Enhancing Clinical Trial Analysis: The Role of Bayes Factors in Evaluating Drug Efficacy

Abstract

Clinical trials are essential for evaluating new medical interventions. However, challenges such as increasing costs and enrolment difficulties often result in inadequate sample sizes. This makes it challenging to obtain reliable efficacy assessments using traditional frequentist methods. Bayesian statistics provides a superior solution to this problem by integrating prior knowledge and managing uncertainty effectively with limited data. In particular, Bayes factor serves as a key metric, offering several advantages over the conventional p-value: 1) explicit incorporation of prior evidence, 2) balancing the assessment of the null and alternative hypotheses, and 3) providing an intuitive probabilistic interpretation of the results. Using a theoretical framework and an empirical case study of acute inflammatory diseases of peripheral nerves, this paper shows that this method can enable early trial termination decisions in small samples and generate more robust efficacy evidence. This approach improves the efficiency of clinical trials and provides new ideas for future regulatory decision-making and post-market drug monitoring.

Key words: prior distributions; Bayesian hypothesis tests; Bayes factor; randomized clinical trials; drug efficacy.

1 Introduction

Randomized controlled trials (RCTs) are widely regarded as the gold standard for evaluating the safety and efficacy of new medical interventions, playing a crucial role in optimizing treatment strategies and improving patient outcomes. However, they often entail high costs, making efficient data utilization and rigorous analysis essential. In this context, Bayesian methods offer distinct advantages by systematically integrating prior evidence and enabling iterative model refinement through sequential data assimilation [1].

Bayesian methods have gained increasing prominence in medical research in recent years. Within this framework, prior distribution specification constitutes a critical methodological consideration. Johnson et al. systematically evaluated approaches for deriving prior distributions, assessing their validity, reliability, and applicability in clinical research [2]. Quintana et al. illustrated the utility of Bayesian methods in clinical trial monitoring through three case studies employing conjugate prior distributions, demonstrating their role in facilitating early stopping decisions via Bayesian predictive probabilities [3]. Berry formalized the theoretical foundations of Bayesian clinical trials while quantifying their potential to enhance drug development efficiency [4]. Sugitani et al. introduced a biomarker-driven Bayesian randomized clinical trial design to refine target population identification and optimize treatment effect evaluation, thereby advancing precision medicine [5].

As a pivotal methodological tool in Bayesian statistics, the Bayes factor has emerged as a robust approach for model comparison and hypothesis testing in clinical trial design [6]. Distinct from conventional null hypothesis significance testing that relies on p-value thresholds, this Bayesian framework provides direct quantification of the evidence strength supporting specific hypotheses [7]. Whereas traditional frequentist methods emphasize binary hypothesis rejection, Bayes factors facilitate comparative assessment of competing models, thereby offering a probabilistic interpretation of how empirical data update prior scientific

knowledge.

The growing adoption of Bayesian methods in pharmacological research has provided a new framework for therapeutic evaluation, particularly through the application of Bayes factors in comparative effectiveness research. This analytical approach offers distinct advantages over traditional frequentist methods by allowing quantitative assessment of competing hypotheses while accounting for uncertainty in the evidence.

Recent applications have highlighted the utility of Bayes factors in pharmacological research [8]. Monden et al. [9] employed hierarchical Bayes factor meta-analysis to evaluate randomized controlled trials of second-generation antidepressants, revealing substantial pharmacological heterogeneity among these treatments. Furthermore, Richard et al. [10] conducted a comprehensive Bayesian reanalysis of anti-amyloid beta (Aβ) immunotherapy trials. Their analysis provided robust evidence supporting the null hypothesis, statistically confirming the absence of clinical efficacy in Aβ-targeted interventions. In addition, Pittelkow et al. [11] extended this analytical approach to oncology drug approvals; their results indicated that the majority of newly approved anticancer drugs lacked robust evidence demonstrating improvements in overall survival, with some agents even showing no evidence of efficacy whatsoever. The methodology provides clinicians with a nuanced decision-making framework that integrates empirical evidence with prior biological plausibility. Furthermore, it enables regulatory bodies to stratify drug approvals based on gradations of evidentiary support, potentially enhancing post-marketing surveillance strategies.

The remainder of this paper is organized as follows: Section 2 discusses key considerations for selecting appropriate Bayesian priors in clinical research applications. Section 3 provides a comprehensive overview of the Bayes factor, detailing its theoretical foundations and interpretation framework. Section 4 presents a comparative analysis between Bayes factor approaches and conventional null hypothesis significance testing (NHST), examin-

ing their respective strengths and limitations. Sections 5 demonstrate concrete applications through detailed case studies comparing Bayes factors with traditional hypothesis testing for discrete clinical trial endpoints. Finally, Section 6 concludes with a discussion of the broader implications of Bayes factor methodology.

2 The selection of appropriate prior

This section provides a systematic examination of prior distribution specification in Bayesian analysis, with particular emphasis on its critical role in clinical research applications. We present a comprehensive overview of commonly employed prior distributions and their methodological implications. In Bayesian inference, prior selection constitutes a fundamental analytical decision that determines both the posterior distribution's properties and the validity of subsequent statistical conclusions. Notably, the choice of prior distribution exerts two principal effects: (1) it directly impacts the numerical precision of Bayes factor computation, and (2) it fundamentally governs the reliability of comparative model assessment [12].

The specification of appropriate prior distributions for model parameters constitutes a fundamental consideration in Bayes factor computation. Kass and Wasserman (1996) conducted a comprehensive review of formal methods for prior selection and examined the associated practical and philosophical challenges [13]. Current methodological approaches offer several systematic solutions to this critical issue as follows.

(1) Informative Priors: When substantial domain knowledge or historical data exists, researchers may construct informative priors that explicitly incorporate this available information. This approach is particularly valuable in clinical research settings where reliable prior evidence exists.

(2) Non-informative Priors: In situations with limited or ambiguous prior information, non-informative priors provide a principled alternative that minimizes subjective influence

on inferential outcomes.

(3) Conjugate Priors: From a theoretical perspective, when the sampling distribution is known, selecting priors that are conjugate to the likelihood function offers dual advantages: (a) they maintain theoretical coherence with the underlying statistical model, and (b) they yield computationally tractable posterior distributions while ensuring methodological rigor in parameter estimation.

