International Statistical Review

International Statistical Review (2026) doi: 10.1111/insr.70031

The Random Power Function for Tests Based on Pivotal Quantities

Fulvio De Santis[1], **Stefania Gubbiotti**[1] and **Francesco Mariani**[2]

[1]*Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy*
[2]*Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Bologna, Italy*
Correspondence: Francesco Mariani, Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Bologna, Italy. Email: francesco.mariani20@unibo.it

Summary

In clinical trials planning, evaluation of the probability of success of an experiment is of central interest, for instance, in sample size determination. This assessment typically involves analyses of the power function of a test on a parameter of interest, such as a relevant treatment effect. In this article, we adopt a hybrid frequentist-Bayesian approach that is lately becoming more and more popular in the literature. Specifically, we focus on superiority trials, and we study the distribution of the power function induced by a design prior assigned to the parameter. Under mild assumptions, we derive general expressions for the cumulative and density functions of the random power in terms of its inverse. We then specialise this result to tests based on pivotal quantities, and we consider some classes of problems, both exact and asymptotic, conventionally employed in clinical trials. Ideas are exposed by resorting to four biomedical settings adapted from real data applications.

Key words: design prior; probability of success; sample size; uniformly most powerful tests; Wald test.

1 Introduction

In the frequentist approach to hypothesis testing on an unknown parameter of interest, the power function of a test is the function of the parameter defined by the probability of rejecting the null hypothesis (see Casella & Berger, 2002, Definition 8.3.1. and Equation 2 for a formal definition). This function is typically employed for evaluating the pre-experimental chance of success of a trial and for design purposes, for instance, to select the sample size. In this case, it is necessary to specify a design-value of the parameter that represents a relevant effect or effects difference; then, the optimal sample size is the smallest number of observations such that the power function evaluated at the design value is larger than a given threshold.

The resulting problem of local optimality can be overcome by specifying a probability distribution (design prior) on a set of design values. This prior induces a distribution on the power function, that is typically summarised by its expected value, a quantity usually—but non-univocally—called probability of success (PoS). See Kunzmann *et al.* (2021) for several alternative definitions of PoS and their interpretations. Despite its broad use in clinical trials applications, PoS is still a point summary of the random power distribution. It does not reflect neither the

variability of the distribution nor its shape, and as a consequence, it may be insufficient as an overall evaluation of the PoS of a trial. Conversely, as argued by Rufibach *et al.* (2016), De Santis & Gubbiotti (2024a), Mariani *et al.* (2024) and, pioneeringly, by Spiegelhalter *et al.* (2004) (Section 6.5.5), the whole density function of the random power shows in a plastic way what the chances of success are for any possible sample size. The shape of this distribution and where it spreads the probability mass of the random power depend crucially on the sample size: As n increases, this probability mass tends to be concentrated on larger and larger values of power. In Rufibach et al.'s words: "the evolution of" the power density function "when data accumulates during the conduct of a trial may yield insights about how much these data are worth." Therefore, the whole distribution of the power and other summaries in addition to PoS can be exploited for sample size calculations. Studying this distribution with respect to the sample size and to the other input elements involved (such as sampling distribution and design prior) is the basic motivation underlying the present article and the previous ones on the topic. Specifically, Rufibach *et al.* (2016) consider one-sided testing on the location parameter of a normal model and derive the expressions of the cumulative distribution function (cdf) and probability density function (pdf) of the power. They also study the qualitative features of this pdf to support the adequacy of PoS as a summary of the distribution. De Santis & Gubbiotti (2024a) extend this approach to the case of the scale parameter of distributions that are exponential families, according to the unifying formulation of Hoshyarmanesh *et al.* (2016). As a further development, Mariani *et al.* (2024) consider the distribution of alternative definitions of the random power.

In this paper, we further extend this idea in the following directions.

1. First of all, we provide general expressions of cdf and pdf of the random power as functions of the design prior cdf and pdf and of the inverse of the power function (Theorem 1).
2. As a second and main result, we obtain explicit formulas for the cdf and pdf of the random power that do not require the determination of a closed-form expression for the power function but solely rely on some mild assumptions on test statistics (Theorem 2). We also highlight that these assumptions are fulfilled by pivotal quantities.
3. Using the latter theorem, we retrieve the results of Rufibach *et al.* (2016) for the location parameter of a normal model and the results of De Santis & Gubbiotti (2024a) for the scale parameter of an exponential family model.
4. As an innovative case, we consider the Wald test statistic that also fulfils the assumptions of Theorem 2. Specifically, we consider the Poisson and the Bernoulli model.
5. Finally, we address the sample size determination (SSD) problem. For general review on the topic, see Wang & Gelfand (2002) and Sahu & Smith (2006).

Before moving on, notice that the approach of the present article falls into the hybrid frequentist-Bayesian category: We average the power of a frequentist test with respect to a (Bayesian) design prior distribution on the unknown parameter. Therefore, inference is frequentist; design is Bayesian. Even though they are not employed here, hybrid approaches have been considered in the literature. For instance, one can consider a fully Bayesian testing procedure and evaluate it for design using the (frequentist) sampling distribution of the data conditional on a fixed design value. For interplays between frequentist and Bayesian methods in SSD, see, among others, Brutti *et al.* (2014).

In order to illustrate our analytical results, we consider the following setups adapted from some classical examples in clinical trials (setups A, B and D) or observational studies (setup C).

Setup A: Rheumatoid arthritis trial (O'Hagan et al., 2005).

The goal of this trial is to assess the effect of a new drug in reducing C-reactive protein (CRP), a marker for disease severity, in patients with rheumatoid arthritis. Patients' reduction in CRP after 4 weeks relative to baseline is the continuous outcome variable of interest. The experiment is successful if this reduction is larger than a given cutoff.

Setup B: Head and neck cancer trial (Efron, 1988).

This study considers the survival time (days) of a group of patients suffering from head and neck cancer disease and treated using a combination of radiotherapy and chemotherapy. The goal of the trial is to prove that a new treatment regimen implies an improvement in terms of survival.

Setup C: Oral health intervention study (Lesaffre & Lawson, 2012).

This setup is based on a longitudinal study on a sample of 7-year-old children born in 1989 in Flanders, Belgium (Vanobbergen *et al.*, 2001). The endpoint is the dmft-index, that is the number of primary teeth that are decayed, missing due to extraction or filled for caries: this index ranges between 0 (no caries) - 20 (all primary teeth with caries). The goal is to assess whether the dmft-index in the population exceeds a relevant threshold.

Setup D: rheumatoid arthritis trial (O'Hagan et al., 2005).

In the same context of setup A, instead of a continuous outcome, we consider the binary outcome variable ACR20 that indicates whether a patient has 20% improvement in symptoms after 6 months of treatment. Success of the trial is achieved if we prove that the probability of response is sufficiently large.

The R code used for the implementation of the corresponding examples A,B,C,D is provided as Supporting Information to ensure full reproducibility of the results and to allow the possibility of taking into account additional design scenarios.

This article is structured as follows. In Section 2, we provide the general results (Theorems 1 and 2) that are later specialised to uniformly most powerful tests in Section 3. The normal location problem with illustration to setup A is addressed in Section 3.1, where both known and unknown variance cases are considered; in Section 3.2, results for the scale exponential case are derived and an application to setup B is illustrated. Section 3.3 is devoted to Wald test for the parameters of two specific models: Poisson (Section 3.3.1, setup C) and Bernoulli (Section 3.3.2, setup D). In Section 3.4, we introduce SSD criteria with an application to the efficacy trial with binary outcome described in setup D. Finally, Section 4 contains some concluding remarks.

2 Methodology

Let $\mathbf{X}_n = (X_1, \ldots, X_n)$ be a random sample from $f_n(\mathbf{x}_n|\theta) = \prod_{i=1}^n f_X(x_i|\theta)$, where θ is a scalar parameter in $\Omega \subseteq \mathbb{R}$. With no loss of generality, let us consider the following one-sided hypotheses:

$$H_0: \theta \in \Omega_0 \qquad \text{vs.} \qquad H_1: \theta \in \Omega_1, \tag{1}$$

where $\Omega_0 = (\theta \in \Omega: \theta \leq \theta_0)$ and $\Omega_1 = (\theta \in \Omega: \theta > \theta_0)$, with $\theta_0 \in \Omega$. Let us now consider a testing procedure, either frequentist or Bayesian, based on a test statistic and on a rejection region R. The frequentist power function of the test is an application $\eta: \Omega \rightarrow [0, 1]$ defined as

$$\eta(\theta) = \mathbb{P}_\theta(R), \quad \theta \in \Omega, \tag{2}$$

where the probability is computed with respect to $f_n(\mathbf{x}_n|\theta)$, the sampling distribution of the data. For SSD, the power function is typically evaluated at a design-value $\theta_d \in \Omega_1$: The optimal sample size is the minimum integer number such that $\eta(\theta_d) \geq \lambda, \ \lambda \in (0,1)$, and it depends on θ_d. To avoid this problem of local optimality, we assume that Θ is a random variable. Specifically, we suppose that Θ is absolutely continuous with density $\pi(\cdot)$. We are here interested in determining and studying the cdf $G(\cdot)$ and pdf $g(\cdot)$ of the random power $Y = \eta(\Theta)$ induced by $\pi(\cdot)$. The features of $g(\cdot)$ will guide us in accepting its expected value PoS as a suitable summary of the random power distribution or in selecting alternative syntheses. The following result provides general expressions of $g(\cdot)$ and $G(\cdot)$.

Theorem 1. *Consider the hypotheses (1). Let R be the rejection region of a test, $\eta(\theta) = \mathbb{P}_\theta(R)$ its power function and $\mathbb{F}_\pi(\cdot)$ and $\pi(\cdot)$ the design prior cdf and pdf of Θ, respectively. Then,*

i) if $\eta(\cdot)$ is strictly increasing and $\eta^{-1}(\cdot)$ is its inverse function, then the cdf of $Y = \eta(\Theta)$ is

$$G(y) = \mathbb{F}_\pi[\eta^{-1}(y)], \ y \in (0,1); \tag{3}$$

ii) if $\eta(\cdot)$ is differentiable and $\eta'(\theta) \neq 0 \ \forall \theta \in \Omega$, then $\eta^{-1}(\cdot)$ is also differentiable and the density function of Y is

$$g(y) = \pi[\eta^{-1}(y)] \cdot \left| \frac{d}{dy} \eta^{-1}(y) \right|, \ y \in (0,1). \tag{4}$$

Proof *We have that*

$$\mathbb{P}_\pi[\eta(\Theta) \leq y] = \mathbb{P}_\pi[\Theta \leq \eta^{-1}(y)] = \mathbb{F}_\pi[\eta^{-1}(y)], \ y \in (0,1).$$

The expression of $g(y)$ follows by deriving $G(y)$ with respect to y. □

Remarks.

1. Monotonicity and differentiability of $\eta(\cdot)$, that guarantee existence and differentiability of $\eta^{-1}(\cdot)$, are typically satisfied in one-sided testing of parameters of standard models.
2. A similar result holds for the reversed hypotheses: $H_0: \theta \geq \theta_0$ vs $H_1: \theta < \theta_0$. In this case, $G(y) = 1 - \mathbb{F}_\pi[\eta^{-1}(y)], \ y \in (0,1)$ and $g(y)$ is still given by Equation (4).
3. The existence of $g(\cdot)$ requires that the design prior $\pi(\cdot)$ is proper.
4. If $\pi(\cdot)$ is a point-mass prior on θ_d, then $Y = \eta(\theta_d)$ with probability one.

