Sample Size Estimation for Trials of Recurrent Events

Kuolung Hu, Amgen Inc., Thousand Oaks, CA

ABSTRACT

In some clinical trials, repeated occurrences of the same type of event are the study endpoint. For example, numbers of adverse event of interest occurred during the study period. There are several standard statistical methods for the analysis of such data, including the Anderson-Gill model (Anderson and Gill, 1982) and the Poisson regression model (Frome et. al. 1973). Signorini (1991) has proposed a method to estimate power for the Poisson model in the similar situation, but no approaches to power calculations are available for the Anderson-Gill model. We illustrate SAS macro to implement Signorini's method and extend it to incorporate unequal randomization between the treatment and control groups. Furthermore, using simulations, we show that the power for both the Poisson model and the Anderson-Gill model are similar under a variety of scenarios, so that this approach to sample size calculations can be used for either method of analysis in a study.

Keywords: Poisson regression model, recurrent event data, Anderson-Gill Model, sample size estimation.

INTRODUCTION

Recurrent event data is where the subject experiences repeated occurrences of the same type of event or events of different types given a follow-up period. The sequence of times of the event constitutes a simple point process observed continuously over the study period. Many clinical trials are designed to demonstrate the treatment efficacy or safety based on the frequency of the outcome over time, defined as the event rate. Although the exposure period may be different from subject to subject, the frequency of event is usually summarized as a rate or number per year of exposure, or per 100 subject years. Examples include trials detecting the reduction in relapses of neurological symptoms, occurrence of blood transfusions, or number of adverse event incidence.

Recurrent events can be analyzed by many statistical methods. The traditional tests are based on either frequency counts or time to the first event. Current methodologies attempt to capture both types of information. Summaries of several approaches and their methodologies to analysis such data can be found in Lachin (2000) and Therneau and Grambsch (2000). In this paper, we will describe two of the widely utilized methods: Anderson-Gill (A-G) model and the Poisson regression model; for the recurrent event analysis. A-G model describes a generalization of the proportional hazard model and the Poisson regression is a member of a class of generalized linear models (McCullagh and Nelder, 1989). Similarity in the inference results from A-G and Poisson model usually can be found (Therneau and Grambsch, 2000). Both methods could be implemented in SAS®. Many examples could be referred.

Clinical trials should have sufficient statistical power to support the study objectives. The size of the trial should be considered early in the planning stage. It is the study statistician’s responsibility to provide the sample size estimate (Friedman et. al. 1998) to support the trial design. In general, as long as the calculated sample size is realistically obtainable, it is better to overestimate the size than to underestimate it. Signorini (1991) proposed the technique using asymptotic variance of the maximum likelihood estimate to calculate the sample size for the Poisson regression model. Shieh (2001) summarized the sample size calculation methods for logistic and Poisson regression method.

In this paper, we will review the power estimation method for Poisson regression proposed by Signorini (1991). A SAS macro based on this method to calculate the sample size will be presented in the clinical trial setting. Furthermore, we will explore the similarity in term of power between these A-G and Poisson regression models based on the simulations under certain circumstances. The simulation methods for the recurrent event process proposed by Metcalfe and Thompson (2006) will be utilized.

METHODS

ANDERSON-GILL MODEL

The Anderson-Gill (A-G) model, an extension of the Cox proportional hazards model as a counting process, is a regression analysis of the intensity of the recurrent event. An individual hazard function at time t is
\[ h_i(t \mid x_i) = h_0(t) \exp \{ \beta x_i \} \]

where \( h_0(t) \) is the hazard over time in individuals on placebo \((x=0)\) and \( \beta \) is the log hazard ratio comparing treatment groups for the first event.

The principal assumption of the A–G approach is that the hazard or risk ratio is proportional over time. The hazard ratio represents the proportionate change in the mortality rate due to a unit change in the respective covariate. Proportional hazards are maintained if the influence of some treatment or other independent variable remains consistent across the duration of the study \( \text{i.e., } \beta(t) = \beta \text{ for all } t \). The assumption can be assessed using plots of the logarithms of the estimated cumulative hazard functions for different treatment groups \( \text{(Andersen 1982)} \). Convergence of curves indicates that the hazard function has a different effect on each group across time.

