arXiv:2304.06164v1 [stat.AP] 12 Apr 2023

用于早期肿瘤学试验概念验证和剂量优化的多试验组两阶段（MATS）设计

Zhenghao Jiang[1], Gu Mi[2], Ji Lin[2], Christelle Lorenzato[3], 和 Yuan Ji*[4]

[1]芝加哥大学统计系，南埃利斯大道5747号，芝加哥，伊利诺伊州60637

[2]赛诺菲生物统计与编程，Water Street 450号，剑桥，马萨诸塞州02141[3]赛诺菲研发生物统计与编程，法国塞纳河畔维特里[4]芝加哥大学公共卫生科学系，南马里兰大道5841号，芝加哥，伊利诺伊州60637*通信请联系 yji@bsd.uchicago.edu

摘要

美国食品药品监督管理局肿瘤学卓越中心发起的Project Optimus计划被广泛视为一项开创性举措，旨在改变肿瘤学中传统剂量探索策略的 现状 。不同于其他治疗领域在剂量范围研究中对多种剂量进行充分评估，早期肿瘤学剂量探索研究的特点是通常确定单一剂量，例如最大耐受剂量（MTD）或

推荐的II期剂量（RP2D）。秉承 Project Optimus 的理念，我们提出了一种用于早期肿瘤学试验中概念验证（PoC）和剂量优化的多臂两阶段设计（MATS），可评估剂量递增试验中选定的两个剂量。该设计在第一阶段首先在多个适应症中评估较高剂量；如果较高剂量表现出有希望的抗肿瘤活性，则针对相应适应症自适应地进入第二阶段。在第二阶段，通过随机比较较高剂量和较低剂量，实现概念验证（PoC）和剂量优化。贝叶斯分层模型通过在不同剂量、适应症和阶段之间借用信息，指导统计推断和决策。模拟研究表明，所提出的 MATS 设计具有良好的性能。我们已开发 R Shiny 应用程序，并发布于 https://matsdesign.shinyapps.io/mats/。

关键词

贝叶斯分层模型；肿瘤学；Project Optimus；概念验证。

1 引言

过去十年间，肿瘤治疗领域取得了令人振奋的治疗进展，其中包括分子靶向药物和免疫疗法的开发；与传统细胞毒性化疗相比，这些疗法基于更有效、更精准的作用机制（MoAs）。尽管涌现出大量可适应不同作用机制的剂量递增设计，以更好地刻画毒性或毒性/疗效特征，但其中大多数仍在寻求单一的最大耐受剂量（MTD）。通常，MTD会成为后续开发中事实上的推荐的II期剂量（RP2D）。相比之下，在非肿瘤学治疗领域，通常会通过大规模的

剂量探索研究对多种剂量进行全面评估。肿瘤学领域传统上将MTD作为RP2D，通常是因为癌症患者的治疗选择相当有限，而且迫切需要推进新型药物，以满足尚未满足的重大医疗需求。另一个原因是，对于大多数细胞毒性肿瘤药物而言，理论上MTD就是最佳剂量，因为在毒性所允许的剂量范围内，它能提供最大疗效。

美国食品药品监督管理局的ProjectOptimus（美国食品药品监督管理局，2022年b）由肿瘤学卓越中心于2021年发起，是对肿瘤学传统剂量探索中现状提出挑战的历史性一步。剂量优化是Project Optimus的核心议题之一，并于2021年4月美国癌症研究协会（AACR）年会的“肿瘤学监管热点”专题会议上被提出。截至本文于2023年初撰写时，已过去近两年，在此期间，我们看到该主题出现了不少值得注意的论文和监管实例，例如Shah 等人（2021）发表于《新英格兰医学杂志·观点》的文章，强调“少即是多”；Blumenthal 等人（2021）发表于癌症研究之友的白皮书，讨论了外界对肿瘤学剂量优化预期的变化；以及Fourie Zirkelbach等人（2022）发表于《临床肿瘤学杂志》特刊：肿瘤学统计学的文章，通过近期案例强调剂量优化策略。ProjectOptimus 的众多倡议之一，是建议在随机化设置下评估多个剂量，以便更好地理解剂量/暴露-活性/毒性关系，并在申报前确定使获益-风险比最大化的剂量。目前正在提出新颖且实用的统计设计，这些设计不仅遵循Project Optimus 的精神，还旨在根据每个项目的独特特征和开发阶段实现“战略适配”。然而，大多数拟议方法仍以“只要安全性特征可接受，剂量越高越好”为思路，按序贯方式评估所有剂量水平。过去，在肿瘤学中使用规模较大的随机化队列同时评估多个候选剂量极为罕见，但预计在……首次亮相后，这一主题将从“例外”变成“常规”，而不是……

Project Optimus。在早期开发阶段，剂量优化中对多个剂量进行随机比较不一定需要针对统计错误进行统计功效设计。然而，解决关键的统计学考量仍然至关重要，例如论证样本量的合理性，以便为剂量选择产生令人满意且稳健的决策，尤其是在存在多种适应症、要求在药物开发早期阶段采用更高效设计的情况下。

为响应 Project Optimus，并结合监管预期与反馈以及工业界申办方视角下的实际考量，我们提出一种名为 MATS（多臂两阶段）的自适应设计，旨在贝叶斯分层模型下于单项研究中同时实现概念验证（PoC）和剂量优化（DO）两项目标。我们假设剂量递增已在首次人体I期试验中完成，并已选出两种剂量用于进一步扩展。例如，这两种剂量可以是最大耐受剂量（MTD）或（MTD-1）剂量，也可以是两种均低于最大耐受剂量的剂量。扩展阶段的目标是：1）进行概念验证，即确定这两种剂量在多个适应症中是否具有抗肿瘤活性；2）进行剂量优化，即确定两种剂量中哪一种在疗效方面更佳。MATS设计允许在第一阶段同时评估多个适应症的初步活动（这是肿瘤学扩展阶段的典型情形），随后仅在第二阶段针对有前景的适应症，将两种剂量随机分配以进行剂量优化。大量模拟表明，MATS设计能够以相当高的概率识别最佳剂量并正确完成概念验证，适当利用队列间、阶段间和剂量水平间可能的信息借用。MATS的一项潜在优势是能够节省样本量：第一阶段中，MATS仅在多个适应症中扩展单一剂量（而非两种或更多剂量），并且只有具有较有前景抗肿瘤活性的适应症才会进入第二阶段进行剂量比较。

本文其余部分安排如下。在第2节中，我们首先概述MATS设计，包括其基础贝叶斯分层模型、所提出的设计、

以及关键调优参数的确定。综合模拟研究的情景在第3节中进行了描述，随后在第4节中介绍模拟结果。我们在第5节中总结了将MATS作为剂量优化数据包一部分的应用、MATS设计如何与近期的FDA剂量优化指南（美国食品药品监督管理局，2023年）保持一致、根据近期监管反馈对MATS进行的潜在扩展，以及从工业界申办方角度出发的一些实际考量。

2 MATS设计

2.1 概率模型

拟议的MATS设计评估一种新型试验药物在 J 种不同适应症、 $K = 2$ 个阶段中 $I = 2$ 个不同剂量水平的疗效。令（i， j， k）表示剂量水平 i、适应症 j 和阶段 k 对应的试验组， $i = 1, \ldots, I$， $j = 1, \ldots, J$， $k = 1, \ldots, K$。将较高剂量称为 “ 高剂量 ”，对应 $i = 1$ ；较低剂量称为“低剂量”，对应 $i = 2$。假设这两个剂量已在I期试验中基于剂量限制性毒性（DLT）评估多个递增剂量的剂量递增阶段中选定。例如，高剂量和低剂量可以分别为最大耐受剂量（MTD）和（MTD-1）。随后可使用MATS设计指导试验的后续阶段，重点关注扩展和剂量优化。

