MAIN PAPER

Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.1764

Published online 21 July 2016 in Wiley Online Library

Bayesian predictive power: choice of prior and some recommendations for its use as probability of success in drug development

Kaspar Rufibach,* Hans Ulrich Burger, and Markus Abt

Bayesian predictive power, the expectation of the power function with respect to a prior distribution for the true underlying effect size, is routinely used in drug development to quantify the probability of success of a clinical trial. Choosing the prior is crucial for the properties and interpretability of Bayesian predictive power. We review recommendations on the choice of prior for Bayesian predictive power and explore its features as a function of the prior. The density of power values induced by a given prior is derived analytically and its shape characterized. We find that for a typical clinical trial scenario, this density has a *u*-shape very similar, but not equal, to a β-distribution. Alternative priors are discussed, and practical recommendations to assess the sensitivity of Bayesian predictive power to its input parameters are provided.

Keywords: Bayesian predictive power; conditional power; prior distribution; probability of technical success

1. INTRODUCTION

The sample size of a pivotal phase 3 superiority trial is usually determined such that for an assumed clinically relevant effect in the primary endpoint, the probability of a type II error is not larger than a given threshold β when using a test that has significance level α.

Adequate sample size and unconditional power are essential for a well-designed trial, because they tell us whether the trial is likely to yield useful and interpretable results [1]. Once results of an interim analysis (IA) become available, unconditional power calculations can be updated by conditioning on the data observed up to the IA, for example, to decide whether to stop a trial for futility. Such *conditional power* (CP) computations remain dependent on an assumed true treatment effect that is hypothesized to generate the data beyond the IA. Noting that simply conditioning on selected hypotheses that are typically specified at the start of the trial, [2] argue that this ignores the knowledge about the treatment effect that has accumulated by the time of IA. They suggest extending the conditional to a predictive power analysis by averaging the CP function with respect to the current knowledge or opinion about the treatment effect. The current knowledge about the underlying treatment effect θ is thereby summarized using a distribution p for θ. Averaging over p avoids having to assume a specific treatment effect as needed for CP. Because a distribution on the parameter of interest is assumed, [2] call the averaged CP *Bayesian predictive power* (BPP), although Bayes's theorem is not necessarily involved. BPP exhibits frequentist and Bayesian aspects and is thus typically considered a 'hybrid classical Bayesian' approach [3]. Alternative names for BPP in the literature are *assurance* [3] or *probability of study success* (PoS) [4], [5]. In what follows, we will use PoS. For a discussion of the different power concepts, with an emphasis on predictive power, we refer to [6].

For ethical and economical reasons, the importance of good decision-making in drug development can hardly be overestimated. We refer to [7] and [8] for identification of drivers in decision-making and quantification of their impact. Unconditional and CP as well as PoS provide a quantitative framework to base decisions on, typically, whether to start a trial at all or whether to stop a trial for futility.

Unconditional power is typically used in the planning phase of a trial to determine sample size. PoS can, besides supporting a potential futility stopping decision, be used to support decisions on funding of additional studies, calculation of project valuations, development of budgets and hiring plans (e.g., for salesforce), or planning of manufacturing capacity.

After introducing details about PoS in Section 2, we will review recommendations in the literature on how to choose a prior when computing PoS in Section 3. Section 4 introduces a generic example that will be used to illustrate our points. Section 5 classifies the shapes of the density of power values and discusses some features of it. We extend this discussion to non-normal priors in the sections to follow. A sensitivity interval for PoS is proposed in Section 8. The paper is concluded with a discussion and some recommendations in Section 9.

We will provide a few mathematical results, the proofs of which are deferred to the Supplementary Materials.

Department of Biostatistics, Hoffmann-La Roche Ltd, Basel, Switzerland

**Correspondence to: Kaspar Rufibach, Department of Biostatistics, Hoffmann-La Roche Ltd, Basel, Switzerland.*
E-mail: kaspar.rufibach@roche.com

2. SETUP

To be specific, we assume that for prespecified probabilities of types I and II errors, the pivotal phase 3 trial was planned assuming a certain alternative effect size and that the variance σ^2_{fin} of the estimated effect size at the final analysis is known. We denote by α the nominal significance level at the final analysis at which we aim to reject the null hypothesis $\theta = 0$ if a suitable two-sided test is significant at α. If the corresponding test statistic is (approximately) normal, we call the critical value θ_{suc} of this test the *minimal detectable effect size* and compute it as follows:

$$\theta_{\text{suc}} = \sigma_{\text{fin}} \cdot z_{\alpha/2},$$

wherein $z_{\alpha/2}$ is the $\alpha/2$-quantile of the standard normal distribution. This definition implies that we use the estimated effect size as our test statistic. The quantity θ_{suc} can be interpreted as the effect size that gives a two-sided p-value of α at the final analysis of the phase 3 trial. The probability of success at the final analysis, for an assumed underlying effect θ, is then given by the power function:

$$T(\theta_{\text{suc}}|\theta) := P(\hat{\theta}_{\text{fin}} \leqslant \theta_{\text{suc}}|\theta) = \Phi\left(\frac{\theta_{\text{suc}} - \theta}{\sigma_{\text{fin}}}\right). \quad (1)$$

Here, $\hat{\theta}_{\text{fin}}$ is the estimate of the treatment effect θ at the final analysis and follows a normal distribution with mean θ and variance σ^2_{fin}. In Section 4, we will make this setup specific to a time-to-event endpoint, implying that $\hat{\theta}_{\text{fin}}$ will be the estimate of a log hazard ratio. That is why in (1), we are actually interested in the probability of the event $\{\hat{\theta}_{\text{fin}} \leqslant \theta_{\text{suc}}\}$, which – with the $\leqslant$ sign – might seem unfamiliar for a power function.

Because θ is not known, PoS is defined as the average over the prior p of the power function T:

$$\text{PoS} = \int T(\theta_{\text{suc}}|\theta)\, p(\theta) d\theta. \quad (2)$$

As discussed in [3], PoS can be defined setting the argument of T equal to any meaningful trial outcome, not just the minimal detectable effect size θ_{suc} as in (2).

If the prior p is normal with mean θ_0 and variance σ^2_0, PoS can be explicitly computed as follows:

$$\text{PoS} = \Phi\left(\frac{\theta_{\text{suc}} - \theta_0}{\sqrt{\sigma^2_{\text{fin}} + \sigma^2_0}}\right). \quad (3)$$

For non-normal priors, numerical integration can be used to compute PoS via its definition (2).

Comparing (1) with (3), one finds that power is equal to 0.5 for $\theta = \theta_{\text{suc}}$ and PoS = 0.5 for a normal prior centered at θ_{suc}. For any prior with mean $\theta_0 < \theta_{\text{suc}}$, PoS is always smaller than the power at $\theta = \theta_0$. For $\theta_0 > \theta_{\text{suc}}$, the order is reversed. In summary, if the prior is normal, PoS considered as a function of the prior mean θ_0 is always closer to 0.5 than the power at θ_0 (Figure 1). In Lemma A.1 in the Supplementary Material, we show that this statement holds even for any unimodal symmetric prior with finite variance. This is in line with the discussion in [3] that 'The assurance figure is often much lower [than the power], because there is an appreciable prior probability that the true treatment difference is less than δ^*.', where in their notation, δ^* is the assumed alternative in sample size planning. Note that because the authors in [3] consider normally distributed endpoints, their power function corresponding to (1) is defined for effects greater or equal θ_{suc}, instead of smaller or equal in our case.

