Sample size calculation for the Andersen-Gill model comparing rates of recurrent events

Yongqiang Tang, Ronan Fitzpatrick

Summary

Recurrent events arise frequently in biomedical research, where the subject may experience the same type of events more than once. The Andersen-Gill (AG) model has become increasingly popular in the analysis of recurrent events particularly when the event rate is not constant over time. We propose a procedure for calculating the power and sample size for the robust Wald test from the AG model in superiority, noninferiority and equivalence clinical trials. Its performance is demonstrated by numerical examples. Sample SAS code is provided in the supplementary material.

KEYWORDS:
Mixed Poisson process; Noninferiority and equivalence trials; Overdispersion; Proportional rates/means model; Sandwich variance

1 INTRODUCTION

Recurrent events are frequently encountered in biomedical research, where the subject may experience the same type of events more than once. Examples include attacks in hereditary angioedema, exacerbations in chronic obstructive pulmonary disease, bleeds in hemophilia, relapses in multiple sclerosis, and infections in chronic granulomatous disease (CGD). In clinical trials, the recurrent events are commonly analyzed by the negative binomial (NB) regression. The NB regression assumes constant event rates over time, which may fail to hold in some applications. The Andersen-Gill (AG) model provides a popular alternative tool for the analysis of recurrent events, and it allows arbitrary event rate functions. The AG model often yields similar treatment effect estimates (i.e. ratio of event rates between groups) to the NB regression in empirical studies when the event rate is roughly constant over time.

Sample size calculation is critical in designing a clinical trial to ensure sufficient power to detect an important treatment effect. Sample size methodology has been well developed for the NB regression; Please see Tang and references therein. Matsui and Song et al derive sample size formulae for the robust log-rank test, which is a nonparametric test suitable only for superiority trials. In this paper, we propose a power and sample size calculation procedure for the robust Wald test from the AG model. It is applicable to superiority, noninferiority (NI) and equivalence trials. Two designs are considered. In one design, the planned treatment duration is the same for all subjects. In the other design, subjects are enrolled at different calendar time, but administratively censored at the same calendar time. We introduce the sample size procedure in Section 2, and assess its performance numerically in Section 3.

The paper was published in Statistics in Medicine 2019 (Volume 38, Issue 24, Pages 4819 - 4827). There was an error in Equation A4 in the appendix. It does not affect design 1, but appears to slightly overestimate the sample size for design 2 with staggered entry. The result becomes better after the correction in the sense that the nominal power generally becomes closer to the simulated power for design 2. The corrected contents were highlighted in red.
2 | POWER AND SAMPLE SIZE FORMULAE

Andersen and Gill\textsuperscript{12} provide a simple extension of the Cox proportional hazards model to the analysis of recurrent events. Suppose \( n \) subjects are randomized to either the active (\( x_i = 1 \)) or control (\( x_i = 0 \)) treatment in a clinical trial. Let \( T_i \) be the follow-up time for subject \( i \), \( Y_i(t) = I(T_i \geq t) \) the indicator function that subject \( i \) is still under observation at time \( t \), and \( N_i(t) \) the number of events experienced by subject \( i \) by time \( t \). Inference for the event rate ratio \( \exp(\beta) \) between treatment groups is based on the following partial likelihood

\[
PL(\beta) = \prod_{i=1}^n \prod_{Y_i(t) = 1} \left[ \frac{\exp(\beta x_i)}{\sum_{j=1}^n Y_j(t) \exp(\beta x_j)} \right] dN_i(t).
\]

An attractive feature of the AG model is that the baseline event rate function can be of arbitrary shape. We assume a constant event rate over time, but the AG model can handle time-varying treatment effects.

To obtain the maximum likelihood estimate (MLE) \( \hat{\beta} \), we solve the score function

\[
U(\beta) = \frac{\partial \log[PL(\beta)]}{\partial \beta} = \sum_{i=1}^n \int_0^\tau [x_i - \tilde{\beta}(\hat{x}_i, t)]dN_i(t) = 0,
\]

where \( S^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n Y_i(t)x_i^k \exp(\beta x_i) \), \( \tilde{\beta}(\hat{x}_i, t) = \frac{\sum_{j=0}^k S^{(j)}(\beta, t) x_j}{\sum_{j=0}^k S^{(j)}(\beta, t)} \), and \( \tau \) is the maximum treatment duration in the trial.

If all covariates are time invariant (the covariates measured after randomization are rarely used to assess the treatment effect in clinical trials since the covariates may be affected by the treatment), the AG model assumes that the time increments between events are independent according to a Poisson process, but the recurrent events are generally dependent within a subject.\textsuperscript{12} The Poisson-type assumption can be relaxed by using the sandwich variance estimator, and the validity of this robust approach is justified by Lin et al\textsuperscript{14} for arbitrary dependence structures among recurrent events if the proportional rate or mean assumption is met. For this reason, the robust approach is also called the proportional rates/means model. The sandwich variance estimate\textsuperscript{11} for \( \hat{\beta} \) is

\[
\hat{\beta} = n^{-1} \hat{V}_\beta = n^{-1} \hat{\tilde{I}}_\beta \hat{\Sigma} \hat{\tilde{I}}_\beta^{-1}, \quad \hat{\Sigma}_\beta = \sum_{i=1}^n \int_0^\tau \left[ x_i - \tilde{\beta}(\hat{x}_i, t) \right]d\hat{M}_i(t), \quad \hat{\tilde{I}}_\beta = n^{-1} \sum_{i=1}^n \hat{\Sigma}_\beta^{-1} \hat{\tilde{I}}_\beta^{-1} \hat{\tilde{I}}_\beta.
\]