2.1 Uniform prior

In Bayesian analysis, the uniform prior distribution represents a canonical non-informative prior that formalizes the state of complete prior ignorance about a parameter of interest. This prior assigns equal probability density to all possible values within the parameter space. For a parameter θ constrained to a finite interval $[a, b] \subset R$, the uniform prior is formally expressed as: $\theta \sim U[a, b]$ with probability density function given by

$$\pi(\theta) = \frac{1}{b-a}, \quad \theta \in [a, b].$$

While the uniform prior offers notable advantages in computational tractability and apparent objectivity, it presents several important limitations: (1) potential impropriety when applied to unbounded parameter spaces, (2) implicit assumptions about parameter scaling that may introduce bias, and (3) dependence on arbitrary bounds in finite intervals. These characteristics make uniform priors most appropriate for preliminary analyses or situations with abundant data where prior sensitivity is minimized.

2.2 Jeffreys' prior

The Jeffreys' prior [14] is distinguished by its invariance property, ensuring consistency under arbitrary reparameterizations. This characteristic solves a fundamental limitation of uniform priors: while a uniform prior may appear non-informative for parameter θ, its form is not

preserved under nonlinear transformations $g(\theta)$, potentially introducing unintended informativeness. Jeffreys resolved this through a theoretically rigorous approach using the Fisher information matrix $I(\theta \mid \boldsymbol{y})$, yielding a prior that maintains its non-informative character across all smooth monotone transformations. The Jeffreys' prior is defined as:

$$\pi(\theta) \propto det\big(I(\theta \mid \boldsymbol{y})\big)^{\frac{1}{2}} \tag{2.1}$$

where $I(\theta \mid \boldsymbol{y})$ denotes the Fisher information matrix of the parameter θ, and $det(\cdot)$ represents the matrix determinant operator.

Jeffreys' prior preserves the theoretically valuable property of invariance under parameter transformations, yet its practical implementation presents several challenges. The requirement to compute model-specific Fisher information matrices can create: (i) substantial computational burdens, particularly for high-dimensional parameters, and (ii) limitations in complex modeling frameworks where analytical solutions may be unavailable. Despite these constraints, Jeffreys' prior remains a principled choice for objective Bayesian analysis when: (1) substantive prior knowledge is absent, (2) a neutral reference prior is preferred to minimize subjective influence, and (3) analyzing small datasets where prior specification has disproportionate impact. In these scenarios, Jeffreys' prior enhances the robustness of inferential conclusions while maintaining theoretical coherence, making it particularly valuable for exploratory analyses and sensitivity assessments.

2.3 Reference prior

Another prominent class of non-informative priors is the reference prior, introduced by Bernardo (1979) as a modification of Jeffreys' approach. A key distinction of this method is its explicit separation between parameters of interest and nuisance parameters. The central idea of the reference prior is to maximize the Kullback-Leibler (K-L) divergence between the prior and posterior distributions upon observing the data, thereby minimizing the impact of

prior information [15, 16]. This ensures that the posterior distribution is primarily driven by the data rather than subjective prior assumptions.

Let $\boldsymbol{D} = (y_1, \cdots, y_n)$ be an independent and identically distributed (i.i.d.) random sample with observed realization $\boldsymbol{y}$. The Kullback-Leibler (K-L) divergence between the posterior and prior distributions is defined as follows

$$KL(\pi(\boldsymbol{\theta} \mid \boldsymbol{D}), \pi(\boldsymbol{\theta})) = \int_{\Theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{D}) log \frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{D})}{\pi(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta} \tag{2.2}$$

Let $I_{\pi(\boldsymbol{\theta})}(\boldsymbol{\theta}, \boldsymbol{y})$ be the expected K-L divergence with respect to $\boldsymbol{y}$, that is

$$I_{\pi(\boldsymbol{\theta})}(\boldsymbol{\theta}, \boldsymbol{y}) = E_{\boldsymbol{X}}\big[KL(\pi(\boldsymbol{\theta} \mid \boldsymbol{D}), \pi(\boldsymbol{\theta}))\big] = \int_{\mathcal{X}^n} p(\boldsymbol{y}) \Big[\int_{\Theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) log \frac{\pi(\boldsymbol{\theta} \mid \boldsymbol{y})}{\pi(\boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta} \Big] \mathrm{d}\boldsymbol{y} \tag{2.3}$$

where $\mathcal{X}^n = \mathcal{X} \times \mathcal{X} \cdots \times \mathcal{X}$ is the sample space, $p(\boldsymbol{y}) = \int_{\Theta} f(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}$ is the marginal density of the sample $\boldsymbol{D}$.

If $\pi^*(\boldsymbol{\theta}) \in \mathcal{P}$, where $\mathcal{P} = \{\pi(\boldsymbol{\theta}) > 0 : \int_{\Theta} \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \mathrm{d}\boldsymbol{\theta} < \infty\}$, and satisfies the condition $I_{\pi^*(\boldsymbol{\theta})}(\boldsymbol{\theta}, \boldsymbol{y}) = \max\limits_{\pi(\boldsymbol{\theta}) \in \mathcal{P}} I_{\pi(\boldsymbol{\theta})}(\boldsymbol{\theta}, \boldsymbol{y})$, then

$$\pi^*(\boldsymbol{\theta}) = arg \max_{\pi(\boldsymbol{\theta}) \in \mathcal{P}} I_{\pi(\boldsymbol{\theta})}(\boldsymbol{\theta}, \boldsymbol{y})$$

is referred to as the reference prior for the parameter $\boldsymbol{\theta}$.

The primary advantage of the reference prior lies in its enhanced flexibility in handling nuisance parameters through structured parameter ordering, which effectively avoids potential theoretical paradoxes inherent in Jeffreys prior. However, its invariance property is contingent upon predefined parameter groupings and remains valid only for within-group transformations. In cases where explicit parameter ordering is absent, an independent optimization approach must be adopted for individual parameters, resulting in substantially increased computational complexity [17].

2.4 The power prior

An informative prior is one that reflects a high degree of certainty regarding the model parameters to be estimated. When sufficient information about θ is available, we may employ methods such as the histogram approach or the relative likelihood method to specify the prior density function. Specifically, one can estimate hyper-parameters or key quantiles of θ, then derive the corresponding cumulative distribution function to construct the prior distribution.

The power prior [18, 19, 20] is a flexible approach that allows researchers to assign a discounting weight to historical data. As an optimal class of informative prior distributions, the power prior achieves its optimality by minimizing the convex combination of the Kullback-Leibler (KL) divergences between two posterior distributions under extreme scenarios [21]: (1) the posterior distribution without borrowing historical data (i.e., $\alpha_0 = 0$), and (2) the posterior distribution with full borrowing of historical data (i.e., $\alpha_0 = 1$). This property ensures the power prior's theoretical optimality in balancing new and historical information.