3. CHOICE OF THE PRIOR *P* FOR PROBABILITY OF STUDY SUCCESS

The purpose of PoS is to compute it taking into account available information about the true underlying effect θ. 'Information' here can be understood in a broad sense and may include data external to our pivotal trial, for example, expert beliefs if suitably quantifiable or modeling of an effect estimate based on early phase clinical trials on a different endpoint using surrogacy models; see [3, Example 4] for an example. We assume that such information is quantified in the prior p.

In the derivations that follow, we focus on the computation of PoS as in (3) at the trial start. The density p can be either a prior based on, for example, data from early phase studies or already a posterior received from updating a prior with data that provided information on θ. Our analysis later does only depend on the shape of p and not on how it was conceptually constructed, and we will therefore always refer to it as 'prior'. The authors in [9] recommend to use priors based on pilot studies. Discussing PoS in the context of sample size computation, [3] point out that '…the synthesis of the available evidence for the treatment effect (and any other required parameters) should be regarded as an important contribution to the design of a clinical trial and this evidence should be incorporated into sample size determination.' They emphasize that sufficient effort should be spent to elicit a suitable prior, based on best available clinical and scientific judgments. As stated in [10], there is no consensus in the literature on how to choose a prior when computing PoS. Because PoS is often applied in early drug development, the authors in [10] suggest that the prior often be flat and as minimally informative as possible. However, they conclude on a cautious note mentioning that '...unless the user of these statistics has a deep understanding of their characteristics important pitfalls may be encountered, especially with the use of predictive power.' In summary, the existing literature concentrates on 'non-informative' priors, and insufficient attention is paid to the relationship between the chosen prior and computed success probabilities. We will explore the sensitivity of PoS to the prior assumption in Sections 5–7. For mean difference as an endpoint and normal priors in each group, [10] provide a recommendation on how to choose the parameters of

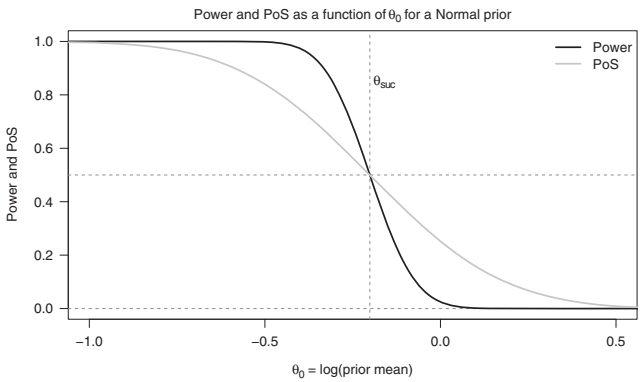

Figure 1. Power and probability of study success (PoS) for a normal prior. The input quantities are those introduced in Section 4.

the priors, depending on the trial's objective, development phase, and patient population.

The connections between predictive probabilities, CP, and PoS are discussed in [10], together with conditions under which these quantities actually coincide.

A comparison of the features of PoS and CP is undertaken in [6]. The authors emphasize that PoS takes into account the uncertainty about the underlying true effect, parts of which are ignored by standard sample size calculations and CP.

In what follows, by defining PoS via (2), we explore properties of PoS when looking at it prior to the start of the pivotal trial.

4. GENERIC EXAMPLE

In what follows, we will illustrate our ideas using a generic example. We assume that the pivotal phase 3 trial has a time-to-event primary endpoint where the final analysis is performed after $d_{\text{fin}} = 380$ events, testing a null hypothesis of a hazard ratio equal to 1 using a two-sided significance level of $\alpha = 0.05$. The chosen number of events implies that the test has 80% power to detect an alternative hazard ratio of 0.75. Additionally, supposing that patients were randomized in a 1:1 ratio, we can assume the standard error of the hazard ratio estimate at the final analysis to be known and compute it as $\sigma_{\text{fin}} = \sqrt{4/380} = 0.103$. The distribution of this estimate at the final analysis can be well approximated by a normal distribution (see, e.g., [11]), and the minimal detectable log hazard ratio is thus $\theta_{\text{suc}} = \sigma_{\text{fin}} \cdot z_{0.05/2} = \log(0.818)$. We assume that we have prior knowledge, for example, from a phase 2 study, with observed hazard ratio $\theta_0 = \log(0.7)$, which we consider to be worth 50 events. With these numbers, we get PoS $= 0.697$. The prior is depicted in Figure 2 (left).

5. DENSITY OF THE POWER IF THE PRIOR IS NORMAL

From (2), we see that in fact, by the law of large numbers, PoS is simply the average of power values $T(\theta_{\text{suc}}|\theta)$ when θ is drawn from a random variable with density p. As an illustration, based on the assumptions in Section 4, Figure 2 (right) shows a histogram of power values of a sample of size $M = 10^6$ with θ's drawn from the prior p.

The resulting PoS value is indicated with the dot. The probability to see very low or very high power values is quite high. The shape of the histogram in Figure 2 can be explained by looking at the prior on the left: The normal prior is centered at $\theta_0 = \log(0.7)$, and the variance is somewhat large, so that the probability assigned to very large beneficial effects (hazard ratio below 0.5) or very large detrimental effects (hazard ratio above 1) is not small. For such effect sizes, the value of the power will be almost 1 or almost 0, respectively, which is reflected in the histogram on the right.

Figure 2 raises three questions: First, is the distribution estimated by the histogram on the right in Figure 2 a β-distribution? Lemma 5.1 later shows that the answer to this question is no, although for a wide range of parameters, the density derived in Lemma 5.1 can be very well approximated by a β-density.

Second, we summarize the distribution that is described by the histogram by just its mean called 'PoS'. How representative for this distribution is the PoS value? As an analogy, to summarize normal data, we generally use the mean because we consider it to be a useful summary of the *unimodal* normal distribution (leaving aside further optimality features of the mean). For PoS with a normal prior as in our generic example, the setup is somewhat opposite, and the distribution we summarize can be far from unimodal.

Third, can conditions on the prior distribution be imposed that ensure that the density estimated via the histogram in Figure 2 is unimodal, so that PoS may be considered a more representative summary statistic of the power value distribution? In order to answer this third question, we provide the analytical form of this density in Lemma 5.1. To this end, let the random variable Θ have the prior distribution p, assume θ_{suc} to be fixed, and define the random variable $Y := T(\theta_{\text{suc}}|\Theta)$.

Lemma 5.1
If $\Theta \sim N(\theta_0, \sigma_0^2)$, the density g of Y is, for $y \in (0,1)$, given by the following:

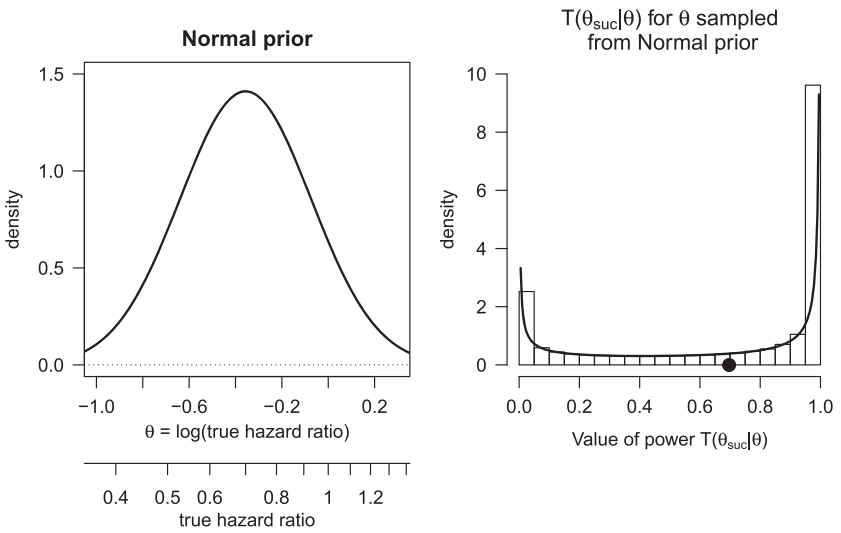

Figure 2. Normal prior and histogram of power values for θ sampled from the prior.

$$g(y) = \tau\phi\left(\psi - \tau\Phi^{-1}(y)\right)\left[\phi(\Phi^{-1}(y))\right]^{-1}, \tag{4}$$

$$= \tau \exp\left(-\frac{1}{2}\left(\psi - \tau\Phi^{-1}(y)\right)^2 + \frac{1}{2}\Phi^{-1}(y)^2\right) \quad y \in (0,1), \tag{5}$$

where ϕ, Φ, and Φ^{-1} are the standard normal PDF, CDF, and quantile function, $\tau := \sigma_{\text{fin}}/\sigma_0 > 0$, and $\psi := (\theta_{\text{suc}} - \theta_0)/\sigma_0$.