The two-sided 100(1 - \( \alpha \))\% confidence interval (CI) for \( \beta \) is

\[
[c_{\alpha}, c_{\alpha}] = [\hat{\beta} - z_{1-\alpha/2} \sqrt{n^{-1} \hat{V}_\beta}, \hat{\beta} + z_{1-\alpha/2} \sqrt{n^{-1} \hat{V}_\beta}],
\]

where \( z_{\alpha} \) is the \( \alpha \)-th percentile of the standard normal distribution \( N(0, 1) \).

In the sample size determination, we assume a mixed Poisson process (MPP) model\textsuperscript{4,9,12} for the event process. Let \( \Lambda_g(t) = E[N_i(t)|x_i = g] \) be the mean event function for group \( g \). The MPP introduces a random effect \( \epsilon_i \) with mean 1 and variance \( \kappa_g \) for each subject. Given \( \epsilon_i \), the subject in group \( g \) follows a Poisson process with mean function \( \epsilon_i \kappa_g \). Subjects with \( \epsilon_i < 1 \) (\( \epsilon_i > 1 \)) tend to experience more (less) events than the average in the population. The dispersion parameter \( \kappa_g \) measures the between-subject heterogeneity. Inclusion of important risk factors in the model may reduce heterogeneity\textsuperscript{12}. The MPP provides a natural way to handle overdispersion in recurrent events in that the variance of \( N_i(t) \) is larger than its mean\textsuperscript{12}. The mixing distribution for the random effect \( \epsilon_i \) is unspecified in the AG model. The NB regression uses a gamma mixing distribution, and the event count \( N_i(t) \) follows the NB distribution\textsuperscript{4}.

In Appendix A\textsuperscript{11}, we show that \( \hat{V}_\beta \) converges in probability to \( V_\beta \)

\[
V_\beta = \frac{p_g[A_1 + \kappa_g B_1] + p_0[A_0 + \kappa_0 B_0]}{\left( \int_0^\tau \left[ \frac{p_g}{p_g \pi_g(t) + p_0 \pi_0(t)} \right] \exp(\beta \pi_g(t))d\lambda_g(t) \right)^2},
\]

where \( p_g \) is the proportion of subjects randomized to treatment group \( g \), \( \pi_g(t) \) is the probability that a subject in group \( g \) remains in the study at time \( t \), \( \omega_g(t) = \frac{p_g \pi_g(t) \exp(\beta \pi_g(t))}{p_g \pi_g(t) + p_0 \pi_0(t)} \), \( \omega_g(t) = 1 - \omega_0(t) \), \( A_g = \int_0^\tau \omega_g(t) \pi_g(t) d\lambda_g(t) \), and \( B_g = 2 \int_0^\tau \left[ \int_{s=0}^t \omega_g(s) d\lambda_g(s) \right] \pi_g(t) \omega_g(t) d\lambda_g(t) \). We allow the loss to follow-up distribution \( G_g(t) = 1 - \pi_g(t) \) and the dispersion parameter \( \kappa_g \) to differ between the two treatment groups.

At the design stage, it is often reasonable to assume the same dropout distribution in the two treatment groups (i.e. \( \pi_1(t) = \pi_0(t) \) for all \( t \)), and \( V_\beta \) reduces to

\[
V_\beta = \frac{1}{p_1 E_1} + \frac{1}{p_0 E_0} + 2 \left( \frac{k_1 F_1}{p_1 E_1^2} + \frac{k_0 F_0}{p_0 E_0^2} \right) = \left[ \frac{1}{p_1 \exp(\beta)} + \frac{1}{p_0} \right] \frac{1}{E_0} + \left[ \frac{k_1}{p_1} + \frac{k_0}{p_0} \right] \frac{2F_0}{E_0^2},
\]
where \( E_g = \int \pi_g(t) d\Lambda_g(t) \) and \( F_g = \int_0^\tau \pi_g(t) d\Lambda_g(t) \). In general, formula (2) can be well approximated by the term between the two equal signs in formula (3) if the dropout distribution differs between the two groups.

In Appendix A.2, we provide analytic expressions of \( E_g \) and \( F_g \) for the Weibull and piecewise constant event rate functions when the dropout pattern is identical in the two groups in two types of clinical trial designs. In practical applications, almost any event rate function can be approximated reasonably well by the piecewise constant function.

### 2.1 Superiority and NI trials

Suppose a lower event rate is desirable. In both superiority and NI trials, the hypothesis can be written as

\[
H_0: \exp(\beta) \geq M_0 \text{ or } \beta \geq \log(M_0) \text{ versus } H_1: \exp(\beta) < M_0 \text{ or } \beta < \log(M_0). \tag{4}
\]

In a superiority trial, the objective is to demonstrate that the experimental treatment can lower the event rate, and we set \( M_0 = 1 \). The NI trial aims to show that the experimental treatment is not worse than the standard control treatment by \( M_0 \), where \( M_0 > 1 \) is the prespecified NI margin on the rate ratio.

