Published in final edited form as:

Biometrics. 2011 December ; 67(4): . doi:10.1111/j.1541-0420.2011.01589.x.

Insights on the Robust Variance Estimator under Recurrent-Events Model

Hussein R. Al-Khalidi¹, Yili Hong^{2,*}, Thomas R. Fleming³, and Terry M. Therneau⁴

¹Department of Biostatistics & Bioinformatics, Duke University, Durham, North Carolina 27705, U.S.A

²Department of Statistics, Virginia Tech, Blacksburg, Virginia 24060, U.S.A

³Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A

⁴Section of Biostatistics, Mayo Clinic, Rochester, Minnesota 55905, U.S.A

Summary

Recurrent events are common in medical research for subjects who are followed for the duration of a study. For example, cardiovascular patients with an implantable cardioverter defibrillator (ICD) experience recurrent arrhythmic events that are terminated by shocks or antitachycardia pacing delivered by the device. In a published randomized clinical trial, a recurrent-event model was used to study the effect of a drug therapy in subjects with ICDs, who were experiencing recurrent symptomatic arrhythmic events. Under this model, one expects the robust variance for the estimated treatment effect to diminish when the duration of the trial is extended, due to the additional events observed. However, as shown in this article, that is not always the case. We investigate this phenomenon using large datasets from this arrhythmia trial and from a diabetes study, with some analytical results, as well as through simulations. Some insights are also provided on existing sample size formulae using our results.

Keywords

Andersen–Gill model; Clinical trials; Recurrent-events data; Robust standard error; Sample size; Sandwich estimator

1. Introduction

Recurrent events are commonly encountered in medical research where events can occur repeatedly over time for each subject. Researchers are often interested in determining whether a therapeutic agent can reduce the rate of recurrent events. For example, in the cardiovascular area, arrhythmias often happen over time in subjects with implantable cardioverter defibrillators (ICDs). Subjects can benefit from a therapy that reduces the symptomatic burden of arrhythmias. In another example, subjects who suffer from type II diabetes often are interested in therapy that provides long-term reduction in the rate of major microvascular and macrovascular events while having a low rate of recurrent hypoglycemia events over the course of treatment period.

Supplementary Materials

Web Appendix A, referenced in Section 3.2, and Web Appendix B, referenced in Section 4.3, are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

To evaluate the effect of a treatment on an outcome measure for single event data, a proportional hazard model (Cox, 1972) is the most popular. To generalize the Cox model to recurrent-event data, the Andersen–Gill (AG) model (Andersen and Gill, 1982) with an appropriate sandwich variance estimator provides a common approach and is implemented in most statistical software. In a two-arm clinical trial with equal group sizes, sample size calculations for a Cox model often use the approximation $\text{var}(\hat{\beta}) = 4/L$, where $\hat{\beta}$ is the estimate of the log relative risk β and L is the total number of events. This is an approximation that is sufficiently accurate in most settings. A natural question is whether this formula extends to studies with multiple events per subject. Two studies described below with a large number of events per subject show that this is not so. In the Shock Inhibition Evaluation with Azimilide (SHIELD) Trial (see Section 4.1 for more details), there was a total of 2124 recurrent symptomatic arrhythmic events in 434 subjects. The robust variance estimate for the treatment effect from an AG model is 0.062, whereas the simple estimate $4/L$ is 0.0019. For the diabetes data in the Exubera Trial (see Section 4.2 for details), there was a total of 3495 hypoglycemia events in 327 subjects; the robust variance is 0.034 as compared to the projection of $4/L = 0.0011$.

In these two datasets, the variance of the treatment effect is clearly not decreasing at a rate of $4/L$ when there is a large number of events per subject. In this article, we decompose the sandwich type robust variance into two terms, the first term is of the order $O(1/L)$ and the second term is of the order $O(1/n)$, where n is the number of subjects under study. When n is fixed but the number of events L increases, the second term eventually dominates.

The outline of the article is as follows. Section 2 of this article gives a brief introduction about recurrent-events models, parameter estimation, and the robust variance estimate of the parameter. Section 3 shows analytical results that provide insights for the robust variance estimates. Section 4 presents the analysis of the SHIELD and Exubera data. Section 5 examines the robust variance estimates under different scenarios through simulations. Section 6 provides insights into sample size calculations under recurrent-events models. Section 7 provides conclusions and areas for future work.

