双臂II期临床试验中二元终点贝叶斯因子的检验效能与样本量计算

Riko Kelter*科隆大学医学院医学统计与计算生物学研究所，德国科隆

arXiv:2603.01715v1 [stat.ME] 2 Mar 2026

2026年3月3日

摘要

在实际应用中，临床试验的贝叶斯样本量计算通常依赖复杂的蒙特卡罗模拟。贝叶斯假阳性率和检验效能概念的界限往往缺乏闭式解或近似数值解。本文聚焦于II期试验双组二项分布情形下贝叶斯因子的检验效能与样本量计算。我们讨论点零假设与复合假设检验以及方向性假设检验，推导相应的贝叶斯因子，并探讨采用本文所介绍的方法进行贝叶斯试验设计时需要考虑的相关问题。基于这些贝叶斯因子，我们提出一种数值方法，可高效地确定必要的样本量，以满足预先设定的贝叶斯功效和I类错误率界限。我们的方法不依赖蒙特卡罗模拟，而仅采用标准数值方法。来自肿瘤学和自身免疫性疾病领域的真实II期试验案例展示了所提出校准方法的优势。总之，我们的方法针对II期临床试验的双组二项分布情形下的多种贝叶斯因子，提供了频率学派检验效能分析的贝叶斯对应量，从而实现贝叶斯—频率学派折衷。相关方法已在我们的 R 软件包 bfbin2arm 中实现。

关键词：*II*期试验、贝叶斯统计学、贝叶斯因子、双臂临床试验、二元终点

1 引言

样本量的确定是临床试验设计中最关键的环节之一，能够确保研究既提供科学信息，又符合伦理要求。样本量不足可能导致结果无法定论，而过度招募则会浪费资源，还可能使患者不必要地接受试验性治疗（Chow & Liu，2008年；Grieve，2022年；Spiegelhal-ter, Abrams, & Myles，2004年）。在统计严谨性与伦理责任之间取得平衡

*有关本文的通信请联系：rkelter@uni-koeln.de。草稿版本 1.0，2025年12月23日。本文尚未经过同行评审。未经作者许可，请勿复制或引用。用于复现本文结果的数据和 R 代码已公开发布于：https://osf.io/zsrfh/overview?viewonly=351aecc896d0468991569c99ae58beaf。R 软件包 bfbin2arm 可在 CRAN 上获取，详见 https://cran.r-project.org/web/packages/bfbin2arm/index.html。我们声明不存在利益冲突。

在早期临床试验中，这一点尤为突出，因为治疗效果存在很大的不确定性，患者群体通常也较小。

在II期肿瘤学试验中，主要终点通常为二元变量（例如肿瘤反应、缓解或毒性），Simon两阶段设计（Si-mon，1989）等经典设计方法仍被广泛使用。这些频率学派方法可以明确控制I类和II类错误率，通常分别设定为 $\alpha = 0.05$ 和 $\beta = 0.2$。然而，这些方法依赖固定样本推断，也无法自然地纳入有关治疗效果的先验信息——这一局限性在现代自适应和确证性情境中日益受到重视（Berry，2006；Spiegelhalteret al.，2004年；Thall & Simon，1994）。

贝叶斯设计提供了一种有原则且灵活的替代方案。它们将来自既往研究、专家意见或机制建模的先验证据正式整合到分析中，并随着数据积累不断更新这些知识。在贝叶斯II期设计中，决策规则通常以后验概率或贝叶斯因子的形式表达。例如，研究者可以建议将某项治疗推进至III期，前提是应答率超过临床相关阈值的后验概率高于预先设定的决策界值（Grieve，2016；Neuenschwander, Branson, & Spiegelhalter，2009；Thall &Simon，1994；Thall& Wathen，2007）。贝叶斯设计还允许因无效性或疗效而自适应停止，同时维持易于解释的长期错误控制（Berry，2006）。

贝叶斯因子尤其能通过比较边际似然，为支持或反对零假设提供一个颇具吸引力的证据指标：

$$\underbrace{\frac{P(H_0 \mid y)}{P(H_1 \mid y)}}_{\text{Posterior odds}} = \underbrace{\frac{f(y \mid H_0)}{f(y \mid H_1)}}_{\text{Bayes factor } \mathrm{BF}_{01}(y)} \cdot \underbrace{\frac{P(H_0)}{P(H_1)}}_{\text{Prior odds}}, \tag{1}$$

这里，$\mathrm{BF}_{01}(y)$量化了观测数据在 H_0 下比在 H_1 下有多大可能性（Kass和Raftery，1995）。例如，$\mathrm{BF}_{01}(y) = 10$ 表示数据对 H_0 的支持力度是对 H_1的十倍。由于贝叶斯因子不受赋予各假设的先验赔率影响，因此它通常被认为比后验概率更客观，是一种贝叶斯证据度量。近年来，它在生物医学研究中日益受到欢迎，成为连接贝叶斯推理与频率学派推理的桥梁（Barto˘s、Aust和Haaf，2022年；Kelter和Pawel，2025b；Pawel和Held，2024年；Pourmohamad和Wang，2023年；Rouder、Speckman、Sun、Morey和Iverson，2009年）。

然而，贝叶斯样本量规划面临一项重大挑战。与基于渐近理论推导的经典效能与样本量公式不同，贝叶斯设计很少能得出所需样本量的闭式表达式。在大多数情况下，必须通过模拟评估检验效能和I类错误率，尤其是当先验分布或决策阈值导致目标参数出现非线性依赖关系时。然而，使用频率学派的汇总指标来

贝叶斯准则——即所谓的“校准贝叶斯”方法（Dawid，1982；Grieve，2016；Little，2006；Rubin，1984年）——提供了一种颇具吸引力的折中方案。美国食品药品监督管理局（FDA）和欧洲药品管理局（EMA）等监管机构同样鼓励开展贝叶斯分析，并证明其具有可接受的频率学派运行特性（欧洲药品管理局人用药品委员会／方法学工作组，2025年；Ionan 等，2023年；美国食品药品监督管理局器械与放射健康中心，2010年）。

这种混合方法涉及量化贝叶斯决策指标（如贝叶斯因子或后验概率）的长期表现。对于支持H_0 的贝叶斯因子$\mathrm{BF}_{01}(y)$以及决策阈值 $k < 1$，I类错误率与检验效能的贝叶斯对应量由下式给出

$$P(\mathrm{BF}_{01}(y) < k \mid H_0) \quad \text{and} \quad P(\mathrm{BF}_{01}(y) < k \mid H_1), \tag{2}$$

分别为。概率 $P(\mathrm{BF}_{01}(y) < k \mid H_0)$ 可解释为在 H_0下至少发现证据 $1/k$ 以支持 H_1 的机会——这类似于假阳性决策——因而构成贝叶斯第一类错误率。[1] 概率 $P(\mathrm{BF}_{01}(y) < k \mid H_1)$ 基于类似论证构成贝叶斯功效的一种概念。

根据杰弗里斯尺度（Jeffreys，1939），$k = 1/10$ 对应于反对H_0的“强”证据。若 $P(\mathrm{BF}_{01}(y) < 1/10 \mid H_0) \leq \alpha$ 且 $P(\mathrm{BF}_{01}(y) < 1/10 \mid H_1) \geq 1-\beta$，则该贝叶斯检验在水平 α 上实现I类错误控制，并达到检验效能 $1-\beta$:

$$P(\mathrm{BF}_{01}(y) < k \mid H_0) \leq \alpha \quad \text{and} \quad P(\mathrm{BF}_{01}(y) < k \mid H_1) > 1-\beta. \tag{3}$$

这些量自然对应于频率学派意义上对假阳性和假阴性决策的控制，同时保留了根本上的贝叶斯解释。

对于二元结局（例如肿瘤学研究中的肿瘤缓解率），依据方程（3）进行样本量规划，在实践和概念层面都具有若干优势。二项似然为使用共轭贝塔先验进行解析处理提供了自然框架，而贝叶斯因子则有助于透明地评估证据，且无需依赖渐近近似（Ibrahim、Chen 和 Sinha，2001年；Neuenschwander 等，2009；Spiegelhalter 等，2004年）。此外，贝叶斯预测检验效能或“保证度”——即在给定设计下获得有利证据的先验平均概率——可用于评估先验不确定性下的试验稳健性（Case、Herrera、Rumps 和 Mulcahey，2022年；Grieve，2016，2022年）。这些特性使贝叶斯试验设计尤其适用于自适应、小样本或罕见病情境（Berry，2006）。下文所说的保证度或

[1] 这是成立的，因为 $\mathrm{BF}_{01}(y) = 1/\mathrm{BF}_{10}(y)$，因此 $\mathrm{BF}_{01}(y) < k \Leftrightarrow \mathrm{BF}_{10}(y) \geq 1/k$。

有力证据的概率 H_0，指的是以下内容：

$$P(\mathrm{BF}_{01}(y) > k_f | H_0) \tag{4}$$

要求 $P(\mathrm{BF}_{01}(y) > k_f|H_0) > 1 - \beta_f$达到某个 β_f意味着，当 H_0成立时，至少能以一个有保证的最低概率发现支持 k_f 的 H_0。这会削弱人们对试验结果的信任，因为不仅考虑了方程（ H_1 2）中 H_0 成立时的贝叶斯功效，也考虑了在后者成立时找到足够证据支持 H_0 的概率。换句话说，若不存在效应，那么针对 H_0的、经过校准的有力证据概率——即 $P(\mathrm{BF}_{01}(y) > k_f|H_0) > 1 - \beta_f$——表明该试验有足够的机会揭示这一点。

本文采用BF_{01} 的 Jeffreys(1939) 方向，决策阈值为$k < 1$，使得较小的贝叶斯因子表示反对零假设的证据。我们的重点是确定满足方程（3）中基于证据的校准标准的最小样本量，其中给定 α 和 β 水平。我们采用BF_{01} 方向，以保持与 Jeffreys 原始表述的一致，并使“较小的值表示反对 H_0的证据”这一解释与熟悉的频率学派对 p值的惯例相一致。

1.1 概述

本文提出了一种新颖的方法，利用贝叶斯因子计算二元终点双臂（临床）试验的贝叶斯功效和样本量。我们的方法无需模拟，仅使用数值积分等标准数值技术。因此，用户无需开展、重复并报告蒙特卡罗研究（包括计算并报告所有估计值的蒙特卡洛标准误；由于多种原因，这项工作颇具挑战性，详情请参见 Morris、White 和 Crowther （2019年）、Siepe 等人（2024年）或Kelter（2023年））。基于这一新方法，二元终点 II 期临床试验中贝叶斯因子的检验效能和样本量计算可以在无需模拟的情况下高效完成，且不依赖渐近理论或近似方法。

因此，第2 节介绍贝塔-二项模型，并详细说明总体方法。第2.1 节介绍双试验组 II 期试验中所考虑的不同假设，并针对双侧假设检验对我们的方法进行完整演示，该检验用于检验

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2$$

其中 p_1 和 p_2 分别是对照组和治疗组的成功概率。

节 n 2.2、第 2.3 节和第 2.4 节涵盖了相关推导和详细内容，涉及 e

双臂二元终点 II 期试验中的单侧检验。第2.5 节随后介绍了两个真实世界的 II 期试验案例。第一个是利奥西呱试验，研究利奥西呱与安慰剂相比对系统性硬化症患者的疗效，相关报告见（Khanna 等人，2020年）。第二个是 ICT-107 试验，评估了新诊断胶质母细胞瘤患者的免疫应答等方面（Wen 等人，2019年）。我们通过采用新方法重新分析这两项试验，展示了该方法的实用性，并利用贝叶斯因子，在信息性和无信息设计先验的不同设定下，提供经过校准的贝叶斯设计。此外，我们还说明了不同证据阈值的影响，并为实际应用该方法提供指导。第 3 节讨论研究结果并总结全文。

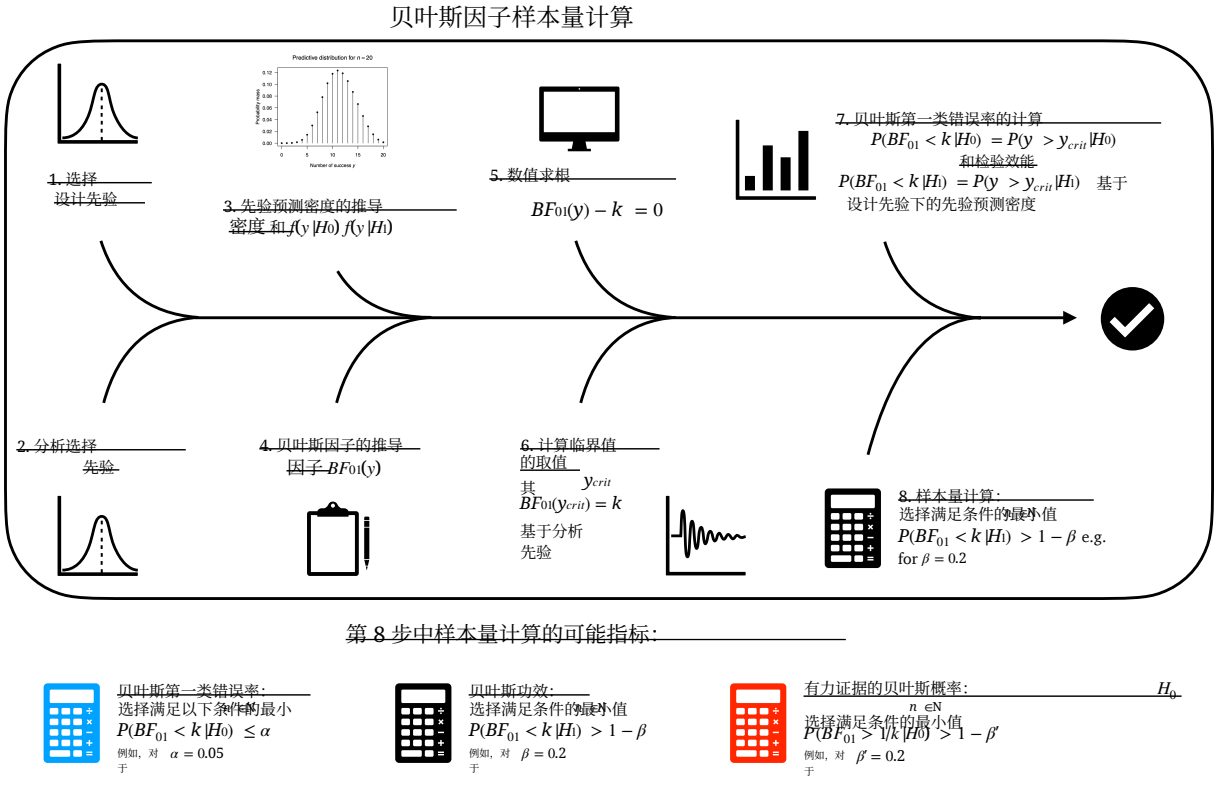

图1：双组二项分布情形下贝叶斯因子的贝叶斯功效与样本量计算方法概览，该方法是在Kelter 和 Pawel（2025b）所述单臂设定的基础上修改和调整而来

2 双臂设定下贝叶斯因子的贝叶斯功效与样本量计算

图1 以图示方式概述了我们的方法，该方法在Kelter 和 Pawel (2025b)的基础上进行了修改和改编，其中对单臂情形作了详细介绍。我们的方法旨在根据选定的设计先验和分析先验，推导先验预测密度 $f(y|H_0)$ 和 $f(y|H_1)$，以及贝叶斯因子$\text{BF}_{01}(y)$（图1中的步骤1至4）。分析先验是用于计算贝叶斯因子的先验，而设计先验则是在试验规划阶段、尚未观测到任何数据之前，用于事先计算检验效能和样本量的先验。我们方法的下一步是求解以下方程的数值解

$$\text{BF}_{01}(y) - k = 0$$

对于指定的证据阈值 k，例如 $k = 1/10$ 或 $k = 1/3$，从而计算临界值 y_{crit}，使得对于我们的分析先验，$\text{BF}_{01}(y_{crit}) = k$ 成立，并且贝叶斯因子达到我们要求的证据阈值，以支持 H_1（图1中的步骤5和6）。根据第三步得到的先验预测，第七步是基于设计先验下的先验预测密度，计算贝叶斯第一类错误率 $P(\text{BF}_{01}(y) < k|H_0)$ 和检验效能 $P(\text{BF}_{01}(y) < k|H_1)$。我们方法的最后一步（第八步）是选择最小样本量 n，使得我们达到所需的检验效能、第一类错误率或支持 H_0 的有力证据概率，亦或达到上述指标的某种组合。

2.1 双侧假设检验

本节将按照图 1 所示流程，详细说明我们针对以下双侧假设检验所采用的方法 $H_0: p_1 = p_2$ 与 $H_1: p_1 \neq p_2$，该检验适用于具有二元终点的双臂II期试验。这里的主要假设是，两组观测数据分别来自两个随机变量 Y_1，Y_2 它们均服从参数为 n_1 和 n_2 以及 p_1 的二项分布 p_2,

$$Y_1 \sim \text{Bin}(n_1, p_1), \qquad Y_2 \sim \text{Bin}(n_2, p_2)$$

概率质量函数如下:

$$f(y_1 \mid p_1) = \binom{n_1}{y_1} p_1^{y_1}(1-p_1)^{n_1-y_1}, \qquad f(y_2 \mid p_2) = \binom{n_2}{y_2} p_2^{y_2}(1-p_2)^{n_2-y_2} \tag{5}$$

对于 $y_1 = 0, \ldots, n_1$ 以及 $y_2 = 0, \ldots, n_2$。我们用 Y_1 表示对照组数据，该组接受标准治疗或可能接受安慰剂；用 Y_2 表示治疗组数据，该组接受

正在研究的新药或治疗。在这种情况下，对于 $i = 1, 2$，我们有 $p_i \in [0,1]$，因此参数空间 $\Theta := [0,1] \times [0,1]$ 是二维单位立方体。完整数据向量$D = (Y_1, Y_2)$在假设两组数据相互独立的情况下，具有如下联合密度：

$$f(y_1, y_2|p_1, p_2) = \prod_{i=1}^{2} p_i^{y_i}(1-p_i)^{n_i - y_i} \tag{6}$$

现在，在 II期试验中，可能会关注两个假设。首先，检验治疗组与对照组的比例是否相等 $p_1 = p_2$；前者接受新药，后者接受标准治疗或安慰剂。检验的假设如下：

$$H_0 : p_1 = p_2 \qquad \text{versus} \qquad H_1 : p_1 \neq p_2 \tag{7}$$

或者，这项检验的一种广为人知的参数化方法会引入差值参数 $\eta = p_2 - p_1$ 和总体均值 $\zeta = \frac{1}{2}(p_1 + p_2)$。采用这种参数化方法，可得：

$$p_1 = \zeta - \frac{\eta}{2}, \qquad p_2 = \zeta + \frac{\eta}{2}$$

并且，方程（7）中的假设可改写为：

$$H_0 : \eta = 0 \qquad \text{versus} \qquad H_1 : \eta \neq 0 \tag{8}$$

这种参数化最初由Gunel 和 Dickey (1974) 提出，另见 Dickey andLientz(1970)、Jamil 等人(2017) 和 Kelter (2025年)。现在，在 II-b期试验背景下，方程（8）中的检验似乎颇具吸引力，因为它可以表示支持以下命题的证据：H_0： $\eta = 0$。在这种情况下，标准治疗或安慰剂与新型治疗被认为疗效相同。另一方面，如果分析试验数据后发现支持以下命题的证据 H_1： $\eta \neq 0$，则有两种可能。第一种是 $p_1 > p_2$，此时对照组（安慰剂或标准治疗）的疗效甚至优于治疗组。第二种是 $p_1 < p_2$，此时治疗组的疗效优于对照组的标准治疗或安慰剂。试验结果可能属于上述任一情况；无论如何，试验完成后估计 p_1 和 p_2，都应作为方程（8）检验结果的补充，以便对检验结果作出全面解读

