Journal of Biopharmaceutical Statistics, 25: 1–15, 2015

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543406.2014.919935

Taylor & Francis
Taylor & Francis Group

SAMPLE SIZE CALCULATION FOR COUNT DATA IN COMPARATIVE CLINICAL TRIALS WITH NONUNIFORM PATIENT ACCRUAL AND EARLY DROPOUT

Huiling Li[1], Lin Wang[2], Lynn Wei[2], and Hui Quan[2]

[1]*Department of Statistical Science, Forest Research Institute, Harborside Financial Center - Plaza V, Jersey City, New Jersey, USA*

[2]*Department of Biostatistics and Programming, Sanofi, Bridgewater, New Jersey, USA*

In clinical trials with counts of recurrent event data, it is often of particular interest to test whether the experimental treatment reduces the event rate in comparison with a control. The sample size calculation for such a trial often assumes fixed follow-up time for each patient. In many trials, however, we follow all patients for a predetermined follow-up time after the end of the accrual period, in which case the follow-up time is variable. This article provides methods for sample size calculation for clinical trials with ordinary Poisson count data and variable follow-up time allowing for nonuniform accrual and early dropouts. We also generalize the sample size formula to count data with overdispersion.

Key Words: Count data; Event rate; Nonuniform accrual; Overdispersion; Sample size.

1. INTRODUCTION

In clinical trials with counts of recurrent event data, the clinically important event is transient and may occur repeatedly over a period of follow-up. Examples include relapses in clinical trials for patients with relapsing–remitting multiple sclerosis (Committee for Medicinal Products for Human Use, 2006), occurrence of seizures in epileptic patients (Albert, 1991), the number of hospitalizations occurring during a follow-up period, and so on. A beneficial experimental treatment would reduce the number of recurrences. Thus, the event rate (number of occurrences per unit of time) is a key parameter in the analysis of such data. The analysis of recurrent events is often implemented by an ordinary Poisson regression model assuming the variance is equal to the mean. However, when the count data are overdispersed due to unobserved heterogeneity and/or excess zeros, the variance is larger than the mean, and analysis by ordinary Poisson regression could lead to underestimation of standard errors. Several methods have been established to account for the overdispersion. One method is based on a quasi-likelihood approach (Wedderburn, 1974; McCullagh and Nelder, 1989) that assumes that the variance is proportional to the mean, with a factor greater than unity representing overdispersion. An alternative method is negative binomial regression model, which assumes the variance is a quadratic function of the mean (Keene et al., 2007; Wang et al., 2009). The validity of the method depends on

Received November 15, 2012; Accepted June 18, 2013
Address correspondence to Huiling Li, Department of Statistical Science, Forest Research Institute, Harborside Financial Center - Plaza V, Jersey City, NJ 07311, USA; E-mail: huiling.li@frx.com

the assumptions about the distribution and variance function form. Another method is the Poisson regression model with robust error variance (Diggle et al., 1994). The variance is based on the robust "sandwich" method to account for the correlated counts.

The sample size calculation for this kind of trials typically assumes fixed follow-up time or constant exposure time for each patient (Thode, 1997; Ng and Tang, 2005; Friede and Schmidli, 2010a, 2010b; Kappos et al., 2010), which fits well for clinical trial designs in which we follow each individual for a fixed length of time. In many clinical trials, however, we follow all patients from recruitment until an additional predetermined follow-up time after the end of the accrual period. We call this a common end time design because all patients finish the study at the same time. In this setting, the follow-up time varies across patients because those who are recruited earlier will be followed for a longer time than those who are recruited later. Thus, in sample size calculations, we use an expected average follow-up time. To ease the calculation of the average follow-up time, one typically assumes a uniform accrual pattern or constant accrual rate. In practice, however, patient entry is not always uniform.

An instructive example of disease area is relapsing multiple sclerosis, in which disease activity is commonly measured by the number of relapses. In clinical trials of this disease, we often experience a lagging patient enrollment at the early stages due to study start-up issues, such as differences in regulatory requirements between countries, site contracting, drug importing, investigator training, and so on, which can significantly delay the initiation of sites and affect recruitment rates. After many sites are open for full enrollment, recruitment typically speeds up. On the other hand, we sometimes observe faster recruitment at the early stages, thanks to pent-up demand for new treatments, followed by a gradual slowing as patient pools are exhausted and some sites close. The sample size calculated without considering the departure from the uniform entry may either underpower or overpower a study.

Another reason that leads to variable follow-up time is early treatment discontinuation or early dropouts due to various reasons among some of the patients. Under the model-based analyses of count data, it is often assumed that the dropout is random and that the dropout process is independent of the event process (Keene et al., 2007). A simple method to account for potential dropouts is to inflate the sample size by the factor $1/(1-\text{dropout rate})$. This approach assumes that the dropout patients will not contribute to achieve the desired power for testing a given hypothesis, although they are included in analysis following the intention-to-treat principle. To account for the partial information from the dropouts, we provide an alternative approach by incorporating the dropout time into the sample size calculation. We consider the special case assuming that the dropout time follows an exponential distribution, with hazard rates independent of the event process. The exponential dropout time assumption is used only for power calculation and not for data analysis.

