JOURNAL OF BIOPHARMACEUTICAL STATISTICS
2016, VOL. 26, NO. 2, 191–201
http://dx.doi.org/10.1080/10543406.2014.972508

Taylor & Francis
Taylor & Francis Group

Sequentially updating the likelihood of success of a Phase 3 pivotal time-to-event trial based on interim analyses or external information

Kaspar Rufibach, Paul Jordan, and Markus Abt

F. Hoffmann-La Roche Ltd., Product Development Biostatistics, Basel, Switzerland

ABSTRACT

When performing a pivotal clinical trial, it may be of interest to assess the probability of success (PoS) of that trial. Initially evaluated when the trial is designed, PoS can be updated as the trial progresses and new information about the drug effect becomes available. Such information can be external to the trial, such as results from trials conducted in parallel, or internal, such as continuing after an interim analysis. We develop a framework to update PoS based on such internal and external information for a time-to-event endpoint and illustrate it using a realistic development program for a new molecule.

ARTICLE HISTORY
Received 8 May 2014
Accepted 30 September 2014

KEYWORDS
Bayesian predictive power; conditional power; interim analysis; prior distribution; probability of technical success

1. Introduction

Clinical development programs require a considerable investment of resources in possibly large and long Phase 3 studies to establish safety and efficacy of new therapeutic agents. Budgets are tight and therefore investment decisions need to carefully assess the likelihood of development programs being successful. To support informed decisions, the evaluation of the probability of success (PoS) of Phase 3 studies plays a key role. PoS, called Probability of Study Success in Wang et al. (2013), is nothing else than Bayesian predictive power (Spiegelhalter et al., 1986), i.e., the expectation of the conditional power with respect to a distribution for the true underlying drug effect θ. The latter quantifies our current knowledge or belief about θ. See Section 2.5 for a more thorough discussion of conditional and Bayesian predictive power.

PoS is typically derived by systematically incorporating all available data regarding the efficacy of the molecule in a quantitative assessment. Two recent publications have been describing statistical frameworks that allow doing so. Zhang and Zhang (2013) assume that already at the end of Phase 2, information on the benefit of the drug on endpoints that will be used in Phase 3 is available. Based on this they describe a mathematical framework to derive the likelihood for one or more Phase 3 studies to be successful. Wang et al. (2013) acknowledge that Phase 2 and Phase 3 studies may have used different endpoints. Yet, modeling the uncertainty in the observed Phase 2 results as well as in available information describing the relationship between the Phase 2 and Phase 3 endpoints will again allow estimation of the PoS.

We describe our proposed framework in terms of a time-to-event setup under the proportional hazards assumption, using the hazard ratio as effect measure. We use the fact that the logarithm of the estimate of the hazard ratio can be approximated by a Normal distribution with known variance. While the mathematical details will be developed for this situation, the conceptual ideas also apply to other endpoints.

CONTACT Kaspar Rufibach kaspar.rufibach@roche.com F. Hoffmann-La Roche Ltd., Product Development Biostatistics, Building 670, CH-4070 Basel, Switzerland.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lbps.

Time-to-event studies needed to establish efficacy and safety can be very large. For example, in cardiovascular research, one large pivotal Phase 3 study may be conducted in a higher risk population and be the basis for the initial regulatory approval. In parallel to this large pivotal Phase 3 study, smaller safety or mechanistic supporting studies may be conducted. The results of these smaller trials may become available while the pivotal Phase 3 study is still ongoing. The information obtained can be used in reassessing the PoS of the pivotal Phase 3 study. In what follows, we call updates based on such information "external," as they are external to the pivotal Phase 3 study.

Within the pivotal Phase 3 study itself, interim analyses may be performed to enable early stopping for overwhelming efficacy or futility, based on prespecified efficacy and/or futility boundaries θ_{eff} and θ_{fut}. Passing such interim analyses based on the recommendation of a Data and Safety Monitoring Board (DSMB) to continue the study as planned may also be seen as information that can be used in a reassessment of the PoS. As this information comes from the pivotal Phase 3 study itself, we call such updates "internal."

However, using DSMB recommendations to update PoS faces the challenge that the trial sponsor remains blinded to the study results. This implies that when receiving the recommendation to continue a trial after an interim analysis, the sponsor does not come to know the actual interim effect estimate $\hat{\theta}_{\text{int}}$, but may only conclude that $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$.

In this article, we will discuss statistical approaches that can be used to sequentially update PoS of a Phase 3 study with information as it becomes available in parallel to the conduct of that trial. In addition, we will illustrate how to generalize the concept of Bayesian predictive power to incorporate a DSMB recommendation leading to internal interval information of the type $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$.

In Section 2, the statistical methodology will be introduced. We will define what we mean by "study success," discuss the framework of Bayesian predictive power, and how to update it by incorporating external information regarding the drug effect that may become available sequentially. In addition, we will discuss how to incorporate interval knowledge about $\hat{\theta}_{\text{int}}$.

