Timeline Prediction for Ongoing Major Cardiovascular Event Trials

Hui Quan*, Xuezhou Mao, Yujun Wu, Meehyung Cho, Peng-Liang Zhao,
Ji Zhang and Deborah Bauer

Biostatistics and Programming, Sanofi, 55 Corporate Drive, Bridgewater, NJ 08807

Abstract

Cardiovascular (CV) event trials are long term, large scale and costly trials. Since the study power for an event-driven trial depends on the total number of events, it is important to use all available information including blinded information acquired during the study to continuously predict the timeline of reaching the target number of events at different study stages. The sponsor thus can timely optimize resources and modify the new drug development strategy. In this paper, we consider the methodology for such a timeline prediction. The background hazard rates of different time intervals used for the prediction are based on the observed internal blinded rates and the external published rates. We first derive the conditional probability of having an event by a specific time point or the common end of the study given the observed information of a patient. We then calculate the total number of events as the sum of the conditional probabilities across all patients. These conditional probabilities are increasing functions of the overall study duration. Thus, we can predict the timeline to reach the target number of events. We use a CV trial example to illustrate the application of the methodology.

Keywords: hazard rate, conditional probability, event time, study duration, intent-to-treat.

*hui.quan@sanofi.com

1. Introduction

Cardiovascular (CV) event trials are often conducted to assess the CV efficacy or safety of experimental drugs particularly since the Food and Drug Administration (FDA) issued the guidance [1] in 2008 for evaluating CV risk with the use of new therapies for treating type 2 diabetes mellitus (T2DM). The FDA guidance recommends that before the submission of an anti-diabetic New Drug Application, the sponsor should demonstrate the Non-Inferiority (NI) of the drug compared to control(s) on CV risk using a 1.8 margin for the relative risk. Unless the pre-marketing data have already shown the NI based on a 1.3 margin with a sufficient number of events, a post-marketing CV safety trial should be conducted to show the 1.3 NI achieved by either data from the single CV safety trial that is adequately powered or by the combination of the pre-marketing trials and the post-marketing safety trial. As a result, a sponsor developing a T2DM drug has to conduct a CV trial if the Phase III data alone cannot demonstrate non-inferiority with the 1.3 margin.

CV event trials are long term, large scale and costly trials. Since the study power for an event-driven trial depends on the total number of events, it is important to use all available information including blinded internal information acquired during the study to continuously predict the timeline to reach the target number of events. The sponsor thus can timely optimize resources and modify the new drug development strategy. Obviously, many factors can impact the timeline for reaching the target number of events. One of them is the background CV event rate or hazard rate that could vary substantially from less than 1.0% per patient-year to more than 10% per patient-year [2, 3] depending on the patient populations. Even within the same group of patients, the event rate could also change dramatically overtime. For example, the event rate in T2DM patients with acute

coronary syndrome (ACS) that includes acute myocardial infarction or unstable angina requiring hospitalization within the previous 15 to 90 days could be reduced from around 10% per patient-year during the first several months to around 3.5% 2 years later [2]. Another factor is the enrollment curve – a quicker enrollment will increase the total exposure for the study given the same total study duration. Sample size for the study is obviously another critical factor. Given these key factors and additional other observed information during the trial, the expected total number of events is an increasing function of the total study duration and the timeline for reaching the target number of events can be predicted.

Like other experimental drugs (e.g., lipid lowering drugs), an anti-diabetes drug could have a CV protective effect. Moreover, due to the difficulty of recruiting patients for a non-inferiority CV trial, the sponsor could design a CV trial with the objective of demonstrating potential superiority or just demonstrating non-inferiority if newly available information including external or internal interim information shows the chance for the superiority is slim. The required numbers of events for the desired power for these two objectives are very different. For example, with 20% true overall risk reduction, 90% power and one-sided 2.5% significance level, the required number of events is 844 for showing the superiority and is only 179 for demonstrating non-inferiority with a 1.3 margin for relative risk. On the other hand, if there is truly no treatment effect, no matter how large the number of events is, there will be at most a 2.5% chance (the Type I error rate) to show the superiority and the required number of events is 612 to have 90% power to show non-inferiority. Within the same study, we can only assume one true treatment effect. Then the required number of events for the superiority assessment is always much larger than that of the non-inferiority assessment.

Recently, a CV trial was designed for an experimental diabetes drug using high risk T2DM patients with recent ACS. The objective of the trial was to demonstrate the superiority or non-inferiority of the experimental drug compared to placebo on CV risk. An independent Date Monitoring Committee was instituted to closely monitor data during the trial. Unless there was a serious safety issue, there was no plan to stop the trial early. At the design stage, the background hazard rates were assumed to be 10% per patient-year during the first year and 7% per patient-year after the first year. A linear enrollment curve of 37 months to enroll a total of 6000 patients was also assumed. To have 90% power based on an Intent-To-Treat (ITT) analysis to detect an overall 20% risk reduction at 2.5% significance level (one-sided), the required number of events was 844. In case it failed to demonstrate the superiority, this number of events would also ensure sufficient power for establishing non-inferiority with a 1.3 margin if the true treatment effect is null. To reach this target of 844 events, the total study duration (from the first patient randomization to the end of the study) was anticipated to be 47 months. After the number of events reached 353, the sponsor's management wanted to make a more precise prediction of the timeline for reaching 844 events using all updated information including information from published articles of CV trials with a similar population as well as internal interim blinded information such as the real enrollment times for enrolled patients, the observed event times for those who had had events and the observed event-free times for those who had not yet had events. The purpose of this paper is to provide the methodology which could also be applied to other future CV trials.

2. Estimate of event rate

Because of the large difference in CV event rate across different patient populations, when a CV trial is designed, the first step is to determine the patient population for the

trial. Besides the target patient population for the indication, the use of a high risk patient population associated with a higher event rate requires a smaller sample size for the study to achieve the same target number of events at a given study duration. However, it may be more difficult and take longer to recruit high risk patients than regular patients. The sponsor has to carefully balance the event rate versus the difficulty of recruiting such patients. In the aforementioned CV trial example, the hazard rates for the high risk patient population at the initial event monitoring and timeline prediction were assumed to be 10% per patient-year for the first year and 7% per patient-year afterward. About 3 years after the first randomization, 353 out of approximately 5000 enrolled patients had at least one positively adjudicated primary CV event. The observed blinded pooled numbers of events, the exposures and the hazard rates for 6-month time intervals up to 2 years based on the 5000-patient data sweep are presented in Table 1.