We calculate the average follow-up time assuming a truncated exponential entry distribution (Johnson and Kotz, 1970) over the accrual period in clinical trial designs with a common end time. The distribution for entry time can be concave (i.e., lagging patient entry), convex (fast patient entry), or uniform depending on the sign of the parameter in the distribution for the patient's accrual pattern. The dropout time is accommodated assuming an exponential distribution with different hazard rates in the two treatment groups. Considering both accrual pattern and patient dropouts, we derive the formula to compute the average follow-up time. The developed method can be used for sample size calculation for trials of count data with or without overdispersion. For illustrative purpose, a quasi-likelihood approach is used for sample size calculation on overdispersed count data as it does not need full knowledge of the distribution of the data. Keene et al. (2007) provided

a sample size formula in a two-arm superiority trial assuming that the overdispersed count data follow a negative binomial distribution. Friede and Schmidli (2010b) developed a method for blinded sample size reestimation with negative binomial counts in superiority and noninferiority trials. Similar to the sample size formulas using the quasi-likelihood approach, the sample size for negative binomial counts also depends on the follow-up time. When the follow-up time is not constant across patients due to either recruitment pattern or patient dropout, the expected average follow-up time developed in this article can be applied. In the Appendix, we provide *R* functions to calculate the sample size.

The remainder of this article is organized as follows. In Section 2, we derive the sample size formula for tests of equality, noninferiority/superiority, and equivalence for ordinary Poisson count data in a clinical trial with constant follow-up. We then generalize the basic formula to count data with overdispersion. In Section 3, we extend the formula to clinical trials with variable follow-up time due to nonuniform accrual pattern and/or early dropouts. Examples are provided in Section 4 to illustrate our methodology and to discuss the impact of nonuniform accrual. In Section 5, we examine the finite-sample properties of the sample size formula by simulation studies. Section 6 presents concluding remarks and discussions.

2. SAMPLE SIZE FORMULA IN CLINICAL TRIALS WITH FIXED FOLLOW-UP TIME

2.1. Test for Equality

2.1.1. Ordinary Poisson count data without overdispersion. Suppose that a two-arm trial is designed to compare an experimental treatment with a control, and that we follow all patients in both arms for the same length of time. Let X_{ij} denote the observed counts of recurrent events for patient j in i th treatment group for $i = 1, 2$ ($i = 1$ for the control group and $i = 2$ for the treatment group) and $j = 1, 2, n_i$ where n_i denotes the number of patients in the i th treatment group. Let X_i denote the sum of the observed counts of the patients in the i th treatment group. Assume X_{ij} follows a Poisson distribution, with mean and variance $E(X_{ij}) = var(X_{ij}) = \lambda_i$. Suppose θ_i is the Poisson event rate (counts of recurrent events per unit of time) in the i th treatment group and h_i is the fixed length of follow-up for each patient in the i th treatment group. Since we follow all patients in both arms for the same length of time, h_1 and h_2 are the same if no differential dropout is considered. Then, $\lambda_i = \theta_i h_i$ for $i = 1, 2$. Let $\kappa = n_2/n_1$ be the patient allocation ratio between the two groups. The total follow-up time in the two treatment groups will be given by $l_1 = n_1 h_1$ and $l_2 = n_2 h_2 = \kappa n_1 h_2$, respectively.

The treatment effect is measured in terms of the logarithm of the risk ratio $\delta = \log(\theta_2/\theta_1)$, with a smaller event rate ratio indicating a better treatment outcome. To test for equality, the following hypotheses are usually considered:

$$H_0 : \delta = 0 \quad \text{vs.} \quad H_1 : \delta \neq 0.$$

The maximum likelihood estimators (MLEs) of the event rates are given by $\hat{\theta}_i = X_i/l_i$, and the variance of the MLEs is θ_i/l_i, for $i = 1, 2$. Asymptotically, $\hat{\theta}_i \sim N(\theta_i, \theta_i/l_i)$. Via the delta method, we have $\log \hat{\theta}_i \sim N\left(\log \theta_i, \frac{1}{\theta_i l_i}\right)$. It follows that the variance of $\log \hat{\theta}_i$ can be estimated by $1/X_i$. Therefore a Wald-type test statistic is given by (Ng and Tang, 2005; Friede and Schmidli, 2010a)

$$Z = \frac{\log \hat{\theta}_2 - \log \hat{\theta}_1}{\sqrt{1/X_2 + 1/X_1}},$$

which has an asymptotic normal distribution $N(\omega,\ 1)$, where

$$\omega = \frac{\delta}{\sqrt{1/(\theta_2 l_2) + 1/(\theta_1 l_1)}} = \frac{\log(\theta_2/\theta_1)}{\sqrt{1/(\theta_2 n_2 h_2) + 1/(\theta_1 n_1 h_1)}} = \frac{\log(\theta_2/\theta_1)}{\sqrt{1/(\theta_2 \kappa n_1 h_2) + 1/(\theta_1 n_1 h_1)}}$$

Under the null hypothesis of no treatment effect, Z asymptotically follows a standard normal distribution $N(0,\ 1)$. Hence, we reject H_0 if $|Z| \geq Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

In the study design, it is of interest to determine the sample size in each treatment group for reaching a desired study power of $1-\beta$ for given event rate and length of follow-up. The power is then given by

$$P\left\{|N(\omega,\ 1)| \geq Z_{1-\alpha/2}\right\} \approx \Phi(|\omega| - Z_{1-\alpha/2}) = \Phi(\frac{|\log(\theta_2/\theta_1)|}{\sqrt{1/(\theta_2 \kappa n_1 h_2) + 1/(\theta_1 n_1 h_1)}} - Z_{1-\alpha/2}),$$

where $\Phi(x)$ is the cumulative probability function of the standard normal distribution. The sample size needed to achieve the desired power of $1-\beta$ can be obtained by solving

$$\frac{|\log(\theta_2/\theta_1)|}{\sqrt{1/(\theta_T \kappa n_1 h_2) + 1/(\theta_1 n_1 h_1)}} - Z_{1-\alpha/2} = Z_{1-\beta}$$