The methodology was developed to update PoS in a large clinical development program, and the numbers used for illustration reflect approximately what we saw in that program. However, for publication purposes, we anonymized the data, i.e., we do not give explicit details of the program.

We show the type of information that can sequentially become available in a development program while the pivotal Phase 3 study is running. Starting from two different hypothetical priors described in Section 4, we will show in Section 5 how PoS may change over time.

In addition to some practical recommendations, the discussion in Section 6 will provide suggestions for possible extensions of this work. Finally, some computations are deferred to the appendix.

2. Statistical methods

2.1. *Time-to-event setup*

Assume that in a large pivotal Phase 3 study, we randomize N patients in equal proportions to two treatment arms. We are interested in a time-to-event endpoint such as time to tumor progression or death in oncology or major adverse cardiac events in the cardiovascular area. In these situations, effectiveness of a therapy is expressed by the hazard ratio λ. The estimated log-hazard ratio $\hat{\theta} = \log \hat{\lambda}$ can be well approximated by a normal distribution with mean true log-hazard ratio θ and variance 4/d, where d is the total number of events in both arms at the time of analysis, see, e.g., Jennison and Turnbull (2000).

Since the proposed methodology depends on the α-spending strategy only through the values of the α spent at the interim and final analyses, in what follows, we assume without loss of generality that one interim and the final analysis are performed in a pivotal Phase 3 trial, with corresponding log-hazard ratio estimates $\hat{\theta}_{\text{int}}$ and $\hat{\theta}_{\text{fin}}$.

2.2. Definition of trial success

Assume that for prespecified overall Type I and Type II errors α and β, the pivotal Phase 3 trial was planned assuming an alternative log-hazard ratio, requiring d events to reject the null hypothesis of no effect. We denote by α_{fin} the nominal significance level at the final analysis resulting from a chosen α-spending function. At the final analysis, we then reject the null hypothesis $\theta = 0$ in our pivotal Phase 3 trial if the two-sided logrank test is significant at α_{fin}, so $\hat{\theta}_{\text{fin}}$ must fulfill

$$\hat{\theta}_{\text{fin}} \leq -z_{1-\alpha_{\text{fin}}/2}\text{SE}(\hat{\theta}_{\text{fin}}) = -z_{1-\alpha_{\text{fin}}/2}\sqrt{4/d_{\text{fin}}} =: \theta_{\text{suc}}. \quad (1)$$

Therein, θ_{suc} ("suc" for success) is the largest observed log-hazard ratio that would still lead to a statistically significant result at the final analysis and z_γ is the γ-quantile of the standard normal distribution. Typically, $\alpha_{\text{fin}} < \alpha$ in a group-sequential trial.

If interim analyses are yet to be performed, boundaries for the hazard ratio can be derived based on the α-spending function and success may be defined as the observed hazard ratio to be below the boundary at any of the (interim or final) analyses still to be conducted.

2.3. Definition and computation of probability of success

The PoS at the final analysis is given by $P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}})$. This quantity depends on the unknown true underlying log-hazard ratio θ, since $\hat{\theta}_{\text{fin}} \sim \text{N}(\theta, 4/d_{\text{fin}})$. While this true effect θ of the drug is not known, a range of plausible values may be possible to describe via a distribution q_0 for θ which then leads to the definition

$$\text{PoS} := \mathbb{E}_\theta\left(P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}})\right) = \int P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}}) q_0(\theta) d\theta \quad (2)$$

for the PoS. To compute (2), Wang et al. (2013) suggest to simulate trials and approximate PoS based on those simulations. However, in many examples, (2) can in fact simply be evaluated using numerical integration (e.g., via the function integrate in R, R Core Team, 2013) or even be expressed in closed form.

In what follows, we will discuss how to update the prior $q_0(\theta)$ once information regarding the treatment effect θ becomes available and how to update PoS with interval information of the type $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$.

2.4. Updating PoS based on external information

As discussed in Section 1, smaller safety, mechanistic, or other supporting studies may be conducted in parallel to the pivotal Phase 3 trial. While the primary endpoint in such trials is often different from the time-to-event endpoint that is evaluated in the Phase 3 study, information on this endpoint may still be collected as a secondary or exploratory objective in these smaller trials. We use the log-hazard ratio estimate $\hat{\theta}_{\text{ext}}$ with the corresponding standard error from the external study to define a normal density $f_{\text{ext}}(\text{data}|\theta)$ with which we update our prior $q_0(\theta)$ using Bayes' Theorem to get

$$q_1(\theta|\text{data}) \propto f_{\text{ext}}(\text{data}|\theta) q_0(\theta).$$

We can then update PoS in (2) by replacing $q_0(\theta)$ by $q_1(\theta\,|\text{data})$. Since the information is assumed to be external to the pivotal Phase 3 trial and hence does not affect the distribution of $\hat{\theta}_{\text{fin}}$ for fixed θ, the term $P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}})$ in (2) does not change. Bayes' Theorem can also be used sequentially (Carlin

and Louis, 2009), i.e., updates can be done over time as further external information becomes available.