Under some assumptions on the test statistic, we can find more specific and explicit expressions for $G(\cdot)$ and $g(\cdot)$.

Theorem 2. *Let $Q(T, \theta_0)$ be a test statistic for the hypotheses (1) based on a (real-valued) statistic $T(\mathbf{X}_n)$ such that*

i) $Q(T, \theta)$ has cdf $\mathbb{F}(\cdot)$ not dependent on $\theta, \ \forall \theta \in \Omega$;
ii) the test rejects H_0 if $Q(t, \theta_0) \geq q$, where $t = T(\mathbf{x}_n)$ and q is a suitable threshold;
iii) $Q(\cdot, \theta)$ and its inverse $Q_1^{-1}(\cdot, \theta)$ are both strictly increasing $\forall \theta \in \Omega$;
iv) $Q(t, \cdot)$ and its inverse $Q_2^{-1}(t, \cdot)$ are both strictly decreasing $\forall t$.

If Θ *has cdf* $\mathbb{F}_\pi(\cdot)$ *and pdf* $\pi(\cdot)$*, then* $G(y)$ *and* $g(y)$ *are given by (3) and (4) with*

$$\eta^{-1}(y) = Q_2^{-1}\left[Q_1^{-1}(q, \theta_0), q_{1-y}\right], \qquad y \in (0,1), \tag{5}$$

where q_ϵ is the $\epsilon-$ quantile of $\mathbb{F}(\cdot)$, $\epsilon \in (0,1)$.

Proof *For the hypotheses (1), the power function is*

$$\begin{aligned}
\eta(\theta) &= \mathbb{P}_\theta[Q(T, \theta_0) \geq q] && \text{from (ii)}\\
&= \mathbb{P}_\theta[T \geq Q_1^{-1}(q, \theta_0)] && \text{from (iii)}\\
&= \mathbb{P}[Q(T, \theta) \geq Q\big(Q_1^{-1}(q, \theta_0), \theta\big)] && \text{from (iii)}\\
&= 1 - \mathbb{F}[Q\big(Q_1^{-1}(q, \theta_0), \theta\big) \quad].
\end{aligned}$$

Hence, for $y \in (0,1)$,

$$\begin{aligned}
G(y) &= \mathbb{P}_\pi[\eta(\Theta) \leq y]\\
&= \mathbb{P}_\pi[\mathbb{F}\big(Q\big(Q_1^{-1}(q, \theta_0), \Theta\big)\big) \geq 1 - y]\\
&= \mathbb{P}_\pi[Q\big(Q_1^{-1}(q, \theta_0), \Theta\big) \geq q_{1-y}]\\
&= \mathbb{P}_\pi[\Theta \leq Q_2^{-1}\Big(Q_1^{-1}(q, \theta_0), q_{1-y}\Big)] && \text{from (iv)}
\end{aligned}$$

which is Equation (3). Then, taking the derivative wrt y, $g(y)$ given in (4) is obtained. □

Remarks.

1. Assumption ***i)*** is satisfied if $Q(t, \theta)$ is a pivotal quantity.
2. Pivotal quantities typically satisfy conditions ***iii)*** and ***iv)***.
3. Theorem 2 can be widely applied since a large range of statistical tests are actually based on pivotal quantities.
4. If T has pdf $f(t|\theta)$ that can be expressed in the form $f(t|\theta) = h(Q(t, \theta))|\frac{\partial}{\partial t}Q(t, \theta)|$, for some function h and some function Q monotone in t for each θ, then $Q(t, \theta)$ is a pivotal quantity. See, for instance, Casella & Berger (2002). As special cases, if θ is a location parameter, $t - \theta$ (and its linear functions) is a pivotal quantity; similarly, if θ is a scale parameter, $\frac{t}{\theta}$ is a pivotal quantity.

2.1 Nuisance Parameters

So far we have considered a one-parameter model, a case which covers most of the specific examples analysed in Sections 3.1–3.3. Suppose now that the model for $\mathbf{X}_n$ depends on the two-dimensional real-valued parameter (θ, ψ), where θ is the parameter of interest and ψ is a nuisance parameter. This case has been considered, with respect to the definition of PoS, by Liu (2010). Let us assume that the power function η, the test statistics Q and their inverse functions η^{-1}, Q_1^{-1} and Q_2^{-1} can be defined conditionally on the nuisance parameter ψ. By assigning a design prior density to the vector (θ, ψ), that is

$$\pi(\theta, \psi) = \pi(\theta|\psi) \cdot \pi(\psi),$$

we can obtain the unconditional pdf of the random power as

$$g(y) = \int_{\mathbb{R}} g(y|\psi) \cdot \pi(\psi) d\psi,$$

where the conditional density $g(y|\psi)$ is derived using Theorems 1 and 2 of the previous section. Similarly, following Liu (2010), the unconditional PoS can be computed as the expected value of the conditional PoS

$$\int_{\mathbb{R}} y \cdot g(y|\psi) dy$$

with respect to $\pi(\psi)$. As an example, see the normal case with unknown variance ($\psi = \sigma^2$) in Section 3.1.

3 Uniformly Most Powerful Tests

For the hypotheses (1), we now consider the size γ-UMP (uniformly most powerful) test that rejects H_0 if $Q(T, \theta_0) > q_{1-\gamma}$, $\gamma \in (0, 1)$, where Q is a pivotal quantity and T is a sufficient statistic such that the likelihood ratio is a monotone function of T (see Karlin & Rubin, 1956). In the following sections, we consider tests for the normal location parameter and scale parameter of exponential families for which, as stated in Remark 3 of Theorem 2, the inversion of the pivotal quantities wrt t and θ is guaranteed and straightforward. By analysing these two cases, we revisit the results obtained in Rufibach *et al.* (2016) and in De Santis & Gubbiotti (2024a), respectively. In Section 3.3, we apply our methodology to Wald tests.

3.1 Normal Location Parameter

Assume that $X_1, \ldots, X_n|\theta$ are iid $N(\theta, \sigma^2)$, where $\sigma^2 = \sigma_0^2$ is known. For the set of hypotheses (1), the pivotal test statistic is $Q(T, \theta_0) = \frac{\sqrt{n}}{\sigma}(T - \theta_0)$, where $T = \overline{X}^n$. The power function of the size γ-UMP test is $\eta(\theta) = 1 - \Phi\left[\frac{\sqrt{n}}{\sigma}(\theta_0 - \theta) + z_{1-\gamma}\right]$, where $\Phi(\cdot)$ and $z_{1-\gamma}$ are the cdf and the $(1-\gamma)$-quantile of the standard normal, $\gamma \in (0, 1)$. Using Theorem 2, we retrieve the specific expression of the cdf and the pdf of Y previously obtained by Rufibach *et al.* (2016), that we now denote by $G(\cdot|\sigma^2)$ and $g(\cdot|\sigma^2)$ to stress the conditioning on the variance (nuisance parameter). In order to use Equation (5) of Theorem 2, note that the inverse functions of $Q(t, \theta)$ are, respectively,

$$Q_1^{-1}(u, \theta) = \theta + \frac{\sigma}{\sqrt{n}}u \qquad \text{and} \qquad Q_2^{-1}(t, w) = t - \frac{\sigma}{\sqrt{n}}w.$$

Therefore,

$$Q_1^{-1}(z_{1-\gamma}, \theta_0) = \theta_0 + \frac{\sigma}{\sqrt{n}}z_{1-\gamma}$$

and

$$\eta^{-1}(y) = \theta_0 + \frac{\sigma}{\sqrt{n}}(z_{1-\gamma} - z_{1-y}). \tag{6}$$

Noting that $\left|\frac{d}{dy}\eta^{-1}(y)\right| = \frac{\sigma}{\sqrt{n}}\frac{1}{\phi(z_{1-y})}$, where $\phi(\cdot)$ is the standard normal pdf, we have that

$$g(y|\sigma^2) = \pi\left(\theta_0 + \frac{\sigma}{\sqrt{n}}(z_{1-\gamma} - z_{1-y})\right)\frac{\sigma}{\sqrt{n}}\frac{1}{\phi(z_{1-y})}, \qquad y \in (0,1). \tag{7}$$

Now, if we assume that $\Theta \sim \mathrm{N}(\theta_d, \sigma^2/n_d)$, $\theta_d \in \mathbb{R}$ and $n_d > 0$, then, for $y \in (0,1)$,

$$G(y|\sigma^2) = \Phi\left(\frac{\sqrt{n}}{\sigma}\tau(\theta_0 - \theta_d) + \tau(z_{1-\gamma} - z_{1-y})\right) \tag{8}$$

and

$$g(y|\sigma^2) = \phi\left(\frac{\sqrt{n}}{\sigma}\tau(\theta_0 - \theta_d) + \tau(z_{1-\gamma} - z_{1-y})\right)\frac{\tau}{\phi(z_{1-y})}, \qquad y \in (0,1), \tag{9}$$

where $\tau = \sqrt{\frac{n_d}{n}}$. In this case, it is also possible to obtain the closed-form expression of the expected value of Y that we refer to as *conditional* PoS (on the value σ^2):

$$\mathbb{E}_\pi[Y|\sigma^2] = 1 - \Phi\left(\frac{\theta_0 - \theta_d + \frac{\sigma}{\sqrt{n}}z_{1-\gamma}}{\xi_d}\right), \qquad \xi_d^2 = \sigma^2\left(\frac{1}{n_d} + \frac{1}{n}\right). \tag{10}$$

Notice that, consistently with Remark 3 of 1, the existence of $g(\cdot)$ and of its expected value $\mathbb{E}_\pi[Y]$ requires $\pi_d(\cdot)$ to be proper, that is, in this normal case, that n_d is strictly larger than zero; this means that, as $n_d \to 0$, the given expression of $g(\cdot)$ and of PoS lose their meaning. On the other hand, consistently with Remark 4 of 1, a point-mass design prior concentrated on θ_d (that, in this normal case, is obtained as $n_d \to \infty$) would make PoS equal to the frequentist power $\eta(\theta_d)$: This means that the frequentist design can be seen as a limiting case of the Bayesian design (see De Santis & Gubbiotti, 2024b). Finally, note that, as n tends to infinity, PoS converges to π_1 that is the probability of Ω_1 under the design prior (see De Santis & Gubbiotti, 2024a, 2024b).

Unknown variance case.

Assume now that σ^2 is unknown and that $\pi(\theta, \sigma^2) = \pi(\theta|\sigma^2) \cdot \pi(\sigma^2)$, where $\pi(\theta|\sigma^2)$ is a $\mathrm{N}(\theta_d, \sigma^2/n_d)$ density and $\pi(\sigma^2)$ is an inverse gamma denoted by $\mathrm{IG}(\alpha_d, \beta_d)$, a standard choice for a prior on (the square of) a scale parameter. We can then exploit the approach sketched in Section 2.1 with $\psi = \sigma^2$. In this case, $g(y)$ is obtained by taking the expectation of the conditional density $g(y|\sigma^2)$ given by (9), with respect to $\pi(\sigma^2)$. Similarly, the *unconditional* PoS can be found by averaging (10) again with respect to $\pi(\sigma^2)$. The unconditional distribution of Y and its expected value can be easily computed by resorting to numerical integration and/or Monte Carlo simulation.

Example A. Let us consider the rheumatoid arthritis trial of setup A, where θ is the expected reduction in CRP after 4 weeks relative to baseline. We are interested in testing the null hypothesis $\theta \leq \theta_0 = 0$, against the alternative $\theta > \theta_0$.