Table 1. Blinded pooled hazard rate for the ongoing CV trial based on 5000-patient dataset

Time interval	# of patients with event	Exposures (PYs)	Annual hazard rates (%)
0–6 months	210	2071.5	10.1
6-12 months	90	1375.7	6.5
12-18 months	40	746.1	5.4
18-24 months	13	349.5	3.7

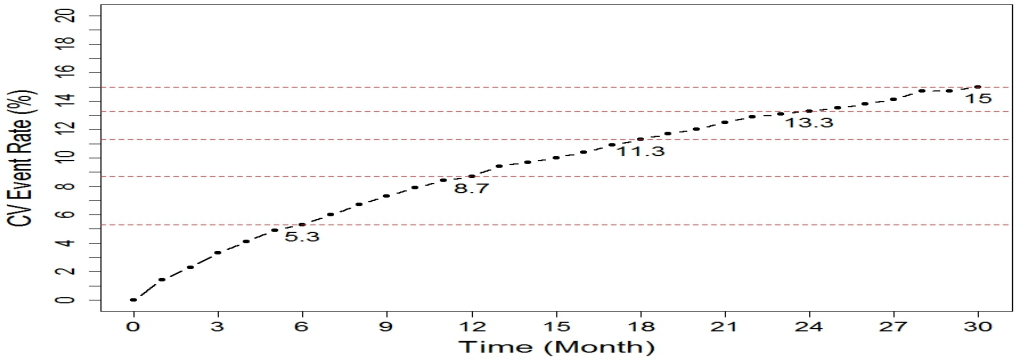
Therefore, the observed pooled event rates were 8.7% (ie, 300 patients in 3447.2 patient-year exposure) and 4.8% (ie, 53 patients in 1095.6 patient-year exposure) per patient-year during the first and second year after the first randomization. There were not enough data for estimating the event rate beyond the second year. These rates were lower than initially assumed, which implied that the total study duration could be longer than what was initially predicted. The use of external data would be helpful to confirm these hazard rates.

Around the same time of the data sweep, results from a similar CV trial called EXAMINE [2] became publicly available. The trial recruited a total of 5380 Type 2 diabetes patients who had had ACS within 15 to 90 days before randomization, which was a similar patient population to the example CV trial. EXAMINE used an adaptive design with several pre-planned interim analyses and the dual objective of demonstrating non-inferiority or superiority of alogliptin compared to placebo on CV risk. The pre-planned interim analysis after 550 positively adjudicated primary CV events demonstrated the non-inferiority of alogliptin compared to placebo. The estimate of the hazard ratio was 0.96. The upper bound of the repeated confidence interval (adjusted for the pre-planned interim and final analyses) was 1.17, which was smaller than the non-inferiority margin of 1.3. Based on these results, the conditional power for the superiority assessment at the final analysis with 650 events was less than 20%. Therefore, on the basis of the protocol, the trial was stopped with non-inferiority but not the superiority claim. An additional 71 patients had CV events after the 550-event interim analysis and before the database lock resulting in a total of 621 patients with CV events in the trial. The hazard ratio from the final analysis was 0.96 and the upper bound of the confidence interval was 1.16. The two Kaplan-Meier (KM) curves of cumulative CV event rates for the two treatment groups (Figure 2 in the EXAMINE publication) were almost identical.

Because of the similarity in populations for EXAMINE and the example CV trial, we certainly could use the results from EXAMINE along with the blinded pooled data from the ongoing CV study to get reliable estimates of the hazard rates to predict the future

events' occurrence and the timeline of reaching the target number of events. Since individual patients' data from EXAMINE were not available from the publication, we used the KM curves to manually measure the approximate cumulative event rates up to some key time points, which are depicted in Figure 1 and Table 2.

Figure 1. Approximate pooled cumulative CV event rates from EXAMINE



With these cumulative event rates, the approximate hazard rates of the corresponding time intervals were derived using a piece-wise exponential distribution for the event time via the following method.

Suppose T_k, k =1, …, K, are the key time points and $P(T_k) = 1 - S(T_k)$, k =1, …, K, are the corresponding cumulative event rates. If a piece-wise exponential distribution is assumed for the event time with hazard rate λ_k for the kth time interval, then the hazard function is

$$\lambda(t) = \lambda_1 1(0 \le t < T_1) + \lambda_2 1(T_1 \le t < T_2) + \lambda_3 1(T_2 \le t < T_3) + \lambda_4 1(T_3 \le t < T_4) + \lambda_5 1(T_4 \le t),$$

where K=5 and $T_k = 0.5 \times k$ years for k =1, …, K. Based on $P(t) = 1 - \exp\left\{-\int_0^t \lambda(s)ds\right\}$ and the observations of $P(T_k)$ from the EXAMINE trial ($P(T_1) = 5.3\%$, $P(T_2) = 8.7\%$, $P(T_3) = 11.3\%$, $P(T_4) = 13.3\%$, $P(T_5) = 15\%$), we can calculate the piece-wise constant annual hazard rates for the time intervals as follows.

$$P(T_1) = 1 - \exp\{-\lambda_1 T_1\} = 0.053 \Rightarrow \lambda_1 = -\log(1 - 0.053)/T_1 = 0.1089$$

$$P(T_2) = 1 - \exp\{-(\lambda_1 - \lambda_2)T_1 - \lambda_2 T_2\} = 0.087 \Rightarrow \lambda_2 = \{\log(1 - 0.087) + \lambda_1 T_1\}/(T_1 - T_2) = 0.0731$$

$$P(T_3) = 1 - \exp\{-(\lambda_1 - \lambda_2)T_1 - (\lambda_2 - \lambda_3)T_2 - \lambda_3 T_3\} = 0.113$$
$$\Rightarrow \lambda_3 = \{\log(1 - 0.113) + \lambda_1 T_1 + \lambda_2 (T_2 - T_1)\}/(T_2 - T_3) = 0.0578$$

