x;\beta_a) \equiv (1+\exp\{-\beta_{0a} - \sum_{j=1}^{p} x_j\,\beta_{ja}\})^{-1}$. In this model, $\theta^T = (\beta_c^T, \beta_t^T)$, where $\beta_a^T = (\beta_{0a}, \ldots, \beta_{pa})$, for $a \in \{c,t\}$. It can be shown that the conditional probability of treatment benefit is

$$p_s(x;\theta) = \frac{e^{\beta_{0t}+\sum_{j=1}^{p} x_j\beta_{jt}} - e^{\beta_{0c}+\sum_{j=1}^{p} x_j\beta_{jc}}}{\left(1+e^{-\beta_{0t}-\sum_{j=1}^{p} x_j\beta_{jt}}\right)\left(1+e^{-\beta_{0c}-\sum_{j=1}^{p} x_j\beta_{jc}}\right)}. \tag{7}$$

The MLE of the NNT(x) is attained by applying g to $p_s(x;\theta)$ and replacing the unknown parameters with their corresponding MLEs. The MLE of the harmonic NNT is calculated by applying g to the corresponding sample average of $p_s(x;\hat{\theta})$ as defined in (6).

Example 2: NNT(x) in normal linear regression

Let $X = (X_1, \ldots, X_p)^T$ be the vector of explanatory variables with a support set $\mathcal{X}$, where $\mathcal{X} \subseteq \mathbb{R}^p$. Assume that the response variable Y, given a realization x and an allocated arm a, follows a normal distribution $N\left(\beta_{0a} + \sum_{j=1}^{p} x_{ja}\beta_{ja}, \sigma^2\right)$. In this model, $\theta^T = (\beta_c, \beta_t, \sigma^2)$, where $\beta_a^T = (\beta_{0a}, \ldots, \beta_{pa})$, for $a \in \{c,t\}$, and the formal model is

$$Y_{i_a} = \beta_{0a} + \sum_{j=1}^{p} x_{ji_a}\beta_{aj} + \epsilon_{i_a}, \quad i_a = 1, \ldots, n_a, \quad a \in \{c,t\}. \tag{8}$$

It can be shown that the conditional probability of treatment benefit is

$$p_s(x;\theta) = \Phi\left(\frac{\tau - \beta_{0c} - \sum_{j=1}^{p} x_j\beta_{jc}}{\sigma}\right) - \Phi\left(\frac{\tau - \beta_{0t} - \sum_{j=1}^{p} x_j\beta_{jt}}{\sigma}\right). \tag{9}$$

The MLE of the conditional NNT(x) is attained by applying g to $p_s(x;\theta)$ and replacing the unknown parameters with their corresponding MLEs. The MLE of the harmonic NNT is calculated by applying g to the corresponding sample average of $p_s(x;\hat{\theta})$ as defined in (6). Note that for these two examples, model parameters can be shared by both arms. In other words, $p(x;\theta_c)$ and $p(x;\theta_t)$ can be modeled either separately with different parameters θ_c and θ_t or jointly with common parameters that are shared by models of the two arms. These specifications change neither the estimation nor the asymptotic properties of the NNT(x) estimators.

2.5 | Theoretical properties of the NNT(x) estimator

To this end, the following assumptions are required.

- A.1 The response variable Y, given $X = x$ and $A = a$, follows a parametric model with a true parameter θ_0. In this case, the true conditional NNT is given by $\text{NNT}(x) \equiv g(p_s(x;\theta_0))$.
- A.2 There are n_a observations in the ath arm, $a \in \{c,t\}$, and $n = n_c + n_t$ is the total number of observations. Assume that $n_c/n_t \to \alpha$, as $n \to \infty$.
- A.3 The probability of treatment benefit, $p_s(x;\theta)$, is a differentiable function w.r.t. θ; hence the composite function $g(p_s(x;\theta))$ where g as defined in (1), is also a differentiable function w.r.t. θ for $p_s(x;\theta) > 0$, and $\nabla g(p_s(x;\theta_0)) \neq 0$.
- A.4 The standard regularity conditions of the MLEs asymptotics hold.[35]
- A.5 $\mathbb{E}[p_s(X;\theta_0)] > 0$, and $\nabla g\left(\mathbb{E}[p_s(X;\theta_0)]\right) \neq 0$.

Theorem 1. *Let $\hat{\theta}_n$ be the MLE of θ_0, $\mathbf{I}(\theta_0)$ its fisher information matrix, and $Z \sim N(0,1)$. Then,*

1. *$g(p_s(x;\hat{\theta}_n))$ and $g\left(\frac{1}{n}\sum_{i_a,a} p_s(x_{i_a};\hat{\theta}_n)\right)$ are the MLEs of the conditional NNT(x) and the harmonic NNT, respectively.*
2. *$g(p_s(x;\hat{\theta}_n)) \xrightarrow{a.s.} \text{NNT}(x)$, and $g\left(\frac{1}{n}\sum_{i_a,a} p_s(x_{i_a};\hat{\theta}_n)\right) \xrightarrow{a.s.} \text{NNT}$.*
3. *For every x, such that $p_s(x) > 0$,*

i. $\sqrt{n}\left(g(p_s(x;\hat{\theta}_n)) - \mathrm{NNT}(x)\right) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\theta_0)\nabla g(p_s(x;\theta_0))\|Z,$

ii. $\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i_a,a} p_s(x_{i_a};\hat{\theta}_n)\right) - \mathrm{NNT}\right) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\theta_0)\nabla g(\mathbb{E}[p_s(X;\theta_0)])\|Z,$

where $\nabla g(p_s(x;\theta_0))$ *and* $\nabla g(\mathbb{E}[p_s(X;\theta_0)])$ *are the gradients of NNT*(x) *and NNT, respectively, evaluated at* θ_0.

For a detailed proof of Theorem 1 please refer to the Appendix. Notably, for every x such that $p_s(x) \leq 0$, $\mathrm{NNT}(x) \equiv \infty$. Therefore, the MLE of $\mathrm{NNT}(x)$ converges almost surely to infinity. In such a case, discussion of the $\mathrm{NNT}(x)$ distribution is meaningless.

2.6 | CIs of harmonic NNT and conditional NNT(x)

In this subsection we present four different methods of CI construction. These methods will be further used to construct asymptotically-correct CIs for the conditional and the harmonic NNT. These results are summarized in Theorem 2.

Approach 1: Transformation

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed (i.i.d) random variables with $\mathbb{E}[X_i] = \mu_0$ and $Var(X_i) = \sigma_0^2$. Denote $\theta_0^T = (\mu_0, \sigma_0^2)$. By the univariate Central Limit Theorem (CLT)

$$\sqrt{n}\frac{(\overline{X}_n - \mu_0)}{\sigma_0} \xrightarrow{D} Z, \tag{10}$$

where $Z \sim N(0,1)$, and $n \to \infty$. For a monotonically decreasing function h, for example, the one defined in (1), the univariate-CLT-based asymptotically-correct $(1-\alpha)100\%$ level CI for $h(\mu_0)$ is

$$h\left(\overline{X}_n + z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right) \leq h(\mu_0) \leq h\left(\overline{X}_n - z_{1-\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right), \tag{11}$$

where z_α is the αth quantile of the standard normal distribution.

Approach 2: Delta method

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed $p-$dimensional random vectors, with $\mathbb{E}[X] = \mu_0$ and $\mathrm{Var}(X) = \mathbf{\Sigma}_0$, where $\mathbf{\Sigma}_0$ is a positive definite covariance matrix of X. Denote $\theta_0^T = (\mu_0, \mathbf{\Sigma}_0)$. By the multidimensional CLT

$$\sqrt{n}(\overline{X}_n - \mu_0) \xrightarrow{D} \mathbf{\Sigma}_0^{1/2} Z_p, \tag{12}$$

where $Z_p \sim N_p(0, \mathbf{I})$. Let $h : \mathbb{R}^p \to \mathbb{R}$ be a differentiable function over the parametric set Θ with a nonzero gradient $\nabla h(\mu_0)$. Consequently, the delta-method-based asymptotically-correct $(1-\alpha)100\%$ level CI for $h(\mu_0)$ is

$$h(\overline{X}_n) - \frac{1}{\sqrt{n}}\|\hat{\mathbf{\Sigma}}^{1/2}\nabla h(\overline{X}_n)\|z_{1-\alpha/2} \leq h(\mu_0) \leq h(\overline{X}_n) + \frac{1}{\sqrt{n}}\|\hat{\mathbf{\Sigma}}^{1/2}\nabla h(\overline{X}_n)\|z_{1-\alpha/2}, \tag{13}$$

where $\|\cdot\|$ is the Euclidean norm.