as

$$\begin{aligned} n_1 &= \left((Z_{1-\alpha/2} + Z_{1-\beta}) \sqrt{\frac{1}{\theta_2 \kappa h_2} + \frac{1}{\theta_1 h_1}} \Big/ \log\left(\theta_2/\theta_1\right) \right)^2 \\ &= \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\left[\log\left(\theta_2/\theta_1\right)\right]^2} \left(\frac{1}{\theta_2 \kappa h_2} + \frac{1}{\theta_1 h_1} \right) \end{aligned} \tag{1}$$

and

$$n_2 = \kappa n_1.$$

2.1.2. Extension to count data with overdispersion. Count data sometimes exhibit overdispersion due to unobserved heterogeneity, which manifests itself as the variance exceeding the mean. Assuming that the variance is proportional to the mean and that overdispersion is the same across treatment groups, one can define the overdispersion factor $\phi = \text{var}(X_{ij})/\lambda_i$. The assumption of constant overdispersion across treatment arms is common, and results are often robust to misspecification (McCullagh and Nelder, 1989; Friede and Schmidli, 2010a).

Introducing the dispersion parameter is a natural way to extend the ordinary Poisson distribution without making specific assumptions on the mechanism of the overdispersion. If $\phi = 1$, then the variance is equal to the mean and the distribution reduces to an ordinary Poisson. If $\phi > 1$, the variance is larger than the mean, and the distribution represents count data with overdispersion. If $\phi < 1$, the variance is smaller than the mean, and the distribution represents count data with under-dispersion. Various estimators are

available for the dispersion parameter. For example, deviance and Pearson Chi-square statistics divided by the degrees of freedom are often used to quantify overdispersion or under-dispersion.

Using the method described above, we have

$\log \hat{\theta}_i \sim N\left(\log \theta_i, \frac{\phi}{\theta_i l_i}\right)$ for $i = 1, 2$ (Friede and Schmidli, 2010a)

The null hypothesis H_0 is tested with the Wald-type test statistic by multiplying the variance term by $\tilde{\phi}$, the estimator of the dispersion parameter. Consequently, the test statistic is $Z^{(O)} = Z/\sqrt{\tilde{\phi}}$. Similar to Section 2.1.1, a normal approximation is used to derive the sample size formula, and we have

$$
\begin{aligned}
n_1{}^{(O)} &= \phi n_1 = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\left[\log\left(\theta_2/\theta_1\right)\right]^2}\left(\frac{\phi}{\theta_2 \kappa h_2} + \frac{\phi}{\theta_1 h_1}\right) \\
n_2{}^{(O)} &= \kappa n_1{}^{(O)}
\end{aligned}
\tag{2}
$$

Equation (2) is a general sample size calculation formula for count data with or without overdispersion. By setting $\phi = 1$, Equation (2) reduces to Equation (1), the sample size formula for ordinary Poisson count data without overdispersion.

The above method to handle overdispersion does not require full knowledge of the distribution of the data. As an alternative, Keene et al. (2007) provided a sample size formula in a two-arm superiority trial assuming that the overdispersed count data follows a negative binomial distribution. For a negative binomial distribution, the relationship between variance and mean is characterized by $\text{var}(X_{ij}) = \lambda_i + \tau {\lambda_i}^2$, where λ_i represents the mean of the count data in the i th treatment group and τ is a shape parameter. The negative binomial distribution reduces to the Poisson as the shape parameter converges to 0. Using the above notations and assuming equal sample size and equal average follow-up time for the two treatment groups (h), the sample size formula developed by Keene et al. (2007) for negative binomial distributed data can be rewritten as

$$
n_1{}^{(O)} = n_2{}^{(O)} = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2}{\left[\log\left(\theta_2/\theta_1\right)\right]^2}\left[\left(\frac{1}{\theta_2} + \frac{1}{\theta_1}\right)/h + 2\tau\right].
$$

As in Equation (2), the sample size is a function of the follow-up time. When the follow-up time for individual patient is not a fixed constant due to either recruitment pattern or patient dropout, the expected average follow-up time developed in Section 3 can be used to compute h and then the sample size.

2.2. Test for Noninferiority/Superiority

As discussed in Section 2.1, the treatment difference is characterized in terms of the logarithm of the risk ratio $\delta = \log(\theta_2/\theta_1)$, where θ_1 and θ_2 are the event rates of the control and experimental treatments, respectively. In practice, a smaller event rate ratio (or negative δ) indicates a beneficial treatment effect. Therefore, the testing of superiority and noninferiority, with a smaller value indicating better treatment benefits, can be unified by the following hypotheses:

$$H_0 : \delta \geq \varepsilon \text{ vs. } H_1 : \delta < \varepsilon,$$

where ε is the superiority or noninferiority margin. When $\varepsilon \leq 0$, the rejection of the null hypothesis implies superiority of the experimental treatment over the control. When $\varepsilon > 0$, the rejection of the null hypothesis implies noninferiority of the experimental treatment against the control. Using similar derivations as described in Section 2.1.2, the sample size needed to achieve power $1 - \beta$ for testing superiority and noninferiority at significance level α is given by

$$\begin{aligned} n_1{}^{(O)} &= \frac{(Z_{1-\alpha} + Z_{1-\beta})^2}{\left[\log\left(\theta_2/\theta_1\right) - \varepsilon\right]^2}\left(\frac{\phi}{\theta_2 \kappa h_2} + \frac{\phi}{\theta_1 h_1}\right) \\ n_2{}^{(O)} &= \kappa n_1{}^{(O)}. \end{aligned} \tag{3}$$