The power for test (4) is given by

\[
\Pr(c_a < \log(M_0)) = \Pr \left[ Z < \frac{-z_{1-\alpha/2} \sqrt{n^{-1} V_\beta} - \beta + \log(M_0)}{\sqrt{n^{-1} V_\beta}} \right] \approx \Phi \left[ \frac{\sqrt{n} \log(M_0) - \beta}{\sqrt{V_\beta}} - z_{1-\alpha/2} \right], \tag{5}
\]

where \( Z = (\hat{\beta} - \beta)/\sqrt{n^{-1} V_\beta} \) is asymptotically distributed as \( N(0, 1) \). The required sample size is

\[
n = \frac{(z_{1-\alpha/2} + z_\alpha)^2 V_\beta}{[\log(M_0) - \beta]^2}. \tag{6}
\]

As mentioned in Tang, Equation (6) is identical to the upper size bound of Tang for the NB regression (the dispersion parameter may differ between the two groups in Tang) under the assumption of constant event rates if the dropout pattern is the same in the two groups since \( F_0 = \lambda_0^2 E(T_1^2)/2, E_0 = \lambda_0 E(T_1) \), and

\[
V_\beta = \left[ \frac{1}{p_1 \exp(\beta)} + \frac{1}{p_0} \right] \frac{1}{\lambda_1 E(T_1)} + \left[ \frac{\kappa_1}{p_1} + \frac{\kappa_0}{p_0} \right] \frac{E(T_1^2)}{E^2(T_1)}. \]

In this special situation, the AG model is almost as powerful as the NB regression when the variation in the patients' follow-up time \( T_1 \) is small, and the two models yield the same power if all subjects have the same follow-up time \( T_1 = \ldots = T_n \). However, the AG model does not require specifying the mixing distribution.

The NI test is one-sided, and the actual type I error is \( \alpha/2 \). In superiority trials, a two-sided test (i.e., \( H_0: \exp(\beta) = 1 \text{ vs } H_1: \exp(\beta) \neq 1 \)) is often used in practice. Formulae (5) and (6) can be used for the two-sided test since there is little chance that the observed outcomes will be significantly better in the control group than in the experimental group if the experimental treatment is truly more effective than the control treatment. The power and sample size formulae (5) and (6) remain the same if higher event rates indicate better health (\( M_0 \leq 1 \)) and the experimental treatment is truly superior or clinically noninferior to the control treatment in improving the event rate.

### 2.2 Equivalence trials

In an equivalence trial, the objective is to demonstrate that the experimental treatment is neither superior nor inferior to the standard control treatment. If the 100(1 - \( a/2 \))% CI for \( \exp(\beta) \) lies completely within the interval \([M_1, M_u]\), we can claim clinical equivalence of the two treatments, where \( M_1 < 1 \) and \( M_u > 1 \) are the prespecified margins. The hypothesis is

\[
H_0: \exp(\beta) \geq M_u \text{ or } \exp(\beta) \leq M_1 \text{ versus } H_1: M_1 < \exp(\beta) < M_u.
\]

The equivalence test can be viewed as the two one-sided tests and the type I error is \( \alpha/2 \). The power is given by

\[
P = \Pr(\hat{\beta} + z_{1-\alpha/2} \sqrt{n^{-1} V_\beta} < \log(M_u) \text{ and } \hat{\beta} - z_{1-\alpha/2} \sqrt{n^{-1} V_\beta} > \log(M_1)) \approx \Phi \left( \frac{\sqrt{n} \log(M_u) - \beta}{\sqrt{V_\beta}} - z_{1-\alpha/2} \right) - \Phi \left( \frac{\sqrt{n} \log(M_1) - \beta}{\sqrt{V_\beta}} + z_{1-\alpha/2} \right). \tag{7}
\]
The trial was terminated early for efficacy on basis of an interim analysis of the time to the first infection. In the trial, a total of 128 patients were randomized to gamma interferon or placebo. The data are simulated using Algorithm 1 and analyzed using the SAS NLMIXED procedure on basis of the likelihood function given in Equation (20) of Dean and Balshaw. Matsui obtained similar point estimates based on the generalized estimating equations (GEE) for the MPP model to the data using the SAS NLMIXED procedure on basis of the likelihood function given in Equation (20) of Dean and Balshaw. Matsui obtained similar point estimates based on the generalized estimating equations (GEE) for the MPP model to the data using the SAS NLMIXED procedure on basis of the likelihood function given in Equation (20) of Dean and Balshaw. Matsui.

To determine the sample size, we assume a common dispersion parameter and identical dropout pattern in the two groups. We set \( \psi = 1.1, \nu = 1.2, \kappa = 0.8 \), which are close to the MLE. The treatment allocation ratio is \( p_1 : p_0 = 1 : 1 \) or \( 2 : 1 \). We also perform sensitivity analyses to calculate the sample sizes at alternative parameter values \( \kappa = 0.4, 1.2, \psi = 1.5, \nu = 0.9 \). Both design 1 (planned treatment duration \( \tau = 1 \) year for all patients) and design 2 (accrual period \( \tau = 0.5 \) year, additional treatment duration \( \tau = 1 \) year, constant enrollment rate \( \eta = 0 \) ) are considered (please refer to Appendix A.2 for details). In both designs, the loss to follow-up distribution is exponential with mean \( 1/\delta = 4 \) years.