结果。另一种选择是进行其中一种方向性检验

$$H_0 : \eta \leq 0 \qquad \text{versus} \qquad H_1 : \eta > 0 \tag{9}$$

$$H_0 : \eta = 0 \qquad \text{versus} \qquad H_1 : \eta > 0 \tag{10}$$

$$H_0 : \eta = 0 \qquad \text{versus} \qquad H_1 : \eta < 0 \tag{11}$$

其中， H_0 在方程（9）中意味着安慰剂或标准治疗的效果至少不逊于新型治疗或药物，而 H_1 可以解释为新药的疗效优于安慰剂或标准治疗。另一方面，方程（10）中的检验假设，先验上 $\eta < 0$是不现实的。也就是说， $p_2 < p_1$ 先验上被排除。这种假设可能是合理的，例如对照组接受标准治疗，而治疗组接受标准治疗加附加治疗，且已知附加治疗不会干扰标准治疗。因此，从理论上讲，治疗组的成功概率至少应与对照组一样大。同样，在二元终点衡量失败的情形下，方程（11）也可能适用。此时，先验上排除 $\eta > 0$这种可能性意味着 $p_2 \leq p_1$ 必须成立。如果对照组接受标准治疗，而治疗组接受标准治疗加上已知不会干扰标准治疗的附加治疗，那么这一假设也是合理的。本节仅关注方程（8）中的双侧检验；但如图 1 所示，我们还将针对第 9 个方程、第 10 个方程和第 11 个方程中的检验，在第 2.2 节、第2.3 节和第 2.4 节中，推导贝叶斯因子、设计先验与分析先验，以及用于检验效能和样本量计算所需的先验预测分布。

2.1.1 设计与分析先验的选择

样本量计算方法的第一步如图1中的步骤1和2所示。因此，必须选择所谓的设计先验和分析先验（Grieve，2022年；Kelter和Pawel，2025b；Pawel和Held，2025年）。我们分别用 $P^d_{p_1}$ 和 $P^d_{p_2}$ 表示前者，用 $P^a_{p_1}$ 和 $P^a_{p_2}$表示后者。设计先验的思路是，基于该先验 $P^d_{p_i} i = 1$，2来规划样本量以及贝叶斯第一类错误率和检验效能等性质——参见公式（2）。然而，规划阶段可能会涉及主观信念，因此分析——在此情形下即贝叶斯因子检验——应基于另一种（更客观的）先验分布 P^a （O'Hagan和Stevens，2001年）。为 $i = 1$选择$P^d_{p_i} = P^a_{p_i}$时，2，设计先验与分析先验相同，因此我们基于同一分布进行研究的规划和分析。在实践中，不同利益相关方对分析先验和设计先验分别有不同要求。例如，从监管机构的视角来看，使用非信息性的、“客观的”分析先验几乎是强制要求。高度信息性或“主观”的分析先验则会受到怀疑。另一方面，研究者可能倾向于使用信息性更强的分析先验，也许基于历史数据，以减少试验所需的患者

。关于设计先验，在试验规划阶段，与信息性分析先验相比，使用信息性更强的设计先验较少引起监管视角下的疑虑，前提是以合理的方式校准由此产生的设计的长期频率学派性质。我们将在第 2.5 节稍后说明这一点。

现在，我们从设计先验开始，必须分别在 H_0 和 H_1下选择先验，然后在第二步使用先验预测密度，计算我们在方程（3）中所需的量。因此，为参数 $p_i \in [0,1], i = 1\,2$ 选择两个先验 $P^d_{p_1}$ 和 $P^d_{p_2}$。 $\text{Beta}(a_0, b_0)$分布是二项似然的共轭先验；将其选作先验时，后验分布 $P_{p_i|Y_i}$ 也服从贝塔分布（Held& Saban´esBov´e,2014）：

$$p_i \mid Y_i = y \sim \text{Beta}(a_0 + y, b_0 + n - y)$$

对于 $i = 1$，2。因此，先验的自然选择是贝塔分布。我们遵循Dickey and Lientz (1970)的做法，在 H_0 下选择如下的 Beta 设计先验：

$$p_1 = p_2 = p \mid H_0 \sim \text{Beta}(a_0^d, b_0^d) \tag{12}$$

因此，在 H_0 下： $\eta = 0$，两个概率相同，即 $p_1 = p_2$，并取某个值 $p \in [0,1],$ ，该值服从贝塔设计先验。同样，我们在H_1 下选择独立Beta设计先验： $\eta \neq 0$

$$p_1 \mid H_1 \sim \text{Beta}(a_1^d, b_1^d) \tag{13}$$
$$p_2 \mid H_1 \sim \text{Beta}(a_2^d, b_2^d) \tag{14}$$

对于 H_1 下的分析先验 $P^a_{p_1}, P^a_{p_2}$，我们同样选择独立 Beta 先验，并为 $i = 1$，2 取可能不同的值 a_i^a 和 b_i^a，其中上标表示这些超参数属于我们的分析，而非设计先验：

$$p_1 \mid H_1 \sim \text{Beta}(a_1^a, b_1^a) \tag{15}$$
$$p_2 \mid H_1 \sim \text{Beta}(a_1^a, b_1^a) \tag{16}$$

最后，对于 H_0 下的分析先验 P^a_p： $\eta = 0$，我们选择一个狄拉克先验，其全部概率质量集中在 $\eta = p_2 - p_1 = 0$ 上，并以总体均值的均匀先验 ζ 为条件，即

$$p_1 = p_2 = p | H_0 \sim 1_{\{\eta=0\}} | \zeta \sim U(0,1)$$

用于贝叶斯因子分析。因此，在试验的规划阶段，我们采用独立Beta设计先验，保持模糊而灵活，并同样在H_0 和 H_1 下选择灵活的先验，以便在贝塔-二项模型中进行双侧贝叶斯因子检验，用于分析

阶段（Dickey 和 Lientz，1970；Jamil 等人，2017）。

2.1.2 先验预测分布的推导

图1所示的第三步是推导零假设和备择假设下的先验预测概率质量函数 $f(y \mid H_0)$ 和 $f(y \mid H_1)$。稍后将结合这些概率质量函数和贝叶斯因子 $\mathrm{BF}_{01}(y)$，计算临界值y_{crit}，据此我们可以判断 BF_{01} 是否超过给定阈值，例如 k。然后，根据预测分布，我们可以计算所需的概率 $P(\mathrm{BF}_{01}(y) < k \mid H_0)$ 和$P(\mathrm{BF}_{01}(y) < k \mid H_1)$，参见公式（3）。接下来，通过选择满足预测概率条件的最小样本量 $n \in \mathbb{N}$，即可达到指定的检验效能

$$P(y > y_{crit} \mid H_1) \geq 1 - \beta$$

对于指定的 β，例如 $\beta := 0.2$，其中 $P(\cdot \mid H_1)$ 表示H_1下的预测分布。

我们基于两个观测样本 Y_1（对照组的二元数据）和 Y_2（治疗组的二元数据），以及 H_0 和 H_1 下的独立 Beta 设计先验，得到先验预测分布。由此得到数据 $D = (Y_1, Y_2)$ 的以下预测概率质量函数，该函数基于固定的 n_1，n_2和独立 Beta 先验 $p_1 \sim \mathrm{Beta}(a_1^d, b_1^d)$；详情请参见附录：

$$f(d|H_1) = \binom{n_1}{y_1} \frac{\mathrm{B}(\mathrm{y}_1 + \mathrm{a}_1^{\mathrm{d}}, \mathrm{n}_1 - \mathrm{y}_1 + \mathrm{b}_1^{\mathrm{d}})}{\mathrm{B}(a_1^d, b_1^d)} \cdot \binom{n_2}{y_2} \frac{\mathrm{B}(\mathrm{y}_2 + \mathrm{a}_2^{\mathrm{d}}, \mathrm{n}_2 - \mathrm{y}_2 + \mathrm{b}_2^{\mathrm{d}})}{\mathrm{B}(a_2^d, b_2^d)} \tag{17}$$

利用贝塔函数 $\mathrm{B}(z_1, z_2)$ 的性质：$= \int_0^1 t^{z-1}(1-t)^{z_2-1}$，即 $\mathrm{B}(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1+z_2)}$，其中 $\Gamma(n) := (n-1)!$ 表示伽马函数，上式也可以写成

$$\begin{aligned} &= \frac{\Gamma(n_1+1)\Gamma(y_1+a_1^d)\Gamma(n_1-y_1+b_1^d)}{\Gamma(n_1+a_1^d+b_1^d)\Gamma(y_1+1)\Gamma(n_1-y_1+1)} \frac{\Gamma(a_1^d+b_1^d)}{\Gamma(a_1^d)\Gamma(b_1^d)} \\ &\cdot \frac{\Gamma(n_2+1)\Gamma(y_2+a_2^d)\Gamma(n_2-y_2+b_2^d)}{\Gamma(n_2+a_2^d+b_2^d)\Gamma(y_2+1)\Gamma(n_2-y_2+1)} \frac{\Gamma(a_2^d+b_2^d)}{\Gamma(a_2^d)\Gamma(b_2^d)} \end{aligned} \tag{18}$$

根据所采用的设计先验，我们进一步将a_1^d 和b_1^d 以及a_2^d 和 b_2^d替换为 H_1 在公式（13）中设计先验下选定的值。需要强调的是，公式（17）中的先验预测分布使用了预先选定的设计先验超参数。在 H_0 下的先验预测密度可推导为

$$f(d|H_0) = \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}}{\mathrm{B}(a_0^d, b_0^d)} \mathrm{B}(a_0^d + y_1 + y_2, b_0^d + n_1 + n_2 - y_1 - y_2) \tag{19}$$

详情请参见附录。

2.1.3 贝叶斯因子的推导

图1 中的下一步（第四步）是推导贝叶斯因子。我们采用上一小节讨论的独立贝塔分析先验。不过，需要强调的是，在 a_i^a 和 b_i^a 的分析中，$i = 1$可以选择与设计先验不同的超参数2（设计先验中的超参数为$P_{p_i}^d$ 和 a_i^d，用于 b_i^d和 $i = 1$、2）。当然，也可以采用相同的超参数，使设计先验与分析先验相同。使用萨维奇–迪基密度比，Dickey 和 Lientz（1970，第 219 页）解析地推导出了这一设定下的贝叶斯因子：

$$\mathrm{BF}_{01}(y_1, y_2) := \frac{\mathrm{B}(a_1^a + y_1 + a_2^a + y_2 - 1, b_1^a + n_1 + b_2^a + n_2 - y_1 - y_2 - 1)\mathrm{B}(a_1^a, b_1^a)\mathrm{B}(a_2^a, b_2^a)}{\mathrm{B}(a_1^a + y_1, b_1^a + n_1 - y_1)\mathrm{B}(a_2^a + y_2, b_2^a + n_2 - y_2)\mathrm{B}(a_1^a + a_2^a - 1, b_1^a + b_2^a - 1)} \tag{20}$$

$\mathrm{BF}_{01}(y_1, y_2)$ 是观测到的对照组成功次数 y_1 和治疗组成功次数 y_2 的函数。当两组的样本量 n_1 和 n_2 确定，并且针对 $i = 1$、2 的分析先验超参数 a_i^a 和 b_i^a 确定后，即可计算用于检验公式（8）的贝叶斯因子。值得注意的是，在 2×2 列联表中检验行与列的独立性时，只要行或列中的任一方固定，所得的贝叶斯因子相同（Gunel 和 Dickey，1974；Jamil 等人，2017）。表1展示了这种 2×2列联表的结构；在治疗组和对照组的II-b期试验中，试验组的样本量以及相应的行样本量，会在完成该试验的样本量计算后，预先固定为 n_1 和 n_2。

	成功	失败	$\sum$
治疗	y_2	$n_2 - y_2$	n_2
对照组	y_1	$n_1 - y_1$	n_1
$\sum$	未知	未知	$\sum = n_1 + n_2$

表1：结构 2×2 列联表，其中公式（20）中的贝叶斯因子可用于检验行列独立性（行或列固定时）。

表1 表明，我们的方法也可用于目标可能并非检验 H_0： $\eta = 0$ 对 H_1： $\eta \neq 0$，而是检验行与列之间的随机独立性的情形，例如，在未随机分配至治疗组和对照组的流行病学研究中。

2.1.4 数值求根

根据贝叶斯因子，图1 中的第五步现在包括数值求根。简而言之，其思路是求解以下方程

$$\mathrm{BF}_{01}(y_1, y_2) = k \tag{21}$$

通过数值方法处理固定样本量 n_1，n_2。然而，使用公式（20）并通过数值方法求出 贝叶斯因子$_{01}(y_1$，$y_2)$ $-$ k $=$ 0 的根，例如使用牛顿法（如统计编程语言 R 中的 `uniroot` 函数所实现的方式）并不简单（R 核心团队，2020年），因为解不止一个$(y_1$，$y_2)$。

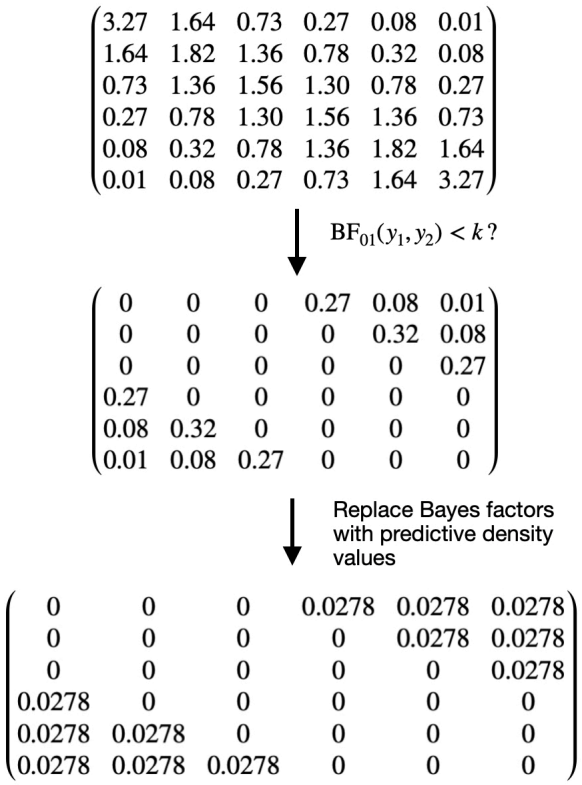

图2：基于 $n_1 = n_2 = 5$ 以及平坦分析先验（ $a_i^a = b_i^a = 1$ ）的贝叶斯因子检验效能计算中临界值计算示意图，适用于 $i = 1$，2。

为便于说明，假设 $n_1 = n_2 = 5$。那么，对于 $y_i = 0$.....5 而言， $i = 1$、2，因为在两组中均可观察到 0 至 5 次成功次数。如图2所示，顶部矩阵将可能的结果可视化为一个 6×6 矩阵，其中行表示对照组中的成功次数 y_1，列表示治疗组中的成功次数 y_2。因此，左上角的条目 3.27 是贝叶斯因子 $\mathrm{BF}_{01}(y_1 = 0, y_2 = 0)$，其右侧的条目是 $\mathrm{BF}_{01}($ $y_1 = 0, y_2 = 1) = 1.64$，依此类推。图2中的顶部矩阵显示了所有取值 $y_i = 0$.....5（对于 $i = 1$、2）时所得的贝叶斯因子 $\mathrm{BF}_{01}(y_1, y_2)$，其中选择了满足 $a_i^a = b_i^a = 1$ 的平坦分析先验

两组中的 $i = 1, 2$。现在，从图 2中的顶部矩阵可以观察到两种现象：首先，矩阵是对称的，这是由于采用了平坦先验和关于 H_0 的双侧检验： $\eta = 0$ 对 H_1： $\eta \neq 0$。其次，最大的贝叶斯因子始终出现在主对角线上，因为此时 $y_1 = y_2$ 成立，且两组中的成功次数相等，从而为 H_0 提供证据： $\eta = 0$。对于固定的行或列，沿行水平方向或沿列垂直方向偏离主对角线条目越远，贝叶斯因子就越小。这是意料之中的，因为在这种情况下， y_1 与 y_2 之间的差异变得更加明显。这两种现象表明，在平坦设计先验下，表示反对H_0： $\eta = 0$ 的较小贝叶斯因子主要出现在矩阵的右上角和左下角边缘。

2.1.5 临界值的计算

在图2中间的第二个矩阵里，所有贝叶斯因子未达到所需证据阈值 $k = 1/3$ 的条目均已设为零。因此，仅保留了以下条目：其中的元组 (y_1, y_2) 得出的结果会导致拒绝 H_0： $\eta = 0$，依据 $\text{BF}_{01}(y_1, y_2) < 1/3$（这意味着 $\text{BF}_{10}(y_1, y_2) > 3$，表明根据 Jeffreys（1961），支持 H_1： $\eta \neq 0$ 的证据至少为中等强度）。因此，非零条目对应的临界值 (y_1, y_2) 能够基于贝叶斯因子提供足够证据 $1/k$ 支持H_1： $\eta \neq 0$。我们将这些元组的集合记为 $\mathcal{Y}$。在下文中，在图2的示例里，我们有 $\mathcal{Y} = \{(y_1 = 0, y_2 = 3)$、$(y_1 = 0, y_2 = 4)$、$(y_1 = 0, y_2 = 5)$、$(y_1 =1, y_2 = 4)$、$(y_1 = 1, y_2 = 5)$、$(y_1 = 2, y_2 = 5)$、$(y_1 = 3, y_2 = 0)$、$(y_1 = 4, y_2 = 0)$、$(y_1 = 5, y_2 =0)$、$(y_1 = 4, y_2 = 1)$、$(y_1 = 5, y_2 = 1)$、$(y_1 = 5, y_2 = 2)\}$。

2.1.6 贝叶斯第一类错误率和检验效能的计算

基于上一步找到的临界值，图1 中的第七步包括计算相关量 $P(\text{BF}_{01}(y) < k \mid H_0)$ 和 $P(\text{BF}_{01}(y) < k \mid H_1)$，其中

$$P(\text{BF}_{01}(y) < k \mid H_0) = P(\mathcal{Y} \mid H_0)$$

and

$$P(\text{BF}_{01}(y) < k \mid H_1) = P(\mathcal{Y} \mid H_1)$$

前者等于贝叶斯第一类错误率，而后者是贝叶斯意义下的频率学派检验效能，参见公式（3）。

因此，请比较图2下方的矩阵。其中，贝叶斯因子值$\text{BF}_{01}(y_1, y_2)$已替换为预测密度的值 $f(y_1, y_2 \mid n_1, n_2, a_1^d, b_1^d, a_2^d, b_2^d)$，该预测密度由方程（17）推导得出。最后，将这些值相加，得到预测概率