Often, external data may not be used directly to update the prior $q_0(\theta)$, because they may have been obtained in a different indication, comparing different treatments, or using an alternative endpoint. More care and work needs then to be used in setting up $f_{\text{ext}}(\text{data} \mid \theta)$, for example, by purposedly decreasing the number of events d_{ext} or using models to assess what possible effects of the treatment on such alternative endpoints may mean for the endpoint used in Phase 3. See Wang et al. (2013) for examples. Weighting of the densities is an alternative, yet technically slightly more complex and less easily understood by clinicians.

2.5. *Updating PoS based on internal information*

In Section 2.4, we have outlined how to update PoS with information emerging outside of the pivotal Phase 3 trial we are interested to compute PoS for. Such an external update leaves the term $P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}})$ in (2) unaffected. If instead an interim analysis for the pivotal Phase 3 trial itself is performed, this part of the PoS formula changes as well. In this work we assume only a single interim analysis was planned and we seek to update PoS based on the study continuing after that interim analysis. The work presented can readily be extended to a situation where an interim analysis was passed but more than one analysis is still planned. Regarding the information that may be available after an interim analysis, we consider two scenarios.

First, suppose the interim estimate $\hat{\theta}_{\text{int}}$ is known, e.g., members of a DSMB may, as part of a planned interim analysis, be provided with a point estimate $\hat{\theta}_{\text{int}}$ of θ and evaluate $P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{int}} = \theta_{\text{int}})$ for various values of θ as part of a futility assessment, see for example Proschan et al. (2006). Alternatively, assuming a (posterior) density q_1 for θ has been prespecified, the DSMB can evaluate the Bayesian predictive power according to (3) below. If $\hat{\theta}_{\text{int}}$ is known, the PoS formula (2) becomes

$$\text{PoS} := \int P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{int}} = \theta_{\text{int}}) q_1(\theta) d\theta. \tag{3}$$

We update the first factor by conditioning on the interim result, thus becoming $\text{CP}_\theta := P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{int}} = \theta_{\text{int}})$. The latter term is known as *conditional power* in the literature, see, e.g., Proschan et al. (2006). Assuming a distribution q_1 for θ and computing the expectation of $P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}} | \hat{\theta}_{\text{int}} = \theta_{\text{int}})$ with respect to that distribution in (3) is known as *Bayesian predictive power*. The concept has been introduced by Spiegelhalter et al. (1986) as a tool to plan and monitor clinical trials, see as well Proschan et al. (2006, Chapter 3). Computation of CP_θ is not straightforward, due to the fact that $\hat{\theta}_{\text{int}}$ and $\hat{\theta}_{\text{fin}}$ are correlated. See Lemma A.1 in the appendix for details.

Second, the sponsor may ask whether the PoS has changed if a DSMB recommends to continue a trial after a preplanned interim analysis for futility and/or efficacy. Such an interim analysis would require prespecified efficacy and futility boundaries θ_{eff} and/or θ_{fut}, i.e., the DSMB is advised to recommend to continue the trial if $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$. Otherwise, either early stopping for futility (if $\hat{\theta}_{\text{int}} \geq \theta_{\text{fut}}$) or efficacy (if $\hat{\theta}_{\text{int}} \leq \theta_{\text{eff}}$) may be recommended. In other words, if the trial is recommended to be continued, the sponsor may conclude that $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$. We set aside here other, possibly disease area-specific considerations, that may lead to a DSMB not following such decision guidance exactly. Choose $\theta_{\text{fut}} = \infty$ or $\theta_{\text{eff}} = -\infty$ for interim analyses allowing stopping for efficacy or futility only.

In order to be able to compute PoS based on interval information about $\hat{\theta}_{\text{int}}$, we make use of the joint bivariate normal distribution of $(\hat{\theta}_{\text{int}}, \hat{\theta}_{\text{fin}})$ given by

$$\begin{pmatrix} \hat{\theta}_{\text{int}} \\ \hat{\theta}_{\text{fin}} \end{pmatrix} \sim N\left(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 4/d_{\text{int}} & 4/d_{\text{fin}} \\ 4/d_{\text{fin}} & 4/d_{\text{fin}} \end{pmatrix} \right), \tag{4}$$

see, e.g., Jennison and Turnbull (2000) for details. Using this result, together with the definition of conditional probability, we obtain

$$\begin{aligned} \text{CP}_\theta &= P_\theta(\hat{\theta}_{\text{fin}} \le \theta_{\text{suc}} | \hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]) \\ &= P_\theta(\hat{\theta}_{\text{fin}} \le \theta_{\text{suc}}, \hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]) / P_\theta(\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]) \\ &= \frac{P_\theta\left(\theta_{\text{eff}} \le \hat{\theta}_{\text{int}} \le \theta_{\text{fut}}, -\infty \le \hat{\theta}_{\text{fin}} \le \theta_{\text{suc}}\right)}{\Phi\left((\theta_{\text{fut}} - \theta)/\sqrt{4/d_{\text{int}}}\right) - \Phi\left((\theta_{\text{eff}} - \theta)/\sqrt{4/d_{\text{int}}}\right)}. \end{aligned} \tag{5}$$