Known variance case. First of all, assume that the population variance is known; for instance, its value may be provided by an expert or suggested by the literature. Alternatively, the known variance assumption may arise from the formulation proposed by Spiegelhalter *et al.* (2004): Data relevant to θ are based on a statistic that is (at least asymptotically) normally distributed with mean θ and variance σ^2/m, where σ^2 is fixed and m is the effective number of observations. In the present example, following O'Hagan *et al.* (2005), we assume standard deviation of the

CRP reduction that is equal to 0.25; that is, we consider $\sigma^2 = \sigma_0^2 = 0.0625$. In addition, the Authors suggest that "*the elicitation of prior information about the unknown treatment effect from a relevant expert gives*" prior mean 0.2 and prior variance 0.06: Therefore, as design prior, we consider a normal density with $\theta_d = 0.2$ and prior sample size $n_d \approx 2$ (see Figure 1, left panel, solid line) which assigns probability $\pi_0 = 0.212$ to Ω_0. For $n = 50$ and $\gamma = 0.05$, this prior yields the u-shaped density function $g(y|\sigma^2)$ shown in the right panel of Figure 1 (solid line), with expected value $\mathbb{E}_\pi(Y|\sigma^2) = 0.709$ and median $\mathbb{M}_\pi(Y|\sigma^2) = 0.999$. PoS falls in an interval of y-values with rather low density, and therefore, it does not seem to properly summarise the distribution of $g(y|\sigma^2)$. In Figure 1 (left panel, dashed line), we consider a second design prior centred on the same mean $\theta_d = 0.2$ but more concentrated ($n_d = 10$). In this case, $\pi_0 = 0.037$ and the u-shape of the (dashed) density $g(y|\sigma^2)$ is now almost negligible; the corresponding expected value is now much closer to the median (0.886 and 1 respectively).

Sensitivity analysis. We are now interested in performing a sensitivity analysis with respect to the choice of the nuisance parameter σ^2. In Table 1, we report the values of the conditional power and PoS for a set of values for the variance ranging approximately from $1/5\sigma_0^2$ to $3\sigma_0^2$. The table also shows the probabilities π_1 that are the maximum values that PoS can reach conditionally on the chosen levels for the variance. The bottom line reports the ratio between PoS and its maximum. As expected, the larger the variance, the smaller the values of power and PoS. Specifically, for values of the variance smaller than σ_0^2 the power reaches its maximum, whereas PoS is at most 95% of π_1. Figure 2 (left panel) shows the curves $g(y|\sigma^2)$ corresponding to the different choices of σ^2: as the variance gets smaller, the u-shape tends to disappear. Overall, in this example, the impact of the choice of σ^2 is not strong for power, but it is more relevant for PoS.

Unknown variance case. Following Sections 2.1 and 3.1, we now account for uncertainty on the nuisance parameter by assigning an $\text{IG}(\alpha_d, \beta_d)$ prior to σ^2. In order to elicit this prior, we set the prior mean equal to $\sigma_0^2 = 0.0625$, and we require that it assigns a probability approximately equal to 0.95 to the interval $[1/5\sigma_0^2, 3\sigma_0^2]$ considered in Table 1. It turns out that $\alpha_d = 2.5$ and $\beta_d = 0.09375$. The resulting prior tends to favour values less than σ_0^2, since it assigns a probability approximately equal to 0.69 to the interval $[0, \sigma_0^2]$. As a result, the unconditional PoS is equal to 0.999, a value larger than 0.709, that is the value of PoS conditional on σ_0^2. Finally, Figure 2 (right panel) displays the unconditional density of the random power.

3.2 Scale Parameter for Exponential Families

Assume that $X_1, \ldots, X_n|\theta$ are iid from a scale-exponential family with density function

$$f_X(x|\theta) = \frac{h(x)}{\theta^m}\exp\left\{-\frac{\omega(x)}{\theta^r}\right\}, \quad x \geq 0, \quad \theta > 0, \tag{11}$$

where $r,\ m \in \mathbb{R},\ mr > 0,\ h(x)$ and $\omega(x)$ are non-negative real-valued functions of the observation x. The joint density function of $\mathbf{X}_n$ is $f_n(\mathbf{x}_n|\theta) = \frac{h_n(\mathbf{x}_n)}{\theta^{mn}}\exp\left\{-\frac{T(\mathbf{x}_n)}{\theta^r}\right\}$, where $h_n(\mathbf{x}_n) = \prod_{i=1}^n h(x_i)$ and $T(\mathbf{x}_n) = \sum_{i=1}^n \omega(x_i)$ is a complete sufficient statistic. It is easy to check that the size γ-UMP test for the hypotheses (1) rejects H_0 if

$$Q(T,\ \theta_0) = \frac{2T}{\theta_0^r} \geq q_{1-\gamma},$$

where under the null hypothesis, $Q(T,\ \theta_0) \sim \chi^2_{2\nu},\ \nu = mn/r$ and where q_ϵ is the ϵ-quantile of the

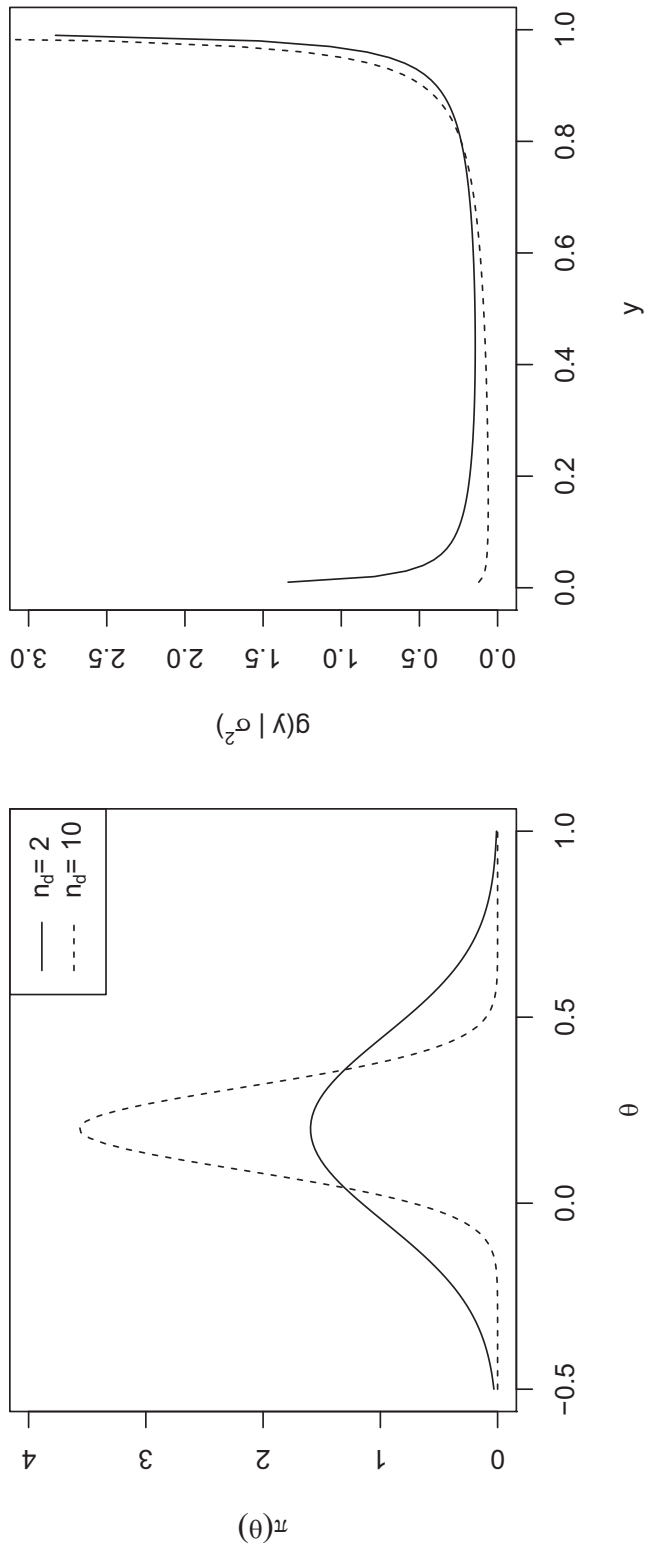

FIGURE 1. *Example A (rheumatoid arthritis trial): plots of the normal densities* $\pi(\theta)$ *for* $\theta_d = 0.2$ *and* $n_d = 2, 10$ *(left panel) and of the corresponding* $g(y|\sigma^2)$ *(right panel), for* $n = 50$, $\theta_0 = 0$, $\gamma = 0.05$, $\sigma^2 = 0.0625$.

Table 1. *Example A (sensitivity analysis with respect to σ^2): Values of conditional power, PoS, π_1 and PoS/π_1 for different values of σ^2.*

σ^2	$1/5\sigma_0^2$ *0.013*	$1/3\sigma_0^2$ *0.021*	$1/2\sigma_0^2$ *0.031*	σ_0^2 *0.0625*	$2\sigma_0^2$ *0.125*	$3\sigma_0^2$ *0.188*
Conditional power	1.000	1.000	1.000	1.000	0.991	0.947
Conditional PoS	0.933	0.873	0.813	0.713	0.627	0.589
π_1	0.960	0.916	0.872	0.788	0.714	0.678
Conditional PoS / π_1	0.971	0.953	0.932	0.904	0.879	0.869

$\chi^2_{2\nu}$ (see Hoshyarmanesh *et al.*, 2016 and De Santis & Gubbiotti, 2024a). Specific choices of r, m and $\omega(\cdot)$ yield well-known scale parameter models, as reported in Table 1 in Hoshyarmanesh *et al.* (2016). For instance, we retrieve the exponential model for $r = 1,\ m = 1$ and $\omega(x) = x$, the normal model with known mean μ and unknown standard deviation θ for $r = 2,\ m = 1$ and $\omega(x) = (x - \mu)^2/2$. The power function is

$$\eta(\theta) = 1 - \mathbb{F}\left[\left(\frac{\theta_0}{\theta}\right)^r q_{1-\gamma}\right],\ \theta > 0,$$

where $\mathbb{F}(\cdot)$ is now the cdf of the $\chi^2_{2\nu}$. Using Theorem 2, we retrieve the specific expression of $G(\cdot)$ and $g(\cdot)$, previously obtained in De Santis & Gubbiotti (2024a). In order to apply (5), note that the inverse functions of $Q(t, \theta)$ are, respectively,

$$Q_1^{-1}(u, \theta) = \frac{u\theta^r}{2} \qquad \text{and} \qquad Q_2^{-1}(t, w) = \left(\frac{2t}{w}\right)^{\frac{1}{r}}.$$

Therefore, $Q_1^{-1}(q_{1-\gamma}, \theta_0) = \dfrac{q_{1-\gamma}\theta_0^r}{2}$ and $\eta^{-1}(y) = \theta_0\left(\frac{q_{1-\gamma}}{q_{1-y}}\right)^{\frac{1}{r}}$. Note that $\left|\dfrac{d}{dy}\eta^{-1}(y)\right| = \dfrac{1}{|r|}\dfrac{1}{f(q_{1-y})}\dfrac{\eta^{-1}(y)}{q_{1-y}}$, where $f(\cdot)$ is the pdf of the $\chi^2_{2\nu}$; we have that

$$g(y) = \frac{1}{|r|}\frac{\pi(\eta^{-1}(y))}{f(q_{1-y})}\frac{\eta^{-1}(y)}{q_{1-y}}, \qquad y \in (0, 1). \tag{12}$$

The above expression is general and can be specialised by choosing any sensible prior for Θ. For instance, De Santis & Gubbiotti (2024a) consider conjugate r-root inverse gamma priors.