In this article, we assume that D represents the current study data, with the corresponding likelihood function denoted by $L(\boldsymbol{\theta} \mid D)$, where $\boldsymbol{\theta}$ is a parameter vector. Suppose we also have historical data D_0 from a similar previous study, with likelihood function $L(\boldsymbol{\theta} \mid D_0)$. Given a discounting parameter $\alpha_0 \in [0, 1]$, the power prior for $\boldsymbol{\theta}$ in the current study is defined as:

$$\pi(\boldsymbol{\theta} \mid D_0, \alpha_0) \propto L(\boldsymbol{\theta} \mid D_0)^{\alpha_0} \pi_0(\boldsymbol{\theta}) \tag{2.4}$$

where $\pi_0(\boldsymbol{\theta})$ denotes the initial prior distribution specified before observing any historical data D_0, typically chosen to be non-informative.

Using the power prior defined in (2.4), the posterior distribution of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta} \mid D, D_0, \alpha_0) \propto L(\boldsymbol{\theta} \mid D) L(\boldsymbol{\theta} \mid D_0)^{\alpha_0} \pi_0(\boldsymbol{\theta})$$

The parameter α_0 quantifies the proportion of historical information incorporated into the current study, as defined by the prior in equation (2.4). When $\alpha_0 = 0$, the historical data are entirely excluded, whereas $\alpha_0 = 1$ assigns equal weight to the historical likelihood $L(\boldsymbol{\theta} \mid D_0)$ and the current study likelihood $L(\boldsymbol{\theta} \mid D)$, effectively fully integrating the historical data. Thus, (2.4) generalizes the standard Bayesian update of $\pi_0(\boldsymbol{\theta})$. The power parameter α_0 can be interpreted as a precision parameter for the historical data, with its estimated value $\hat{\alpha}_0$ reflecting the reliability or relevance of D_0. For further discussion, see, for example, [22, 23].

In many applications of the power prior, α_0 may be treated as a fixed value, with sensitivity analyses conducted across different specified values. Alternatively, a hierarchical prior specification can be implemented by treating α_0 as random - for instance, by assigning it a Beta distribution. In this hierarchical framework, the complete prior specification becomes

$$\pi(\boldsymbol{\theta}, \alpha_0 \mid D_0) \propto L(\boldsymbol{\theta} \mid D_0)^{\alpha_0} \pi_0(\boldsymbol{\theta}) \pi_0(\alpha_0) \tag{2.5}$$

We refer to (2.5) as the **joint power prior**. Given the current data D, the joint posterior distribution of $\boldsymbol{\theta}$ and α_0 is given by

$$\pi(\boldsymbol{\theta}, \alpha_0 \mid D, D_0) \propto L(\boldsymbol{\theta} \mid D) L(\boldsymbol{\theta} \mid D_0)^{\alpha_0} \pi_0(\boldsymbol{\theta}) \pi_0(\alpha_0)$$

A natural choice for the prior distribution of α_0 is the Beta distribution, $Beta(\alpha, \beta)$ or more simply, a uniform distribution over $[0, 1]$, given the constraint that $0 \leq \alpha_0 \leq 1$.

Power priors provide a formal mechanism for incorporating historical data via a discounting factor, but they require careful specification of the power parameter. Additionally, if the historical and current data distributions differ significantly, power priors may introduce bias.

2.5 Conjugate prior

In Bayesian statistics, a conjugate prior is a prior distribution that, when combined with a specific likelihood function, yields a posterior distribution belonging to the same probability distribution family. This mathematical convenience simplifies Bayesian updating, enabling closed-form analytical solutions. Conjugate priors remain foundational in Bayesian methodology, bridging theoretical elegance and practical utility. While computational advances have reduced their necessity, they persist as vital tools for prototyping, education, and theoretical analysis.

A class $\mathcal{P}$ of prior distributions is called a conjugate family if, for every prior density $\pi(\theta) \in \mathcal{P}$ the posterior density

$$\pi(\theta \mid \boldsymbol{y}) \propto \pi(\theta) L(\theta \mid \boldsymbol{y}) \tag{2.6}$$

also belongs to $\mathcal{P}$ for all observed data $\boldsymbol{y}$.

Conjugate priors are a mathematically convenient class of prior distributions that provide analytical tractability for sampling distributions from the exponential family. In Bayesian inference, a well-defined correspondence exists between probability distributions and their conjugate priors. For example: (1) the Beta distribution serves as the conjugate prior for the binomial distribution (a discrete case); (2) the normal distribution acts as the conjugate prior for the mean of another normal distribution (a continuous case).

By ensuring the prior and posterior belong to the same distribution family, conjugate priors simplify posterior derivation, parameter estimation, and predictive analysis. However, despite their computational advantages in Bayesian inference, the rigid structure of conjugate priors may restrict prior flexibility and undermine model robustness [17]. Thus, their application requires careful consideration in practice.

3 Conceptual and theoretical foundations of Bayes factors

Within Jeffreys' Bayesian framework, hypothesis testing serves primarily to assess the evidentiary support for scientific theories. The Bayes factor, as a fundamental statistical instrument, provides the methodological basis for this evaluation by accomplishing two key objectives: (1) quantifying the evidence strength for competing hypotheses, and (2) systematically incorporating prior knowledge to improve hypothesis assessment reliability. Mathematically, the Bayes factor represents the ratio of posterior to prior odds for H_0 versus H_1, precisely measuring how much the observed data should shift our belief between the competing hypotheses [24]. Importantly, this evidentiary measure maintains complete symmetry - the evidence may favor either H_0 or H_1 equally, as neither hypothesis holds a privileged position in the analysis.