Table 1 reports the sample size and power estimates at the target 90% power and one-sided type I error \( \alpha/2 = 0.025 \). The empirical power is evaluated based on 40,000 trials. The data are simulated using Algorithm 2 of Tang and analyzed using the SAS PHREG procedure. There is more than 95% chance that the simulated power lies within \( 2 \sqrt{0.9} = 0.14/0.0000 = 0.3 \) of the true power. In both designs, in design 1, the simulated power is within 1% of the nominal power in nearly all cases. The performance slightly deteriorates in design 2 possibly because of larger variation in the follow-up time and higher overall dropout rate.

3.2 Example 2

We conduct simulations to assess the performance of the proposed method in the presence of unequal dispersion or differential dropout. Two scenarios are considered. In one scenario, the dispersion parameters in the two groups are different. In the other scenario, we assume different loss to follow-up distributions for the two groups. The setup is otherwise similar to that in the example 1. The parameter values and simulation results are presented in Table 2. The performance of the power and sample size method is almost as good as that in Example 1.
Simulation is conducted to assess the proposed sample size method for NI and equivalence trials. For illustration purposes, we assume a piecewise constant event rate function for the control arm $\lambda_0(t) = 1.0I(0 \leq t < 0.4) + 1.25I(0.4 \leq t < 0.8) + 1.5I(0.8 \leq t \leq 1)$, the event rate ratio between the active and control arm is $\exp(\beta) = \lambda_1(t)/\lambda_0(t) = 0.9$ or 1.0, and the dispersion parameter is $\kappa = 0.8$ or 1.2. Only design 1 is considered, and the planned treatment duration is $\kappa = 1$ year for all patients. The treatment allocation ratio is $1:1$. The loss to follow-up is exponentially distributed with mean $1/\delta = 4$ years (annual dropout rate 22.1%) in both arms.

| $\kappa$ | $\psi$ | $\nu$ | balanced size | unbalanced size | balanced size | unbalanced size |
|----------|--------|-------|---------------|----------------|---------------|----------------|
| 0.4      | 1.1    | 0.9   | 289           | 90.05          | 91.03         | 304           | 90.00         | 89.05         |
| 1.2      | 294    | 90.01 | 310           | 90.03          | 89.33         | 321           | 90.03         | 89.13         |
| 1.5      | 0.9    | 231   | 244           | 90.03          | 89.13         | 255           | 90.05          | 89.11          |
| 1.2      | 235    | 90.07 | 249           | 90.07          | 88.90         | 240           | 90.13          | 89.30          |
| 0.8      | 1.1    | 0.9   | 358           | 90.02          | 89.15         | 328           | 90.08          | 89.35          |
| 1.2      | 365    | 90.03 | 390           | 90.05          | 89.23         | 324           | 90.03          | 89.34          |
| 1.5      | 0.9    | 300   | 322           | 90.03          | 89.05         | 278           | 90.02          | 89.59          |
| 1.2      | 306    | 90.08 | 328           | 90.01          | 89.34         | 277           | 90.04          | 89.42          |
| 1.2      | 1.1    | 0.9   | 428           | 90.06          | 89.49         | 399           | 90.05          | 89.38          |
| 1.2      | 436    | 90.04 | 469           | 90.01          | 89.55         | 398           | 90.06          | 89.38          |
| 1.5      | 0.9    | 369   | 400           | 90.03          | 89.17         | 349           | 90.00          | 89.51          |
| 1.2      | 376    | 90.02 | 408           | 90.04          | 89.28         | 351           | 90.07          | 89.31          |

Table 1 reports the sample size and power estimates at the target 80% power and one-sided type I error $a/2 = 0.025$. The empirical power is evaluated based on 10,000 simulated trials. There is more than 95% chance that the simulated power lies within $2\sqrt{0.8 \times 0.2/10000} = 0.8\%$ of the true power. The simulated power at the calculated sample size is generally close to the target 80% power, indicating the accuracy of the proposed method.

### 4 DISCUSSION

We derive the power and sample size formulae for comparing recurrent rates in superiority, NI and equivalence trials using the robust Wald test from the AG model. The method allows the dispersion parameter, dropout rate, and/or sample size to differ between treatment groups. Numerical examples demonstrate the accuracy of the proposed method in moderate-to-large samples. It is always recommended to run simulation studies to verify the power calculation particularly when the sample size is relatively small.

We calculate the variance $V_\beta$ and the sample size at given event rate function, dispersion parameter and dropout rate. These parameters may be estimated from the historical trials using parametric methods. It is flexible to adjust the parameter values and conduct sensitivity analyses to examine how the sample size estimates vary with these parameter values. Please see Example 1 for illustration. It is possible to estimate $V_\beta$ from the historical trials by nonparametric methods. However, the nonparametric approach may require that the new trial is sufficiently similar to the historical trial in terms of the study population, treatment duration, drop rates, etc.
TABLE 2 Estimated sample size and simulated power (SIM) at the nominal 90% power in the presence of unequal dropout or dispersion

[1] SIM is evaluated using 10,000 simulated trials
[2] The treatment allocation ratio is 1 : 1

(a) Losses to follow-up are exponentially distributed with mean $1 / \delta = 4$ years (annual dropout rate 22.1%) in both arms.