2. Model and Parameter Estimation

Let $N_i(t)$ be the counting process for subject i and $N(t) = \sum_{i=1}^n N_i(t)$ be the overall counting process for the n subjects. The process $N_i(t)$ has a jump of size one when an event occurs. Let $Y_i(t)$ be the left continuous at-risk process, which is unity whenever the subject is at risk for an event. Let x_i be the treatment group indicator, which has value 1 for the active and 0 for the control. Let

$$\begin{aligned} r(\beta, t) &= \sum_{i=1}^n Y_i(t) e^{x_i \beta}, \\ \bar{x}(\beta, t) &= \frac{\sum_{i=1}^n Y_i(t) e^{x_i \beta} x_i}{r(\beta, t)}, \\ v(\beta, t) &= \frac{\sum_{i=1}^n Y_i(t) e^{x_i \beta} \{x_i - \bar{x}(\beta, t)\}^2}{r(\beta, t)}. \end{aligned} \quad (1)$$

be the weighted number at risk, weighted mean, and weighted variance at each time t , respectively.

The Cox model assumes that the observed hazard for each subject is

$$\lambda_i(t) = Y_i(t) \lambda_0(t) e^{x_i \beta}. \quad (2)$$

The estimated treatment effect $\hat{\beta}$ is obtained as the maximizer of the log-partial likelihood (Cox, 1975) given below:

$$l(\beta) = \sum_{i=1}^n \int_0^\infty [Y_i(t)x_i\beta - \log\{r(\beta, s)\}]dN_i(s). \quad (3)$$

The model based variance of $\hat{\beta}$ is the inverse of the information matrix $I(\beta)$ of the partial-likelihood score equations. In particular,

$$I(\beta) = \sum_{i=1}^n \int_0^\infty v(\beta, s)dN_i(s), \quad \text{and} \quad V(\beta) = I^{-1}(\beta). \quad (4)$$

In a two-arm clinical trial $v(0, 0) = p(1 - p)$, where p is the fraction of subjects assigned to the active arm. This result leads, under a small β approximation, to the simple variance formula commonly used in sample size calculations. A more sophisticated approximation based on the partial likelihood score equations when $p = 1/2$ is

$$V(\hat{\beta}) \approx \frac{4}{L} \left\{ \frac{(1+B_0)^2}{8B_0} + \frac{(1+B_1)^2}{8B_1} \right\}, \quad (5)$$

where B_0 is the relative risk under the null hypothesis H_0 and B_1 is the relative risk under the alternative hypothesis H_1 . More details on the approximation in (5) can be found in Fleming and Harrington (1991, pp. 394–395). The term in the right side of equation (5), $\{(1+B_0)^2/(8B_0) + (1+B_1)^2/(8B_1)\}$, ranges between 1 and 1.125 when both B_0 and B_1 are between 0.5 and 2.0, a common range for medical studies. Note that when $B_0 = B_1 = 1$, equation (5) reduces to $4/L$. We will use $4/L$ as an approximation henceforth.

For recurrent-events data it is common to use the AG model extension of the Cox model. The intensity function and partial likelihood for the AG model are as in (2) and (3), respectively. A model based variance $V(\beta)$ requires independent increments in $N_i(t)$ for each subject, which in turn implies that all relevant covariates affecting each subject's risk need to be included in the model. To account for possible model misspecification, robust variance estimates of $\hat{\beta}$ are often used in practice. Lin and Wei (1989) introduced a sandwich type robust variance for $\hat{\beta}$. In particular, the sandwich type robust variance is defined as

$$R = I(\hat{\beta})^{-1} \left\{ \sum_{i=1}^n U_i(\hat{\beta})U_i'(\hat{\beta}) \right\} I(\hat{\beta})^{-1}, \quad (6)$$

where $U_i(\hat{\beta})$ is the score residual for subject i . In particular, $U_i(\hat{\beta}) = U_i(\hat{\beta}, \infty)$, where

$$U_i(\hat{\beta}, t) = \int_0^t \{x_i - \bar{x}(\hat{\beta}, s)\} d\widehat{M}_i(s)$$

and $\widehat{M}_i(t) = N_i(t) - \int_0^t Y_i(s)\lambda_0(s)\exp(x_i\hat{\beta})ds$ is the martingale residual. See Therneau and Grambsch (2000, Chapter 8.2) for applications, including an interpretation of the sandwich type robust variance estimate as an infinitesimal jackknife (Jaeckel, 1972), and Fleming and Harrington (1991) for technical details.

3. Analytical Results

The AG model together with the robust variance provides a simple and valid approach to analyze recurrent events. It is available in most commercial software and widely used by practitioners. The purpose of this article is to show how the formula in (6) for the robust variance behaves under some possible model misspecifications. The AG model assumes the true intensity is $d\Lambda_i(t) = \lambda_0(t)e^{x_i\beta}$. The true intensity function for subject i , however, may also depend on covariates Z_i (e.g., age, body mass index) that potentially could be included in the model, and an unobservable variable w_i (i.e., a frailty with mean 1 and variance σ_w^2), in addition to the treatment covariate x_i . In particular, we assume the true intensity function is

$$d\Lambda_i(t) = \mathbf{E} \{ dN_i(t) | N_i(s), 0 \leq s < t, x_i, Z_i, w_i \}. \quad (7)$$

We use (7) as the true intensity function for subject i in the development of our analytical results.