Consider the general case of two competing hypotheses

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1$$

where $\Theta_0 \bigcup \Theta_1 = \Theta$ and $\Theta_0 \bigcap \Theta_1 = \Phi$. Bayesian hypothesis testing requires prior probabilities $\pi_0 = \boldsymbol{P}(H_0)$ and $\pi_1 = \boldsymbol{P}(H_1)$ that sum to 1. Most people just use $\pi_0 = \pi_1 = 1/2$. For observable data $\boldsymbol{y}$, denote the likelihood functions under the two hypotheses by $p(\boldsymbol{y} \mid \theta, H_0)$ and $p(\boldsymbol{y} \mid \theta, H_1)$, and $p(\theta \mid H_k)$ is the prior of the parameter θ if H_k is true. For each $k = 0, 1$, the marginal likelihood is

$$p(\boldsymbol{y} \mid H_k) = \int_{\Theta_k} p(\boldsymbol{y} \mid \theta, H_k) p(\theta \mid H_k) d\theta,$$

which is called the prior predictive distribution of $\boldsymbol{y}$ conditional on H_k. According to Bayes' rule, the posterior probability of H_k given the observable data $\boldsymbol{y}$ is

$$p(H_k \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid H_k)\pi_k}{p(\boldsymbol{y} \mid H_0)\pi_0 + p(\boldsymbol{y} \mid H_1)\pi_1} := p_k.$$

Computing the posterior odds on H_1 against H_0 gives the equation

$$\frac{p_1}{p_0} = \frac{p(\boldsymbol{y} \mid H_1)}{p(\boldsymbol{y} \mid H_0)} \times \frac{\pi_1}{\pi_0}.$$

Then the Bayes factor in favour of H_1 against H_0, denoted BF_{10}, is defined by

$$BF_{10} = \frac{p(\boldsymbol{y} \mid H_1)}{p(\boldsymbol{y} \mid H_0)} = \frac{p_1/p_0}{\pi_1/\pi_0} = \frac{p_1\pi_0}{p_0\pi_1}. \tag{3.1}$$

It follows that Bayes factor BF_{10} represents the likelihood ratio of H_1 relative to H_0, as determined by the observed data. If the priors of the hypotheses are set to $\pi_0 = \pi_1 = 1/2$, then BF_{10} equals the posterior odds of the hypotheses.

The Bayes factor offers a direct quantitative evaluation of competing hypotheses and serves as a robust measure of evidential strength. Different magnitudes of the Bayes factor indicate varying degrees of support: (1) if $BF_{10} > 1$, the data support H_1 over H_0, (2) if $BF_{10} < 1$, the data favor H_0 over H_1, and (3) if $BF_{10} = 1$, the evidence is equivocal, providing equal support for both hypotheses. Beyond merely indicating the relative strength of evidence, the Bayes factor explicitly quantifies the degree of support. Following established conventions [25], Table 1 presents a classification framework for interpreting Bayes factor values.

Jeffreys' scale of evidence enables researchers to qualitatively assess the strength of evidence for either the null or alternative hypothesis, using predefined thresholds that reflect varying degrees of evidential support. For example, a Bayes factor of 5 indicates that the data support H_1 five times more strongly than H_0, while a Bayes factor of 0.2 suggests that the data favor H_0 five times more than H_1.

3.1 Computation of the Bayes Factor

The computation of the Bayes factor is fundamentally based on three core components of Bayesian analysis: the prior distribution, the likelihood function, and the posterior distri-

Table 1: Jeffreys' scale of evidence for interpreting Bayes factor BF_{10}.

Bayes factor BF_{10}	Interpretation
> 100	Extreme evidence for H_1
30 - 100	Very strong evidence for H_1
10 - 30	Strong evidence for H_1
3 - 10	Moderate evidence for H_1
1 - 3	Anecdotal evidence for H_1
1	No evidence
1/3 - 1	Anecdotal evidence for H_0
1/10 - 1/3	Moderate evidence for H_0
1/30 - 1/10	Strong evidence for H_0
1/100 - 1/30	Very strong evidence for H_0
$< 1/100$	Extreme evidence for H_0

bution. The prior distribution encapsulates our initial beliefs about the parameters before observing the data, while the likelihood function quantifies the probability of the observed data under specific parameter values. The posterior distribution, derived by updating the prior with the observed data through Bayes' theorem, represents our revised understanding of the parameters. Together, these components form the mathematical foundation for calculating Bayes factors, which provide a rigorous means for quantifying the evidence in favor of competing hypotheses.

The calculation of Bayes factors frequently involves complex numerical integration problems, for which traditional numerical integration methods often prove computationally inefficient. To address this challenge, researchers have developed various approximation and simulation techniques to streamline the process. For instance, Schwarz's Bayesian Information Criterion (BIC) provides an effective approximation of the logarithmic Bayes factor when dealing with large model spaces and sufficient sample sizes [26]. Kass and Vaidyanathan [27] further refined the approximation of Bayes factors by simplifying the marginal likelihood

computation. Additionally, Smith and Roberts [28] pioneered Bayesian computation techniques utilizing Gibbs sampling and Markov chain Monte Carlo (MCMC) methods, offering efficient numerical solutions for complex Bayesian model inference.

Notably, the growing adoption of Bayesian statistics has been accompanied by significant advancements in computational tools (e.g., JASP, Stan, JAGS, BayesFactor, brms, bain, BANOVA, PyMC). These developments have substantially enhanced researchers' capacity to perform Bayesian analyses with greater efficiency and accessibility.

3.2 Some examples

This section illustrates the derivation of Bayes factors through concrete examples using binomial and normal distributions, incorporating both non-informative and conjugate prior specifications. Through these specific cases, we present a systematic derivation process and examine the resulting statistical interpretations.

3.2.1 The Bayes factor derivation of the binomial data

We first focus on binary outcomes as they represent one of the most prevalent forms of response data in clinical research. In clinical trials, binary endpoints are particularly valuable as they enable direct comparison of success rates between treatment and control groups.

In the control group, the number of cured patients y_c follows a binomial distribution with sample size n_c and fixed success probability p_c: $y_c \mid p_c \sim \text{Bin}(n_c, p_c)$, where p_c is a known constant established by prior studies. Similarly, in the treatment group, the number of cured patients y_t follows: $y_t \sim \text{Bin}(n_t, p_t)$, where p_t represents the treatment's cure probability. Our objective is to test the null hypothesis $H_0 : p_t = p_c$, against the one-sided alternative $H_1 : p_t > p_c$.