(b) Losses to follow-up are exponentially distributed with $\delta_1 = 0.15$ and $\delta_0 = 0.35$ (annual dropout rates 13.9% and 29.5%) in the two arms.

[3] The sample size and nominal power estimates are updated for design 2 with staggered entry. The simulated power may be different from the previously reported values after re-running the simulation for design 1.

| $\psi$ | $\nu$ | $\kappa_0$ | $\kappa_1$ | size | nominal | SIM | power (%) | $\psi$ | $\nu$ | $\kappa_0$ | $\kappa_1$ | size | nominal | SIM | power (%) |
|-------|-------|------------|------------|------|---------|-----|-----------|-------|-------|------------|------------|------|---------|-----|-----------|
| 1.1   | 0.9   | 0.4        | 0.8        | 324  | 90.08   | 91.12| 292       | 90.05 | 0.4   | 0.4        | 0.4        | 287  | 90.06   | 90.89| 254       |
| 0.4   | 1.2   | 3.58       | 90.02      | 91.07| 328     | 90.08| 91.10     | 0.8   | 0.8   | 356       | 90.02      | 90.55| 326     | 90.07| 90.65     |
| 0.8   | 1.2   | 3.93       | 90.04      | 90.84| 363     | 90.03| 90.80     | 1.2   | 1.2   | 426       | 90.06      | 90.46| 397     | 90.04| 90.26     |
| 1.1   | 1.2   | 0.4        | 0.8        | 330  | 90.06   | 90.79| 287       | 90.01| 0.4   | 0.4        | 0.4        | 292  | 90.05   | 90.37| 248       |
| 0.4   | 1.2   | 3.65       | 90.03      | 91.02| 324     | 90.03| 90.99     | 0.8   | 0.8   | 363       | 90.06      | 90.58| 322     | 90.03| 90.66     |
| 0.8   | 1.2   | 4.00       | 90.00      | 90.75| 361     | 90.04| 90.69     | 1.2   | 1.2   | 434       | 90.06      | 90.29| 396     | 90.04| 90.33     |
| 1.5   | 0.9   | 0.4        | 0.8        | 265  | 90.04   | 90.82| 243       | 90.09| 0.4   | 0.4        | 0.4        | 229  | 90.06   | 90.55| 206       |
| 0.4   | 1.2   | 3.00       | 90.07      | 91.26| 278     | 90.02| 91.12     | 0.8   | 0.8   | 298       | 90.01      | 90.36| 277     | 90.05| 90.47     |
| 0.8   | 1.2   | 3.34       | 90.00      | 90.70| 314     | 90.06| 90.89     | 1.2   | 1.2   | 368       | 90.06      | 90.33| 348     | 90.02| 90.26     |
| 1.5   | 1.2   | 0.4        | 0.8        | 270  | 90.03   | 90.63| 240       | 90.02| 0.4   | 0.4        | 0.4        | 233  | 90.05   | 90.51| 202       |
| 0.4   | 1.2   | 3.06       | 90.08      | 91.13| 277     | 90.04| 90.72     | 0.8   | 0.8   | 304       | 90.05      | 90.63| 276     | 90.08| 90.48     |
| 0.8   | 1.2   | 3.41       | 90.05      | 90.67| 314     | 90.06| 90.79     | 1.2   | 1.2   | 375       | 90.06      | 90.37| 349     | 90.01| 90.45     |

TABLE 3 Estimated sample size at the nominal 80% power and simulated power (SIM) at the calculated sample size based on 10,000 NI or equivalence trials

(a) NI margin is $M_0 = 1.25$

(b) Equivalence margin is $(M_L, M_U) = (0.75, 1.25)$

The robust AG approach has several limitations. First, the AG model uses a common baseline hazard function for all events, and assumes that the risk of an event is unaffected by any early events that occur within the same subject. Therefore, the AG model is not suitable if the occurrence of early events increases the risk for subsequent ones. The AG model provides a convenient way to estimate an overall treatment effect, but it would be difficult to estimate the event specific treatment effect, which is useful for studying whether the treatment effect reduces after the patients experience one or more events. Second, when the sample size is small, the sandwich variance estimator tends to underestimate the true variance and have large sampling
variability, leading to inflated type I error rate.\textsuperscript{18,19} In the GEE methodology, the bias corrected sandwich variance estimator has been proposed for small sample inferences.\textsuperscript{18,19,20} It is possible to extend the bias correction method to the analysis of recurrent events. An alternative strategy for the analysis of small trials is to use the robust score test instead of the robust Wald test.\textsuperscript{20}