The density g is added as a solid line to the histogram in Figure 2. Note that once the prior and the critical value for success at the final analysis of the pivotal trial are specified, τ and ψ are determined. The values of τ and ψ for g in Figure 2 amount to 0.363 and 0.550, respectively. For illustration, Figure 3 provides shapes of g as a function of τ (left, for fixed ψ) and as a function of ψ (right, for fixed τ).

The plots reveal that, as a function of τ and ψ, there is a transition point where the density g is uniform and changes shape from u-shaped to unimodal. This is made rigorous in Theorem 5.2. Furthermore, g is symmetric around 0.5 if, and only if, $\psi = 0$; see Lemma B.1 in the Supplementary Material.

Theorem 5.2 (Qualitative features of g)
We have the following statements:

(1) If $\tau = 1$, then g is as follows:

$$\begin{cases} \text{strictly decreasing for} & \psi < 0, \\ \text{constant for} & \psi = 0, \\ \text{strictly increasing for} & \psi > 0 \end{cases}$$

on [0, 1]. Minima and maxima of g are accordingly either at 0 or 1.

(2) If $\tau \neq 1$ then g has a

$$\begin{cases} \text{minimum at } y_m \text{ if } & \tau < 1 \text{and} \\ \text{maximum at } y_m \text{ if } & \tau > 1, \end{cases}$$

for

$$y_m = \Phi\left(\frac{\tau\psi}{\tau^2 - 1}\right) = \Phi\left(\frac{\sigma_{\text{fin}}(\theta_{\text{suc}} - \theta_0)}{\sigma_{\text{fin}}^2 - \sigma_0^2}\right).$$

Furthermore, g

$$\begin{cases} \text{is decreasing for } y < y_m \text{ and increasing for } y > y_m \text{ if } & \tau < 1 \text{and} \\ \text{is increasing for } y < y_m \text{ and decreasing for } y > y_m \text{ if } & \tau > 1. \end{cases}$$

The interpretation as applied to PoS is that we get a u-shaped density g if $\sigma_{\text{fin}} < \sigma_0$, that is, the amount of information in the prior is less than what phase 3 will provide. Applied to our generic example of a time-to-event endpoint, that would imply that the number of events the prior information is worth, were smaller than the number of events at the final analysis in the phase 3 trial. If we wanted to use a normal prior and ensure that the power values follow a unimodal distribution, so that PoS is somewhat representative of the distribution it is supposed to summarize, as discussed previously, we would need to choose a prior such that $\sigma_{\text{fin}} > \sigma_0$, that is, a prior information that is worth more events than specified for the final analysis of the pivotal trial. Such a setup is rarely the case in clinical development, because one can argue whether a phase 3 trial is even necessary if one were rather sure about the effect prior to initiating a phase 3 trial.

Typically, increasing the variance of a normal prior is associated with 'non-' or 'minimal' informativeness. This is true for PoS because from (3), we see that PoS $\to 0.5$ if $\sigma_0 \to \infty$ for any fixed $\sigma_{\text{fin}}, \theta_{\text{suc}}, \theta_0$, and $y \in (0, 1)$. In this sense, a normal prior with large variance is indeed non-informative for PoS. Of note, as was illustrated in the case of a time-to-event endpoint, choosing a small prior number of events implies that the probability to see hazard ratios corresponding to large detrimental or beneficial effects becomes high, implying a u-shaped density g. In many trials with a time-to-event endpoint, it is questionable whether large beneficial effects are realistic at all, so assigning substantial probabilities to such hazard ratios might not reflect the true prior belief.

If the prior approaches a point mass at θ_0, PoS becomes either power or CP at θ_0. The latter result is noted in passing in [1, p. 62] as well as again in [10] and [6].

Based on these examples, we conclude that assuming a normal prior when computing PoS requires hard-to-meet criteria to be fulfilled in order to obtain a non-u-shaped density g of power values. Interestingly, as discussed in [10], [9] already found that for binary endpoints, uniform priors give too much weight on

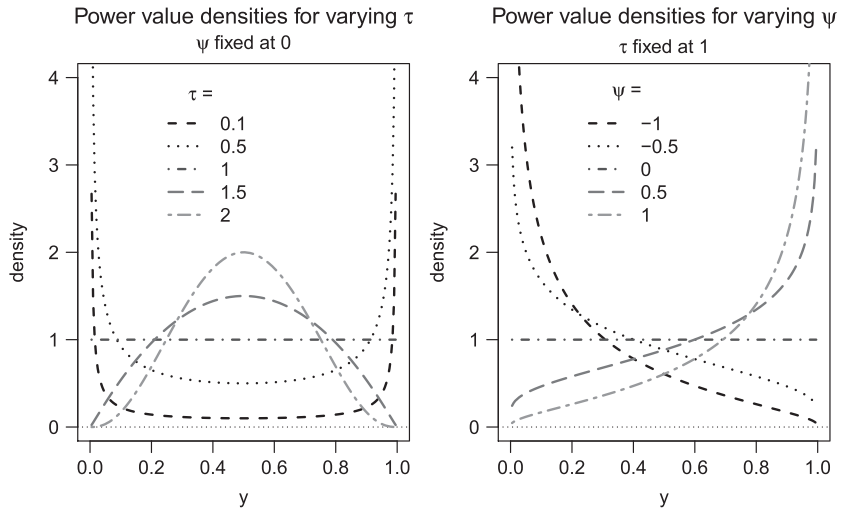

Figure 3. Density g as a function of $\tau = \sigma_{\text{fin}}/\sigma_0 > 0$ and $\psi = (\theta_{\text{suc}} - \theta_0)/\sigma_0$.

extreme cases. In passing, [10] mention that 'Same is true for Normal case'.

In what follows, we will broaden the scope, discuss PoS based on non-normal prior distributions, and explore whether more general setups are able to provide alternative properties of g and thus PoS.