Approach 3: Nonparametric bootstrap

Assume that the original sample is of n realizations of a random vector X. The constant of interest is a real valued scalar function h at some point x, that is, $h(x)$. The bootstrap algorithm[37] takes n samples with replacements from the n

aforementioned observations. This is repeated B times with B samples, each of size n. For each of these samples $b = 1, \ldots, B$, an estimator $h^b(x)$ is calculated. Using the $\alpha/2$ and the $1-\alpha/2$ quantiles of the empirical distribution of $h^1(x), \ldots, h^B(x)$ we obtain a nonparametric-bootstrap-based asymptotically-correct $(1-\alpha)100\%$ level CI

$$h^B(x)_{\alpha/2} \leq h(x) \leq h^B(x)_{1-\alpha/2}. \tag{14}$$

Approach 4: Parametric bootstrap

Assume that the original sample consists of n observations, which are assumed to be realizations of a parametric distribution that depends on the unknown vector of parameters θ_0. This vector is estimated by $\hat{\theta}_n$. Assume that the sample distribution of the estimator $\hat{\theta}_n$ is $N(\theta_0, \mathbf{\Sigma}_n(\theta))$. The parametric bootstrap algorithm[37] takes the estimator $\hat{\theta}_n$ and its sampled covariance matrix as a replacement of θ_0, and the corresponding covariance matrix $\mathbf{\Sigma}(\theta_0)$. Then we sample B times, where B is some large positive integer, from $N(\hat{\theta}_n, \mathbf{\Sigma}_n(\hat{\theta}_n))$. For each of the B samples of parameters $\hat{\theta}^b$, a function $h(\hat{\theta}^b)$, for $b = 1, \ldots, B$, is calculated, where h is a scalar function. This results in a sample of B estimators of $h(\theta_0)$. Using the normality of this sample, the parametric-bootstrap-based asymptotically-correct $(1-\alpha)100\%$ level CI is

$$h(\hat{\theta}_n) - z_{1-\alpha/2}\hat{\sigma}_{h(\hat{\theta}^B)} \leq h(\theta_0) \leq h(\hat{\theta}_n) + z_{1-\alpha/2}\hat{\sigma}_{h\left(\hat{\theta}^B\right)}, \tag{15}$$

where $\hat{\sigma}^2_{h(\hat{\theta}^B)}$ is the bootstrap sample standard deviation. An advantage of the parametric bootstrap is that in order to obtain an effective CI, usually significantly fewer samples are required, compared to its nonparametric counterpart.

Theorem 2. *Let assumptions A.1-A.5 hold, and let* $g(p_s(x;\hat{\theta}_n))$, *and* $g\left(\frac{1}{n}\sum_{i_a,a} p_s\left(x_{i_a};\hat{\theta}_n\right)\right)$ *be the MLE of NNT*(x) *and NNT, respectively. Therefore, for a fixed* x, *where* $p_s(x) > 0$, *the asymptotically-correct transformation-based, delta-method-based, nonparametric-bootstrap-based, and parametric-bootstrap-based* $(1-\alpha)100\%$*-level CIs are given in (11), (13), (14), and (15), respectively. In this case,* $h(\theta_0) = g(p_s(x;\theta_0))$ *and* $h(\theta_0) = g(\mathbb{E}[p_s(X;\theta_0)])$, *for NNT*$(x)$ *and NNT, respectively. The function* g *is defined in (1), and* $\hat{\mathbf{\Sigma}}_0$ *is the inverse of the observed Fisher information matrix evaluated at* $\hat{\theta}$, *that is,* $\hat{\mathbf{I}}^{-1}(\theta_0) \equiv \mathbf{I}^{-1}(\hat{\theta}_n)$.

Note that since NNT$(x) \geq 1$ for all x, the lower confidence limit of NNT(x) can be truncated at 1 with no effect on the coverage rate.

3 | HARMONIC NNT(Y) AND CONDITIONAL NNT(Y|X) IN SURVIVAL ANALYSIS

3.1 | Introduction

Laubender and Bender,[38] Austin,[34] and Yang and Yin[39] presented and elaborated on an NNT for right-censored data in the Cox proportional hazard and parametric accelerated failure models. Laubender and Bender[38,40] derived the harmonic NNE and EIN in the framework of the Cox model, with four types of CIs. Three types were based on the resampling method conditional on the covariates:[41] the normal approximation, the basic bootstrap and the percentile bootstrap. The fourth type was a multivariate delta-method CI based on the theory of martingales.[40] Yang and Yin[39] proposed an alternative measure to the NNT that is based on the restricted mean survival time instead on the absolute risk difference. All these works analyzed model-specific scenarios with naturally dichotomous outcomes without providing a general modeling framework. The singularity point of the original definition often led to categorization of the NNT into different measures as a function of the allocated arm and the sign of its estimator, and resulted in CIs that consist of a union of two disjoint infinite intervals. In addition, different authors used varying approaches and terminology to define and compute the conditional and the harmonic NNTs.

3.2 | Definition of NNT(y) and NNT($y|x$)

Let the outcome variable Y be the time to event or time until death. For a fixed time point $y \geq 0$, define the beneficial outcome as $I = I\{Y > y\}$. Let the cumulative distribution of Y be $F(y)$, and define the survival function as $S(y) = 1 - F(y)$.

In survival analysis we may not observe the realization Y due to loss of follow-up or for other reasons. Namely, we assume that all clinical trial study participants start at time 0, but some are censored before the end of the trial. Therefore, we observe $T = \min\{Y, C\}$, where C is right-censoring time. The beneficial outcome, given allocated arm a, is $I_a(y) \equiv I\{Y > y|A = a\}$, where $a \in \{c, t\}$. The marginal probability of a beneficial outcome in the ath arm is $S_a(y)$; thus, the marginal probability of treatment benefit is $p_s(y) \equiv S_t(y) - S_c(y)$. Consequently, we define the NNT(y) as

$$NNT(y) \equiv g(p_s(y)), \quad y \geq 0. \tag{16}$$

Let $X = (X_1, \ldots, X_p)^T$ be the risk factors. Define the conditional NNT($y|x$) as

$$NNT(y|x) = g(p_s(y|x)), \quad \forall x \in \mathcal{X}, \quad y \geq 0, \tag{17}$$

where $p_s(y|x) = p_t(y|x) - p_c(y|x)$, and g is the function defined in (1). Similarly to the harmonic NNT, the harmonic NNT(y) is calculated by applying g to $\mathbb{E}[p_s(y|X)]$, where the expectation is taken w.r.t. X.

3.3 | Estimation of NNT(y)

Let $S_a(y)$ be an unspecified survival function of the ath arm, $a \in \{c, t\}$, at a fixed time point y. To estimate the NNT(y), we need a suitable estimator of $S_a(y)$, which is also the marginal probability of a beneficial outcome in the ath arm. Let $n = n_c + n_t$ be the total number of clinical trial study participants that are observed at the baseline time point $y = 0$. The next example and the subsequent theorems present the nonparametric MLE of NNT(y) and its theoretical properties.

Example 3: NNT(y) using the Kaplan-Meier nonparametric MLE

Let $\hat{S}_a(y) \equiv p_a^{KM}(y)$ be the product limit Kaplan-Meier's nonparametric MLE;[42] formally, for any fixed time $y \geq 0$ we have

$$p_a^{KM}(y) = \prod_{y_j \leq y} \left(1 - \frac{d_{ja}}{n_{ja}}\right), \quad a \in \{c, t\}, \tag{18}$$

where n_{ja} is the number of clinical trial study participants in risk at time y_j, and d_{ja} is the total number of failures at time y_j in the ath arm. Therefore, we define the $\text{NNT}_{KM}(y)$ to be the estimator of NNT(y) by applying g as defined in (1) to the difference between the corresponding Kaplan-Meier's estimators. Formally,

$$NNT_{KM}(y) = g\left(p_s^{KM}(y)\right), \tag{19}$$

where $p_s^{KM}(y) = p_t^{KM}(y) - p_c^{KM}(y)$. Notably, $\text{NNT}_{KM}(y)$ can be stratified by more than two levels. Furthermore, when there is no censoring, it can be shown that $\text{NNT}_{KM}(y)$ coincides with Laupacis NNT_L, where the beneficial outcome is defined as $I\{Y > y\}$. See Lemma 2 in the Appendix for detailed proof. Similar results for the parametric logistic regression and the parametric linear regression approaches are presented in Section 2.4.

3.4 | Theoretical properties of the NNT(y)'s estimator

The asymptotic properties of NNT(y)'s parametric estimator are presented in Theorem 1, by replacing x with y. Therefore, we focus only on the theoretical properties of the nonparametric MLE. For further theorems we need another assumption.

A.6 Let $[0, \eta)$ be the follow-up time interval, where η is a time point such that there is at least one more observation to its right in each of the arms.

Theorem 3. *Let assumptions A.2 and A.6 hold, and assume that the survival time Y is independent of the right-censoring time C, and $Z \sim N(0, 1)$. Then, for any fixed time point y in $[0, \eta)$*

1. $NNT_{KM}(y)$ *is the nonparametric MLE of* $NNT(y)$.
2. $NNT_{KM}(y) \xrightarrow{a.s.} NNT(y)$.
3. *For every fixed* y, *such that* $p_s(y) > 0$, $\sqrt{n}(NNT_{KM}(y) - NNT(y)) \xrightarrow{D} \sigma_0(y)Z$, *where* $\sigma_0^2(y)$ *is the asymptotic variance of* $\sqrt{n}NNT_{KM}(y)$.

For a detailed proof of Theorem 3 please refer to the Appendix. We proceed to the construction of CIs that are based on the large-sample distribution of the $\text{NNT}_{KM}(y)$.

