免疫肿瘤学剂量探索研究中疗效生物标志物与毒性终点的整合

Yiding Zhang[1], Zhixing Xu[2], Hui Quan[2], Ji Lin[1]

[1]生物统计与程序设计，赛诺菲，450 Water Street，剑桥，马萨诸塞州 02141

[2]生物统计与程序设计，赛诺菲，55 Corporate Drive，布里奇沃特，新泽西州 08807

摘要

I期肿瘤学研究的主要目标是确定新疗法的安全性特征，并确定最大耐受剂量（MTD）。这一目标源于细胞毒性药物的开发，其依据的基本假设是剂量越高，疗效和毒性发生的可能性就越大。然而，近期旨在刺激患者免疫系统抗癌的癌症免疫疗法开发证据对这一假设提出了挑战，尤其是达到某一剂量水平后继续递增剂量，未必能提高疗效。分子靶向药物（MTA）的剂量递增研究通常不再仅依赖安全性特征。本文提出一种简单灵活的模型，利用多元高斯潜变量整合毒性终点和疗效生物标志物。该模型还可轻松扩展，以纳入其他免疫生物标志物。通过同时考虑多个结局，所提方法能够更好地识别生物学最优剂量，从而改善决策。模拟研究表明，所提方法具有理想的运行特性，能够确定风险-获益权衡最优的目标剂量。我们还将所提方法实现为一款易于使用的 R Shiny 工具。**关键词：** 贝叶斯自适应设计、剂量探索、免疫疗法、I–II期试验、生物标志物、最优剂量

1. 引言

传统上，I期肿瘤学研究的目标通常是寻找最大耐受剂量（MTD）。安全性通常以剂量限制性毒性（DLT）的发生率衡量。DLT“传统上定义为治疗期间首个周期内发生的、至少可能与治疗相关的3/4级非血液学毒性或4级血液学毒性”，但也存在一些调整（Paoletti等人 2014年）。仅关注毒性的肿瘤剂量递增试验方法大致分为两类：基于规则的设计，包括经典的3 + 3 设计及其变体；基于模型的设计，包括持续再评估法（CRM）（O’Quigley等人 1990年）、过量控制剂量递增（EWOC）（Babb等人 1998年）以及贝叶斯逻辑回归模型（BLRM）（Neuenschwander等人 2008年）。在这些标准I期设计中，通常假定毒性和疗效的概率会随剂量水平升高而增加。对于细胞毒性药物而言，这一假设可能成立。然而，随着肿瘤药物开发格局的变化，分子靶向药物（MTA）日益普及，其剂量-疗效曲线和剂量-毒性曲线可能不同于细胞毒性药物。近期，美国食品药品监督管理局（FDA）鼓励“……申办方开展剂量探索研究，以评估暴露-反应、疗效和安全性，并为注册性试验的剂量选择提供依据。”（Blumenthal等人 2022年）。这一举措被称为Optimus项目，人们更加重视早期剂量探索研究，以确保所选剂量能够为癌症患者提供最佳治疗（Blumenthal等人 2022年；Shah等人 2021年）。在这种情况下，旨在确定最大耐受剂量的传统设计将不再适用；免疫治疗试验的目标应是确定生物学最优剂量（BOD），即风险获益权衡最佳的剂量。已有多种设计被提出用于在剂量探索研究中同时纳入毒性和疗效，例如 EffTox（Thall 和 Cook 2004年）、TriCRM（Zhang等人 2006年）、EBE-CRO（Colin等人 2017年）等。其次，免疫疗法是一种创新疗法，可刺激患者的免疫系统抗癌，并能产生多种免疫反应；免疫/疗效生物标志物可提供有关免疫疗法激活免疫系统的生物学疗效的信息。免疫反应通常与免疫疗法的治疗效果相关（Galon等人 2006年；Guo、Li和Yuan 2018年）。第三，毒性数据和替代疗效/免疫生物标志物数据可能同时获得的情况相当常见

与主要疗效数据相比，生物标志物数据的获取速度更快。一种可能的解决方案是，在剂量探索分析中直接用单个或多个生物标志物的数据替代主要疗效数据。因此，需要开发更复杂但易于实施的方法，以恰当利用所有可用信息。

我们开发了一种新型I/II期设计，用于确定免疫疗法的生物学最优剂量。该设计可以轻松整合毒性终点与疗效/免疫生物标志物。对于毒性，我们考虑了多种剂量-毒性曲线，包括早发和迟发毒性；对于疗效/免疫生物标志物，我们考虑二元或有序疗效终点，并且可以直接扩展至连续疗效/免疫生物标志物终点。我们还将该方法应用于R Shiny，便于使用。

近期已有一些设计被提出，用于结合免疫反应、疗效和毒性数据，例如Liu等（2018）提出的整合式I/II期设计、SPIRIT（Guo等，2018）、BDFIT（Guo等，2018）等。所提出的方法更加灵活。BDFIT和SPIRIT考虑的疗效终点均为无进展生存期（PFS），但两种设计采用不同的效用函数，并根据研究需求适用于不同的试验场景。