Data for the A–G model are structured so that each subject is treated as 1 to many observations. The interval of risk for each observation is defined by variables describing the start and end times of successive periods. An event variable is coded as “1” for mortality or “0” for right-censored intervals. Right-censored intervals are not considered as incomplete data, but as subjects whose event counts are still 0. Here are the example of the data from the cancer bladder data from Wei, Lin, and Weissfeld \( \text{(1989)} \) reference.

<table>
<thead>
<tr>
<th>ID</th>
<th>TSTART</th>
<th>TSTOP</th>
<th>STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>16</td>
<td>18</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>23</td>
<td>0</td>
</tr>
</tbody>
</table>

The time of previous event start time \( (t_{\text{start}}) \) and the current event start time \( (t_{\text{stop}}) \) construct each interval of the risk. The following SAS statements use PROC PHREG to fit the A-G model to the data with treatment \( \text{(trt)} \), number of initial tumor \( \text{(number)} \) and initial tumor size \( \text{(size)} \).

```sas
proc phreg data=bladder covs(aggregate) covm;
  model (tstart, tstop) * status(0) = trt number size;
run;
```

PROC PHREG can compute both the regression estimates \( \beta \) and their robust covariance estimates by maximum likelihood estimates. Please refer to Johnston and So \( \text{(2003)} \) and SAS/STAT\textsuperscript{®} PHREG procedures for description and examples of this procedure.

**POISSON REGRESSION**

The Poisson regression model is straightforward to analyze this type of data and had been illustrated in many references \( \text{(Zou, 2004 and Hujoel, et. al. 1994)} \). Poisson regression assumes that the dependent variable follows a Poisson distribution, a distribution that we frequently encounter when we are counting a number of events. It can be used to model the number of occurrences of an event of interest or the rate of occurrence of an event of interest, as a function of some independent variables. Using a log transformation which adjusts for the skewness and prevents the model from producing negative predicted values, the generalized linear model is:

\[ \log(E(Y_i)) = \log T_i + \beta x_i \]

where \( \beta \) is a vector of regression coefficients, \( x_i \) is a vector of covariates for subject \( i \), an offset variable \( \log T_i \) is needed to account for possible different observation period for different subjects.

Cochran \( \text{(1954)} \) suggested that the homogeneous Poisson process assumption be tested using a simple chi-square goodness of fit test. Under the assumption of a common intensity for all subjects, then the mean number of events in the population \( E(Y_i) = V(Y_i) = \lambda T_i \) and the test is provided by:
\[ \chi^2 = \sum_{i=1}^{N} \frac{(Y_i - \hat{\lambda}T_i)^2}{\hat{\lambda}T_i} \]

which is distributed as chi-square on \( N-1 \) df under the hypothesis of homogeneity.

Pedan (2001) provides an example from a clinical trial analyzed by Poisson regression

```sas
data Hypo_Cramp;
  input id devices $ count fu_time;
  logt=log(fu_time);
  datalines;
  1 control 2 2.5
  2 control 0 3
  ..................
  68 test 1 3
  69 test 3 3
  70 test 0 2.7
  ;
run;
```

The variable `device` represents treatment assignment, the variable `count` contains the number of hypotensions and crampings for each subject during the follow up period. The variable `fu_time` represents the follow up time for each subject measured in months. The `logt` is the log transformation of `fu_time`.

The PROC GENMOD of SAS can fit a wide range of generalized linear models. The following SAS statements use PROC GENMOD to fit the Poisson regression

```sas
proc genmod data=hypo_cramp;
  class device;
  model count=device/offset=logt dist=poisson link=log;
run;
```

For further details about PROC GENMOD, please refer to SAS/STAT® GENMOD procedure.