The key idea for the analytical results is to decompose the observed-expected martingale residual process for each subject as

$$N_i(t) - \hat{\Lambda}_i(\hat{\beta}, t) = \{N_i(t) - \Lambda_i(t)\} + \{\Lambda_i(t) - \hat{\Lambda}_i(\hat{\beta}, t)\}. \quad (8)$$

Each subject has a true (but unknown) cumulative intensity function $\Lambda_i(t)$, which is estimated in the fitted model by $\hat{\Lambda}_i(\hat{\beta}, t)$. The first term in the decomposition is the difference between the number of observed and expected events for each subject. It has a Poisson-like variance, which is proportional to the expected number of events for the subject, decreasing at the rate of $O(1/L)$, where L is the total number of events. The second term reflects the variability in the event rate for each subject and has components of both model convergence and model correctness.

3.1 Simple Case

In the diabetes data almost all subjects have complete follow-up through the end of the study. This motivates a simple theoretical case where all subjects have the same follow-up through some time τ , and treatment is balanced with an equal number of subjects on each arm. Simplifying further, we consider the variance of the robust score statistic (i.e., at $\beta = 0$).

In this simple case we have

$$\bar{x}(\beta, s) = p = 1/2, \quad v(\beta, s) = p(1 - p) = 1/4$$

at any follow-up time $s < \tau$, and $V(0) = 4/L$ exactly. Here p is the proportion of subjects that were randomized to the active group. The martingale residual for each subject is simplified to $N_i - L/n$, giving a score residual of $U_i = \pm 1/2(N_i - L/n)$ for those with $x = 0$ or 1 . The robust variance is

$$R = \frac{4}{L^2} \sum_{i=1}^n (N_i - L/n)^2 = \frac{1}{n} \frac{4}{(L/n)^2} \frac{\sum_{i=1}^n (N_i - L/n)^2}{n}. \quad (9)$$

In an idealized experiment, the variation in λ_i is attributed to the variability among subjects entering the trial. The random quantity $N_i|\lambda_i$ follows a Poisson process (i.e., assuming constant event rate). Thus, we have

$$E(N_i) = E(N_i - \Lambda_i)^2 = \Lambda_i = \lambda_i \tau, \quad E(L) = \sum_{i=1}^n \Lambda_i = n\tau \bar{\lambda}$$

where $\bar{\lambda}$ is the mean of λ_i 's, and

$$\begin{aligned} \sum_{i=1}^n E(N_i - L/n)^2 &= \sum_{i=1}^n \left\{ E(N_i - \tau\lambda_i)^2 + E(\tau\lambda_i - L/n)^2 \right\} \\ &\approx n\bar{\lambda}\tau + \tau^2 n \text{var}(\lambda). \end{aligned} \quad (10)$$

The approximation in (10) is obtained by assuming $(N_i - \tau\lambda_i)$ and $(\tau\lambda_i - L/n)$ are essentially independent (because the covariance between $(N_i - \tau\lambda_i)$ and $(\tau\lambda_i - L/n)$ is of a smaller order than the variances of $(N_i - \tau\lambda_i)$ and $(\tau\lambda_i - L/n)$). By (9) and the approximation $E(x/y) \approx E(x)/E(y)$, we obtain

$$R \approx \frac{4}{n} \left\{ \frac{1}{\tau\bar{\lambda}} + \frac{\text{var}(\lambda)}{\bar{\lambda}^2} \right\}. \quad (11)$$

When the expected number of events per subject $\tau\bar{\lambda}$ is small, as is often the case for a clinical trial, the first term of R dominates leading to the usual Cox model result. By rescaling λ and τ , and assuming without loss of generality that $\bar{\lambda} = 1$, we recognize the second term as equivalent to the variance of a random effect or frailty w_i for each subject. As τ increases this term remains constant for fixed n .

3.2 General Case

The theorem below formalizes the heuristic derivation in Section 3.1 for the robust variance of $\hat{\beta}$. For the general case, each subject i has a true event rate $d\Lambda_i(t)$ as in (7). To make the presentation simple, we assume type I censoring, that is all subjects are followed to time τ . For the decomposition in (8) when $t = \tau$, we have,

$$\left\{ N_i(\tau) - \hat{\Lambda}_i(\hat{\beta}, \tau) \right\} = \left\{ N_i(\tau) - \Lambda_i(\tau) \right\} + \left\{ \Lambda_i(\tau) - \hat{\Lambda}_i(\hat{\beta}, \tau) \right\}.$$

The first term is the variation of each subject from their true expected value, with mean 0 and variance $\Lambda_i(\tau)$.