在第1阶段，MATS在所有适应症中均以高剂量招募受试者。假设 n_{1j1} 名受试者已在试验组 (1, j,1) 中接受治疗，其中 y_{1j1} 名显示出初步抗肿瘤活性（例如，实体瘤中依据 RECIST 1.1 判定为完全缓解或部分缓解）。令 p_{ij} 表示剂量 i 和适应症 j 下真实但未知的响应率。我们假设 y_{1j1} 服从二项分布，即

$$y_{1j1}|n_{1j1}, p_{1j} \sim \text{Bin}(n_{1j1}, p_{1j}).$$

令 $D_{j1} \in \{0,1\}$ 表示在适应症 j的第1阶段结束时作出的决策，其中$D_{j1}=1$ 表示“进入第2阶段”，且适应症 j 被认为“有前景”；而 $D_{j1}=0$ 表示“不要进入第2阶段，因无效而停止”，因此不会再为适应症 j 招募受试者。

在第2阶段，如果 $D_{j1}=1$（即，仅针对第1阶段确定为“有前景”的适应症），我们会将受试者随机分配至高剂量 DL-H 和低剂量 DL-L。假设已在试验组 $(i,j,2)$ 中接受治疗的受试者有 n_{ij2} 名，其中 y_{ij2} 名为应答者。我们假设在 $D_{j1}=1$（即第1阶段的决策为继续）的条件下， y_{ij2}服从二项分布。于是

$$y_{ij2}|n_{ij2},p_{ij},D_{j1}=1 \sim \mathrm{Bin}(n_{ij2},p_{ij}).$$

在此阶段，我们希望基于两组假设检验概念验证和剂量优化。对于概念验证，我们考虑以下原假设和备择假设，

$$H_{0,ij}: p_{ij} \le p_{0j} \quad \text{versus} \quad H_{1,ij}: p_{ij} > p_{0j},$$

其中， p_{0j} 是适应症 j 的参考应答率，该数值根据临床医生和/或历史数据确定。对于剂量优化，我们通过以下假设检验适应症 j 的两个剂量水平之间的疗效差异是否足够大，

$$H_{0,j}: p_{1j}-p_{2j} \le \epsilon \quad \text{versus} \quad H_{1,j}: p_{1j}-p_{2j} > \epsilon.$$

其中， ϵ 是预设阈值，反映两个剂量水平之间理想的疗效差异。我们进一步假设，每种剂量的疗效率均随剂量水平增加而单调不减。因此， $p_{1j} \ge p_{2j}$。这一假设适用于多种药物，例如细胞毒性药物、抗体药物偶联物、分子靶向药物和免疫疗法等；不过，由于作用机制不同，它们的治疗指数可能相对较窄或较宽。当这一假设不成立时，MATS设计可能并不适用。

令 θ_{ij} =logit(p_{ij}) 表示 $i = 0$,1, 2 的响应率的对数优势，其中 logit(x) =log$[x/(1-x)]$。我们定义

$$\eta_j = \theta_{1j} - \theta_{0j} \quad \text{and} \quad \gamma_j = \theta_{1j} - \theta_{2j}$$

分别表示高剂量与参考响应率之间，以及高剂量与低剂量这两个剂量之间的响应率差异（在 logit 尺度上）。根据定义，($\eta_j - \gamma_j$)量化了低剂量与参考值之间的差异。我们通过收缩先验分布对 η_j 和γ_j 进行建模，具体如下。

$$\eta_j|\eta_0, \sigma_\eta^2 \sim \mathrm{N}(\eta_0, \sigma_\eta^2) \quad \text{and} \quad \gamma_j|\gamma_0, \sigma_\gamma^2 \sim \mathrm{LogNormal}(\gamma_0, \sigma_\gamma^2),$$

它们在适应症 j 之间相互独立。为实现收缩，我们假设参数 η_0、 σ_η^2、 γ_0 和 σ_γ^2 服从以下超先验分布，即:

$$\eta_0|\mu_{\eta_0}, \sigma_{\eta_0}^2 \sim \mathrm{N}(\mu_{\eta_0}, \sigma_{\eta_0}^2) \quad \text{and} \quad \sigma_\eta^2|\alpha_\eta, \beta_\eta \sim \text{Inv-Gamma}(\alpha_\eta, \beta_\eta)$$
$$\gamma_0|\mu_{\gamma_0}, \sigma_{\gamma_0}^2 \sim \mathrm{N}(\mu_{\gamma_0}, \sigma_{\gamma_0}^2) \quad \text{and} \quad \sigma_\gamma^2|\alpha_\gamma, \beta_\gamma \sim \text{Inv-Gamma}(\alpha_\gamma, \beta_\gamma)$$

试验结束时，我们做出最终决策 $D_{j2} \in \{0, 1, 2\}$，其中 $D_{j2} = 1$ 表示选择较高剂量高剂量作为最佳剂量， $D_{j2} = 2$ 表示选择较低剂量低剂量作为最佳剂量，而 $D_{j2} = 0$ 表示两种剂量均未被选中，即在任一剂量下，适应症 j 均未观察到足够的疗效信号。

MATS设计的完整层级模型总结如下:

第1阶段似然：$y_{1j1}|n_{1j1}, p_{1j} \sim$ Bin(n_{1j1}, p_{1j}) 第2阶段似然：$y_{ij2}|n_{ij2}, p_{ij}$, $D_{j1} = 1 \sim$ Bin($n_{ij2} \cdot D_{j1}$, p_{ij}) 转换：$\theta_{0j} =$ logit变换(p_{0j}) $\theta_{ij} =$ logit变换(p_{ij}) θ_{ij} 的先验分布：适应症特异性治疗效应：$\eta_j|\eta_0, \sigma_\eta^2 \sim$ N(η_0, σ_η^2) 超先验分布：$\eta_0|\mu_{\eta_0}, \sigma_{\eta_0}^2 \sim$ N(μ_{η_0}, $\sigma_{\eta_0}^2$)$\sigma_\eta^2|\alpha_\eta, \beta_\eta \sim$ Inv-Gamma(α_η, β_η)$\gamma_0|\mu_{\gamma_0}, \sigma_{\gamma_0}^2 \sim$ N(μ_{γ_0}, $\sigma_{\gamma_0}^2$) $\sigma_\gamma^2|\alpha_\gamma, \beta_\gamma \sim$ Inv-Gamma(α_γ, β_γ)

$$\theta_{1j} = \theta_{0j} + \eta_j \text{ and } \theta_{2j} = \theta_{1j} - \gamma_j$$

$$\gamma_j|\gamma_0, \sigma_\gamma^2 \sim \text{LogNormal}(\gamma_0, \sigma_\gamma^2) \tag{1}$$

模型的后验推断通过马尔可夫链蒙特卡罗（马尔可夫链蒙特卡罗）模拟进行。具体而言，获取每个参数的马尔可夫链蒙特卡罗样本，并据此计算后续章节中的后验概率和决策。

图1通过一个示例性试验说明了 MATS 的整体设计方案。第1阶段共测试四种适应症（$J = 4$），受试者接受较高剂量 DL-H 治疗。中期分析（IA）后，两种适应症（1和3）显示出有希望的抗肿瘤活性，进入第2阶段剂量优化阶段，在这两种适应症中分别随机分配两个剂量水平（$I = 2$）：DL-H 和 DL-L。由于疗效不足，其余两种适应症（2和4）终止试验（NG：不继续）。第2阶段结束时，将为每种适应症选择 DL-H、DL-L 中的一种剂量，或不选择其中任何一种，作为最佳剂量。