Theorem 4. *For every fixed* $y \in [0, \eta)$, *such that* $p_s(y) > 0$, *the asymptotically-correct transformation-based, delta-method-based, nonparametric-bootstrap-based, and parametric-bootstrap-based* $(1-\alpha)100\%$*-level CIs are given in (11), (13), (14), and (15), respectively. In this case,* $h(p_s(y)) = g(p_s(y))$, *and* $\hat{\boldsymbol{\Sigma}}(y)$ *is* $NNT_{KM}^4(y)\hat{\sigma}_s^2(y)$, *where* $\hat{\sigma}_s^2(y)$ *is the estimated variance of* $p_s^{KM}(y)$ *that is based on Greenwood's formula.*[43]

3.5 | Estimation of the NNT($y|x$)

If the structure of $p_s(y|x)$ is assumed to be fully parametric, then the estimation is done in the same manner as in the parametric models of conditional NNT(x), that is, by replacing the unknown parameters with their estimators. Therefore, we focus our attention only on the semiparametric structures. We start with a definition of the hazard function. The hazard function is the instantaneous rate of occurrence of the event. Formally,

$$\lambda(y) = \lim_{dy \to 0} \frac{\mathbb{P}(y \le Y < y + dy | Y \ge y)}{dy} = -\frac{\partial}{\partial y} \ln S(y). \tag{20}$$

Therefore, the probability of beneficial outcome in the ath arm, $S_a(y)$, can be expressed as a function of the cumulative hazard function up to time y, $\Lambda_a(y)$, that is,

$$S_a(y) = \exp\{-\Lambda_a(y)\}, \quad a \in \{c, t\}, \tag{21}$$

where $\Lambda_a(y) \equiv \int_{[0,y]} \lambda_a(x)dx$. Next, define $\lambda_a(y|x)$ to be the hazard function given the risk factors $x \in \mathcal{X}$, and an allocated arm $a \in \{c, t\}$. Therefore, the conditional probability of beneficial outcome in the ath arm is $p_a(y|x) = \exp\{-\Lambda_a(y|x)\}$. The most commonly-used model for the conditional hazard function is Cox's proportional hazard model.[44]

Example 4: NNT($y|x$) in Cox regression

Let $X = (X_1, \dots, X_p)^T$ be the vector of risk factors with a support set $\mathcal{X} \subseteq \mathbb{R}^p$. Assume that given a realization $x = (x_1, \dots, x_p)^T$, and an allocated arm a, the hazard function is

$$\lambda_a(y|x) = \lambda(y) \exp\left\{\beta_a^T x\right\}. \tag{22}$$

Namely, the baseline hazard $\lambda(y)$ has a nonparametric structure, while the adjustment for the risk factors follows a parametric model. In this model, $\theta_0 = (\Lambda_0(y), \beta_0^T)$, where $\Lambda_0(y)$ is the cumulative baseline hazard function up to time y, $\beta_0^T = (\beta_c^T, \beta_t^T)$, and $\beta_a^T = (\beta_{0a}, \dots, \beta_{pa})$, for $a \in \{c, t\}$. The conditional cumulative hazard function in the ath arm is $\Lambda_a(y|x) = \Lambda_0(y) \exp\left\{\beta_a^T x\right\}$. Therefore, using (22), the conditional probability of beneficial outcome in the ath arm is

$$p_a(y|x) = \exp\left\{-\Lambda_0(y) \exp\left\{\beta_a^T x\right\}\right\} = S(y)^{\exp\{\beta_a^T x\}},$$

where $S(y)$ is the baseline survival probability. Following (2), the conditional probability of treatment benefit is

$$p_s(y|x; \theta_0) = S(y)^{\exp\{\beta_t^T x\}} - S(y)^{\exp\{\beta_c^T x\}}. \tag{23}$$

The semiparametric estimator of the conditional NNT($y|x$), is attained by applying the function g from (1) to the estimated $p_s(y|x; \theta_0)$, where β_0^T is replaced with its maximum partial likelihood estimators (MPLE),[44] and $\Lambda_0(y)$ with Breslow's

nonparametric MLE.[45] The partial likelihood does not depend on the baseline hazard function $\lambda(y)$, which allows estimating the model coefficients β_0 without dealing with the nonparametric baseline hazard, which is estimated afterward using the calculated MPLEs. The calculated semiparametric estimator of the conditional NNT$(y|x)$ we denote by NNT$_{COX}(y|x)$ since it is based on the Cox proportional hazard regression model and is estimated via the Cox partial likelihood method. Under this model, the estimator of the harmonic NNT(y), which is denoted by NNT$_{COX}(y)$, is calculated by applying g to the corresponding sample average of $p_s(y|x;\hat{\theta})$ as defined in (6).

3.6 | Theoretical properties of the NNT$(y|x)$'s estimator

In this section we present the asymptotic analysis of the conditional NNT$(y|x)$ semiparametric estimator and its corresponding CIs. The generalization of the presented results to other survival models is straightforward using the principles presented below. Thus, we will focus only on the Cox model. To this end, additionally to A.1-A.6., we require the next two assumptions.

A.7 $\mathbb{E}[X\exp\{\beta_0^T X\}]^2$ is bounded uniformly in the neighborhood of β_0.
A.8 $\Lambda_0(y) < \infty$ for all finite and positive y.

Theorem 5. *Assume that A.1 to A.8 hold, and that the true model follows the Cox proportional hazard structure with an unspecified baseline hazard, such that C and Y are independent conditionally on the risk factors X. Let $\hat{\beta}_n^T$ and $\hat{\Lambda}(y)$ be the MPLE and Breslow's nonparametric MLE, respectively. Then, for any fixed time point $y \in [0,\eta)$, and for all $x \in \mathcal{X}$,*

1. $NNT_{COX}(y|x) \xrightarrow{a.s.} NNT(y|x)$, *and* $NNT_{COX}(y) \xrightarrow{a.s.} NNT(y)$.
2. *For every fixed time point $y \in [0,\eta)$, such that $p_s(y|x) > 0$,*
   i. $\sqrt{n}\,(NNT_{COX}(y|x) - NNT(y|x)) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\beta_0)\nabla g(p_s(y|x;\beta_0))\|Z$,
   ii. $\sqrt{n}\,(NNT_{COX}(y) - NNT(y)) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\beta_0)\nabla g\,(\mathbb{E}[p_s(y|X;\beta_0)])\|Z$,
   where $\nabla g(p_s(y|x;\beta_0))$ and $\nabla g(\mathbb{E}[p_s(y|X;\beta_0)])$ are the gradients of $NNT(y|x)$ and $NNT(y)$, respectively, evaluated at β_0, $\mathbf{I}(\beta_0)$ is the Fisher information matrix, and $Z \sim N(0,1)$.

For a detailed proof of Theorem 5 please refer to the Appendix. The construction of the CIs is done in the same manner as in Theorem 2, and presented in Theorem 6.

Theorem 6. *Let assumptions A.1 to A.8 hold. For every $x \in \mathcal{X}$, and a fixed time point $y \in [0,\eta)$, such that $p_s(y|x;\beta_0) > 0$, the asymptotically-correct transformation-based, delta-method-based, nonparametric-bootstrap-based, and parametric-bootstrap-based $(1-\alpha)100\%$-level CIs are given in (11), (13), (14), and (15), respectively. In this case, $h(\theta_0) = g(p_s(y|x;\beta_0))$, and $h(\theta_0) = g(\mathbb{E}[p_s(y|X;\beta_0)])$, for $NNT(y|x)$ and $NNT(y)$, respectively. The sample covariance matrix $\hat{\mathbf{\Sigma}}_0$ is estimated using Fisher's observed information matrix $\mathbf{I}^{-1}(\hat{\beta}_n)$ that is based on Cox MPLEs of β_0.*

For a detailed derivation of the sample variance of NNT$_{COX}(y|x)$ in Theorem 6, please refer to the Appendix. For every fixed $x \in \mathcal{X}$, both empirical processes $\sqrt{n}(NNT_{COX}(y|x) - NNT(y|x))$, and $\sqrt{n}(NNT_{COX}(y) - NNT(y))$ converge weakly to a corresponding Gaussian process.[46] These processes have zero mean and covariance structures that can be derived from lemma 6.1 in Tsiatis.[47]

4 | SIMULATION STUDY

In this section, a simulation analysis of the aforementioned four examples (logistic regression, linear regression, Kaplan-Meier, and Cox regression) to compute the conditional and the harmonic NNTs and their corresponding 95% level CIs is presented. The harmonic NNTs were calculated, as described in (6), using the empirical distribution of the covariates. For all settings, sample sizes of 200, 400, and 800 were implemented, with 400 iterations for each sample size. The simulation results of the point estimators and the lengths of the corresponding finite CIs are summarized using boxplot charts. In addition, tables of the CIs' mean coverage rates are presented. This simulation can be replicated using the R package that has been developed and made available for users from the author's GitHub repository. Setting I illustrates the

logistic regression model with a continuous explanatory variable. Setting II illustrates the Kaplan-Meier and the Cox proportional hazard models under the Weibull distribution. Setting III (please see Appendix) illustrates the linear regression model with a continuous explanatory variable. For all settings, the true harmonic NNT was computed by Monte Carlo numerical integration.

Setting I: Logistic regression

In this setting, the explanatory variable X is a normally distributed variable with $\mu_X = 2$ and $\sigma^2 = 1$. The response variable Y, given a realization x and an allocated arm $a \in \{c, t\}$, follows a Bernoulli distribution with probability $p(x; \beta_a) = (1 + \exp\{-\beta_{0a} - \beta_{1a}x\})^{-1}$, where $\beta_c^T = (\beta_{0c}, \beta_{1c}) = (-2, 1)$, and $\beta_t^T = (\beta_{0t}, \beta_{1t}) = (0, 0.5)$. Without loss of generality, the beneficial outcome is defined as $I\{Y = 1\}$. The conditional NNT(x) can be calculated for every x. However, for this illustration we chose only three representative values: 1.5, 2, and 2.5. The true values of NNT(1.5), NNT(2), and NNT(2.5) are 3.32, 4.33, and 6.46, respectively. The true value of the harmonic NNT, computed over the full covariate distribution used in the simulation, is 4.54. The conditional NNT(x)'s, $x = 1.5, 2, 2.5$, point estimators and the corresponding CIs with their mean coverage rates are illustrated in Figures 1 and 2, and Table 1. The point estimators of the harmonic NNTs with the corresponding CIs and their mean coverage rates are presented in Figures 3 and 4, and Table 2.