\section*{APPENDIX}

\subsection*{APPENDIX: TECHNICAL DETAILS}

\subsection*{A.1 A brief proof of equations (2) and (3)}

By Lin et al, \( \hat{V}_\beta \) is a consistent estimate of \( V_\beta \)

\begin{equation}
V_\beta = \frac{E[E(U^2_i|x_i)]}{E^2(I_\beta)} = \frac{p_1 \Sigma_1 + p_0 \Sigma_0}{\left( \int_0^\tau \omega_1(t)\omega_0(t)[p_1\pi_1(t)d\Lambda_1 + p_0\pi_0(t)d\Lambda_0] \right)^2}, \tag{A1}
\end{equation}

where \( d M_i(t) = d N_i(t) - Y_i(t) \exp(\beta x_i) d \Lambda_0(t) \), \( U_i = \int_0^\tau [x_i - \tilde{x}(\beta, t)] d M_i(t) \), \( \Sigma_g = E(U_i^2|x_i = g) \), and \( I_\beta = n^{-1} \sum_{i=1}^n \int_0^\tau [\tilde{x}(\beta, t) - \tilde{x}^2(\beta, t)] d N_i(t) \). By Lemma 1 in the web-based supplementary material of Song et al,\textsuperscript{2,3} we get

\begin{equation}
\Sigma_g = E \left[ \int_0^\tau (g - \tilde{x}(\beta, t))d M_i(t) \int_0^\tau (g - \tilde{x}(\beta, s))d M_j(s) \right] = A_g + k B_g \tag{A2}
\end{equation}

for subjects in group \( g \), where

\begin{align*}
A_g &= E \left[ \int_0^\tau Y_i(t)\omega^2_x(t)d\Lambda_0(t) \right] = \int \omega^2_\delta(t)\pi_\delta(s)d\Lambda_\delta(t), \\
B_g &= E \left[ \int_0^\tau \int_0^\tau Y_i(t)Y_j(s)\omega_\delta(t)\omega_\delta(s)d\Lambda_\delta(t)d\Lambda_\delta(s) \right] = 2 \int_0^\infty \int_0^\infty \omega_\delta(t)\omega_\delta(s)d\Lambda_\delta(t)d\Lambda_\delta(s).
\end{align*}

Inserting Equation (A2) into Equation (A1) yields Equation (2).

Equation (3) holds under equal dropout since \( \omega_0(t) = p_1 \exp(\beta)/D, \omega_1(t) = p_0/D \), \( I_\beta = p_0\omega_0(t)E_0 = p_1\omega_1(t)E_1 \), \( A_i = \omega^2_i(t)E_i \), \( B_i = 2\omega^2_i(t)F_i \), \( E_1 = E_0 \exp(\beta) \), \( F_1 = F_0 \exp(2\beta) \) and \( F_1/E_1^2 = F_0/E_0^2 \), where \( D = p_0 + p_1 \exp(\beta) \).

\subsection*{A.2 Asymptotic variance expressions in two designs under equal dropout}

\subsection*{A.2.1 Design 1}

The planned treatment duration is \( \tau_x \) years for each subject (the accrual period is irrelevant in the sample size calculation). The loss to follow-up is exponentially distributed with mean \( \delta^{-1} \). The probability that a subject is in the trial at time \( t \) after randomization is \( \pi(t) = \exp(-\delta t) \).

\textbf{Weibull event rate function}

Suppose the rate function is \( \lambda_0(t) = \psi v t^{v-1} \) and the mean function is \( \Lambda_0(t) = \psi t^v \) for the recurrent event in the control group, where \( \psi \) is a scale parameter and \( v \) is a shape parameter. Let \( \text{IG}(\nu, a) = \int_{t=0}^a t^{v-1} \exp(-t)dt \) be the incomplete gamma function and \( \text{IG}(\nu, a, b) = \int_{t=b}^a t^{v-1} \exp(-t)dt \). We have

\begin{align*}
E_0^1 &= \int_0^{\tau_x} \pi(t)d\Lambda_0(t) = \psi v \int_0^{\tau_x} \exp(-\delta t)t^{v-1}dt = \begin{cases} 
\frac{\psi v}{\delta^v} \text{IG}(v, \delta \tau_x) & \text{if } \delta \neq 0 \\
\frac{\psi v}{\tau_x^v} & \text{at } \delta = 0,
\end{cases}
\\
F_0^1 &= \int_0^{\tau_x} \pi(t)\Lambda_0(t)d\Lambda_0(t) = \psi^2 v \int_0^{\tau_x} \exp(-\delta t)t^{2v-1}dt = \begin{cases} 
\frac{\psi^2 v}{\delta^v} \text{IG}(2v, \delta \tau_x) & \text{if } \delta \neq 0 \\
\psi^2 \tau_x^{2v}/2 & \text{at } \delta = 0.
\end{cases}
\end{align*}
Piecwise constant event rate function

Let \( \lambda_0(t) = \sum_{k=1}^{d} \lambda_k I(l_{k-1} \leq t < l_k) \), where \( l_0 = 0, l_d = \tau_c \). Then \( \Lambda_0(t) = \Lambda_0(l_{k-1}) + \int_{l_{k-1}}^{t} \exp(-\delta(t-l_{k-1})) \) \( (t-l_{k-1})^{\delta} \) \( dt \) \( \int_l^{\Delta} \exp[-\delta t] t^{\delta} dt \) for \( m = 0, 1, 2 \). Then

\[
\begin{align*}
G_{k0} &= \frac{1-\exp(-\delta \Delta_k)}{\delta} , \quad G_{k1} = \frac{1-(1+\delta \Delta_k) \exp(-\delta \Delta_k)}{\delta}, \quad G_{k2} = \frac{2-(\delta^2 \Delta_k^2+2\delta \Delta_k+2) \exp(-\delta \Delta_k)}{\delta^2} \quad \text{if } \delta > 0 \\
G_{k0} &= \Delta_k , \quad G_{k1} = \frac{\Delta_k^3}{3}, \quad G_{k2} = \frac{\Delta_k^5}{3} \quad \text{at } \delta = 0.
\end{align*}
\]