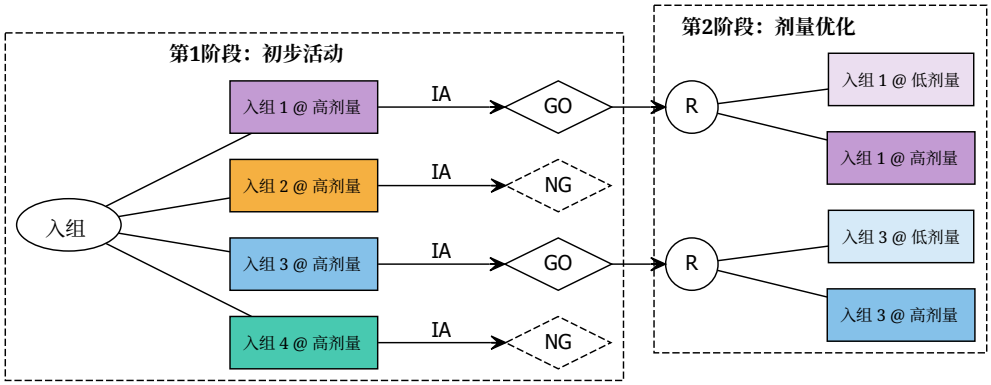


图 1：多臂两阶段（MATS）设计示意图。“NG”表示“不继续”，“R”表示“随机化”。

2.2 决策标准

本节将详细说明 MATS 中的决策标准及相关调优参数，因为在两个阶段结束时，均会针对概念验证或剂量优化作出多项决策。记 τ_{1j} 为我们希望区分的新试验药物响应率与参考响应率（例如标准治疗）之间的目标（对数优势）差异。为简便起见，下文将省略“对数优势”一词，除非省略后会造成混淆。用 data_1，表示第1阶段结束时的第1阶段数据，如果

$$\mathrm{P}(\eta_j \geq \tau_{1j} | \text{data}_1) > s_1, \tag{2}$$

然后，若试验药物被认为对适应症 j 有效，该适应症将继续进入第2阶段（继续）；否则，将不再招募适应症 j 的受试者，并因无效而停止（NG）。此时作出的决定记为 $D_{j1} \in \{0, 1\}$，分别对应继续（GO）和停止（NG）决定。这里， s_1 是一个调优参数，可通过模拟研究确定，以获得理想的运行特征。为确定 τ_{1j}，我们首先定义目标响应率 p^*_{1j}，即判定试验药物有效所需的最低响应率。根据 p_{0j}的设定

以及 p_{1j}^*，我们将 τ_{1j} 定义为

$$\tau_{1j} = \text{logit}\left(\frac{p_{1j}^* + p_{0j}}{2}\right) - \text{logit}(p_{0j}).$$

也可以采用其他方式定义 τ_{1j}，例如用 $\text{logit}(p_{1j}^*)$ 代替 $\text{logit}\left(\frac{p_{1j}^*+p_{0j}}{2}\right)$。

第 2 阶段结束时（即试验结束时），将进行最终分析。将第 I 阶段和第 II 阶段的数据记为 data_{12},，若满足以下条件，则判定适应症 j 中的高剂量 DL-H 有效:

$$\text{P}(\eta_j \geq \tau_{1j} | \text{data}_{12}) > s_2. \tag{3}$$

若满足以下条件，则判定适应症 j 中的低剂量 DL-L 有效:

$$\text{P}(\eta_j - \gamma_j \geq \tau_{1j} | \text{data}_{12}) > t_2. \tag{4}$$

若满足以下条件，则判定适应症 j 中的高剂量 DL-H 为最优剂量，且疗效优于低剂量 DL-L:

$$\text{P}(\gamma_j \geq \tau_2 | \text{data}_{12}) > w_2. \tag{5}$$

以下表 1 总结了四种不同的决策标准。

表 1：MATS设计中的决策标准。最后一行中的“剂量优化”是指检验高剂量的疗效是否显著优于低剂量。

决策标准	描述	指标
P($\eta_j \geq \tau_{1j}$|数据$_1$) $> s_1$	确定第1阶段是否继续/不继续	D_{j1}
P($\eta_j \geq \tau_{1j}$|数据$_{12}$) $> s_2$	在第2阶段确定高剂量的概念验证	POC-H$_j$
P($\eta_j - \gamma_j \geq \tau_{1j}$|数据$_{12}$) $> t_2$	在第2阶段确定低剂量的概念验证	POC-L$_j$
P($\gamma_j \geq \tau_2$|数据$_{12}$) $> w_2$	确定第2阶段的剂量优化	DO_j

对于第2阶段的三项决策标准，s_2，t_2 和 w_2 是调优参数，τ_2 表示我们希望区分两个剂量水平响应率的目标差异。τ_2 的取值取决于 p_{1j} 和 p_{2j}的具体取值，后文第2.3节将对此进行讨论。

令 $\mathbb{I}\{\}$ 为指示函数，令 $D_{j1} = \mathbb{I}\{\text{Equation (2) is true}\}$ 为第1阶段的 GO/NG 决策。此外，进一步定义 POC-H$_j$ $= \mathbb{I}\{\text{Equation (3) is true}\}$、POC-L$_j$ $=\mathbb{I}\{\text{Equation (4) is true}\}$和 DO$_j = \mathbb{I}\{\text{Equation (5) is true}\}$，分别作为条件 (3)、(4) 和 (5) 的三个指示变量。利用 POC-H$_j$、POC-L$_j$和 DO$_j$，我们针对每种适应症 j 估计高剂量和低剂量的概念验证结果，以及高剂量相较于低剂量的优越性，即剂量优化。为此，令 $D_{j2} \in \{0, 1, 2\}$ 表示适应症 j 在第2阶段结束时要做出的决策，其中 $D_{j2} = 1$ 表示选择高剂量 DL-H 作为最佳剂量，$D_{j2} = 2$ 表示选择低剂量 DL-L 作为最佳剂量，而 $D_{j2} = 0$ 表示不选择任何剂量。下面的算法1给出了 MATS设计，重点关注决策 D_{j1} 和 D_{j2}。

算法1 MATS试验设计

进行试验第1阶段并记录数据$_1 = \{y_{1j1}, n_{1j1}\}$。通过马尔可夫链蒙特卡罗从后验分布中抽样 η_j。计算 $D_{j1} = \mathbb{I}$(P($\eta_j \geq \tau_{1j}$|数据$_1$) $> s_1$)。**如果** $D_{j1} = 1$ **那么**进行第2阶段试验并记录数据 $\{y_{ij2}, n_{ij2}\}$，将数据$_{12} = \{$data$_1, y_{ij2}, n_{ij2}\}$记为。通过马尔可夫链蒙特卡罗从后验分布中抽样 η_j 和 γ_j。计算 POC-H$_j$、POC-L$_j$和剂量优化$_j$，按如下方式作出第2阶段决策 D_{j2}：**如果** POC-H$_j$ $=$POC-L$_j = 0$，**那么**$D_{j2} = 0$ 且不选择任何剂量；**否则，如果** POC-H$_j$ $=$POC-L$_j$ $=$ DO$_j = 1$ **那么**$D_{j2} = 1$ 并选择高剂量 DL-H；**否则，如果** POC-H$_j = 1$ 且 POC-L$_j = 0$ **那么**$D_{j2} = 1$ 并选择高剂量 DL-H；**否则**$D_{j2} = 2$ 并选择低剂量 DL-L。