本文组织如下。第2节介绍相关符号并阐述所提出的方法学。随后，我们在第3节开展模拟，以考察运行特性，并在第4节进行讨论。

2. 方法

2.1 概率模型

假设正在研究 J 个剂量。令 dose_k 表示给予患者 k 的剂量，并令d_k =log (dose$_k$/参考剂量) ，其中参考剂量可以是计划最大剂量，也可以对 d_k 进行标准化以改善模型性能。可以使用相对剂量的对数变换值或原始值，但我们发现，尤其是在剂量水平差异较大时，对数变换的性能更好

令 Y_{Tox} 表示二元毒性终点。如果观察到患者出现毒性，则Y_{Tox} 等于 1；否则，Y_{Tox} 等于 0。令 Y_{Eff} 表示基于最佳总体疗效的有序疗效终点。若患者的疗效生物标志物表达量或测量值相对较高、中等、较低或未检测到，则令 Y_{Eff} 分别等于 2、1 和 0。例如，可使用细胞角蛋白18（CK）裂解型（M30）和完整型（M65）酶联免疫吸附测定来描述上皮细胞和内皮细胞的细胞凋亡（Greystoke 等，2008）。

令 Y_{Imm} 表示生物标志物反应。免疫反应通常被评定为一个二元变量。若存在具有生物学意义的响应（例如，CD8+ T 细胞增加了某个百分比），则令 Y_{Imm} 等于 1；否则等于 0。

令 $p^{Tox} = \Pr\left(Y^{Tox} = 1|d\right)$，即给定剂量下发生剂量限制性毒性的概率。我们考虑使用潜变量 $Z^{Tox} = \alpha_1 + \beta_1 \cdot d + \varepsilon$的 Probit 回归，

$$\Pr\left(Y^{Tox} = 1|d\right) = \Phi\{E(Z^{Tox}|d)\},$$

其中 $\varepsilon \sim N(0,1)$，$E(\cdot)$ 是期望算子，$\Phi(\cdot)$ 是标准正态分布的累积分布函数。因此，Y^{Tox}可视为指示该潜变量是否为正的指标：

$$Y^{Tox} = \begin{cases} 0, & Z^{Tox} \le 0 \\ 1, & otherwise \end{cases}.$$

这就得出 $\Pr(Y^{Tox} = 1|d) = \Pr(\alpha_1 + \beta_1 \cdot d + \varepsilon > 0) = \Phi(\alpha_1 + \beta_1 \cdot d)$。

参数 α_1 和 β_1 通常通过极大似然进行估计，相应的似然函数为

$$L(\alpha_1, \beta_1|D_n) = \prod_{i=1}^{n} [\Phi(\alpha_1 + \beta_1 \cdot d_i)]^{Y_i^{Tox}} [1 - \Phi(\alpha_1 + \beta_1 \cdot d_i)]^{1-Y_i^{Tox}}$$

其中，$D_n = \{\left(Y_i^{Tox}、\ d_i\right)、\ i = 1,\ \ldots,\ n\}$表示观测数据，$Y_i^{Tox}$ 表示第i位患者的剂量限制性毒性状态，d_i 表示给予第i位患者的对数转换剂量。

类似地，对于有序疗效生物标志物终点，我们可以采用潜变量方法。也就是说，令Z^{Eff}表示与某一剂量水平d 下的Y^{Eff}相关的连续潜变量，如下所示

$$Z^{Eff} = \alpha_2 + \beta_2 \cdot d \quad + \gamma_2 \cdot d^2 \; + \epsilon$$

$$Y^{Eff} = \begin{cases} 0, & Z^{Eff} < 0 \\ 1, & 0 \le Z^{Eff} < \zeta \\ 2, & Z^{Eff} \ge \zeta \end{cases}$$

其中，ζ 是未知的疗效切点，$\epsilon \sim N(0,1)$。Z^{Eff} 可解释为患者的潜在特质。当Z^{Eff} 超过某个阈值时，就会观察到某些临床结局，例如完全缓解（CR）或部分缓解（PR）。我们假设，对于特定剂量水平d，$[Z^{Tox}, Z^{Eff}|d]$ 服从二元正态分布

$$Z \quad = [Z^{Tox}, Z^{Eff}]'$$

$$Z \quad \sim N\left(\begin{bmatrix}\eta^{Tox} \\ \eta^{Eff}\end{bmatrix}, \Sigma\right),$$

其中，$\eta^{Tox} = \mathrm{E}(Z^{Tox}|\mathrm{d})$，$\eta^{Eff} = \mathrm{E}(Z^{Eff}|\mathrm{d})$，且$\Sigma$ 是一个2×2 协方差矩阵，其方差分量为1，相关系数为ρ。可以很容易地将上述二元正态分布扩展为多元正态分布，从而将更多生物标志物信息纳入模型。例如，对于包含额外生物标志物的研究，可以将潜变量定义如下：

$$Z \quad \sim N\left(\begin{bmatrix}\eta^{Tox} \\ \eta^{Eff} \\ \eta^{Bio}\end{bmatrix}, \Sigma\right),$$

其中，η^{Bio} =E$\left(Z^{B}{}_{io}\mathrm{od7}\right.$，$\Sigma$ 是一个3×3 协方差矩阵，其方差分量为1，相关系数为 ρ.