2.3. Test for Equivalence

Treatment equivalence can be established by testing the following hypotheses:

$$H_0 : |\delta| \geq \varepsilon \text{ vs. } H_1 : |\delta| < \varepsilon.$$

The objective of the test is to reject the null hypothesis of non-equivalence and conclude the alternative hypothesis of equivalence. Similarly, the sample size needed to achieve power $1 - \beta$ is given by

$$\begin{aligned} n_1{}^{(O)} &= \frac{(Z_{1-\alpha} + Z_{1-\beta/2})^2}{\left[\varepsilon - \left|\log\left(\theta_2/\theta_1\right)\right|\right]^2}\left(\frac{\phi}{\theta_2 \kappa h_2} + \frac{\phi}{\theta_1 h_1}\right) \\ n_2{}^{(O)} &= \kappa n_1{}^{(O)}. \end{aligned} \tag{4}$$

3. SAMPLE SIZE CALCULATION FOR CLINICAL TRIAL WITH VARIABLE FOLLOW-UP

Equations (1)–(4) reveal that the sample size depends on the follow-up time. In clinical trials with predetermined fixed follow-up time, follow-up time is supposed to be the same for all patients in both arms. However, the length of follow-up changes due to different dropout patterns. In the trial design with a common end time that follows all patients for an additional predetermined follow-up time after the end of accrual, or until a predetermined total number of events has been observed, follow-up time varies across patients. As a result, h_i is essentially the average follow-up time across all patients in the i th treatment group. In Section 3.1, we derive a formula for the average follow-up time for the trial design with common end time, considering different distributions for entry time and dropout time. In Section 3.2, we derive the formula to compute the average follow-up time considering early dropouts with predetermined fixed follow-up. One obtains general sample size formulas by replacing h_i in Equations (2), (3), and (4) with the calculated average follow-up times given in Sections 3.1.3, 3.1.4, and 3.2.

3.1. Average Follow-Up Time for Clinical Trials with Common End Time and Early Dropout

3.1.1. Patient accrual pattern. Consider a two-arm parallel-group clinical trial with accrual time period T_0 and additional follow-up time $T - T_0$ after the accrual period. The total length of the trial is then T, and $T > T_0$. Using the notation given in Section 2, let h_i denote the average follow-up time (or expected follow-up time) for the i th treatment group, $i = 1, 2$.

Assume that patient entry times are independent, and that patient j in arm i enters the study with entry time a_{ij}, $0 \le a_{ij} \le T_0$. Assume moreover that a_{ij} follows a truncated exponential distribution over the interval 0 to T_0 with the density function given by

$$g(a_{ij}) = \begin{cases} \frac{\gamma e^{-\gamma a_{ij}}}{1-e^{-\gamma T_0}}, 0 \le a_{ij} \le T_0, \gamma \ne 0 & \text{(Non-uniform accrual, Johnson and Kotz, 1970)} \\ 1/T_0 \text{ when } \gamma = 0 & \text{(Uniform accrual)} \end{cases}.$$

Here, γ is a parameter indexing the accrual pattern. For $\gamma > 0$, the pattern is convex, indicating a declining accrual rate over time. For $\gamma < 0$, the accrual pattern is concave, indicating an increasing accrual rate. Uniform enrollment is indicated by $\gamma = 0$, or $g(a) = 1/T_0$.

3.1.2. Early dropout. Some patients might discontinue the study early for various reasons. Let t_{ij} be the dropout time for the j th patient in the i th treatment group, with $i = 1, 2$ and $j = 1, 2, n_i$. We assume that the dropout time follows an exponential distribution with hazard rate μ_i and is independent of the event process. Therefore, the probability density function is $f(t_{ij}) = \mu_i e^{-\mu_i t_{ij}}$, and the survival function is $S(t_{ij}) = e^{-\mu_i t_{ij}}$.

3.1.3. Average follow-up time with nonuniform accrual pattern and exponential dropout. Given the patient entry distribution $g(a_{ij})$ and dropout time distribution $f(t_{ij})$, the average follow-up time considering both the accrual pattern and early dropout is

$$\begin{aligned} h_i = E(t_{ij}) &= \int_0^{T_0} g(a)da \left[\int_0^{T-a} t\mu_i e^{-\mu_i t} dt + (T-a)e^{-\mu_i(T-a)} \right] \\ &= \int_0^{T_0} \frac{\gamma e^{-\gamma a}}{1-e^{-\gamma T_0}} da \left[(-te^{-\mu_i t} - \tfrac{1}{\mu_i} e^{-\mu_i t})\Big|_0^{T-a} + (T-a)e^{-\mu_i(T-a)} \right] \\ &= \int_0^{T_0} \frac{\gamma e^{-\gamma a}}{1-e^{-\gamma T_0}} \left[\frac{1-e^{-\mu_i(T-a)}}{\mu_i} \right] da = \int_0^{T_0} \frac{\gamma e^{-\gamma a} - \gamma e^{-\gamma a - \mu_i T + \mu_i a}}{\mu_i(1-e^{-\gamma T_0})} da \\ &= \frac{\int_0^{T_0} \gamma e^{-\gamma a} da - \int_0^{T_0} \gamma e^{-\gamma a - \mu_i T + \mu_i a} da}{\mu_i(1-e^{-\gamma T_0})} \\ &= \frac{(1-e^{-\gamma T_0}) - \frac{\gamma}{\mu_i - \gamma} e^{-\mu_i T}(e^{(\mu_i - \gamma)T_0} - 1)}{\mu_i(1-e^{-\gamma T_0})} \\ &= \frac{1}{\mu_i} + \frac{\gamma e^{-\mu_i T}(1 - e^{(\mu_i - \gamma)T_0})}{\mu_i(\mu_i - \gamma)(1-e^{-\gamma T_0})} \end{aligned} \tag{5}$$

The sample size is then evaluated by substituting (5) into formulas (2), (3), or (4). According to Equation (5), given the same accrual pattern parameter (γ), dropout hazard rate (μ_i), and the total length of trial (T), the average follow-up time (h_i) decreases with the increase of the accrual time (T_0). Therefore, one will need to increase the sample size to maintain the same power. On the other hand, the average follow-up time (h_i) increases with the increase of the total length of trial (T), and the sample size needed will be smaller.