结束如果 否则在第1阶段终止
适应症 j 。

结束 if

2.3 确定 τ_2

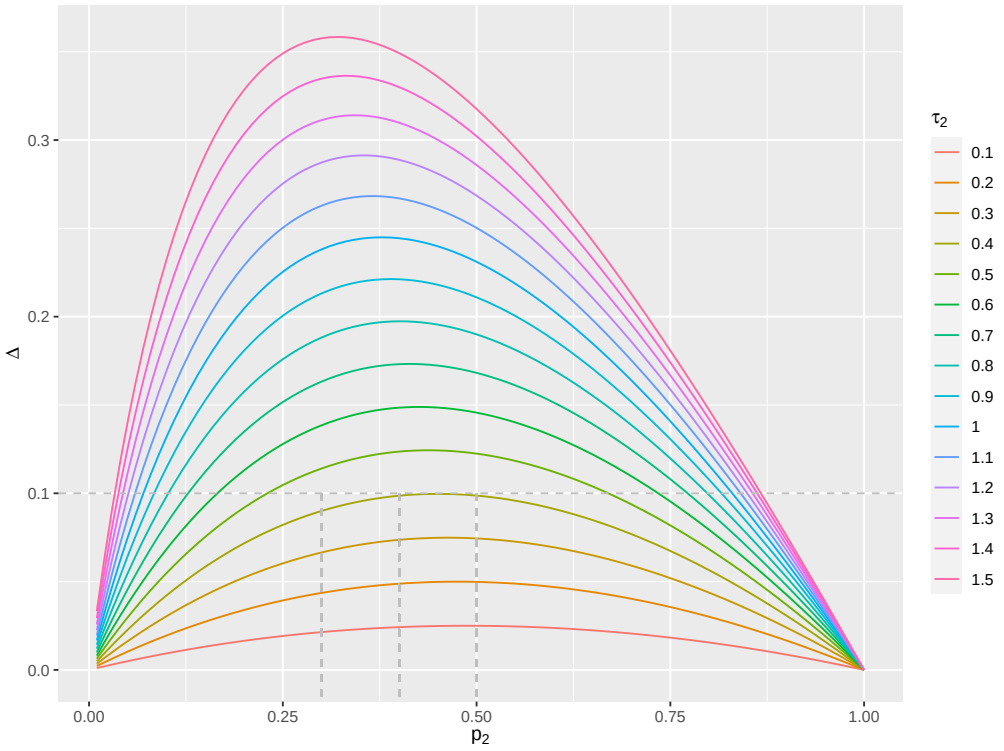


图 2： p_2 与给定 τ_2对应的差值之间的关系

在表 1 中，需要确定调优参数 τ_2 的取值。如算法1所述， τ_2 用于在第2阶段做出剂量优化决策，即 $\text{DO}_j = \mathbb{I}\{\text{P}(\gamma_j \geq \tau_2 | \text{数据}_2) > w_2\}$。因此，它用于确定两种剂量间的响应率差异是否足够大。然而， γ_j 表示 logit 尺度上的差异，这给校准 γ_j 带来了挑战，因为它与($p_{1j} - p_{2j}$)之间呈非线性关系。

为此，我们定义 $\Delta_j = \text{logit}^{-1}\{\tau_2 + \text{logit}(p_{2j})\} - p_{2j}$，它表示根据模型 (1) 得出的 p_{1j} 与 p_{2j} 之间的差异。图 2 展示了以下变量之间的关系

p_{2j} 与 Δ_j 在 τ_2 取值介于 0.1 和 1.5 之间的一组网格值下的关系。根据该图，我们提出了一种确定实际试验中 τ_2 的方法；研究者需要确定高剂量与低剂量之间的最小差异，以及低剂量下可能的响应率。例如，如果我们想区分 p_{1j} 与 p_{2j} 之间大于或等于 0.1 的差异，可以在图 2 上画一条对应于 $\Delta_j = 0.1$ 的水平线。假设低剂量下可能的响应率为 0.3、0.4 和 0.5，则分别画出 $p_{2j} = 0.3$、$p_{2j} = 0.4$ 和 $p_{2j} = 0.5$ 的垂直线。所提出的方法是，找出使相应曲线与这些垂直线的交点均位于水平线下方的最大 τ_2 值。结果是图 2 中对应于 $\tau_2 = 0.4$ 的曲线。该方法给出了网格中的最大 τ_2 值；在低剂量响应率约为 0.3、0.4 或 0.5 时，如果高剂量的真实响应率比低剂量高 0.1，则会被判定为更优（即 $\gamma_j \geq \tau_2$ 成立）。换言之，如果研究者希望高、低剂量之间的响应率相差 0.1，那么指定的 τ_2 值适用于作出剂量优化$_j$ 决策。

3 模拟设置

3.1 情景

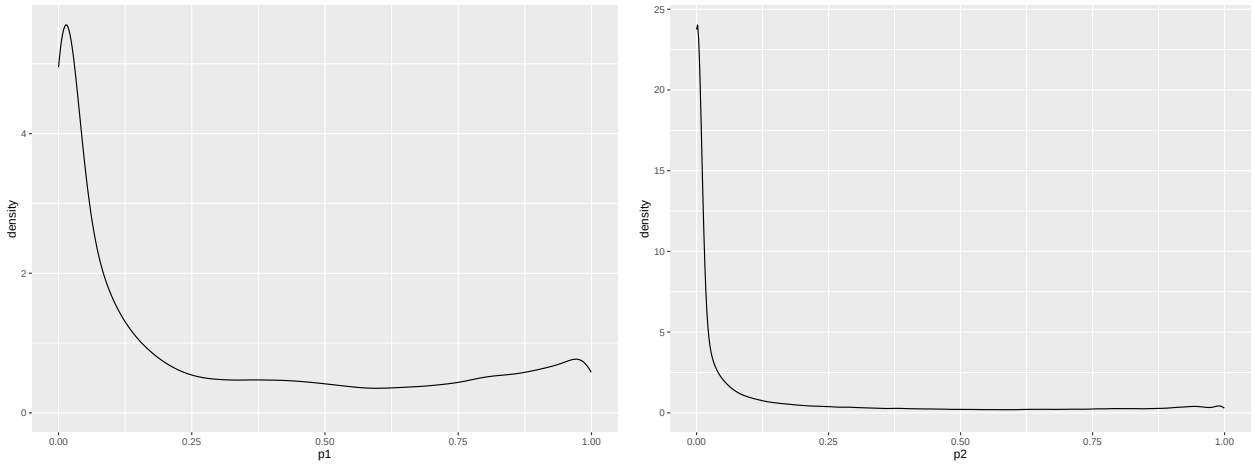
我们考虑了八种情景，并分别指定每种剂量/适应症组合的真实响应率。情景列表见表2。每种情景的名称都有特定含义：“GN”指“全局零假设”，即两种剂量在任何适应症中均无效。“GA-NS”指“无优效性的全局备择假设”，即两种剂量在所有适应症中均有效，且疗效相同。“GA-S”指“具有优效性的全局备择假设”，即两种剂量在所有适应症中均有效，但高剂量（DL-H）的疗效优于低剂量（DL-L）。“Pick-H-All”指“在所有适应症中选择较高剂量”，即只有高剂量（DL-H）在所有适应症中有效（而低剂量（DL-L）无效）。

“Pick-H-Partial”指“在部分适应症中选择较高剂量”，即只有高剂量（DL-H）在部分适应症中有效（而低剂量（DL-L）无效）。“Pick-L-Partial”指“在部分适应症中选择较低剂量”，即两种剂量在部分适应症中均有效，且疗效相同。“Mixed”是前两种情景的混合。“Intermediate”是“Pick-H-Partial”的一种变体：对于部分适应症，低剂量（DL-L）的疗效略优于安慰剂，但尚未达到与高剂量（DL-H）疗效相同的程度。

表2：所有情景下四种适应症中两种剂量的真实响应率。剂量水平1和2分别表示高剂量 DL-H 和低剂量 DL-L。情景缩写详见正文“模拟设置”部分。

情景	剂量水平	适应症1	适应症2	适应症3	适应症4
GN	1	0.1	0.2	0.1	0.2
	2	0.1	0.2	0.1	0.2
GA-NS	1	0.4	0.5	0.4	0.5
	2	0.4	0.5	0.4	0.5
GA-S	1	0.5	0.6	0.5	0.6
	2	0.4	0.5	0.4	0.5
Pick-H-All	1	0.4	0.5	0.4	0.5
	2	0.1	0.2	0.1	0.2
Pick-H-Partial	1	0.4	0.5	0.1	0.2
	2	0.1	0.2	0.1	0.2
Pick-L-Partial	1	0.4	0.2	0.1	0.5
	2	0.4	0.2	0.1	0.5
混合	1	0.4	0.2	0.1	0.5
	2	0.4	0.2	0.1	0.2
中间	1	0.4	0.2	0.1	0.5
	2	0.3	0.2	0.1	0.4