$P(\mathrm{BF}_{01}(y) < k \mid H_1)$。例如，在图2 所示的情形下，我们得到贝叶斯功效

$$P(\mathrm{BF}_{01}(y) < k \mid H_1) = P(\mathcal{Y} \mid H_1) = \sum_{(y_1,y_2)\in\mathcal{Y}} f(y_1, y_2 \mid n_1, n_2, a_1^d, b_1^d, a_2^d, b_2^d)$$
$$= 12 \cdot 0.0278 \approx 0.33$$

同样，我们可以在上述计算中用 $f(y_1, y_2|H_0)$ 替代$f(y_1, y_2|H_1)$，从而计算贝叶斯第一类错误率。

2.1.7 贝叶斯因子的样本量计算

现在的最终目标是求出样本量 n，使我们能够断定 $P(\mathrm{BF}_{01}(y) < k \mid H_1) = P(\mathcal{Y} \mid H_1)$ 超过给定阈值，例如 $1-\beta$ 对于 $\beta := 0.2$，从而使我们至少有 80% 的贝叶斯功效，能够发现至少有证据 $1/k$ 支持 H_1（或同样地，有证据 k反对 H_0）。可能会出现这样的情况：对于固定的 n_1 和 n_2，集合 $\mathcal{Y}$ 不够大，导致满足阈值 $\mathrm{BF}_{01}(y) < k$ 的条目数量不足，使得最终的 $\sum_{(y_1,y_2)\in\mathcal{Y}} f(y_1, y_2 \mid n_1, n_2, a_1^d, b_1^d, a_2^d, b_2^d)$ 大于预期阈值 $1-\beta$（例如 80% ）。在最极端的情况下，集合 $\mathcal{Y}$ 甚至可能变为空集，因为不存在任何 y_1 和 y_2 取值组合能产生贝叶斯因子 $\mathrm{BF}_{01}(y) < k$（在这种情况下，图2中的中间矩阵将全部由零组成）。不过，在此类情况下，增加 n_1 和 n_2 最终会使贝叶斯因子足够小。这是因为当样本量 n_1 和 n_2足够大时，必然存在成功次数 y_1 和 y_2，使得 $\mathrm{BF}_{01}(y_1, y_2) < k$ 成立；这是由贝叶斯因子的渐近性质以及后验分布的一致性所决定的（Kleijn，2022年）。

尽管如此，图1 表明，在非信息性的平坦设计先验下，对检验效能有贡献的概率质量主要位于表示该情形的相应矩阵的右上角和左下角边缘。对于较大的 n_1 和 n_2，这些边缘处的预测概率将越来越小，因此在完全非信息性的设计先验选择下，检验效能可能只会非常缓慢地增加。因此，使用更具信息性的设计先验来区分假设 H_0： $\eta = 0$ 与 H_1： $\eta \neq 0$ 或许是合理的。我们将在第2.5节中进一步说明这一现象。

综上所述，我们已选定设计先验与分析先验，推导了贝叶斯因子，并开发了一种简单方法，用矩阵搜索算法替代标准的数值求根。临界值的计算基于该方法，而贝叶斯第一类错误率与检验效能的计算则利用设计先验与分析先验以及贝叶斯因子。综合起来，这使得我们可以针对递增的样本量 (n_1, n_2) 重复这些步骤，并考察贝叶斯功效、I类错误率或有力证据的概率何时达到期望水平。在结束本小节时，我们指出无需渐近理论或蒙特卡罗

因此需要进行模拟；除此之外，仅使用标准数值方法，例如数值积分。

2.2 单侧假设检验 $H_0:\ \eta = 0$ 与 $H_+:\ \eta > 0$

上一节说明了我们针对图 1所示双组二项分布情形中贝叶斯因子的检验效能与样本量计算方法的全部八个步骤。在上一节中，我们重点讨论了 $H_0:\ \eta = 0$ 对 $H_1:\ \eta \neq 0$ 的双侧检验。本节给出 $H_0:\ \eta = 0$对 $H_+:\ \eta > 0$的单侧检验的相应推导。

2.2.1 二项模型与假设

我们观察到

$$Y_1 \sim \mathrm{Bin}(n_1, p_1), \qquad Y_2 \sim \mathrm{Bin}(n_2, p_2),$$

并希望检验

$$H_0 : \eta = 0 \quad \text{vs.} \quad H_+ : \eta > 0, \qquad \eta = p_2 - p_1.$$

在 H_0 下，我们有 $p_1 = p_2 =: p$。在 H_+ 下，我们施加序约束 $p_2 > p_1$。

2.2.2 H_0 与 H_+下的设计先验与分析先验

贝塔密度由下式给出

$$\pi_{\text{Beta}}(p \mid a, b) = \frac{1}{B(a,b)}\, p^{a-1}(1-p)^{b-1}, \qquad 0 < p < 1,$$

其中贝塔函数为

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\, dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

（未正则化的）不完全贝塔函数及其正则化版本为

$$B_x(a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\, dt, \qquad I_x(a,b) = \frac{B_x(a,b)}{B(a,b)}.$$

在 H_0 下：$p_1 = p_2 =: p$ 我们取

$$p \sim \mathrm{Beta}(a_0^d, b_0^d),$$

具有先验密度

$$\pi_0(p) = \frac{1}{B(a_0^d, b_0^d)}\, p^{a_0^d - 1}(1-p)^{b_0^d - 1}, \qquad 0 < p < 1.$$

其中，上标 d 表示这是在 H_0 下用于预测密度的先验。同样，用于计算贝叶斯因子的分析先验也服从 Beta 分布，

$$p \sim \mathrm{Beta}(a_0^a, b_0^a),$$

其中上标 a 表示分析先验。在 H_+ 下，我们从独立Beta设计先验开始

$$p_1 \sim \mathrm{Beta}(a_1^d, b_1^d), \qquad p_2 \sim \mathrm{Beta}(a_2^d, b_2^d),$$

其联合（未截断）密度为

$$\pi_{\text{untr}}(p_1, p_2) = \frac{1}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)}\, p_1^{a_1^d-1}(1-p_1)^{b_1^d-1}\, p_2^{a_2^d-1}(1-p_2)^{b_2^d-1}, \qquad 0 < p_1, p_2 < 1.$$

然后，我们将其截断到序约束区域

$$A = \{(p_1, p_2) \colon 0 < p_1 < p_2 < 1\}.$$

截断先验的归一化常数为

$$C = \iint_A \pi_{\text{untr}}(p_1, p_2)\, dp_1\, dp_2 = P(p_2 > p_1)$$

在独立Beta(a_1^d， b_1^d)和Beta(a_2^d， b_2^d)先验下。用$I_x(a,\ b)$表示的一个便捷形式为

$$C = \int_0^1 \frac{p_2^{a_2^d-1}(1-p_2)^{b_2^d-1}}{B(a_2^d, b_2^d)}\, I_{p_2}(a_1^d, b_1^d)\, dp_2. \tag{22}$$

因此， H_+ 下的截断先验密度为

$$\pi_+(p_1, p_2) = \frac{\pi_{\text{untr}}(p_1, p_2)\, \mathbf{1}\{0 < p_1 < p_2 < 1\}}{C}.$$

图 3 展示了 H_+下未截断先验与截断先验的一个示例。图 3 右侧面板显示，未截断先验下的概率质量分布在整个集合 $[0,1] \times [0,1]$上，而左侧面板中只有集合 A上具有正概率质量，这与 $p_2 > p_1$ 下施加的限制 H_+ 一致： $\eta > 0$。

2.2.3 C 的有限和形式（整数形状）

当 a_1^d 和 a_2^d 为正整数时，我们可以写出 C 的有限和表达式，因此也可以写出 p(y_1, $y_2|H_1$)的有限和表达式。详情见附录。

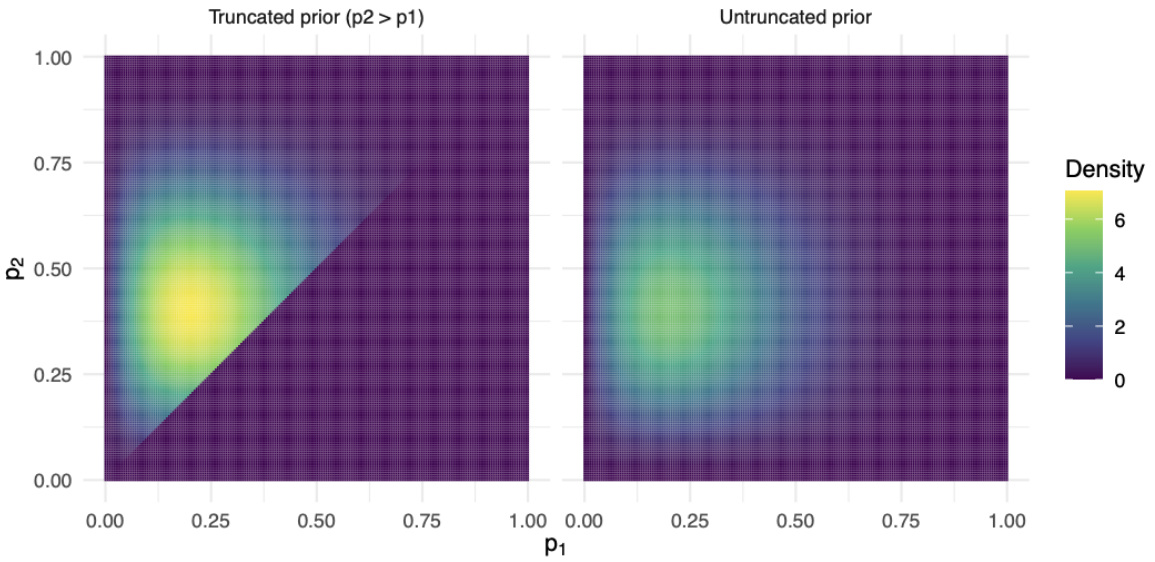

图 3：未截断的 Beta(a_1^d, a_2^d) 设计（或分析）先验（右）和截断的 Beta(a_1^d, a_2^d) 先验（左），适用于 $a_1^d = 2$、$b_1^d = 5$、$a_2^d = 3$ 和 $b_2^d = 4$，其中截断至集合 $\iint_A \pi_{\text{untr}}(p_1, p_2)\, dp_1\, dp_2 = P(p_2 > p_1)$。

2.2.4 H_0 下的预测密度

在 H_0、$p_1 = p_2 = p$ 下，并且在以 p为条件时，

$$Y_1 \mid p \sim \text{Bin}(n_1, p), \qquad Y_2 \mid p \sim \text{Bin}(n_2, p),$$

相互独立。给定 p 的概率质量函数为

$$f(p \mid y_1, y_2) = \binom{n_1}{y_1}\binom{n_2}{y_2} p^{y_1+y_2}(1-p)^{(n_1-y_1)+(n_2-y_2)}.$$

采用先验 $p \sim \text{Beta}(a_0^d,\ b_0^d)$,

$$\pi_0(p) = \frac{1}{B(a_0^d, b_0^d)}\, p^{a_0^d-1}(1-p)^{b_0^d-1},$$

H_0 下的预测密度（边际似然）为

$$\begin{aligned} p(y_1, y_2 \mid H_0) &= \int_0^1 f(p \mid y_1, y_2)\, \pi_0(p)\, dp \\ &= \binom{n_1}{y_1}\binom{n_2}{y_2} \int_0^1 \frac{1}{B(a_0^d, b_0^d)}\, p^{y_1+y_2+a_0^d-1}(1-p)^{n_1+n_2-y_1-y_2+b_0^d-1}\, dp \\ &= \binom{n_1}{y_1}\binom{n_2}{y_2} \frac{B\big(y_1+y_2+a_0^d,\ n_1+n_2-y_1-y_2+b_0^d\big)}{B(a_0^d, b_0^d)}. \end{aligned} \tag{23}$$

2.2.5 H_+下的预测密度

给定（p_1， p_2），概率质量函数可分解为

$$f(p_1, p_2 \mid y_1, y_2) = \binom{n_1}{y_1}\binom{n_2}{y_2} p_1^{y_1}(1-p_1)^{n_1-y_1} p_2^{y_2}(1-p_2)^{n_2-y_2}.$$

H_+ 下的预测密度可推导为

$$p(y_1, y_2 \mid H_+) = \binom{n_1}{y_1}\binom{n_2}{y_2} \frac{I(y_1, y_2)}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)\, C}. \tag{24}$$

更新后的形状参数为

$$A_1 = y_1 + a_1^d, \quad B_1 = n_1 - y_1 + b_1^d, \qquad A_2 = y_2 + a_2^d, \quad B_2 = n_2 - y_2 + b_2^d. \tag{25}$$

其中

$$J = \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1} I_{p_1}(A_2, B_2)\, dp_1 \qquad \text{and} \qquad I(y_1, y_2) = B(A_2, B_2)\big[B(A_1, B_1) - J\big]$$

详情请参见附录。

2.2.6 I 的有限和形式$(y_1,\ y_2)$（整数形状参数）

当 A_2 为正整数时， $I(y_1, y_2)$ 再次存在有限和形式，详情请参阅附录。

2.2.7 H_0 下 $\eta = 0$: $\eta > 0$ 相对于 H_+: $\eta > 0$ 的贝叶斯因子

贝叶斯因子（ H_+ 对比 H_0 ）为

$$\mathrm{BF}_{+0}(y_1, y_2) = \frac{p(y_1, y_2 \mid H_+)}{p(y_1, y_2 \mid H_0)},$$

其中 $p\left(y_1,\ y_2 \mid H_0\right)$ 和 $p\left(y_1,\ y_2 \mid H_+\right)$ 如上所推导，可通过计算公式（24）并将其除以公式（23）得到：

$$
\begin{aligned}
\mathrm{BF}_{+0}(y_1, y_2) &= \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{I(y_1,y_2)}{B(a_1^a,b_1^a)B(a_2^a,b_2^a)\,C}}{\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{B(y_1+y_2+a_0^a,\ n_1+n_2-y_1-y_2+b_0^a)}{B(a_0^a,b_0^a)}} \\
&= \frac{B(a_0^a, b_0^a) I(y_1, y_2)}{B(y_1+y_2+a_0^a,\ n_1+n_2-y_1-y_2+b_0^a) B(a_1^a, b_1^a) B(a_2^a, b_2^a) C} \\
&= \frac{B(a_0^a, b_0^a) B(A_2, B_2)\left[B(A_1, B_1) - J\right]}{B(y_1+y_2+a_0^a,\ n_1+n_2-y_1-y_2+b_0^a) B(a_1^a, b_1^a) B(a_2^a, b_2^a) C} \\
&= \frac{B(a_0^a, b_0^a) B(A_2, B_2)\left[B(A_1, B_1) - \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1} I_{p_1}(A_2, B_2)\, dp_1\right]}{B(y_1+y_2+a_0^a,\ n_1+n_2-y_1-y_2+b_0^a) B(a_1^a, b_1^a) B(a_2^a, b_2^a) C}
\end{aligned}
\tag{26}
$$

其中，附录中的公式（22）可用于替代；若截断贝塔分析先验的形状参数取整数值，则可用附录中的公式（29）替代公式（26）分母中的 C。请注意，在上面的公式（26）中，我们将预测密度里的设计先验超参数 a_i^d, b_i^d 替换为用于计算贝叶斯因子的、可能不同的分析先验的超参数 a_i^a, b_i^a。还请注意，观测数据y_1, y_2 以及分析先验超参数 a_1^a, b_1^a, a_2^a, b_2^a 会以更新后的形状参数的形式影响贝叶斯因子，如公式（25）所示。计算上，贝叶斯因子 BF_{+0} 可通过数值积分和贝塔函数计算。形状参数为整数时，计算 C 不再需要数值计算，从而进一步简化计算工作。

2.3 $H_0:\ \eta = 0$ 与 $H_-:\ \eta < 0$的单侧假设检验

针对 $H_0:\ \eta = 0$ 与 $H_-:\ \eta < 0$的单侧假设检验与前文讨论的$H_0:\ \eta = 0$ 与 $H_-:\ \eta > 0$ 检验非常相似。因此，所需的推导全部放在附录中，我们仅总结主要差异。首先，由于现在施加的序约束为 $p_1 < p_2$，截断区域A 也相应发生变化，这会影响 $H_-:\ \eta < 0$下的预测区域。H_0 下的预测密度保持不变，贝叶斯因子仍可通过标准数值积分求得。

2.4 $H_0:\ \eta \leq 0$与 $H_1:\ \eta > 0$的单侧假设检验

$H_0:\ \eta \leq 0$ 与 $H_1:\ \eta > 0$ 的单侧假设检验，也与上文处理的 $H_0:\ \eta = 0$与 $H_-:\ \eta > 0$ 的检验类似。因此，所需的推导过程全部放在附录中。再次说明，在 H_+ 和 H_- 下，设计先验与分析先验的截断区域会受到所施加的序约束的影响。也就是说，在 H_+ 下，我们将先验限制为

prior to $p_2 > p_1$，而在 H_-下，我们将先验限制为 $p_1 \geq p_2$。

2.5 II期试验中的真实数值示例

上一节详细介绍了我们针对二元终点双臂临床试验中贝叶斯因子的检验效能和样本量计算方法。正如前几节所述，我们不会在 H_0 或 H_1条件下模拟试验数据。此外，所提出的方法不依赖渐近理论，因此给出的样本量和检验效能值都是精确的。本节将演示如何将我们的方法应用于二元终点双臂II期试验。因此，我们采用了两项已发表随机对照试验的数据，这些试验报告了每组成功和失败的确切人数。我们对这两项试验进行了再分析，并解读了该方法的输出结果，展示了在标准贝叶斯分析之外采用我们的方法所带来的益处。此外，我们还说明，如何运用我们的方法从头规划试验，以达到预期的检验效能和I类错误率，是很简单的。[2]

2.5.1 利奥西呱II期试验

利奥西呱试验是一项针对早期弥漫性皮肤型系统性硬化症的IIb期研究，评估了利奥西呱治疗52周的疗效和安全性（Khanna等，2020年）。主要终点是改良Rodnan皮肤评分（mRSS）病情进展，定义为第52周时较基线增加 >5个单位，且增加 $\geq 25\%$。成功定义为无病情进展。上述计数（见表2 ）——（18.6% 对 36.7%）表明治疗获益（OR 0.40，名义

表2：RISE-SSc IIb期试验第52周改良Rodnan皮肤评分（mRSS）病情进展（>5 个单位且 $\geq$25%）

	病情进展	无病情进展
利奥西呱（n=59）	11	48
安慰剂（n=60）	22	38

$p = 0.024$）（Khanna 等人，2020年）。

该试验检验了利奥西呱的优效性，类似于 H_0： $p_1 = p_2$ 与 H_+： $p_1 < p_2$。对于皮肤病情进展这一终点，成功定义为没有病情进展，因此，支持 H_+的证据意味着利奥西呱治疗可能使没有病情进展的患者比例更高（或者说，使皮肤病情进展的患者比例更低）。因此，理想情况下，试验数据应提供支持 H_+的证据。

[2]所有分析和图表均可通过 R 包`binbf2arm`重新生成，并运行我们在以下网址提供的 Quarto复现脚本：https://osf.io/zsrfh/overview?viewonly=351aecc896d0468991569c99ae58beaf。该R 包可通过 CRAN 获取，网址为https://cran.r-project.org/web/packages/bfbin2arm/index.html。