3.1.4. Average follow-up time with uniform accrual and exponential dropout. Under the uniform accrual pattern, $\gamma = 0$. As γ goes to 0, formula (5) converges to

$$h_i = \frac{1}{\mu_i} + \frac{e^{-\mu_i T}(1 - e^{\mu_i T_0})}{T_0 {\mu_i}^2}. \tag{6}$$

The sample size is then evaluated by substituting (6) into formulas (2), (3), or (4).

3.2. Average Follow-Up Time for Clinical Trials with Predetermined Fixed Follow-Up Time for Each Individual and Early Dropout

For the clinical trial with predetermined fixed follow-up time for each patient, the average follow-up time depends not on the accrual pattern but only on the follow-up time and dropout time, if any. The predetermined follow-up time for each patient is constant, and it is $T - T_0$, with T_0 as the accrual time and T as the total length of the trial. Using the same dropout distribution as that in Section 3.1, the average follow-up time for the i th treatment group is

$$\begin{aligned} h_i = E(t_{ij}) &= \int_0^{T-T_0} t\mu_i e^{-\mu_i t} dt + (T - T_0)e^{-\mu_i(T-T_0)} \\ &= -(T - T_0)e^{-\mu_i(T-T_0)} - \frac{1}{\mu_i}e^{-\mu_i(T-T_0)} + \frac{1}{\mu_i} + (T - T_0)e^{-\mu_i(T-T_0)} \\ &= \frac{1}{\mu_i} - \frac{1}{\mu_i}e^{-\mu_i(T-T_0)} \end{aligned} \tag{7}$$

The sample size is then evaluated by substituting (7) into formulas (2), (3), or (4).

4. APPLICATIONS

4.1. Effects of Patient Accrual Pattern on Power and Sample Size

Suppose we plan to design a study for relapsing multiple sclerosis to compare an experimental therapy with a control based on a two-sided test. The annualized relapse rate (ARR), defined as the total number of relapses divided by the total person-years of follow-up, is the commonly used primary efficacy endpoint in this disease. Nicholas et al. (2011) carried out a systematic literature review in relapsing multiple sclerosis and characterized the uncertainty of ARR in the placebo group by modeling the ARR over time. In all trials included in the review (apart from one small trial), placebo ARRs were in the range

0.3–1.5 with a decrease of 6.2% per year, resulting in substantial uncertainty in the planning of future trials. Even for trials published within a short span of time, as identified in the review, considerable variation remains. For instance, the placebo ARR for the seven trials published in 2008 varied between 0.44 and 0.84. In our example, we assume ARR values of 0.40 in the control group and 0.28 in the experimental therapy group, representing a treatment effect of 40%. The length of study is expected to be 3 years ($T = 3$) with 2 years of enrollment ($T_0 = 2$). All patients will be followed until the last patient enrolled is followed for 1 year, and we assume a 25% dropout rate (with a corresponding hazard rate of 0.1438) for both treatment groups. The overdispersion factor is assumed to be 1.3 for both arms. The overall two-sided type I error rate $\alpha = 0.05$ is used, and the study power is set to 90% ($\beta = 0.1$). The allocation ratio is 1:1 between the two treatment groups ($n_1{}^{(O)} = n_2{}^{(O)} = n$).

Assuming a uniform accrual pattern and calculating the average follow-up time by (6) and the sample size by (2), we obtain a total sample size of 380/arm. Assuming a concave (i.e., lagging) accrual pattern with 25% recruitment (rather than 50%) observed at year 1 gives $\gamma = -1.1$. Using formulas (5) and (2) gives a sample size of 446/arm, an increase of 17% over uniform accrual.

Using the same parameters and assumptions, we investigate the effects of nonuniform patient entry on the power of a study designed assuming uniform accrual. Assuming n=380/arm as arrived at above, we calculate the power for various values of γ ranging from –1.5 to 0.5. The upper panel of Fig. 1 plots power against γ. As we can see, the power is decreased when γ is negative (a lagging enrollment), while it is increased when γ is positive (a faster enrollment than uniform). The lower panel in Fig. 1 plots the sample size required to achieve power of 0.90 against the accrual pattern parameter γ. The required sample size increases when γ is negative (a lagging enrollment) while it decreases when γ is positive (a faster enrollment than uniform). For example, we need $n = 446$/arm to provide a power of 0.90 when $\gamma = -1.1$. The power would be 0.849 if we did not adjust the sample size.

4.2. Effects of Dropout on Power and Sample Size

Using the same parameters and assumptions as above, and setting $\gamma = -1.1$, we investigate the effects of dropout on power and sample size. If there are no dropouts, we need $n = 393$/arm to provide a power of 0.9 to detect a reduction in event rate from $\theta_1 = 0.40$ to $\theta_2 = 0.28$ with $T = 3$, $T_0 = 2$, and $\gamma = -1.1$. Table 1 shows the sample size required to maintain the power of 0.9 under different dropout scenarios. In every case, time to dropout is assumed to be exponentially distributed. For example, for a large dropout rate of 25% in both arms, the sample size must be increased from $n = 393$/arm to $n = 446$/arm to retain 90% power. If we increase the sample size by the factor 1/(1–dropout rate), we need $n = 393/(1 - 0.25) = 524$ patients per arm. Thus, this simple adjustment is unnecessarily conservative. Table 1 also provides the sample size in the event of differential dropout rates between the two treatment groups. It would be unusual to see 0% dropout in one arm and 25% dropout in the other, but we provide these data to explore the limits of what one would be likely to observe in practice.