Setting II: Survival analysis

In this setting, the data were generated with the survsim R package.[48] Specifically, the explanatory variable X is a normally distributed variable with $\mu_X = -1.5$ and $\sigma^2 = 1$. The allocated arm A is a binomial random variable with probability of 0.5, and $n = n_c + n_t$. The response variable Y, given a realization x and an allocated arm $a \in \{c, t\}$, follows the Weibull

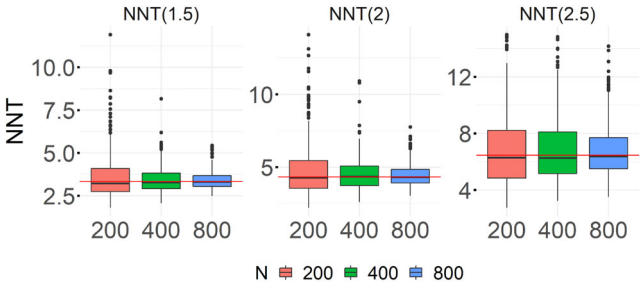

FIGURE 1 Parametric MLEs of the Conditional NNT(x), for $x = 1.5, 2, 2.5$, in the logistic regression model (for a formal definition of NNT(x) in the logistic regression model see Equation (7)), for $n = 200,400, 800$

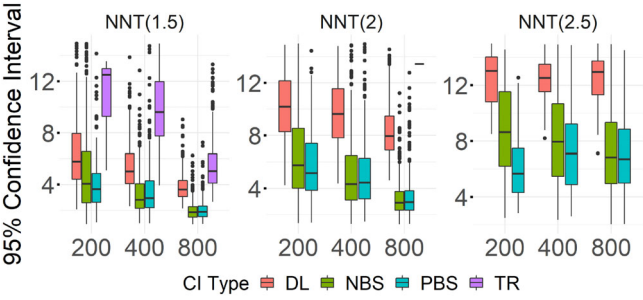

FIGURE 2 CI lengths by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation) of the conditional NNT(x), $x = 1.5, 2, 2.5$, in the logistic regression model, for $n = 200,400, 800$. The box-plots of certain transformation-based CIs were not displayed, since they are either infinite or too large, and thus distort the figure

TABLE 1 Setting I: conditional NNT(x), for $x = 1.5, 2, 2.5$, in the logistic regression model (for a formal definition of NNT(x) in the logistic regression model see Equation (7))

	NNT(1.5)				NNT(2)				NNT(2.5)			
N	DL	TR	NBS	PBS	DL	TR	NBS	PBS	DL	TR	NBS	PBS
200	1.00	1.00	0.94	0.93	1.00	1.00	0.95	0.93	1.00	1.00	0.94	0.59
400	1.00	1.00	0.94	0.93	1.00	1.00	0.94	0.92	1.00	1.00	0.95	0.80
800	1.00	1.00	0.94	0.94	1.00	1.00	0.94	0.94	1.00	1.00	0.94	0.92

Note: Mean coverage rates of the pointwise CIs by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation), and sample sizes of $n = 200,400,800$.

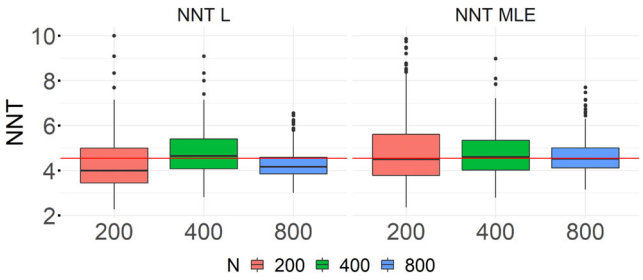

FIGURE 3 Harmonic NNT in the logistic regression model, for $n = 200,400,800$. Parametric MLE; NNT MLE. Laupacis' nonparametric MLE; NNT L

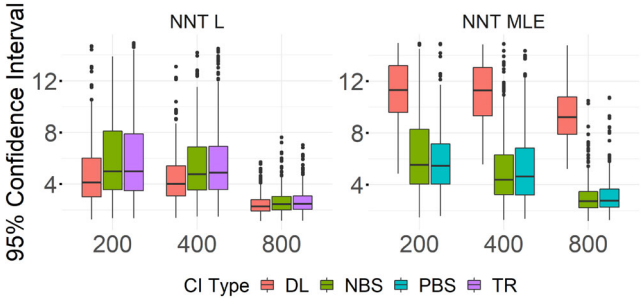

FIGURE 4 CI lengths by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation) of the harmonic NNT in the logistic regression model, using the parametric (NNT MLE) and the nonparametric (NNT L) MLEs, for $n = 200,400,800$. For the parametric MLE, the box-plots of transformation-based CIs were not displayed, since they are either infinite or too large, and thus distort the figure

TABLE 2 Setting I: harmonic NNTs in the logistic regression model

	NNT_{MLE}				NNT_L		
N	DL	TR	NBS	PBS	DL	TR	NBS
200	1.00	1.00	0.95	1.00	0.86	0.95	0.95
400	1.00	1.00	0.95	1.00	0.94	0.96	0.96
800	1.00	1.00	0.94	0.94	0.85	0.91	0.91

Note: Mean coverage rates of the pointwise CIs by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation), and sample sizes of $n = 200,400,800$. Based on the parametric maximum likelihood estimator, NNT_{MLE}, and the nonparametric maximum likelihood estimator, NNT_L.

distribution with $\lambda(x) = \exp\left\{-\frac{\beta_0+\beta_a+\beta_1 x}{\alpha}\right\}$, where $\beta_0 = 1$, $\beta_a = 2$, $\beta_1 = 1$, and $\alpha = 0.5$. Consequently, the conditional hazard function is

$$\lambda_a(y|x) = \lambda(y)\exp\left\{-\frac{\beta_{0a} + \beta_1 x}{\alpha}\right\},$$

such that $\beta_{0c} = 1$, and $\beta_{0t} = 3$. The beneficial outcome is defined as $I\{Y > y\}$. The conditional NNT$(y|x)$ can be calculated explicitly for every realization x, and any time point y. We chose a representative time point $y = 8$. For each time point y, we chose three representative values of x: -2.5, -2, and -1.5. The censoring mechanism follows the Weibull distribution with $\alpha = 0.01$, and $\beta_0 = 12$. The overall resulting censoring rate was approximately 58%. Particularly, the event rate in the control arm was approximately 40%, and in the treatment arm was approximately 45%. However, the mean follow-up time in the treatment arm was 5.75 time units, while in the control arm 1.03 time units. The true values of the conditional NNT$(8|x)$, are 9.9, 6.18, and 4.40, for x of -2.5, -2, and -1.5, respectively. The true value of the harmonic NNT(8), computed over the full covariate distribution used in the simulation for this time point, is 4.43. The point estimators of the conditional NNT$(8|x)$ and the corresponding CIs with their mean coverage rates are illustrated in Figures 5 and 6, and Table 3. The point estimators of the harmonic NNT(8) with the corresponding CIs and their mean coverage rates are

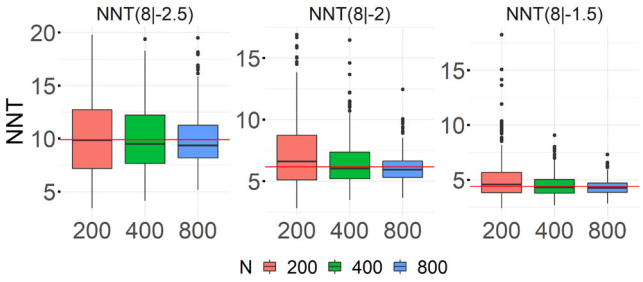

FIGURE 5 Semiparametric estimators of the conditional NNT$(y|x)$, for $y = 8$, and $x = -2.5, -2, -1.5$, using the Cox regression model (for a formal definition of NNT$(y|x)$ in survival analysis see Equation (23)), for $n = 200,400, 800$

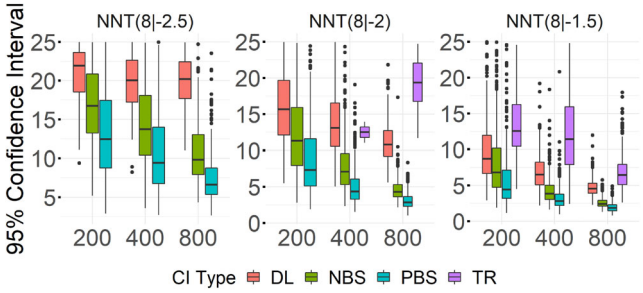

FIGURE 6 CI lengths by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation) of the conditional NNT$(y|x)$, for $y = 8$, and $x = -2.5, -2, -1.5$, using the Cox regression model, for $n = 200,400, 800$. The box-plots of certain transformation-based CIs were not displayed, since they are either infinite or too large, and thus distort the figure

TABLE 3 Setting II: conditional NNT$(y|x)$s, for $y = 8$, and $x = -2.5, -2, -1.5$, in the Cox regression model (for a formal definition of NNT$(y|x)$ in survival analysis see Equation (23))

	NNT(8| − 2.5)				NNT(8| − 2)				NNT(8| − 1.5)			
N	**DL**	**TR**	**NBS**	**PBS**	**DL**	**TR**	**NBS**	**PBS**	**DL**	**TR**	**NBS**	**PBS**
200	1.00	0.99	0.96	0.91	1.00	1.00	0.96	0.90	0.99	0.99	0.95	0.91
400	1.00	1.00	0.94	0.86	1.00	1.00	0.94	0.82	1.00	0.99	0.94	0.86
800	1.00	1.00	0.95	0.83	1.00	1.00	0.94	0.80	0.99	1.00	0.94	0.81

Note: Mean coverage rates of the pointwise CIs by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation), and sample sizes of $n = 200,400, 800$.