我们采用第2.2 节推导的理论，并为贝叶斯功效和在 H_0条件下有力证据的概率均采用适中的阈值，即 $k = 1/3$ 和 $k_f = 3$分别用于公式（3）和公式（4）。

在分析中，我们为每个组使用独立的均匀 Beta$(1, 1)$ 先验（对共同p 零假设模型也使用均匀 Beta BF_{+0} 先验）。在这些先验下，支持零假设的贝叶斯因子$\mathrm{BF}_{+0} = 4.32$, 表明有中等程度的证据支持 $H_+ : p_2 > p_1$ 根据 Jeffreys (1939年)的量表。因此，人们可能会倾向于得出结论：与安慰剂相比，在利奥西呱治疗下病情进展发生得更少。尽管这一结论是可以接受的，但与之相关的检验效能或错误控制保证并不存在，从监管视角来看，这可能会带来问题。此类视角可能要求进行第一类错误控制或保证最低的预先设定的功效，参见欧洲药品管理局人用药品委员会／方法学工作组 (2025年)；美国食品药品监督管理局器械与放射健康中心 (2010年)以及 Ionan 等 (2023年)。因此，出现了三个问题：

1. 每组需要招募多少患者，才能确保检验效能至少达到（例如）80%，且 I 类错误率为（例如）5%？该试验招募的患者是否足以确保检验效能和第一类错误的长期保证？

2. 根据该试验招募的119名患者，贝叶斯功效和 I 类错误率分别是多少？也就是说，根据治疗组59名患者中48人成功、对照组60名患者中38人成功的情况，我们可能会关注基于这些样本量得出的检验效能。

3. 与方程（*3*）中的贝叶斯功效和 I 类错误率对应量相比，采用我们的方法校准的设计，其完全频率学派I 类错误率（或检验效能）是多少？从监管视角来看，可能需要进行完全频率学派校准。仅提供贝叶斯功效和 I 类错误率对应量或许并不足够。

图4 展示了我们针对第一个问题进行检验效能和样本量校准的结果。上图显示了用于 p_1 和 p_2的设计先验与分析先验，两者均为均匀分布，因此属于非信息性先验（上图中的实线和虚线重叠，因此在这种情况下，视觉上无法区分设计先验与分析先验）。中图给出了贝叶斯功效、I类错误率以及频率学派功效（下文将进一步讨论）随样本量变化的情况。[3] 因此，中图展示了第2节概述的样本量计算方法的关键结果。下图则展示了有力证据的概率（见公式（4））随样本量变化的情况。

[3]重要的是，由于贝塔-二项模型存在波动（这是数据离散性所致——详情另见Kelter 和 Pawel （2025b）——校准可确保接下来十个样本量下的检验效能不会低于所需阈值（同样，对于 I 类错误率，校准可确保其在接下来十个样本量下不会超过指定阈值）。

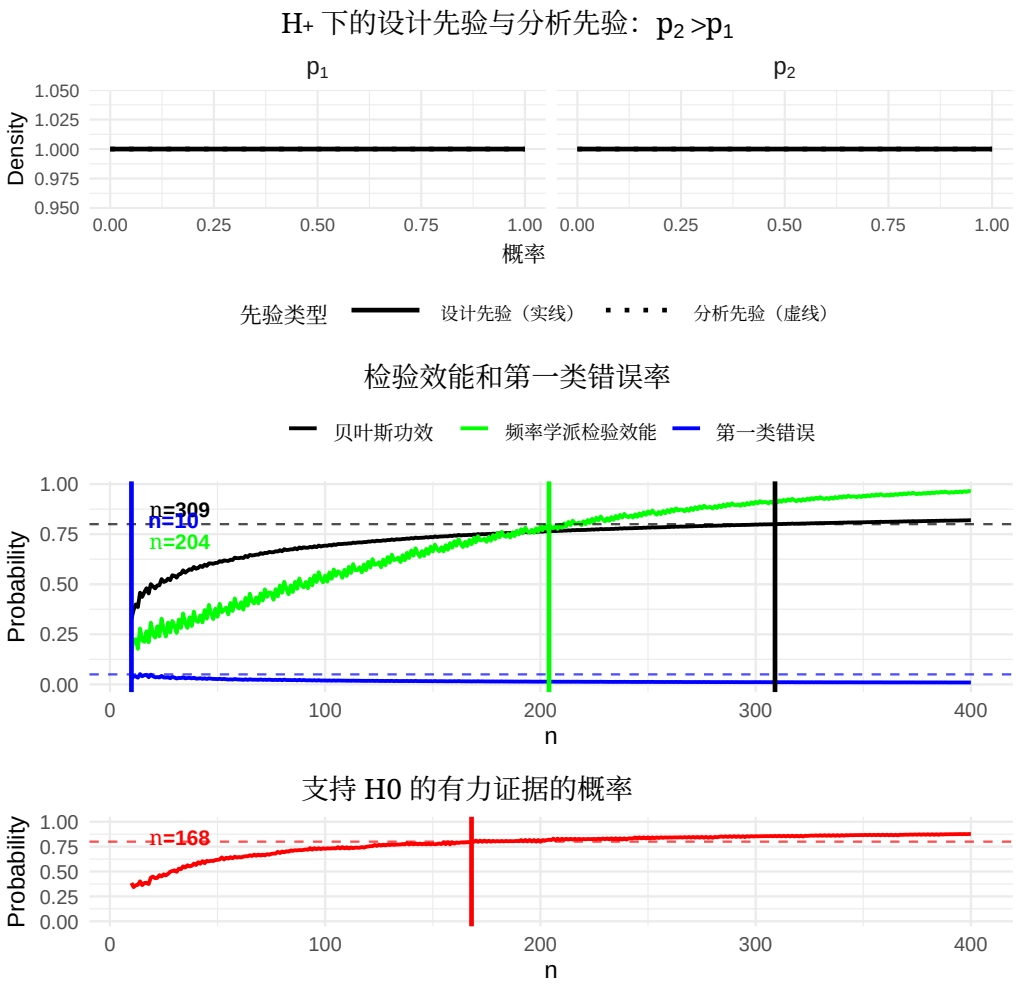

图4：利奥西呱试验的贝叶斯功效与样本量计算，其中检验 H_0： $p_1 = p_2$ 与 H_1： $p_2 > p_1$。采用平坦的设计先验和分析先验，中等证据阈值为 $k = 1/3$、 $k_f = 3$。校准结果显示，在 H_0下，检验效能为 80%、I类错误率为 5%，且支持 H0 的有力证据的概率为 80%。频率学派功效是在 $p_1 = 0.4$和 $p_2 = 0.6$下得出的。

现在，关于上述第一个问题，图4 的中图显示，检验效能达到0.80 时的样本量为 $n = 309$；贝叶斯第一类错误 ≤ 0.05 在 $n = 10$时即可达到；而针对 H_0， $P(\mathrm{BF}_{+0} > k_f | H_0) \geq 0.80$时，有力证据的概率达到 $n = 168$。请注意，样本量始终指两个试验组各自所需的患者人数。因此，在假定随机化概率相等的情况下，要达到80%的检验效能，两个组都需要309 名患者，远多于实际招募的119 名患者。因此，尽管贝叶斯因子表明零假设 H_0： $p_1 = p_2$的证据强度为中等，但我们的方法显示，该试验纳入的患者不足，无法达到80%的检验效能来提供支持 H_1： $p_1 < p_2$ 的贝叶斯统计证据

因素。[4]

这引出了上文的第二个问题：基于 119 名入组患者，检验效能实际上有多高。尽管两组人数并不均衡（图 4中的检验效能计算假设两组人数均衡），但使用我们的方法计算不均衡试验组的检验效能十分简单。第 2节详述的方法绝不需要假设$n_1 = n_2$；对于治疗组的 $n = 59$ 名患者和对照组的 $n = 60$ 名患者，我们设定 $n_1 = 60$、 $n_2 = 59$、 $y_1 = 38$以及 $y_2 = 48$，并计算相应的（贝叶斯）检验效能。计算结果为 71。04%，低于期望的80%。我们还可以计算（贝叶斯）第一类错误率；对于利奥西呱试验的 $0.017 < 0.05$ 名119 名患者，计算结果为此。

现在，图 4 中的检验效能和第一类错误率是按照公式（3）中的贝叶斯版本计算的。美国食品药品监督管理局或 EMA 等监管机构可能要求严格的频率学派检验效能，尤其是要求控制第一类错误。因此，也有必要研究这些情形下所需的样本量。要计算频率学派I类错误率，只需注意，在 H_0： $p_1 = p_2$下，原假设集合是单位立方体 $[0, 1]$中的主对角线。因此，使用足够精细的网格，并计算该网格中所有参数组合 (p_1, p_2) 下拒绝 H_0 的第一类错误率最大值，即可近似求得该集合下的上确界，也就是频率学派I类错误率。[5] 通过网格近似得到的最大值对应的第一类错误率为 $0.09 > 0.05$。因此，从纯粹的频率学派视角来看，该设计未经过校准。采用更严格的证据阈值 k，例如 $k = 1/10$，可以改善频率学派I类错误率，但同时也会增加达到预期检验效能所需的患者人数。

因此，图5 展示了我们的方法在将证据阈值调整为有力证据的阈值 $k = 1/10$并采用略具信息性的设计先验时所得的结果。在图4中，设计先验是平坦且非信息性的。然而，我们先验地假设，利奥西呱治疗组中未出现病情进展的患者比例低于安慰剂组。因此，在对照组中，为接近零的小概率赋予比接近一的大概率更高的先验密度值是合理的。在治疗组中，我们会为接近一的大概率赋予比接近零的小概率更高的密度值，以表达我们对该药物有效的预期。因此，我们采用 a_1^d =1， $b_1^d = 2$和 $a_2^d = 2$， $b_2^d = 1$，如图5顶部的图所示。现在，图5中间的图显示，要达到 80% 的贝叶斯功效，所需样本量大约减半：现在只需$n = 136$ 名患者——即每个试验组68名患者。 此外，频率学派的第一类错误

[4]尽管第一类错误是从贝叶斯视角校准的，该试验仍未招募足够的患者，以达到针对 H_0所需的80% 有力证据概率。因此，本需要168名患者。[5]对于 H_0： $p_1 = p_2$ 对 H_-： $p_2 \leq p_1$ 或 H_+： $p_2 > p_1$ 的方向性检验，计算方式完全相同；而对于 H_-： $p_2 \leq p_1$ 对 H_+： $p_2 > p_1$ 的检验，原假设集合 $\{p_2 \leq p_1 | p_1, p_2 \in [0, 1]\}$ 是单位立方体 $[0, 1]$中的一个矩形。在该矩形上使用网格同样可以近似上确界。

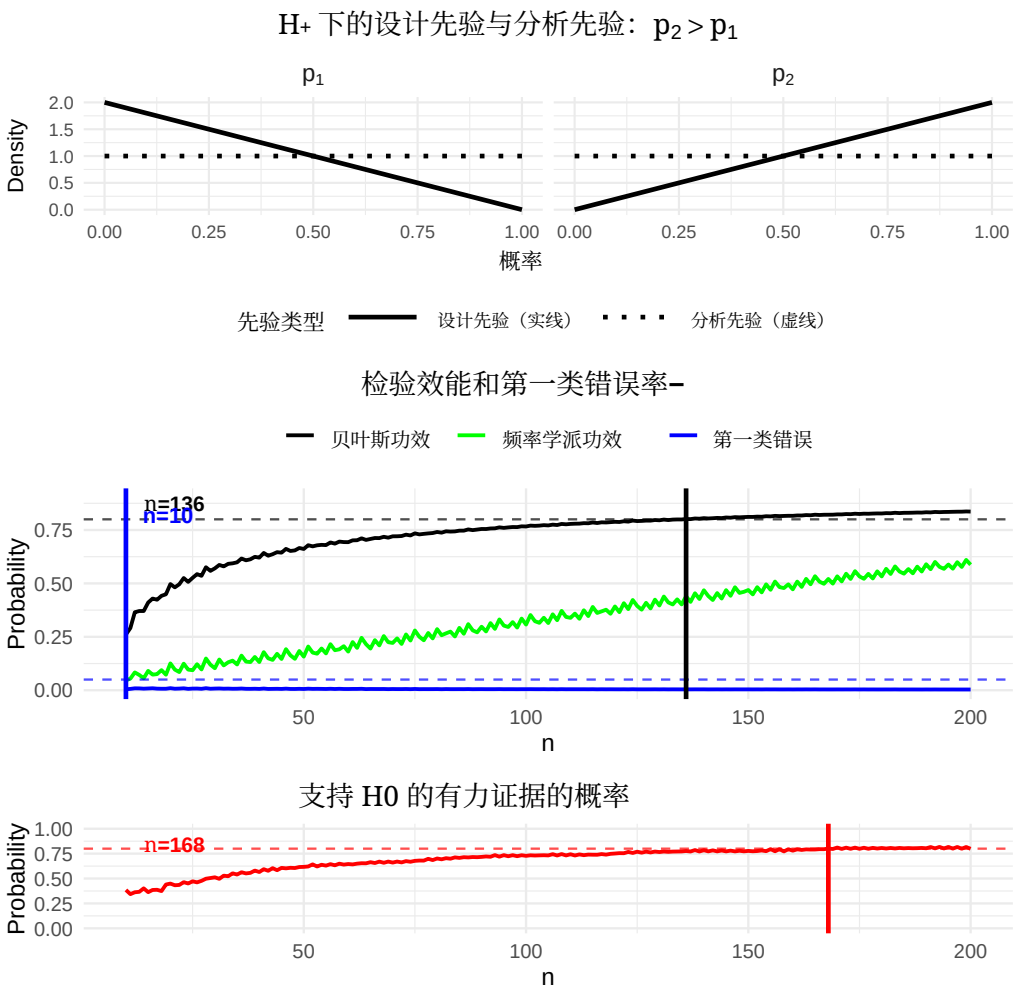

图5：利奥西呱试验的贝叶斯功效与样本量计算，其中检验 H_0：$p_1 = p_2$ 与 H_1：$p_2 > p_1$。采用信息性 $B(1,2)$和 $B(2,1)$设计先验及平坦分析先验，强证据阈值为 $k = 1/10$，在 H_0 为 $k_f = 3$时，有力证据的证据阈值为 $k = 1/10$。校准结果显示，在 H_0下，检验效能为80%、I类错误率为5%，且有力证据的概率为80%时的结果。频率学派功效是在 $p_1 = 0.4$和 $p_2 = 0.6$下获得的。

率现在已完成校准，且在原假设集合上求得的最大值为 $0.021 < 0.05$。其他指标，例如贝叶斯第一类错误率以及在 H_0 下有力证据的概率，并未改变，因为 H_+ 下的设计先验仅与贝叶斯功效有关。[6]

最后，请看图4 和图5中间的绿色线。监管机构也可能要求达到一定的频率学派检验效能。因此，我们允许指定一个参数组合(p_1, p_2)，用于表示治疗组与

[6]当然，也可以在 H_0下选择信息性设计先验，也就是说，为 a_0^d和 b_0^d选择不同的值。例如，相较于该区域之外，假设 $p_1 = p_2$ 在区域 $[0.25, 0.75]$ 内成立可能更符合实际。将设计先验的中心设在 $p = 0.5$，可以表达在 H_0下的这种先验信念。

对照组之间的预期效应，并根据这些固定的参数选择计算贝叶斯因子的频率学派检验效能。在图4 和图5中，我们选择了 $p_1 = 0.4$ 和 $p_2 = 0.6$，表示我们预计安慰剂组中有 40% 的患者不会出现病情进展。我们假设治疗组中的这一比例进一步升至约 60%。[7] 图4 显示，在这种情况下，需要$n = 102$ 名患者/组，或总计204 名患者，才能以完全频率学派的方式达到所需的 80% 检验效能。转到图5时，设计先验不会影响频率学派检验效能计算，但更严格的证据阈值 $k = 1/10$ 会产生影响。因此，图5中间图所需的患者数有所增加，即使总计 $n = 200$ 名患者，也无法达到频率学派检验效能要求。

总的来看，利奥西呱试验中贝叶斯因子所表明的支持 H_+ 的中等证据，并没有可靠的长期错误保证作为支撑。在采用平坦或信息性设计先验时，无论使用中等还是强证据阈值，要达到足够高的（贝叶斯）检验效能或频率学派功效，需要的患者人数都将超过实际招募的 n =119人；同时，要使 H_0有力证据的概率足够高，也需要更多患者。

我们的方法表明，在略具信息性的设计先验下，总共需要136名患者才能实现完全校准的设计；但在假定 $p_1 = 0.4$ 和 $p_2 = 0.6$且使用强证据阈值时，检验效能的频率学派校准除外。

2.5.2 ICT-107试验

ICT-107 II期试验将124名新诊断胶质母细胞瘤患者随机分组（81人接种ICT-107疫苗，43人接受安慰剂），以评估免疫应答（Wen 等人，2019年）。ICT-107项目采用自体树突状细胞（DC）疫苗方案，在新诊断胶质母细胞瘤患者中进行测试；其关键随机证据来自一项纳入124名患者的II期双盲安慰剂对照试验。补充表格报告了确切的二元免疫学应答者人数：

表3：二元终点：免疫应答（ICT-107试验）

	应答者（成功）	无应答者
ICT-107（n=81）	49	32
安慰剂（n=43）	12	31

这些计数——见表3 ——（60.5% 对比 27.9%）表明治疗获益（OR 0.29,$p < 0.001$）（Wen 等人，2019年）。

该试验检验了 ICT-107疫苗的优效性，类似于 H_-： $p_2 \leq p_1$ 与 H_+： $p_2 > p_1$。对于免疫应答这一终点，这意味着支持 H_+ 的证据表示治疗

[7]请注意，根据表2，试验数据表明，真实比例甚至更高。利奥西呱组中约有 $\approx 80\%$ 的患者未出现病情进展，而对照组的这一比例约为 $\approx 66\%$。然而，在规划试验时，通常无法获得这些信息，或只能通过历史信息或其他试验数据获取。因此，我们建议针对p_1 和 p_2的多个现实参数取值计算频率学派功效。