Theorem 1—We assume that the true event rate for subject i is $d\Lambda_i(t)$ as in (7), all subjects are followed for a fixed amount of time τ , and a model with a single binary covariate x_i . Let β^* be the asymptote of $\hat{\beta}$ as $n \rightarrow \infty$, and

$$\begin{aligned} \bar{\Lambda}(\tau) &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Lambda_i(\tau), \\ \bar{\Lambda}_j(\tau) &= \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{\{i: x_i = j\}} \Lambda_i(\tau), \\ V_{\lambda_j} &= \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{\{i: x_i = j\}} \left\{ \Lambda_i(\tau) - \bar{\Lambda}_j(\tau) \right\}^2. \end{aligned}$$

Here $m = n/2$ is the number of subjects in each arm. Then, the asymptote β^* is given by $\beta^* = \log \{ \Lambda_1(\tau) / \Lambda_0(\tau) \}$, and the robust variance in (6) can be expressed as

$$R = R_1 + R_2 + o_p \left(\frac{1}{n \{ \bar{\Lambda}(\tau) \}} \right) \quad (12)$$

where

$$\begin{aligned} R_1 &= \frac{4 \{ \bar{\Lambda}(\tau) \}^2}{\Lambda_0(\tau) \Lambda_1(\tau)} \times \frac{1}{n \{ \bar{\Lambda}(\tau) \}}, \\ R_2 &= \frac{2 \{ \exp(\beta^*) V_{\lambda 0} + \exp(-\beta^*) V_{\lambda 1} \}}{\Lambda_0(\tau) \Lambda_1(\tau)} \times \frac{1}{n}. \end{aligned} \quad (13)$$

The proof of Theorem 1 is available in Web Appendix A. The first term R_1 is of the order $O(1/L)$, decreasing in the total number of events L . The second term R_2 is of the order $O(1/n)$, decreasing in the total number of subjects, and does not depend on the number of events. The key term in R_2 is the variance term $V_{\lambda j}$. The quantity $V_{\lambda j}$ is the variance of each true Λ_i around the mean value Λ_0 or Λ_1 for the control or active groups. Thus $V_{\lambda j}$ reflects the variation in Λ_i 's.

When the population is nonhomogeneous, $V_{\lambda j} > 0$ and thus $R_2 > 0$. Theorem 1 implies that there is a law of diminishing returns for a multievent study when there is variation in the true rates beyond the covariates in the model. This is because the increasing of number of events resulting from extending the study does not have effect on R_2 , if $R_2 > 0$. When R_2 is negligible (i.e., $R_2 \approx 0$), the first term R_1 dominates the robust variance R and we have the usual Cox model behavior with respect to the number of events. If R_2 is nonnegligible, different behavior is exhibited: As the length of the study is extended and the number of events increases, R_1 diminishes to 0 whereas R_2 remains essentially fixed, so that R_2 eventually becomes the dominant term. Extending the length of the study does not cause R to diminish to 0 because R_2 depends only on the number of subjects and not on the length of the study. Section 5 presents simulations that illustrate these results.

The following corollaries consider several situations of true event rates $d\Lambda_i(t)$ and give corresponding expressions for those terms in R .

Corollary 1—Suppose the true event rate $d\Lambda_i(t) = w_i \lambda(t; x_i, Z_i)$, where w_i is a frailty with mean 1 and variance σ_w^2 and $\lambda(t; x_i, Z_i)$ is a positive function depending on covariates Z_i as well as x_i and t . Let $\Lambda(t; x_i, Z_i) = \int_0^t \lambda(s; x_i, Z_i) ds$. We have

$$V_{\lambda j} = \sigma_w^2 \{ \bar{\Lambda}(\tau; j) \}^2 + \text{var} \{ \Lambda(\tau; j, Z) \}$$

where $\bar{\Lambda}(\tau; j) = \lim_{m \rightarrow \infty} m^{-1} \sum_{\{i: x_i = j\}} \Lambda(t; j, Z_i)$ and $\text{var} \{ \Lambda(\tau; j, Z) \}$ is the variance of $\Lambda(\tau; j, Z)$ caused by the variability in Z . Conditional on covariate Z (i.e., treating Z as

deterministic), $V_{\lambda j} = \sigma_w^2 \{ \bar{\Lambda}(\tau; j) \}^2$. In a special case where $d\Lambda_i(t) = w_i \exp(x_i \beta)$, $V_{\lambda j} = \sigma_w^2 \{ \exp(x_j \beta) \}^2 \tau^2$ and $R_2 = 4\sigma_w^2/n$.