An interesting connection of our finding of u-shaped densities g in the context of PoS to CP is discussed in [12]. These authors look at the density function of CP, that is, the update after an IA of the *a priori* calculated power in the planning phase. Because CP depends on the data up to the interim, this is a random quantity, and its distribution in their Equation (1) is constructed as that of the random variable Y introduced in Lemma 5.1. It thus gives rise to a density of the same structure; compare Equations (2) and (3) in [12] with Equation 4 here. In the simple normal model they consider, the corresponding density depends on the timepoint of the interim, the effect size μ_d the study has been powered for, and the effect size μ that generates the data after the interim. The latter has to be assumed, and [12] discuss the case of either taking for μ the effect size the study had been powered with or the estimated effect size at the interim. In the latter situation, they find that, irrespective of when the IA is performed, the density function of CP exhibits a u-shaped behavior as in our case of PoS. The differences between the setup in [12] and the one treated here can be summarized as follows:

- CP is the quantity that exhibits the u-shaped density, whereas PoS is the mean of a random variable with a u-shaped density, where the u-shape only appears for certain parameter configurations. As discussed in [12], this shape may be informative for the statistician when analyzing CP.
- Once a model for the data has been assumed, the properties of CP are determined by that model, that is, the shape of its density is directly induced by the model assumption. Somewhat more flexibly, for PoS, this shape is induced by the assumption on the prior. In Section 7, we discuss the implication of different prior shapes on the properties of PoS.

In ongoing research, we are exploring features of these densities if both effects are combined, that is, if PoS is updated at an IA, be it in an unblinded manner as analyzed in [12] or blinded as in [5], and the density of the power values exhibits a u-shape.

To conclude the discussion on u-shapedness, we assess the influence of the prior on an entire portfolio of trials. Assume we have a portfolio consisting of $n = 25$ molecules that could be brought forward to phase 3. Developing a molecule is assumed to be associated to a cost c, say $c = 10$. If the molecule is approved, a profit of $w = 100$ can be expected. The number of successful trials X follows a binomial distribution of size n and success probability p. Now assume p is not fixed but is itself drawn from the distribution with density g, where the same random p is assumed to apply for all 25 trials. Typically in clinical development, phase 2 trials are smaller than phase 3 trials, so that we can assume g to have a u-shaped density. Marginalizing over g, we get a 'g-binomial' distribution (reminiscent of β-binomial) for X. The gain S associated to the portfolio can be computed as $S = w \cdot X - l \cdot n$. From the distribution of X, we can easily derive the distribution of S. Using this distribution, we find that the probability to have a gain of $S \leqslant 0$, that is a loss, amounts to < 0.001, 0.017, and 0.267 for the three priors 'g is 0.5 with probability 1', 'g is bell shaped', and 'g is u-shaped'. In general, looking at the entire distribution of S reveals that a portfolio with success probabilities drawn from a u-shaped prior may be more volatile than portfolios with success probabilities either following a point mass or bell-shaped prior. This applies in this very simple model where we assume the same, though random, p for all 25 trials that we acknowledge is not realistic. Exploring implications and the relevance of success probabilities drawn from g for portfolio decision-making is subject of further research.

6. DENSITY OF THE POWER FOR A GENERAL PRIOR DISTRIBUTION

Before we consider non-normal priors, we derive the analytical expression for g as a function of the prior. To this end, assume we have a prior distribution that we do not specify further at the moment, with density q and CDF Q. From (11) in the proof of Lemma 5.1, we find that the CDF G in this case is as follows:

$$G(y) = 1 - Q(\theta_{\text{suc}} - \sigma_{\text{fin}}\Phi^{-1}(y)). \tag{6}$$

Taking the derivative with respect to y wherever Q is differentiable, we get the density:

$$g(y) = q\left(\theta_{\text{suc}} - \sigma_{\text{fin}}\Phi^{-1}(y)\right)\frac{\sigma_{\text{fin}}}{\phi\left(\Phi^{-1}(y)\right)}. \tag{7}$$

Once y approaches either 0 or 1, the factor $1/\phi(\Phi^{-1}(y))$ tends to infinity thereby dominating the shape of g. It is worth nothing that this factor actually comes from the power function, not the prior. This implies that the effect of this factor cannot be removed by a smart choice of the prior, only compensated.

7. ALTERNATIVE PRIOR DISTRIBUTIONS

In Section 5, it was discussed that the u-shape of g is due to high probabilities for large beneficial or detrimental effects allowed by the prior. Here, we will explore the properties of PoS for alternative prior distributions. Candidate alternative distributions potentially not inducing a u-shape for g are thus as follows:

- Truncated normal, typically with large variance.
- Uniform prior on $[a, b]$.
- Uniform prior with normal tails. This is the 'pessimistic' prior introduced in [5].

The idea behind a truncated prior density is that the statistician elicits an interval in which the subject matter experts expect the effect to be in. In our experience, such a discussion might be easier than one about the center and variability of a normal prior distribution.

However, for both the truncated normal and the uniform prior with support $[a, b]$, the posterior will always be constrained to $[a, b]$, regardless of the data. Given that our prior assumption on the support might not always be fully accurate, the uniform prior with normal tails was developed to still basically elicit a uniform prior from the subject matter expert but provide the flexibility of an unrestricted support of the prior density.

For the first two of the aforementioned prior distributions, the function g will be constrained to the support $[\bar{a}, \bar{b}]$ where

$$\bar{a} = \Phi((\theta_{\text{suc}} - b)/\sigma_{\text{fin}}) \qquad \bar{b} = \Phi((\theta_{\text{suc}} - a)/\sigma_{\text{fin}}).$$

Note that they depend on the specifications of the final analysis and the support endpoints of the prior or the endpoints of the

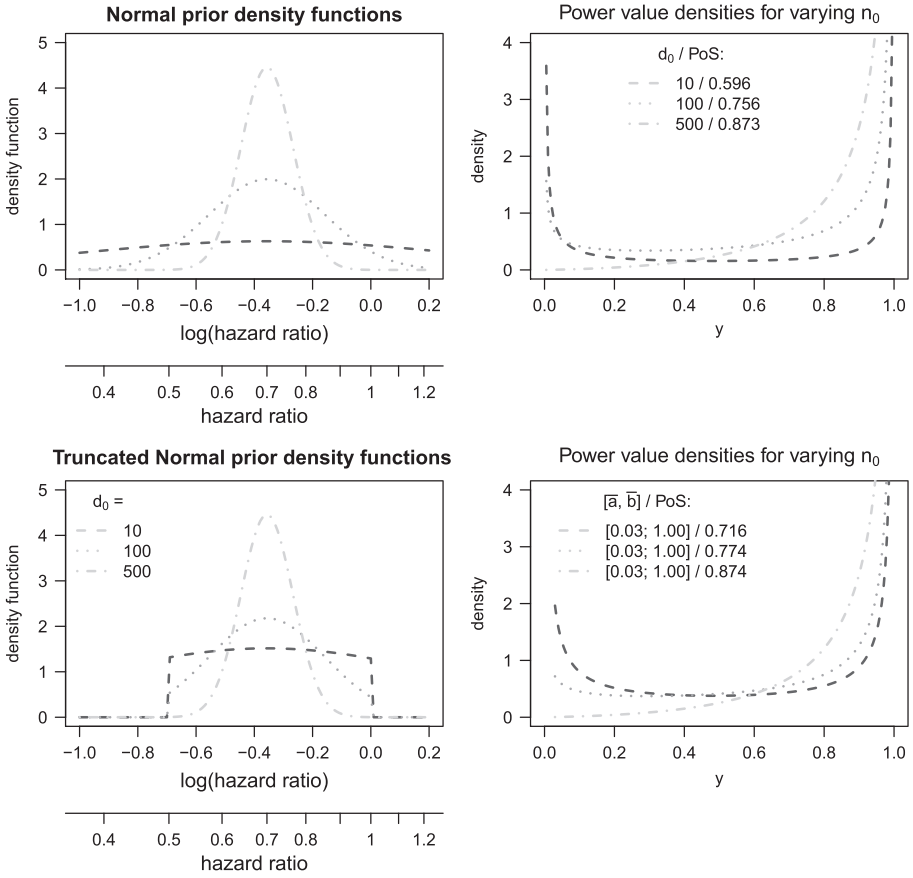

Figure 4. The functions g and $\bar{g}$ for normal and truncated normal prior, for different assumptions on prior variance (parameterized through prior number of events d_0, also for truncated normal for simplicity, via $4/d_0$).

constant part for the pessimistic prior. This implies that we can tune $\bar{a}$ and $\bar{b}$ by choosing a prior accordingly.