Example B. In setup B, let θ be the expected survival time of head and neck cancer patients treated by a combination of radiotherapy and chemotherapy. We model the survival time with the rate exponential model obtained from (11) by setting $r = -1,\ m = -1$ and $\omega(x) = x$. Historical data from Efron (1988) are used to set the mean $\theta_d = 222$ of a gamma design prior. In particular, we consider a $\text{Ga}(a = 2,\ b = 111)$ (b scale) that assigns 95% probability to the equal tail interval $(26.9, 618.4)$. This prior yields $\pi_0 = 0.23$ and $\pi_1 = 0.77$. Assuming that $\theta_0 = 100$, for a sample size $n = 50$, we obtain the density function represented in Figure 3 (solid line). Note that the expected value of Y is quite different from the median ($\mathbb{E}_\pi(Y) = 0.683$ and $\mathbb{M}_\pi(Y) = 0.995$ respectively), so that PoS cannot be considered a representative summary of $g(y)$ (in fact $G(0.683) \approx 0.32$). Conversely, the discrepancy between $\mathbb{E}_\pi(Y) = 0.898$ and $\mathbb{M}_\pi(Y) = 0.999$ reduces if we consider an alternative design prior with the same mean, but smaller variance, for example, $\text{Ga}(a = 8,\ b = 27.75)$, that assigns 95% probability to the equal

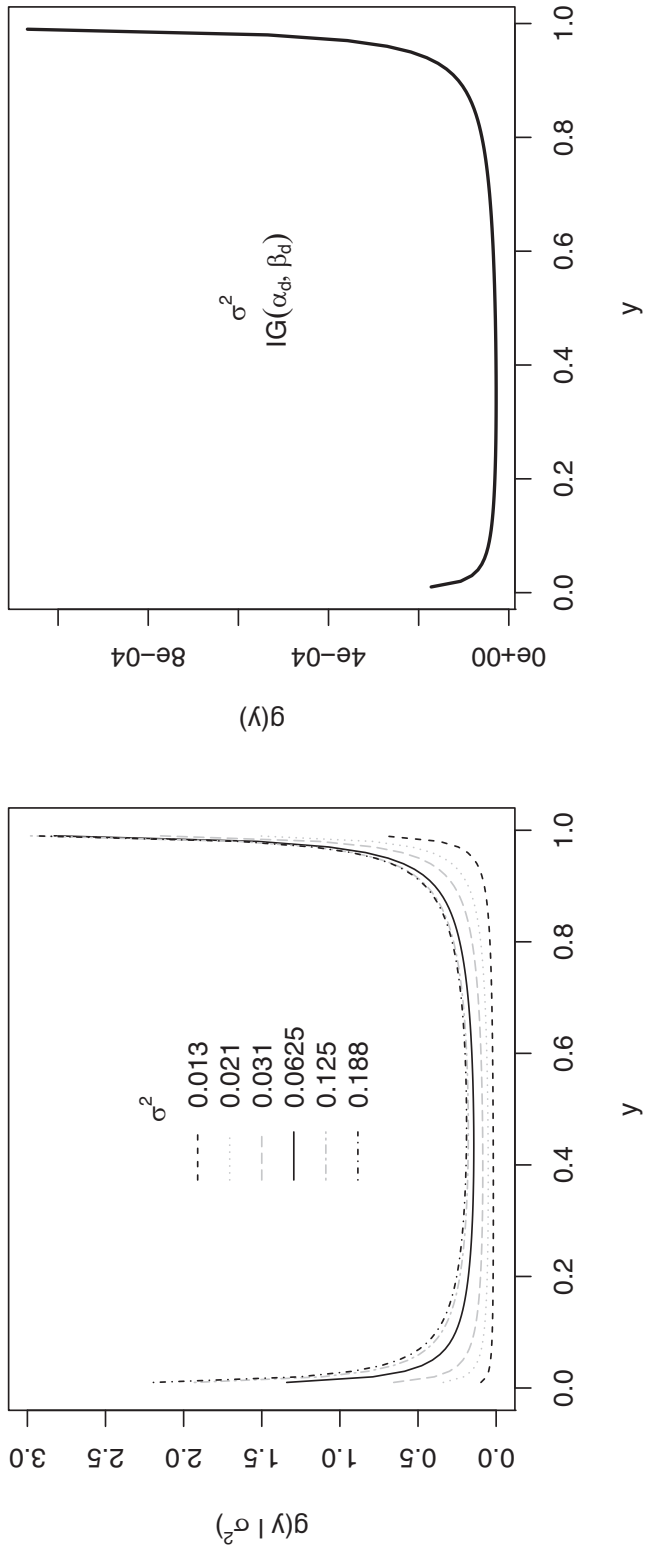

FIGURE 2. *Example A (rheumatoid arthritis trial): plots of the densities* $g(y|\sigma^2)$ *conditional on different values of* σ^2 *(left panel) and of the unconditional density* $g(y)$ *when* $\sigma^2 \sim \text{IG}(\alpha_d = 2.5, \beta_d = 0.09375)$ *(right panel), for* $n = 50$, $\theta_0 = 0$, $\gamma = 0.05$.

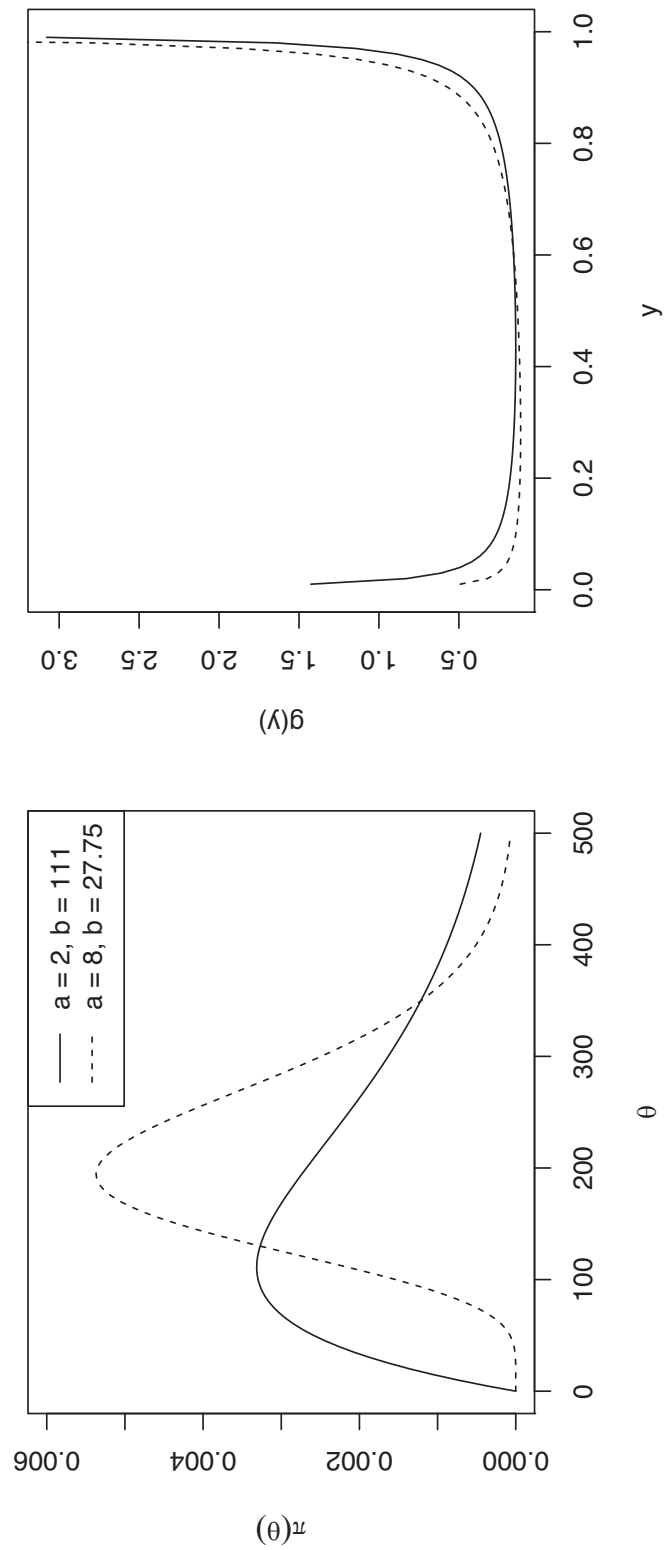

FIGURE 3. *Example B (head and neck cancer trial): plots of the gamma densities* $\pi(\theta)$ *with mean* $\theta_d = 222$ *(left panel) and of the corresponding* $g(y)$ *(right panel), for* $n = 50$, $\theta_0 = 100$, $\gamma = 0.05$.

tail interval $(95.8, 400.2)$. In this case, we have $\pi_0 = 0.03$ and $\pi_1 = 0.97$, and the resulting density function $g(y)$ (dashed line in Figure 3) shows a better behaviour.

3.3 Wald Test

Consider now a generic (non-normal) model with unknown one-dimensional parameter $\theta \in \mathbb{R}$. For the one-sided hypotheses (1), consider the Wald test based on the pivotal test statistic

$$Q(T, \theta_0) = \frac{T - \theta_0}{v_n(\theta_0)}, \tag{13}$$

where T is an estimator of θ that, for any generic $\theta \in \Theta$, is asymptotically normal $\mathrm{N}[\theta, v_n^2(\theta)]$ and where $v_n^2(\theta)$ is the asymptotic variance of T. If we consider regular models and we assume that T is the MLE of θ, then $v_n^2(\theta) = I_n(\theta)^{-1}$, where $I_n(\theta)$ is the expected Fisher information. Under regularity conditions, for iid data, we have that $I_n(\theta) = nI_1(\theta)$: Hence, $v_n^2(\theta) = \frac{1}{n} v_1^2(\theta)$, with $v_1^2(\theta) = I_1(\theta)^{-1}$. The asymptotic size $\gamma -$ Wald test rejects the null hypothesis if $Q(T, \theta_0) \geq z_{1-\gamma}$. Theorem 2 holds if (13) satisfies assumptions ***iii)*** and ***iv)***. Note that

$$Q_1^{-1}(u, \theta) = \theta + v_n(\theta)u, \tag{14}$$

whereas $Q_2^{-1}(t, w)$ does not have a general expression and must be determined in each specific example. In the following, we consider two cases: the Poisson and the Bernoulli model.

3.3.1 Poisson model

Assume that $X_1, \ldots, X_n|\theta$ are iid Pois(θ). In this case, $T(\mathbf{X}_n) = \overline{X}_n$ and its asymptotic variance is $v_n^2(\theta) = \theta/n$. It can be checked that

$$Q_1^{-1}(u, \theta) = \theta + \sqrt{\frac{\theta}{n}}u \qquad \text{and} \qquad Q_2^{-1}(t, w) = \left(\frac{-w + \sqrt{w^2 + 4nt}}{2\sqrt{n}}\right)^2, \tag{15}$$

where $Q_1^{-1}(u, \theta)$ follows from (14); the expression of $Q_2^{-1}(t, w)$ is derived in Appendix A. Therefore, from Theorem 2, it follows that

$$\eta^{-1}(y) = \frac{[h(y)+z_y]^2}{4n}$$

and that

$$g(y) = \pi\left(\frac{[h(y)+z_y]^2}{4n}\right)\left|\frac{[h(y)+z_y]^2}{2nh(y)\phi(z_y)}\right|, \tag{16}$$

where $h(y) = h(y, n, \theta_0, \gamma) = \sqrt{z_y^2 + 4n\left(\theta_0 + z_{1-\gamma}\sqrt{\frac{\theta_0}{n}}\right)}$. The proof is provided in Appendix A. In the following example, we consider a Ga(a, b) design prior, where the parameters a and b chosen so that Θ has expected value and variance respectively equal to m and v. It is easy to check that $a = \frac{m^2}{v}$ and $b = \frac{m}{v}$.