This result shows that the nonhomogeneity can arise from two sources, the frailty and the covariates. If the nonhomogeneity is caused by the variability in covariates and there is information on those covariates, adding those covariates to the model will make the model-

based variance close to the robust variance. For example in the chronic granulomatous disease dataset found in Therneau and Grambsch (2000, pp. 318–320), some very strong predictors are available. The variance of the residual random effect is modestly large for the treatment only model, but close to zero for the model with other covariates. The robust and standard variance estimates become almost identical when other covariates are included. Theorem 1 suggests that the gain from extra events can be larger if we have good predictors.

The following corollary considers a situation that can be related to the simple case considered in Section 3.1. In addition, in Section 4.3, we use this corollary to illustrate the main theorem.

Corollary 2—*Suppose the true event rate $d\Lambda_i(t) = w_i\lambda(x_i, Z_i) = \lambda_i$, that is the event rate λ_i is constant over time. We have $\beta^* = \log(\lambda_1/\lambda_0)$ and*

$$R = R_1 + R_2 + o_p\left(\frac{1}{n\tau}\right), \quad (14)$$

where

$$R_1 = \frac{4\bar{\lambda}}{\lambda_0\bar{\lambda}_1} \times \frac{1}{n\tau}, \quad R_2 = \frac{2\{\exp(\beta^*)V_{\lambda_0}^c + \exp(-\beta^*)V_{\lambda_1}^c\}}{\bar{\lambda}_0\bar{\lambda}} \times \frac{1}{n}.$$

Here $\lambda_j^- = \lim_{m \rightarrow \infty} m^{-1} \sum_{\{i: x_i = j\}} \lambda_i$, $\bar{\lambda} = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \lambda_i$, and $V_{\lambda_j}^c = \lim_{m \rightarrow \infty} m^{-1} \sum_{\{i: x_i = j\}} (\lambda_i - \bar{\lambda}_j)^2$.

When $\lambda_0^- = \lambda_1^-$ (i.e., $\beta^* = 0$), the results in (14) reduces to (11) as a special case.

4. Analysis of Two Clinical Datasets

In this section, we present the analyses of two datasets from randomized clinical trials, which motivated the analytical results in Section 3.

4.1 SHIELD Data

The SHIELD Trial (Dorian et al., 2004) was a 12-month randomized placebo-controlled double-blinded study to evaluate the efficacy of azimilide in subjects with ICDs. Azimilide is an investigational class III antiarrhythmic agent which is expected to reduce the rate of recurrent symptomatic arrhythmias that require termination via antitachycardia pacing or shocks delivered by the device. SHIELD randomized 633 subjects to three doses (214 in the placebo group, 220 in the 75 mg azimilide group, 199 in the 125 mg azimilide). For illustration in this article, we will use data from placebo and 75 mg azimilide. Among the 434 subjects randomized to placebo or 75 mg azimilide, there were 2124 recurrent symptomatic arrhythmia events over 12 months of follow-up.

For the SHIELD data, the estimated treatment effect is $\hat{\beta} = -0.836$, and the estimated robust variance in (6) is 0.062. For insights about how the estimates of β and the variance change when the duration of the trial varies, we assume the trial was terminated at week 1 (i.e., all subjects are censored on day 7 if not censored earlier), week 2, and so on. Figure 1 shows the estimates of β and the corresponding robust variance as the last follow-up time increases in the SHIELD data.

Figure 2a gives a plot of the average number of events per week for subjects in the SHIELD data. The majority of the subjects experienced less than three events per week but there were 12 subjects with more than seven events per week and 2 subjects with over 30 events per week. The large variability in the number of events per week provides an indication of the lack of homogeneity in the population. Another interesting feature about the SHIELD data is that those who completed the follow-up period show smaller numbers of events per week. Those with larger numbers of events per week tend to drop out earlier. This is not uncommon in ICD trials where a subject could have such high intensity of shocks that the physician may switch the subject to another therapy (dropout).

4.2 Type 2 Diabetes Data

The Exubera Trial was a placebo-controlled evaluation of inhaled versus injected insulin conducted in subjects with type 2 diabetes mellitus. The clinical setting is similar to Hollander et al. (2004) but for a different population. Although longterm sequelae due to vascular events such as retinopathy and neuropathy are of concern, incidents of hypoglycemia provide some indication of disease control, stability, and safety. A total of 327 subjects was randomized with 162 subjects in treatment A and 165 in treatment B (here the active treatment/placebo information is blinded). Most of the subjects were followed for 6 months. Our analysis includes episodes of moderate and severe hypoglycemia.