Table 2 provides the actual power in the presence of dropout if there is no sample size adjustment. Using the sample example, the actual study power would be 0.861 in the presence of 25% dropout rate in both arms without sample size increase. In summary, the loss in power or increase in required sample size is roughly proportional to the total dropout

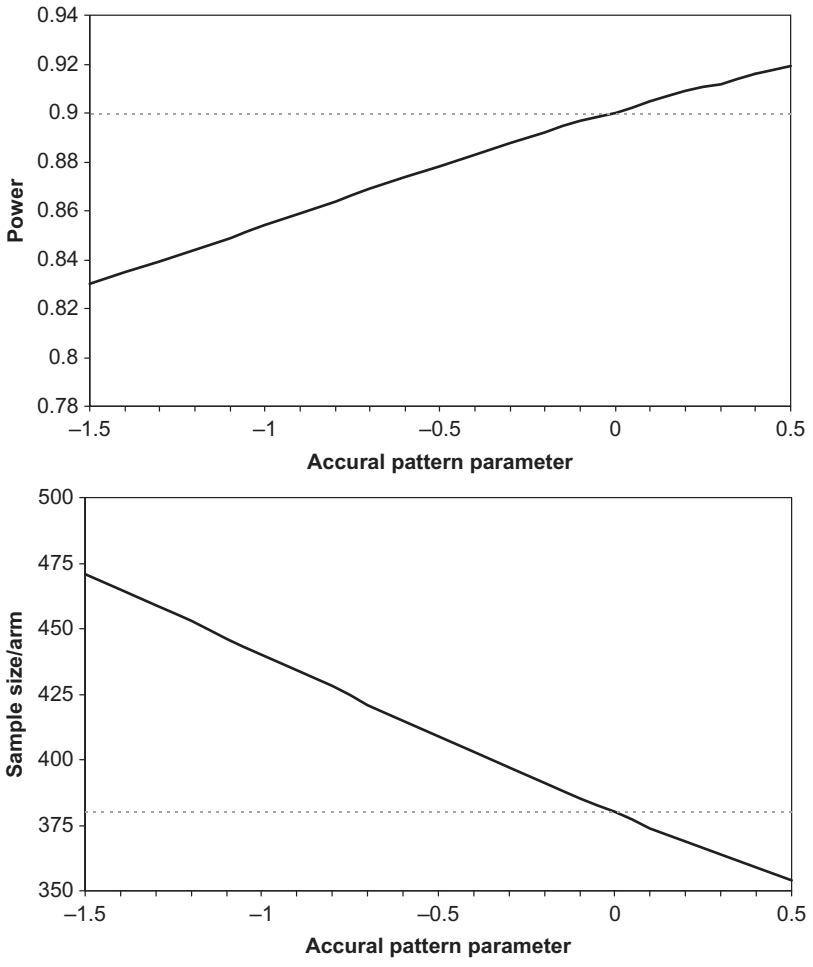

Figure 1 Plot of power vs. accrual pattern parameter γ with $n = 380$/arm (upper panel) and plot of sample size/arm vs. accrual pattern parameter γ for power = 0.90 (lower panel) ($T = 3$, $T_0 = 2$, $\theta_2 = 0.28$, $\theta_1 = 0.40$, two-sided $\alpha = 0.05$, 25% dropout).

Table 1 Sample size/arm as a function of dropout rates to achieve a power of 0.9

	Dropout rate in experimental treatment					
Dropout rate in control	0%	5%	10%	15%	20%	25%
0%	393	399	404	411	417	425
5%	397	402	408	414	421	428
10%	401	407	412	419	425	432
15%	405	411	417	423	430	437
20%	410	416	421	428	434	441
25%	415	421	426	433	439	446

Note. ($T = 3$, $T_0 = 2$, $\theta_2 = 0.28$, $\theta_1 = 0.40$, $\alpha = 0.05$ (two-sided), $\gamma = -1.1$).

Table 2 Power as a function of dropout rates with a sample size of 393/arm

	Dropout rate in experimental treatment					
Dropout rate in control	0%	5%	10%	15%	20%	25%
0%	0.900	0.896	0.892	0.887	0.882	0.877
5%	0.897	0.893	0.889	0.885	0.880	0.874
10%	0.894	0.890	0.886	0.882	0.877	0.871
15%	0.891	0.887	0.883	0.878	0.873	0.868
20%	0.888	0.884	0.880	0.875	0.870	0.864
25%	0.884	0.880	0.876	0.871	0.866	0.861

Note. ($T = 3$, $T_0 = 2$, $\theta_2 = 0.28$, $\theta_1 = 0.40$, $\alpha = 0.05$ (two-sided), $\gamma = -1.1$).

rates from the two treatment groups. The larger the dropout rate, the smaller the study power or the larger the sample size required to achieve the same power.

5. SIMULATION STUDY TO EXAMINE THE FINITE-SAMPLE PROPERTIES

The sample size formulas derived in this article are based on asymptotic results using a normal approximation. We investigate the finite-sample properties and compare the asymptotic power with the exact power by Monte Carlo simulation studies. The asymptotic power is computed according to formula (2) given the parameters, while the exact power is computed by simulation. We choose $\theta_1 = 0.8$ in all simulations and θ_2 with two values of 0.3 and 0.4. We consider equal follow-up time ($h_2 = h_1$) for both treatment groups with two values of 1.5 and 2.0. The sample size in the control group is fixed ($n_1 = 20$) and the sample size in the treatment group is either $n_2 = 20$ or $n_2 = 40$. The two-sided significance level is 5%. It is not possible to simulate overdispersed count data based on a quasi-likelihood model because the data do not follow a specific distribution. Thus, we simulated data from Poisson and negative binomial models. For the latter, we transformed the shape parameter τ_i to the overdispersion parameter ϕ by $\phi = (\lambda_i + \tau_i\lambda_i{}^2)/\lambda_i = 1 + \tau_i\lambda_i$ in order to calculate the asymptotic power using (2). If the rate parameters differ, the shape parameters must also differ in order to have a common overdispersion parameter between arms.