Note that CP_θ is a function of the true underlying log-hazard ratio θ. For a given θ, the right-hand side of (5) can be explicitly computed, for example, using the function mvtnorm in the R (R Core Team, 2013) package of the same name (Genz and Bretz, 2009; Genz et al., 2013).

In addition to the likelihood of success now becoming a conditional probability, also the prior $q_0(\theta)$ needs to be updated using that knowledge. Using Bayes' theorem we get

$$q_1(\theta) \propto \left(\Phi\left(\frac{\theta_{\text{fut}} - \theta}{\sqrt{4/d_{\text{int}}}} \right) - \Phi\left(\frac{\theta_{\text{eff}} - \theta}{\sqrt{4/d_{\text{int}}}} \right) \right) q_0(\theta). \tag{6}$$

Again, $q_1(\theta)$ is a function of θ. Combining (5) and (6), we can now (numerically) compute PoS according to (3) if a planned interim analysis has led to continuation of the trial.

2.6. *Prior distribution*

It remains to quantify the prior knowledge on the underlying effect θ. In general, the prior should be agreed upon before any updates of PoS are made and any (internal or external) information becomes available. The prior should reflect the overall current knowledge regarding the drug effect. Sensitivity assessments to illustrate how the choice of the prior impacts PoS should be performed and discussed with clinical colleagues.

Whenever a Phase 2 study is completed before the pivotal Phase 3 is started, it seems sensible to use the knowledge learned in Phase 2 to inform the Phase 3 trial. A straightforward way to incorporate such prior knowledge is by assuming $q_0(\theta)$ to be a normal density with mean $\hat{\theta}_{\text{Ph2}}$ and variance $4/d_{\text{Ph2}}$. This choice carries the advantage that we can update $q_0(\theta)$ in a conjugate normal Bayesian model. However, in our example in Section 5 we will also illustrate an alternative prior that might more appropriately match clinician's expectations about the true effect.

Note that a typical "noninformative" or vague prior, e.g., a normal centered at some log-hazard ratio and a variance corresponding to a small number of events, may in this setup not be a sensible choice. It would put approximately equal weight on hazard ratios above and below 1, which may not be realistic.

A rich literature discusses how to transform historical information into prior distributions, see e.g. Neuenschwander et al. (2010). Alternatively, methodology to derive confidence distributions based on meta-analyses of historic information may be considered to derive prior distributions (Xie et al., 2011).

3. A hypothetical development program

Assume that based on epidemiologic data patients with a higher value for some prognostic biomarker had lower risk compared to patients with low values. This leads to the hypothesis that an

intervention with a molecule increasing the biomarker may result in a decrease of a given risk beyond what may be achieved with existing drugs.

The pivotal study intended to lead to approval is assumed to be powered at 90% to detect a 15% risk reduction at a two-sided 5% significance level. It is planned to randomize patients 1:1 to experimental treatment or placebo; primary endpoint is the time from randomization to a defined event. The study would stop latest when 1600 patients experienced an event. We assume one planned interim analysis after 800 events is specified and that observed p values below 0.001 (interim) and 0.049 (final analysis) would lead to the conclusion of superiority.

Assume two smaller substudies are performed in parallel to the pivotal trial, assessing related questions. The first (Sub1) randomizes 400 and the second 150 patients (Sub2) 1:1 to experimental treatment or placebo. Three milestones lead to information becoming available during the course of the pivotal study prior to its termination. In Sub1, seven patients (5 on placebo and 2 on experimental) experience an event. Second, later, the results of Sub2 show that six patients (five on placebo and one on experimental) report an event. The third important milestone is the first interim analysis in the pivotal trial itself. We assume that the DSMB recommends to continue the trial. Albeit limited, information regarding the efficacy of the new treatment is obtained at each of these milestones. We will illustrate how this could change our belief regarding the drug effect and how this in turn may impact the likelihood for the pivotal study to ultimately be successful.