$$P(T_4) = 1 - \exp\{-(\lambda_1 - \lambda_2)T_1 - (\lambda_2 - \lambda_3)T_2 - (\lambda_3 - \lambda_4)T_3 - \lambda_4 T_4\} = 0.133$$
$$\Rightarrow \lambda_4 = \{\log(1 - 0.133) + \lambda_1 T_1 + \lambda_2 (T_2 - T_1) + \lambda_3 (T_3 - T_2)\}/(T_3 - T_4) = 0.0456$$

$$P(T_5) = 1 - \exp\{-(\lambda_1 - \lambda_2)T_1 - (\lambda_2 - \lambda_3)T_2 - (\lambda_3 - \lambda_4)T_3 - (\lambda_4 - \lambda_5)T_4 - \lambda_5 T_5\} = 0.15$$
$$\Rightarrow \lambda_5 = \{\log(1 - 0.15) + \lambda_1 T_1 + \lambda_2 (T_2 - T_1) + \lambda_3 (T_3 - T_2) + \lambda_4 (T_4 - T_3)\}/(T_4 - T_5) = 0.0396$$

With a large K, the piece-wise exponential distribution will be close to a smooth non-parametric distribution. Since the purpose here is to predict the timeline for study

planning rather than performing formal data analysis to make claims, *K* of 3-5 should be sufficient. Table 2 shows the annual hazard rates of individual 6-month intervals with *K*=5 based on the KM plot of the EXAMINE trial. They were slightly higher than the observed hazard rates of the example CV trial as shown in Table 1 likely due to a difference in the duration between qualifying ACS event and randomization. Note that if the experimental drug has a CV protective effect, the hazard rate of the experimental drug arm would be smaller than the background rate of the placebo control arm. Thus, the pooled blinded hazard rate would be smaller than the background hazard rate. As anticipated, the hazard rate for an ACS patient population decreased over time.

Table 2. Observed hazard rates for EXAMINE

Time interval	Cumulative rate (%)*	Annual Hazard rate (%)
0–6 months	5.3	10.89
6-12 months	8.7	7.31
12-18 months	11.3	5.78
18-24 months	13.3	4.56
24-30 months	15.0	3.96

*cumulative event rate (%) up to the end of the interval.

3. **Prediction of timeline for ongoing trial**

Intent-to-treat (ITT) design and analysis are often used for major CV trials. With an ITT design, all patients including those who discontinue study medication will be followed until the common end of the study. Their events occurring after treatment discontinuation will still be included in the analysis. Suppose the *i*th patient enters the study at time t_i (relative to the first randomization) and the first event time relative to the randomization of the patient is S_i. For a piece-wise exponential distribution for the event time, the hazard rate is a constant within a pre-specified time interval. For simplification of presentation, let's consider the case of a total of 4 fixed time intervals $[0, y_1)$, $[y_1, y_2)$, $[y_2, y_3)$ and $[y_3, \infty)$ where y_k could be year k or other time value. The ideas can be generalized to any number of time intervals if a smoother non-parametric distribution for event time is more preferable. Again, these assumptions should be reasonable for the purpose of predicting the timeline rather than making a claim. The hazard function is a step function over time

$$\lambda(s) = \lambda_1 1(0 \le s < y_1) + \lambda_2 1(y_1 \le s < y_2) + \lambda_3 1(y_2 \le s < y_3) + \lambda_4 1(y_3 \le s) \qquad (1)$$

where $\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \lambda_4 > 0$ for the example CV trial with a high risk ACS patient population as suggested by the results in Tables 1 and 2. For the other patient population particularly relatively low risk populations, $0 < \lambda_1 \le \lambda_2 \le \lambda_3 \le \lambda_4$ may hold to more realistically reflect the reality as patients age gradually. Here, we consider a simple case of using the pooled hazard rate of the two treatment groups for the overall timeline prediction. This is particularly reasonable if the treatment is non-inferior to the control or has a small impact on the hazard rate (e.g., around 15% overall risk reduction) and treatment is still blinded for the ongoing trial. Similarly, we do not incorporate the impact of treatment discontinuation on the hazard rate for the ITT analysis to avoid the complication of the prediction even though it is achievable [4]. With hazard function (1), the corresponding survival function can be derived as

$$S(s)=\exp\left\{-\int_0^s \lambda(s)ds\right\}=\begin{cases}\exp(-\lambda_1 s), & 0\le s<y_1\\ \exp\{-(\lambda_1-\lambda_2)y_1-\lambda_2 s\}, & y_1\le s<y_2\\ \exp\{-(\lambda_1-\lambda_2)y_1-(\lambda_2-\lambda_3)y_2-\lambda_3 s\}, & y_2\le s<y_3\\ \exp\{-(\lambda_1-\lambda_2)y_1-(\lambda_2-\lambda_3)y_2-(\lambda_3-\lambda_4)y_3-\lambda_4 s\}, & y_3\le s\end{cases}$$

and the density function as

$$f(s)=\lambda(s)S(s)=\begin{cases}\lambda_1 S(s), & 0\le s<y_1\\ \lambda_2 S(s), & y_1\le s<y_2\\ \lambda_3 S(s), & y_2\le s<y_3\\ \lambda_4 S(s), & y_3\le s\end{cases}.$$

Suppose the timeline prediction is performed while the study is still ongoing. For a patient who has already been randomized into the study and had an event, regardless of the randomization time t_i (relative to the first randomization) and event time for the patient, the conditional probability for the patient to have an event during the study is 1 or the patient can be directly counted to contribute one event to the study. For a patient who was randomized at time t_i and has not yet had an event up to time C_i (relative to t_i and time of the prediction), the conditional probability for the patient to have an event by the common study end time *T* (relative to the first randomization) is

$$P_i(T\mid C_i)\equiv P(S_i\le T-t_i\mid S_i>C_i)=\frac{P(C_i<S_i\le T-t_i)}{P(S_i>C_i)},$$

where probability $P(S_i>C_i)$ is the survival function $S(C_i)$.

We need to calculate $P_i(T\mid C_i)$ under all possible scenarios of $T-t_i$ and C_i.