Under Bayesian framework, we specify a conjugate prior for p_t: $p_t \sim \text{Beta}(\alpha_t, \beta_t)$. The

Bayes factor is computed as the ratio of marginal likelihoods as follows

$$BF_{10} = \frac{P(y_t \mid H_1)}{P(y_t \mid H_0)} = \frac{\displaystyle\int_{p_c}^{1} \binom{n_t}{x_t} p_t^{x_t}(1-p_t)^{n_t-x_t} \cdot \frac{p_t^{\alpha_t-1}(1-p_t)^{\beta_t-1}}{B(\alpha_t,\beta_t)}\,\mathrm{d}p_t}{\binom{n_t}{x_t} p_c^{x_t}(1-p_c)^{n_t-x_t}}. \tag{3.2}$$

Using beta-binomial conjugation, the marginal likelihood of H_1 can be expressed as:

$$P(y_t \mid H_1) = \binom{n_t}{x_t} \cdot \frac{B(x_t+\alpha_t, n_t-x_t+\beta_t)}{B(\alpha_t,\beta_t)} \cdot I^1_{p_c}(\alpha_t,\beta_t,x_t,n_t),$$

Where $I^1_{p_c}$ is the truncated Beta integral. If the truncation ($p_t \in [0,1]$) is ignored, this simplifies to:

$$P(y_t \mid H_1) = \binom{n_t}{x_t} \cdot \frac{B(x_t+\alpha_t, n_t-x_t+\beta_t)}{B(\alpha_t,\beta_t)}.$$

The Bayes factor is simplifies to:

$$BF_{10} = \frac{B(x_t+\alpha_t, n_t-x_t+\beta_t)}{B(\alpha_t,\beta_t)\cdot p_c^{x_t}(1-p_c)^{n_t-x_t}}. \tag{3.3}$$

When the uniform prior $\alpha_t = \beta_t = 1$ is adopted, it is further simplified as

$$BF_{10} = \frac{B(x_t+1, n_t-x_t+1)}{p_c^{x_t}(1-p_c)^{n_t-x_t}} = \frac{x_t!(n_t-x_t)!}{(n_t+1)!} \cdot \frac{1}{p_c^{x_t}(1-p_c)^{n_t-x_t}}. \tag{3.4}$$

If $BF_{10} > 1$, H_1 is supported; Otherwise, H_0 is supported.

3.2.2 The Bayes factor derivation of the normal distribution

Within the context of clinical trials characterised by the utilisation of continuous outcome measures (for instance, alterations in blood pressure or blood glucose levels as quantitative endpoints), the Bayes factor may be employed to quantitatively assess the magnitude of the evidence for treatment effects between the treatment and control groups.

Assume the treatment group outcomes $\boldsymbol{y}$ follows a normal distribution with mean μ and variance σ^2,i.e., $\boldsymbol{y} \sim \mathrm{N}(\mu, \sigma^2)$, where σ^2 is a known. The purpose is to undertake a comparative evaluation between the null hypothesis $H_0 : \mu = \mu_0$ and the alternative hypothesis $H_1 : \mu \neq \mu_0$.

For parameter μ, we specify a conjugate prior: $\mu \sim \mathrm{N}(\mu_1, \tau^2)$, where μ_1 is prior mean and τ^2 is the prior variance. Under the alternative hypothesis H_1, the marginal likelihood is obtained by integrating the parameter μ:

$$\begin{aligned} P(\boldsymbol{y} \mid H_1) &= \int p(\boldsymbol{y} \mid \mu, H_1)\, p(\mu \mid H_1)\, \mathrm{d}\mu \\ &= (2\pi\sigma^2)^{-n/2} \left(\frac{\sigma^2}{\sigma^2 + n\tau^2} \right)^{1/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + \frac{n\sigma^2(\bar{y} - \mu_1)^2}{\sigma^2 + n\tau^2} \right] \right\}, \end{aligned}$$

Under the null hypothesis $H_0 : \mu = \mu_0$ is given by:

$$P(\boldsymbol{y} \mid H_0) = (2\pi\sigma^2)^{-n/2} \exp\left\{ -\frac{\sum_{i=1}^{n} (y_i - \mu_0)^2}{2\sigma^2} \right\},$$

The Bayes factor is

$$BF_{10} = \left(\frac{\sigma^2}{\sigma^2 + n\tau^2} \right)^{\frac{1}{2}} \cdot \exp\left\{ \frac{n}{2\sigma^2} \left[(\bar{y} - \mu_0)^2 - \frac{\sigma^2(\bar{y} - \mu_1)^2}{\sigma^2 + n\tau^2} \right] \right\}. \tag{3.5}$$

When the parameter μ follows non-informative prior (such as Jeffrey's prior), taking $\tau^2 \to \infty$, the simplified Bayes factor is:

$$BF_{10} \approx \sqrt{\frac{2\pi\sigma^2}{n}} \cdot \exp\left\{ \frac{(\bar{y} - \mu_0)^2}{2\sigma^2/n} \right\}. \tag{3.6}$$

If the $BF_{10} > 1$, H_1 is accepted, and the greater the BF_{10} value, the more the hypothesis is supported.

4 Bayesian hypothesis testing versus null hypothesis significance testing

In clinical trials, Bayes factors and Null Hypothesis Significance Testing (NHST) are two widely utilized statistical inference tools, each possessing distinct applications and interpretations in the context of hypothesis testing. Although both methodologies ultimately seek to assist researchers in determining whether the data supports a specific hypothesis, they

exhibit significant differences in their foundational principles, inference processes, and interpretations of results. This section will examine the key distinctions between Bayes factors and NHST within the framework of hypothesis testing.

4.1 Null hypothesis significance testing

The conventional method of statistical inference is the NHST, which utilizes the P value to determine whether to reject the null hypothesis (H_0) or accept the alternative hypothesis (H_1). The P value represents a specific observed value or a more extreme value that occurs under the assumption that the null hypothesis is true. It quantifies how anomalous the data are under this null hypothesis, reflecting evidence against it. A smaller P value provides stronger evidence against the null hypothesis.

Under the framework of the frequentist school, researchers control the type I error rate by setting the significance level (α, usually 0.05). If $P \leq \alpha$, the null hypothesis is not supported. Nonetheless, this dichotomous decision-making approach does not fully reflect the spectrum of evidence strength. As shown in Table 2, the Fisher evidence stratification framework divides P values into seven distinct categories. For example, when $P = 0.05$ (corresponding to a coverage level of 0.95), it indicates "moderate" evidence; in contrast, when $P \leq 0.005$ (with a coverage level ≥ 0.995), it represents "very strong" evidence. Historically, the strength of evidence linked to an α threshold of 0.05 is notably weaker compared to that associated with more rigorous standards, such as $P = 0.005$. Consequently, in research practice, it is crucial to perform comprehensive evaluations by incorporating multiple levels of P value (e.g., treating values between 0.05 and 0.10 as "borderline", while those below 0.001 are classified as "overwhelming"), along with effect size and confidence intervals, to steer clear of overly simplistic binary interpretations.

Table 2: Fisher's scale of evidence against null hypothesis H_0 and in favor of H_1, as a function of coverage level (1 minus the P value).

