生物統計学/疫学・予防保健学 抄読会　2013/09/11

SPH2 年　鄭　日川

# Within-trial prediction for the number of future events

## Introduction

In randomized clinical trials, mid-course monitoring of sample size is becoming increasingly commonplace and important. For the studies with a failure time outcome, sample size estimates depend mainly on the proposed effect of treatment on the difference in failure rates, usually measured by the hazard ration, but sample size is also dependent on the number of events that will be observed during follow-up which depends on the overall failure rate. A departure from the estimated failure rate at the design stage may reduce the power unacceptably for a clinically important hazard ration. This might be resolved by making appropriate design modifications (usually increasing the follow-up time until the enough events are observed). To inform the decision about whether and how to make mid-course correction, predicting the number of events in the future course of the trial is desirable.

Prediction has long been recognized as an important practical aspect of interim monitoring of randomized clinical trials. It is nature and potentially more accurate to make the prediction by using the accumulating data from the trial itself, rather than the results from prior trials with similar design. In addition to the point prediction, for many applications, it is important to provide lower and upper bounds of the prediction (prediction interval) at some specified level of confidence to reflect the uncertainty of prediction.

In this paper, I will present 2 separate approaches to predict the number of events for the future course of the trial given the accumulating observed data. Either method will estimate prediction intervals by a straightforward simulation. 2 methods will be demonstrated with data from a clinical trial of REAL-CAD.

## REAL-CAD trial context

REAL-CAD trial is an ongoing randomized, open-labeled, investigator-initiated, large scale clinical trial, of stable CAD (coronary artery disease) patients with hypercholesterolemia, comparing the intensive (Pitavastatin, 4 mg/day) and moderate (Pitavastatin, 2 mg/day) Statin therapy on cardio-cerebrovascular outcomes. The primary endpoint is defined to be a composite of cardiovascular death, nonfatal myocardial infarction, nonfatal stroke and hospitalization for unstable angina. This

trial intended to randomly assign 12600 patients in a 1:1 ratio, as it eventually randomized 13054 patients from January 2010 to July 2013. All patients are to be followed up until March 2016 to obtain 1033 primary events, which would give the study a statistical power of 80% at 2-tailed $\alpha$ level of 0.05, assuming a constant overall event rate of 0.025 person-year and an hazard ratio (high/low) throughout follow-up of 0.84.

At the time accrual was complete (August 1, 2013), 193 primary events occurred. From this data, we found that the overall event rate was approximately 0.009 person-year, rather than 0.025 as was proposed. Because the planned sample size is already quite large and accrual of patients was closed, increasing the sample size is not realistic. So we would have to rely on extending follow-up to achieve significance, given this much lower event rate than expected would last until the end of follow up period. But we also have chance to prove our hypothesis without changing our plan, if overall hazard rate is not constant but getting higher with time, and then 1033 required events are still possible to reach within the planned follow up period. Or since the assumption of constant overall event rate no longer hold and hence, the estimated sample size before the study commenced might not so accurate, a smaller sample size instead of 1033 events may also enable us to detect the treatment difference.

*(A paragraph to be inserted: summarize the review on event rates in the light of prior trials to explain possible factors accounting for lower event rates)*

With the context described above, to aid decisions to modify study designs or not, firstly we have to be focused on predictions regarding results to be observed in the future.

## Methods

The objective is to predict at the future time T* (in this paper, 12, 24 and 32 months later. But it can be user-defined.) how many events will occur, that is, how many patients who are alive at current time t0 (August 1, 2013) are expected to fail between t0 and T*, given the current data.

Following this objective, .a method like Method A using the survival outcome data only, which place the assumption that the survival function is under a parametric survival time distribution and prior trial survival data are relevant with distribution assumption, can be utilized. Method A can make the prediction simpler, more convenient and potentially efficient. But there is concern that the predictions may be less reliable if relations between the survival outcome and the other variables (covariates) are not extrapolated. To address this concern, the second method – Method

B, which is incorporating several covariates into the survival model to predict the point hazard rate, is also proposed. Convergence of the two methods will add confidence to the predictions.

**Method A: Bayesian Weibull prediction based on prior survival knowledge combined with current survival data**

Unlike the exponential, the two-parameter Weibull distribution describes the hazard function increased or decreased over time. Thus, the Weibull, which includes the exponential as a special case, can offer a promising compromise position between the exponential and nonparametric approaches.