Given each set of parameters, we generated 10,000 replications of the dataset. We analyzed the Poisson data by Poisson regression, and the negative binomial data by quasi-likelihood. The exact power was computed as the proportion of the 10,000 replications with p-values less than 0.05. Table 3 provides a comparison of the exact and asymptotic powers. Results show that the asymptotic power is very close to the exact power for all scenarios. Thus, the sample size method is valid in practical trials, which are likely to be several times larger than the simulation examples.

6. DISCUSSION

When accrual lags in the early part of a count-based trial, the ultimate power of the study may not achieve its target, even if enrollment eventually reaches the planned target. In this article, we have derived sample size formulas for clinical trials with counts of recurrent event data and variable follow-up time, with adjustments for both the patient entry distribution and early dropout. Our method can be applied in sample size calculation

Table 3 Comparisons of exact power (based on simulation studies with 10,000 replications) and asymptotic power

θ_1	θ_2	τ_1	τ_2	ϕ	$h_2 = h_1$	n_1	n_2	Exact power	Asymptotic power
Ordinal Poisson distribution									
0.8	0.3	0.0	0.0	0.00	2.0	20	20	0.852	0.825
0.8	0.3	0.0	0.0	0.00	2.0	20	40	0.950	0.952
0.8	0.3	0.0	0.0	0.00	1.5	20	20	0.734	0.708
0.8	0.4	0.0	0.0	0.00	2.0	20	20	0.617	0.619
0.8	0.4	0.0	0.0	0.00	2.0	20	40	0.782	0.791
0.8	0.4	0.0	0.0	0.00	1.5	20	20	0.493	0.500
Negative binominal distribution									
0.8	0.3	0.3	0.8	1.48	2.0	20	20	0.691	0.663
0.8	0.3	0.3	0.8	1.48	2.0	20	40	0.848	0.847
0.8	0.3	0.3	0.8	1.36	1.5	20	20	0.604	0.576
0.8	0.4	0.4	0.8	1.64	2.0	20	20	0.433	0.423
0.8	0.4	0.4	0.8	1.64	2.0	20	40	0.601	0.581
0.8	0.4	0.4	0.8	1.48	1.5	20	20	0.373	0.363

for testing equality, noninferiority/superiority, and equivalence, and for count data with or without overdispersion.

Through examples, we investigated the effects of a concave accrual pattern on power and sample size. As shown in Section 4, the larger the magnitude of the patient accrual pattern parameter γ, the more significantly the distribution deviates from uniform and the larger the reduction in power. Additionally, both accrual time and total length of trial affect the sample size. The sample size increases with the increase of the accrual time, while it decreases with the increase of the total length of trial. When the accrual pattern is convex, with more rapid accrual early in the enrollment period, the effects are reversed, and power will exceed nominal levels.

We also investigated the effects of dropout on power and sample size. The decline in power (increase in required sample size) is roughly proportional to the total dropout rates from the treatment groups. To analyze the count data using model-based approaches, it is often assumed that the dropout process is independent of the event process. In practice, the dropout mechanism needs to be further investigated. If the dropout is associated with the event process, estimates may be biased. One can determine the potential magnitude of these effects through a sensitivity analysis.

Patient recruitment is obviously a critical element of any clinical trial. Departures from the planned or anticipated accrual pattern can affect study duration, sample size, and power. Carter (2004) considered a Poisson process model for patient recruitment. Anisimov and Fedorov (2005, 2007) constructed a Gamma-Poisson model, which allows the recruitment rate to vary according to a Gamma distribution. Gajewski et al. (2008) developed a Bayesian model to include prior probabilities and refine the predictions by bringing new evidence as data accumulates. We investigated the truncated exponential entry distribution (Johnson and Kotz, 1970) over the recruitment period. When planning a clinical trial, however, it is often difficult for parameters to be pre-specified especially when no previous studies have been done. Sensitivity power calculations can be performed to assess the impact of parameter assumptions on study power. Additionally, interim review of the

recruitment curve can be done, and the estimates based on the interim will be used to adjust the sample size.

Constant event rates over time are assumed in our sample size calculation. In practice, the event rates might increase or decrease as the trials progress, which results in substantial uncertainty in the event rate assumptions. As noted in the recent systematic reviews carried out by Nicholas et al. (2011, 2012), the relapse rate decreased over time in a series of randomized, placebo-controlled trials in relapsing multiple sclerosis. In the planning stage of clinical trials, overtime average event rate can be considered for sample size calculation. Additionally, adaptive strategies such as blinded sample size reestimation procedure (Friede and Schmidli, 2010a, 2010b) can be applied to maintain the targeted power.