Truncated normal prior. Assume the random variable Θ is normal with mean θ_0 and variance σ_0^2. Then, Θ conditional on $a \leqslant \Theta \leqslant b$ for $a, b \in \mathbb{R}$ has a truncated normal distribution with density:

$$q(\theta) = \frac{\sigma_0^{-1}\phi((\theta-\theta_0)/\sigma_0)}{\Phi((b-\mu_0)/\sigma_0) - \Phi((a-\mu_0)/\sigma_0)}$$

for $\theta \in [a, b]$ leading to the density $\bar{g}$ of the power values:

$$\bar{g}(y) = \frac{g(y)}{\Phi\left((b-\mu_0)/\sigma_0\right) - \Phi((a-\mu_0)/\sigma_0)},$$

where $y \in [\bar{a}, \bar{b}]$.

Recall the generic example from Section 4. We center a normal prior at $\theta_0 = \log(0.7)$ and assume for a truncated normal prior that $a = \log(0.5), b = \log(1)$. The truncated normal prior also has its mode at θ_0. Figure 4 displays those two prior distributions (left) for assumed prior number of events 10/100/500 and the corresponding functions $\bar{g}$ (right).

For comparison and as a benchmark, on the top row of Figure 4, the corresponding functions for a normal prior as discussed in Section 5 are added.

Looking at Figure 4, we find that the qualitative features of $\bar{g}$ are similar to those of g. However, extreme power values close to 0 and 1 are now down-weighted. If the prior variance is decreasing, the truncated normal prior basically becomes a normal one inheriting the same shape for g. Note that for a normal prior on the log hazard ratio, we typically parameterize the prior variance using a suitable number of events, either corresponding to the amount of information we want the prior mean to have or the number of events this prior mean is based on. For the truncated normal, the variance does not correspond anymore to $\sqrt{4/d_0}$ with d_0 the number of events. As a consequence, the prior variance should rather be considered a tuning parameter inducing desirable properties of PoS and not be interpreted anymore based on number of events. However, for simplicity and for ease of comparison, we parameterize in Figure 4 the variance σ_0^2 of the truncated normal prior via $\sigma_0^2 = 4/d_0$ and provide the number of events, d_0, in the figure.

Uniform prior on $[a, b]$. Now, using a uniform prior, defined in $[a, b]$ as follows:

$$q_u(\theta) = (b-a)^{-1} 1\{\theta \in [a, b]\},$$

implies that the support of g is constrained to $[\bar{a}, \bar{b}]$. Using (7), we get

$$g_u(y) = \frac{\sigma_{\text{fin}}}{(b-a)\phi\left(\Phi^{-1}(y)\right)} 1\left\{y \in \left[\bar{a}, \bar{b}\right]\right\},$$

which is decreasing for $y \leqslant 1/2$ and increasing for $y > 1/2$. Figure 5 (top row) provides the corresponding plots from which this feature can also be seen. In order to be aligned with the normal and truncated normal priors, the uniform priors were chosen to be symmetric on the log scale around log(0.7). Computing g_u''

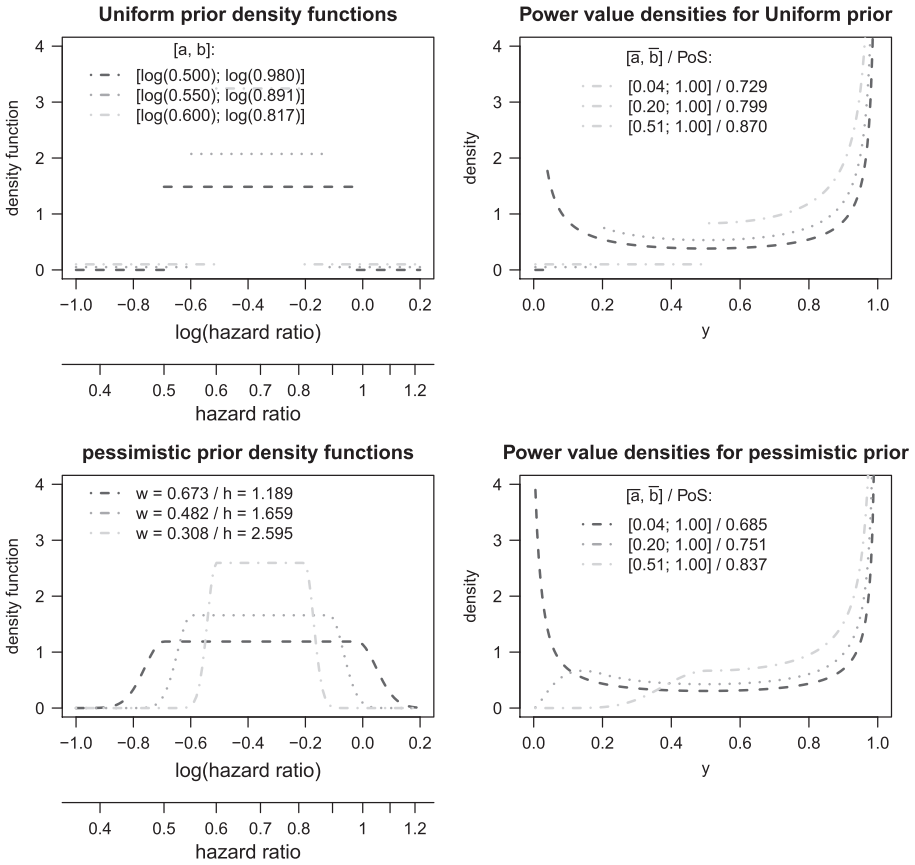

Figure 5. The functions g and $\bar{g}$ for various uniform priors as well as for various pessimistic priors. In the top row, stretches at height 0 slightly jiggled in height to ensure visibility of all functions.

(not shown) further reveals that g_u is always convex on the interval where it is larger than 0, irrespective of our choices of a and b. So in contrast to a (truncated) normal prior, where unimodality with an increasing piece left and a decreasing piece right of the mode can be achieved in principle (although only in practically hardly relevant cases) through suitable choice of a prior; this is not possible for a uniform prior. For a uniform prior, we can only achieve unimodality of g if it is either exclusively decreasing to the right of 0 or increasing to the left of 1, as in the third example in Figure 5 (top row).

However, the resulting PoS values for comparable scenarios (variance computed from prior sample size 10/100/500 for normal and truncated normal, support [log(0.50), log(0.98)] for uniform) are very similar.