If $0\le C_i<T-t_i<y_1$,

$$P_i(T\mid C_i)=\left(\int_{C_i}^{T-t_i}\lambda_1\exp\{-\lambda_1 s\}ds\right)\Big/\exp(-\lambda_1 C_i)=1-\exp\{-\lambda_1(T-t_i-C_i)\}.$$

If $y_1\le T-t_i<y_2$, then

if $0\le C_i<y_1$,

$$P_i(T\mid C_i)=1-\exp\{-\lambda_1(y_1-C_i)-\lambda_2(T-t_i-y_1)\};$$

otherwise if $y_1\le C_i<T-t_i$,

$$P_i(T\mid C_i)=1-\exp\{-\lambda_2(T-t_i-C_i)\}.$$

If $y_2\le T-t_i<y_3$, then

if $0\le C_i<y_1$,

$$P_i(T\mid C_i)=1-\exp\{-\lambda_1(y_1-C_i)-\lambda_2(y_2-y_1)-\lambda_3(T-t_i-y_2)\};$$

otherwise if $y_1\le C_i<y_2$,

$$P_i(T\mid C_i)=1-\exp\{-\lambda_2(y_2-C_i)-\lambda_3(T-t_i-y_2)\};$$

or if $y_2\le C_i<T-t_i$,

$$P_i(T\mid C_i)=1-\exp\{-\lambda_3(T-t_i-C_i)\}.$$

If $y_3\le T-t_i$, then

if $0\le C_i<y_1$,

$$P_i(T\mid C_i)=1-\exp\{-\lambda_1(y_1-C_i)-\lambda_2(y_2-y_1)-\lambda_3(y_3-y_2)-\lambda_4(T-t_i-y_3)\};$$

if $y_1 \le C_i < y_2$,

$$P_i(T \mid C_i) = 1 - \exp\{-\lambda_2(y_2 - C_i) - \lambda_3(y_3 - y_2) - \lambda_4(T - t_i - y_3)\};$$

if $y_2 \le C_i < y_3$,

$$P_i(T \mid C_i) = 1 - \exp\{-\lambda_3(y_3 - C_i) - \lambda_4(T - t_i - y_3)\};$$

if $y_3 \le C_i < T - t_i$,

$$P_i(T \mid C_i) = 1 - \exp\{-\lambda_4(T - t_i - C_i)\}.$$

For a patient who has not yet been randomized/entered into the study, suppose the predicted randomization time is t_i, then the conditional probability for the patient to have an event by the common study end time *T* is a special case of $P_i(T \mid C_i)$

$$P_i(T \mid C_i = 0) \equiv P(S_i \le T - t_i \mid S_i > 0) = P(S_i \le T - t_i)$$

or by setting the event-free time C_i=0 in the formulas of $P_i(T \mid C_i)$. Clearly, the larger the $T - t_i$, the larger the conditional probability for a patient to have an event during the study.

After obtaining the conditional probabilities of having events by time *T* (the common study end time) for individual patients, we can sum them up to get the total expected number of events by time *T* as

$$\Omega(T) = \sum_{i=1}^{N} P_i(T \mid C_i), \qquad (2)$$

where *N* is the total number of patients in the study. Note that $P_i(T \mid C_i)$=1 for a patient who has had an event observed by the time when the prediction is performed. Again, $\Omega(T)$ of (2) is an increase function of *T*. Therefore, for any specific target number of events *A*, $\Omega(T) = A$ has a unique solution T^*. This T^* is the predicted timeline for reaching the target number of events.

Total exposure is another indicator for the amount of data from the study. With this specific T^*, we can calculate the observed and expected total exposure for the study. The exposure E_i for a patient with an event already observed when the timeline prediction is performed is the time between the randomization and the patient's event time. For a patient who has not yet been randomized into the study, the expected exposure is

$$E_i = E\min(T^* - t_i, S_i)$$

where again t_i is the predicted patient's randomization time relative to the first randomization and S_i is the event time from the patient's randomization that follows the piece-wise exponential distribution. For a patient who has been randomized into the study and has not yet had an event by time C_i (relative to the t_i so $0 \le C_i \le T^* - t_i$), the conditional expected exposure is $E_i = E\min(T^* - t_i, S_i \mid S_i \ge C_i)$. The total exposure for the whole study is the sum of all the exposures of individual patients $E = \sum_{i=1}^{N} E_i$.

4. **Prediction for on-treatment analysis**

Because of the concern of dilution of treatment effect after treatment discontinuation, the on-treatment analysis or treatment emergent analysis rather than the ITT analysis may be the primary analysis for safety evaluation. For the on-treatment analysis, only an event which occurs within the on-treatment period and the corresponding exposure between

randomization and earliest of event time and treatment discontinuation are included in the analysis. When a CV event is considered as an adverse event for non-inferiority analysis, on-treatment analysis in addition to an ITT analysis should also be performed. Formulas for calculating the conditional probability of having an event during the study for the on-treatment analysis are different from those of the ITT analysis.

Suppose the discontinuation time for the *i*th patient D_i (relative to t_i) follows a distribution $G(d)$ for both treatment groups to simplify the presentation particularly for blinded pooled analysis. If a patient has been observed to have an event before treatment discontinuation, the conditional probability for the patient to have an event for the on-treatment analysis is 1. If a patient has been randomized into the study with the observed event-free time C_i and has not yet discontinued treatment when the timeline prediction is performed, the conditional probability for the patient to have an event for the on-treatment analysis is

$$P_i^o(T \mid C_i) \equiv P(S_i \le \min(T - t_i, D_i) \mid S_i > C_i, D_i > C_i) = \frac{P(C_i < S_i \le \min(T - t_i, D_i) \mid C_i)}{P(S_i > C_i, D_i > C_i \mid C_i)}, \quad (3)$$

where as before, S_i is assumed to follow the same piece-wise exponential distribution for both treatment groups for the purpose of timeline prediction particularly for non-inferiority assessment. Nonetheless, the derived hazard rates based on the pooled blinded data of the on-treatment analysis could be different from those of the ITT analysis (presented in Table 1 for the example CV trial).

To ease the calculation of (3), we can assume that D_i and S_i are independent, which isusually the case for survival analysis. Then $P(S_i > C_i, D_i > C_i \mid C_i) = P(S_i > C_i \mid C_i)$ $P(D_i > C_i \mid C_i)$. We may further assume D_i to follow an exponential distribution with a constant hazard rate of η. Under these assumptions, the calculation of $P_i^0(T \mid C_i)$ (3) under all possible scenarios of $T - t_i$ and C_i can be found as follows.