APPENDIX: *R* PROGRAM FOR SAMPLE SIZE CALCULATION

```
## R program to calculate the sample size (test for
equality)
## entry: T for uniform patient entry and F for nonuniform
patient entry
## drop_trt: dropout rate in the treatment group
## drop_pbo: dropout rate in the placebo group
## r: The parameter indicating patient's accrual pattern
## t0: accrual time of the trial
## t: total length of the trial
## r_trt: the event rate in the treatment group
## r_pbo: the event rate in the placebo group
## disp: overdispersion parameter
## alpha: 2-sided type I error rate
## beta: study power
## ratio: allocation ratio between treatment group and
placebo group
sz.count<-function(entry, drop_trt, drop_pbo, r, t0, t,
r_trt, r_pbo, disp, alpha, beta, ratio) {
    ## assume dropout follow an exponential distribution
with hazard rate mu
    mu_trt<- -log(1-drop_trt)/(t-0.5 × t0)
    mu_pbo<- -log(1-drop_pbo)/(t-0.5 × t0)
    ## Nonuniform patient entry
      if(entry==F)
      {up_trt<- r × exp(-mu_trt × t) × (1-exp((mu_trt-r) ×
        t0))
      down_trt<- mu_trt × (mu_trt-r) × (1-exp(-r ×
        t0))
      up_pbo<- r × exp(-mu_pbo × t) × (1-exp((mu_pbo-r) ×
        t0))
        down_pbo<- mu_pbo × (mu_pbo-r) × (1-exp(-r × t0)) }
    ## Uniform patient entry
        if(entry ==T)
            {up_trt<- exp(-mu_trt × t) × (1-exp(mu_trt × t0))
            down_trt<- t0 × mu_trt × mu_trt
```

```
        up_pbo<- exp(-mu_pbo × t) × (1-exp(mu_pbo × t0))
        down_pbo<- t0 × mu_pbo × mu_pbo }
t_trt<- 1/mu_trt + up_trt/down_trt
t_pbo<- 1/mu_pbo + up_pbo/down_pbo
delta<- log(r_trt/r_pbo)
n1<-((qnorm(beta)+qnorm(1-alpha/2)) × sqrt(disp ×
(1/r_trt/t_trt/ratio+1/r_pbo/t_pbo))/delta)^2
n1<-ceiling(n1)
n2<-n1 × ratio
list(exposure.placebo=t_pbo, exposure.treatment=t_trt,
sample.size.placebo = n1, sample.size.treatment = n2) }
sz.count(entry=F, drop_trt=0.25, drop_pbo=0.25, r=-1.1,
t0=2, t=3, r_trt=0.28, r_pbo=0.4, disp=1.3, alpha=0.05,
beta=0.90, ratio=1)
sz.count(entry=T, drop_trt=0.25, drop_pbo=0.25, r=,
t0=2, t=3, r_trt=0.28, r_pbo=0.4, disp=1.3, alpha=0.05,
beta=0.90, ratio=1)
```

ACKNOWLEDGMENTS

The authors are grateful to the three referees, the editor, and Daniel Heitjan for insightful comments that improved the presentation of this article.

REFERENCES

Albert, P. S. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* 47:1371–1381.

Anisimov, V., Fedorov, V. (2005). Modeling of enrolment and estimation of parameters in multicentre trials. *GSK BDS Technical Report, 2005-01*.

Anisimov, V., Fedorov, V. (2007). Modeling, prediction and adaptive adjustment of recruitment in multicentre trials. *Statistics in Medicine* 26:4958–4975.

Carter, R. E. (2004). Application of stochastic processes to participant recruitment in clinical trials. *Controlled Clinical Trials* 25:429–436.

Committee for Medicinal Products for Human Use (CHMP, 2006). Guideline on clinical investigation of medicinal products for the treatment of multiple sclerosis. Doc. Ref. EMEA/CHMP/EWP/561/98 Rev1, London.

Diggle, P. J., Liang, K. Y., Zeger, S. L. (1994). *Analysis of Longitudinal Data*. Oxford: Clarendon Press.

Friede, T., Schmidli, H. (2010a). Blinded sample size reestimation with count data: methods and applications in multiple sclerosis. *Statistics in Medicine* 29:1145–1156.

Friede, T., Schmidli, H. (2010b). Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials. *Methods of Information in Medicine* 49:618–624.

Gajewski, B. J., Simon, S. D., Carlson, S. E. (2008). Predicting accrual in clinical trials with Bayesian posterior predictive distributions. *Statistics in Medicine* 27:2328–2340.

Johnson, N. L., Kotz, S. (1970). *Continuous Univariate Distributions*, Vol. 1. New York: Wiley.

Kappos, L, Radue, E. W., O'Connor, P., Polman, C., Hohlfeld, R., Calabresi, P., et al. (2010). A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *The New England Journal of Medicine* 362:387–401.

Keene, O. N., Jones, M. R. K., Lane, P. W., Anderson J. (2007). Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: Example from the TRISTAN study. *Pharmaceutical Statistics* 6:89–97.

McCullagh, P., Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed., New York, NY: Chapman and Hall.

Ng, H. K. T., Tang, M. L. (2005). Testing the equality of two Poisson means using the rate ratio. *Statistics in Medicine* 24:955–965.

Nicholas, R., Straube, S., Schmidli, H., Pfeiffer, S., Friede, T. (2012). Time-patterns of annualized relapse rates in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: a systematic review and meta-analysis. *Multiple Sclerosis Journal* 18:1290–1296.

Nicholas, R., Straube, S., Schmidli, H., Schneider, S., Friede, T. (2011). Trends in annualized relapse rates in relapsing–remitting multiple sclerosis and consequences for clinical trial design. *Multiple Sclerosis Journal* 17:1211–1217.

Thode, H. C. (1997). Power and sample size requirements for test of difference between two Poisson rates. *The Statistician*, 46:227–230.

Wang, Y. C., Meyerson, L., Tang, Y. Q., Qian, N. (2009). Statistical methods for the analysis of relapse data in MS clinical trials. *Journal of the Neurological Sciences* 285:206–211.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:439–447.