### POWER ESTIMATION FOR POISSON REGRESSION

Signorini (1991) proposed a method for calculating sample size for Poisson regression model with various distributions of a single covariate. The method to determining the sample size is based on the resulting asymptotic variance of the maximum likelihood estimator of the parameters. The reference also presents the method for a family of multivariate exponential-type distributions for multiple covariates. There are two examples illustrated in the article and we will utilize one of the examples in this paper.

Suppose there are \( N \) subjects, each with the exposure time \( t_i (i=1, \ldots, N) \). Considering \( \beta \) as the parameter of interest, we wish to test the null hypothesis \( H_0: \beta=(0, \beta_2, \ldots, \beta_p) \) against the alternative hypothesis \( H_1: \beta=(\beta_1, \beta_2, \ldots, \beta_p) \) at a significance level of \( \alpha \) and with power at least 1-\( \gamma \). For a single covariate \( X \), for simplified purpose, the least \( N \) needed can be calculated by

\[
N \geq \frac{z_{\alpha} \{ \text{var}(\lambda) \}^{\frac{1}{2}} + z_{\gamma} \sqrt{V^2(\hat{\beta})}}{u_T e^{p \hat{\beta}^2}}
\]

where \( u_T \) is the mean exposure time. The derivation of the above formula can be referred in the article for more details. In this paper, we will focus on the Bernoulli parameter (treated vs. non-treated) for covariate \( X \) due to the commonly use in the clinical trial setting. Therefore, the \( \text{var}(X) = \frac{1}{\pi} + \frac{1}{1-\pi} \) where \( \pi \) is the allocation probability for a subject assigned to the active treatment arm in a trial. The formula for \( V(\beta) \) estimation for the Bernoulli with parameter \( \pi \) is as follow:
The formula for other common distributions to calculate $V(\beta)$ can be found in the article.

Usually it is common to set as $\pi=0.5$, meaning equal allocation for enrollment. Sometimes, the rational for a 2:1 or 3:1 allocation is that the study may gain more information about participant responses to the new intervention, such as toxicity and side effects. Additionally, if the intervention turns out to be beneficial, more study subjects would benefit than under an unequal allocation design.

**RECURRENT EVENT SIMULATION**

Power estimation can be evaluated using simulated data sets. Simulation is still the preferred method if the study statistician would like to confirm the accuracy of these approximations in the given study assumptions. We will utilize the simulation process for recurrent events proposed by Metcalfe and Thompson (2006). The reference describes the simulation of recurrent events data based on four different types of generation process: Poisson, mixed Poisson, autoregressive and Weibull. In this paper, we will adopt the Poisson process to generate the simulated data and will not consider drop-out in order to simplify the illustration. A-G model and Poisson regression model both perform well for the simulated datasets based on Poisson process.

Each simulation is of $n$ individuals. For each subject, the follow-up is a maximum of 1 unit of time, or 12 events. A limit on the number of events simplifies simulation and is rarely reached for a subject. Therefore, 12 inter-event times were simulated as independent realization of an exponential distribution of rate $\lambda=\exp(x\beta)$, obtained from the inversion method

$$t_{ik} = -\frac{\ln u_i}{\exp(x_i\beta)} = -\frac{\ln u_i}{\lambda}$$

where $i=1, ..., n$ individuals; $k=0, ..., k_i$ events observed from a individual; $t_{ik}$ denoted as the interval time from $T_{i,k-1}$ until $T_{i,k}$ and $u_i \sim U[0,1]$, a uniform distribution. All simulations are based on random realization from a uniform distribution. By transforming $u_i$, we can generate interval-event time $t_{ik}$ for each subject until the maximum follow-up reached or $k$, total number of events, is above 12.

For each simulation, we will obtain the parameters estimates and their test statistics. The power of the specific model will be determined as the total number of time that observing the p-value for the treatment effect is less than 0.05 at two sided divided by the total simulation time.