Pessimistic prior. The input parameters for the pessimistic prior q_p, introduced in [5], are the width w of the constancy interval (on the log scale) and the height h of the constant part. Assuming that the tails are both half-normal with means at $a := \theta_0 - w/2, b := \theta_0 + w/2$ and that q_p is a density, the standard deviation σ_p of the normal tails can be computed from

$$(1 - wh)\phi_{a,\sigma_p^2}(a) = h,$$

where ϕ_{μ,σ^2} is the density function of a normal distribution with mean μ and variance σ^2. This yields

$$\sigma_p = (1 - wh)\left(h\sqrt{2\pi}\right)^{-1}$$

and

$$q_p(\theta) := \begin{cases} (1 - wh)\phi_{a,\sigma_p^2}(\theta) & \theta < a, \\ h & \theta \in [a, b], \\ (1 - wh)\phi_{b,\sigma_p^2}(\theta) & \theta > b. \end{cases}$$

Via (7), we compute the corresponding density g_p as follows:

$$g_p(y) = \frac{\sigma_{\text{fin}}}{\phi\left(\Phi^{-1}(y)\right)} \begin{cases} (1 - wh)\phi_{b,\sigma_p^2}\left(\theta_{\text{suc}} - \sigma_{\text{fin}}\Phi^{-1}(y)\right) & y < \bar{a}, \\ h & y \in \left[\bar{a}, \bar{b}\right], \\ (1 - wh)\phi_{a,\sigma_p^2}\left(\theta_{\text{suc}} - \sigma_{\text{fin}}\Phi^{-1}(y)\right) & y > \bar{b}. \end{cases}$$

Figure 5 (bottom row) depicts q_p and g_p for three different scenarios. The latter were chosen such that the support of the pessimistic prior matched those of the uniform priors in the top row of the plot. The height h was reduced by 20% compared with the uniform priors and redistributed to the normal tails. The advantages of using q_p over q_u are that the posterior in the latter model will not be constrained to the chosen prior support and g_p is continuous over [0, 1].

With a non-normal prior that assumes some knowledge on the underlying effect size, as the lightest gray variant in Figure 5, the resulting density g_p has a shape so that the corresponding PoS seems not an unreasonable summary measure. Compared with normal priors, quantifying the amount of prior knowledge using a prior with constrained (truncated) or clearly defined (pessimistic) support seems more straightforward than via the number of events for a normal prior. It also avoids the need to have $\sigma\text{fin} < \sigma_0$ to get a unimodal g.

When trying to achieve a benign form of g by tuning the prior, the following points should be considered:

- Such a prior typically assumes more knowledge about the underlying effect size than a simple normal prior with large variance.
- With the priors we explored, the shape of g could be made slightly more amenable to be summarized with the mean, but with the considered priors, we are in fact not fully successful in generating a satisfactory density g.

8. SENSITIVITY INTERVAL FOR PROBABILITY OF STUDY SUCCESS

As introduced in Section 2, PoS is the expectation of the power T as a function of the true effect with respect to the prior p, see (2). Note that PoS is neither a population parameter to be estimated in a frequentist sense nor a parameter with a prior distribution in a Bayesian framework, but an average of a transformed effect size (the power function) with respect to a prior on that effect size. In fact, PoS is a measure of the practical utility of a proposed trial, and PoS can thus be considered a utility function [3]. Once we have computed PoS for a given prior, we still would like to provide a measure of 'uncertainty', or rather 'sensitivity' to the chosen prior in the following sense: For two prior sample sizes 50 and 500 and a prior log hazard ratio θ_0 smaller than the minimal detectable hazard ratio, it seems sensible that PoS is higher for the larger compared with the smaller prior sample size. It seems sensible to use the density g of the distribution of the power values for that purpose. As a simple first approach, we define an equally tailed *sensitivity interval* for PoS corresponding to a level of γ (typically chosen to be 0.95) as follows:

$$\left[G^{-1}\left((1-\gamma)/2\right), G^{-1}\left((1+\gamma)/2\right)\right],$$

where G is the CDF with derivative g, as introduced in (6). Setting $\gamma = 0.95$, we find for our generic example an interval ranging from 0.00005 to 1.000. The interpretation of this interval is that 95% of the power values $T(\theta_{\text{suc}}|\theta)$ lie in this interval when θ is drawn from the prior p. As a matter of fact, this interval is of no practical use as it does not restrict the interval of plausible PoS values that are compatible with the prior. The reason is that the prior does not carry much information on the underlying true effect, implying a u-shaped density g for the distribution of the power values, as displayed in Figure 2 (right). For comparison, for $\gamma = 0.8$, we get an interval ranging from 0.022 to 1.000.

If in our generic example we increased the number of events, the prior information about the effect, log(0.7), is considered to be worth from 50 to 500, PoS increases to 0.873, and the two intervals (for $\gamma = 0.95, 0.8$) become [0.424, 0.999] and [0.655, 0.996]. Note that PoS increases because in our generic example, we indeed assumed that $\theta_0 < \theta - \text{suc}$. By increasing the number of events, θ_0 is considered to be worth, and thus, its weight in PoS computation the latter number has to go up.

As a matter of fact, the high-level conclusions from this exercise also could have been drawn by just looking at the function g. However, we found the sensitivity interval to be a useful tool when communicating PoS to non-statistician colleagues. To conclude, given the potential u-shape of g, alternative approaches to construct a sensitivity interval, for example, non-equally tailed with minimal coverage, could be explored as well.

9. DISCUSSION

In this paper, we have focused on properties of PoS as a function of the prior. Although the framework is generally applicable for binary, continuous, and time-to-event endpoints, for the purpose of illustration in our examples, we have focused on the latter. The probability of success, PoS, is the expectation of the power function, where expectation is with respect to a weighting density on the unknown effect size that could either be a prior (compute PoS before trial start) or posterior (compute PoS at an IA or update with external data).

For commonly used priors – notably normal priors – the density g of power values when sampling from the prior may have a shape for which PoS might not be very representative.

The u-shapedness is essentially a consequence of priors that, compared with the distribution of the effect estimate at the final analysis of the pivotal trial, do reflect less information on the underlying effect size, so a very plausible scenario in clinical development.

Using the analytical form of g, we can provide conditions under which, for a normal prior, a density g of more amenable, notably unimodal, shape appears. However, we consider these conditions rarely to be met in practice. However, unimodality of g and thus statistical properties of PoS should not be the only criteria that guide the selection of the shape of a prior on θ. Rather, statistical properties of a method together with clinical judgment assessing all the prior evidence should influence that choice.

Moving away from normal to priors with restricted support, shapes of g that seem better summarizable in one number can be achieved, although none of the priors considered in Section 7 provides a fully satisfactory approach.

To conclude, our recommendations when using PoS to quantify the probability of success of a clinical trial are the following:

- To always look at the density g when discussing PoS.
- To carefully assess whether a normal prior is the best choice and consider alternative priors as discussed in this paper.
- The evolution of g when data accumulates during conduct of a trial may yield insights about how much these data are worth beyond just looking at the prior.
- Prior distributions that represent as much information as the final effect estimate will lead to unimodal shapes of g. Wide prior distributions will not and hence reflect the fact that little is known about θ. Therefore, we encourage drug developers to give equal attention to setups with different densities g all yielding the same PoS.

More practical experience with the alternative priors proposed in this paper is needed to see whether these can contribute to higher-quality decision-making in drug development.

Acknowledgements

We thank Willi Maurer and Martin Posch for independently pointing us to [12] and Simon Wandel for helpful comments. In addition, we thank two reviewers for valuable comments that led to an improvement of the paper. We acknowledge that the discussion of the ordering of power and PoS at the end of Section 2 and Lemma A.1 were only added after a reviewer asked whether a result along the lines of Lemma A.1 could be proven.

Software

R code for all the methodology described in this paper is available upon request from the authors. This document was created using `Sweave` [13], LaTeX [14,15], and R [16].