Example C. In the oral intervention study described in setup C, data relative to the dmft-index are assumed to be realisations of Poisson iid random variables whose unknown parameter θ is the average number of caries experienced in the population. In Lesaffre & Lawson (2012), a Ga($a = 3,\ b = 1$) based on historical data is considered as a prior for Θ, which corresponds to the information of a single sample observation. This prior has expected value equal to 3 and assigns probability 0.08 to Ω_0. Figure 4 shows the plot of the prior density (left panel, solid line) and the corresponding $g(y)$ for $n = 10$ (right panel, solid line). In this case $\mathbb{E}_\pi(Y) = 0.784$, whereas $\mathbb{M}_\pi(Y) = 0.987$: Hence, by using the expected value as a summary of $g(y)$, we slightly underestimate the PoS of the trial. As a second example, we consider a design prior with the same mean $\theta_d = 3$ and prior sample size 10 that is a Ga($a = 30,\ b = 0.1$) (see Figure 4, left panel, dashed line) corresponding to $\pi_0 \approx 0$. The resulting $g(y)$ (dashed curve) is now strictly increasing with almost coinciding summaries $\mathbb{E}_\pi(Y) = 0.982$ and $\mathbb{M}_\pi(Y) = 0.996$.

3.3.2 *Bernoulli model*

Assume that $X_1, \ldots, X_n|\theta$ are iid Ber(θ). In this case, $T(\mathbf{X}_n) = \overline{X}_n$ and its asymptotic variance is $v_n^2(\theta) = \dfrac{\theta(1 - \theta)}{n}$. From (14), it follows that

$$Q_1^{-1}(u, \theta) = \theta + \sqrt{\frac{\theta(1 - \theta)}{n}}u. \tag{17}$$

In Appendix B, we show that

$$Q_2^{-1}(t, w) = \begin{cases} \dfrac{2nt + w^2 - \operatorname{sgn}(w)\sqrt{w^4 + 4nt(1 - t)w^2}}{2(n + w^2)} & w \neq 0 \\ t w = 0, \end{cases} \tag{18}$$

where $\operatorname{sgn}(\cdot)$ is the sign function. Then, from Theorem 2, it follows that

$$\eta^{-1}(y) = \begin{cases} \dfrac{2nt_0 + z_{1-y}^2 + \operatorname{sgn}\left(y - \dfrac{1}{2}\right)h(y)}{2(n + z_{1-y}^2)} & y \neq \dfrac{1}{2} \\ t_0 & y = \dfrac{1}{2} \end{cases}, \tag{19}$$

where $t_0 = \theta_0 + \sqrt{\dfrac{\theta_0(1 - \theta_0)}{n}}z_{1-y}$ and $h(y) = \sqrt{z_{1-y}^4 + 4nt_0(1 - t_0)z_{1-y}^2}$. Finally, we have that, for $y \in (0, 1)$,

$$g(y) = \begin{cases} \pi\left(\dfrac{2nt_0 + z_{1-y}^2 + \operatorname{sgn}\left(y - \dfrac{1}{2}\right)h(y)}{2(n + z_{1-y}^2)}\right)\left|\dfrac{d}{dy}\eta^{-1}(y)\right| & y \neq \dfrac{1}{2} \\ 0 & y = \dfrac{1}{2} \end{cases}, \tag{20}$$

where

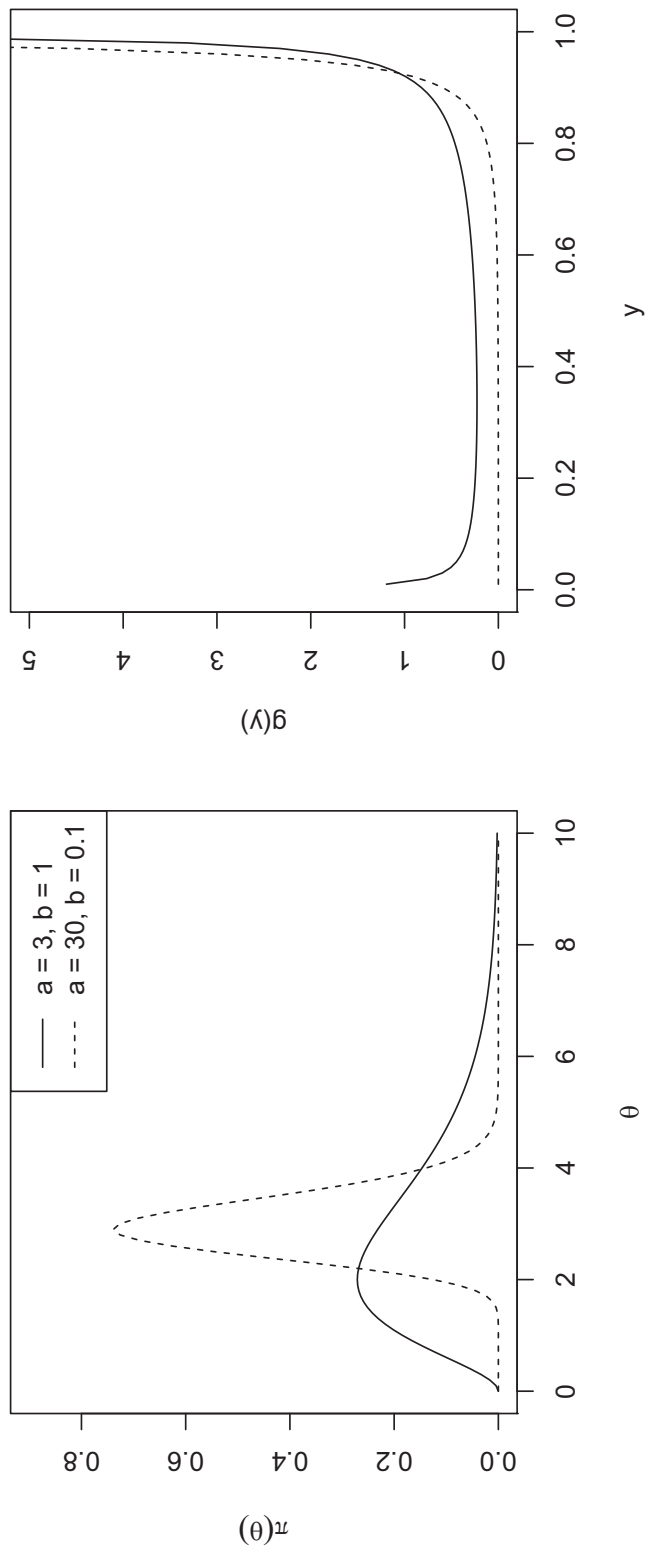

FIGURE 4. *Example C (oral health intervention): plots of the gamma densities $\pi(\theta)$ with mean $\theta_d = 3$ and prior sample sizes 1 and 10 (left panel, solid and dashed line, respectively); plots of the corresponding $g(y)$ for $n = 10$, $\theta_0 = 1$, $\gamma = 0.05$ (right panel).*

$$\frac{d}{dy}\eta^{-1}(y) = \frac{z_{1-y}}{(n+z_{1-y}^2)^2}\left[\left(-1+\text{sgn}\left(y-\frac{1}{2}\right)\frac{\phi(z_{1-y})}{2z_{1-y}}h'(y)\right)(n+z_{1-y}^2)+\right.$$
$$\left.+\left(2nt+z_{1-y}^2+\text{sgn}\left(y-\frac{1}{2}\right)h(y)\right)\right]$$

and where $h'(y) = -\dfrac{2z_{1-y}^3+4nt(1-t)z_{1-y}}{h(y)\,\phi(z_{1-y})}$. In the following example, we consider a beta design prior density, denoted by $\text{Be}(a, b)$, where the parameters a and b are chosen so that Θ has expected value and variance, respectively, equal to m and v. It is easy to check that $a = \dfrac{m[m(1-m)-v]}{v}$ and $b = \dfrac{(1-m)[m(1-m)-v]}{v}$.

Example D. In the rheumatoid arthritis trial of setup D, the binary outcome variable ACR20 is considered, and we are interested in testing whether θ is less than or equal to $\theta_0 = 0.2$ against the opposite alternative, with θ indicating the probability that a patient shows ACR20 equal to one. In doing so, we consider the design priors employed by O'Hagan *et al.* (2005) that are based on the performance of the new drug during the development phase and on the efficacy of related drugs. In the original example, the Authors propose to use a $\text{Be}(a, b)$ density with expected value $m = \theta_d = 0.4$ and variance $v = 0.17^2$, obtained by setting $a = 3$ and $b = 4.5$ (see Figure 5, left panel, solid line). The prior sample size is then $a + b = 7.5$ and $\pi_0 = 0.123$. For a sample size $n = 50$, the density function of the random power is plotted in the right panel of Figure 5 (solid line). Correspondingly, the expected value and the median of $g(y)$ are $\mathbb{E}_\pi(Y) = 0.691$ and $\mathbb{M}_\pi(Y) = 0.916$, respectively. For comparison, as a second prior, we consider a beta density with same expected value $m = \theta_d = 0.4$ but with a 50% of reduction in standard deviation with respect to the previous prior. It turns out that $a = 12.9$, $b = 19.3$, with a prior sample size ≈ 32 and $\pi_0 = 0.005$. In this case, $g(y)$ is strictly increasing almost everywhere (see Figure 5, right panel, dashed line) with expected value $\mathbb{E}_\pi(Y) = 0.833$ and median $\mathbb{M}_\pi(Y) = 0.934$. As in the previous examples, when the prior is not adequately concentrated and induces a u-shaped $g(y)$, the representativeness of PoS is questionable.