**RESULTS**

**MACRO FOR POWER ESTIMATION**

The %POISSONSAMPLE macro for SAS is found in the appendix. The macro can produce a listing of total sample size estimates under 80%, 85%, 90% and 95% of the power. It also provides the options for selecting one-sided or two-sided test and the allocation ratio. To use the macro, we can simply enter all the corresponding parameters properly:

```sas
%poissonsampel(R0,RR,TT,Alpha,Side,PO);
```

where $R_0=\exp(\beta_0)$, $RR=\exp(\beta_1)/\exp(\beta_0)$, $TT$ is the study follow-up time, $Alpha$ is alpha level of the significance level, $Side$ is one-sided or two sided test, and $PO$ is proportion of subjects enrolled into compared treatment arm.

For validation, we will replicate the sample size calculation for the study presented in Signorini reference in order to validate our macro. It is a study of water pollution in terms of number of illnesses and infections contracted per swimming season for ocean swimmers versus non-ocean or infrequent swimmers around Sydney, Australia. Applying the simple Poisson regression for the number of infections, with the covariate indicating ocean swimmer $(X=1)$ or not and the sample size are balanced (ie. 1:1). In this case, the estimated rate of non-ocean or infrequent swimmers is 0.85 and it is required to detect an increase of at least 30% in the infection rate for ocean swimmer $(\exp(\beta_1)=1.3)$. The unit of time will be one single swimming season. We will submit the following the following statement:

$$\left(\pi e^{\beta} \right)^{-1} + (1 - \pi)^{-1}$$
and the outputs are:

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80</td>
<td>406</td>
</tr>
<tr>
<td>0.85</td>
<td>470</td>
</tr>
<tr>
<td>0.90</td>
<td>556</td>
</tr>
<tr>
<td>0.95</td>
<td>697</td>
</tr>
</tbody>
</table>

For a one-sided, significance level of 5%, the required sample size to detect a 30% or greater increase is 406 at 80% power, 556 at 90% power or 697 for 95% power from the reference. The results from the macro do match with estimates in the article.

In addition to estimating the total sample size with 1:1 enrollment ratio (non-ocean swimmers: ocean swimmers), this macro will also handle the sample size estimation for the trial with unbalanced allocation. By increasing the allocation ratio of ocean swimmers, for example, the sample size does increase and by approximate 14%, 36% and 60%, compared to the balanced design (see table 1). Also, by adapting this strategy, we will collect more information from the ocean swimmer in order to properly estimate the infection rate of the ocean swimmer. In the other way, similar trend is observed if the study team decided to increase the enrollment of non-ocean swimmer or non-frequent swimmer.

<table>
<thead>
<tr>
<th>Power</th>
<th>1:1</th>
<th>1:2</th>
<th>1:3</th>
<th>1:4</th>
</tr>
</thead>
<tbody>
<tr>
<td>80%</td>
<td>406</td>
<td>463</td>
<td>553</td>
<td>651</td>
</tr>
<tr>
<td>85%</td>
<td>470</td>
<td>536</td>
<td>641</td>
<td>755</td>
</tr>
<tr>
<td>90%</td>
<td>556</td>
<td>636</td>
<td>761</td>
<td>897</td>
</tr>
</tbody>
</table>

Table 1. Sample size estimation from Signorini’s example and different allocation ratios

If the study is considering an unbalance design, the estimate based on different allocation ratios are useful for study operation and enrollment feasibility considerations. However, we should also point out that unbalanced allocation appears to degrade the accuracy of sample size calculation (Shieh, 2001).

SIMULATION RESULTS

Furthermore, we utilized the simulation method described above to explore the similarity in term of power between the A-G and Poisson regression models. By simulating 1000 replications with, say, n=280 subjects for each group, we will obtain the probability mass function for the number of infections observed in each treatment group with censoring at 1 year. Table 2 presents the results across these 1000 replications.

<table>
<thead>
<tr>
<th>Number of infection</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>≥4</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-ocean swimmers</td>
<td>0.43</td>
<td>0.36</td>
<td>0.15</td>
<td>0.04</td>
<td>0.02</td>
</tr>
<tr>
<td>ocean swimmers</td>
<td>0.33</td>
<td>0.37</td>
<td>0.2</td>
<td>0.07</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 2. Probability mass function for number of infections observed for both groups across 1000 replications based on Signorini’s example

and the mean cumulative frequencies (MCFs) plot for non-ocean swimmer and ocean swimmer shown in figure 1. We plot this MCFs with one of the 1000 replications to show the separation of the two groups during the observation period. The SAS statement for creating MCFs plot can be found in Johnston and So (2003).
The summary of parameter estimates for the Poisson and A-G model from the 1000 replicates is shown in table 3.