疗效/免疫生物标志物模型中的二次项用于考虑疗效和/或免疫反应可能不会随剂量单调递增的情况。$\gamma_2 < 0$ 是使剂量-

疗效曲线符合实际所必需的。当$\gamma_2 < 0$ 和$\beta_2 > 0$时，剂量-疗效曲线变为单调递增。所提出的模型易于实施，无需像 EffTox（Thall 和 Cook，2004）那样指定权衡轮廓，也无需像 TriCRM（Zhong、Koopmeiners 和 Carlin，2012）那样为 Copula 联合分布设置额外参数。为简化起见，假设潜变量服从二元正态分布。在回归模型中为θ =（α_1、 α_2、 β_1、 β_2、 γ_2, ρ、 ζ）指定先验分布时，基本思路是，协变量的典型变化不太可能导致响应变量的概率发生巨大变化。根据先验分布和中期数据，我们可以计算后验分布。令N 表示试验样本量，令Y_i =（Y_i^{Tox}、 Y_i^{Eff}） 表示第 i 位患者的观测结果，d_i 表示其分配剂量，其中$i = 1, \ldots, N$。假设在某个中期剂量递增时点有 k 位患者，令D_k =（Y_1、 Y_2、 …、 Y_k）表示这 k 位患者的观测数据。这 k 位患者的似然为L（$D_n|\theta$）= $\prod_i$ L（$Y_i|d_i$、 θ）$_{n=1}$ 。令 p（θ）表示θ的联合先验分布。基于这 k 位患者的联合后验分布为p（$\theta|D_n$） ∝ L（$D_n|\theta$）p（θ）。

2.2 控制潜在过量给药风险

Babb等人（1998）提出了 EWOC 方法，通过引入贝叶斯可行水平$1-\delta$ 来加快剂量递增，使接受某一剂量水平治疗且该剂量“毒性过高”的患者预测比例等于 δ。过毒剂量水平定义为超过最大耐受剂量（MTD）的剂量水平；最大耐受剂量是指剂量限制性毒性的概率不超过特定值的最高剂量水平。对于某一剂量水平，超过预先定义的 MTD 的剂量限制性毒性的后验概率，可用于估计该剂量水平下“过量给药”患者的预测比例。对于不含潜变量的概率模型，可直接根据耐受分布的后验样本计算剂量限制性毒性的后验概率。常见的耐受分布示例包括

$$F(Z) = \begin{cases} \Phi(Z), \ Probit\ model \\ \frac{e^Z}{1+e^Z}, \ logistic\ model \end{cases},$$

其中，Z 是Z^{Tox}的推广。获得参数$\alpha\ and\ \beta$的后验样本后，可以通过与第2.1节定义的剂量进行线性组合，得到$Z^* = \{z_1^*, z_2^*, \ldots\}$的后验样本，例如，$z_1^* = \alpha_1^* + \beta_1^* \cdot d + \varepsilon$ 代入后验样本Z^*，即可得到剂量限制性毒性概率的一个后验样本：

$F(Z^*)$。无论是否包含潜变量，Probit模型中剂量限制性毒性概率的后验均值都相同，因为 $E(Z^*_{latent}) = E(Z_{no\ laten}{}^*{}_t + \varepsilon) = E(Z_{no\ laten}{}^*{}_t)$。然而，包含潜变量的模型所得到的剂量限制性毒性概率后验样本的方差会额外受到误差项ε的随机性影响，因此，剂量限制性毒性概率后验样本的方差会大于不包含潜变量的Probit模型所得方差。此后，对于固定的过量剂量控制水平，后验样本的方差越大，预测中将在“毒性过高”的剂量水平接受治疗的患者比例就越高。因此，剂量限制性毒性概率的后验概率方差较大的剂量水平，被选为最大耐受剂量的可能性较低。

为说明 Probit模型中过量给药风险的控制方式有所不同，我们简要证明如下。对于不含潜变量的Probit模型，假设 $\alpha^* and\ \beta^*$是来自其后验分布的随机变量，则 $Var\{P(Y=1|\alpha^*,\beta^*,d)\} = Var\{\Phi(\alpha^*+\beta^* d)\} = Var\left[\frac{1}{2}\left\{1+\mathrm{erf}\left(\frac{\alpha^*+\beta^* d}{\sqrt{2}}\right)\right\}\right] = \frac{1}{4}Var\{\mathrm{erf}(\alpha^*+\beta^* d/\sqrt{2})\}$，其中 $\mathrm{erf}(\cdot)$ 是标准正态分布的误差函数，定义为 $\mathrm{erf}(x) = 2\Phi(\sqrt{2}x) - 1$。对于含潜变量的Probit模型，剂量限制性毒性的概率的方差可表示为 $Var\{P(Y=1|\alpha^*,\beta^*,d)\} = Var\{\Phi(\alpha^*+\beta^* d+\varepsilon)\} = \frac{1}{4}Var\{\mathrm{erf}(\alpha^*+\beta^* d/\sqrt{2+\sigma_\varepsilon^2})\}$。显然，无论 α 和 β的后验样本方差如何，ε的随机性始终会带来一个固定的方差来源。因此，传统的过量给药控制方法无法在含潜变量的Probit模型中为过量给药风险提供类似的容忍度，反而会使剂量递增过于保守。