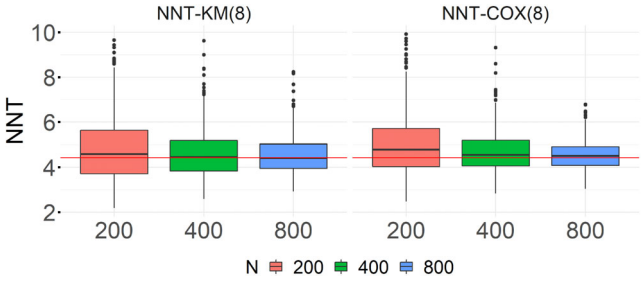

FIGURE 7 Harmonic NNT in the Cox regression model, for $y = 8$, and $n = 200,400,800$. Cox semiparametric estimator $\text{NNT}_{COX}(y)$; NNT-COX(8). Kaplan-Meier nonparametric MLE $\text{NNT}_{KM}(y)$; NNT-KM(8)

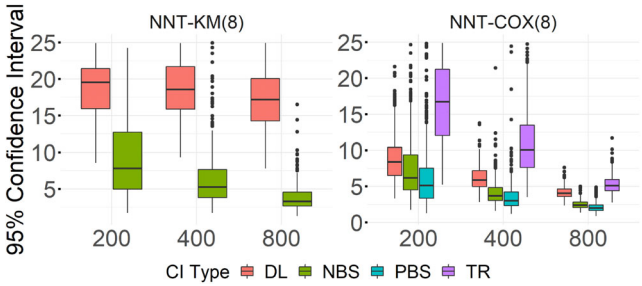

FIGURE 8 Pointwise CI lengths by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation) of the harmonic NNT(8) in survival analysis, based on the semiparametric estimator (NNT(8) Cox) and the nonparametric MLE (NNT(8) KM), for $n = 200,400,800$. For the nonparametric NNT(8), the box-plots of transformation-based CIs were not displayed, since they are either infinite or too large, and thus distort the figure

TABLE 4 Setting II: harmonic NNT(y), for $y = 8$, in the Cox regression model

	$\text{NNT}_{COX}(8)$				$\text{NNT}_{KM}(8)$		
N	**DL**	**TR**	**NBS**	**PBS**	**DL**	**TR**	**NBS**
200	0.99	0.97	0.94	0.96	1.00	1.00	0.94
400	1.00	0.99	0.95	0.92	1.00	1.00	0.95
800	0.99	0.99	0.94	0.92	1.00	1.00	0.93

Note: Mean coverage rates of the pointwise CIs by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation), and sample sizes of $n = 200,400,800$. Based on the semiparametric estimator, $\text{NNT}_{COX}(y)$, and the nonparametric maximum likelihood estimator, $\text{NNT}_{KM}(y)$, for $y = 8$.

presented in Figures 7 and 8, and Table 4. For Setting III of Simulation study that illustrates the linear regression model, please refer to the Appendix.

Simulation summary

1. In all settings: For the conditional NNTs, the parametric estimators converge to their true values. For the harmonic NNT, where the model is correctly specified, both the parametric and the nonparametric estimators converge to the true value of the parameter. The correctly specified parametric estimators appear to be more stable compared to the nonparametric alternative.
2. In all settings, the transformation-based CIs appear to be consistently larger compared to other alternatives. Frequently, infinitely large. Consequently, their coverage rates were frequently 100%. This can be explained by the behavior of the function g as defined in (1), in the vicinity of $x = 0$. Namely, for small positive values of x, g behaves as $1/x$; hence

large standard errors of the estimated probability of treatment benefit result in wide CIs. All the three other methods of CIs construction mitigate this sensitivity by taking the standard error of the transformation g itself, rather than transforming the standard error of the estimated probability of treatment benefit.

3. For all settings the bootstrap-based CIs, either parametric or nonparametric, were the most efficient CIs. Specifically, the parametric method mostly produced the tightest CIs, however with a certain extent of undercoverage, while the nonparametric-bootstrap-based CIs tended to be slightly larger and more accurate. The delta-method-based CIs were usually larger with a perfect or near perfect coverage rate. In addition, similarly to Laubender and Bender,[38] we observed that bootstrap-based CIs are computationally demanding tasks that consume considerable processing time.
4. For all settings, the larger the true NNT, the more biased and less stable were the point estimators and thus the CIs were less accurate. This stems from the convexity of g as defined in (1). The larger bias resulted in less accurate CIs. Since all CIs are symmetric w.r.t. the point estimators, their limits were also biased upwards, which resulted in lower coverage rates. The larger the true NNT, either conditional or harmonic, the larger sample size that is required in order to obtain more accurate point estimates and associated confidence limits.
5. The overall simulation results are consistent with the theoretical considerations (in Sections 1,2, and 3). Although all CI types are asymptotically-correct, the parametric CIs exploit the asymptotic normality of the MLE. In particular, the asymptotic normality of the ML estimators of NNT is derived from the asymptotic normality of the regression coefficients. There are three layers of approximation. The first layer is for the regression coefficients themselves since we use sample sizes under 1000 which may represent some RCTs in certain domains of research. The second is for the normality of the estimated probability of treatment benefit which is a nonlinear function of the regression coefficients. The third is for the normality of the estimated NNT itself, since it is a convex transformation g of the probability of treatment benefit. The normal approximation of the sample distribution of the NNT may thus require very large sample sizes to be accurate. The nonparametric BS-based CIs do not use the asymptotic normal distribution, rather just the empirical quantiles of the NNT's sample distribution. Therefore, they are not sensitive to deviations from the normality assumption. CIs constructed by nonparametric bootstrap conformed to their nominal confidence coefficient in the scenarios described.

5 | SUMMARY AND DISCUSSION

We have introduced a systematic framework to model and estimate the conditional and the harmonic NNT in the presence of explanatory variables, in various models, with dichotomous and nondichotomous outcomes. The conditional NNT was illustrated in a series of four examples: Logistic and linear regressions, alongside Kaplan-Meier and Cox-regression models. We established the relationship between the conditional and the harmonic NNT in the presence of explanatory variables. We introduced four different methods to calculate the asymptotically-correct CIs for both NNT measures. Additionally, we conducted a simulation study to provide a numerical illustration of the theoretical results with the four examples. The results indicate that the parametric MLE with nonparametric bootstrap-based CIs are the preferable estimators in all settings. For large NNT values, the point estimators were more biased and less stable. Transformation-based CIs tend to be too wide or even infinite with a perfect coverage rates. The delta-method CIs were usually finite, however larger than the bootstrap counterparts, with perfect or near perfect coverage rate. The bootstrap-based CIs were usually smaller, with the parametric bootstrap-based CIs frequently suffering from undercoverage, and the nonparametric attaining the expected coverage rates. For smaller NNT values, < 5, the bootstrap-based CIs, both parametric and nonparametric, performed well with approximately expected coverage rates, while the transformation-based and delta-method based CIs were usually much larger with perfect coverage rates. An R package and a corresponding web application[†] to calculate the conditional and the harmonic NNTs with their corresponding CIs, has been developed and made available for users online.

The NNT is not without limitations. NNT is an index for presenting results and not analyzing data. As pointed out by several authors,[7,11] the NNT is a one-dimensional index for conveying particular information and not a magical number to summarize comprehensive data analysis. Hence, it should not be the only statistic presented in a summary of analysis, and the user should acknowledge its caveats. However, the popularity of the NNT in various applications indicates the usefulness of this index and the need for an easy-interpretable statistic to convey the efficacy of an intervention. Unfortunately, as no clear methodological recommendations regarding the use of NNT have been formulated, calculations and interpretations of the NNT are sometimes misleading and even erroneous. Our work aims to address these problems by

providing concise statistical analysis, recommendations, and practical tools for appropriate use of the NNT in clinical trials.

Some scenarios that may arise in certain real-world data applications were not addressed in this article. Such scenarios include complex patterns of missing data and irregular or less frequent data structures, for example, longitudinal studies and time series with informative or nonmonotonic missingness. Another possible aspect that can be addressed is bias correction of the point estimators of the conditional and harmonic NNTs. These are directions for future research. Nonetheless, we have demonstrated how the NNT with its corresponding asymptotically-correct CIs can be effectively estimated in various widely-used statistical models, and provided the users with an R package and a web application to implement these calculations.

ENDNOTES

*The R package and the simulations source code: https://github.com/vancak/NNTcalculator.

†The web application: https://nntcalc.iem.technion.ac.il.

ORCID

Valentin Vancak https://orcid.org/0000-0001-8732-7353

REFERENCES

1. Newcombe RG. *Confidence Intervals for Proportions and Related Measures of Effect Size*. Boca Raton, FL: CRC Press; 2012.
2. Garg V, Shen X, Cheng Y, Nawarskas JJ, Raisch DW. Use of number needed to treat in cost-effectiveness analyses. *Ann Pharmacother*. 2013;47(3):380-387.
3. Lee TY, Kuo S, Yang CY, Ou HT. Cost-effectiveness of long-acting insulin analogues vs intermediate/long-acting human insulin for type 1 diabetes: A population-based cohort followed over 10 years. *Brit J Clin Pharmaco*. 2020;86(5):852-860.
4. Verbeek JG, Atema V, Mewes JC, et al. Cost-utility, cost-effectiveness, and budget impact of Internet-based cognitive behavioral therapy for breast cancer survivors with treatment-induced menopausal symptoms. *Breast Cancer Res Tr*. 2019;178(3):573-585.
5. de Costa BR, Rutjes AW, Johnston BC, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *Int J Epidemiol*. 2012;41(5):1445-1459.
6. Mendes D, Alves C, Batel-Marques F. Number needed to treat (NNT) in clinical literature: an appraisal. *BMC Med*. 2017;15(1):1-13.
7. Kristiansen IS, Gyrd-Hansen D, Nexøe J, Nielsen JB. Number needed to treat: easily understood and intuitively meaningful? theoretical considerations and a randomized trial. *J Clin Epidemiol*. 2002;55(9):888-892.
8. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New Engl J Med*. 1988;318(26):1728-1733.
9. Grieve AP. The number needed to treat: a useful clinical measure or a case of the Emperor's new clothes? *Pharm Stat*. 2003;2(2):87-102.
10. Hutton J. Number needed to treat: properties and problems. *J R Stat Soc A Stat*. 2000;163(3):381-402.
11. Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. *Stat Med*. 2011;30(19):2341-2348.
12. Sønbø Kristiansen I, Gyrd-Hansen D. Cost-effectiveness analysis based on the number-needed-to-treat: common sense or non-sense? *Health Econ*. 2004;13(1):9-19.
13. Röhmel J. On confidence bounds for the ratio of net differences in the "gold standard" design with reference, experimental, and placebo treatment. *Biom J*. 2005;47(6):799-806.
14. Vancak V, Goldberg Y, Levine SZ. Systematic analysis of the number needed to treat. *Stat Methods Med Res*. 2020;29(9):2393-2410.
15. Duncan BW, Olkin I. Bias of estimates of the number needed to treat. *Stat Med*. 2005;24(12):1837-1848.
16. Walter S. Number needed to treat (NNT): estimation of a measure of clinical benefit. *Stat Med*. 2001;20(24):3947-3962.
17. Senn S. Disappointing dichotomies. *Pharm Stat*. 2003;2(4):239-240.
18. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharm Stat*. 2009;8(1):50-61.
19. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407-415.
20. Senn S. Statistical pitfalls of personalized medicine; 2018.
21. Suissa S. The number needed to treat: 25 years of trials and tribulations in clinical research. *Rambam Maimonides Med J*. 2015;6(3):e0033.
22. Smeeth L, Haines A, Ebrahim S. Numbers needed to treat derived from meta-analyses—Sometimes informative, usually misleading. *BMJ*. 1999;318(7197):1548-1551.
23. Vancak V. nntcalc: the number needed to treat (NNT) calculator; 2020. R package version 1.1.
24. Vancak V, Goldberg Y, Levine SZ. Guidelines to understand and compute the number needed to treat. *Evid-Based Ment Health*. 2021;24(4):131-136.
25. Riegelman R, Schroth WS. Adjusting the number needed to treat: incorporating adjustments for the utility and timing of benefits and harms. *Med Decis Mak*. 1993;13(3):247-252.
26. Ebrahim S. The use of numbers needed to treat derived from systematic reviews and meta-analysis: caveats and pitfalls. *Eval Health Prof*. 2001;24(2):152-164.
27. Misselbrook D, Armstrong D. Thinking about risk. Can doctors and patients talk the same language? 2002.

28. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ*. 1999;319(7223):1492-1495.
29. Ola B, Papaioannou S, Afnan MA, Hammadieh N, Gimba S. Recombinant or urinary follicle-stimulating hormone? A cost-effectiveness analysis derived by particularizing the number needed to treat from a published meta-analysis. *Fertil Steril*. 2001;75(6):1106-1110.
30. Bender R, Blettner M. Calculating the "number needed to be exposed" with adjustment for confounding variables in epidemiological studies. *J Clin Epidemiol*. 2002;55(5):525-530.
31. Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. *J Clin Epidemiol*. 2010;63(1):2-6.
32. Bender R, Kuss O, Hildebrandt M, Gehrmann U. Estimating adjusted NNT measures in logistic regression analysis. *Stat Med*. 2007;26(30):5586-5595.
33. Bender R, Vervölgyi V. Estimating adjusted NNTs in randomised controlled trials with binary outcomes: a simulation study. *Contemp Clin Trials*. 2010;31(5):498-505.
34. 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(1):46-55.
35. Van der Vaart AW. *Asymptotic Statistics*. Vol 3. Cambridge, UK: Cambridge University Press; 1998.
36. Holland PW. Statistics and causal inference. *J Am Stat Assoc*. 1986;81(396):945-960.
37. Efron B. Bootstrap methods: another look at the jackknife. Vol 2. *Springer*. 1992;569-593.
38. 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(7-8):851-859.
39. Yang Z, Yin G. An alternative approach for estimating the number needed to treat for survival endpoints. *PLoS One*. 2019;14(10):e0223301.
40. Laubender RP, Bender R. A note on calculating asymptotic confidence intervals for the adjusted risk difference and number needed to treat in the Cox regression model. *Stat Med*. 2014;33(5):798-810.
41. Burr D. A comparison of certain bootstrap confidence intervals in the Cox model. *J Am Stat Assoc*. 1994;89(428):1290-1302.
42. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-481.
43. Greenwood M. A report on the natural duration of cancer; 1926:33.
44. Cox DR. *Analysis of Survival Data*. Boca Raton, FL: Chapman and Hall/CRC Press; 2018.
45. Breslow NE. Discussion of the paper by Dr. *Cox J Roy Stat Soc B Methodol*. 1972;34:216-217.
46. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;4:1100-1120.
47. Tsiatis AA. A large sample study of Cox's regression model. *Ann Stat*. 1981;9(1):93-108.
48. Moriña D, Navarro A. The R package survsim for the simulation of simple and complex survival data. *J Stat Softw*. 2014;59(2):1-20.

How to cite this article: Vancak V, Goldberg Y, Levine SZ. The number needed to treat adjusted for explanatory variables in regression and survival analysis: Theory and application. *Statistics in Medicine*. 2022;41(17):3299-3320. doi: 10.1002/sim.9418

APPENDIX

Lemma 1. *Let $\hat{\theta}_n$ be the MLE of θ_0, and let $g(\mathbb{E}[p_s(X;\theta_0)]$ and $g(p_s(x;\theta_0))$ be the NNT and the NNT(x), respectively. Therefore,*

i. $Var\left(\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i_a,a} p_s(x_{i_a};\hat{\theta}_n\right) - g\left(\mathbb{E}[p_s(X;\theta_0)]\right)\right)\right) \xrightarrow{a.s.} \|\mathbf{I}^{-1/2}(\theta_0)\nabla g(\mathbb{E}[p_s(X;\theta_0)])\|,$

ii. $Var\left(\sqrt{n}\left(g\left(p_s(x;\hat{\theta}_n\right)\right) - g\left(p_s(x;\theta_0)\right)\right) \xrightarrow{a.s.} \|\mathbf{I}^{-1/2}(\theta_0)\nabla g(p_s(x;\theta_0))\|,$

where g is defined in (1), and $\mathbf{I}(\theta_0)$ is the Fisher information matrix.

Proof. Let assumptions A.1-A.5 hold. For every $x \in \mathcal{X}$, such that $p_s(x;\theta_0) > 0$, the gradient of g is

$$\nabla_\theta g(p_s(x;\theta)) = -g^2(p_s(x;\theta))\begin{pmatrix} \frac{\partial}{\partial\theta_1}p(x;\theta_t) - \frac{\partial}{\partial\theta_1}p(x;\theta_c) \\ \vdots \\ \frac{\partial}{\partial\theta_p}p(x;\theta_t) - \frac{\partial}{\partial\theta_p}p(x;\theta_c) \end{pmatrix}$$

which is estimated by replacing θ_0 with its MLE. Since $g(\mathbb{E}[p_s(X;\theta)])$ is bounded in the neighborhood of θ_0, we can interchange the derivative and the expectation, that is, $\partial/\partial\theta_j g(\mathbb{E}[p_s(X;\theta)]) = -g^2(\mathbb{E}[p_s(X;\theta)])\mathbb{E}[\partial/\partial\theta_j p_s(X;\theta)]$, or explicitly

$$\nabla_\theta g(\mathbb{E}[p_s(X;\theta)]) = -g^2(\mathbb{E}[p_s(X;\theta)])\mathbb{E}\begin{pmatrix} \frac{\partial}{\partial\theta_1}p(X;\theta_t) - \frac{\partial}{\partial\theta_1}p(X;\theta_c) \\ \vdots \\ \frac{\partial}{\partial\theta_p}p(X;\theta_t) - \frac{\partial}{\partial\theta_p}p(X;\theta_c), \end{pmatrix}.$$

This gradient can be estimated by replacing the expectation operator with a corresponding sample mean, namely,

$$\nabla g\left(\widehat{\mathbb{E}[p_s(X;\theta)]}\right) = -g^2\left(\frac{1}{n}\sum_{i_a,a} p_s(x_{i_a};\hat{\theta}_n)\right)\frac{1}{n}\sum_{i_a,a}\nabla p_s(x_{i_a};\theta)\Big|_{\theta=\hat{\theta}_n}. \tag{A1}$$

Therefore, the asymptotic variance of NNT and NNT(x) can be derived by an application of the delta method to the asymptotic distribution of $\hat{\theta}_n$. ∎

Proof of Theorem 1

Proof. Let assumptions A.1-A.5 hold, and let $\hat{\theta}_n$ be the MLE of θ_0. Then,

1. Both $g(p_s(x;\hat{\theta}_n))$ and $g\left(\frac{1}{n}\sum_{i_1,a} p_s(x_{i_a};\hat{\theta}_n)\right)$ are functions of $\hat{\theta}_n$; hence by the invariance property of the MLEs,[35] they are the MLEs of NNT(x) and NNT, respectively.
2. Both estimators are continuous mappings of $\hat{\theta}_n$. Therefore, by the continuous mapping theorem,[35] $g(p_s(x;\hat{\theta}_n))$ and $g\left(\frac{1}{n}\sum_{i_1,a} p_s(x_{i_a};\hat{\theta}_n)\right)$ are strongly consistent estimators of NNT(x) and NNT, respectively.
3. Since $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \mathbf{I}^{-1}(\theta_0))$, hence by the delta method,[35] and Lemma 1, for every $x \in \mathcal{X}$, such that $p_s(x) > 0$,
   i. $\sqrt{n}\left(g\left(\frac{1}{n}\sum_{i_a,a} p_s(x_{i_a};\hat{\theta}_n)\right) - NNT\right) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\theta_0)\nabla g(\mathbb{E}[p_s(X;\theta_0)])\|Z$,
   ii. $\sqrt{n}\left(g\left(p_s(x;\hat{\theta}_n)\right) - NNT(x)\right) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\theta_0)\nabla g(p_s(x;\theta_0))\|Z$,
   where $\nabla g(\mathbb{E}[p_s(X;\theta_0)])$ and $\nabla g(p_s(x;\theta_0))$ are the gradients of NNT and NNT(x), respectively, w.r.t. θ, evaluated at the true parameter(s) θ_0.