Coverage	(P-value)	Evidence for H_1
.80	(.20)	null
.90	(.10)	borderline
.95	(.05)	moderate
.975	(.025)	substantial
.99	(.01)	strong
.995	(.005)	very strong
.999	(.001)	overwhelming

4.2 A Bayesian/frequentist comparison list

The widespread misuse and misinterpretation of P value in NHST have drawn increasing criticism in recent years [29, 30]. Some Flaws in frequentist inference are as follows [31]: (1) fundamental limitations of NHST include the prevalent misconception that P value represent the probability of the null hypothesis being true, when they actually indicate the probability of observing the data under the null assumption; (2) NHST fails to adequately account for sample size effects, where large samples may produce statistically significant yet practically meaningless results; (3) the methodology is also vulnerable to P-hacking selective reporting and analytical flexibility that inflates false positive rates and proves particularly inadequate for complex modeling scenarios involving multiple comparisons, where it tends to generate excessive type I errors. This has prompted a shift toward more comprehensive analytical approaches incorporating effect sizes, confidence intervals, and Bayesian methods.

The Bayes factor can directly quantify the relative evidence supporting either the null hypothesis or the alternative hypothesis, thereby providing a measure of the strength of each hypothesis in light of the observed data. Consequently, in statistical decision-making, the Bayes factor offers a more comprehensive and informative approach compared to the P

value. Each method has unique advantages and disadvantages that will be systematically compared below [32].

1. Prior specification challenges. In Bayesian inference, the process of defining a prior distribution $\pi(\theta)$ explicitly can pose practical challenges. Conversely, frequentist approaches substitute this requirement with the creation of an algorithmic procedure $t(x)$, which is specifically designed to tackle particular research questions. While an ideal $t(x)$ seeks to reduce subjectivity, contemporary computational techniques frequently surpass classical optimality paradigms, integrating ad hoc components into frequentist evaluations.

2. The simplicity argument cuts both ways. Bayesian methodologies confirm their conclusions based on the suitability of the selected prior distribution, highlighting the significance of prior determination. In contrast, frequentist strategies adopt a cautious approach by emphasizing robustness across diverse potential parameter values θ, thereby ensuring reliable performance under different circumstances.

3. Operational characteristics. The single execution of Bayesian analysis delivers extensive solutions to numerous inferential inquiries, providing adaptability and effectiveness. On the other hand, frequentist analysis usually involves constructing separate estimators for distinct questions, enabling a more targeted exploration of specific issues but demanding greater analytical effort.

4. Subjectivity considerations. When genuine prior information is absent, Bayesian outcomes obtained from non-informative priors might still unintentionally incorporate nuanced forms of subjectivity. Classical frequentism has established itself as the benchmark for scientific impartiality, a position especially highlighted in controversial domains like drug testing and approval, where both skeptics and supporters meticulously

assess statistical intricacies.

An increasing number of studies highlight the practical advantages of Bayesian methods, responding to the overly simplistic criticism of the frequentist methodology [33]. Nonetheless, it would be unwise to completely overlook the P value, as it requires thorough analysis. A low P value does not definitively disprove the null hypothesis; instead, it could be explained by random variations or additional influencing factors. In certain scenarios, Bayes factors provide deeper and more enlightening perspectives, especially in the area of model evaluation, where Bayesian approaches exhibit significant competence in systematically and effectively combining evidence [34, 35].

5 A clinical case analysis of Bell's palsy trials to assess drug efficacy

The Bayes factor has increasingly drawn interest and is now extensively utilized in multiple fields, such as clinical trials, epidemiology, psychology, neuroscience, and economics. Its main applications are prominently seen in areas like model evaluation, hypothesis assessment, and decision-making analysis. Within clinical research, the examination of drug effectiveness is vital for determining treatment results, improving therapy approaches, and boosting patient outcomes. This section focuses on exploring how the Bayes factor can be applied in clinical trials tailored to assess the efficacy of drug treatments.

Bell's palsy, also referred to as Bell's paralysis or Bell's facial paralysis, is characterized by an acute, idiopathic unilateral facial nerve palsy [36, 37]. While most patients have a favorable prognosis, a subset may experience persistent residual dysfunction to varying degrees. Bell's palsy not only alters the facial appearance of patients but also imposes significant negative effects on their daily functioning and psychological well-being [38]. Consequently, timely and effective management of Bell's palsy holds substantial clinical significance.

Prednisolone and acyclovir are commonly employed either individually or in combination for treating Bell's palsy [39]. Research has demonstrated that early administration of prednisolone markedly enhances the probability of complete recovery at both 3 and 9 months. However, acyclovir, whether used alone or in conjunction with prednisolone, did not exhibit any additional benefits [40]. Leveraging patient data derived from real-world clinical trials, we conducted an analysis of the trial outcomes utilizing synthetic data.

Final outcomes were assessed in 494 of the 551 randomized patients. The study population consisted of 256 males and 238 females. In terms of treatment allocation, 250 patients received prednisolone, while 244 did not. Additionally, 249 patients were treated with acyclovir, whereas 245 remained untreated with acyclovir. Among these, 127 patients received both prednisolone and acyclovir, and the placebo group comprised 122 patients.

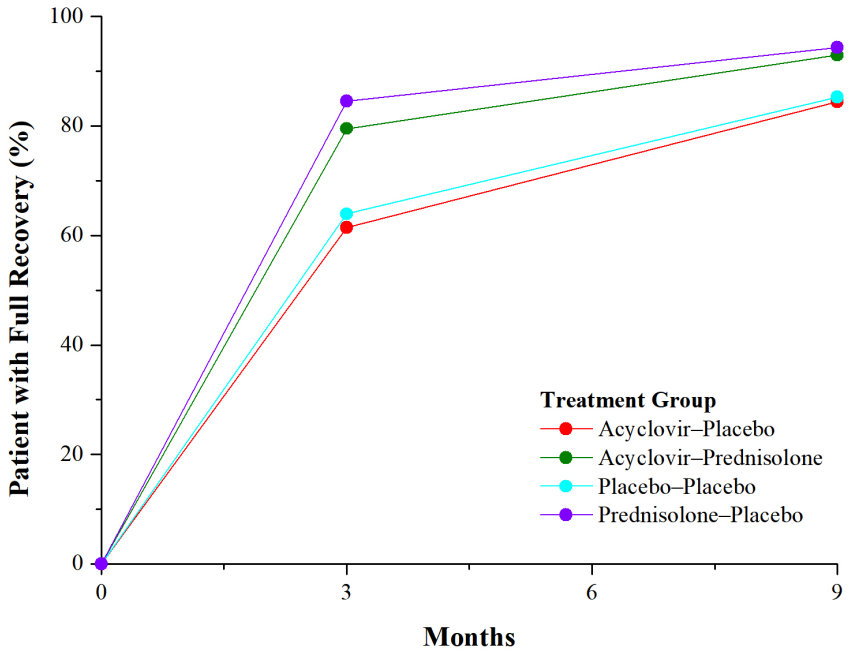

Figure 1: Patients Who Had a Full Recovery at 3 Months and 9 Months,According to Study Group. (Full recovery was defined as grade 1 on the House-Brackmann facial-nerve grading scale, which ranges from 1 to 6, with higher grades indicating worse facial paralysis.)