REFERENCES

[1] Proschan M, Lan K, Wittes J. *Statistical monitoring of clinical trials: a unified approach*. Springer: New York, 2006.
[2] Spiegelhalter DJ, Reedman LS, Blackburn PR. Monitoring clinical trials – conditional power or predictive power. *Control Clin Trials* 1986; **7**(1):8–17.
[3] O'Hagan A, Stevens JW, Campbell MJ. Assurance in clinical trial design. *Pharmaceutical Statistics* 2005; **4**(3):187–201.
[4] Wang Y, Fu H, Kulkarni P, Kaiser C. Evaluating and utilizing probability of study success in clinical development. *Clin Trials* 2013; **10**(3):407–413.
[5] Rufibach K, Jordan P, Abt M. Sequentially updating the likelihood of success of a phase 3 pivotal time-to-event trial based on interim analyses or external information. *Journal of Biopharmaceutical Statistics* 2016; **26**(2):191–201.
[6] Dallow N, Fina P. The perils with the misuse of predictive power. *Pharmaceutical Statistics* 2011; **10**(4):311–317.
[7] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve R & D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* 2010; **9**(3):203–214.
[8] Ringel M, Tollman P, Hersch G, Schulze U. Does size matter in R&D productivity? If not, what does?. *Nature Reviews Drug Discovery* 2013; **12**(12):901–902.
[9] Herson J. Predictive probability early termination plans for phase II clinical trials. *Biometrics* 1979; **35**(4):775–783.
[10] Dmitrienko A, Wang MD. Bayesian predictive approach to interim monitoring in clinical trials. *Statistics in Medicine* 2006; **25**(13):2178–2195.
[11] Jennison C, Turnbull B W. *Group sequential methods with applications to clinical trials*. Chapman & Hall/CRC, Boca Raton, FL, 2000.
[12] Bauer P, Koenig F. The reassessment of trial perspectives from interim data – a critical view. *Statistics in Medicine* 2006; **25**(1):23–36.
[13] Leisch F. Dynamic generation of statistical reports using literate data analysis. In *COMPSTAT 2002 – Proceedings in Computational Statistics*, Härdle W, Rönz B (eds). Physica Verlag: Heidelberg; 2002; pp. 575–580.
[14] Knuth DE. Literate programming. *Computer Journal* 1984; **27**(2):97–111.
[15] Lamport L. *LaTeX: a document preparation system* (2nd ed.) Addison-Wesley: Reading Massachusetts, 1994.
[16] R Core Team. R: a language and environment for statistical computing. R *Foundation for Statistical Computing*, Vienna, Austria, 2014.
[17] Shaked M, Shanthikumar G. *Stochastic Orders*. Springer: New York, 2007.
[18] Elbarmi H, Mukerjee H. Peakedness and peakedness ordering in symmetric distributions. *Journal of Multivariate Analysis* 2009; **100**(4):594 –603.
[19] Javier R, José B. On the estimation of symmetric distributions under peakedness order constraints. In vol.57 *of Lecture Notes Monograph Series*, Rojo J (ed.) Institute of Mathematical Statistics: Beachwood, Ohio, USA; 2009; pp. 147–172.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Appendix

A Ordering of Power and PoS for a symmetric unimodal prior

The discussion at the end of Section 2 concluded with the statement that PoS, as a function $\text{PoS}(\theta_0)$ of the prior mean θ_0, is always closer to 0.5 than power at θ_0 not only for a Normal but for any unimodal symmetric prior. Lemma A.1 makes that statement precise.

Lemma A.1 (Ordering of power and PoS)**.** *Assume the prior random variable Θ is independent of $\hat{\theta}_{fin}$ and has a continuous CDF. Its density p is assumed to be unimodal and symmetric around θ_0 with finite variance. Then,*

$$\begin{aligned} PoS(\theta_0) &\leq T(\theta_{suc}|\theta = \theta_0) \quad \textit{for any } \theta_0 \leq \theta_{suc}, \\ PoS(\theta_0) &\geq T(\theta_{suc}|\theta = \theta_0) \quad \textit{for any } \theta_0 \geq \theta_{suc}, \end{aligned}$$

with equality for $\theta_0 = \theta_{suc}$.

As discussed at the end of Section 2 for a Normal prior, the statement of the lemma implies that PoS, considered a function of the prior mean θ_0, is always closer to 0.5 than the power at θ_0, see Figure 1. For a Normal prior on θ_0 this can easily be derived from the closed formulas for the power and PoS. Lemma A.1 extends this result to any prior that is unimodal and symmetric around θ_0

Proof of Lemma A.1. Rewrite the power as

$$\begin{aligned} T(\theta_{\text{suc}}|\theta = \theta_0) &= P\Big(Y \leq \frac{\theta_{\text{suc}} - \theta_0}{\sigma_{\text{fin}}}\Big) \\ &= P(\theta_0 + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}}) \end{aligned}$$

for a standard Normal random variable Y. As for PoS, using the law of total probability, we get

$$\begin{aligned} \text{PoS}(\theta_0) &= \int T(\theta_{\text{suc}}|\theta) p(\theta) d\theta \\ &= \int P(\Theta + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}} \Big| \Theta = \theta) p(\theta) d\theta \\ &= P(\Theta + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}}). \end{aligned}$$

To prove the first statement of the Lemma we thus need to show for any choice of $\theta_0 \leq \theta_{\text{suc}}$ that

$$P(\Theta + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}}) \leq P(\theta_0 + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}}) \tag{8}$$

with $\Theta \sim p$ and $Y \sim N(0,1)$. Rearranging (8) to

$$P\Big((\Theta - \theta_0) + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}} - \theta_0\Big) \leq P\Big((\theta_0 - \theta_0) + \sigma_{\text{fin}} Y \leq \theta_{\text{suc}} - \theta_0\Big)$$

we define

$$X_1 := \theta_0 - \theta_0 \equiv 0 \quad X_2 := \sigma_{\text{fin}} Y \quad Y_1 := \Theta - \theta_0 \quad Y_2 := \sigma_{\text{fin}} Y.$$

Then, X_2, Y_1 and Y_2 have unimodal densities symmetric around 0 and Y_1 is independent of $X_2 = Y_2$. $Y_1 = \Theta - \theta_0$ is a random variable larger in the peakedness order than $X_1 \equiv 0$, so that the assumptions of Theorem 3.D.4 in [17] are met. We can thus conclude that the random variable $\theta_0 + \sigma_{\text{fin}} Y$ is smaller in the peakedness order than $\Theta + \sigma_{\text{fin}} Y$, or equivalently, because of our symmetry assumption, $|\sigma_{\text{fin}} Y|$ is stochastically smaller than $|\Theta - \theta + \sigma_{\text{fin}} Y|$. This in turn implies, using (1.A.2) in [17],

$$P(|\theta_0 + \sigma_{\text{fin}} Y| \leq x) \quad \geq \quad P(|\Theta + \sigma_{\text{fin}} Y| \leq x)$$

for any $x \geq 0$. Exploiting symmetry of the involed densities again we finally conclude that the CDF of $\Theta + \sigma_{\text{fin}} Y$ is not larger than that of $\theta_0 + \sigma_{\text{fin}} Y$ for $x = \theta_{\text{suc}} - \theta_0 \geq 0$, in turn implying (8). The statement for $\theta_0 \geq \theta_{\text{suc}}$ can be proven similarly. □

In general, peakedness ordering does not imply stochastic ordering. The additional assumption of symmetry of the densities is a sufficient additional assumption for the implication to hold, see e.g. [18] or Definition 3.D.1 in [17]. The unimodality is needed in addition to symmetry to be able to show closedness of peakedness ordering under convolution of random variables in Theorem 3.D.4 of [17]. By a slightly more careful argumentation, the continuity assumption on the CDF of Θ could likely be dropped, using the definitions in [19].

B Further properties of g when the prior is Normal

Lemma B.1 specifies conditions under which g is symmetric.