4. Possible prior distributions

4.1. *Normal prior based on literature data*

Assume that a random effects meta-analysis summarizing early phase data and data from related molecules leads to an overall hazard ratio estimate of 0.85, reflecting the 15% risk reduction assumed for the pivotal trial. The 95% confidence interval for the pooled effect estimate is assumed to be [0.83, 0.87], whereas the pooled standard deviation of the effect estimates is $\sigma_{\text{prior}} = 0.11$. We thus assume a normal prior centered at $\theta_{\text{prior}} = \log(0.85)$ with standard deviation $\sigma_{\text{prior}} = 0.11$. The resulting density function is depicted as red line in Fig. 1 (left).

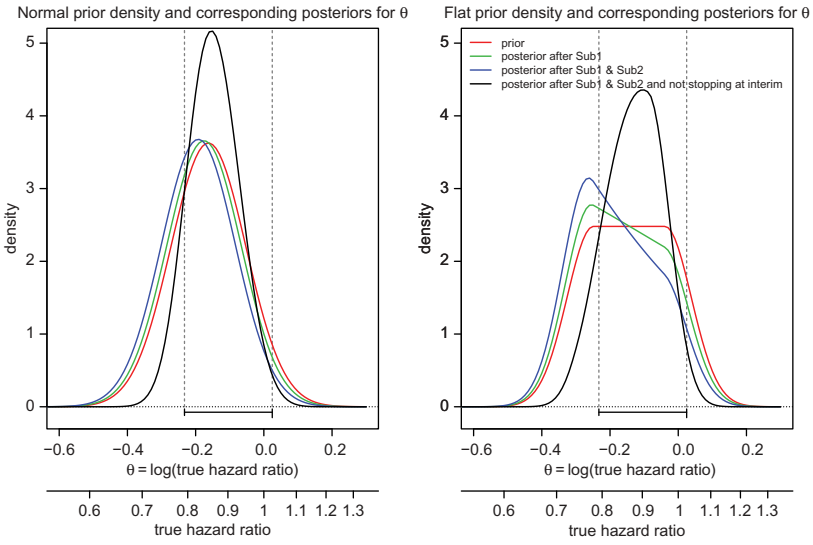

Figure 1. Prior densities for θ and posterior densities after update with Sub1, Sub1 & Sub2, and after not stopping at the interim (sequentially updated after the Sub1 & Sub2 update). The interval below the x-axis indicates the interval $[\theta_{\text{eff}}, \theta_{\text{fut}}]$.

Table 1. PoS values after each update, as well as for $\hat{\theta}_{\text{int}}$ equal to the efficacy and futility boundary, for both priors. Results in the first three lines are based on prior distributions only, whereas the other results are based on posterior distributions reflecting internal or external information that had become available.

	Normal prior	Flat prior
Probability for hazard ratio to be ≤ 0.7	0.039	0.039
Probability for hazard ratio to be ≥ 1	0.070	0.147
PoS based on prior distribution	0.702	0.612
PoS after Sub1	0.740	0.665
PoS after Sub1 and Sub2	0.783	0.727
PoS after not stopping at interim, assuming $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$	0.705	0.617
PoS after not stopping at interim, assuming $\hat{\theta}_{\text{int}} \in [-\infty, \theta_{\text{fut}}]$	0.822	0.782
PoS after not stopping at interim, assuming $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \infty]$	0.653	0.547
PoS after not stopping at interim, assuming $\hat{\theta}_{\text{int}} = \theta_{\text{eff}}$	0.997	0.997
PoS after not stopping at interim, assuming $\hat{\theta}_{\text{int}} = \theta_{\text{fut}}$	0.024	0.016

The choice of this prior entails that the probability for the hazard ratio to be above 1 or below 0.7 amounts to 0.07 and 0.04, respectively. Risk reductions of 30% or more may indeed be unlikely considering that in many indications, new treatments are on top of an already effective therapy. On the other hand, the prior allows for a 7% probability that the treatment effect may be on the harmful side. In the next section, we suggest a more pessimistic prior.

4.2. *More pessimistic prior*

A more pessimistic alternative prior is potentially justified since the biomarker might only be prognostic and there remains uncertainty whether changing it via our intervention indeed translates into a risk reduction. We also recommend to consider alternative priors as part of a sensitivity assessment. Basically retaining the lower tail of the prior discussed in Section 4.1, we move the upper tail slightly to the right allowing for a plateau in the center of the distribution and leading to a 15% probability for the true hazard ratio to exceed 1. The red line in Fig. 1 (right) illustrates this alternative prior and Table 1 provides the probabilities for the hazard ratio to be below 0.7 and above 1 for this prior. In what follows, we will illustrate all computations using the normal and this more pessimistic prior. Based on these priors, evaluation of (2) leads to the prior PoS also shown in Table 1.

5. Results

As outlined in Section 3, an interim analysis after 800 events had been planned, using $\alpha_{\text{int}} = 0.001$ for the interim and $\alpha_{\text{fin}} = 0.049$ for the final analysis. From this, the logarithm of the minimal detectable hazard ratio at the final analysis, based on a two-sided test, can be derived from (1) as $\theta_{\text{suc}} = \log(0.906)$. Applying (1) to the corresponding quantities $\alpha_{\text{int}} = 0.001$ and $d_{\text{int}} = 800$ associated with the interim analysis, we obtain $\theta_{\text{eff}} = \log(0.792)$ as the efficacy boundary at the interim analysis.

