Journal of the Neurological Sciences 285 (2009) 206–211

Contents lists available at ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

Statistical methods for the analysis of relapse data in MS clinical trials

Y.C. Wang [a,*], L. Meyerson [a], Y.Q. Tang [a], N. Qian [b]

[a] Department of Biometrics, Biogen Idec, 14 Cambridge Center, Cambridge, MA 02142, USA
[b] Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA

ARTICLE INFO

Article history:
Received 5 March 2009
Received in revised form 15 July 2009
Accepted 22 July 2009
Available online 25 August 2009

Keywords:
Poisson regression
Negative binomial regression
Cox proportional hazards model
Over-dispersion
Count data
Log-link function

ABSTRACT

Patients with multiple sclerosis (MS) often experience unpredictable recurrent relapses with periods of remission. The modeling of MS relapse data is complicated because both within-subject serial dependence between relapses and between-patient heterogeneity may exist. We compare six statistical methods for assessing the treatment efficacy in reducing the frequency of relapses in MS clinical trials. All methods can be implemented in SAS®, and are grouped into two classes, one based on Poisson-type regressions for count data and the other on Cox proportional hazards models for time to relapse. We apply these models to the data of a Tysabri® (Natalizumab) MS trial and interpret the differences in results based on the underlying assumptions. Negative binomial regression is recommended for evaluating the overall treatment effect because of its simplicity and efficiency.

1. Introduction

Relapses are common recurrent events with serious consequences among patients with multiple sclerosis (MS). A relapse is defined as new or worsening neurological symptoms with duration greater than 24 h, preceded by a minimum of 30 days of clinical stability or improvement. Relapse-related outcome measures in MS clinical trials often include the *number of relapse-free patients* and *time to first relapse* [1]. However, both these measures are inefficient because the information following the first relapse is ignored, resulting in a misleading conclusion if the treatment fails to influence the first relapse but reduces the risk of subsequent relapses [2]. *Annualized relapse rate* (ARR) has been commonly used as a primary efficacy endpoint in phase III MS clinical trials [3] as it takes into account all relapses experienced during the period of observation.

A variety of statistical methods have been developed for analyzing recurrent events and can be used to assess whether a treatment can reduce the frequency of relapses in MS clinical trials. These include approaches based on counts of relapses and survival-time approaches based on time to relapse. It is unclear to many MS clinicians how these models differ from each other and which are the most appropriate for MS relapse data. This paper attempts to explain the differences in assumptions, parameter estimates and interpretations of these methods, and to provide guidelines for analyzing relapse data in MS clinical trials. We focus on several popular approaches implemented in the SAS® statistical package.

For approaches based on counts, the endpoint is the number of relapses that occurred in a given period of time. A count can take only non-negative integer values. Such data are non-normally distributed, and the variance varies with the mean. It might be inappropriate to analyze count data using ordinary linear regression because the linear model assumes homogeneity of variance and could produce non-sensible, negative predicted values [4]. Some researchers may rescale the counts to a dichotomy (e.g., "relapsed / did not relapse") or a set of ordered categories (e.g., 0, 1, 2, and ≥3), and analyze the data using the logistic regression model or the generalized Cochran–Mantel–Haenszel (CMH) test. Reduction of counts into categories wastes information and may dilute statistical power [4]. A simple model for analyzing count data is to assume that they are distributed according to a Poisson distribution. However, the Poisson distribution cannot account for the over-dispersion (i.e., the variance is larger than the mean) typically exhibited in MS relapse data.

Over-dispersion is the rule rather than the exception in practice [5] and can arise in several ways [6]. It can be a result of heterogeneity among patients, that is, each patient has a constant relapse rate, but some patients may be more prone than others to relapse, partly due to genetic and environmental differences or unmeasured covariates. Over-dispersion can also be a consequence of contagion, that is, occurrence of an early relapse increases a patient's risk for subsequent ones. Both heterogeneity and contagion mechanisms cause statistical correlation between relapses. Specifically, a patient with a history of relapse is likely to continue to experience more relapses, while subjects with no history of relapse tend to experience fewer future relapses. So the data contain more zero and large counts, and are more spread than that expected under the assumption of a Poisson distribution. The two mechanisms for over-dispersion are

* Corresponding author. Tel.: +1 617 914 1230; fax: +1 617 769 3280.
E-mail address: yong-cheng.wang@biogenidec.com (Y.C. Wang).

0022-510X/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jns.2009.07.017

indistinguishable given only the total counts. We will introduce quasi-Poisson and negative binomial (NB) models for over-dispersed count data. The NB distribution has been recommended for analysis of counts and rates in psychology [4], MS lesion data [7], and fall prevention trials [8].

While the aggregated count may be a good primary endpoint for assessing overall treatment efficacy, the use of a time-to-relapse endpoint allows the study of trend in relapse rates and treatment effects, as well as the possibility to distinguish heterogeneity and contagion [6]. In the Cox-based models, the serial dependence between successive relapses can be modeled by conditioning on the appropriate aspects of previous history of the patients. The dependence due to subject heterogeneity can be modeled via the subject-specific random-effects models or population-averaged (PA) proportional hazards models. The random-effects model introduces a random variable to explain dependence among relapses. The PA approaches estimate parameters by assuming independence between relapses, and estimate the standard errors (SE) using the robust "sandwich" method [9] to account for the dependence between relapses. While the random-effects model focuses more on how treatment affects individuals' relapse rates, the PA models evaluate how treatment affects the population-averaged relapse rates [10]. We will review several PA models since inference about population-averaged treatment effects is often of major interest.