Lemma B.1 (Symmetry of g around 0.5)**.** *We assume that $\tau > 0$ (corresponding to $\sigma_{fin} > 0$) and $\sigma_0 > 0$. Then, $g(y) = g(1-y)$ if either $\theta_{suc} - \theta_0 = 0$ or $y = 1/2$.*

Proof of Lemma B.1. We have for any $y \in \mathbb{R}$

$$\phi(y) = \phi(-y) \qquad \Phi^{-1}(y) = -\Phi^{-1}(1-y), \tag{9}$$

$$\phi(y) = (2\pi)^{-1/2} \exp(-y^2/2). \tag{10}$$

With this, we compute

$$g(y)/g(1-y) = \frac{\phi\Big(\psi - \tau\Phi^{-1}(y)\Big)}{\phi\Big(\psi - \tau\Phi^{-1}(1-y)\Big)} \frac{\phi(\Phi^{-1}(1-y))}{\phi(\Phi^{-1}(y))}.$$

Using (9), the second term is equal to 1. As for the first term we have, again using (9) as well as (10),

$$\begin{aligned}
&\phi\Big(\psi - \tau\Phi^{-1}(y)\Big) \,/\, \phi\Big(\psi - \tau\Phi^{-1}(1-y)\Big) = \\
&= \exp\Big(-\Big[\psi - \tau\Phi^{-1}(y)\Big]^2/2 + \Big[\psi + \tau\Phi^{-1}(y)\Big]^2/2\Big) \\
&= \exp\Big(2\tau\psi\Phi^{-1}(y)\Big),
\end{aligned}$$

where the last equality follows from $-(a-b)^2 + (a+b)^2 = 4ab$ for any $a, b \in \mathbb{R}$. From the latter expression we see that the exponent is equal to 0 if, and only if, $\psi = 0$ which is equivalent to $\theta_{\text{suc}} - \theta_0 \;=\; 0$ or $y = 1/2$. □

C Proofs of Lemmas

Proof of Lemma 5.1 We start with computing the CDF of T.

$$\begin{aligned} G(y) &= P(T(\theta_{\text{suc}}|\theta) \leq y) \\ &= 1 - P\Big(\Theta \leq \theta_{\text{suc}} - \sigma_{\text{fin}}\Phi^{-1}(y)\Big). \end{aligned} \tag{11}$$

Up to here, we have not yet used the Normality assumption for the prior. Using this, we continue the above computation to get

$$G(y) \;=\; 1 - \Phi\Big(\psi - \tau\Phi^{-1}(y)\Big).$$

To receive the density of T, we need to take the derivative of G. To prepare for that computation, use the rule about the derivative of an inverse to get

$$\frac{d}{dy}\Phi^{-1}(y) \;=\; \Big[\phi(\Phi^{-1}(y))\Big]^{-1}.$$

Using this, we have

$$g(y) \;=\; \frac{d}{dy}G(y) \;=\; \tau\exp\Big(-\frac{1}{2}\Big(\psi - \tau\Phi^{-1}(y)\Big)^2 + \frac{1}{2}\Phi^{-1}(y)^2\Big). \qquad \square$$

Proof of Theorem 5.2. To prove Theorem 5.2 we first compute the first derivative of g. Using (5) we get

$$\begin{aligned} g'(y) &= g(y)\frac{d}{dy}\Big[-\frac{1}{2}\Big(\psi - \tau\Phi^{-1}(y)\Big)^2 + \frac{1}{2}\Phi^{-1}(y)^2\Big] \\ &= \frac{g(y)}{\phi(\Phi^{-1}(y))}\Big[\tau\psi - \tau^2\Phi^{-1}(y) + \Phi^{-1}(y)\Big]. \end{aligned} \tag{12}$$

Setting $\tau = 1$ in g' we find that $g'(y) > 0$ for $\psi > 0$, i.e. g is strictly monotonely increasing, and $g'(y) < 0$ for $\psi < 0$, i.e. g is strictly decreasing. If $\tau = 1$ and $\psi = 0$ then $g' = 0$, so g is constant. See Figure 3 for an illustration.

To prove the first part of the second statement of the lemma, compute y_m where $g'(y_m) = 0$ as

$$y_m \;=\; \Phi\Big(\frac{\tau\psi}{\tau^2 - 1}\Big) \;=\; \Phi\Big(\frac{\sigma_{\text{fin}}(\theta_{\text{suc}} - \theta_0)}{\sigma_{\text{fin}}^2 - \sigma_0^2}\Big).$$

In order to determine whether y_m is a minimum or maximum, we need the second derivative of g. Defining the abbreviation $z = z(y) := \Phi^{-1}(y)$ we first note that

$$\begin{aligned} \frac{d}{dy}\phi(\Phi^{-1}(y)) &= \phi'(z)/\phi(z) \\ &= -z = -\Phi^{-1}(y) \end{aligned} \tag{13}$$

by $\phi'(x) = -x\phi(x)$ for any $x \in \mathbb{R}$. Using the quotient and product rule for derivatives we find

$$\begin{aligned} g''(y) &= \frac{d}{dy}\Big(\frac{g(y)}{\phi(\Phi^{-1}(y))}\Big)(\tau\psi - \tau^2 z + z) + \frac{g(y)}{\phi(z)}\Big[-\tau^2\frac{d}{dy}\Phi^{-1}(y) + \frac{d}{dy}\Phi^{-1}(y)\Big] \\ &= \frac{g'(y)\phi(z) - g(y)\frac{d}{dy}\phi(\Phi^{-1}(y))}{\phi(z)^2}(\tau\psi - \tau^2 z + z) + \frac{g(y)}{\phi(z)^2}(1-\tau^2). \end{aligned}$$

Now, plugging in g' and using (13),

$$\begin{aligned} g''(y)\phi(z)^2 &= \Big(g'(y)\phi(z) + g(y)z\Big)(\tau\psi - \tau^2 z + z) + g(y)(1-\tau^2) \\ &= \Big(g(y)\Big[\tau\psi - \tau^2 z + z\Big] + g(y)z\Big)(\tau\psi - \tau^2 z + z) + g(y)(1-\tau^2). \end{aligned}$$

Wrapping up we can finally write

$$g''(y) = \frac{g(y)}{[\phi(z)]^2}\Big[\Big(\tau\psi + z(1-\tau^2)\Big)^2 + z\Big(\tau\psi + z(1-\tau^2)\Big) + (1-\tau^2)\Big]. \tag{14}$$

Now, observe that for $y = y_m$,

$$z(y_m) = \Phi^{-1}(y_m) = \tau\psi(\tau^2-1)^{-1}$$

and thus $z(1-\tau^2) = -\tau\psi$. Plug this into (14) to get

$$g''(y_m) = \frac{g(y_m)}{[\phi(z)]^2}(1-\tau^2).$$

From this we see that if $\tau \leq 1$ then $g''(y_m) \geq 0$, so that y_m is a local minimum of g. If on the other hand $\tau \geq 1$ then y_m is a local maximum of g. This is consistent with Figure 3.

Finally, let us prove that for $\tau < 1$ and $y < y_m$ we have that $g'(y) < 0$, what implies that g is monotonely decreasing to the left of y_m. This is basically a consequence of the monotonicity of the Normal quantile function, i.e. $\Phi^{-1}(y) < \Phi^{-1}(y_m)$ for $y < y_m$. The first factor in (12) is positive, so it suffices to show that

$$\begin{aligned} \tau\psi - \tau^2\Phi^{-1}(y) + \Phi^{-1}(y) &= \tau\psi + \Phi^{-1}(y)(1-\tau^2) \\ &< \tau\psi + \Phi^{-1}(y_m)(1-\tau^2) \\ &= 0 \end{aligned}$$

by the definition of y_m. The inequality follows since $1-\tau^2 > 0$. The case for $y > y_m$ as well as the two cases for $\tau > 1$ can be proven similarly. □