In addition to the interim efficacy boundary, assume that a futility analysis was prespecified in the DSMB charter, with interim futility boundary of $\theta_{\text{fut}} = \log(1.025)$. Thus, if the trial is not stopped at the interim analysis, we may assume for the log-hazard ratio estimate at the interim analysis that $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}] = [\log(0.792), \log(1.025)]$.

In what follows we illustrate how the prior PoS will change as information becomes sequentially available. Assume that to compute $f_{\text{ext}}(\text{data}|\theta)$, a Cox proportional hazards regression model was fit to the seven events in Sub1, leading to $\hat{\theta}_{\text{ext}} = \log(0.40)$ with 95% confidence interval ranging from log(0.08) to log(2.04). Formula (3) then allows derivation of the posterior distribution reflecting the additional information inherent to the Sub1 data, which is shown as the green line in Fig. 1. From Table 1, PoS increases slightly in both cases, but slightly more for the pessimistic prior.

Further assume that upon availability, Sub2 data were analyzed together with Sub1 data using a Cox model stratified by study. The resulting log-hazard ratio is assumed to be $\hat{\theta}_{\text{ext}} = \log(0.29)$ with 95% confidence interval extending from log(0.08) to log(1.05), leading to another small increase in PoS, see Table 1. This illustrates that the two assumed external updates based on Sub1 alone and later based on Sub1 and Sub2 together alter the normal prior very slightly only. This is because the normal prior itself already carries substantial information. For the more pessimistic prior, the data change the prior more relevantly, resulting in a higher updated PoS. Overall, under both priors the change in PoS is below or just above 10%, which we typically consider the minimal change required to communicate a modification to PoS.

As postulated, based on assessing the interim analysis results, the DSMB recommends to the Sponsor to continue the pivotal trial, leading to the assumption $\hat{\theta}_{\text{int}} \in [\log(0.792), \log(1.025)]$. Using the most recent posterior obtained after accounting for the information provided by both, Sub1 and Sub2, as new prior distribution, the posterior distribution resulting from this interval information is derived according to (6). However, as this information is based on the trial itself, the PoS changes from $P_\theta(\hat{\theta}_{\text{fin}} \leq \theta_{\text{suc}})$ to the conditional probability given in (5). Table 1 displays the results for PoS. Based on both priors, the updated PoS decreases, more so for the pessimistic prior. The main reason is that the posterior assigns little weight—relative to the prior before the update—to log-hazard ratios outside $[\theta_{\text{eff}}, \theta_{\text{fut}}]$ (black curves in Fig. 1). This implies even more pronounced effects when continuing after an interim analysis for efficacy only, i.e., when we know after the interim that $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \infty]$. How large the drop in PoS is can be seen for both priors in Table 1. Similarly, if we do not stop a trial after an interim for futility only, we get an increase in PoS.

Finally, we would like to emphasize the sensitivity of PoS to the prior assumptions. Assuming we do not stop the pivotal trial after an interim recommendation provided by a DSMB, we only know that $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$. We strongly advise to accompany this with a sensitivity analysis by looking at different values of $\hat{\theta}_{\text{int}}$ that a DSMB may have seen. The two most extreme cases certainly are either $\hat{\theta}_{\text{int}} = \theta_{\text{eff}}$ or $\hat{\theta}_{\text{int}} = \theta_{\text{fut}}$. In our example, these extreme values for $\hat{\theta}_{\text{int}}$ yield equally extreme PoS values as shown in Table 1. Fig. 2 shows how the posterior distributions would look like under these assumptions. Figure 3 depicts updated conditional powers for our example.

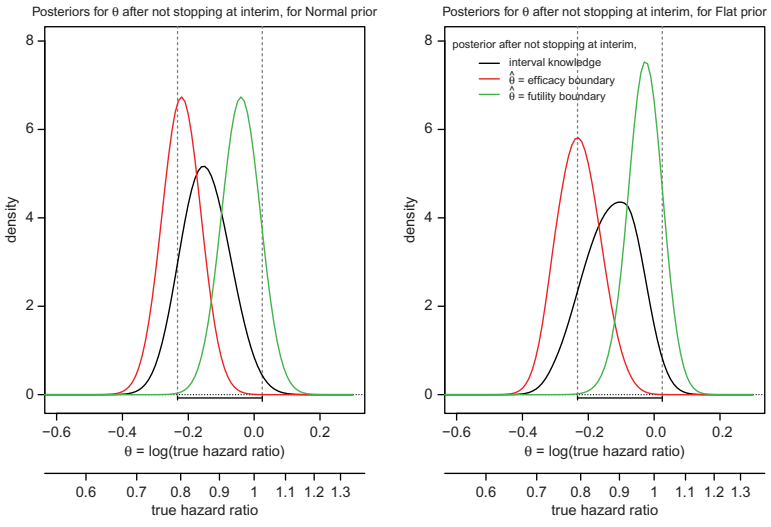

Figure 2. Posterior densities comparing "interval knowledge" as well as assuming that $\hat{\theta}_{\text{int}}$ is either equal to θ_{eff} or θ_{fut}.