The rest of this paper is organized as follows. Section 2 reviews several major statistical methodologies. Their differences will be illustrated through the analysis of relapse data of a Tysabri® (Natalizumab) MS trial in Section 3. Finally, we discuss our findings, and make a recommendation for future MS studies. The recommendation can be applied to any analysis of count data where over-dispersion is evident.

2. Statistical methods

This section will review two Poisson-type regressions for the count outcomes and four Cox proportional hazard models for the time-to-relapse endpoints. The quasi-Poisson and NB regressions are implemented in the SAS® procedure, GENMOD. All four Cox survival models can be fitted using the SAS® PHREG procedure.

2.1. Poisson regression for counts outcomes

Poisson regression is frequently used to model data in the form of counts. In MS trials, the outcome is the number of relapses (y) during a follow-up period (t), the time from study entry to the end of the study or loss to follow-up. The primary question of interest is whether treatment can reduce the ARR. Poisson regression requires the assumption that the successive relapses occur independently at a constant rate (λ) among all patients in each subgroup stratified according to treatment assignments and important prognostic factors [5]. Thus, response y follows a Poisson distribution with mean $\mu=\lambda t$, so that the expected number of relapses is proportional to the length of follow-up time. The logarithm of the relapse rate is assumed to be a linear function of the treatment ($x=1$ if a patient receives the new treatment and $x=0$ if the patient receives control) and other possible explanatory variables (z), that is,

$$\log(\lambda) = \alpha + \beta_1 x + \beta_2 z \text{ or equivalently } \log(\mu) = \log(t) + \alpha + \beta_1 x + \beta_2 z,$$

where α denotes intercept, and $\exp(\beta_1)$ is the ratio of ARR between treatment groups, while controlling for covariates z. The model implies that the ratio of ARR between treatment groups is common across the covariates.

A property of the Poisson distribution is that its variance equals its mean. However, relapse data often exhibit over-dispersion with a variance much greater than the mean, as discussed in Section 1. Ignoring over-dispersion still yields unbiased parameter estimates, but their precision and statistical significance will be overestimated [7].

A simple way to account for over-dispersion is to inflate the variance by a factor ϕ via the quasi-Poisson approach [5],

$$Var(y) = \phi E(y) = \phi\mu.$$

The dispersion parameter ϕ can be estimated by the Pearson χ^2 statistic divided by its degrees of freedom. The test statistic will be adjusted correspondingly. The correction for over-dispersion does not change the parameter estimate, and no probability distribution is specified in the quasi-Poisson approach (this will be discussed further in Section 2.2).

2.2. Negative binomial regression for the analysis of numbers of relapses

An alternative way to handle over-dispersion is to model it directly via negative binomial (NB) regression. Two features [2] of the NB distribution make it particularly suitable for MS relapse data: (1) it allows the count frequencies to be highly skewed and decrease monotonically from a modal value (i.e., the most frequently occurring value) of zero; and (2) its variance is always bigger than its mean.

The NB model can be derived as a Poisson–gamma mixture: if each patient relapses according to a constant Poisson rate, which varies in the subgroup according to a gamma distribution, the marginal distribution of the total counts in the subgroup follows a NB distribution [2]. Specifically, each patient is assumed to have their own relapse rate (λ_i) that varies around the average value (λ) in the subgroup, which can be modeled via a random variable (ε_i) with mean 1 and variance κ_i

$$\lambda_i = \varepsilon_i\lambda, \text{ and } \mu_i = \varepsilon_i\lambda t = \varepsilon_i\mu.$$

The variable ε_i reflects uncertainty about the individuals' rates. Subjects with $\varepsilon_i>1$ are more likely to relapse than subjects with $\varepsilon_i<1$. Under the above assumption, the mean of the total count is still $E[y]=\mu=\lambda t$, but the variance becomes

$$Var(y) = \mu + \kappa_i\mu^2,$$

where the first term on the right side denotes the Poisson variance, and the second term is the variance due to heterogeneity. If the ε_i's are independently identically distributed as gamma random variables, then y follows a NB distribution.

The NB distribution involves two parameters: the mean μ reflects the average rate while the dispersion parameter κ measures the degree of heterogeneity in the subgroup. A large κ indicates great variability in individuals' risk to relapse. If κ is small, the individuals' rates are more homogeneous and the process is approximately a Poisson. Indeed, the NB model becomes a Poisson model when $\kappa=0$.

In the NB model, the variance-to-mean ratio changes linearly with the mean. In the quasi-Poisson model, the variance-to-mean ratio is assumed to be constant, which cannot hold exactly in the presence of subject heterogeneity. If the variance is linear in the mean, then κ_i, which measures the degree of heterogeneity in individuals' rates, shall be proportional to the mean of the total count $\mu=\lambda t$, which depends on the length of follow-up time. That is, there is greater uncertainty about the risk of relapse in patients with smaller expected counts or shorter follow-up time, which sounds unreasonable.

2.3. Cox proportional hazard models for time to event outcomes

The Cox proportional hazard model takes into account time to relapse and does not assume a constant hazard rate (i.e., the instantaneous relapse rate at a specific time). Instead, it assumes that the ratio of risk for a relapse between two patients is constant over time [11].