4 模拟结果



(a) p_{1j} (b) p_{2j}

图3：基于模拟中的超参数值，p_{1j} 和 p_{2j} 的先验分布。

在我们的模拟中，我们设置了 $s_1 = s_2 = t_2 = w_2 = 0.5$ 和 $\mu_{\eta_0} = 0$、$\sigma^2_{\eta_0} = 1$、$\alpha_\eta = 10$、$\beta_\eta = 1$、$\mu_{\gamma_0} = 0$、$\sigma^2_{\gamma_0} = 1$、$\alpha_\gamma = 2$、$\beta_\gamma = 1$。这些参数经过校准，以形成模糊先验。例如，图3展示了基于上述超参数值的 p_{1j} 和 p_{2j} 的先验分布。

我们将每个阶段每个剂量/适应症的样本量设为20。表3和表4总结了所提出的MATS设计的运行特征。

在表3的右侧，对于情景2至8，第2阶段有三个不同的性能指标：“完美”表示在每个适应症中都选出了正确的最佳剂量。“PoC”（“概念验证”的缩写）表示至少在一个适应症中选出了有效剂量。“剂量优化”（“剂量优化”的缩写）表示至少在一个适应症中选出了正确的最佳剂量。全文均使用相同的标记。在附录A中，给出了这些指标针对每个情景的确切定义。基本上，三个指标的取值都介于0和1之间，数值越大表示性能越好。它们的作用类似于“统计功效”。总体而言，MATS设计在所有情景中表现良好，同时对I类/II类错误进行了合理控制。

表3：每个阶段的每个剂量/适应症样本量为20时，MATS设计在八种情景下的运行特征。“完美”、“概念验证”和“剂量优化”指标表示各情景下的剂量选择，汇总于附录 A 的表 A.1。I 类和 II 类错误率是考虑多重性后的族错误率。

No.	情景	第1阶段		第2阶段		
		I 类错误率	II 类错误率	I 类错误率		
1	GN	0.124	NA	0.009		
				完美	PoC	DO
2	GA-NS	NA	0.324	0.380	1	0.986
3	GA-S	NA	0.06	0.056	1	0.689
4	Pick-H-All	NA	0.337	0.586	1	1
5	Pick-H-Partial	0.062	0.214	0.717	0.980	0.980
6	Pick-L-Partial	0.072	0.206	0.507	0.993	0.887
7	混合	0.075	0.229	0.506	0.988	0.932
8	中间	0.072	0.225	0.238	0.673	0.673

三个指标的错误率以及理想的大数值。这表明，MATS设计能够以较高的概率剔除无前景的剂量或适应症，并选出有前景的剂量或适应症。“完美”指标的数值有时会降得很低。这主要是因为，“完美”要求在所有剂量（“剂量优化”）和所有适应症（“概念验证”）上都作出正确决策，而早期临床试验的样本量较小（在我们的模拟中，每个试验组为20例），因此很难在所有决策上都做到完美。还需注意的是，这一表现指标高度依赖于所假设的情景。以“Pick-H-Partial”（完美率 $= 0.717$）为例：作出“完美”决策相对更容易，因为对于适应症3和4，适应症早期无效性的“概念验证”很容易判断，这是因为这两种适应症的两个剂量均无疗效；而对于适应症1和2，“剂量优化”也不难，因为一个剂量有效而另一个无效。不过，在此处的早期临床试验中，“概念验证”和“剂量优化”的数值更为相关，而表3中的数值都相当合理。

表4按适应症呈现了I/II类错误率。对于早期临床试验，这些数值都相当小。

图4和图5展示了模拟中MATS设计在第1阶段针对各适应症的平均样本量和继续率。请注意，在某一阶段，对于某一适应症的一种剂量，最大样本量为20。因此，对于高剂量和适应症，如果其没有前景，则不应进入第2阶段，且应在第1阶段结束后终止该试验组。否则，应在第2阶段对其进行进一步检验，在高剂量和低剂量下再随机分配40名患者。换言之，没有前景的试验组的样本量应约为20，而有前景的试验组应约为60。我们可以在图4中看到这一点。图5证实了不同情景下的继续决策率。所有结果看起来都很合理。

更多敏感性模拟结果见附录B，其中我们改变了每种剂量/适应症对应的样本量。结果表明，MATS设计对样本量的小幅变化相当稳健。

表4：每个阶段每种剂量/适应症的样本量为20时，MATS设计在八种情景下的运行特征。报告了每种适应症对应的第1阶段MATS I 类或 II 类错误率，其中每个阶段每种剂量/适应症的样本量均为20。斜体表示 I 类错误率，正体表示 II 类错误率。

No.	情景	适应症1	适应症2	适应症3	适应症4
1	GN	*0.032*	*0.033*	*0.041*	*0.026*
2	GA-NS	0.045	0.150	0.052	0.127
3	GA-S	0.007	0.028	0.006	0.021
4	Pick-H-All	0.064	0.150	0.053	0.125
5	Pick-H-Partial	0.056	0.168	*0.036*	*0.026*
6	Pick-L-Partial	0.056	*0.033*	*0.040*	0.157
7	混合	0.052	*0.027*	*0.050*	0.185
8	中间	0.060	*0.029*	*0.045*	0.174