ICT-107可使免疫应答患者的比例高于安慰剂（即取得成功）。因此，理想情况下，试验数据应提供支持 H_+的证据。

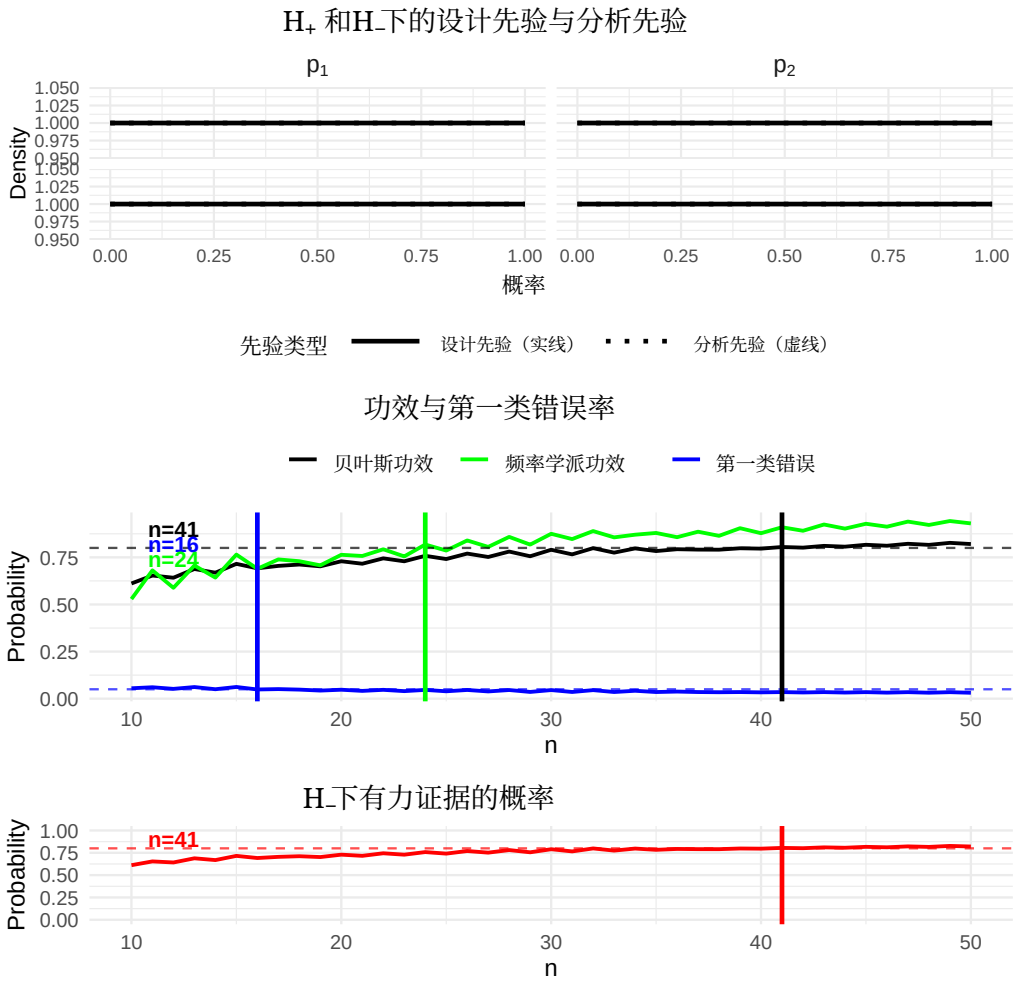

图6：ICT-107试验的贝叶斯功效与样本量计算，其中检验 H_-： $p_1 \geq p_2$ 与 H_+： $p_2 > p_1$。采用平坦的设计先验和分析先验，中等证据阈值为 $k = 1/3$，有力证据的证据阈值为 $k_f = 3$，用于判断 H_-。校准结果展示了在 H_0下，检验效能为80%、I类错误率为5%，且出现有力证据的概率为80%时的结果。频率学派功效是在 $p_1 = 0.3$和 $p_2 = 0.6$下获得的。

我们采用第2.2 节推导的理论，并在贝叶斯功效和 H_-下有力证据的概率计算中均使用中等阈值，即 $k = 3$ 和 $k_f = 3$分别用于方程（3）和方程（4）。

在分析中，我们对每个试验组均采用相互独立的均匀分析先验 Beta(1,1)，无论在 H_- 还是 H_+ 下都是如此。在这些先验下，支持 H_+ 的贝叶斯因子 BF_{+-} 得到 BF_{+-}=3702。65，表明有极强证据支持 H_+： $p_2 > p_1$。因此，人们可能会倾向于得出结论，认为接种疫苗后的免疫应答率更高。不过，如果后者属实，试验是否有足够的检验效能来找到支持 H_- 的有力证据，仍未可知。

此外，要使 H_+ 的（贝叶斯或频率学派）检验效能至少达到 80%，需要多大的样本量，以及要控制第一类错误率（例如控制在 5%）需要多大的样本量，这些问题仍未明确。图6展示了在两组均采用平坦的设计先验和分析先验、且分别在 H_- 和 H_+ 下时，这些问题的答案。假设采用平衡随机分配，图6中间的图显示，总共 $n = 41$ 名患者就足以达到 80% 的贝叶斯检验效能，因此每个试验组有 $n = 21$名患者就足以达到 80% 的贝叶斯检验效能。总共 $n = 24$名患者时，频率学派检验效能已达到校准水平；这里假设对照组和治疗组的真实比例分别为 $p_1 = 0.3$ 和 $p_2 = 0.5$ 。因此，每个试验组有 12 名患者就足以达到 80% 的频率学派检验效能。贝叶斯第一类错误率（参见图6中间图中的蓝色竖线）在总共 $n = 16$名患者时达到校准水平，因此每个试验组有 8 名患者就足够了。有力证据的概率（见图6下图）要求总共 $n = 41$ 名患者，这与原试验的样本量相符，参见表3。[8] 当 $n_1 = n_2 = 8$ 时，贝叶斯第一类错误率已校准至 5%，但频率学派第一类错误率（图中未显示）峰值达到 32.7%，从完全频率学派视角来看，这是无法接受的。

因此，图7 展示了将证据阈值提高到强证据阈值时的结果 k =1/10 而不是 $k = 1/3$。此时，（频率学派）I类错误率降至12%，约为采用 $k = 1/3$时的三分之一，但仍高于预期的5%。图7还显示，为达到80%（贝叶斯或频率学派）功效，现在需要更多患者（两组总计$n = 115$ 名患者，以达到80%的贝叶斯功效；两组各需 $n = 50$名患者，以达到80%的频率学派功效），因为目前的证据阈值更加严格。根据这些中间设计结果，还需采取两个步骤，才能制定出从频率学派角度也经过完全校准（I类错误率为5%）的设计：进一步提高证据阈值，使其更加严格，从而进一步降低I类错误率。

图8 展示了将证据阈值提高到更严格水平时的结果 $k = 1/30$（且 $k_f = 30$），根据 Jeffreys （1939）的量表，这相当于非常强的证据。因此，要达到所需的 80% 检验效能，所需样本量应当增加。然而，由于在平坦设计先验下，每个试验组所得的 $n = 115$ 名患者，以及图$k = 1/10$ 中的7 ，所得试验样本量已接近 ICT-107 试验中使用的 $n = 124$ 名患者——参见表3——我们也转而采用信息性稍强的设计先验。[9]这些先验显示于图8 顶部的图中，与我们对疫苗疗效的预期相似。因此， p_1 在 H_+ 下的设计先验形状： $p_2 > p_1$ 反映了我们的信念：对于 p_1 而言，较小的概率先验上比对于 p_2 更有可能（比较顶部图）

[8]频率学派I类错误率显示在`powerbinbf01()` 函数的数值输出中。该函数已在我们的R软件包`bfbin2arm`中实现。有关如何使用该函数的详情，请参阅附录。[9]重要的是，分析先验仍为平坦先验，贝叶斯因子也基于这些非信息性分析先验计算。从监管机构的角度来看，后一项很重要，例如EMA或美国食品药品监督管理局（欧洲药品管理局人用药品委员会／方法学工作组，2025年；美国食品药品监督管理局器械与放射健康中心，2010年）。

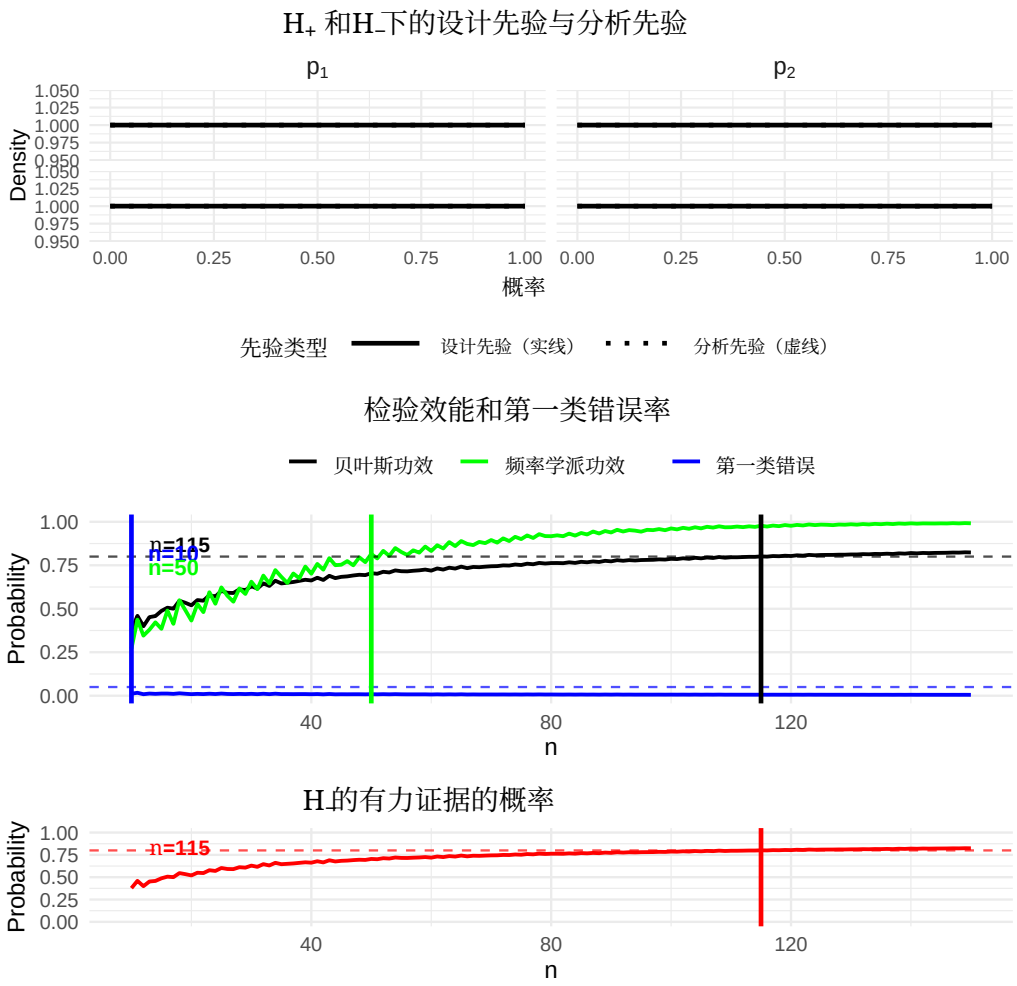

图7：ICT-107试验的贝叶斯功效与样本量计算，其中检验 H_- ： $p_1 \geq p_2$ 对比 H_+ ： $p_2 > p_1$ 。采用平坦的设计先验和分析先验，中等证据阈值为 $k = 1/10$，有力证据的证据阈值为 $k_f = 10$，用于 H_-。校准结果显示，在 H_0下，检验效能为80%、I类错误率为5%、获得有力证据的概率为80%时的结果。频率学派功效是在 $p_1 = 0.3$和 $p_2 = 0.6$下获得的。

图8中）。同样，对于 p_2 而言，较大的概率先验上比对于 p_1 在 H_+下更有可能。采用贝叶斯因子的校准 II 期设计所得样本量显示于图8的中图和底图中。

对于80%的贝叶斯功效， $n = 36$ 名患者足以满足每个试验组。对于80%的频率学派功效， $n = 77$ 名患者足以满足每个试验组。贝叶斯第一类错误率甚至在$n = 10$ 名患者总计时即已校准，而频率学派I类错误率峰值达到0.041在 H_- 此处下，并且同样得到校准。[10] 然而，图8底部的图显示，即使对于 $n = 100$,

[10]这同样是通过对 H_-下的集合进行网格近似计算得出的。对于该网格中的每个点——即单位立方体 $[0, 1$ 中参数值 p_1 和 p_2 的一种组合——都会计算分析先验分布（平坦）下贝叶斯因子在阈值 k 处的I类错误率。随后，将所有参数值（或该网格中的点）对应结果的最大值报告为频率学派I类错误率。

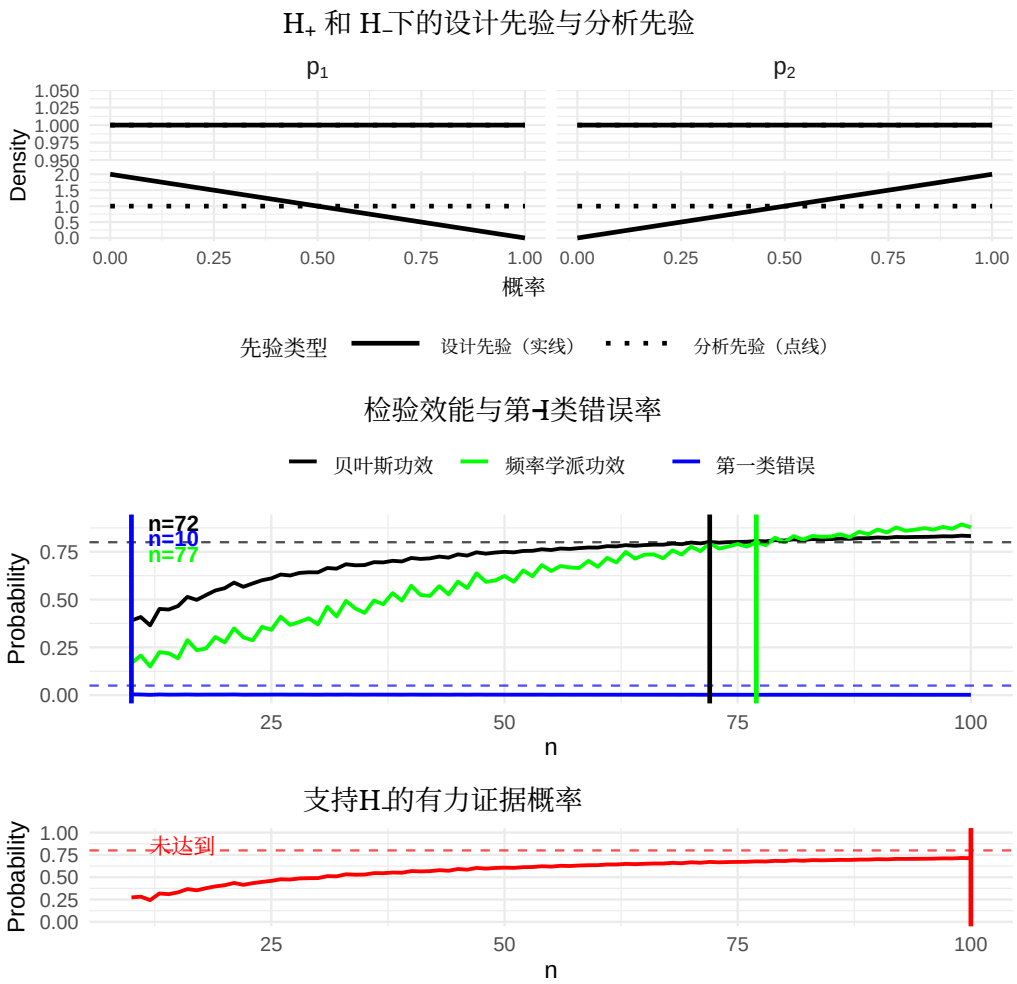

图8：ICT-107试验的贝叶斯功效与样本量计算，其中检验 H_-： $p_1 \geq p_2$ 与 H_+： $p_2 > p_1$。对照组和治疗组分别采用信息性设计先验 $B(1,2)$和 $B(2,1)$，并选用平坦分析先验。结果基于非常强的证据阈值 $k = 1/30$ 以及针对 H_-有力证据的证据阈值 $k_f = 30$。校准展示了在 H_0下80%检验效能、5% I类错误率和80%有力证据概率的结果。频率学派功效是在 $p_1 = 0.3$和 $p_2 = 0.6$下得到的。

有力证据的概率对于 H_0 未达到80%的要求。[11] 现在，有力证据的概率对于 H_0 所需患者远多于检验效能或I类错误率的原因在于，在 H_- 下设计先验仍然是平坦的。这使得难以积累有利于 H_- 的证据，因为关于 p_1 和 p_2 及其在 H_- 下关系的任何预期均未反映在设计先验的形状中。

因此，图9 展示了在 H_-下将设计先验调整为略具信息性的形式后得到的最终结果，这反映了在 H_- 下： $p_2 \leq p_1$ 我们认为

[11]对于每个试验组有 $n = 100$ 名患者的情况，我们的方法得出的 H_0 有力证据的概率为 71。5%，略低于要求的 80%。

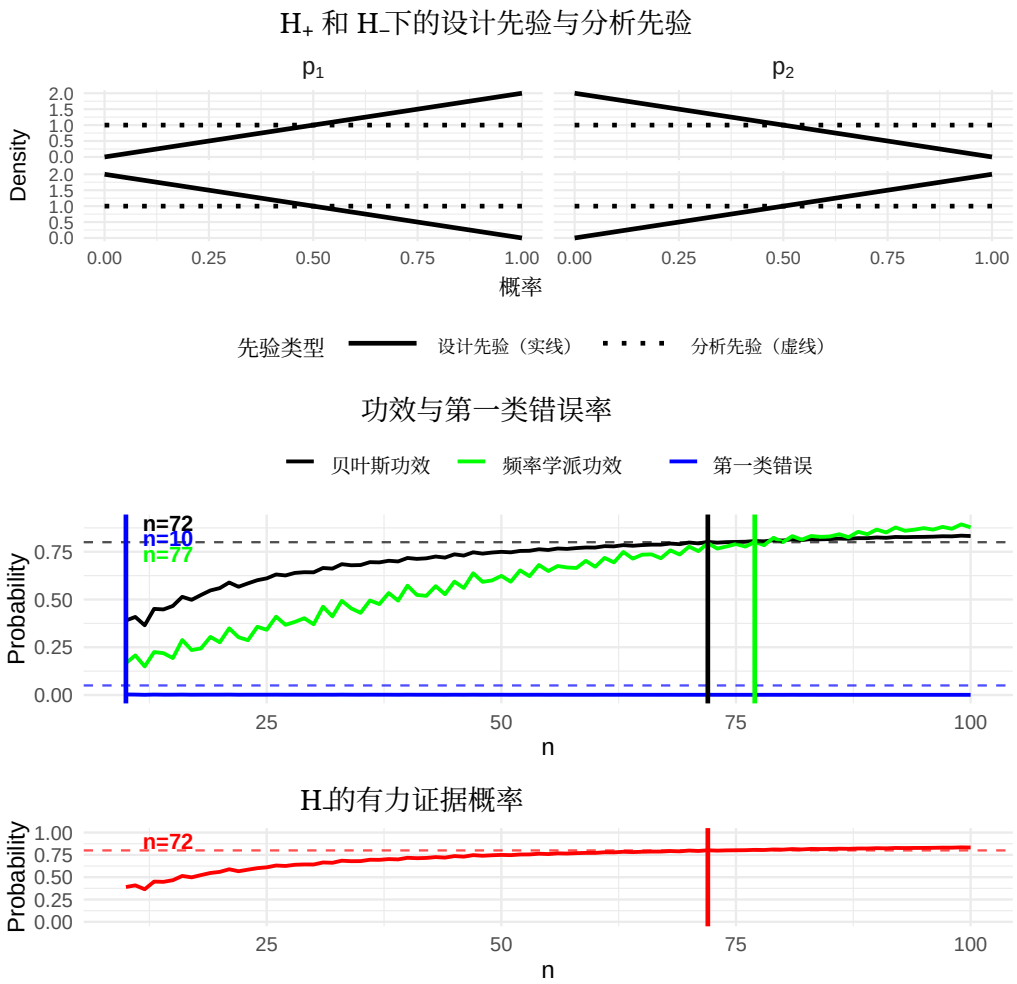

图 9：ICT-107试验的贝叶斯功效与样本量计算，其中检验 H_-： $p_1 \geq p_2$ 与 H_+： $p_2 > p_1$。对照组和治疗组分别采用信息性设计先验 $B(1,2)$ 和 $B(2,1)$，并选用平坦分析先验。此外，现在 H_-： $p_1 \geq p_2$ 下的设计先验也具有信息性，对照组和治疗组分别选用 $B(2,1)$ 和 $B(1,2)$。结果基于非常强的证据阈值 $k = 1/30$，以及针对 H_- 的有力证据阈值 $k_f = 30$。校准结果显示，在 H_0 下，功效为 80%、I类错误率为 5%、有力证据的概率为 80% 时的结果。频率学派功效是在 $p_1 = 0.3$ 和 $p_2 = 0.6$ 下获得的。