Let $t'_i = T - t_i$, if $0 \le C_i < t'_i < y_1$,

$$P_i^0(T \mid C_i) = \frac{\lambda_1}{\lambda_1 + \eta}\left(1 - \exp\{-(\lambda_1 + \eta)(t'_i - C_i)\}\right).$$

If $y_1 \le t'_i < y_2$, then

if $0 \le C_i < y_1$,

$$P_i^0(T \mid C_i) = \frac{\lambda_1}{\lambda_1 + \eta} + \left(\frac{\eta}{\lambda_1 + \eta} - \frac{\eta}{\lambda_2 + \eta}\right)\exp\{-(\lambda_1 + \eta)(y_1 - C_i)\} - \frac{\lambda_2}{\lambda_2 + \eta}\exp\{-(\lambda_1 - \lambda_2)y_1 - (\lambda_2 + \eta)t'_i + (\lambda_1 + \eta)C_i)\}$$

otherwise if $y_1 \le C_i < t'_i$,

$$P_i^0(T \mid C_i) = \frac{\lambda_2}{\lambda_2 + \eta}\left(1 - \exp\{-(\lambda_2 + \eta)(t'_i - C_i)\}\right).$$

If $y_2 \le t'_i < y_3$, then

if $0 \le C_i < y_1$,

$$P_i^0(T \mid C_i) = \frac{\lambda_1}{\lambda_1 + \eta} + \left(\frac{\eta}{\lambda_1 + \eta} - \frac{\eta}{\lambda_2 + \eta}\right)\exp\{-(\lambda_1 + \eta)(y_1 - C_i)\}$$
$$+ \left(\frac{\eta}{\lambda_2 + \eta} - \frac{\eta}{\lambda_3 + \eta}\right)\exp\{-(\lambda_1 - \lambda_2)y_1 - (\lambda_2 + \eta)y_2 + (\lambda_1 + \eta)C_i\}$$

$$-\frac{\lambda_3}{\lambda_3+\eta}\exp\{-(\lambda_1-\lambda_2)y_1-(\lambda_2-\lambda_3)y_2-(\lambda_3+\eta)t'_i+(\lambda_1+\eta)C_i\};$$

otherwise, if $y_1 \le C_i < y_2$,

$$P_i^0(T \mid C_i)$$

$$=\frac{\lambda_2}{\lambda_2+\eta}+\left(\frac{\eta}{\lambda_2+\eta}-\frac{\eta}{\lambda_3+\eta}\right)\exp\{-(\lambda_2+\eta)(y_2-C_i)\}-\frac{\lambda_3}{\lambda_3+\eta}\exp\{-(\lambda_2-\lambda_3)y_2-(\lambda_3+\eta)t'_i+(\lambda_2+\eta)C_i\}$$

or if $y_2 \le C_i < t'_i$,

$$P_i^0(T \mid C_i) = \frac{\lambda_3}{\lambda_3+\eta}\left(1-\exp\{-(\lambda_3+\eta)(t'_i-C_i)\}\right).$$

If $y_3 \le t'_i$, then

if $0 \le C_i < y_1$,

$$P_i^0(T \mid C_i) = \frac{\lambda_1}{\lambda_1+\eta}+\left(\frac{\eta}{\lambda_1+\eta}-\frac{\eta}{\lambda_2+\eta}\right)\exp\{-(\lambda_1+\eta)(y_1-C_i)\}$$

$$+\left(\frac{\eta}{\lambda_2+\eta}-\frac{\eta}{\lambda_3+\eta}\right)\exp\{-(\lambda_1-\lambda_2)y_1-(\lambda_2+\eta)y_2+(\lambda_1+\eta)C_i\}$$

$$+\left(\frac{\eta}{\lambda_3+\eta}-\frac{\eta}{\lambda_4+\eta}\right)\exp\{-(\lambda_1-\lambda_2)y_1-(\lambda_2-\lambda_3)y_2-(\lambda_3+\eta)y_3+(\lambda_1+\eta)C_i\}$$

$$-\frac{\lambda_4}{\lambda_4+\eta}\exp\{-(\lambda_1-\lambda_2)y_1-(\lambda_2-\lambda_3)y_2-(\lambda_3-\lambda_4)y_3-(\lambda_4+\eta)t'_i+(\lambda_1+\eta)C_i\};$$

if $y_1 \le C_i < y_2$,

$$P_i^0(T \mid C_i) = \frac{\lambda_2}{\lambda_2+\eta}+\left(\frac{\eta}{\lambda_2+\eta}-\frac{\eta}{\lambda_3+\eta}\right)\exp\{-(\lambda_2+\eta)(y_2-C_i)\}$$

$$+\left(\frac{\eta}{\lambda_3+\eta}-\frac{\eta}{\lambda_4+\eta}\right)\exp\{-(\lambda_2-\lambda_3)y_2-(\lambda_3+\eta)y_3+(\lambda_2+\eta)C_i\}$$

$$-\frac{\lambda_4}{\lambda_4+\eta}\exp\{-(\lambda_2-\lambda_3)y_2-(\lambda_3-\lambda_4)y_3-(\lambda_4+\eta)t'_i+(\lambda_2+\eta)C_i\};$$

if $y_2 \le C_i < y_3$,

$$P_i^0(T \mid C_i) = \frac{\lambda_3}{\lambda_3+\eta}+\left(\frac{\eta}{\lambda_3+\eta}-\frac{\eta}{\lambda_4+\eta}\right)\exp\{-(\lambda_3+\eta)(y_3-C_i)\}$$

$$-\frac{\lambda_4}{\lambda_4+\eta}\exp\{-(\lambda_3-\lambda_4)y_3-(\lambda_4+\eta)t'_i+(\lambda_3+\eta)C_i\};$$

if $y_3 \le C_i < t'_i$,

$$P_i^0(T \mid C_i) = \frac{\lambda_4}{\lambda_4+\eta}\left(1-\exp\{-(\lambda_4+\eta)(t'_i-C_i)\}\right).$$

For a patient who has not been randomized into the study, the conditional probability to have an event during the study before treatment discontinuation can be obtained by setting C_i=0 in (3) or directly calculating $P(S_i \le \min(T-t_i, D_i))$. For patients who had no events and discontinued before the timeline prediction, their conditional probability to

have an event for the on-treatment analysis is zero. The timeline for reaching a target number of events *B* for the on-treatment analysis is the solution T^o of

$$\Omega^o(T) = \sum_{i=1}^{N} P_i^o(T \mid C_i) = B$$

where $\Omega^o(T)$ is a monotone increasing function of *T*. As discussed previously, for non-inferiority analysis, the required number of events should be much smaller than that of the superiority analysis. That is, *B*<*A*. Therefore, T^o could be smaller than $T*$ even though events after treatment discontinuation are not counted in deriving T^o. The total exposure for the on-treatment analysis can be derived similarly.