Following a three-month monitoring period, a marked rise in the percentage of patients achieving full recovery was noted in all treatment groups. More specifically, the Prednisolone-Placebo group achieved the highest recovery rate at 84.55%, outperforming all other groups. The Acyclovir-Prednisolone group came in a close second with a recovery rate of 79.53%. Conversely, the Acyclovir-Placebo and Placebo-Placebo groups reported lower recovery rates of 61.48% and 63.93%, respectively. By the conclusion of the nine-month follow-up, recovery rates across all groups had further improved. The Prednisolone-Placebo group maintained its lead with a recovery rate of 94.31%. The Acyclovir-Prednisolone group also demonstrated progress, attaining a recovery rate of 92.91% and placing second. Moreover, the Placebo-Placebo and Acyclovir-Placebo groups saw their recovery rates rise to 85.25% and 84.43%, respectively. Fisher's exact test was utilized to evaluate differences in complete recovery rates at both 3 months and 9 months between drug-treated patients and those receiving placebo, as detailed in Table 3.

Table 3: Comparison of full recovery at 3 and 9 months

Recovery Period	Comparison Group	P value	Odds Ratio	95% Confidence Interval
3 Months	A vs D	$< 0.001^{**}$	3.07	(1.61, 6.04)
	B vs D	0.791	1.11	(0.64, 1.93)
	C vs D	0.007^{*}	0.46	(0.25, 0.83)
9 Months	A vs D	0.021^{*}	2.86	(1.08, 8.43)
	B vs D	1.000	1.07	(0.50, 2.29)
	C vs D	0.066	0.44	(0.17, 1.09)

[a] **A**: Prednisolone-Placebo, **B**: Acyclovir-Placebo, **C**: Prednisolone-Acyclovir, **D**: Placebo-Placebo

Prednisolone showed markedly greater complete cure rates in comparison to placebo at both the 3 months and 9 months marks. When administered independently, acyclovir failed to produce a notable impact. Furthermore, the combination of prednisolone and acyclovir

proved to be considerably less efficacious than placebo after 3 months and could potentially have been less effective even after 9 months.

Table 4 displays the outcomes of the Bayes factor analysis, providing a numerical assessment of treatment effects across the various treatment groups relative to the placebo group. As illustrated in Table 4, the Bayes factor comparing Group A with Group D at 3 months is 46,877.41, offering overwhelmingly strong support for the hypothesis that Group A's treatment effect exceeds that of Group D. At 9 months, this Bayes factor diminishes to 30.08 but continues to endorse the hypothesis that Group A maintains a superior treatment effect compared to Group D. In contrast, the Bayes factors for Group B versus Group D remain below 1 at both 3 and 9 months, signifying that the data align with the null hypothesis, implying no notable difference in treatment effects between Group B and Group D. Additionally, the Bayes factor for Group C versus Group D at 3 months is 236.85, furnishing substantial evidence in favor of the hypothesis that Group C's treatment effect surpasses that of Group D. At 9 months, the Bayes factor for Group C versus Group D declines to 8.16, continuing to uphold the hypothesis that Group C's treatment effect is greater than that of Group D, though with less compelling evidence compared to the earlier time point. Through the computation of Bayes factors among the groups, we achieved a quantitative evaluation of the strength of evidence regarding treatment effects in relation to the placebo. The significant differences in Bayes factors between the treatment groups and the placebo group deliver robust statistical backing for the enhanced efficacy of the treatments.

Table 4: Bayes factor comparison between treatment groups and placebo

Comparison Group	3 Months Bayes Factor	9 Months Bayes Factor
A vs D	46,877.41	30.08
B vs D	0.26	0.31
C vs D	236.85	8.16

Table 3 demonstrates the contrast in recovery rates among various treatment groups and the placebo group at both 3 and 9 months. Conventional hypothesis testing depends on P value to evaluate the significance of differences, while the Bayes factor serves as a measure of the strength of evidence backing a particular hypothesis. For instance, for the comparison between Group A and Group D at 3 months, $P < 0.001$ alongside the $BF_{10} = 46,877.41$ clearly shows that Group A is substantially more effective than Group D, with the Bayes factor offering an extra level of quantifiable support. On the other hand, for Group B versus Group D, both the P value and the Bayes factor correspond to the null hypothesis of no significant difference. Regarding Group C versus Group D at 3 months, $P = 0.007$ and the $BF_{10} = 236.85$ validate that Group C outperforms Group D significantly. At 9 months, although the $P > 0.05$, the $BF_{10} = 8.16$ still endorses the idea that Group C is more efficacious than Group D. In summary, the Bayes factor delivers a richer and more detailed quantitative assessment compared to traditional hypothesis testing methods.

6 Conclusion and outlook

In Bayesian approaches, the selection of prior information is crucial, as developing a suitable prior distribution greatly enhances the accuracy and reliability of inferences. However, computing posterior distributions often involves tackling complex numerical integration problems, where Markov Chain Monte Carlo (MCMC) methods provide a feasible solution. Despite these developments, applying Bayesian methods in practical scenarios requires careful consideration, particularly in specifying prior distributions and ensuring computational efficiency. As such, the development of more efficient and user-friendly computational tools continues to be vital for further progress in this field.

In clinical trials, the p value and Bayesian factor serve different functions. The P value provides a clear method for evaluating result significance, while the Bayesian factor allows

for a more nuanced integration and updating of evidence. Researchers should select the most suitable statistical technique based on the particular circumstances of their research, while also recognizing the inherent constraints of each methodology.

In conclusion, Bayes factors hold considerable promise in clinical trials through their ability to offer more flexible and transparent statistical assistance for drug development and medical decision-making processes. With the ongoing development of computational capabilities and the gradual improvement of methodological approaches, Bayes factors are expected to play a critical role in clinical trials, further facilitating the progress and innovation within medical research.

Acknowledgments

The authors thank the editors and anonymous reviewers for their insightful and helpful comments, which help improve the manuscript a lot. The research is supported by Natural Science Foundation of Henan Province (252300421487).