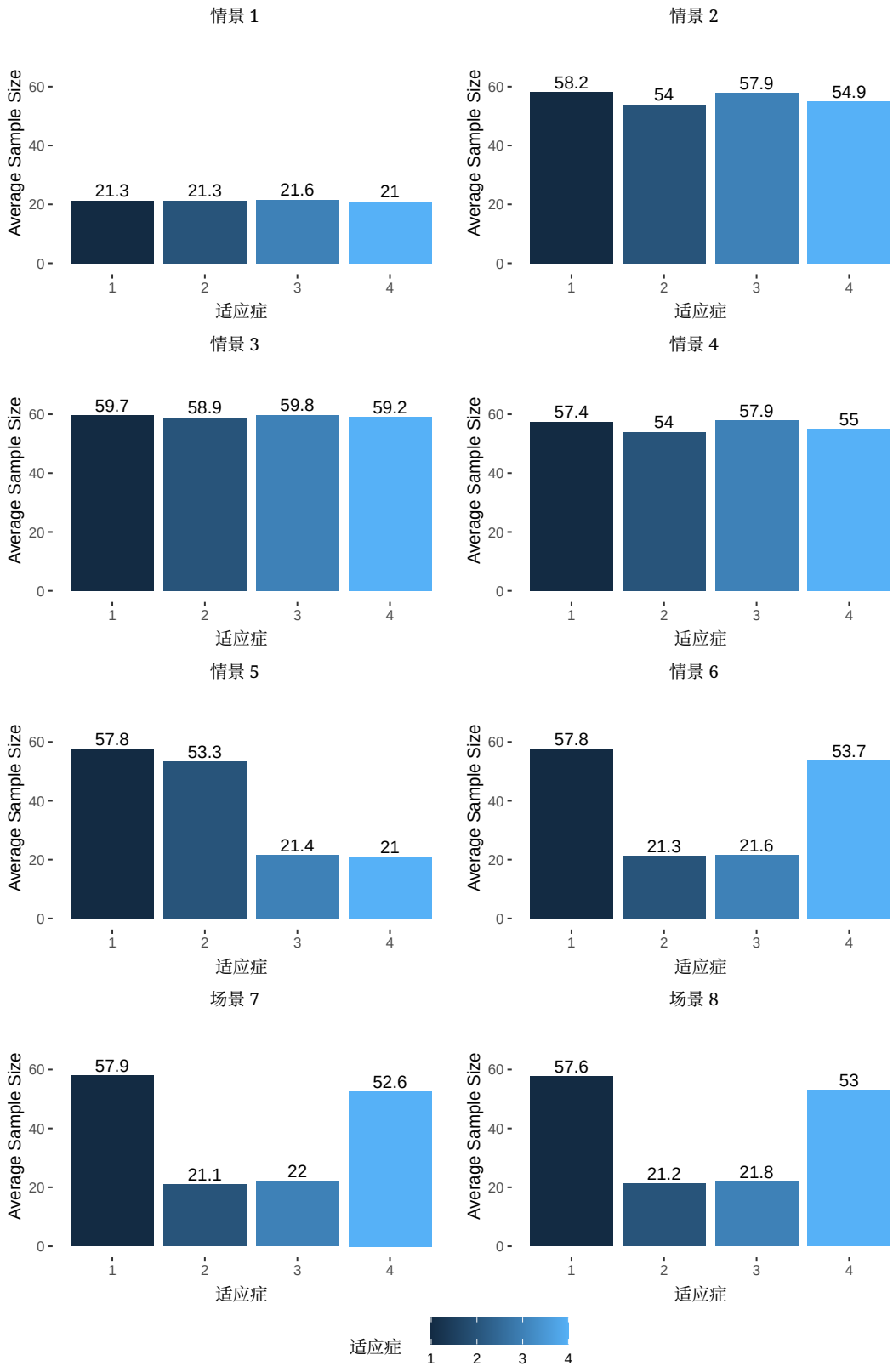


图 4：模拟中 MATS设计的平均样本量。

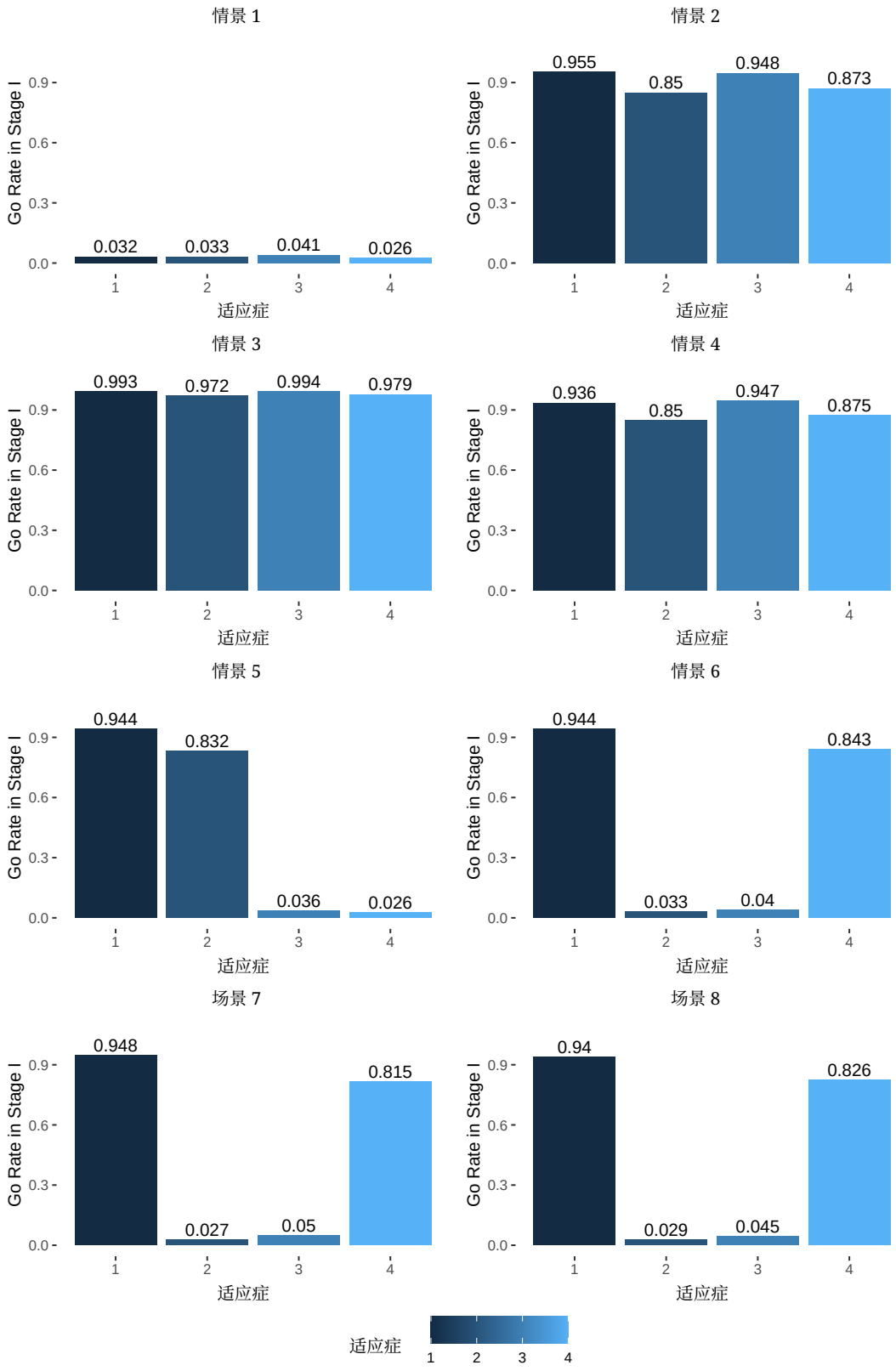


图 5：MATS设计模拟中第1阶段的平均GO率

5 讨论

当疗效是决策的主要关注点时，MATS设计为剂量递增后的剂量优化提供了一种可行方案。在这一阶段，安全性、药代动力学/药效学、暴露等终点已在剂量递增阶段得到考察，且一些剂量（如最大耐受剂量或最大耐受剂量-1）似乎相对可比，值得进一步探索。MATS设计的一个理想特点是，在第1阶段巧妙地只扩展高剂量（DL-H），而不是同时扩展两种剂量。这种选择适用于研究者认为高剂量的疗效优于低剂量的情形。因此，如果根据第1阶段数据，高剂量前景不佳，就没有必要在第2阶段进一步研究低剂量。显然，对于疗效并非被认为会随剂量单调变化的治疗药物，MATS设计并不适用。为确定MATS设计中的两种剂量，研究者应根据较早期剂量递增研究的结果，在多个逐步递增的候选剂量水平上获得初步的安全性和疗效信号；随后可利用传统剂量递增算法根据剂量限制性毒性的发生率所推荐的最大耐受剂量，确定高剂量（DL-H）的候选剂量（或者在整体数据支持的情况下，选择比最大耐受剂量低一个剂量水平的剂量作为DL-H）。可将具有初步疗效的中间剂量水平，或首次显示“作用机制验证”的剂量水平，作为低剂量（DL-L）的候选剂量。还应考虑药代动力学/药效学方面的其他证据，以刻画剂量/暴露-反应曲线，即使这些证据尚属初步。

所提出的 MATS 设计符合 FDA行业指南中关于剂量优化的若干关键考量（美国食品药品监督管理局，2023年）：

1. 指南指出：“临床试验应纳入适当广泛的目标人群，以便评估相关亚群中的剂量。”MATS设计在随机化环境中同时考虑多个适应症，从而允许.....