概率 p_1 大于或等于 p_2。图9 显示，中间图中的（贝叶斯或频率学派）检验效能结果保持不变，因为这些结果仅取决于证据阈值 k 以及 H_+下的设计先验。此外，频率学派和贝叶斯的第一类错误率结果也保持不变。底部图显示，现在只需 $n = 72$ 名患者总计，或每个试验组 $n = 36$ 名患者，即可校准H_-的有力证据概率。因此，图9中的最终结果可概括如下：当在 H_+ 和 H_-下均选择略具信息性的设计先验时，从贝叶斯视角来看，只要至少有 $k = 1/30$名患者，设计就能充分校准，以达到 80% 的检验效能和 5% 的第一类错误率（针对

每个试验组有$n = 36$ 名患者。此外，有力证据的概率也校准为 80%（对应 $k = 30$）。从纯频率学派视角来看，当每个试验组纳入 $n = 36$ 名患者时，I类错误率也校准为 5%。重要的是，假设 $p_1 = 0.3$ 和 $p_2 = 0.6$，频率学派功效仅在总计纳入 $n = 77$名患者时校准为 80%，这意味着在均衡随机分配的假设下，每个试验组有 $n = 39$ 名患者。

综上，ICT-107试验得到了极强的贝叶斯因子贝叶斯因子$_{+0}$ =3702。65，表明有极强证据支持 H_+： $p_2 > p_1$；根据表3中的样本量，该试验已得到全面校准。所有相关指标均已校准，尤其是贝叶斯功效、贝叶斯和频率学派I类错误。此外，从贝叶斯视角来看， H_0 下有力证据概率的要求也已满足。这进一步巩固了表3中已令人信服的结果。[12]

3 讨论

本文介绍了一种无模拟数值精确方法，用于在二元终点双臂临床试验中，基于贝叶斯因子计算贝叶斯功效与样本量。我们的方法弥补了贝叶斯试验设计中的一项关键空白：无需蒙特卡罗模拟或渐近近似，即可提供计算高效的解决方案，对长期错误率和检验效能进行校准。针对点零假设与复合假设（H_0： $p_1 = p_2$与 H_1： $p_1 \neq p_2$）以及方向性假设（H_0： $\eta = 0$ 与 H_+： $\eta > 0$, H_0： $\eta = 0$ 与H_-： $\eta < 0$, H_-： $\eta \leq 0$与 H_+： $\eta > 0$）的推导，使II期试验能够得到精确的规划与设计。

3.1 相较于基于模拟方法的主要优势

传统的贝叶斯样本量确定依赖蒙特卡罗模拟，需要进行数千次迭代，才能以可接受的精度估计错误率（Kelter，2023年；Kelter &Pawel，2025b；Morris 等，2019年；Pawel& Held，2024年；Seibold、Charlton、Boulesteix 和 Hoff- mann，2021；Siepe 等，2024年）。我们的基于矩阵的求根方法（见图1）通过先验预测密度和贝塔函数对所有可能的数据组合进行精确计算，从而消除蒙特卡罗标准误，并确保结果可复现。在现代台式电脑上，调用我们的校准函数，对每组患者数在 $n = 1$ 至 $n = 100$ 范围内的情况进行计算，只需几秒钟；而模拟则需要数小时，且精度毫无损失。

该 R package bfbin2arm 通过有限和表示和数值积分，以解析方式实现所有贝叶斯因子，并支持整数和非整数先验形状参数

[12]我们承认，原始试验随机化方案并未采用均衡随机分配，而是将三分之一的患者随机分配至对照组，三分之二随机分配至治疗组。我们在 binbf2arm R 软件包中的实现支持针对不均衡随机化设计和校准贝叶斯因子设计，并在附录中提供了原始 ICT-107 随机化方案的结果及示例代码。

。[13] 由于随机种子和收敛诊断，模拟代码的监管申报颇具挑战；我们采用的确定性方法可提供透明且可验证的结果。

3.2 真实世界校准的启示

我们对利奥西呱和 ICT-107 试验的重新分析揭示了在 II 期试验中使用贝叶斯因子时常见的设计挑战。尽管贝叶斯因子颇具前景（贝叶斯因子$_{+0} = 4.32$，贝叶斯因子$_{+-} = 3702$），但在无信息先验下，这两项试验的样本量都无法保证检验效能。平坦设计先验在这两个案例中都要求不切实际的大样本量；而反映临床预期的略具信息性先验，则能为完全校准的设计带来现实可行的样本量。这一现象可归因于先验几何，参见图 3：平坦先验无法以任何合理方式区分所考虑的假设，因此，达到强或极强证据阈值所需的样本量 k 会变得很大。一旦略具信息性的先验能够以有意义的方式区分这些假设，所需样本量就会迅速降至双臂 II 期试验的现实水平。与此同时，分析仍采用平坦分析先验，并符合美国食品药品监督管理局（FDA）和欧洲药品管理局（EMA）的监管指南（详情见下一小节）。

在我们的重新分析中，方向性检验在优效性主张方面表现更佳：BF$_{+-}$ for H_- : $\eta \leq 0$ vs H_+ : $\eta > 0$在全面贝叶斯校准下（检验效能为 80%，贝叶斯第一类错误率为 $\leq 5\%$，有力证据的概率为 80% $P(\text{BF}_{10} > 30|H_-)$），每组需要 $n = 36$ 名患者（采用略具信息性的设计先验，$k = 1/30$），且频率学派第一类错误率峰值为 4%。因此，ICT-107 的 124 名患者提供了超越频率学派分析的有力证据（OR=0.29,$p < 0.001$）。

3.3 监管一致性

美国食品药品监督管理局和 EMA 的指南认可能够展现“足够稳健”频率学派特性的贝叶斯方法（欧洲药品管理局人用药品委员会／方法学工作组，2025年；美国食品药品管理局器械与放射健康中心，2010年）。我们的校准结果超出了这些要求：贝叶斯第一类错误 $\leq 5\%$ 与 $\alpha = 0.05$一致，而针对原假设区域的频率学派第一类错误计算则提供了严格的频率学派第一类错误控制。校准 H_0，$P(\text{BF}_{01} > k_f|H_0) \geq 80\%$ 有力证据概率的选项，可确保在治疗或药物无效时检测出无效性；而固定阈值后验概率设计往往缺少这一点。美国食品药品监督管理局发布的《医疗器械临床试验中贝叶斯统计使用指南》指出

[13]有关如何使用该软件包的指南，请参阅附录。

> “纯粹”的贝叶斯统计方法不一定像传统的频率学派方法那样重视控制第一类错误。不过，文献中已有一些提议，认为贝叶斯方法应经过“校准”，以具备良好的频率学派性质（例如 *Rubin，1984*年；*Box，1980*年）。本着这一理念，并遵循监管惯例，美国食品药品监督管理局建议您提供所提出的贝叶斯分析计划的第一类和第二类错误率。（美国食品药品监督管理局器械与放射健康中心，2010年，第29页）

同样，当主要推断以贝叶斯方法为基础时，EMA要求申办方证明其具备适当的I类错误控制和充分的检验效能（欧洲药品管理局，2025年，第24页）。

鉴于我们提出的方法为校准贝叶斯设计提供了广泛的选择，所开发的方法与监管建议相一致。此外，杰弗里斯尺度阈值（$k = 1/3,\ 1/10,\ 1/30$）能够在适用于II期试验的现实样本量下，提供可解释的证据等级（中等/强/非常强）。贝叶斯—频率学派折衷既符合倾向于采用客观分析先验的监管机构要求，又允许在规划阶段采用信息性设计先验纳入对治疗疗效的先验信念。

最近，美国食品药品监督管理局发布了一份关于在药物和生物制品临床试验中使用贝叶斯方法学的指南草案，并强调了以下内容：

> 对于（.....）贝叶斯方法，成功判据通常基于真实治疗效应大小超过某个阈值的后验概率。用数学符号表示，此类判据可写作 $Pr(d > a) > c$，其中 d 是治疗效应大小的总体水平汇总，a 是认定治疗有益所需的治疗效应最小阈值，c是支持有效性结论所需的最小概率水平。（.....）因此，选择此类成功判据，意味着要为 a 和 c选定具体数值。贝叶斯分析中，设定这些阈值有多种方法。具体采用哪种方法，取决于试验目标和所使用的具体贝叶斯方法。(美国卫生与公众服务部，美国食品药品监督管理局，药品评价与研究中心（CDER）、生物制品评价与研究中心（CBER），2026年，第6页)

与后验概率不同，我们的设计采用贝叶斯因子作为成功判据。这里值得一提的是两点：

- 首先，如有需要，我们的设计可以扩展为使用后验概率，因为后验概率就是贝叶斯因子乘以这些假设的先验赔率

所考虑的。虽然目前尚未在 bfbin2arm R 软件包中实现，但扩展起来并不困难。

▶ 其次，我们的假设形式有所不同：例如，对于 H_+： $p_2 - p_1 > 0$相对于 H_-： $p_2 - p_1 \leq 0$，上述最小阈值 a 等于零，而 $p_2 - p_1$ 是治疗效应大小的总体水平汇总。 a的不同取值需要不同的假设，例如 H_+： $p_2 - p_1 > 0.1$ 相对于 H_-： $p_2 - p_1 \leq 0.1$。这进而要求对两种假设下所使用的截断先验进行修改，参见图 3。不过，此类修改同样很直接。例如，对于假设 H_+： $p_2 - p_1 > 0.1$ 相对于 H_-： $p_2 - p_1 \leq 0.1$，截断先验并非受二维单位正方形中主对角线的限制，而是受图 3 中该对角线沿 p_2轴向上平移 0.1 后的限制。

因此，这表明我们的设计符合美国食品药品监督管理局目前关于在临床试验中使用贝叶斯方法学的观点。是否有必要按照上文第二点所述对假设进行修改，由用户决定，因为“在将贝叶斯成功标准校准至I类错误率时， a 对于优效性设计，美国卫生与公众服务部、美国食品药品监督管理局、药品评价与研究中心（CDER）、生物制品评价与研究中心（CBER），2026年，第6页）选择

> 对于某些贝叶斯方法， a 和 c 可以进行选择，使总体家族错误率在给定水平上得到控制，通常为单侧*0.025*。这被称为将成功标准校准至*I*类错误率。对于使用贝叶斯方法并非为了综合多个信息来源，而是为了促进复杂自适应设计的设计，这种方法可能较为适用。对于采用表达与分析相关先验信息缺乏的非信息性先验分布的设计，将其校准至*I*类错误率也可能有用。（美国卫生与公众服务部、美国食品药品监督管理局、药品评价与研究中心（CDER）、生物制品评价与研究中心（CBER），2026年，第6页）

这与我们设计中采用的方法完全一致。我们设计的一个显著特点是，完全不使用蒙特卡罗模拟来校准贝叶斯结果，使其符合 I 类错误率。美国食品药品监督管理局的草案提到，这目前仍是实践中的主流方法：

> 对于将贝叶斯结果校准至*I*类错误率的试验设计，其主要操作特性与上述采用频率学派推断的试验所述相同（.....）。临床试验模拟通常用于估计或

> 证明操作特性得到控制。简言之，在给定所选先验分布和样本量的条件下，假设零假设成立或备择假设成立，生成大量模拟试验。随后，根据零假设被拒绝的模拟试验所占比例，分别估计*I*类错误率和检验效能。 (美国卫生与公众服务部 美国食品药品监督管理局 药品评价与研究中心 (CDER)，生物制品评价与研究中心 (CBER)，2026年，第 8 页)

相比之下，我们提出的校准方法仅采用数值方法，无需进行计算量巨大的模拟，也无需报告蒙特卡罗标准误；参见 Kelter (2023年) 和Seibold等人 (2021)。需要强调的是，尽管原则上可以进行模拟，但我们的方法能够以大幅降低的计算成本得出精确结果。

美国食品药品监督管理局关于复杂创新试验设计的指南草案也提出了类似要求，其中建议“详细评估设计的操作特性，包括其得出错误结论的可能性以及治疗效果估计的可靠性。应在适用情况下讨论第一类错误概率的控制和检验效能。”(美国卫生与公众服务部、美国食品药品监督管理局生物制品评价与研究中心和药品评价与研究中心，2020年，第8页），而分析这些操作特性的关键方法是基于模拟的方法。

总而言之，我们的方法完全符合当前的监管建议，甚至避免了采用计算成本高昂的模拟，以及由此给申办方和临床医生带来的试验设计实施、沟通和重复性方面的问题。

3.4 局限性与扩展

离散二项模型会因格点效应而使功效曲线呈现阶梯函数特征，这一点在本文所有图中都清晰可见。我们在`binbf2arm`中实现的前瞻10个样本量的方法确保了稳定性，并且在所有示例中均运行正常，因此用户可以合理确信，接下来的十个样本量对应的检验效能不会低于指定阈值（I类错误率和有力证据的概率也同样如此）。我们方法的一个主要局限是假设样本量固定，因此无法进行期中分析。不过，序贯扩展可以自然地基于递归先验预测分布实现（Kelter & Pawel，2025b），可在未来的研究中加以探讨。对于二元终点的单臂试验，Kelter 和 Pawel（2025a）最近开发了一种高效的数值方法

用于开展两阶段序贯贝叶斯试验设计，并在多种设定下将 Simon 两阶段设计作为一个特例纳入其中。该文详述的方法也可推广至双臂设定。binbf2arm 软件包的实现可以处理非等比例随机化（$n_1 \neq n_2$），但假设分配比例已知。

该框架将贝叶斯证据量化与频率学派操作特性相衔接，通过使用贝叶斯因子，促进资源受限环境下高效II期设计获得监管认可。

致谢

作者感谢Samuel Pawel就该方法学提供的反馈，并感谢Silke J¨orgens就官方监管机构对临床试验中使用贝叶斯方法学所持立场的相关文献提出的有益建议。

附录

A bfbin2arm R 包

该bfbin2arm 包可在CRAN上获取（ https://cran.r-project.org/web/packages/bfbin2arm/index.html），并支持使用贝叶斯因子对双臂II期临床试验设计进行单行校准：

install.packages("bfbin2arm"); #从 CRAN 安装软件包 library(bfbin2arm); # 加载软件包 ntwoarmbinbf01(检验效能=0.8, alpha=0.05,pce_H0=0.8, test="BF+-", k=1/10,nrange=c(20,200))

例如，可以使用以下一组参数，通过 'ntwoarmbinbf01' 函数重新分析 ICT-107试验：

```
ntwoarmbinbf01(
    k = 1/30, k_f = 30, 检验效能 = 0.8, alpha = 0.
    05, pce_H0 = 0.8, test ="BF+-", nrange =
    c(10, 100), n_step = 1, progress = TRUE, a
    _1_d = 1, b_1_d = 2,
```

```
    a_2_d = 2,b_2_d = 1, a_1_d_Hminus = 2, b_1_d_
Hminus = 1, a_2_d_Hminus = 1, b_2_d_Hminus = 2,
compute_freq_t1e = TRUE, p1_检验效能 = 0.3, p2_检验效能
 = 0.6, output = "plot" # 返回每组建议的样本量 )
```

有关该软件包使用方法的详细信息与进一步说明，请参阅 CRAN 上提供的软件包指南。为便于说明，我们展示如何设计并校准 ICT-107试验（第2.5.2节中有详细讨论），采用原始随机化方案，其中三分之一的患者被随机分配至对照组，三分之二被随机分配至治疗组。

```
ntwoarmbinbf01(k = 1/30, k_f = 30, 检验效能 = 0.
8, alpha = 0.05, pce_H0 = 0.8, test ="BF+-",
nrange =c(10, 100), n_step = 1, progress = T
RUE, a_1_d = 1, b_1_d = 2, a_2_d = 2, b_2_d
 = 1, a_1_d_Hminus = 2, b_1_d_Hminus = 1, a
_2_d_Hminus = 1, b_2_d_Hminus = 2, compute_
freq_t1e = TRUE, p1_power = 0.3,p2_power = 0.
6, output = "plot", alloc1 = 1/3, alloc2 = 2/3 )
```

除了此前调用`ntwoarmbinbf01`的函数外，现在还提供了`alloc1`和`alloc2`参数，用于指定随机化方案。`alloc1`是对照组的随机化概率，而`alloc2` 是治疗组的随机化概率。图10 展示了函数调用的输出，并表明总共需要 $n = 83$名患者即可达到80%的贝叶斯功效。这意味着，对照组需要 $1/3 \cdot 83 \approx 28$ 名患者，治疗组需要$2/3 \cdot 83 \approx 56$ 名患者，才能达到80%的贝叶斯功效；原始ICT-107样本量已满足这一要求，参见表3。