由于误差函数没有闭式解，我们无法推导出方差的精确公式，因此无法考察：要达到与不含潜变量的模型相近的风险耐受水平，过量给药风险应采用多大的概率。可以通过模拟来估计超过某个给定最大耐受剂量（MTD）的剂量限制性毒性（DLT）后验概率，前提是我们知道DLT后验概率均值的估计值。例如，从 $\Phi\{\Phi^{-1}(Z_\pi)+\varepsilon\}$ 中生成大量随机样本，其中$\varepsilon \sim N(0,1)$，$\phi(Z_\pi) = \pi, \pi \in (0,1)$。可以将随机样本的均值作为DLT后验概率均值的估计值，并将超过某个预先设定的MTD（例如0.3）的随机样本比例作为潜在的过量给药风险。过量给药控制水平可根据研究者的需求或研究设计通过模拟来选择。例如，如果目标MTD为0.3，EWOC或贝叶斯逻辑回归模型（BLRM）中经典的0.25过量剂量控制水平，在含潜变量的Probit模型中可能过于保守

。模拟表明，如果目标MTD约为0.3，那么在含潜变量的Probit模型中，0.4是一个合适的候选项。

2.3 剂量寻找规则

剂量递增按以下步骤进行:

1) 将目标剂量限制性毒性发生率区间设定为 $[T_l, T_u]$ ，或仅设定上限 T_u ，并为疗效/免疫生物标志物设定一个阈值。目标剂量限制性毒性发生率的选择可根据研究的具体需求而有所不同。在大多数情况下，早期肿瘤学研究中的剂量递增仍旨在寻找最大耐受剂量，因此可以选择采用目标剂量限制性毒性发生率区间，而非单一上限，使剂量探索以安全性为导向。如果研究旨在寻找疗效较高的生物学最优剂量，同时将剂量限制性毒性发生率控制在低于某一水平且尽可能低，则可采用单一的剂量限制性毒性发生率上限。由于可靠且良好的疗效生物标志物通常尚未得到充分研究，尤其是在首次人体试验中，因此应谨慎确定疗效生物标志物在临床上有意义的下限，特别是在选择剂量限制性毒性发生率单一上限的情况下。

2) 每次将一个大小为 c 的队列（例如，3 名受试者）分配至一个剂量水平，从最低剂量水平开始。剂量递增过程中不得跳过剂量水平。

3) 根据试验每个中期时点累积的数据，使用贝叶斯MCMC算法更新后验分布。找出 j^* ，使其在 $[T_u,1)$ 中的剂量限制性毒性发生率后验概率低于某一阈值（例如 0.4 ）。实际操作中，只有当当前队列中至少 3 名患者完成了最低观察期（例如 1 个月）后，我们才会更新后验分布，以作出剂量递增决策。

4) 根据目标区间内或低于截止值的毒性，以及高于阈值的疗效/免疫生物标志物，计算目标区域内后验概率最大的剂量水平，记为 j_{target}。

5) 如果 $j_{recommend} > j_{curr}$，，则将 $j_{curr} + 1$ 剂量水平分配给下一队列。如果 $j_{recommend} = j_{curr}$，，则将 $j_{recommend}$ 剂量水平分配给下一队列。如果 $j_{recommend} < j_{curr}$，，则将 $j_{curr} - 1$ 剂量水平分配给下一队列。

6) 重复步骤 2) 至 4)，直至满足停止规则。试验停止时，$j_{recommend}$ 即为下一阶段研究（例如 II期扩展研究）的目标剂量水平。如果最低剂量的毒性过高，试验将停止，且不会为下一阶段研究推荐目标剂量。停止规则可根据研究需要制定。例如，达到预先设定的样本量时可以停止试验；也可以在某一剂量已测试两次且再次被推荐时停止剂量递增。

由于阈值的选择，剂量递增研究确定的目标剂量水平应能确保安全性特征，同时使疗效生物标志物反应尽可能高。也可以考虑指示安全风险的生物标志物。在这种情况下，可为安全性生物标志物设定上限，以确保安全性特征。

3. 模拟研究

本节通过模拟研究评估运行特性。参照示例试验中为开发所提出的设计而选取的剂量，我们设定九个剂量（mg）如下：60、75、90、105、120、135、150、165、180。最大样本量为54，队列大小为3。首先考虑两个结局 $\{Y^{Tox}, Y^{Eff}\}$:Y^{Tox}，其中剂量限制性毒性为二元变量，$Y^{Tox} = 1$ 代表剂量限制性毒性，$Y^{Tox} = 0$ 代表非剂量限制性毒性；Y^{Eff} 为有序型变量，取值为{0,1,2}，分别代表{无应答、轻度应答、高度应答}。目标毒性区间$[T_l, T_u] = [0.16,\ 0.33]$,疗效/免疫生物标志物应答（$Y^{Eff} = 1\ or\ 2$）下限$\theta_E$= 0.2。过量给药的潜在风险控制在0.4。我们假设，在作出递增决策时，已能获得疗效/免疫生物标志物测量结果。在本模拟研究中，每种情景生成1000个数据集。 $\theta =$（α_1、α_2、β_1、β_2、γ_2,ρ、ζ）的先验分布可设定如下：$\alpha_1 \sim N$（-1，1.25），$\alpha_2 \sim N$（-0.5，1.25）$\beta_1 \sim N$（0.5，1.25），$\beta_2 \sim N$（0.7，1.25）$\gamma_2 \sim N$（-0.4，1），$\zeta \sim U$（0，6），以表示对剂量限制性毒性发生率和疗效/免疫生物标志物应答率的一般合理估计，并具有较宽的置信区间