We have

\[
F_0^1 = \int_0^{\tau_c} \pi(t) d\Lambda_0(t) = \sum_{k=1}^{d} \int_{l_{k-1}}^{l_k} \pi(t) dt = \left\{ \begin{array}{ll}
\sum_{k=1}^{d} \int_{l_{k-1}}^{l_k} \pi(t) \Lambda_0(l_{k-1}) G_{k0} & \text{if } \delta \neq 0 \\
\sum_{k=1}^{d} \int_{l_{k-1}}^{l_k} \pi(t) \Lambda_0(l_{k-1}) / 2 & \text{at } \delta = 0.
\end{array} \right.
\]

A.2.2 Design 2

Subjects are enrolled during an accrual period of \( \tau_a \) years, and followed for an additional \( \tau_c \) years after the closure of recruitment. The total study duration is \( \tau = \tau_a + \tau_c \) years. Suppose the entry time for a subject is distributed with density function given by

\[
f(e_i) = \frac{\eta \exp(-\eta e_i)}{1-\exp(-\eta \tau_c)} \quad \text{where } 0 \leq e_i \leq \tau_a.
\]

The entry distribution is convex (faster patient entry at the beginning) if \( \eta > 0 \), and concave (lagging patient entry) if \( \eta < 0 \), and uniform \( f(e_i) = 1/\tau_a \) if \( \eta \rightarrow 0 \). In terms of the sample size calculation, design 1 can be viewed as a special case of design 2 by setting \( \tau_a = 0 \).

Given the entry time \( e_i \), the maximum follow-up for an individual is \( \tau - e_i \). We assume the loss to follow-up is exponentially distributed with mean \( \delta^{-1} \). The probability that a subject is still in the trial at time \( t \) after randomization is

\[
\pi(t) = \Pr(T_i > t) = \Pr(T_i > t|e_i + t > \tau) \Pr(e_i + t > \tau) + \Pr(T_i > t|e_i + t > \tau) \Pr(e_i + t > \tau)
\]

\[
= \left\{ \begin{array}{ll}
\exp(-\delta t) & \text{if } t \leq \tau_c \\
\exp(-\delta t) \frac{1-\exp[-\eta (t-t_c)]}{1-\exp[-\eta \tau_c]} & \text{if } \tau_c < t \leq \tau.
\end{array} \right.
\]

When \( \eta \rightarrow 0 \), we shall replace \( \frac{1-\exp[-\eta (t-t_c)]}{1-\exp[-\eta \tau_c]} \) by its limiting value \( (t-t_c)/\tau_c \) in Equation (A4). In design 2, it is easy to see that

\[
\int_0^{\tau} \pi(t) d\Lambda_0(t) = E^1 + \int_{\tau_c}^{\tau} \pi(t) d\Lambda_0(t) \quad \text{and} \quad \int_0^{\tau} \pi(t) \Lambda_0(t) d\Lambda_0(t) = F_0^1 + \int_{\tau_c}^{\tau} \pi(t) \Lambda_0(t) d\Lambda_0(t),
\]

where \( E^1 \) and \( F_0^1 \) are defined in Appendix A.2.1. Below we give analytic expression for \( \int_{\tau_c}^{\tau} \pi(t) d\Lambda_0(t) \) and \( \int_{\tau_c}^{\tau} \pi(t) \Lambda_0(t) d\Lambda_0(t) \) at \( \eta = 0 \). The expressions are omitted when \( \eta \neq 0 \) due to limited space.

Weibull event rate function

Suppose \( \lambda_0(t) = \psi vt^{\psi-1} \). When \( \eta = 0 \), we get

\[
\int_{\tau_c}^{\tau} \pi(t) d\Lambda_0(t) = \left\{ \begin{array}{ll}
\frac{\psi v}{\tau_c^\psi} \frac{1}{\tau_c^\psi} \frac{1}{\tau_c^\psi} IG(v, \delta \tau_c, \delta \tau_c) - \frac{\psi v}{\tau_c^\psi} \frac{1}{\tau_c^\psi} IG(v + 1, \delta \tau_c, \delta \tau_c) & \text{if } \delta > 0 \\
\frac{\psi v}{\tau_c^\psi} \frac{1}{\tau_c^\psi} \frac{1}{\tau_c^\psi} [v t^{\psi-1} t^{\psi-1} - \tau_c^\psi] & \text{at } \delta = 0
\end{array} \right.
\]

\[
\int_{\tau_c}^{\tau} \pi(t) \Lambda_0(t) d\Lambda_0(t) = \left\{ \begin{array}{ll}
\frac{\psi v}{\tau_c^\psi} \frac{1}{\tau_c^\psi} \frac{1}{\tau_c^\psi} IG(2v, \delta \tau_c, \delta \tau_c) - \frac{\psi v}{\tau_c^\psi} \frac{1}{\tau_c^\psi} IG(2v + 1, \delta \tau_c, \delta \tau_c) & \text{if } \delta > 0 \\
\frac{\psi v}{\tau_c^\psi} \frac{1}{\tau_c^\psi} \frac{1}{\tau_c^\psi} [v t^{2\psi-1} t^{2\psi-1} - \tau_c^2] & \text{at } \delta = 0
\end{array} \right.
\]