A fit of the AG model to this data yields $\hat{\beta} = 0.125$ (used treatment A as the control), and the estimated robust variance in (6) is 0.034. Figures 3 and 2b are similar to Figures 1 and 2a and provide results for the diabetes data. The variability in the number of events per week is much less than that of the SHIELD data.

4.3 Estimation of Variance Terms

For illustration, we apply Corollary 2 to the datasets presented in Sections 4.1 and 4.2. The application of the corollary is only approximate because the percentage of subjects with complete follow-up was only 60% in the SHIELD Trial and 90% in the Exubera Trial. One may use $\hat{\lambda}_i = N_i(\tau_i)/\tau_i$ as an estimate of λ_i . Here $N_i(\tau_i)$ is the observed number of events at the last follow-up time τ_i for subject i . The variance terms in Corollary 2 can be estimated correspondingly by substituting in $\hat{\lambda}_i$. The entire dataset was used to estimate those $\hat{\lambda}_i$'s.

Figure 4a shows the estimated variance terms in the SHIELD data under the assumption that the trial was stopped at week 1, week 2, ..., and so on. The estimate $\hat{\beta}_k$ is obtained using the data up to Week k . We use $\hat{\beta}_k$ as estimates for the quantity β^* defined in Corollary 2. Those estimated variance terms were obtained by substituting $\hat{\lambda}_i$'s into variance terms in Corollary 2. As shown in Figure 2a, there is large variability in $\hat{\lambda}_i$, $i = 1, \dots, n$ in the SHIELD data. The large variability in $\hat{\lambda}_i$'s and the early dropouts cause the gap between the robust variance and the estimated $R_1 + R_2$ as shown in Figure 4a (a small simulation study that mimics the setting of the SHIELD Trial did verify this observation; details are available in Web Appendix B). Figure 4b gives a similar plot for the diabetes data. The robust variance and the estimated $R_1 + R_2$ agree quite well for the diabetes data.

5. Simulation Studies

5.1 Design of the Simulation Studies

We simulated data with various hazard functions. Note that the time scale (t) is in days throughout these simulations. We simulated 400 subjects with 200 in each group (placebo/active) in each trial. Each subject is followed for 200 days. Each trial was replicated $B = 1000$ times and the results shown in Section 5.2 are stable. The choice of the number of subjects and the length of study are close to the clinical trial settings in Section 4.

- Scenario A: *Proportional Hazard Case with Homogeneous Subjects.* Each counting process $N(t)$ is associated with hazard rate functions $\lambda_0(t)$ and $\lambda_0(t)\exp(\beta)$ for placebo and active groups, respectively. Here we chose the power-law model with $\lambda_0(t) = (\alpha/\eta)(t/\eta)^{\alpha-1}$. In this simulation, we set $\alpha = 1.5$, $\eta = 20$, $\beta = 0.5$.
- Scenario B: *Nonproportional Hazard Case with Homogeneous Subjects.* Each counting process $N(t)$ is associated with hazard rate functions $\lambda_0(t)$ and $\lambda_1(t)$ for placebo and active groups, respectively. Here we chose $\lambda_0(t) = \lambda_0$ and $\lambda_1(t) = (\alpha/\eta)(t/\eta)^{\alpha-1}$, which is the power-law model. In this simulation, we set $\lambda_0 = 0.3$, $\alpha = 1.5$, $\eta = 15$.
- Scenario C: *Proportional Hazard Case with Nonhomogeneous Subjects.* Each counting process $N(t)$ is associated with hazard rate functions $\lambda_0(t) = \lambda_0$ and $\lambda_1(t) = \lambda_0\exp(\beta)$ for placebo and active groups, respectively. Here we chose $\lambda_0 \sim \text{exponential}(\gamma_0)$ with event rate γ_0 . In this simulation, we set $\gamma_0 = 20$, $\beta = 0.3$. Note that this scenario covers a simple frailty model, where the frailty distribution is an exponential distribution.
- Scenario D: *Nonproportional Hazard Case with Nonhomogeneous Subjects.* Each counting process $N(t)$ is associated with hazard rate functions $\lambda_0(t) = \lambda_0$ and $\lambda_1(t) = (\alpha/\eta)(t/\eta)^{\alpha-1}$ (which is the power-law model) for placebo and active groups, respectively. Here we chose $\lambda_0 \sim \text{exponential}(\gamma_0)$. In this simulation, we set $\gamma_0 = 20$, $\alpha = 1.5$, $\eta = 15$. Note that this scenario allows the events within subject to be correlated.