图1和图2。协方差矩阵∑ 服从具有相同尺度矩阵且自由度为3的Wishart分布。

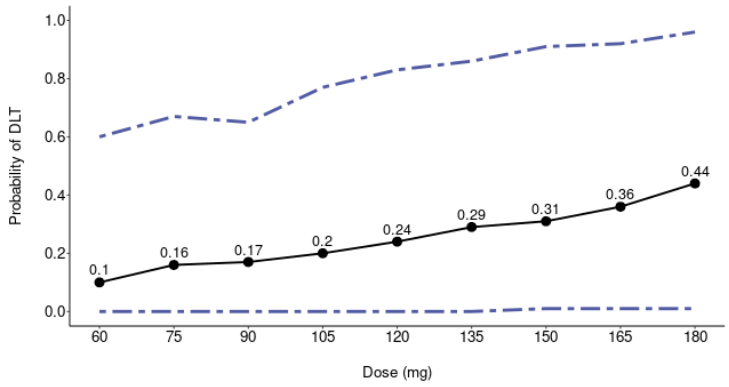


图1： 先验剂量限制性毒性发生率

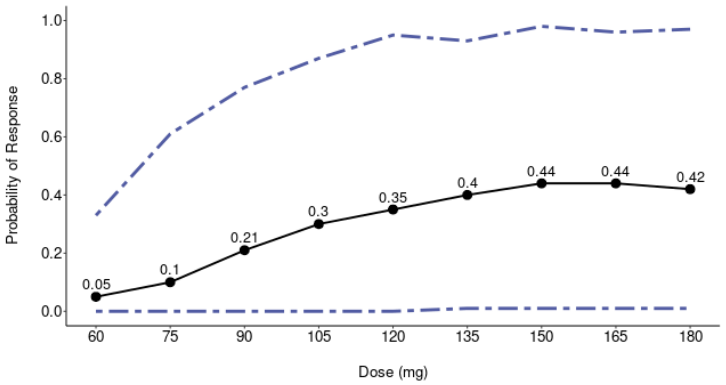


图2： 先验疗效/免疫生物标志物应答率

毒性和疗效/免疫生物标志物应答通过伯努利分布进行模拟，其剂量限制性毒性的边际概率和疗效/免疫生物标志物的应答概率见表1。粗体标示的概率是在目标毒性区间内或高于疗效/免疫生物标志物下限阈值的概率。请注意，情景1—4的毒性设定相同，以评估经典基于毒性的剂量探索设计所带来的边际增益；采用相同的目标剂量限制性毒性发生率区间和过量毒性控制水平时，我们

将模拟表现与仅考虑毒性的BLRM方法和所提出的潜变量Probit回归模型进行比较。对于剂量-毒性和剂量-疗效/免疫生物标志物反应关系，我们考虑目标毒性区间内包含一个或两个剂量水平的情况，并考虑剂量-疗效/免疫生物标志物反应呈单调递增、先递增后趋于平台以及钟形的情况。这6种情景的关键特征总结于表2。

表1：剂量限制性毒性的概率和疗效/免疫生物标志物概率的模拟场景

Dose	60	75	90	105	120	135	150	165	180	目标剂量水平
剂量水平	1	2	3	4	5	6	7	8	9	
情景1										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4, 5
疗效概率	0.05	0.1	0.18	**0.22**	**0.23**	**0.24**	**0.25**	**0.26**	**0.27**	
情景2										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4, 5
疗效概率	0.05	0.1	0.18	**0.38**	**0.4**	**0.42**	**0.44**	**0.45**	**0.46**	
场景 3										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4, 5
疗效概率	0.05	0.1	0.18	**0.28**	**0.4**	**0.42**	**0.44**	**0.45**	**0.46**	
场景4										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4, 5
疗效概率	0.05	0.15	**0.25**	**0.38**	**0.25**	**0.2**	**0.19**	**0.19**	**0.19**	
场景 5										
剂量限制性毒性概率	0.01	0.05	0.08	0.12	**0.18**	**0.27**	0.38	0.5	0.55	5, 6
疗效概率	0.05	0.15	**0.25**	**0.38**	**0.25**	**0.2**	0.19	0.19	0.19	
情景6										
剂量限制性毒性概率	0.01	0.05	0.10	0.14	**0.25**	0.35	0.5	0.55	0.7	5
疗效概率	0.05	0.1	0.18	**0.25**	**0.38**	**0.28**	**0.2**	0.19	0.19	

表2： 模拟场景的关键特征

	目标毒性	**疗效生物标志物曲线**	**关键特征**
情景1	两个候选方案	平台，对目标剂量而言相似	应答率刚好高于临界值
情景2	两个候选方案	平台期，目标剂量相似	高响应率
场景 3	两个候选方案	平台，目标剂量单调	目标剂量的应答率递增
场景4	两个候选方案	钟形，峰值位于中等剂量水平	目标剂量的应答率递减
场景 5	两个候选方案	钟形，峰值位于较低剂量水平	目标剂量的响应率递减
情景6	一个候选项	钟形	响应最高的目标剂量

对于两种结局，我们在表3中按被选为目标剂量的剂量水平百分比（平均入组患者数）汇总模拟结果。BLRM、仅针对毒性的含潜变量Probit模型，以及对毒性和疗效/免疫生物标志物进行联合建模，均采用2.4节所述的同一停止规则。我们将同时考虑毒性和主要疗效生物标志物的设计，与采用权衡等高线进行剂量推荐的常用EffTox设计进行比较。EffTox(v5.2.1)软件可从MD安德森癌症中心生物统计软件网站获取。EffTox的模拟结果基于固定的54名患者人数进行评估。EffTox的参数详见补充材料。