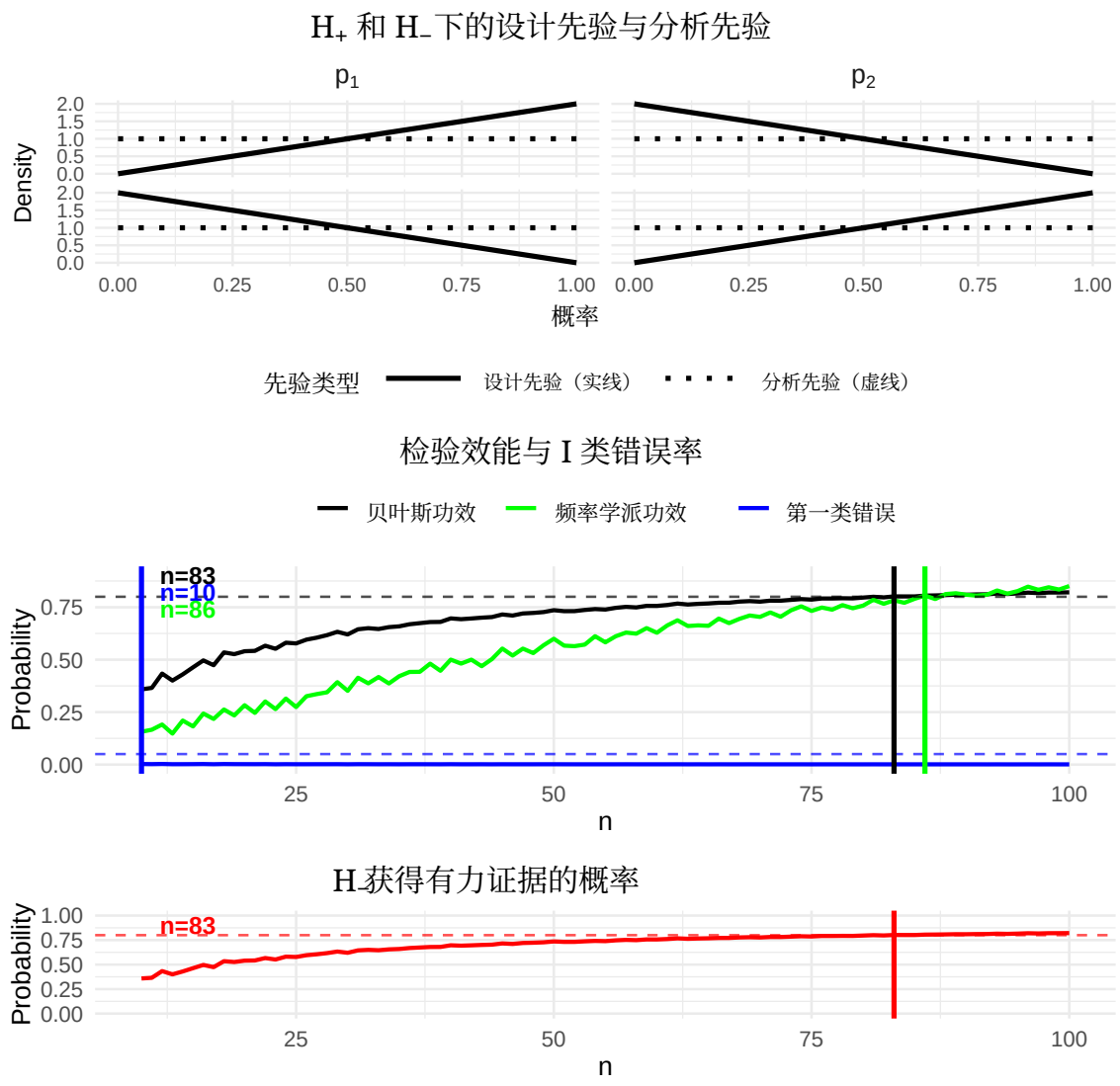

图10：ICT-107试验的贝叶斯功效与样本量计算，其中检验 H_-： $p_1 \geq p_2$ 与 H_+： $p_2 > p_1$。对照组和治疗组分别采用信息性设计先验 $B(1,2)$ 和 $B(2,1)$，并选择平坦分析先验。此外，在 H_-： $p_1 \geq p_2$ 下，设计先验现也具有信息性，并分别为对照组和治疗组选择 $B(2,1)$ 和 $B(1,2)$。结果基于非常强的证据阈值 $k = 1/30$，以及针对 H_- 获得有力证据的阈值 $k_f = 30$。校准结果显示，在 H_0 下，检验效能为 80%、I 类错误率为 5%，且获得有力证据的概率为 80%。频率学派功效是在 $p_1 = 0.3$ 和 $p_2 = 0.6$ 下获得的。随机化概率为对照组 $1/3$，治疗组 $2/3$。

B 双侧假设检验的推导 $H_0:\quad \eta=0$与 $H_1:\quad \eta\neq 0$

在 $H_1:\quad \eta\neq 0$ 下的先验预测。

$$
\begin{aligned}
f(d|H_0) &= f(y_1,y_2 \mid n_1,n_2,a_0^d,b_0^d) \\
&= \int \mathrm{Bin}(y_1|n_1,p)\mathrm{Bin}(y_2|n_2,p)\mathrm{Beta}(p|a_0^d,b_0^d)dp \\
&= \int \binom{n_1}{y_1} p^{y_1}(1-p)^{n_1-y_1}\binom{n_2}{y_2} p^{y_2}(1-p)^{n_2-y_2}\frac{p^{a_0^d-1}(1-p)^{b_0^d-1}}{\mathrm{B}(a_0^d,b_0^d)}dp \\
&= \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}}{\mathrm{B}(a_0^d,b_0^d)}\int p^{y_1+y_2+a_0^d-1}(1-p)^{n_1+n_2-y_1-y_2+b_0^d-1}dp \\
&= \frac{\binom{n_1}{y_1}\binom{n_2}{y_2}}{\mathrm{B}(a_0^d,b_0^d)}\mathrm{B}(a_0^d+y_1+y_2,b_0^d+n_1+n_2-y_1-y_2)
\end{aligned}
\tag{27}
$$

即公式（19）。 □

在 $H_0:\quad \eta=0$ 下的先验预测。

$$
\begin{aligned}
f(d|H_1) &= f(y_1,y_2 \mid n_1,n_2,a_1^d,b_1^d,a_2^d,b_2^d) \\
&= \int\int \mathrm{Bin}(y_1|n_1,p_1)\mathrm{Beta}(p_1|a_1^d,b_1^d)\mathrm{Bin}(y_2|n_2,p_2)\mathrm{Beta}(p_2|a_2^d,b_2^d)dp_1dp_2 \\
&= \int\left[\int \mathrm{Bin}(y_1|n_1,p_1)\mathrm{Beta}(p_1|a_1^d,b_1^d)dp_1\right]\mathrm{Bin}(y_2|n_2,p_2)\mathrm{Beta}(p_2|a_2^d,b_2^d)dp_2 \\
&= \int \mathrm{Bin}(y_1|n_1,p_1)\mathrm{Beta}(p_1|a_1^d,b_1^d)dp_1\cdot\int \mathrm{Bin}(y_2|n_2,p_2)\mathrm{Beta}(p_2|a_2^d,b_2^d)dp_2 \\
&= \int \binom{n_1}{y_1}\frac{p_1^{y_1}(1-p_1)^{n_1-y_1}p_1^{a_1^d-1}(1-p_1)^{b_1^d-1}}{\mathrm{B}(a_1^d,b_1^d)}dp_1 \\
&\cdot\int \binom{n_2}{y_2}\frac{p_2^{y_2}(1-p_2)^{n_2-y_2}p_2^{a_2^d-2}(1-p_2)^{b_2^d-1}}{\mathrm{B}(a_2^d,b_2^d)}dp_2 \\
&= \binom{n_1}{y_1}\frac{\mathrm{B}(\mathrm{y}_1+\mathrm{a}_1^{\mathrm{d}},\mathrm{n}_1-\mathrm{y}_1+\mathrm{b}_1^{\mathrm{d}})}{\mathrm{B}(a_1^d,b_1^d)}\cdot\binom{n_2}{y_2}\frac{\mathrm{B}(\mathrm{y}_2+\mathrm{a}_2^{\mathrm{d}},\mathrm{n}_2-\mathrm{y}_2+\mathrm{b}_2^{\mathrm{d}})}{\mathrm{B}(a_2^d,b_2^d)}
\end{aligned}
\tag{28}
$$

即公式（17）。 □

C 的有限和形式（整数形状参数）。使用展开式（对于整数 a_1）

$$
I_x(a_1,b_1)=\frac{1}{B(a_1,b_1)}\sum_{k=0}^{a_1-1}\binom{a_1+b_1-1}{k}x^k(1-x)^{a_1+b_1-1-k},
$$

我们得到

$$
\begin{aligned}
C &= \int_0^1 \frac{p_2^{a_2-1}(1-p_2)^{b_2-1}}{B(a_2,b_2)}\, I_{p_2}(a_1,b_1)\, dp_2 \\
&= \frac{1}{B(a_1,b_1)B(a_2,b_2)} \sum_{k=0}^{a_1-1} \binom{a_1+b_1-1}{k} \int_0^1 p_2^{a_2-1+k}(1-p_2)^{b_2-1+a_1+b_1-1-k}\, dp_2 \\
&= \frac{1}{B(a_1,b_1)B(a_2,b_2)} \sum_{k=0}^{a_1-1} \binom{a_1+b_1-1}{k} B\big(a_2+k,\ b_2+a_1+b_1-1-k\big). \qquad (29)
\end{aligned}
$$

□

C De双侧假设检验的推导 $H_0:\ \eta = 0$对 $H_+:\ \eta > 0$

$H_+:\ \eta > 0$下的预测密度。

$$
p(y_1,y_2 \mid H_+) = \iint_A f(p_1,p_2 \mid y_1,y_2)\, \pi_+(p_1,p_2)\, dp_1\, dp_2.
$$

代入截断先验,

$$
\pi_+(p_1,p_2) = \frac{1}{C}\, \frac{p_1^{a_1^d-1}(1-p_1)^{b_1^d-1}}{B(a_1^d,b_1^d)}\, \frac{p_2^{a_2^d-1}(1-p_2)^{b_2^d-1}}{B(a_2^d,b_2^d)}\, \mathbf{1}\{0<p_1<p_2<1\},
$$

得到

$$
\begin{aligned}
p(y_1,y_2 \mid H_+) = \binom{n_1}{y_1}\binom{n_2}{y_2} \frac{1}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)\, C} \int_0^1 \int_{p_1}^1 & p_1^{y_1+a_1^d-1}(1-p_1)^{n_1-y_1+b_1^d-1} \\
& \times p_2^{y_2+a_2^d-1}(1-p_2)^{n_2-y_2+b_2^d-1}\, dp_2\, dp_1.
\end{aligned}
$$

定义更新后的形状参数

$$
A_1 = y_1 + a_1^d, \quad B_1 = n_1 - y_1 + b_1^d, \qquad A_2 = y_2 + a_2^d, \quad B_2 = n_2 - y_2 + b_2^d. \qquad (30)
$$

那么，对于固定的 p_1，内层积分为

$$
\int_{p_1}^1 p_2^{A_2-1}(1-p_2)^{B_2-1}\, dp_2 = B(A_2,B_2) - B_{p_1}(A_2,B_2) = B(A_2,B_2)\big[1 - I_{p_1}(A_2,B_2)\big].
$$

因此

$$p(y_1, y_2 \mid H_+) = \binom{n_1}{y_1}\binom{n_2}{y_2}\frac{1}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)\,C}\int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1} \times B(A_2, B_2)\big[1 - I_{p_1}(A_2, B_2)\big]\,dp_1.$$

定义

$$I(y_1, y_2) = \int_0^1 \int_{p_1}^1 p_1^{A_1-1}(1-p_1)^{B_1-1} p_2^{A_2-1}(1-p_2)^{B_2-1}\,dp_2\,dp_1,$$

so

$$p(y_1, y_2 \mid H_+) = \binom{n_1}{y_1}\binom{n_2}{y_2}\frac{I(y_1, y_2)}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)\,C}. \tag{31}$$

利用内层积分的表达式,

$$\begin{aligned} I(y_1, y_2) &= \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}\big[B(A_2, B_2) - B_{p_1}(A_2, B_2)\big]\,dp_1 \\ &= B(A_2, B_2)\int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}\,dp_1 - \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}B_{p_1}(A_2, B_2)\,dp_1 \\ &= B(A_2, B_2)B(A_1, B_1) - B(A_2, B_2)\,J, \end{aligned}$$

其中

$$J = \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}I_{p_1}(A_2, B_2)\,dp_1.$$

Thus

$$I(y_1, y_2) = B(A_2, B_2)\big[B(A_1, B_1) - J\big].$$

可将其代入方程（24）。 □

的有限和形式 $I(y_1,\ y_2)$（整数形状参数）。当 A_2 为正整数时，我们可以再次展开

$$I_{p_1}(A_2, B_2) = \frac{1}{B(A_2, B_2)}\sum_{k=0}^{A_2-1}\binom{A_2 + B_2 - 1}{k}p_1^k(1-p_1)^{A_2+B_2-1-k},$$

获得

$$\begin{aligned} J &= \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}I_{p_1}(A_2, B_2)\,dp_1 \\ &= \frac{1}{B(A_2, B_2)}\sum_{k=0}^{A_2-1}\binom{A_2 + B_2 - 1}{k}\int_0^1 p_1^{A_1-1+k}(1-p_1)^{B_1-1+A_2+B_2-1-k}\,dp_1 \\ &= \frac{1}{B(A_2, B_2)}\sum_{k=0}^{A_2-1}\binom{A_2 + B_2 - 1}{k}B\big(A_1 + k,\ B_1 + A_2 + B_2 - 1 - k\big). \end{aligned}$$

因此

$$I(y_1,y_2)=B(A_2,B_2)B(A_1,B_1)-\sum_{k=0}^{A_2-1}\binom{A_2+B_2-1}{k}B\big(A_1+k,\ B_1+A_2+B_2-1-k\big).$$

将其代入公式（24），即代入

$$p(y_1,y_2\mid H_+)=\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{I(y_1,y_2)}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)\,C},$$

并使用式（ C 29）中的有限和形式，可得 H_+ 完全由贝塔函数表示的预测密度显式表达式。 □

D 双侧假设检验的推导： $H_0:\ \ \eta=0$与 $H_-:\ \ \eta<0$

D.1 单侧假设检验： $H_0:\ \ \eta=0$ 与 $H_-:\ \ \eta<0$

二项模型与假设

我们观察到

$$Y_1\sim\mathrm{Bin}(n_1,p_1),\qquad Y_2\sim\mathrm{Bin}(n_2,p_2),$$

并希望检验

$$H_0:\eta=0\quad\text{vs.}\quad H_-:\eta<0,\qquad \eta=p_2-p_1.$$

在 H_0 下，我们有 $p_1=p_2=:\ p$。在 H_- 下，我们施加序约束 $p_2<p_1$。

D.2 H_0 和 H_- 下的设计与分析先验

H_0 下的先验与上一节完全相同：

$$p\sim\mathrm{Beta}(a_0^d,b_0^d)\quad\text{(design)},\qquad p\sim\mathrm{Beta}(a_0^a,b_0^a)\quad\text{(analysis)},$$

先验密度为

$$\pi_0(p)=\frac{1}{B(a_0^d,b_0^d)}\,p^{a_0^d-1}(1-p)^{b_0^d-1},\qquad 0<p<1.$$

在 H_- 下，我们从独立Beta设计先验开始

$$p_1\sim\mathrm{Beta}(a_1^d,b_1^d),\qquad p_2\sim\mathrm{Beta}(a_2^d,b_2^d),$$

具有联合（未截断）密度

$$\pi_{\text{untr}}(p_1,p_2)=\frac{1}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)}\,p_1^{a_1^d-1}(1-p_1)^{b_1^d-1}\,p_2^{a_2^d-1}(1-p_2)^{b_2^d-1},\qquad 0<p_1,p_2<1.$$

然后，我们将其截断至序约束区域

$$A_-=\{(p_1,p_2)\colon 0<p_2<p_1<1\}.$$

截断先验的归一化常数为

$$C_-=\iint_{A_-}\pi_{\text{untr}}(p_1,p_2)\,dp_1\,dp_2=P(p_2<p_1)$$

在独立 Beta 先验下。一个简便的表达式为

$$C_-=\int_0^1\frac{p_1^{a_1^d-1}(1-p_1)^{b_1^d-1}}{B(a_1^d,b_1^d)}\,I_{p_1}(a_2^d,b_2^d)\,dp_1. \tag{32}$$

因此， H_- 下的截断先验密度为

$$\pi_-(p_1,p_2)=\frac{\pi_{\text{untr}}(p_1,p_2)\,\mathbf{1}\{0<p_2<p_1<1\}}{C_-}.$$

D.3 C_- 的有限和形式（整数形状）

当 a_2^d 为正整数时，

$$I_{p_1}(a_2^d,b_2^d)=\frac{1}{B(a_2^d,b_2^d)}\sum_{k=0}^{a_2^d-1}\binom{a_2^d+b_2^d-1}{k}p_1^k(1-p_1)^{a_2^d+b_2^d-1-k},$$

从而得到

$$C_-=\frac{1}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)}\sum_{k=0}^{a_2^d-1}\binom{a_2^d+b_2^d-1}{k}B\big(a_1^d+k,\ b_1^d+a_2^d+b_2^d-1-k\big). \tag{33}$$

D.4 H_0下的预测密度

H_0 下的预测密度保持不变:

$$p(y_1,y_2\mid H_0)=\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{B\big(y_1+y_2+a_0^d,\ n_1+n_2-y_1-y_2+b_0^d\big)}{B(a_0^d,b_0^d)}. \tag{34}$$

D.5 H_-下的预测密度

似然保持不变

$$f(y_1, y_2 \mid p_1, p_2) = \binom{n_1}{y_1}\binom{n_2}{y_2} p_1^{y_1}(1-p_1)^{n_1-y_1} p_2^{y_2}(1-p_2)^{n_2-y_2}.$$

H_- 下的预测密度为

$$p(y_1, y_2 \mid H_-) = \iint_{A_-} f(y_1, y_2 \mid p_1, p_2)\pi_-(p_1, p_2)\, dp_1\, dp_2.$$

代入截断先验可得

$$p(y_1, y_2 \mid H_-) = \binom{n_1}{y_1}\binom{n_2}{y_2} \frac{1}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)\, C_-} \int_0^1 \int_0^{p_1} p_2^{y_2+a_2^d-1}(1-p_2)^{n_2-y_2+b_2^d-1} \times p_1^{y_1+a_1^d-1}(1-p_1)^{n_1-y_1+b_1^d-1}\, dp_2\, dp_1.$$

使用公式（25）中的更新后形状参数，内层积分为

$$\int_0^{p_1} p_2^{A_2-1}(1-p_2)^{B_2-1}\, dp_2 = B_{p_1}(A_2, B_2) = B(A_2, B_2)\, I_{p_1}(A_2, B_2).$$

因此

$$p(y_1, y_2 \mid H_-) = \binom{n_1}{y_1}\binom{n_2}{y_2} \frac{B(A_2, B_2)}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)\, C_-} \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1} I_{p_1}(A_2, B_2)\, dp_1.$$

定义

$$J(y_1, y_2) = \int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1} I_{p_1}(A_2, B_2)\, dp_1,$$

so

$$p(y_1, y_2 \mid H_-) = \binom{n_1}{y_1}\binom{n_2}{y_2} \frac{B(A_2, B_2)J(y_1, y_2)}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)\, C_-}. \tag{35}$$

D.6 $J(y_1, y_2)$ 的有限和形式（整数形状参数）

当 A_2 为正整数时，

$$J(y_1, y_2) = \frac{1}{B(A_2, B_2)} \sum_{k=0}^{A_2-1} \binom{A_2+B_2-1}{k} B\big(A_1+k,\ B_1+A_2+B_2-1-k\big).$$

D.7 H_0： $\eta = 0$ 相对于 H_-： $\eta < 0$的贝叶斯因子

针对 H_- 与 H_0 的贝叶斯因子为

$$\mathrm{BF}_{-0}(y_1, y_2) = \frac{p(y_1, y_2 \mid H_-)}{p(y_1, y_2 \mid H_0)},$$

使用方程（34）和（35）中的表达式:

$$\mathrm{BF}_{-0}(y_1, y_2) = \frac{B(a_0^a, b_0^a)B(A_2, B_2)J(y_1, y_2)}{B(y_1 + y_2 + a_0^a,\ n_1 + n_2 - y_1 - y_2 + b_0^a)B(a_1^a, b_1^a)B(a_2^a, b_2^a)C_-}, \tag{36}$$