Assume the event times in the two arms follow 2 independent Weibull distribution with the scale parameters of $\lambda_1$ and $\lambda_2$ and the identical shape parameter of $\alpha$. Then the likelihood function in terms of $\alpha$ and $\lambda$ is as follows:

$$\begin{aligned} L(\alpha,\lambda|t) &= \Pi_{i=1}^{n} p(t_i|\alpha,\lambda_i)^{v_i} S(t_i|\alpha,\lambda_i)^{1-v_i} \\ &= \Pi_{i=1}^{n} (\alpha t_i^{\alpha-1}\exp(\lambda_i - \exp(\lambda_i)t_i^{\alpha}))^{v_i} (\exp(-\exp(\lambda_i)t_i^{\alpha}))^{1-v_i} \\ &= \Pi_{i=1}^{n} (\alpha t_i^{\alpha-1}\exp(\lambda_i))^{v_i} (\exp(-\exp(\lambda_i)t_i^{\alpha})) \end{aligned}$$

where n=2 and v=1/$\lambda$. If the arm covariate is linked to $\lambda$ with $\lambda_i = x_i' \beta$ where $x_i$ is the vector of treatment arm corresponding to the $i$th observation and $\beta$ is a vector of regression coefficients, the log-likelihood function becomes:

$$l(\alpha,\beta|t,x) = \sum_{i=1}^{n} v_i(\log(\alpha) + (\alpha-1)\log(t_i) + x_i'\beta) - \exp(x_i'\beta)t_i^{\alpha})$$

Taking advantage of prior survival information, a multivariate normal with mean equal to numbers proposed at the design stage are used for intercept and $\beta$ as our prior. Meanwhile, to explore the sensitivity of predictions to the prior, non-informative diffuse priors are also applied.

Informative moderate priors (time unit of month):

intercept ~ normal(mean= -6.09, sd = 0.60); corresponding to an event rate of 0.027 person-year in reference group ranged from about 0.008 to 0.090.

$\beta$ ~normal(mean=-0.1743, sd = 0.125); corresponding to an HR of 0.84 ranged from about 0.65 to 1.08.

$\alpha$ ~gamma(shape =4, scale =0.25); corresponding to a shape parameter of 1 (exponential survival) ranged from about 0 to 2.

Non-informative diffuse priors:

intercept ~ normal(mean= 0, var=10000);

$\beta$ ~normal(mean= 0, var=10000);

$\alpha$ ~gamma(shape = 0.001, inverse scale = 0.001);

Once the priors are specified, we can resort MCMC (Markov Chain Monte Carlo) to get a list of posterior sampled values of (intercept, $\beta$, $\alpha$). Specifically, to simulate the values of the numbers of events at T*, the following steps are executed:

(i) Take the next element from the list of sampled values of (intercept, $\beta$, $\alpha$). The scale parameter $\lambda_1$ for reference group can be estimated as exp(intercept) and then for non-reference group, scale parameter $\lambda_2$ can be estimated as exp(intercept+ $\beta$).

(ii) Conditional on the time spent so far and treatment arm, impute event times for currently censored subjects given the sampled Weibull parameters.

(iii) Calculate the failure subjects whose expected event time exceeds T*, by group.

By repeating steps (i)–(iii) 1000 times, we obtain a list of values representing a set of 1000 draws from the predictive distribution of number of events at T*, while the 5 and 95 quartiles of this list give the limits of a 90% prediction interval for the predictive number of events.

**Method B: Assuming a model incorporating several covariates to extrapolate beyond survival data** ***(Development is Ongoing.)***

## Results from a semi-simulated dataset

Methods are evaluated on a semi-simulated dataset based upon the actual accrual by month and the numbers of events up to 31 December, 2012 (140 events) and August 1, 2013 (193 events). Group data is randomly generated from the Bernoulli distribution with p=1/2 and survival times are simulated from 2 exponential distributions for 2 discrete time grids (~31 December, 2012 and January 1, 2013~ August 1, 2013), provided a constant hazard ratio (high/low) of 0.84.