3.4 Sample size determination

The analysis of the distribution of Y has a practical application in the context of SSD, which is a crucial step of experimental design. In order to highlight the dependence of Y on the sample size, from now on, we use the notation $Y_n = \eta_n(\Theta)$. As before, we focus on $\mathbb{E}_\pi(Y_n)$ and $\mathbb{M}_\pi(Y_n)$. Noting that they are both non-decreasing sequences with respect to n, we define the following SSD criteria:

$$n^\star_{\text{PoS}} = \{n \in \mathbb{N}: \mathbb{E}_\pi(Y_n) > \lambda\}, \tag{21}$$

$$n^\star_{\text{med}} = \{n \in \mathbb{N}: \mathbb{M}_\pi(Y_n) > \lambda\}, \tag{22}$$

where $\lambda \in (0, 1)$. For comparison, we also consider the usual criterion based on the power function evaluated at a design-value θ_d, that is

$$n^\star_{\text{pow}} = \{n \in \mathbb{N}: \eta_n(\theta_d) > \lambda\}. \tag{23}$$

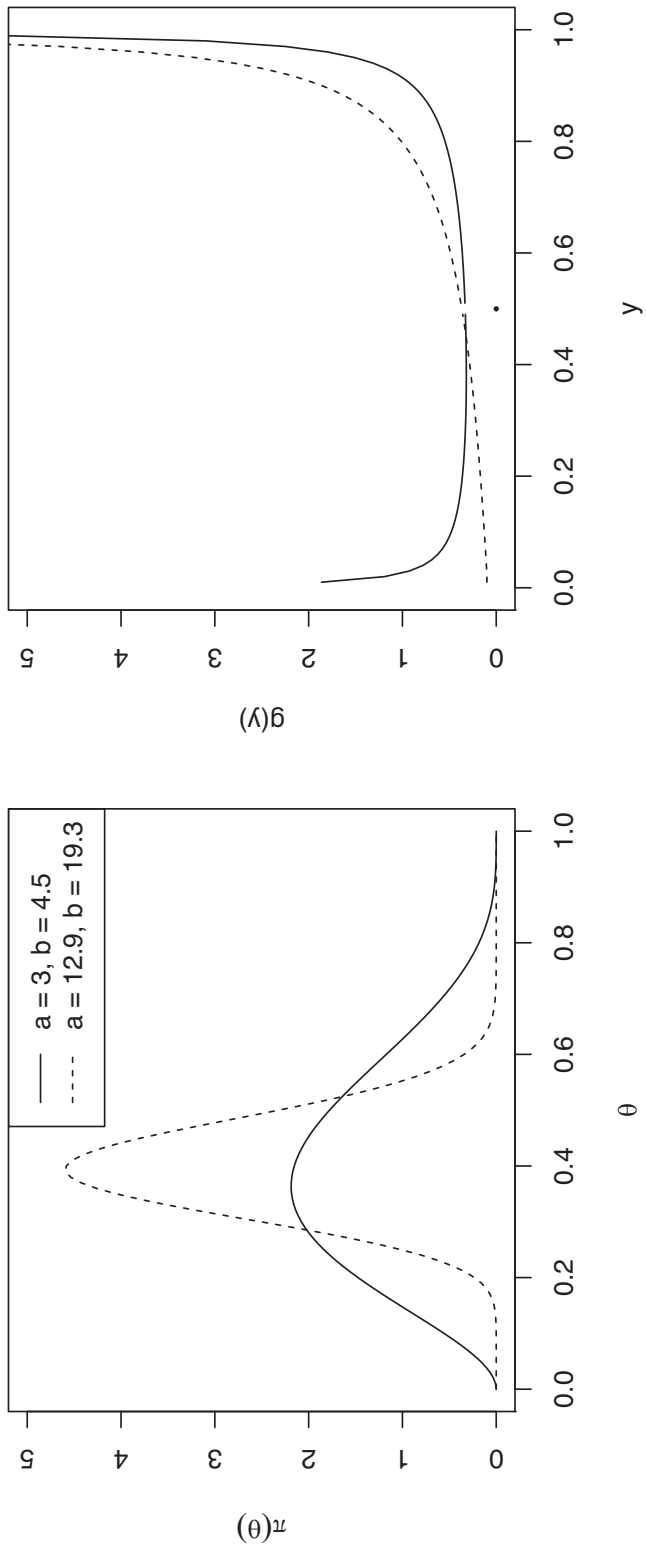

FIGURE 5. *Example D (rheumatoid arthritis): plots of the beta densities $\pi(\theta)$ with mean $\theta_d = 0.4$ and standard deviation 0.17 and 0.08 (left panel, solid and dashed line respectively); plots of the corresponding $g(y)$, for $n = 50$, $\theta_0 = 0.2$, $\gamma = 0.05$ (right panel).*

Notice that, $\forall \theta_d \in \Omega_1$ and $\forall \lambda \in (0,1)$, existence of $n^{\star}_{\text{pow}}$ is guaranteed by the convergence to 1 of $\eta_n(\theta_d)$, that is a well known property of the power function. Conversely, the choice of λ must take into account that

$$\lim_{n \longrightarrow \infty} \mathbb{E}_{\pi}(Y_n) = \pi_1$$

and that

$$\lim_{n \longrightarrow \infty} \mathbb{M}_{\pi}(Y_n) = \begin{cases} 0 & \pi_1 < \dfrac{1}{2} \\ 1 & \pi_1 \geq \dfrac{1}{2} \end{cases},$$

as shown in De Santis & Gubbiotti (2024b) and De Santis & Gubbiotti (2024a). Therefore, $n^{\star}_{\text{med}}$ exists finite for any $\lambda \in (0,1)$ provided that $\pi_1 > \dfrac{1}{2}$, as it is always the case in practical applications; whereas $n^{\star}_{\text{PoS}}$ exists finite as long as $\lambda < \pi_1$. In the examples, to ensure existence of $n^{\star}_{\text{PoS}}$ and a fair comparison with the optimal sample sizes based on the other critria, we set $\lambda = \delta \times \pi_1$, where $\delta \in (0,1]$ is the amount of power one is willing to reach at the chosen sample size.

Example D (continued). For illustration, we consider again setup D and we assume that we want to select the sample size for a new trial aimed at testing $\theta \leq \theta_0 = 0.2$, where θ is the probability that a patient has ACR20 is equal to one. Under the design prior assumptions of Example D in Section 3.3, we obtain the plots of $\mathbb{E}_{\pi}(Y_n)$, $\mathbb{M}_{\pi}(Y_n)$ and $\eta_n(\theta_d)$ as functions of n, shown in Figure 6. As expected, the most concentrated design prior (right panel), that induces a strictly increasing $g(y)$ almost everywhere (see again Figure 5), determines that the three curves are closer than in the other case (left panel), with $\mathbb{M}_{\pi}(Y_n)$ and $\eta_n(\theta_d)$ substantially coincident. Recall that, under the two design priors, π_1 is 0.877 and 0.995 respectively, that are the two limiting values of PoS, as Figure 6 illustrates. As mentioned before, these values are used in order to set the threshold $\lambda = \delta \times \pi_1$, for suitable choices of δ. Optimal sample sizes are reported in Table 2. Consistently, $n^{\star}_{\text{PoS}} > n^{\star}_{\text{med}} \approx n^{\star}_{\text{pow}}$, regardless of the considered prior. As δ (and therefore λ) increases, the optimal sample sizes get larger, for each criterion and for each prior.

In the example under consideration, curves $\mathbb{M}_{\pi}(Y_n)$ and $\eta_n(\theta_d)$ are remarkably close, despite their intrisic difference: On the one hand, $\eta_n(\theta_d)$ is just the power function evaluated at a single design-value, and therefore, it can be interpreted as the expected value or the median of the random power $\eta_n(\Theta)$ for a design prior that assigns probability one to θ_d; on the other hand, $\mathbb{M}_{\pi}(Y_n)$ is the median of $\eta_n(\Theta)$ with respect to a non–point-mass design prior. It is interesting to note that, if we keep the same design prior mean, a significant discrepancy between the two curves can be observed only if we force the standard deviation to be fairly large. As an example, we consider an additional beta prior where θ_d is still 0.4, but we double the standard deviation, so that $\pi_0 = 0.397$. Figure 7 (right panel) shows that $\mathbb{M}_{\pi}(Y_n)$ and $\eta_n(\theta_d)$ are now notably separated at least for values of n smaller than 150. However, the resulting prior is the u-shaped beta density of Figure 7 (left panel) that can hardly be accepted as a prior describing the design goal of the trial. As a final remark, the values of PoS are dramatically smaller than those of $\mathbb{M}_{\pi}(Y_n)$ and $\eta_n(\theta_d)$, as an obvious consequence of the large probability assigned to Ω_0.

4 Discussion

In their seminal article on the (Bayesian) probability distribution of the (frequentist) power function, Rufibach *et al.* (2016) recommend "*to always look at the density* $g(\cdot)$ *when*

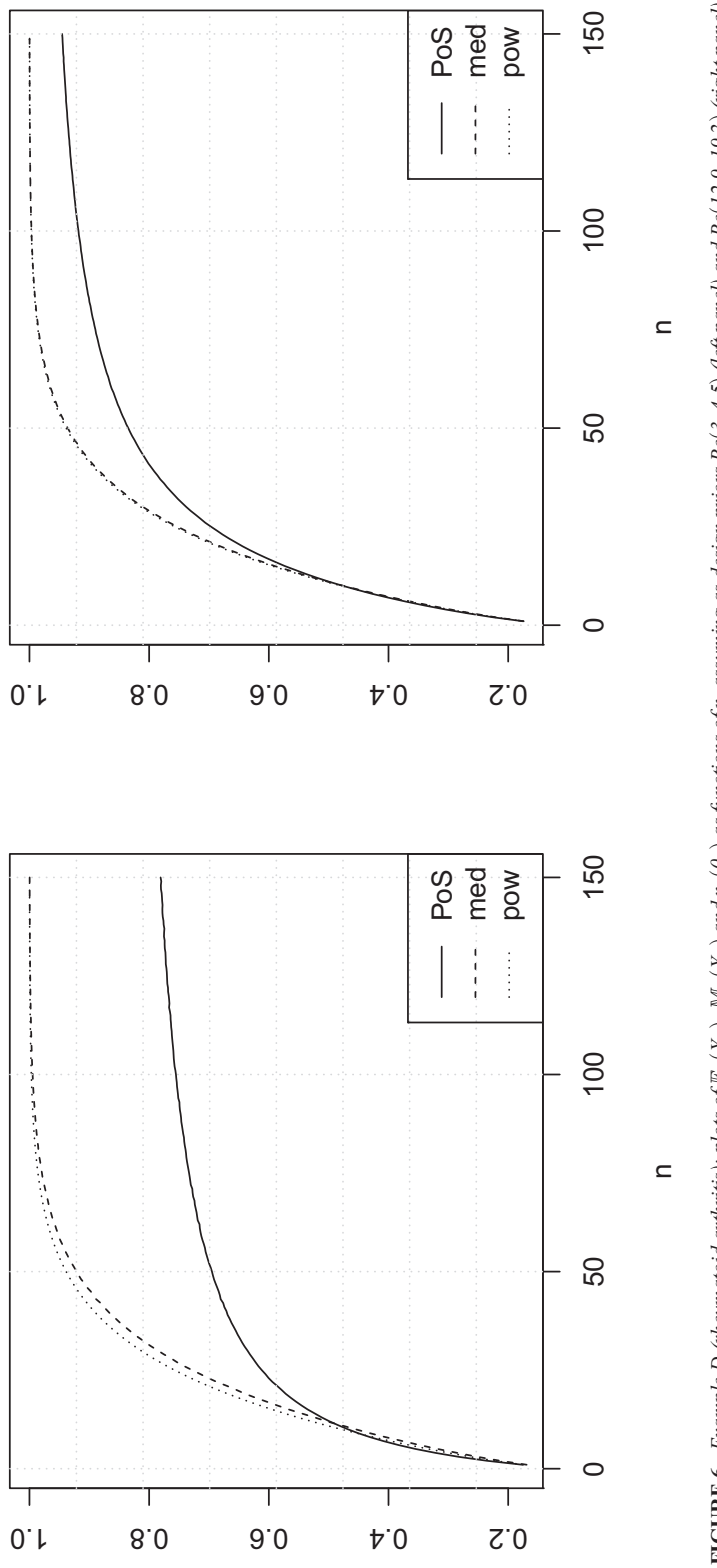

FIGURE 6. *Example D (rheumatoid arthritis): plots of* $\mathbb{E}_\pi(Y_n)$, $\mathbb{M}_\pi(Y_n)$ *and* $\eta_n(\theta_d)$ *as functions of n, assuming as design priors* $Be(3, 4.5)$ *(left panel) and* $Be(12.9, 19.3)$ *(right panel).*

Table 2. *Example D (rheumatoid arthritis): Optimal sample sizes* $n^{\star}_{\text{pow}}$, $n^{\star}_{\text{med}}$ *and* $n^{\star}_{\text{PoS}}$ *using the threshold* $\lambda = \delta \times \pi_1$ *in Equations (21–23) for several values of* δ.