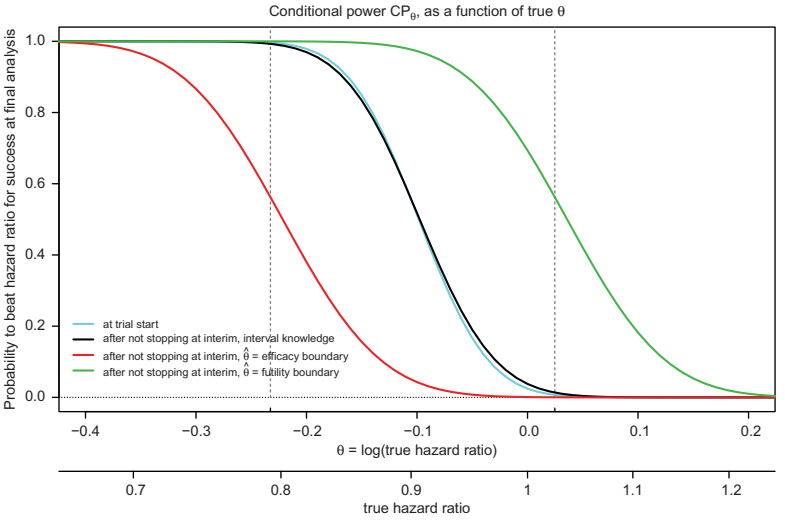

Figure 3. Conditional power at study start and after not stopping at the interim. Vertical lines indicate the efficacy and futility boundary.

6. Discussion

In this article, we discussed methodology how the PoS of a large and long Phase 3 pivotal study can be reassessed as new information becomes available. While this may not typically lead to any changes a Sponsor would make to the ongoing study, it could impact the decision on whether or not further pivotal trials may be started at risk. Additional activities related to a potential launch may be front-loaded as well based on a PoS assessment, e.g., securing salesforce resources.

Information that becomes available can be the results from other studies conducted in parallel, but also be based on continuing the Phase 3 pivotal trial after a planned interim analysis.

Assuming we do not stop the pivotal trial after an interim recommendation provided by a DSMB, we only know that $\hat{\theta}_{\text{int}} \in [\theta_{\text{eff}}, \theta_{\text{fut}}]$. Updating PoS with precisely this knowledge seems sensible and we provide the methodology to do so. We illustrate that the PoS typically goes down after not stopping at an interim for efficacy only, and goes up after an interim for futility only. We recommend to perform sensitivity analyses assuming different scenarios that the DSMB might have seen, as we do in Section 5.

For simplicity, we have restricted our attention to assessing a hazard ratio in a 1:1 randomized trial. This implies that we assume the variance of the estimated log-hazard ratio to be determined by the number of events, i.e., known. Extension to unbalanced randomization is readily possible and the conceptual ideas also apply to binary or normally distributed endpoints. Further modifications to the prior to better align it with clinician's expectations can be considered as well.

Success here was, for simplicity, defined as statistical significance at the final analysis. More general, in a study with k interim analyses the approach can easily be adapted. If $\ell < k$ interim analyses have been passed already, the likelihood of the study being statistically significant has to be expressed as the likelihood of reaching statistical significance at any of the remaining $k - \ell$ planned analyses conditional on having not stopped the study at the ℓ earlier interim analyses. As the joint distribution of the estimated log-hazard ratios follows a multivariate normal distribution (as shown in (4) for $k = 2$), these conditional probabilities can be readily derived.

7. Software

R code for all the methodology described in this paper is available upon request from the authors. This document was created using Sweave (Leisch, 2002), LaTEX (Knuth, 1984; Lamport, 1994), and R (R Core Team, 2013). This means that all of the code has been checked by R.

Acknowledgments

We thank Anders Svensson and Hans-Ulrich Burger for valuable comments on an earlier version the manuscript. We also gratefully acknowledge thoughtful comments by two reviewers and the editor.