**Method A:**

**Table 1 Posterior summaries using informative moderate priors**

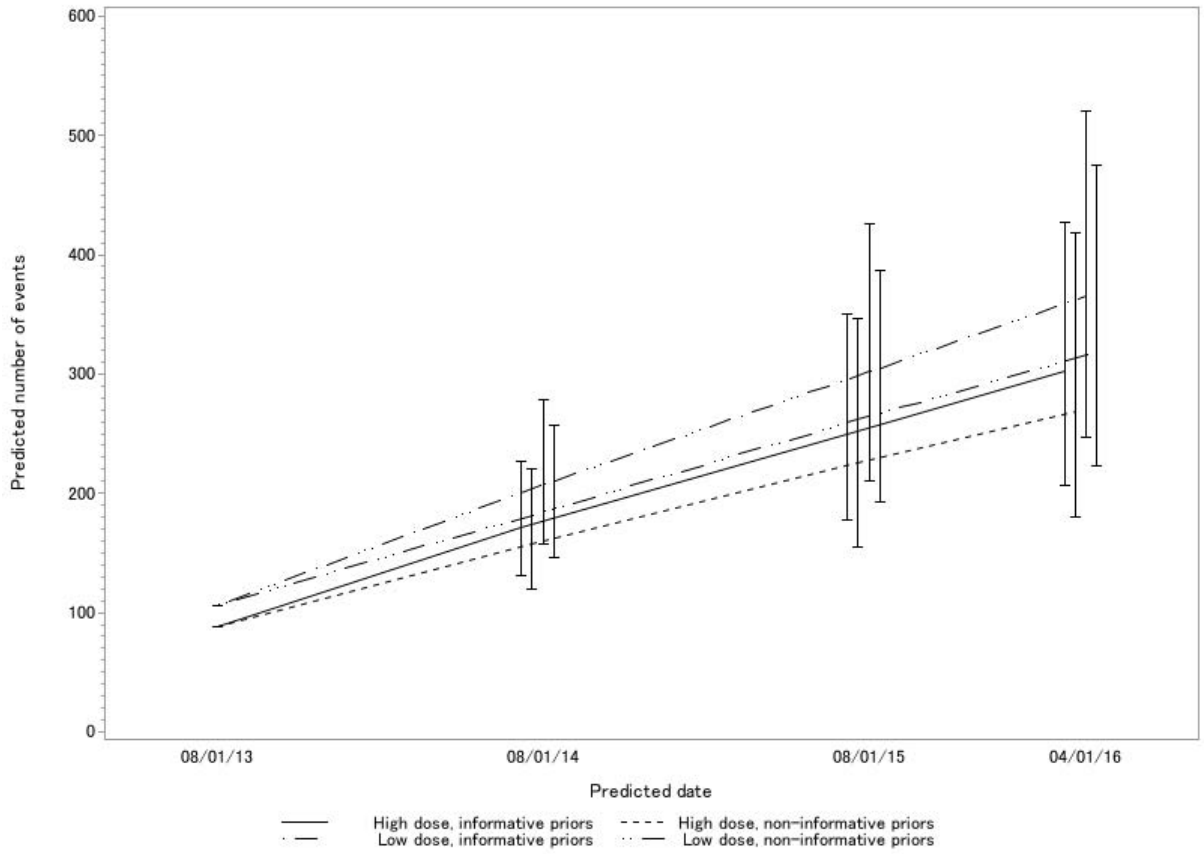
| | | | Standard | 95% HPD | |
|---|---|---|---|---|---|
| **Parameter** | **N** | **Mean** | **Deviation** | **Lower** | **Upper** |
| **alpha** | 1000 | 0.9297 | 0.0618 | 0.8322 | 1.0494 |

| | | | | | |
|---|---|---|---|---|---|
| **beta0** | 1000 | -6.849 | 0.213 | -7.2531 | -6.5016 |
| **beta1** | 1000 | -0.1814 | 0.0977 | -0.3555 | -0.00345 |

Table 2 Posterior summaries using non-informative diffuse priors

| | | | Standard | 95% HPD | |
|---|---|---|---|---|---|
| **Parameter** | **N** | **Mean** | **Deviation** | **Lower** | **Upper** |
| **alpha** | 1000 | 0.9592 | 0.0645 | 0.8312 | 1.087 |
| **beta0** | 1000 | -6.9625 | 0.2065 | -7.391 | -6.5745 |
| **beta1** | 1000 | -0.1711 | 0.1377 | -0.4065 | 0.1042 |

Figure 1 Predictions of the number of events




The vertical lines are 90% prediction intervals, and medians are connected by a line. Change of power (log-rank test, two sided α=0.05): informative: 33%→49%→56%; non-informative: 30%→46%→52%.

【参考文献】


1） Weibull WA. A statistical distribution of wide applicability. Journal of Applied Mechanics 1951; 18:293–297.

2） Ying GS, Heitjan DF. Weibull prediction of event times in clinical trials. Pharmaceutical Statistics 2008 Apr-Jun; 7(2):107-20.

3） Bagiella E, Heitjan DF. Predicting analysis times in randomized clinical trials. Statistics in Medicine. 2001 Jul 30; 20(14):2055-63.

4） Cook TD. Methods for mid-course corrections in clinical trials with survival outcomes. Statistics in Medicine. 2003 Nov 30; 22(22):3431-47.

5） Ying GS, Heitjan DF, Chen TT. Nonparametric prediction of event times in randomized clinical trials. Clinical Trials. 2004;1(4):352-61.

6） SAS/STAT(R) 9.2 User's Guide, Second Edition
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#titlepage.htm