在早期阶段探索多个可能有前景的亚群，同时将预期总样本量控制在合理范围内。

2. 指南指出：“申办方……应规划其开发计划，以便能够在建立药物的安全性和有效性之前或同时确定最佳剂量。”MATS设计将用于评估初步活动和安全性的第1阶段与随后开展的剂量优化阶段无缝衔接，显著提高开发效率并加快开发进程。

3. 指南指出：“可以考虑采用适应性设计，在对疗效和/或安全性进行期中评估后，停止受试者进入临床试验的一个或多个剂量组。”MATS设计具有适应性，因此，只有在期中分析（第1阶段结束时）观察到初步活动且安全性可接受后，才会启动涉及更多受试者的剂量优化。

通过适当校准阈值，MATS设计可以灵活地平衡期望的 I 类错误率和 II 类错误率。由于此类剂量优化研究并非用于注册（即并非旨在证明药物优于对照药），因此可根据申办方的研发管线优先级、可预见的竞争格局等因素，设定更宽松的统计错误控制目标。一方面，可以设置调节参数，确保有较高的概率将有前景的药物从第1阶段推进至第2阶段（即在继续情景下，降低 II 类错误率），代价是在 GN 情景下，第1阶段按适应症划分的 I 类错误率略有升高。另一方面，也可以设置调节参数，确保将缺乏前景的药物从第1阶段推进至第2阶段的概率较低（即在 GN 情景下，将第1阶段按适应症划分的 I 类错误率控制在 10% 以下），代价是在继续情景下，第1阶段

按适应症划分的 II 类错误率略有升高）。我们认为，这种灵活性是 MATS设计一个非常吸引人的特点。

MATS 中用于评估初步活动的第1阶段是必要的，因为根据 Wong 等人（2019）的研究，肿瘤学领域从第1阶段进入第2阶段的成功概率仅为 57.6%，从第2阶段进入第3阶段的成功概率为 32.7%；这两项数据在各治疗领域中均为最低。若药物没有显示出初步活动信号，就无法基于临床疗效进行剂量优化，而且可能会让超出必要数量的受试者接受无效的治疗方案。在剂量优化之前设置一个小型扩展队列，也是我们从美国食品药品监督管理局收到的反馈之一（不过，这项建议未必适用于所有研究，因为剂量优化设计应根据具体情况逐案考虑）。本文假设总响应率（ORR）是衡量药物疗效的可靠指标，因此选择其作为两个阶段所关注的终点。如果发现 ORR 无法很好地预测最终临床结局，则应考虑采用其他终点，但整体的两阶段自适应设计框架保持不变。

尽管第1阶段的样本量最低为20，但具体数值可根据研究的特定考量进行调整；例如，FDA行业指南《急性髓系白血病》建议，在推定的RP2D剂量水平设置较小的队列*FDA*行业指南（美国食品药品监督管理局，2022a）。例如，为避免让过多受试者暴露于可能比预期毒性更高的较高剂量，可在第1阶段制定连续安全性监测规则（Ivanova et al., 2005）。尽管这类安全性预防措施并未通过联合建模与疗效终点进行定量整合，但仍是剂量优化过程中不可或缺的一部分。

MATS设计的另一项优势，是能够灵活处理不同适应症采用不同剂量的情况，这在联合治疗方案中相当常见。剂量递增阶段结束时，可能会确定不同的最大耐受剂量（MTD）/最大给药剂量（MAD），因为通常会针对目标适应症评估每种特定联合方案。以 ABRAXANE®（紫杉醇）为例

（美国食品药品监督管理局，2020），不同联合方案/适应症的获批剂量有所不同：转移性乳腺癌单药治疗的推荐剂量为 260 mg/m^2；与卡铂联合用于非小细胞肺癌时为 100 mg/m^2；与吉西他滨联合用于胰腺癌时为 125 mg/m^2。MATS无需针对每种适应症分别进行多项剂量优化，而是可以通过允许不同适应症采用不同剂量（即，第1阶段的“高剂量”不一定在各适应症间相同），自然地应对这类情况；同时，基于每种适应症中所测试的两个剂量水平之间的对数优势差恒定这一假设，实现不同联合方案/适应症间的信息借用。因此，预计MATS可大幅缩短开发时间线，使新型药物更快惠及亟需治疗的患者。

我们想强调，MATS设计并非旨在为剂量优化提供“放之四海而皆准”的解决方案：在实际操作中，研究团队仍应将全部数据（或可定义某种复合终点）纳入数据包，以便与监管机构讨论剂量优化。除美国食品药品监督管理局关于剂量优化的行业指南外，以下指南文件也与此主题相关，并有助于准备完整的数据包：群体药代动力学（美国食品药品监督管理局，2019）以及暴露-反应关系——研究设计、数据分析和监管应用 （美国食品药品监督管理局，2003）。

基于 MATS 设计，我们可以考虑混合模型。这样，不同适应症之间借用信息的强度便可由数据决定。为此，我们可以在 γ_j 中使用有限混合模型，并将其先验分布修改为

$$\gamma_j|x_j = t \sim \text{LogNormal}(\gamma_0, \sigma^2_{\gamma,t}), t = 0, 1$$

其中，$x_j \in \{0,1\}$ 是一个潜变量，用于表示 γ_j 的先验分布是否具有信息性。具体而言，$x_j = 0$ 表示针对 γ_j 的信息性先验分布，而 $x_j = 1$ 表示非信息性先验分布。超参数 γ_0 和 $\sigma^2_{\gamma,\ t}$ 的超先验分布设定为

$$\gamma_0|\mu_{\gamma_0},\ \sigma^2_{\gamma_0} \sim \text{N}(\mu_{\gamma_0},\ \sigma^2_{\gamma_0}) \text{ 和 } \sigma^2_{\gamma,\ t}|\alpha_{\gamma,\ t},\ \beta_{\gamma,\ t} \sim \text{Inv-Gamma}(\alpha_{\gamma,\ t},\ \beta_{\gamma,\ t})\ t = 0,\ 1$$

如图所示，为了区分上述信息含量的差异，针对与先验分布方差相关的超参数 $\sigma^2_{\gamma,\ t}$ 赋予了不同的超先验分布。

根据我们近期就多个肿瘤学项目中的剂量优化问题与美国食品药品监督管理局沟通所获得的反馈，可以对 MATS设计进行潜在扩展，使其“适合特定目的”：一种扩展方案是将 MATS 的两个阶段“反转”，首先针对每个所考虑的适应症，对两个剂量水平进行随机剂量优化；随后，对于部分或全部适应症，在观察到有希望的疗效和良好的安全性后，针对一个剂量水平设置扩展队列。与 MATS 相比，这种设计变体可能导致总样本量更大，尤其是在“有希望”的适应症较少时（因为这些适应症招募的受试者数量约为 MATS设计中的两倍）。不过，当较低剂量最终被证明是最佳剂量时，这种方案可能更有优势，因为 MATS设计要到较晚的第2阶段才会评估较低剂量。未来还需开展研究，将 MATS 与不同的设计变体进行比较，以满足不同情景下的需求。

为便于在实践中实施 MATS 设计，我们开发了一款 R Shiny 应用程序，其中包含试验模拟、数据分析和调优参数确定。应用程序可通过 https://matsdesign.shinyapps.io/mats/ 访问。

参考文献

Gabriel Blumenthal、Lakhmir Jain、Amy L Loeser、Yogesh K Pithaval、Arshad Rahman、Mark J Ratain、Manish Shah、Laura Strawn 和 Marc R Theoret。肿瘤学药物开发中的剂量优化。*Friends Cancer Res*，第 1–4 页，2021 年。

Jennifer FourieZirkelbach、Manish Shah、Jesus Vallejo、Janet Cheng、Ali Ayyoub、James Liu、Rachel Hudson、Rajeshwari Sridhara、Glenn Ison、Laleh Amiri-Kordestani 和 Sheng Tang。改进肿瘤学药物开发中的剂量优化流程，以尽量降低毒性并最大限度地惠及患者。临床肿瘤学杂志，40(30):3489–3500，2022年10月20日。