5. **Adaptation**

A long term and large scale CV trial could be designed to allow pre-planned adaptation during the trial [4]. For example, EXAMINE used an adaptive design with a potential to stop the trial for futility or make an early claim of non-inferiority or superiority based on interim results. The trial was actually stopped early with a non-inferiority claim. Another CV trial, SAVOR-TIMI 53 [5], was designed with the primary objective of demonstrating the superiority of saxagliptin compared to placebo on the reduction of CV risk. A total of 16492 T2DM patients who had a history of or who were at risk for cardiovascular events were randomized into the study. This population is different from the ACS population. Final results from the study with 1222 events showed the hazard ratio of saxagliptin versus placebo of 1.00 (95% CI: 0.89 to 1.12; p=0.99). The result could be used only to claim non-inferiority of saxagliptin. If an adaptation was pre-specified and performed for the trial, non-inferiority could be claimed and the study stopped sooner with probably half of 1222 events and a smaller sample size.

A trial could always be modified including adding an adaptation strategy based on blinded interim results or external information. Such a trial modification will not introduce bias to trial results. After the publications of the results of both EXAMINE and SAVOR- TIMI 53 that showed only non-inferiority of two DPP-4 drugs, the sponsor of a diabetes drug, particularly a DPP-4 drug, may want to have a potential adaptation step in a CV trial. One example adaptation strategy can be described as follows.

If the interim result with 612 events fails to show non-inferiority, the trial will be stopped and no claim will be made. If the result with 612 events shows non-inferiority of the drug at a significance level of 2.5% (one-sided) and conditional power for the originally planned superiority assessment with 844 events is small (say less than 15%), the trial is stopped with only the non-inferiority claim. Otherwise, if the observed interim result is consistent with what was expected at the design stage, the trial will be continued and the superiority assessment will be performed with 844 events at a significance level of 2.5% (one-sided). Such an adaptive strategy will not inflate the overall type I error rate even though the nominal significance level is used for testing two hypotheses with different datasets as shown below.

Suppose the point estimate of the log hazard ratio from the interim analysis is $\hat{\delta}_1$ and the total number of events for the interim analysis is E_1 (=612 in our case). Similarly, let $\hat{\delta}_2$ be the point estimate of the log hazard ratio and E_2 be the total number of events of the second stage (=844-612=232 in our case). Then, the total number of events for the final

analysis is $E = E_1 + E_2$ and the final point estimate is approximately $\hat{\delta} = \frac{E_1\hat{\delta}_1 + E_2\hat{\delta}_2}{E}$. The non-inferiority null and alternative hypotheses are

$$H_{01} : \delta \geq \log(\Delta) \quad \text{versus} \qquad H_{a1} : \delta < \log(\Delta)$$

where δ is the true log hazard ratio and Δ is the non-inferiority margin (=1.3 in our case). The corresponding superiority null and alternative hypotheses are

$$H_{02} : \delta \geq 0 \quad \text{versus} \qquad H_{a2} : \delta < 0 .$$

Asymptotically, $\hat{\delta}_i \sim N(\delta, 4/E_i)$, i=1, 2, and $\hat{\delta} \sim N(\delta, 4/E)$. Non-inferiority will be claimed with E_1 events at a significance level α if the upper bound $\hat{\delta}_1 + z_\alpha\sqrt{4/E_1} < \log(\Delta)$, where z_α is the $1-\alpha$ percentile of the standard normal distribution and is 1.96 when $\alpha = 0.025$. If conditional power for the superiority claim for the E -event analysis is sufficient, the trial will be continued to E events. Then the superiority will be claimed if $\hat{\delta} + z_\alpha\sqrt{4/E} < 0$. With such a strategy, the overall type I error rate of making any claims is controlled.

First, clearly,

$$\Pr(\hat{\delta}_1 + z_\alpha\sqrt{4/E_1} < \log(\Delta) \mid H_{01}) \leq \alpha \qquad (4)$$

and

$$\Pr(\hat{\delta} + z_\alpha\sqrt{4/E} < 0 \mid H_{02}) \leq \alpha . \qquad (5)$$

Note that $H_{01}H_{02} = H_{01}$ and

$$\begin{aligned}\Pr(\hat{\delta} + z_\alpha\sqrt{4/E} < 0 \mid H_{01}H_{02}) &= \Pr(\hat{\delta} + z_\alpha\sqrt{4/E} < 0 \mid H_{01}) \\ &\leq \Pr((\hat{\delta} - \log(\Delta))/\sqrt{4/E} < -z_\alpha - \log(\Delta)/\sqrt{4/E} \mid \delta = \log(\Delta)) \\ &\approx \Pr(Z < -z_\alpha - \log(\Delta)/\sqrt{4/E}) = \Pr(Z < -5.77) \approx 0 \qquad (6)\end{aligned}$$

where Z has a standard normal distribution; for Δ =1.3, E =844 and z_α =1.96, $z_\alpha + \log(\Delta)/\sqrt{4/E} = 5.77$. Therefore,

$$\begin{aligned}&\Pr(\hat{\delta}_1 + z_\alpha\sqrt{4/E_1} < \log(\Delta) \text{ or } \hat{\delta} + z_\alpha\sqrt{4/E} < 0 \mid H_{01}H_{02}) \\ &\approx \Pr(\hat{\delta}_1 + z_\alpha\sqrt{4/E_1} < \log(\Delta) \mid H_{01}) = \alpha . \qquad (7)\end{aligned}$$

Combining (4), (5) and (7), the overall error rate of making any claims is basically controlled at no more than α. EXAMINE used a 1% significance level for the interim analysis and a 1.5% significance level for the final analysis to control the overall type I error rate at 2.5%. Such a Bonferroni type multiplicity adjustment strategy ignoring the correlation between the statistics of the two stages is very conservative.