表3：两个结局的模拟结果

剂量水平	1	2	3	4	5	6	7	8	9	无[1]	目标剂量	过度毒性剂量
情景*1*												
BLRM	0(3)	2(3.2)	8(3.5)	30(4.2)	39(4.2)	16(2.6)	4(1.3)	1(0.4)	0(0.1)	0	69(8.4)	21(4.4)
仅 Probit-tox	0(3)	0(3)	2(3.1)	24(4)	46(4.6)	24(3.1)	4(1)	0(0.2)	0(0)	0	70(8.6)	28(4.3)
Probit 联合	0(3)	1(3)	5(3.3)	28(4.1)	46(4.5)	17(2.5)	2(0.8)	0(0.1)	0(0)	1	74(8.6)	19(3.4)
EffTox	3(5.1)	15(7.5)	41(13.4)	27(11.8)	8(6.9)	2(4.1)	2(2.6)	1(1.3)	1(0.9)	1	35(18.7)	6(8.9)
情景*2*												
Probit 联合	0(3)	0(3)	5(3.3)	30(4.2)	43(4.4)	19(2.5)	2(0.8)	0(0.2)	0(0)	1	73(8.6)	21(3.5)
EffTox	1(3.7)	6(5.3)	19(8.7)	26(9.8)	26(10.4)	15(8.2)	4(4.5)	1(2.1)	1(1.3)	0	52(20.2)	21(16.1)
场景 *3*												
Probit 联合	0(3)	0(3)	6(3.2)	27(4)	44(4.4)	20(2.6)	2(0.8)	0(0.2)	0(0)	1	71(8.4)	22(3.6)
EffTox	1(3.7)	6(5.2)	24(9.5)	23(9.9)	24(9.6)	15(7.7)	5(4.6)	1(2.3)	1(1.4)	0	47(19.5)	22(16)
场景*4*												
Probit 联合	0(3)	0(3)	6(3.3)	28(4.1)	47(4.5)	16(2.4)	1(0.8)	0(0.2)	0(0)	1	75(8.6)	17(3.4)
EffTox	6(6.0)	21(9.0)	40(14.1)	27(11.8)	4(6.0)	1(3.6)	0(2.0)	0(0.8)	0(0.5)	0	31(17.8)	1(6.9)
场景 *5*												
BLRM	0(3)	1(3.1)	3（3.3）	15（3.7）	32（4.1）	26（4.1）	17（3.6）	5（2.7）	1(1.3)	0	58(7.7)	23(5.2)
仅 Probit-tox	0(3)	0(3)	1(3.1)	10(3.5)	33(4.2)	37(4)	14(2.2)	2(0.8)	1(0.3)	0	67(8.2)	17(3.3)
Probit 联合	0(3)	0(3)	3(3.1)	24(3.6)	45(4.5)	24(3.6)	3(1.9)	0(0.7)	0(0.2)	2	69(8.1)	3(2.8)
EffTox	4(5.6)	12(6.2)	31(10.7)	37(12.8)	12(8.3)	3(4.9)	1(3.1)	0(1.5)	0(1.0)	0	15(13.2)	1(5.6)
情景*6*												
BLRM	0(3)	2(3.2)	5(3.4)	24(4)	41(4.3)	21(3)	5(3)	2(1.7)	0(0.6)	0	41(4.3)	28(7.7)
仅 Probit-tox	0(3)	0(3)	2(3.2)	21(3.9)	45(4.5)	26(3.2)	5(1.3)	1(0.3)	0(0)	0	45(4.5)	32(4.8)
Probit 联合	0(3)	0(3)	3(3.2)	24(3.9)	47(4.5)	24(2.8)	3(1)	0(0.2)	0(0)	2	47(4.5)	27(4)
EffTox	3(4.9)	12(6.5)	35(12.3)	33(12.2)	12(8.0)	3(4.9)	1(2.8)	0(1.4)	1(0.7)	1	12(8.0)	6(9.8)

[1]无：未选择任何剂量水平

对于场景1–4，由于它们具有相同的剂量-毒性曲线，因此可以评估纳入疗效/免疫生物标志物后，在目标剂量水平中被选为最大耐受剂量的比例所带来的额外提升。总体而言，引入疗效/免疫生物标志物后，所提出的方法能够提高正确选择目标剂量水平的比例，同时对过量给药的控制略有改善。接受过度毒性剂量治疗的患者平均总人数较少，这表明与采用相同过量剂量控制水平的贝叶斯逻辑回归模型（BLRM）相比，使用过量剂量控制水平0.4的含潜变量的Probit模型通常更安全。所提出的方法对过量给药的容忍程度，与采用过量剂量控制水平0.25的BLRM几乎相当。当疗效/免疫生物标志物能够区分候选剂量时，所提出的方法明显优于仅考虑毒性的方法。在场景3和5中，疗效/免疫生物标志物反应较高的候选剂量更受青睐。在场景4中，剂量水平4应比剂量水平5更常被选中；但由于剂量-反应曲线的估计会有所延迟，估计曲线的峰值会较晚出现，因此剂量水平5的疗效/免疫生物标志物反应估计值会高于剂量水平4。在情景6中，选择目标剂量的优势也有所提高。总体而言，EffTox 方法选择过度毒性剂量的比例最低，但由于其较为保守，正确选择目标剂量的比例也最低。