Piecwise constant event rate function

Suppose \( \lambda_0(t) = \sum_{k=1}^{d} \lambda_k I(l_{k-1} \leq t < l_k) \), where \( l_{d_r} = \tau + \tau_c \) and \( l_d = \tau_c \). For notational convenience, if \( \tau_c \) is not a knot,
it can be added as a knot. When \( \eta = 0 \),

\[
\int_{r_c}^{\tau} \pi(t) d \Lambda_0(t) = \sum_{k=1}^{d} \int_{l_k}^{l_{k+1}} \frac{\lambda_k \exp(-\delta t) \tau - t}{\tau} dt = \sum_{k=1}^{d} \frac{\lambda_k}{\tau} \exp(-\delta l_{k-1}) \left[ (\tau - l_{k-1})G_{k0} - G_{k1} \right],
\]

\[
\int_{r_c}^{\tau} \pi(t)\Lambda_0(t)d\Lambda_0(t) = \sum_{k=1}^{d} \frac{\lambda_k}{\tau} \exp(-\delta l_{k-1}) \left\{ \Lambda_{k-1}[(\tau - l_{k-1})G_{k0} - G_{k1}] + \lambda_k[(\tau - l_{k-1})G_{k1} - G_{k2}] \right\},
\]

where \( G_{k0} \), \( G_{k1} \) and \( G_{k2} \) are defined in Equation (A3).

References

1. Wang Y, Meyerson L, Tang Y, Qian N. Statistical methods for the analysis of relapse data in MS clinical trials. Journal of the Neurological Sciences 2009; 285: 206 - 11.
2. Aban IB, Cutter GR, Mavinga N. Inferences and power analysis concerning two negative binomial distributions with an application to MRI lesion counts data. Computational Statistics & Data Analysis 2009; 53: 820 -33.
3. Tang Y. Sample Size Estimation for Negative Binomial Regression Comparing Rates of Recurrent Events with Unequal Follow-Up Time. Journal of Biopharmaceutical Statistics 2015; 25: 1100 - 13.
4. Matsui S. Sample size calculations for comparative clinical trials with over-dispersed Poisson process data. Statistics in Medicine 2005; 24: 1339 - 56.
5. Nicholas R, Straube S, Schmidli H, Schneider S, Friede T. Trends in annualized relapse rates in relapsing-remitting multiple sclerosis and consequences for clinical trial design. Multiple sclerosis 2011; 17: 1211 - 17.
6. Inusah S, Sormani MP, Cofield SS, et al. Assessing changes in relapse rates in multiple sclerosis. Multiple sclerosis 2010; 16: 1414 - 21.
7. Andersen PK, Gill RD. Cox’s regression model counting process: a large sample study. Annual of Statistics 1982; 10: 1100 - 20.
8. Tang Y. Negative Binomial Regression: Sample Size with Unequal Follow-Up Times. In: CRC Press; 2017.
9. Song R, Kosorok MR, Cai J. Robust covariate-adjusted log-rank statistics and corresponding sample size formula for recurrent events data. Biometrics 2008; 92: 741 -50.
10. Lawless JF, Nadeau C. Some Simple Robust Methods for the Analysis of Recurrent Events. Technometrics 1995; 37: 158 - 68.
11. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric Regression for the Mean and Rate Functions of Recurrent Events. Journal of the Royal Statistical Society B 2000; 62: 711 - 30.
12. Tang Y. Algorithms for imputing partially observed recurrent events with applications to multiple imputation in pattern mixture models. Journal of Biopharmaceutical Statistics 2018; 28: 518-533.
13. Tang Y. Sample size for comparing negative binomial rates in noninferiority and equivalence trials with unequal follow-up time. Journal of Biopharmaceutical Statistics 2018; 28: 475-491.
14. Tang Y. Exact and approximate power and sample size calculations for analysis of covariance in randomized clinical trials with or without stratification. Journal of Biopharmaceutical Research 2018; 10: 274-286.
15. Tang Y. A Noniterative Sample Size Procedure for Tests Based on t Distributions. Statistics in Medicine 2018; 37: 3197 - 3213.
16. Dean CB, Balshaw R. Efficiency Lost By Analyzing Counts Rather Than Event Times in Poisson and Overdispersed Poisson Regression Models. *Journal of the American Statistical Association* 1997; 92: 1387 - 98.

17. Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *International journal of epidemiology* 2015; 44: 1 - 10.

18. Kauermann G, Carroll RJ. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association* 2001; 96: 1387 - 96.

19. Lu B, Preisser JS, Qaqish BF, Suchindran C, Bangdiwala S, Wolfson M. A comparison of two bias-corrected covariance estimators for generalized estimating equations. *Biometrics* 2007; 63: 935-41.

20. Guo X, Pan W, Connett JE, Hannan PJ, French SA. Small-sample performance of the robust score test and its modifications in generalized estimating equations. *Statistics in Medicine* 2005; 24: 3479 - 95.