∎

Proof of Theorem 3

Proof. Let assumptions A.2 and A.6 hold, and assume that the survival time Y is independent of the right-censoring time C. Then for every fixed time point $y \in [0, \eta)$,

1. By the invariance property of the MLE,[35] $\text{NNT}_{KM}(y)$ is the nonparametric MLE of the NNT(y).
2. By the continuous mapping theorem,[35] strong consistency of the MLE is preserved under continuous mappings; hence $\text{NNT}_{KM}(y) \xrightarrow{a.s.} \text{NNT}(y)$.
3. By the asymptotic normality of the $\text{NNT}_{KM}(y)$ and the delta method,[35] for every fixed time point $y \in [0, \eta)$, such that $p_s(y) > 0$, $\sqrt{n}(NNT_{KM}(y) - NNT(y)) \xrightarrow{D} \sigma_0(y)Z$, where $\sigma_0^2(y)$ is the asymptotic variance of $\sqrt{n}\text{NNT}_{KM}(y)$.

∎

Lemma 2. *In the absence of censoring, for every fixed time point $y \in [0, \eta)$, $NNT_{KM}(y)$ coincides with NNT_L, where the beneficial outcome is defined as $I\{Y > y\}$.*

Proof. Let the time to event be Y, and let the $p_a^{KM}(y)$, for $a \in \{c, t\}$, be defined as in (18). If no censoring occurs, $n_{ja} - d_{ja} = n_{j+1,a}$, namely the number of clinical trial study participants at risk in time y_{j+1} is the number of clinical trial study participants at risk minus the total number of events (deaths) in time y_j. Therefore,

$$\begin{aligned} p_a^{KM}(y) &= \prod_{y_j \le y}\left(1 - \frac{d_{ja}}{n_{ja}}\right) \\ &= \prod_{y_j \le y}\left(\frac{n_{ja} - d_{ja}}{n_{ja}}\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{n_{1a} - d_{1a}}{n_{1a}} \cdot \frac{n_{2a} - d_{2a}}{n_{2a}} \cdots \frac{n_{j-1,a} - d_{j-1,a}}{n_{j-1,a}} \cdot \frac{n_{j,a} - d_{j,a}}{n_{j,a}} \\
&= \frac{n_{2a}}{n_{1a}} \cdot \frac{n_{3a}}{n_{2a}} \cdots \frac{n_{j,a}}{n_{y-1,a}} \cdot \frac{n_{j+1,a}}{n_{j,a}} \\
&= \frac{n_{j+1,a}}{n_{1,a}} \\
&= \frac{1}{n_a} \sum_{i=1}^{n_a} I\{Y_i > y\},
\end{aligned}
$$

the last equality stems from the fact that the total number of clinical trial study participants at risk at time y_1, denoted by n_{1a}, is the total number of clinical trial study participants in the ath arm. In addition, $n_{y+1,a}$ is the total number of clinical trial study participants in the ath arm that did not have an event up to time y, that is, $n_{y+1,a} = \sum_{i=1}^{n_a} I\{Y_i > y\}$. Therefore, for uncensored data, $\text{NNT}_{KM}(y)$ is the function (1) applied to the sample proportions, which is, by definition, the NNT_L with threshold $\tau = y$. ∎

Proof of Theorem 5

Proof. Let assumptions A.1-A.8 hold, and let $\hat{\beta}_n$ be the MPLEs of β_0 in the Cox model. Therefore, for any fixed time point $y \in [0, \eta)$, and for all $x \in \mathcal{X}$,

1. Since $\hat{\beta}_n \xrightarrow{a.s} \beta_0$, thus by the continuous mapping theorem[35] $\text{NNT}_{COX}(y|x) \xrightarrow{a.s.} \text{NNT}(y|x)$.
2. Since $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\beta_0)\|Z$,[47] therefore by the delta method,[35] and Lemma 1, for every fixed time point $y \in [0, \eta)$, such that $p_s(y|x) > 0$, and $\mathbb{E}[p_s(y|X)] < \infty$,
   i. $\sqrt{n}\,(NNT_{COX}(y|x) - NNT(y|x)) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\beta_0)\nabla g(p_s(y|x;\beta_0))\|Z$,
   ii. $\sqrt{n}\,(NNT_{COX}(y) - NNT(y)) \xrightarrow{D} \|\mathbf{I}^{-1/2}(\beta_0)\nabla g\,(\mathbb{E}[p_s(y|X;\beta_0)])\,\|Z$,

   where $\nabla g(p_s(y|x;\beta_0))$ and $\nabla g(\mathbb{E}[p_s(y|X;\beta_0)])$ are the gradients of $\text{NNT}(y|x)$ and $\text{NNT}(y)$, respectively, evaluated at β_0, and $\mathbf{I}(\beta_0)$ is the true Fisher information matrix. ∎

Sample variance of NNT(y) and NNT($y|x$) in Theorems 4, and 6

In this section we calculate the sample variances in the CIs that were presented in Theorems 4 and 6. For the calculation of the sample variance that was presented in Theorem 2, see Lemma 1.

Calculation of the sample variance of $\text{NNT}_{KM}(y)$ in Theorem 4

Let assumptions A.2 and A.6 hold. Then, by application of the delta method[35] to the sample variance of $p_t^{KM}(y) - p_c^{KM}(y)$, which is estimated using the Greenwood's formula,[43] we derive the sample variance of $\text{NNT}_{KM}(y)$

$$
\widehat{Var}(NNT_{KM}(y)) = NNT_{KM}^4(y) \left(\left(p_t^{KM}(y)\right)^2 \sum_{y_j \le y} \frac{d_{jt}}{n_{jt}(n_{jt} - d_{jt})} + \left(p_c^{KM}(y)\right)^2 \sum_{y_j \le y} \frac{d_{jc}}{n_{jc}(n_{jc} - d_{jc})} \right).
$$

Consequently, it can be shown[44] that in the absence of censoring, $\text{NNT}_{KM}(y)$'s Greenwood-type CI coincides with NNT_L's Wald-type CI with threshold $\tau = y$.

Derivation of the sample variance of $\text{NNT}_{COX}(y|x)$ in Theorem 6

We divide the proof into two different models: (1) no treatment by covariate interaction model, and (2) treatment by covariate interaction model.

Proof. Let assumptions A.1-A.8 hold, and let the probability of treatment benefit be defined as in (23), namely $p_s(y|x;\beta_0) = S(y)^{\exp\{\beta_t^T x\}} - S(y)^{\exp\{\beta_c^T x\}}$, where $\beta_0^T = (\beta_c^T, \beta_t^T)$. The variance of $\text{NNT}_{COX}(y|x)$ is based on the sample variance of $\hat{p}_s$. Consequently, the sample variance is based both on the estimators of the baseline hazard $S(y)$, and the regression coefficients β_0.

Since the Breslow estimator of the baseline hazard is based on the estimators of β_0, the sample variance of $\text{NNT}_{COX}(y|x)$, at a fixed time point y, is based on the sample variance of β_0's MPLEs and the functional form of $g(p_s(y|x;\beta_0))$. ■

No treatment by covariate interaction model

Let $\beta_0^T = (\beta_1, \ldots, \beta_p, \beta_\alpha)$ be the regression coefficients of the model. These coefficients correspond to a set of explanatory variables $(X_1, \ldots, X_p, A)^T$, where A is an indicator variable, such that A equals 1 if the clinical trial study participant is in the treatment arm t, and 0 if the clinical trial study participant is in the control arm c. The gradient of $p_s(y|x;\beta_0)$ consists of two main summands (parts). The first part, designated by $m_1(y;x,\hat{\beta})$, is derived from the parametric structure of the Cox model, and is given by

$$m_1(y;x,\hat{\beta}) \equiv \nabla_\beta p_s(y|x)\Big|_{\beta_0^T=\hat{\beta}^T} = -\hat{\Lambda}(y;\hat{\beta})\big(k_1(y|\hat{\beta};x), \ldots, k_p(y|\hat{\beta};x), k_\alpha(y|\hat{\beta};1)\big)^T,$$

where

$$k_l(y|\hat{\beta};x) = \left(p_t(y|x,\hat{\beta})e^{x^T\hat{\beta}} - p_c(y|x,\hat{\beta})e^{x^T\hat{\beta}}\right)x_l, \quad l = 1, \ldots, p$$

and

$$k_\alpha(y|\hat{\beta};1) = p_t(y|x,\hat{\beta})e^{x^T\hat{\beta}},$$

for $A = 1$. The second part, designated by $m_2(y;x,\hat{\beta})$, is derived from the Breslow[45] NPMLE of the baseline hazard, and is given by

$$m_2(y;x,\hat{\beta}) = \big(k_1(y|x,\hat{\beta}), \ldots, k_p(y|x,\hat{\beta}), k_t(y|x,\hat{\beta})\big)^T,$$

where the first p entries of this vector are

$$k_l(y|x,\hat{\beta}) = \left(p_t(y|x,\hat{\beta})e^{x'\hat{\beta}} - p_c(y|x,\hat{\beta})e^{x'\hat{\beta}}\right)\Lambda'_{\beta_l}(y;\hat{\beta}), \quad l = 1, \ldots, p,$$

and the last term is

$$k_t(y|x,\hat{\beta}) = p_t(y|x,\hat{\beta})e^{x'\hat{\beta}}\Lambda'_{\beta_\alpha}(y;\hat{\beta}).$$