	Be($a = 3$, $b = 4.5$)				Be($a = 12.9$, $b = 19.3$)			
δ	λ	$n^{\star}_{\text{pow}}$	$n^{\star}_{\text{med}}$	$n^{\star}_{\text{PoS}}$	λ	$n^{\star}_{\text{pow}}$	$n^{\star}_{\text{med}}$	$n^{\star}_{\text{PoS}}$
0.7	0.614	17	18	26	0.696	21	22	25
0.8	0.702	22	24	54	0.796	29	29	40
0.9	0.790	28	31	178	0.895	41	42	79

discussing PoS" and to study the evolution of this distribution as the sample size increases. Following this spirit, that is acknowledging the importance of determining the closed-form distribution of the power function, in the present article, we provide a general expression for cdf and pdf of the power function for "pivotal" test statistics. Our main contributions are the following.

1. *Formal results*. Theorem 2 provides explicit expressions of $G(\cdot)$ and $g(\cdot)$ in terms of the inverse functions $Q_1^{-1}(\cdot, \theta)$ and $Q_2^{-1}(t, \cdot)$ of the the pivotal quantity $Q(t, \theta)$. Since a large range of statistical tests are indeed based on pivotal quantities, these results have wide applicability, as we exemplify in the article. Casella & Berger (2002) observe that "perhaps one of the most elegant methods of constructing set estimators and calculating coverage probabilities is the use of pivotal quantities". In a certain sense, we feel that our result is an additional reflection of this formal elegance transferred to the testing context and, specifically, to the structural properties of the distribution of the random power.
2. *Biomedical examples and discussion on* PoS. We have exploited the formal results of Theorem 2 in a series of standard yet relevant testing problems illustrated with classical examples from the biomedical literature. The goal is to show the usefulness of exploring the whole distribution of the power function instead of considering only its expected value. In this regard, we have stressed the crucial relevance of the choice of the design prior in determining the shape of $g(y)$ and, ultimately, the representativeness of PoS as a synthesis of this density. Specifically, we have pointed out the discrepancy between the expected value and the median of the distribution of the random power, due to the unfortunate u-shape of $g(y)$ that is observed when the precision of the design prior is not sufficiently high.
3. *SSD*. One of the main uses of the power function in practice is for sample size selection. The SSD criterion based on PoS has been proposed as an alternative to the criterion based on the standard power in order to avoid its local optimality. However, as discussed in the present article and elsewhere, PoS may present some drawbacks, such as its lack of representativeness (see previous comment) and the fact that its limiting value (as n increases) is stricly less than one when the design prior assigns positive probability to $\boldsymbol{\Omega}_0$, a situation that is not so infrequent in real clinical trials (see, e.g., Temple & Robertson, 2021; Crisp *et al.*, 2018). In the examples of Section 3.4, we have shown how the sample sizes selected using PoS can be strikingly different from those obtained from the criteria based on the standard power and the median of $g(\cdot)$ that, conversely, in many cases are remarkably close.

This work can be extended in several directions. Some of them are the following.

- *Two-sided testing*. In the present article, we focus uniquely on the one-sided testing problem, as in (Rufibach *et al.*, 2016) and in De Santis & Gubbiotti (2024a). One first natural extension would be addressing the case of a point-null hypothesis against a two-sided alternative. This is a common testing problem for which, so far, explicit formulas for PoS and the distribution of the random power function have not been provided. Even though the analytical

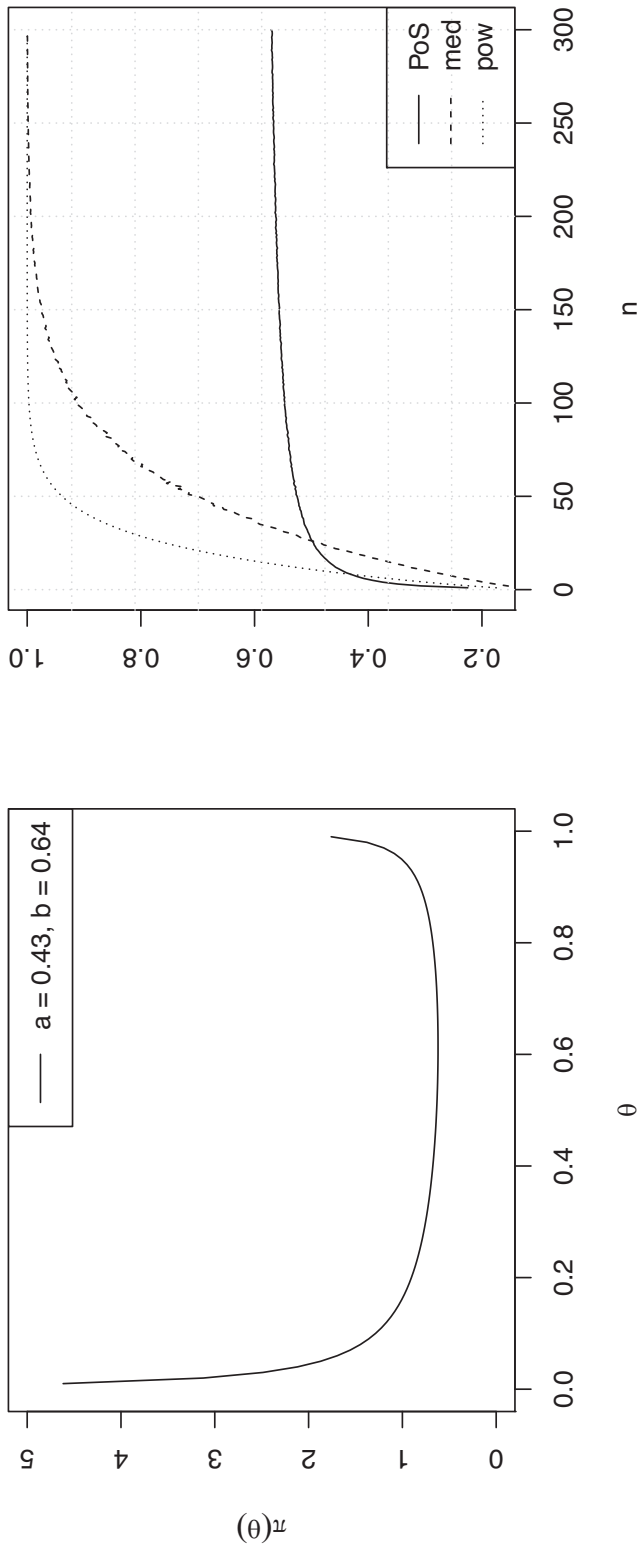

FIGURE 7. *Example D (rheumatoid arthritis): plot of* $\mathbb{E}_\pi(Y_n)$, $\mathbb{M}_\pi(Y_n)$ *and* $\eta_n(\theta_d)$ *as functions of n (right panel), using the* $\mathrm{Be}(0.43, 0.64)$ *design prior (left panel).*

problem for the two-sided testing case is substantially different from the one-sided test and, in a certain sense, less straightforward, we believe that it can still be tractable at least in several standard yet relevant models, such as the normal location and scale problems.

- *Simulation-based approach*. One characterizing feature of the present article is that it provides closed-form formulas for the density of the random power function in several common models. One exception is the unknown variance case for the normal model (Section 3.1), where we have exploited Monte Carlo simulation. In general, as soon as models get more complex than those used here for illustration, we expect that analytical expressions are not available any longer and that simulation is the only way for proceeding. We believe that the availability of a general simulation-based approach, potentially endowed with a user-friendly Shiny app, would significantly widen its applicability.
- *Bayesian test statistic*. As mentioned in Section 1, in the present article, we follow a specific hybrid frequentist-Bayesian approach: It is frequentist in the use of the test statistics and power function; it is Bayesian in exploiting a prior to evaluate the power function and for SSD. This approach can also be employed replacing the frequestist test with a Bayesian testing procedure and properly defining `success` in Bayesian terms as, for instance, in Brutti *et al.* (2008). This fully Bayesian procedure demands an additional prior distribution (*analysis prior*) that incorporates the information on the quantity of interest to be merged with experimental data via the Bayes theorem. In this case, the interplay between design prior and analysis prior and their impact on the distribution of the Bayesian power should be carefully assessed.
- *Other designs*. The use of PoS has been proposed also for designs that are more complex that those proposed in the present article. As an example, Grieve (2024) extends the definition of PoS for group sequential designs. Even in this case, we believe that the analysis of the density of the random power could provide a relevant insight in the evaluation of the PoS of the trial. Specifically, one should consider the power function of a test statistic based on the interim treatment effect statistic and then study its distribution (induced by a design prior). Of course, the model is now more intricate and availability of pivotal test statistics and closed-form expressions of PoS and its density is not guaranteed.

We hope to elaborate on these topics in future research.

Acknowledgements

This research was funded by Sapienza University of Rome, Italy.
Open access publishing facilitated by Universita degli Studi di Bologna, as part of the Wiley - CRUI-CARE agreement.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

Brutti, P., De Santis, F. & Gubbiotti, S. (2008). Robust Bayesian sample size determination in clinical trials. *Statistics in Medicine*, **27**, 2290–2306.

Brutti, P., De Santis, F. & Gubbiotti, S. (2014). Bayesian-frequentist sample size determination: a game of two priors. *Metron*, **72**, 133–151.

Casella, G. & Berger, R.L. (2002). *Statistical inference Edited by* 2nd. Duxbury Press.

Crisp, A., Miller, S., Thompson, D. & Best, N. (2018). Practical experiences of adopting assurance as a quantitative framework to support decision making in drug development. *Pharmaceutical Statistics*, **17**, 317–328.

De Santis, F. & Gubbiotti, S. (2024a). On the distribution of the power function for the scale parameter of exponential families. *Statistics in Medicine*, **43**, 1973–1992.
De Santis, F. & Gubbiotti, S. (2024b). On the limit distribution of the power function induced by a design prior. *Statistical Papers*, **65**, 1927–1945.
Efron, B. (1988). Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association*, **83**, 414–425.
Grieve, A.P. (2024). Probability of success and group sequential designs. *Pharmaceutical Statistics*, **23**, 185–203.
Hoshyarmanesh, H., Karami, A. & Mohammadpour, A. (2016). Confidence intervals for the scale parameter of exponential family of distributions. *The American Statistician*, **70**, 134–137.
Karlin, S. & Rubin, H. (1956). The theory of decision procedures for distributions with monotone likelihood ratio. *Annals of Mathematical Statistics*, **27**, 272–299.
Kunzmann, K., Grayling, M.J., Lee, K.M., Robertson, D.S., Rufibach, K. & Wason, J.M.S. (2021). A review of Bayesian perspectives on sample size derivation for confirmatory trials. *The American Statistician*, **75**(4), 424–432.
Lesaffre, E. & Lawson, A.B. (2012). *Bayesian biostatistics*. Wiley.
Liu, F. (2010). An extension of Bayesian expected power and its application in decision making. *Journal of Biopharmaceutical Statistics*, **20**, 941–953.
Mariani, F., De Santis, F. & Gubbiotti, S. (2024). The distribution of power-related random variables (and their use in clinical trials). *Statistical Papers*, **65**, 5555–5574.
O'Hagan, A., Stevens, J.W. & Campbell, M.J. (2005). Assurance in clinical trial design. *Pharmaceutical Statistics*, **4**, 187–201.
Rufibach, K., Burger, H.U. & Abt, M. (2016). Bayesian predictive power: choice of prior and some recommendations for its use as probability of success in drug development. *Pharmaceutical Statistics*, **15**, 438–446.
Sahu, S. & Smith, T.M.F. (2006). A Bayesian method of sample size determination with practical applications. *Journal of the Royal Statistical Society A*, **169**, 235–253.
Spiegelhalter, D.J., Abrams, K.R. & Myles, J.P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Wiley.
Temple, J. & Robertson, J. (2021). Conditional assurance: the answer to the questions that should be asked within drug development. *Pharmaceutical Statistics*, **20**, 1102–1111.
Vanobbergen, J., Martens, L., Lesaffre, E. & Declerck, D. (2001). Caries prevalence in belgian children: a review. *International Journal of Paediatric Dentistry*, **11**, 164–170.
Wang, F. & Gelfand, A.E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, **17**, 193–208.