其中， C_- 对于整数形状参数，使用方程（32）或方程（33），而 J（y_1， y_2）使用上面的有限和形式。当形状参数为整数时，贝叶斯因子BF_{-0} 可完全使用贝塔函数计算，无需进行数值积分。请注意，在方程（36）中，我们已将方程（34）和（35）中的设计先验超参数 a_i^d和 b_i^d替换为分析先验超参数 a_i^a， b_i^a， $i = 1$， 2。

E H_-： $\eta \leq 0$相对于 H_+： $\eta > 0$的双侧假设检验推导

E.1 二项模型与假设

我们观测到

$$Y_1 \sim \mathrm{Bin}(n_1, p_1), \qquad Y_2 \sim \mathrm{Bin}(n_2, p_2),$$

并希望检验

$$H_0 : \eta \leq 0 \quad \text{vs.} \quad H_1 : \eta > 0, \qquad \eta = p_2 - p_1.$$

在 H_0 下，我们施加序约束 $p_2 \leq p_1$。在 H_1 下，我们施加 $p_2 > p_1$。

E.2 H_0 和 H_1下的设计与分析先验

在 H_0 下： $p_2 \leq p_1$ 我们从独立Beta设计先验开始

$$p_1 \sim \mathrm{Beta}(a_1^d, b_1^d), \qquad p_2 \sim \mathrm{Beta}(a_2^d, b_2^d),$$

具有联合（未截断）密度

$$\pi_{\text{untr}}(p_1, p_2) = \frac{1}{B(a_1^d, b_1^d)B(a_2^d, b_2^d)}\, p_1^{a_1^d - 1}(1 - p_1)^{b_1^d - 1}\, p_2^{a_2^d - 1}(1 - p_2)^{b_2^d - 1}, \qquad 0 < p_1, p_2 < 1.$$

我们将其截断至序约束区域

$$A_0 = \{(p_1, p_2) \colon 0 \le p_2 \le p_1 \le 1\}.$$

H_0 下截断先验的归一化常数为

$$C_0 = \iint_{A_0} \pi_{\text{untr}}(p_1, p_2)\, dp_1\, dp_2 = P(p_2 \le p_1) = 1 - P(p_2 > p_1).$$

使用 $P\,(p_2 > p_1)$ 在方程（22）中的表达式,

$$C_0 = 1 - C, \qquad C = \int_0^1 \frac{p_2^{a_2^d - 1}(1 - p_2)^{b_2^d - 1}}{B(a_2^d, b_2^d)}\, I_{p_2}(a_1^d, b_1^d)\, dp_2. \tag{37}$$

H_0 下的截断先验密度为

$$\pi_0(p_1, p_2) = \frac{\pi_{\text{untr}}(p_1, p_2)\, \mathbf{1}\{0 \le p_2 \le p_1 \le 1\}}{C_0}.$$

在 H_1 下： $p_2 > p_1$ 先验与上一节中的 H_+ Case 完全相同：

$$\pi_+(p_1, p_2) = \frac{\pi_{\text{untr}}(p_1, p_2)\, \mathbf{1}\{0 < p_1 < p_2 < 1\}}{C},$$

归一化常数为 C，见方程（22）。

E.3 C_0 的有限和形式（整数形状）

利用 C 的有限和形式，由公式（29），我们得到

$$C_0 = 1 - \frac{1}{B(a_1^d, b_1^d) B(a_2^d, b_2^d)} \sum_{k=0}^{a_1^d - 1} \binom{a_1^d + b_1^d - 1}{k} B\big(a_2^d + k,\ b_2^d + a_1^d + b_1^d - 1 - k\big). \tag{38}$$

E.4 在 H_0： $\eta \le 0$下的预测密度

在 H_0： $\eta \le 0$下的预测密度可推导如下。似然为

$$f(y_1, y_2 \mid p_1, p_2) = \binom{n_1}{y_1} \binom{n_2}{y_2} p_1^{y_1} (1 - p_1)^{n_1 - y_1} p_2^{y_2} (1 - p_2)^{n_2 - y_2}.$$

H_0 下的预测密度为

$$p(y_1, y_2 \mid H_0) = \iint_{A_0} f(y_1, y_2 \mid p_1, p_2) \pi_0(p_1, p_2)\, dp_1\, dp_2.$$

代入截断先验后得到

$$p(y_1,y_2\mid H_0)=\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{1}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)\,C_0}\int_0^1\int_0^{p_1} p_2^{\,y_2+a_2^d-1}(1-p_2)^{n_2-y_2+b_2^d-1} \times p_1^{\,y_1+a_1^d-1}(1-p_1)^{n_1-y_1+b_1^d-1}\,dp_2\,dp_1.$$

使用方程（25）中的更新后形状参数，内层积分为

$$\int_0^{p_1} p_2^{A_2-1}(1-p_2)^{B_2-1}\,dp_2=B_{p_1}(A_2,B_2)=B(A_2,B_2)\,I_{p_1}(A_2,B_2).$$

因此

$$p(y_1,y_2\mid H_0)=\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{B(A_2,B_2)}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)\,C_0}\int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}I_{p_1}(A_2,B_2)\,dp_1.$$

定义

$$J(y_1,y_2)=\int_0^1 p_1^{A_1-1}(1-p_1)^{B_1-1}I_{p_1}(A_2,B_2)\,dp_1,$$

so

$$p(y_1,y_2\mid H_0)=\binom{n_1}{y_1}\binom{n_2}{y_2}\frac{B(A_2,B_2)J(y_1,y_2)}{B(a_1^d,b_1^d)B(a_2^d,b_2^d)\,C_0}. \tag{39}$$

E.5 J（y_1，y_2）的有限和形式（整数形状参数）

当 A_2 为正整数时，可得

$$J(y_1,y_2)=\frac{1}{B(A_2,B_2)}\sum_{k=0}^{A_2-1}\binom{A_2+B_2-1}{k}B\big(A_1+k,\;B_1+A_2+B_2-1-k\big).$$

E.6 H_1：$\eta>0$下的预测密度

在 H_1 下的预测密度与上一节中的 H_+ 相同。对于假设H_+：$p_2>p_1$，联合预测概率 $f(y_1,y_2|H_+)$ 的推导与 H_-情形类似，但积分上限需反映不等式 $p_2>p_1$。我们有

$$p_1\sim\text{Beta}(a_1^d,b_1^d),\qquad y_1\sim\text{Binomial}(n_1,p_1)$$
$$p_2\sim\text{Beta}(a_2^d,b_2^d),\qquad y_2\sim\text{Binomial}(n_2,p_2)$$

在 H_+ 下，联合先验预测概率可写为

$$f(y_1,y_2|H_+)=\iint_{0\le p_1\le1,\,p_1\le p_2\le1} f(y_1|p_1)f(p_1|a_1^d,b_1^d)\times f(y_2|p_2)f(p_2|a_2^d,b_2^d)\,dp_2\,dp_1,$$

其中

$$f(y_i|p_i) = \binom{n_i}{y_i} p_i^{y_i}(1-p_i)^{n_i-y_i},$$

and

$$f(p_i|a_i^d, b_i^d) = \frac{p_i^{a_i^d-1}(1-p_i)^{b_i^d-1}}{\mathrm{B}(a_i^d, b_i^d)}.$$

现在，我们可以将关于 p_2 的内层积分简化如下：

$$\begin{aligned}\int_{p_1}^1 f(y_2|p_2) f(p_2|a_2^d, b_2^d)\, dp_2 &= \binom{n_2}{y_2} \frac{1}{\mathrm{B}(a_2^d, b_2^d)} \int_{p_1}^1 p_2^{y_2+a_2^d-1}(1-p_2)^{n_2-y_2+b_2^d-1}\, dp_2 \\ &= \binom{n_2}{y_2} \frac{\mathrm{B}_1(y_2+a_2^d, n_2-y_2+b_2^d) - \mathrm{B}_{p_1}(y_2+a_2^d, n_2-y_2+b_2^d)}{B(a_2^d, b_2^d)},\end{aligned}$$

其中，$\mathrm{B}_x(a,\ b)$ 是不完全贝塔函数，且 $\mathrm{B}_1(a,\ b) = \mathrm{B}(a,\ b)$。积分可简化为：

$$\int_{p_1}^1 \text{Beta PDF}(p_2; a', b')\, dp_2 = 1 - I_{p_1}(a', b'),$$

其中，$I_x(a,\ b)$ 是正则化不完全贝塔函数。因此，$f(y_1,\ y_2|H_+)$ 的最终表达式为：

$$\begin{aligned}f(y_1, y_2|H_+) = &\binom{n_1}{y_1}\binom{n_2}{y_2} \frac{1}{\mathrm{B}(a_1^d, b_1^d)\mathrm{B}(a_2^d, b_2^d)} \\ &\times \int_0^1 p_1^{y_1+a_1^d-1}(1-p_1)^{n_1-y_1+b_1^d-1}\left[1 - I_{p_1}(y_2+a_2^d, n_2-y_2+b_2^d)\right] dp_1. \quad (40)\end{aligned}$$

E.7 贝叶斯因子：H_-：$\eta \le 0$ 对比 H_+：$\eta > 0$

H_1 (H_+) 对比 H_0 (H_-) 的贝叶斯因子为

$$\mathrm{BF}_{10}(y_1, y_2) = \frac{p(y_1, y_2 \mid H_1)}{p(y_1, y_2 \mid H_0)},$$

表达式取自方程（39）和（40）：

$$\mathrm{BF}_{10}(y_1, y_2) = \frac{\frac{I(y_1,y_2)}{B(a_1^a,b_1^a)B(a_2^a,b_2^a)\,C}}{\frac{B(A_2,B_2)J(y_1,y_2)}{B(a_1^a,b_1^a)B(a_2^a,b_2^a)\,C_0}} = \frac{I(y_1, y_2)C_0}{B(A_2, B_2)J(y_1, y_2)C}, \quad (41)$$

其中，$C_0 = 1 - C$ 使用公式（37）和（38），$I(y_1, y_2)$使用 H_+ 有限和形式；当 C_0 或 A_2 为（正）整数时，$J(y_1, y_2)$使用上述有限和形式。当形状参数为整数时，因此可以完全使用贝塔函数计算 BF_{10}。请注意，在公式（41）中，我们将 $i = 1$的设计先验超参数 a_i^d 和 b_i^d 替换为分析先验超参数 a_i^a和 b_i^a, $i = 1, 2$。

参考文献

Bartoˇs, F., Aust, F., & Haaf, J. M.（2022年9月）。知情贝叶斯生存分析。*BMC*医学研究方法 *2022*年 *22:1*，*22*(1), 1–22。数字对象标识符：10.1186/S12874-022 - 01676-9 Berry, D. A.（2006年2月）。贝叶斯临床试验。《自然·药物发现评论》*2006*年 *5:1*，*5*(1), 27–36。数字对象标识符：10.1038/nrd1927 Case, M., Herrera, M., Rumps, M. V., & Mulcahey, M. K.（2022年3月）。导师指导对医学界女性学术职业成功的影响：系统综述。《学术医学》,*97*(3), 444–458。数字对象标识符：10.1097/ACM.0000000000004563 Chow, S.-C., & Liu, J.-P.（2008年）。生物利用度和生物等效性研究的设计与分析。博卡拉顿：Chapman & Hall/CRC出版社。Dawid, A. P.（1982年）。校准良好的贝叶斯主义者。《美国统计学会杂志》，*77*(379), 605–610。数字对象标识符：10.1080/01621459.1982.10477856 Dickey, J. M., & Lientz, B. P.（1970年）。加权似然比、关于概率的精确假设及马尔可夫链的阶数。《数学统计年刊》，*41*(1), 214–226。数字对象标识符：10.1214/AOMS/1177697203 欧洲药品管理局。（2025年6月）。*ICH E20* 临床试验自适应设计——科学指南 | 欧洲药品管理局（*EMA*）。欧洲药品管理局。欧洲药品管理局人用药品委员会／方法学工作组。（2025年）。关于在非干预性研究中使用真实世界数据以生成用于监管用途的真实世界证据的反思文件。欧洲药品管理局。Grieve, A. P.（2016年3月）。临床药物开发中一位“校准良好的”贝叶斯主义者的闲思。《药学统计学》,*15*(2), 96–108。数字对象标识符：10.1002/PST.1736 Grieve, A. P.（2022年）。临床试验规划与实施中的频率学派与贝叶斯混合检验效能及贝叶斯功效。美国佛罗里达州博卡拉顿：Chapman & Hall, CRC出版社。Gunel, E., & Dickey, J.（1974年）。列联表中独立性的贝叶斯因子。《生物计量学》,*61*(3), 545–557。数字对象标识符：10.2307/2334738 Held, L., & Saban´es Bov´e, D.（2014年）。应用统计推断。柏林，海德堡：施普林格。数字对象标识符：10.1007/978-3-642-37887-4 Ibrahim, J. G., Chen, M.-H., & Sinha, Debajyoti.（2001年）。贝叶斯生存分析。纽约：施普林格纽约。Ionan, A. C., Clark, J., Travis, J., Amatya, A., Scott, J., Smith, J. P., . . . Rothmann, M.（2023年）。CDER和CBER在人用药品和生物制品开发中的贝叶斯方法。*Ther InnovRegul Sci*, *57*(3), 436–444。数字对象标识符：10.1007/s43441-022-00483-0 Jamil, T., Ly, A., Morey, R. D., Love, J., Marsman, M., & Wagenmakers, E. J.（2017年4月）。列联表的默认“Gunel和Dickey”贝叶斯因子。《行为研究方法》,*49*(2), 638–652。数字对象标识符：10.3758/S13428-016-0739-8/FIGURES/5

Jeffreys, H. (1939). 概率论. 牛津：克拉伦登出版社. Jeffreys, H. (1961). 概率论. 牛津：牛津大学出版社. Kass, R. E., & Raftery, A. E. (1995). 贝叶斯因子. 《美国统计学会杂志》, *90*(430), 773–795. doi: 10.1002/ BIMJ.202200095 Kelter, R. (2023). 用于统计与方法学研究中模拟研究的贝叶斯模拟研究（BASIS）框架. 生物计量学杂志, 2200095. doi: 10.1002/ BIMJ.202200095 Kelter, R. (2025). 2×2列联表中优势比方向性假设的校准贝叶斯假设检验. 统计生物科学, *17*(2), 410–441. doi: 10.1007/ s12561-024-09425-w Kelter, R., & Pawel, S. (2025a). 基于贝叶斯因子的临床*II*期试验贝叶斯最优两阶段设计 (No. arXiv:2511.23144). arXiv. doi: 10.48550/arXiv.2511 . 23144 Kelter, R., & Pawel, S. (2025b). 二项分布情境下贝叶斯因子的贝叶斯功效与样本量计算. *arXiv* 预印本. Khanna, D., Allanore, Y., Denton, C. P., Kuwana, M., Matucci-Cerinic, M., Pope, J. E., . . . Distler, O. (2020, May). 早期弥漫性皮肤型系统性硬化症患者使用利奥西呱（RISE-SSc）：随机、双盲、安慰剂对照多中心试验。风湿病年鉴, *79*(5), 618–625. doi: 10.1136/annrheumdis-2019 -216823 Kleijn, B. (2022). 贝叶斯统计的频率学派理论. 阿姆斯特丹：施普林格. Little, R. J. (2006, August). 校准贝叶斯. 美国统计学家, *60*(3), 213–223. doi: 10.1198/000313006X117837 Morris, T. P., White, I. R., & Crowther, M. J. (2019). 使用模拟研究评估统计方法. 医学统计学,*38*(11), 2074–2102. doi: 10.1002/SIM.8086 Neuenschwander, B., Branson, M., & Spiegelhalter, D. J. (2009, December). 关于幂先验的一则说明. 医学统计学,*28*(28), 3562–3566. doi: 10.1002/SIM.3722 O’ Hagan, A., & Stevens, J. W. (2001). 成本效益临床试验的样本量贝叶斯评估. 医学决策, *21*(3), 219–230. doi: 10.1177/ 0272989X0102100307 Pawel, S., & Held, L. (2024). 贝叶斯因子的闭式功效与样本量计算. *arXiv* 预印本. Pawel, S., & Held, L. (2025, April). 贝叶斯因子的闭式功效与样本量计算. 美国统计学家, 1–15. doi: 10.1080/00031305.2025.2467919 Pourmohamad, T., & Wang, C. (2023). 临床前实验中二元结局样本量缩减的序贯贝叶斯因子. 生物制药研究统计学,*15*(4), 706–715. doi: 10.1080/19466315.2022.2123386

R 核心团队. (2020). *R*：一种用于统计计算的语言与环境。 R 统计计算基金会. 奥地利维也纳. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). 用于接受和拒绝零假设的贝叶斯 t 检验. 《心理学通报与评论》, *16*(2), 225–237. doi: 10.3758/PBR.16.2.225 Rubin, D. B. (1984, December). 《应用统计学家可辩护且相关的贝叶斯频率计算》. *https://doi.org/10.1214/aos/1176346785*, *12*(4), 1151–1172. doi: 10.1214/AOS/1176346785 Seibold, H., Charlton, A., Boulesteix, A. L., & Hoffmann, S. (2021). 统计学家们，卷起袖子来！有一场危机亟待解决. 《*Significance*》, *18*(4), 42–44. doi: 10.1111/ 1740-9713.01554 Siepe, B. S., Bartoˇs, F., Morris, T. P., Boulesteix, A.-L., Heck, D. W., & Pawel, S. (2024, November). 心理学方法学研究中的模拟研究：用于规划、预注册和报告的标准化模板. 《心理学方法》. doi: 10.1037/MET0000695 Simon, R. (1989). II 期临床试验的最优两阶段设计 . 《对照临床试验》,*10*(1), 1–10. doi: 10.1016/0197-2456(89)90015-9 Spiegelhalter, D. J., Abrams, K. R. K. R., & Myles, J. P. (2004). 临床试验与医疗保健评估的贝叶斯方法. 纽约：Wiley. Thall, P. F., & Simon, R. (1994, June). 《IIb 期临床试验实用贝叶斯指南》. 《生物统计学》 ,*50*(2), 337. doi: 10.2307/2533377 Thall, P. F., & Wathen, J. K. (2007, March). 临床试验中的实用贝叶斯自适应随机化. 《欧洲癌症杂志》（英格兰牛津：*1990*）, *43*(5), 859. doi: 10.1016/J.EJCA.2007.01.006 美国卫生与公众服务部、美国食品药品监督管理局、生物制品评价与研究中心和药品评价与研究中心. (2020, December). 与美国食品药品监督管理局就药品和生物制品复杂创新试验设计进行沟通——行业指南。美国卫生与公众服务部、美国食品药品监督管理局、药品评价与研究中心（CDER）、生物制品评价与研究中心（CBER）. (2026, January). 药物和生物制品临床试验中的贝叶斯方法——行业指南 (技术报告). 美国食品药品监督管理局器械与放射健康中心. (2010). 医疗器械临床试验中贝叶斯统计使用指南。
h
t
t
p
s://
www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-use-bayesian-statistics-medical-device-clinical-trials（访问日期：*16/08/2021*）. Wen, P. Y., Reardon, D. A., Armstrong, T. S., Phuphanich, S., Aiken, R. D., Landolfi, J. C.,

. . . Yu, J. S.（2019年10月）。树突状细胞疫苗 ICT-107 用于新诊断胶质母细胞瘤患者的随机、双盲、安慰剂对照II期临床试验。临床癌症研究，*25*(19)，5799–5807。数字对象标识符：10.1158/1078-0432.CCR-19-0261