Richard Iannone。*DiagrammeR*：图形/网络可视化，2022年。网址：https://CRAN.R-project.org/package=DiagrammeR。R 软件包版本 1.0.9.9000。

Anastasia Ivanova、Bahjat F Qaqish 和 Michael J Schell。肿瘤学 II 期试验中的持续毒性监测。生物计量学，61(2):540–545，2005年6月。

Manish Shah、Arshad Rahman、Marc R Theoret 和 Richard Pazdur。肿瘤学中的药物剂量难题——少即是多。新英格兰医学杂志，385(16): 1445–1447，2021年10月14日。

美国食品药品监督管理局。关于暴露-反应关系的行业指南——研究设计、数据分析和监管应用。https://www.fda.gov/media/ 71277/download，2003年4月。

美国食品药品监督管理局。群体药代动力学行业指南。https://www.fda.gov/media/128793/download，2019年7月。

美国食品药品监督管理局。Abraxane 的处方信息。https://www.accessdata.fda.gov/drugsatfda_docs/label/year/label.pdf，2020 年。

美国食品药品监督管理局。急性髓系白血病行业指南：开发用于治疗的药物和生物制品。https://www.fda.gov/media/ 162362/download，2022年10月a。

美国食品药品监督管理局。Project Optimus。https://www.fda.gov/about-fda/oncology-center-excellence/project-optimus，2022b。

美国食品药品监督管理局。关于优化用于治疗肿瘤疾病的人用处方药和生物制品剂量的行业指南。https://www.fda.gov/media/164555/download，2023年1月。

Chee H Wong、Keng Wei Siah 和 Andrew W Lo。临床试验成功率及相关参数的估计。*Biostatistics*，20(2):273–286，2019年4月1日。

附录 A 一些符号

附录B 敏感性分析

B.1 增加样本量

我们可以将样本量从20增加到30。

B.2 第2阶段样本量不平衡

在第2阶段开始时，我们已从第1阶段的试验中获得了一些关于高剂量效果的信息，但没有低剂量的信息。因此，在第2阶段为低剂量分配比高剂量更多的样本是合理的。

表 A.1：不同适应症下性能指标的具体含义

情景	性能指标
GN	I 类错误率：在任一适应症中，于第2阶段选择任一剂量的百分比
GA-NS	“完美”：在所有适应症中均选择 DL2 的百分比 “概念验证”：在任何适应症中选择 DL1 或 DL2 的百分比 “剂量优化”：在任意适应症中选择 DL2 的比例
GA-S	“完美”：在所有适应症中选择 DL1 的百分比 “概念验证”：在任何适应症中选择 DL1 或 DL2 的百分比 “DO”：在任何适应症中选择 DL1 的百分比
Pick-H-All	“完美”：在所有适应症中选择 DL1 的百分比 “概念验证”：在任一适应症中选择 DL1 的比例 “剂量优化”：在任何适应症中选择 DL1 的百分比
Pick-H-Partial	“完美”：在适应症 1 和 2 中选择 DL1 的百分比 “概念验证”：在适应症1或2中选择 DL1 的百分比 “剂量优化”：在适应症1或2中选择 DL1 的比例
Pick-L-Partial	“完美”：在适应症1和4中选择 DL2 的比例 “概念验证”：在适应症1或4中选择 DL1 或 DL2 的比例 “剂量优化”：在适应症1或4中选择 DL2 的比例
混合	“完美”：在适应症4中选择DL1且在适应症1中选择DL2的百分比 “概念验证”：在适应症1中选择DL1或DL2，或在适应症4中选择DL1的百分比 “剂量优化”：在适应症4中选择DL1或在适应症1中选择DL2的比例
中间	“完美”：在适应症1和4中选择 DL1 的百分比 “PoC”：在适应症1中选择DL1的百分比或4 “剂量优化”：在适应症1或4中选择DL1的百分比

表 B.1：八种情景下 MATS设计的运行特征，每个阶段的每个剂量/适应症样本量为 30。“完美”、“概念验证”和“剂量优化”指标分别指每种情景下的剂量选择，其汇总见附录 A 中的表 A.1。I 类和 II 类错误率是考虑多重性后的族错误率。

No.	情景	第1阶段		第2阶段		
		I 类错误率	II 类错误率	I 类错误率		
1	GN	0.052	NA	0		
				完美	PoC	DO
2	GA-NS	NA	0.231	0.568	1	0.998
3	GA-S	NA	0.014	0.057	1	0.666
4	Pick-H-All	NA	0.221	0.741	1	1
5	Pick-H-Partial	0.037	0.124	0.851	0.997	0.997
6	Pick-L-Partial	0.030	0.134	0.687	0.995	0.947
7	混合	0.030	0.131	0.653	0.996	0.972
8	中间	0.028	0.130	0.253	0.665	0.665

表 B.2：每个阶段中每种剂量/适应症的样本量为 30 时，MATS设计在八种情景下的运行特征。报告的是每种适应症对应的 MATS 第1阶段 I 类或 II 类错误率，其中每个阶段中每种剂量/适应症的样本量为 30。斜体表示 I 类错误率，正体表示 II 类错误率。

No.	情景	适应症1	适应症2	适应症3	适应症4
1	GN	*0.018*	*0.008*	*0.022*	*0.007*
2	GA-NS	0.012	0.112	0.018	0.109
3	GA-S	0.001	0.006	0	0.007
4	Pick-H-All	0.012	0.105	0.019	0.103
5	Pick-H-Partial	0.024	0.1	*0.023*	*0.014*
6	Pick-L-Partial	0.023	*0.009*	*0.021*	0.116
7	混合	0.024	*0.009*	*0.021*	0.111
8	中间	0.028	*0.009*	*0.019*	0.104

表 B.3：对于每种适应症，第1阶段样本量为 20，第2阶段高剂量组样本量为 10、低剂量组样本量为 20 时，MATS设计在八种情景下的运行特征。“完美”“概念验证”和“剂量优化”指标表示各情景下的剂量选择，汇总见附录 A 中的表 A.1。考虑多重性后，I 类和 II 类错误率为族错误率。

No.	情景	第1阶段		第2阶段		
		I 类错误率	II 类错误率	I 类错误率		
1	GN	0.12	NA	0.025		
				完美	PoC	DO
2	GA-NS	NA	0.336	0.377	1	0.989
3	GA-S	NA	0.056	0.058	1	0.662
4	Pick-H-All	NA	0.332	0.578	1	1
5	Pick-H-Partial	0.062	0.194	0.723	0.977	0.977
6	Pick-L-Partial	0.079	0.219	0.493	0.989	0.887
7	混合	0.071	0.216	0.502	0.99	0.937
8	中间	0.072	0.219	0.231	0.679	0.679

表 B.4：MATS设计在八种情景下的运行特征，每种适应症在第1阶段的样本量为20，第2阶段高剂量组的样本量为10，低剂量组的样本量为20。报告了MATS设计中各适应症的第1阶段 I 类或 II 类错误率，其中每种适应症在第1阶段的样本量为20，第2阶段高剂量组的样本量为10，低剂量组的样本量为20。斜体表示 I 类错误率，常规字体表示 II 类错误率。

No.	情景	适应症1	适应症2	适应症3	适应症4
1	GN	*0.038*	*0.035*	*0.037*	*0.020*
2	GA-NS	0.051	0.161	0.049	0.131
3	GA-S	0.009	0.031	0.004	0.012
4	Pick-H-All	0.045	0.142	0.054	0.133
5	Pick-H-Partial	0.052	0.151	*0.035*	*0.027*
6	Pick-L-Partial	0.059	*0.036*	*0.045*	0.170
7	混合	0.048	*0.034*	*0.039*	0.177
8	中间	0.062	*0.032*	*0.042*	0.168