Appendix A: Wald Test—Poisson Example

Consider the Wald pivotal quantity for the Poisson model, $Q(T, \theta) = \dfrac{\sqrt{n}(T - \theta)}{\sqrt{\theta}}$.

A.1 Determination of the Inverse $Q_2^{-1}(t, w)$

In order to determine $Q_2^{-1}(t, w)$, we need to solve the following equation with respect to θ:

$$\frac{\sqrt{n}(t - \theta)}{\sqrt{\theta}} = w, \qquad \theta > 0$$

Note that this can be rewritten as

$$\sqrt{n}\theta + w\sqrt{\theta} - \sqrt{n}t = 0,$$

and setting $\lambda = \sqrt{\theta}$, with $\lambda > 0$, we have the following quadratic equation in λ

$$\sqrt{n}\lambda^2 + w\lambda - \sqrt{n}t = 0$$

that admits the following two solutions:

$$\lambda_{1,2} = \frac{-w \pm \sqrt{w^2 + 4nt}}{2\sqrt{n}}.$$

It is straightforward to check that $\lambda_1 < 0$. Hence, we only retain λ_2, and we conclude

$$\theta = \lambda_2^2 = \left(\frac{-w + \sqrt{w^2 + 4nt}}{2\sqrt{n}}\right)^2,$$

which yields the expression of $Q_2^{-1}(t, w)$.

A.2 Determination of $\eta^{-1}(y)$

Using the expressions of $Q_1^{-1}(u, \theta)$ and $Q_2^{-1}(t, w)$ given in Equation (15), it is straightforward to find $\eta^{-1}(y)$ from Equation (5), as follows

$$\begin{aligned}\eta^{-1}(y) &= Q_2^{-1}\left[Q_1^{-1}(z_{1-\gamma}, \theta_0), z_{1-y}\right] = Q_2^{-1}\left[\theta_0 + z_{1-\gamma}\sqrt{\frac{\theta_0}{n}}, z_{1-y}\right] \\ &= \frac{1}{4n}\left[z_y + \sqrt{z_y^2 + 4n\left(\theta_0 + z_{1-\gamma}\sqrt{\frac{\theta_0}{n}}\right)}\right]^2 = \frac{[h(y)+z_y]^2}{4n},\end{aligned}$$

where $h(y) = \sqrt{z_y^2 + 4n\left(\theta_0 + z_{1-\gamma}\sqrt{\frac{\theta_0}{n}}\right)}$.

A.3 Derivative of $\eta^{-1}(y)$ wrt y

Taking the derivative of $\eta^{-1}(y)$ wrt y, we have

$$\begin{aligned}\frac{d}{dy}\eta^{-1}(y) &= \frac{2}{4n}[z_y + h(y)]\left[\frac{1}{\phi(z_y)} + \frac{1}{2}[h(y)]^{-1}\frac{2z_y}{\phi(z_y)}\right] = \\ &= \frac{1}{2n}[z_y + h(y)]\left[\frac{1 + [h(y)]^{-1}z_y}{\phi(z_y)}\right] = \\ &= \frac{[h(y)+z_y]^2}{2nh(y)\phi(z_y)}.\end{aligned}$$

The expression of $g(y)$ given in Equation (16) follows.

Appendix B: Wald Test—Bernoulli Example

Consider the Wald pivotal quantity for the Binomial model, $Q(T, \theta) = \dfrac{\sqrt{n}(T - \theta)}{\sqrt{\theta(1 - \theta)}}$.

B.1 Existence of the Inverse $Q_2^{-1}(t, w)$

To prove the existence of the inverse function $Q_2^{-1}(t, w)$, we show that $Q(t, \theta)$ is a decreasing function of $\theta \in (0, 1)$, for any $t = T(\mathbf{x}_n) = \bar{x}_n \in (0, 1)$. To make derivation easier, we take the logarithm of $Q(t, \theta)$ and consider the two cases: (i) $\theta < t$, (ii) $\theta > t$.

For Case (i), we have

$$\ln(Q(t, \theta)) = \ln\left(\sqrt{n}\right) + \ln(t - \theta) - \frac{1}{2}\ln(\theta) - \frac{1}{2}\ln(1 - \theta), \quad 0 < \theta < t.$$

Then, taking the derivative wrt θ, we have

$$\begin{aligned}\frac{d}{d\theta}\ln(Q(t, \theta)) &= -\frac{1}{t - \theta} - \frac{1}{2\theta} + \frac{1}{2(1 - \theta)} \\ &= \frac{-2\theta(1 - \theta) - (1 - \theta)(t - \theta) + \theta(t - \theta)}{2\theta(1 - \theta)(t - \theta)} \\ &= \frac{2\theta t - \theta - t}{2\theta(1 - \theta)(t - \theta)}.\end{aligned}$$

Note that if $t > \dfrac{1}{2}$, then $2\theta t - \theta - t < 0 \Leftrightarrow \theta < \dfrac{t}{2t - 1}$, which is always true since $\dfrac{t}{2t - 1} > 1$; conversely, if $t < \dfrac{1}{2}$, then $2\theta t - \theta - t < 0 \Leftrightarrow \theta > \dfrac{t}{2t - 1}$, which is always true since $\dfrac{t}{2t - 1} < 0$. In summary, $\forall \theta \in (0, t)$, $2\theta t - \theta - t < 0$; moreover, since $\theta > 0$, $1 - \theta > 0$ and $t - \theta > 0$, $\dfrac{d}{d\theta}\ln(Q(t, \theta)) < 0$ for $\theta \in (0, t)$. As a consequence, $\ln(Q(t, \theta))$ and $Q(t, \theta)$ are both decreasing functions of $\theta \in (0, t)$.

Case (ii) follows similarly.

B.2 Determination of the Inverse $Q_2^{-1}(t, w)$

In order to determine $Q_2^{-1}(t, w)$, we need to solve the following irrational equation wrt θ:

$$\frac{\sqrt{n}(t - \theta)}{\sqrt{\theta(1 - \theta)}} = w, \qquad \theta \in (0, 1)$$

that can be rewritten as

$$\sqrt{n}(t - \theta) = w\sqrt{\theta(1 - \theta)}.$$

Hence, raising to the square both members of the previous equation, we have

$$n(t - \theta)^2 = w^2\theta(1 - \theta),$$

whose solutions are as follows:

$$\theta_{1,2} = \frac{(2nt + w^2) \pm \sqrt{w^4 + 4nt(1 - t)w^2}}{2(n + w^2)}.$$

Monotonicity of $Q(t, \theta)$ with respect to θ—shown in the previous subsection B.1—implies that the above equation admits a unique solution. Furthermore, note that if $w > 0$, then $\theta < t$; if $w < 0$, then $\theta > t$; finally, if $w = 0$, then $\theta = t$. Consequently, we obtain the expression of $Q_2^{-1}(t, w)$ given in Equation (18).

B.3 Determination of $\eta^{-1}(y)$

Using the expressions of $Q_1^{-1}(u, \theta)$ and $Q_2^{-1}(t, w)$ given, respectively, in Equations (17) and (18) and recalling that $z_{1-y} > 0$ if and only if $y < \frac{1}{2}$, it is straightforward to find $K(y)$ from Equation (5), as follows:

$$\eta^{-1}(y) = Q_2^{-1}\left[Q_1^{-1}(z_{1-\gamma}, \theta_0), z_{1-y}\right] = Q_2^{-1}\left[\theta_0 + \sqrt{\frac{\theta_0(1 - \theta_0)}{n}} z_{1-\gamma}, z_{1-y}\right] =$$

$$= \begin{cases} \dfrac{2nt_0 + z_{1-y}^2 - h(y)}{2(n + z_{1-y}^2)} & y < \dfrac{1}{2} \\ t_0 & y = \dfrac{1}{2} \\ \dfrac{2nt_0 + z_{1-y}^2 + h(y)}{2(n + z_{1-y}^2)} & y > \dfrac{1}{2} \end{cases},$$

where $t_0 = \theta_0 + \sqrt{\frac{\theta_0(1 - \theta_0)}{n}} z_{1-\gamma}$ and $h(y) = \sqrt{z_{1-y}^4 + 4nt_0(1 - t_0)z_{1-y}^2}$, which yields Equation (19).

B.4 Derivative of $\eta^{-1}(y)$ wrt y

In order to take the derivative of $\eta^{-1}(y)$ wrt y, consider first the derivative of the function $h(y)$, that is, $h'(y) = -\frac{2z_{1-y}^3 + 4nt(1 - t)z_{1-y}}{h(y)\phi(z_{1-y})}$. Then, for $y < \frac{1}{2}$, it is

$$\frac{d}{dy}\eta^{-1}(y) = \frac{1}{4(n + z_{1-y}^2)^2}\left[\left(-\frac{2z_{1-y}}{\phi(z_{1-y})} - h'(y)\right)(2n + 2z_{1-y}^2) - \left(2nt + z_{1-y}^2 - h(y)\right)\left(-\frac{4z_{1-y}}{\phi(z_{1-y})}\right)\right]$$

$$= \frac{z_{1-y}}{(n + z_{1-y}^2)^2\phi(z_{1-y})}\left[\left(-1 - \frac{\phi(z_{1-y})}{2z_{1-y}}h'(y)\right)(n + z_{1-y}^2) + \left(2nt + z_{1-y}^2 - h(y)\right)\right].$$

Similarly, for $y > \frac{1}{2}$, it is

$$\frac{d}{dy}\eta^{-1}(y) = \frac{1}{4(n+z_{1-y}^2)^2}\left[\left(-\frac{2z_{1-y}}{\phi(z_{1-y})}+h'(y)\right)(2n+2z_{1-y}^2) - \left(2nt+z_{1-y}^2+h(y)\right)\left(-\frac{4z_{1-y}}{\phi(z_{1-y})}\right)\right]$$
$$= \frac{z_{1-y}}{(n+z_{1-y}^2)^2\phi(z_{1-y})}\left[\left(-1+\frac{\phi(z_{1-y})}{2z_{1-y}}h'(y)\right)(n+z_{1-y}^2)+\left(2nt+z_{1-y}^2+h(y)\right)\right].$$

Finally note that if $y = \frac{1}{2}$, then $\eta^{-1}(y) = t_0$ which implies $\frac{d}{dy}\eta^{-1}(y) = 0$. The expression of $g(y)$ given in Equation (20) follows.

[*Received February 2025; accepted February 2026*]