Some algebra yields

$$\Lambda'_{\beta_l}(y;\hat{\beta}) = \sum_{y_{ja}\le y}\left[\left(\frac{d_{ja}}{\sum_{a\in\{c,t\}}\sum_{j\in R_a(y_j)}\exp\left\{\sum_{l=1}^p\hat{\beta}_l x_{alj}\right\}}\right)^2\sum_{a\in\{c,t\}}\sum_{j\in R_a(y_j)}\exp\left\{\sum_{l=1}^p\hat{\beta}_l x_{alj}\right\}x_{alj}\right],$$

for $l = 1, \ldots, p$, and $a \in \{c,t\}$, and

$$\Lambda'_{\beta_\alpha}(y;\hat{\beta}) = \sum_{y_{ja}\le y}\left[\left(\frac{d_{ja}}{\sum_{a\in\{c,t\}}\sum_{j\in R_a(y_j)}\exp\left\{\sum_{l=1}^p\hat{\beta}_l x_{alj}\right\}}\right)^2\sum_{j\in R_t(y_j)}\exp\left\{\sum_{l=1}^p\hat{\beta}_l x_{tlj}\right\}\right].$$

Treatment by covariate interaction model

Let β_t^T and β_c^T be variationally independent p-dimensional vectors of unknown parameters. In this case

$$m_1(y;x;\hat{\beta}) = \nabla_\beta p_s(y|x)|_{\beta_0^T=\hat{\beta}^T} = \hat{\Lambda}(y;\hat{\beta})\big(-k_{t1}(y|\hat{\beta}_t;x_1), \ldots, -k_{tp}(y|\hat{\beta}_t;x_p), k_{c1}(y|\hat{\beta}_c;x_1)\ \ldots, k_{cp}(y|\hat{\beta}_c;x_p)\big)^T,$$

where

$$k_{al}(y|\hat{\beta}_a;x_l) = p_a(y|x;\hat{\beta}_a)e^{x'\hat{\beta}_a}x_l, \quad l = 1, \ldots, p, \quad \text{and} \quad a \in \{c,t\},$$

and

$$m_2(y;x,\hat{\beta}) = \left(p_t(y|x;\hat{\beta}_t)e^{x^T\hat{\beta}_t} - p_c(y|x;\hat{\beta}_c)e^{x^T\hat{\beta}_c}\right)\nabla_\beta\Lambda(y;\hat{\beta})^T.$$

The generalization to include interactions is straightforward. Notably, the second part of $p_s(y|x;\beta)$'s estimated gradient, $m_2(y;x;\hat{\beta})$, was originally neglected in the $\text{NNT}_{COX}(y|x)$'s delta-method-based CI in Laubender and Bender,[38] and only later on Reference 40 estimated using the counting processes and the theory of martingales.

Consequently, the sample variance of the estimated $p_s(y|x;\beta_0)$ in the Cox model, can be approximated by $\|\mathbf{I}^{-1/2}(\hat{\beta})(m_1(y;x,\hat{\beta})+m_2(y;x,\hat{\beta}))\|^2$. Therefore, by application of the delta method[35] to the asymptotic distribution of $p_s(y|x,\hat{\beta})$,[47] we derive the estimated sample variance of the conditional $\text{NNT}_{COX}(y|x)$

$$\widehat{Var}(NNT_{COX}(y|x)) = NNT^4_{COX}(y|x,\hat{\beta})\|\mathbf{I}^{-1/2}(\hat{\beta})(m_1(y;x,\hat{\beta})+m_2(y;x,\hat{\beta}))\|^2. \tag{A2}$$

In order to obtain the estimated sample variance of the harmonic $\text{NNT}_{COX}(y)$, one needs to replace $p_s(y|x;\hat{\beta})$ and $\nabla p_s(y|x;\beta)\Big|_{\beta^T=\hat{\beta}^T}$ with their corresponding sample averages as presented in (A1).

Simulation study Setting III: Linear regression

In this setting, the explanatory variable X is a normally distributed variable with $\mu_X = 3$ and $\sigma^2 = 1.5^2$. The response variable Y, given realization x and an allocated arm $a \in \{c,t\}$, follows a normal distribution $N(\beta_0+\beta_a x, 1)$, where $\beta_0 = 1$, $\beta_c = 0.5$, and $\beta_t = 1$. The MCID threshold τ is 3. In other words, we define the beneficial outcome as $I\{Y > 3\}$. The conditional NNT(x) can be calculated for every x. However for this illustration we chose only three representative values: 1.2, 1.3, and 1.4. The true values of NNT(1.2), NNT(1.3), and NNT(1.4) are 7.63, 6.52, and 5.64, respectively. The true value of the harmonic NNT, computed over the full covariate distribution used in the simulation, is 2.65. The NNT(x)'s, $x = 1.2, 1.3, 1.4$, point estimators and the corresponding CIs with their mean coverage rates are illustrated in Figures A1 and A2, and Table A1. The point estimators of the harmonic NNTs with the corresponding CIs and their mean coverage rates are presented in Figures A3 and A4, and Table A2.

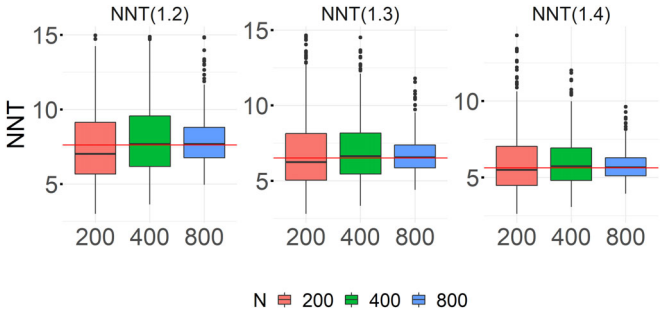

FIGURE A1 Parametric MLEs of the conditional NNT(x), $x = 1.2, 1.3, 1.4$, in the linear regression model with normally distributed error term (for a formal definition of NNT(x) in the linear regression model see Equation (9)), for $n = 200,400,800$

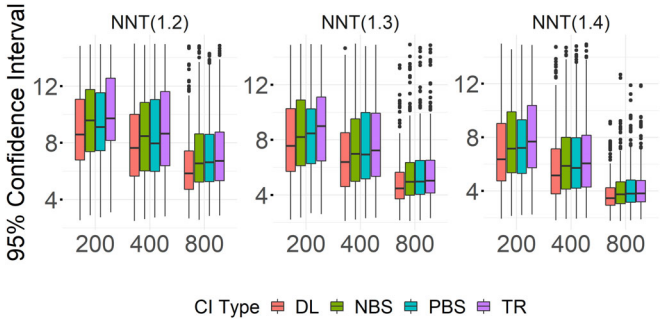

FIGURE A2 CI lengths by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation) of the conditional NNT(x), $x = 1.2, 1.3, 1.4$, in the linear regression model with normally distributed error term, for $n = 200,400,800$. The box-plots of certain transformation-based CIs were not displayed, since they are either infinite or too large, and thus distort the figure

TABLE A1 Setting III: conditional NNT(x), for $x = 1.2, 1.3, 1.4$, in the linear regression model with normally distributed error (for a formal definition of NNT(x) in the linear regression model see Equation (9))

	NNT(1.2)				NNT(1.3)				NNT(1.4)			
N	DL	TR	NBS	PBS	DL	TR	NBS	PBS	DL	TR	NBS	PBS
200	0.91	0.93	0.93	0.87	0.91	0.93	0.93	0.87	0.92	0.93	0.94	0.92
400	0.91	0.94	0.93	0.89	0.91	0.94	0.92	0.89	0.91	0.94	0.93	0.91
800	0.96	0.96	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.96

Note: Mean coverage rates of the pointwise CIs by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation), and sample sizes of $n = 200,400, 800$.

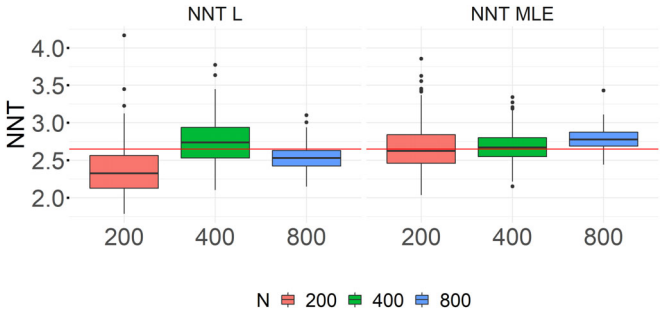

FIGURE A3 Harmonic NNT in the linear regression model with normally distributed error term, for $n = 200,400, 800$. Parametric MLE; NNT MLE. Laupacis' nonparametric MLE; NNT L

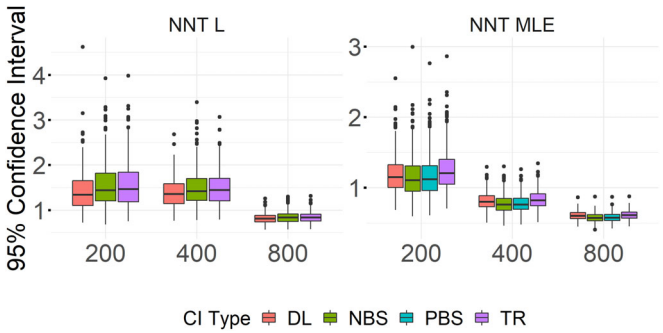

FIGURE A4 Pointwise CI lengths by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation) of the harmonic NNT in the linear regression model with normally distributed error term, using the parametric (NNT MLE) and the nonparametric (NNT L) MLEs, for $n = 200,400, 800$

TABLE A2 Setting III: harmonic NNTs in the linear regression model with normally distributed error

	NNT_{MLE}				NNT_L		
N	DL	TR	NBS	PBS	DL	TR	NBS
200	0.90	0.95	0.92	0.89	0.73	0.82	0.83
400	0.93	0.95	0.92	0.92	0.97	0.99	0.98
800	0.98	0.97	0.95	0.98	0.82	0.90	0.90

Note: Mean coverage rates of the pointwise CIs by construction method (DL, delta method; NBS, nonparametric bootstrap; PBS, parametric bootstrap; TR, transformation), and sample sizes of $n = 200,400, 800$. Based on the parametric maximum likelihood estimator, NNT_{MLE}, and the nonparametric maximum likelihood estimator, NNT_L.