If the proposed procedure fails to demonstrate non-inferiority with 612 events at a significance level of 2.5%, $\hat{\delta}_1 + z_\alpha\sqrt{4/E_1} > \log(1.3)$ and $\hat{\delta}_1 > \log(1.3) - 1.96/\sqrt{153}$ =0.262-0.158=0.104. The estimated hazard ratio would be approximately $e^{\hat{\delta}_1} > e^{0.104} = 1.11$. Increasing the CV risk by 11% for patients with around a 6% patient-year rate implies the NNH (number needed to treat to harm one additional patient) of 152. In such a case, health authorities may have concern when reviewing the NDA package. Separately, power to show non-inferiority at a significance level of 2.5% with 612 events when the true log hazard ratio is zero is already

$$\Pr(\hat{\delta}_1 + z_{0.025}\sqrt{4/E_1} < \log(\Delta) \mid \delta = 0) = 0.9 .$$

Spending all alpha for the non-inferiority assessment with 612 events is a reasonable option. If the sponsor really wants to have another chance to assess non-inferiority at the final analysis with 844 events, the entire 2.5% alpha should not be spent at the 612-event analysis. When a similar approach of Li et al. [6] is applied, non-inferiority will be claimed at level α_1 (<0.025) for the 612-event (Stage 1) analysis. If it fails, non-inferiority will be claimed at the final analysis with 844-events (stage 2) if both the 612-event and 844-event analyses demonstrate non-inferiority at a significance level α or the 844-event analysis demonstrates non-inferiority at a significance level α_2 (<0.025). This is a Hochberg [7] type multiplicity adjustment approach and is much more powerful than the Bonferroni approach used in EXAMINE. Given α_1, taking into account the correlation between $\hat{\delta}_1$ and $\hat{\delta}$, α_2 can be derived through

$$\Pr[UB_{1,\alpha_1} < \log(\Delta) \text{ or } (UB_{1,\alpha} < \log(\Delta) \text{ and } UB_{2,\alpha} < \log(\Delta)) \text{ or } UB_{2,\alpha_2} < \log(\Delta) \mid H_{01}) = \alpha$$

where $UB_{l,a}$ is the level *a* confidence upper bound for δ of stage l analysis. The sponsor would like to demonstrate superiority at the 844-event analysis. Therefore, no superiority assessment is performed at the 612-event analysis and all alpha for the superiority assessment is reserved for the final analysis. Since the trial could potentially be stopped at the 612-event analysis with the non-inferiority claim, it would also be of interest to predict the timeline to reach this number of events based on the on-treatment analysis.

6. **Example CV trial**

In the blinded 5000-patient data set for the ongoing CV trial that was planned to randomize a total of 6000 patients, 356 patients (353 adjudicated and confirmed cases in the database plus 3 likely new cases) had had at least one primary CV outcome event. For the primary objective of the study of assessing the,superiority, the target number of events for 90% power to detect a 20% risk reduction was 844. Therefore, we needed to predict the timeline T^* for the other patients to have 488 additional events. Separately, for the non-inferiority assessment, the target number of events was 612, so we needed also to predict the timeline of having 266 additional events. For the prediction, we considered 4 time intervals for the piece-wise exponential distribution for the event time and four scenarios for the hazard rate configurations as in Table 3. Scenarios 1 and 2 mimicked the combination of the observed blinded pooled hazard rates from the ongoing study and the derived hazard rates from EXAMINE. Scenarios 3 and 4 reflected slightly higher and lower hazard rates for sensitivity predictions, respectively. In addition to the 356 observed events, the prediction utilized all available information in the database

including the observed randomization times and event-free times for those randomized. For patients who had not yet been randomized, their randomization times were predicted based on a refined enrollment model and their event-free times were set to zero.

Table 3. Hazard rate scenarios for the timeline predictions

Time interval	Hazard rate (% per patient-year)			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
0-1 Year	8.7	8.7	10.0	7.0
1-2 Year	4.8	4.8	6.0	4.0
2-3 Year	4.0	3.5	4.5	3.5
>Year 3	3.5	3.0	3.5	3.0

For the ITT analysis, the timeline prediction T^* was based on

$$\Omega(T^*) = \sum_{i=1}^{6000-356} P_i(T^* \mid C_i) + 356 = 844$$

where the summation was over those 6000-356=5664 patients who had not yet had an event before the prediction. Figure 2 shows the cumulative number of primary CV events over time for the 4 scenarios of hazard rates based on the ITT analysis. It was predicted that 844 patients would have the primary CV events by 55 months from the first randomization under scenario 1, by 56 months under scenario 2, by 51 months under scenario 3 and by 60 months under scenario 4. The timeline gap between scenarios 1 or 2 and 3 was about 4 to 5 months, and the gap between scenarios 3 and 4 was 9 months. Note that these gaps were for getting 844-356=488 additional events. The predicted time was the time when 844 primary CV events occurred. Additional time (e.g. 2-3 months) would still be needed to get all these events reported, adjudicated, and confirmed as well as all data cleaned for the database lock and analysis. Most likely, the final number of events would exceed 844 as in the case of the EXAMINE trial where the target number of events was 550 but the final number was 621.

Figure 2. Cumulative number of CV events over time based on ITT analysis

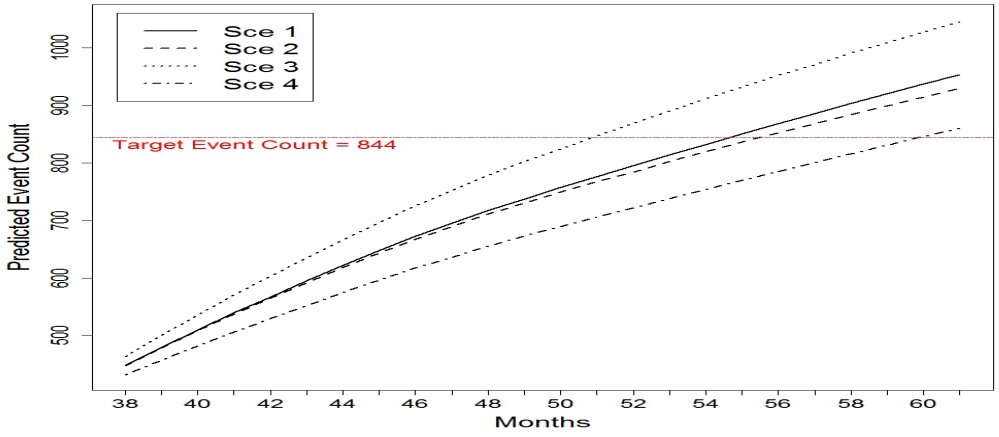


Table 4 presents the other summary statistics for the predictions under the 4 hazard rate scenarios. Since the hazard rate was a function of time, this average hazard rate was also a function of the study duration and other factors.