Competing Interests

No potential conflict of interest was reported by the author.

References

[1] van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., & Tadesse, M. G. et al. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1),1.

[2] Johnson, S. R., Tomlinson, G. A., Hawker, G. A., Granton, J. T.,et al. & Feldman, B. M. (2010). Methods to elicit beliefs for Bayesian priors: a systematic review. *Journal of clinical epidemiology*, 63(4), 355-369.

[3] Quintana, M., Viele, K., & Lewis, R. J. (2017). Bayesian Analysis: Using Prior Information to Interpret the Results of Clinical Trials. *Journal of the American Medical Association*, 318(16), 1605-1606.

[4] Berry, D. A. (2006). Bayesian clinical trials. *Nature reviews Drug discovery*, 5(1), 27-36.

[5] Sugitani, Y., Morita, S., Nakakura, A., & Yamamoto, H. (2023). Biomarker-based Bayesian randomized clinical trial design for identifying a target population. *Statistics in Medicine*, 42(16), 2797-2810.

[6] Jack Lee, J., & Chu, C. T. (2012). Bayesian clinical trials in action. *Statistics in medicine*, 31(25), 2955-2972.

[7] Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, 5(1), 393-419.

[8] Beard, E., Jackson, S. E., Anthenelli, R. M., Benowitz, N. L., Aubin, L. S., & McRae, T. et al. (2021). Estimation of risk of neuropsychiatric adverse events from varenicline, bupropion and nicotine patch versus placebo: secondary analysis of results from the EAGLES trial using Bayes factors. *Addiction*, 116(10), 2816-2824.

[9] Monden, R., Stijn, D. V., Klaas, W., Annelieke, R., Richard, M., & Eric-Jan, W. (2016). Toward evidence-based medical statistics: Re-evaluate the efficacy of antidepressants by using Bayes factors. *European Psychiatry*, 33(S1), S418-S419.

[10] Richard, E., den Brok, M. G., & van Gool, W. A. (2021). Bayes analysis supports null hypothesis of anti-amyloid beta therapy in Alzheimer's disease. *Alzheimer's & Dementia*, 17(6), 1051-1055.

[11] Pittelkow, M. M., Linde, M., de Vries, Y. A., Hemkens, L. G., Schmitt, A. M., & Meijer, R. R. et al. (2024). Strength of statistical evidence for the efficacy of cancer drugs: a Bayesian reanalysis of randomized trials supporting Food and Drug Administration approval. *Journal of Clinical Epidemiology*, 174, 111479.

[12] Gelman, A. (2002). Prior distribution. *Encyclopedia of environmetrics*, 3(4), 1634-1637.

[13] Kass, R. E., & Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American statistical Association*, 91(435), 1343-1370.

[14] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London,Ser.A*, 186(1007), 453-461.

[15] Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society, Ser. B*, 41(2), 113-128.

[16] Berger, J. O., & Bernardo, J. M. (1992). Ordered group reference priors with application to the multinomial problem. *Biometrika*, 79(1), 25-37.

[17] Robert, C.P. (2007). From Prior Information to Prior Distributions. In: The Bayesian Choice. Springer Texts in Statistics. Springer, New York.

[18] Ibrahim, J. G., & Chen, M. H. (2000). Power prior distributions for regression models. *Statistical Science*, 15(1), 46-60.

[19] Ibrahim, J. G., Ryan, L. M., & Chen, M. H. (1998). Using historical controls to adjust for covariates in trend tests for binary data. *Journal of the American Statistical Association*, 93(444), 1282-1293.

[20] Chen, M. H., Ibrahim, J. G., & Shao, Q. M. (2000). Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference*, 84(1-2), 121-137.

[21] Ibrahim, J. G., Chen, M. H., Gwon, Y., & Chen, F. (2015). The power prior: theory and applications. *Statistics in medicine*, 34(28), 3724-3749.

[22] Gravestock, I., Held, L., & COMBACTE-Net consortium. (2017). Adaptive power priors with empirical Bayes for clinical trials. *Pharmaceutical statistics*, 16(5), 349-360.

[23] Bennett, M. S. (2018). Improving the efficiency of clinical trial designs by using historical control data or adding a treatment arm to an ongoing trial (Doctoral dissertation).

[24] Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795.

[25] Lee, M. D., & Wagenmakers, E. J. (2014). Bayesian cognitive modeling: A practical course. Cambridge university press.

[26] Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.

[27] Kass, R. E., & Vaidyanathan, S. K. (1992). Approximate Bayes factors and orthogonal parameters, with application to testing equality of two binomial proportions. *Journal of the Royal Statistical Society, Ser.B*, 54(1), 129-144.

[28] Smith, A. F., & Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Seri.B*, 55(1), 3-23.

[29] Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: context, process, and purpose. *The American Statistician*, 70(2), 129-133.

[30] Nuzzo, R. (2014). Scientific method: Statistical errors. *Nature*, 506, 150-152.

[31] Marden, J. I. (2000). Hypothesis testing: from p values to Bayes factors. *Journal of the American Statistical Association*, 95(452), 1316-1320.

[32] Efron, B., & Hastie, T. (2021). Computer age statistical inference, student edition: algorithms, evidence, and data science (Vol. 6). Cambridge University Press.

[33] Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological methodology*, 111-163.

[34] Box, G. E. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society, Ser.A*, 143(4), 383-404.

[35] Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). Bayesian data analysis. Chapman and Hall/CRC.

[36] Gilden, D. H. (2004). Bell's palsy. *New England Journal of Medicine*, 351(13), 1323-1331.

[37] Eviston, T. J., Croxson, G. R., Kennedy, P. G., Hadlock, T., & Krishnan, A. V. (2015). Bell's palsy: aetiology, clinical features and multidisciplinary care. Eviston TJ, Croxson GR, Kennedy PGE, et al. Bell's palsy: aetiology, clinical features and multidisciplinary care. *Journal of Neurology, Neurosurgery & Psychiatry*, 86(12), 1356-1361.

[38] Zhang, W., Xu, L., Luo, T., Wu, F., Zhao, B., & Li, X. (2020). The etiology of Bell's palsy: a review. *Journal of neurology*, 267, 1896-1905.

[39] Rowlands, S., Hooper, R., Hughes, R., & Burney, P. (2002). The epidemiology and treatment of Bell's palsy in the UK. *European journal of neurology*, 9(1), 63-67.

[40] Sullivan, F. M., Swan, I. R., Donnan, P. T., Morrison, J. M., Smith, B. H., & McKinstry, B. et al. (2007). *New England Journal of Medicine*, 357(16), 1598-1607.