<table>
<thead>
<tr>
<th>Statistical Method</th>
<th>Parameters</th>
<th>Mean of the Estimates</th>
<th>SD of the Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poisson</td>
<td>$\beta_0$</td>
<td>0.095</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>$\beta_1$</td>
<td>-0.257</td>
<td>0.08</td>
</tr>
<tr>
<td>A-G</td>
<td>$\beta$</td>
<td>-0.257</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Table 3. Parameter Estimates for Poisson Regression and A-G model cross 1000 replications based on based on Signorini’s example

From table 3, we can obtain the estimated event rate for non-ocean swimmer is $\exp(0.095-0.257)=0.86$ and for ocean swimmers is $\exp(0.095)=1.1$ from the Poisson regression. The A-G model will estimate the hazard ratio $\exp(-0.257)=0.77$, which is the event rate of non-ocean swimmer vs. ocean swimmer and is expected to be less than 1. Therefore, we observe similar parameter estimate from these two models based on the simulations.

Furthermore, we also observed that in about 85% of replicates was rejected the null hypothesis of the treatment effect at two sided with alpha level=0.05. Both models reach the same conclusion under each of replicates. In addition, if we apply the macro to estimate the sample sizes needs 518 subjects at 80%, 589 at 85%, 685 at 90% and 841 at 95%. Thus, the macro is over–estimating the sample size needed (560 vs. 589) for this trial, compared to the simulation results, but is within a reasonable range.

CONCLUSION

We had discussed the methods determine the sample size for a study with recurrent event. A SAS macro was developed and illustrated based on commonly used in the clinical trial design. With this macro, the study statistician could easily provide the sample size estimate under different power scenarios for the study team to discuss. This macro also provides the option with different allocation ratio selection to be more flexible in study design. In addition, we present the simulation method to further assess the accuracy of the estimate made from this macro. If it is feasible, a larger sample size could provide the more accurate information regarding to efficacy and safety. Other generation processes for the simulation may be considered in order to support our findings toward the similarity between A-G model and Poisson regression in power estimation.

APPENDIX

The following SAS codes to calculate the sample size method for Poisson regression proposed by Signorini.

/* Program : Sample Size evaluation for Poisson Regression Model
R0 = baseline rate, Exp(Bo);
RR = Rate Ratio, denoted as exp(B1)/exp(BO)
tt = exposure length
Alpha = alpha level
side = 1 or 2 side 
P0 = Subject allocation factor for the compared (B1), the range of value is (0, 1) */;

%macro poisson(ro, rr, tt, alpha, side, po);

data samplesize;
ro=&ro;
rr=&rr;
tt=&tt;
alpha=&alpha;
side=&side;
po=&po;
do power=.8, .85, .9, .95;
varbo=(1/po)+(1/(1-po));
varb1= (1/(1-po))+(1/(po*rr));
alphal=1-(alpha/side);
denon=tt*ro*(log(rr)**2);
numer=(quantile('NORMAL',alpha1)*sqrt(varbo)+quantile('NORMAL',power)*sqrt(varb1))**2;
totaln=ceil(numer/denon);
output;
label po='Enrolled Ratio' totaln='Total N';
;
end;
title "Rate= &ro., Odds=&rr., Time Interval=&tt., &side. - sided, Proportion=&po." ;
proc print data=samplesize label noobs double;
var power totaln;
run;
%mend;

REFERENCES


ACKNOWLEDGMENTS

The author is grateful to Bob Parker of Amgen Inc. for comments during the preparation of this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Kuolung Hu
Amgen Inc.
One Amgen Center Drive
Thousand Oaks, CA 91320-1799
Work Phone: 805-313-4409
Fax: 805-447-3365
E-mail:kuolungh@amgen.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.