References

Carlin, B., Louis, T. (2009). *Bayesian Methods for Data Analysis*. Boca Raton, FL: Chapman & Hall.
Genz, A., Bretz, F. (2009). *Computation of Multivariate Normal and t-Probabilities*. Lecture Notes in Statistics. Heidelberg: Springer-Verlag.
Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., Hothorn, T. (2013). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9–9996. http://CRAN.R-project.org/package=mvtnorm
Jennison, C., Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Boca Raton, FL: Chapman & Hall/CRC.
Jones, D., Whitehead, J. (1979). Sequential forms of the log rank and modified Wilcoxon tests for censored data. *Biometrika* 66:105–113.
Knuth, D. E. (1984). *The TEXbook, of Computers and Typesetting* Vol. A. Reading, MA: Addison-Wesley.
Lamport, L. (1994). *LATEX: A Document Preparation System*. 2nd ed. Reading, MA: Addison-Wesley.
Leisch, F. (2002). Dynamic generation of statistical reports using literate data analysis. In *COMPSTAT 2002 – Proceedings in Computational Statistics*, Härdle, W., Rönz, B. (eds.). Heidelberg: Physica Verlag.
Neuenschwander, B., Capkun-Niggli, G., Branson, M., Spiegelhalter, D. J. (2010). Summarizing historical information on controls in clinical trials. *Clinical Trials* 7:5–18.
Proschan, M., Lan, K., Wittes, J. (2006). *Statistical Monitoring of Clinical Trials: A Unified Approach*. New York, NY: Springer.
R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/
Spiegelhalter, D., Reedman, L., Blackburn, P. (1986). Monitoring clinical trials—Conditional power or predictive power. *Controlled Clinical Trials* 7:8–17.
Wang, Y., Fu, H., Kulkarni, P., Kaiser, C. (2013). Evaluating and utilizing probability of study success in clinical development. *Clinical Trials* 10(3):407–413.
Xie, M., Singh, K., Strawderman, W. (2011). Confidence distributions and a unifying framework for meta-analysis. *Journal of the American Statistical Association* 106:320–333.
Zhang, J., Zhang, J. J. (2013). Joint probability of statistical success of multiple Phase III trials. *Pharmaceutical Statistics* 12:358–365.

Appendix A. Computations

A.1. Computation of PoS if $\hat{\theta}_{\text{int}}$ is exactly known

In Table 1 and Figure 3, we provide PoS assuming that we knew the value of $\hat{\theta}_{\text{int}}$. To compute PoS in these situations, we need formulas for the conditional power functions and for completeness of this article, these are provided in Lemma A.1. The formulas are available in Proschan et al. (2006). However, we prove Lemma A.1 not via the theory of "*B*-values," but directly using the independent increments property of the logrank statistic and the bivariate normal distribution (4).

Lemma A.1 (PoS when knowing $\hat{\theta}_{\text{int}}$). *Assume we observe at an interim analysis the log-hazard ratio* θ_{int}. *The conditional power as a function of* θ *then amounts to*

$$CP_\theta = P(\hat{\theta}_{fin} \leq \theta_{suc} \mid \hat{\theta}_{int} = \theta_{int}) = \Phi\left(\frac{d_{fin}\theta_{suc} - d_{int}\theta_{int} - d_2\theta}{2\sqrt{d_{fin}}}\right)$$

for $d_2 := d_{fin} - d_{int}$.

A.1.0.1. Proof. If $\hat{\theta}$ is an estimate of a log-hazard ratio based on d events, then $\ell(\hat{\theta}, d) := d \cdot \hat{\theta}/4$ is the corresponding logrank test statistic, see, e.g., Jennison and Turnbull (2000). For this statistic we know that (Jones and Whitehead, 1979)

$$\ell(\hat{\theta}_{\text{fin}}, d_{\text{fin}}) - \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) \quad \text{and} \quad \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) \quad \text{are independent and,} \tag{A1}$$

$$\ell(\hat{\theta}_{\text{fin}}, d_{\text{fin}}) - \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) \sim N(\theta d_2/4, d_2/4). \tag{A2}$$

We can thus compute

$$\begin{aligned} P(\hat{\theta}_{\text{fin}} \le \theta_{\text{suc}} | \hat{\theta}_{\text{int}} = \theta_{\text{int}}) &= P\Big(\ell(\hat{\theta}_{\text{fin}}, d_{\text{fin}}) \le \theta_{\text{suc}} d_{\text{fin}}/4 | \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) = \ell(\theta_{\text{int}}, d_{\text{int}})\Big) \\ &= P\Big(\ell(\hat{\theta}_{\text{fin}}, d_{\text{fin}}) - \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) \le \\ &\qquad \theta_{\text{suc}} d_{\text{fin}}/4 - \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) | \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) = \ell(\theta_{\text{int}}, d_{\text{int}})\Big) \\ &= P\Big(\ell(\hat{\theta}_{\text{fin}}, d_{\text{fin}}) - \ell(\hat{\theta}_{\text{int}}, d_{\text{int}}) \le \theta_{\text{suc}} d_{\text{fin}}/4 - \ell(\theta_{\text{int}}, d_{\text{int}})\Big) \quad \text{by (A1).} \end{aligned}$$

Using (A2) and rearranging terms we get the result. □