5.2 Results and Discussions

Figure 5 presents the mean estimated log relative risk $\hat{\beta}$ and mean estimated variance over B simulated trials. For the case of a homogenous population, Figures 5a and 5b show that R_2 is negligible. Nonproportionality, however, causes $\hat{\beta}$ to vary as the trial length is extended (Figure 5b). For the case of a nonhomogenous population, Figures 5c and 5d show that R_2 is positive and nearly constant as a function of the trial length. A nonhomogenous population leads, as expected, to a diminishing returns effect as the trial length is extended. Nonproportionality again causes $\hat{\beta}$ to vary as the trial length is extended (Figure 5d). The results in Figure 5 also show that the approximation in Theorem 1 approximates the robust variance quite well. Figure 5 also shows the empirical variance of $\hat{\beta}$ over the B replications. The robust variance formula tracks the variance of the $\hat{\beta}$ estimates well.

6. Relationship to Sample Size Formulae

In the design stage of a clinical trial with a recurrent event as an endpoint, it is often required to estimate how many subjects are needed and how long subjects are to be followed. Literature on this topic includes Cook (1995); Matsui (2005); and Song, Kosorok, and Cai (2008), where sample size formulae were given for data with overdispersed Poisson or extra-Poisson variation. Both Cook (1995) and Matsui (2005) provided formulae on sample size calculations based on a nonhomogeneous Poisson process. Their proposed methods are parametric. The intensity of the nonhomogeneous Poisson process is required as an input parameter in the sample size formulae. The method of Song et al. (2008) is based on a covariate-adjusted robust log-rank statistic. Their method requires knowledge of relevant covariates describing the baseline risk of each subject.

The results in this article on the robust variance decomposition can be applied to sample size determination. Our decomposition of the variance terms provides an intuitive explanation why a larger sample size is needed when the population is nonhomogeneous. As a simple illustration, we show how to use Corollary 2 for planning. Using an approach similar to that of Shieh (2001), the power of the study can be approximated by

$$\gamma = 1 - \Phi \left\{ \frac{z_{\alpha/2} R(0)^{\frac{1}{2}} - |\beta^*|}{R(\beta^*)^{\frac{1}{2}}} \right\},$$

which is a function of sample size n and study duration τ . Here $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution,

$$R(0) = \frac{4}{n} \left\{ \frac{1}{\tau\lambda} + \frac{\text{var}(\lambda)}{\lambda^2} \right\} \text{ and } R(\beta^*) = \frac{2\{1+\exp(\beta^*)\}}{\lambda_0 \exp(\beta^*)} \cdot \frac{1}{n\tau} + \frac{2\{\exp(\beta^*)V_{\lambda_0}^c + \exp(-\beta^*)V_{\lambda_1}^c\}}{\lambda_0^2 \exp(\beta^*)} \cdot \frac{1}{n}.$$

With $\hat{\lambda}_i = N_i(\tau_i)/\tau_i$, the quantities λ , λ_0 , $\text{var}(\lambda)$, $V_{\lambda_0}^c$ and $V_{\lambda_1}^c$ can be estimated from preliminary data. In particular, $\hat{\lambda} = n^{-1} \sum_{i=1}^n \hat{\lambda}_i$, $\widehat{\text{var}}(\lambda) = (n-1)^{-1} \sum_{i=1}^n (\hat{\lambda}_i - \hat{\lambda})^2$, $\hat{\lambda}_j = m^{-1} \sum_{\{i: x_i=j\}} \hat{\lambda}_i$, and $\hat{V}_{\lambda_j}^c = m^{-1} \sum_{\{i: x_i=j\}} (\hat{\lambda}_i - \hat{\lambda}_j)^2$. For illustration, we use the preliminary data from the diabetes study and set $\beta^* = -0.356$, which is approximately a 30% relative risk reduction; in the original dataset, $\beta = 0.125$, which is a rather neutral result and is not interesting for illustration. Figure 6 shows a contour plot of the power as a function of n and τ . Figure 6 also shows, for a given sample size, the gain in power diminishes when τ is extended. Together with other constraints, such as study cost and the recruitment rate, n and τ can be solved from the power function.

One can also use Corollary 1 to derive a sample size formula that is similar to the formula in equation (2) in Song et al. (2008). To obtain planning values needed in sample size calculation, one can fit a mixed-effects Cox model to preliminary data and read off the estimates of the variance σ_w^2 and cumulative intensity functions $\Lambda(\tau, x_i, Z_i)$.