Table 4. Other summary statistics for the ITT analysis

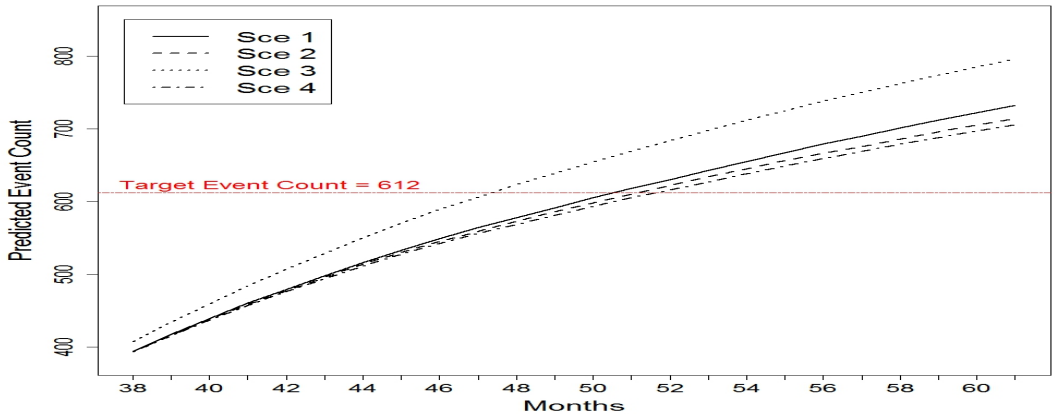
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Predicted timeline	55 months	56 months	51 months	60 months
Predicted Total Exposure (Year)	13746.42	14189.95	11944.73	16014.69
Estimated Overall Event Rate (Event Count/Total Exposure)	6.18%	6.00%	7.09%	5.28%

For the on-treatment analysis, hazard rates based on blinded data of the 5000-patient dataset from the example CV trial were 8.8%, 4.6%, 3.4% and 3.0% per patient-year for the first year, second year, third years and afterward, respectively. They were used as scenarios 4 along with the hazard rates of the other 3 scenarios of Table 3 for the timeline prediction for the on-treatment analysis. Furthermore, by the time of the prediction, 320 patients had had events observed during the on-treatment period. Among those who had not had events or not been randomized, 4433 patients had not yet discontinued treatment and the other 403 patients had discontinued treatment. The timeline prediction T^o for a total of 612 events based on the on-treatment analysis should be the solution of

$$\Omega^o(T^o) = \sum_{i=1}^{4433} P_i^o(T^o \mid C_i) + 320 = 612 .$$

Figure 3 presents the cumulative number of events over time for the on-treatment analysis for the 4 scenarios of hazard rates. For example, based on the on-treatment hazard rates of the blinded data from the study (scenario 4), 612 patients would have on-treatment primary CV events by around 52 months from the first randomization of the study. It was 3 months ahead of the corresponding timeline for reaching 844 events for the ITT analysis (scenario 1 in Figure 2).

Figure 3. Cumulative number of CV events over time based on on-treatment analysis



7. Discussion

In this paper, we discussed timeline prediction for a major CV trial. The hazard rate of CV event trials can have a big impact on the timeline. Therefore, the first step is to obtain a reliable estimate of the hazard rate preferably through internal data and to confirm the estimate by external published results of a similar population. In addition, the hazard rate should be varied slightly in the prediction to assess the robustness of the prediction

results. It is reasonable to apply a piece-wise exponential distribution for event time to the prediction. With such a distribution, the conditional probability for a patient to have an event during the trial given the patient's available information can be easily derived. The sum of all the conditional probabilities across all patients is the expected total number of events and is a monotone increasing function of total study duration. Thus, the derivation of the timeline for reaching the target number of events becomes very straightforward.

The predicted timeline is the theoretical timeline of the occurrence of the target number of events. For events which need go through an adjudication process for confirmation, we should keep in mind that there is always a time gap between the occurrence and confirmation of the target number of events. This non-ignorable time gap should be taken into account when the predicted timeline is communicated to the trial sponsor's management.

Ideas and thinking provided here can be applied to other event-driven trials with some modification and simplifications. For a trial with a relatively constant hazard rate throughout the whole study, the number of pieces for the piece-wise exponential distribution for event time can be reduce. For a trial with a fixed treatment duration for all patients (e.g., 2 years), sample size rather than the T should adjusted for reaching the target number of events. Also, with such a design, t_i for all patients should be set to 0 in all the formulas.

References

[1] FDA, Guidance for Industry: diabetes mellitus-evaluating cardiovascular risk in new antidiabetic therapies to treat Type 2 diabetes, December 2008.

[2] Willaim et al. Alogliptin after acute coronary syndrome in patients with type 2 diabetes. The New England Journal of Medicine 2013.

[3] Tricoci et al. Thrombin-receptor antagonist Vorapaxar in acute coronary syndromes. The New England Journal of Medicine 2011.

[4] Quan H, Li M, Zhao PL, Cho M, Zhang J and Wu Y. Considerations for design and data analysis of noninferiority/superiority cardiovascular trials. Journal of Biopharmaceutical Statistics 2013; 23: 239-260.

[5] Benjamin et al. Saxagliptin and cardiovascular outcomes in patients with type 2 diabetes mellitus. The New England Journal of Medicine 2013.

[6] Li S, Hussey MA, Schwartz TA and Koch GG. A multistage analysis strategy for a clinical trial to assess successively more stringent criteria for a primary endpoint with a low event rate. Pharmaceutical Statistics, 2013, 12: 65-73.

[7] Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988; 75, 800-802.