对于毒性和疗效/免疫生物标志物进行联合建模，我们考虑采用包含额外生物标志物的模型。额外生物标志物的一种选择可以是安全性生物标志物。对于此类生物标志物，通常测量值越低，安全性问题/毒性就越小。例如，细胞因子释放综合征相关生物标志物。也可以选择另一种生物标志物/测量指标，其测量值越低，安全性问题/毒性反而越高；例如，某些生物学测量指标需要维持在特定阈值以上才算正常。场景7、8和9的截断值分别为0.3、0.25和0.5。模拟场景列于表4中。模拟结果以选择为最大耐受剂量的百分比汇总于表5中。

表4：三种结局的模拟场景

Dose	60	75	90	105	120	135	150	165	180	目标剂量水平
剂量水平	1	2	3	4	5	6	7	8	9	

场景 7										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4
疗效概率	0.05	0.10	0.18	**0.22**	**0.23**	**0.24**	**0.25**	**0.26**	**0.27**	
概率生物学	**0.08**	**0.12**	**0.23**	**0.28**	0.35	0.41	0.46	0.51	0.54	
场景 8										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4, 5
疗效概率	0.05	0.10	0.18	**0.38**	**0.4**	**0.42**	**0.44**	**0.45**	**0.46**	
概率生物学	0.08	0.12	0.23	**0.28**	**0.35**	**0.41**	**0.46**	**0.51**	**0.54**	
场景 9										
剂量限制性毒性概率	0.01	0.05	0.10	**0.18**	**0.27**	0.38	0.5	0.55	0.7	4
疗效概率	0.05	0.10	0.18	**0.28**	**0.4**	**0.42**	**0.44**	**0.45**	**0.46**	
概率生物学	**0.80**	**0.78**	**0.65**	**0.55**	0.43	0.35	0.30	0.2	0.2	

表5：三种结局的模拟结果

剂量水平	1	2	3	4	5	6	7	8	9	无[1]	目标剂量	过度毒性剂量
场景 7	1	2	4	45	36	10	1	0	0	1	45	11
场景 8	0	1	7	36	40	13	2	0	0	1	76	15
场景 9	0	0	7	45	34	12	1	0	0	1	45	13

[1]无：未选中任何剂量水平

额外生物标志物将进一步提高目标剂量的选择比例，并降低过量给药风险。然而，当存在不同的截断值及相应的剂量关系时，趋势也会有所不同。还可以向模型中引入不同的权重，以平衡不同结局之间的权衡。例如，Li 等，2017提出使用联合单位概率质量，并通过效用函数为安全性和疗效终点的后验概率分配权重。可以通过选择估计后验期望联合效用最高的剂量，获得安全性与疗效之间的最佳权衡。

4. 讨论

我们提出了一种简单灵活的模型，可在早期剂量递增研究中整合毒性终点和疗效生物标志物信息，并恰当地进行过度毒性控制。所提出的潜变量Probit回归模型可轻松扩展以纳入额外生物标志物，也可进行调整以适应特定研究中的生物标志物使用方式。我们的模拟研究表明，该模型具有理想的运行特性，目标剂量的选择率相对较高，且

这些目标剂量具有最佳的风险获益权衡，而过度毒性风险较高的剂量选择率较低。

生物标志物信息被用作二元或有序型结局，因此，关于生物标志物测量切点的临床/生物学知识至关重要。不同于Li等（2017）提出的联合模型，我们直接在联合多元正态分布中对毒性终点和生物标志物信息建模；当不同结局之间的潜在关系不明确，但关于生物标志物的初步知识仍可用于设定阈值，以区分具有临床意义的边界时，这种做法更为可取。例如，将生物标志物被归类为“响应”的比例设定为大于0.2。因此，阈值的选择也是一个关键点。毒性终点与生物标志物信息之间的相关系数通过联合多元正态分布的协方差矩阵建模。假设协方差矩阵服从Wishart分布，通过调整Wishart分布的自由度，可以为协方差矩阵指定不同的先验分布。例如，若自由度设为4，Wishart先验会将大部分权重置于极端值$\rho_{ij} = 1$ 或-1；若自由度设为3，Wishart先验则会在区间$[-1,1]$ 至ρ_{ij} 上服从均匀分布，且具有$i \neq j$。当模型中不同结局之间的潜在关系不明确时，为方便计算，可以直接对联合多元正态分布的协方差矩阵采用自由度为3的Wishart先验。如果有关于剂量选择的其他信息，我们的方法还可以扩展为贝叶斯方法，从历史数据中借用信息，以更准确地估计参数，例如Quan等（2021）和Xu和Quan（2022）所述的方法。在某些情况下，生物标志物和毒性终点的模型可能有所不同，此时可以采用联合模型方法（例如Xu等（2021）），以更一般的方式捕捉这两个终点之间的联系。