7. Conclusions and Areas for Future Research

It might be expected that the precision of estimates of the relative risk parameters ($\hat{\beta}$) in the AG model would improve when information on the duration of the trial is extended. However, the datasets of two large well-controlled randomized trials (the SHIELD Trial and Exubera Trial) show that the returns on the robust variance of the parameter by extending the trial duration diminish. In this article, we investigated this phenomenon and illustrated the law of diminishing returns. Adding events beyond a certain time point in the trial may not lead to a reduced robust variance. We found this phenomenon arises especially when the study population is highly nonhomogenous such as in the SHIELD population. We developed analytical results, showing that one can decompose the robust variance R into two terms and determine which term dominates the variance, depending on the degree of nonhomogeneity of the population. Insights from our theoretical results are expanded through simulation studies in various settings. We also relate our analytical results to sample size calculation in clinical trial planning.

Our results show there is an interesting trade-off between the number of subjects and the length of the study duration for clinical trials with recurrent events as an endpoint. This trade-off between the number of subjects and the number of observations per subject arises in other contexts as well, including longitudinal studies with a continuous endpoint (e.g., Donner and Klar, 2000; Roy et al., 2007), and in classical survey sampling theory in the setting of cluster sampling (e.g., Särndal, Swensson, and Wretman, 2003, Chapter 12).

For future research, it will be interesting to use simulations to study the performance of the sample size formula derived from the results in this article. It will also be of interest to consider properties of $\text{var}(\hat{\beta})$ under the frailty model. The frailty can address the nonhomogeneity issue to some extent. When the frailty cannot fully address the nonhomogeneity in the population or the distribution of the frailty is misspecified, a study on the behavior of the $\text{var}(\hat{\beta})$ as the length of a trial is extended will be interesting and useful in practice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank an anonymous referee, an associate editor, and the editor for providing helpful comments and suggestions that improved this article.

References

- Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*. 1982; 10:1100–1120.
- Cook RJ. The design and analysis of randomized trials with recurrent events. *Statistics in Medicine*. 1995; 14:2081–2098. [PubMed: 8552888]
- Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*. 1972; 34:187–220.
- Cox DR. Partial likelihood. *Biometrika*. 1975; 62:269–276.
- Donner, A.; Klar, N. Design and Analysis of Cluster Randomised Trials in Health Research. London: Arnold; 2000.
- Dorian P, Borggrete M, Al Khalidi HR, Hohnloser SH, Brum JM, Tatla DS, Brachmann J, Myerburg RJ, Cannon DS, van der Laan M, Holroyde MJ, Singer I, Pratt CM. Placebo-controlled, randomized clinical trial of azimilide for prevention of ventricular tachyarrhythmias in patients with an implantable cardioverter defibrillator. *Circulation*. 2004; 110:3646–3654. [PubMed: 15533855]
- Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. Hoboken, New Jersey: John Wiley & Sons; 1991.
- Hollander R, Blonde L, Rowe R, Mehta A, Milburn J, Hershon K, Chiasson J, Levin S. Efficacy and safety of inhaled insulin in patients with type 2 diabetes: A 6-month, randomized, comparative trial. *Diabetes Care*. 2004; 27:2356–2362. [PubMed: 15451900]
- Jaekel, LA. The infinitesimal jackknife. Murray Hill, New Jersey: Bell Laboratories; 1972. Memorandum MM 72-1215-11
- Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*. 1989; 84:1074–1078.
- Matsui S. Sample size calculations for comparative clinical trials with over-dispersed Poisson process data. *Statistics in Medicine*. 2005; 24:1339–1356. [PubMed: 15568188]
- Roy A, Bhaumik DK, Aryal S, Gibbons RD. Sample size determination for hierarchical longitudinal designs with differential attrition rates. *Biometrics*. 2007; 63:699–707. [PubMed: 17825003]
- Särndal, C-E.; Swensson, B.; Wretman, J. Model Assisted Survey Sampling. New York: Springer-Verlag; 2003.
- Shieh G. Sample size calculations for logistic and Poisson regression models. *Biometrika*. 2001; 88:1193–1199.
- Song R, Kosorok MR, Cai J. Robust covariate-adjusted log-rank statistics and corresponding sample size formula for recurrent events data. *Biometrics*. 2008; 64:741–750. [PubMed: 18162107]
- Therneau, TM.; Grambsch, PM. Modeling Survival Data: Extending the Cox Model. New York: Springer; 2000.

Figure 1.
Relationship between the robust variance of $\hat{\beta}$ and the follow-up time in the SHIELD data.

Figure 2.
Number of events per week in the SHIELD data (a) and diabetes data (b).

Figure 3. Relationship between the robust variance of $\hat{\beta}$ and the follow-up time in the diabetes data.

Figure 4.
Estimated variance terms in SHIELD data (a) and diabetes data (b).

Figure 5. Results showing the mean of the estimated relative risk and variance from $B = 1000$ simulated trials.

Figure 6.
Contour plot of the power as a function of n and τ .