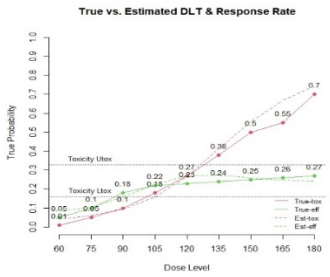
与仅依赖毒性的方法（例如贝叶斯逻辑回归模型）相比，目标剂量水平选择率更高且毒性过高剂量水平选择率更低所带来的边际性能提升，源自额外生物标志物信息。当生物标志物剂量-反应曲线迅速达到平台时，由于每个剂量水平的样本量较小，模型无法迅速捕捉斜率变化的趋势。因此，在剂量-反应曲线的变化点附近，对于变化点之后剂量水平的响应率，可能会出现一定延迟，表现为轻微高估

。如果目标剂量水平恰好位于剂量-反应曲线变化点之后，决策时可能会将目标剂量水平附近的生物标志物剂量-反应曲线视为上升，而不是平台。在具有钟形剂量-反应曲线的情景中，如果剂量-反应曲线的峰值位于目标剂量之前，则估计的剂量-反应曲线峰值可能会稍晚出现，曲线形状也可能更加平缓。这是合理的，因为在过量给药控制下，受试者被分配至较高剂量水平的可能性较低。高剂量水平的剂量-反应估计值更可能受到先验的影响。

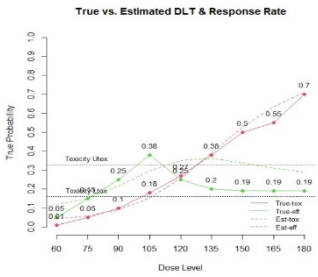
在获得剂量-毒性和剂量-生物标志物反应曲线的后验估计后，可使用效用函数作为临时的安全性与生物标志物权衡。然而，效用函数应经过谨慎选择，并具有合理的临床和医学依据。通常，效用评分量表或效用函数曲线并不能被直接解释，且最终的剂量推荐可能对效用函数或评分的选择较为敏感。与基于效用的 BOIN12（Lin 等人，2020）和两阶段基于效用的 U-BOIN（Zhou、Lee 和 Yuan，2019）相比，我们提出的方法的剂量推荐依赖于直接推断每个剂量下满足特定临床意义要求（例如区间或阈值）的剂量限制性毒性的概率和生物标志物反应的完整联合后验分布。无模型方法 Ji3+3 （Lin 和 Ji，2020）易于实施，但我们模型中生物标志物反应曲线的二次形式能够捕捉广泛的剂量-生物标志物关系，从而可以研究纳入不同类型的生物标志物或其他测量指标，例如给药时的正常生命体征率。

补充材料

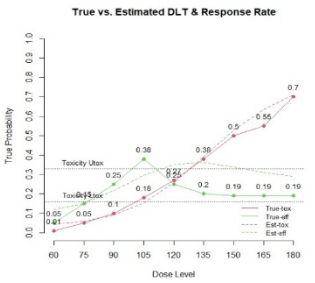
包含毒性和疗效生物标志物的模型的其他模拟结果。



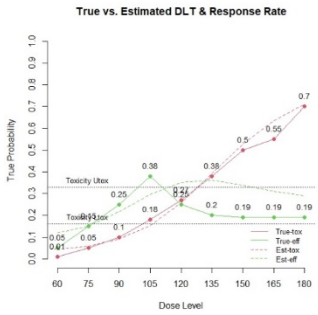
情景1



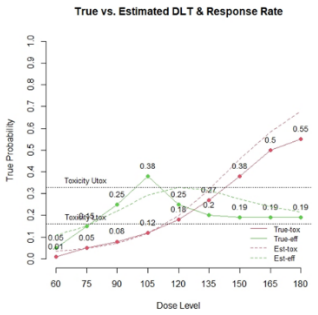
情景2



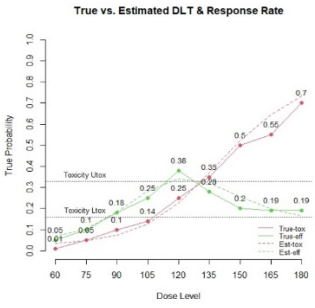
场景 3



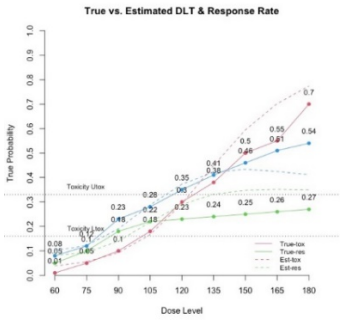
场景4



场景 5



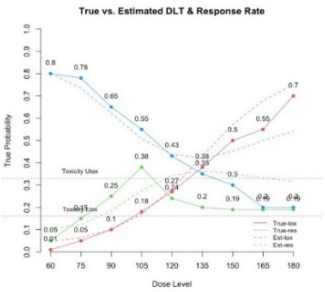
情景6



场景 7



场景 8



场景 9

图 S1：所有情景的模拟结果

图 S1 中的虚线为估计的后验平均剂量-反应/毒性曲线，实线为真实曲线。

表 S1：EffTox 方法的模拟参数

设计参数	描述
可接受剂量	
$\pi_T^* = 0.33\ (= T_u)$	可接受毒性概率上限
$p_{T,L} = 0.6$	剂量安全的最低可接受风险（1-过量剂量对照）
$\pi_E^* = 0.2\ (= \theta_E)$	可接受疗效（活性）概率下限
$p_{E,\mathrm{E}} = 0$	剂量所需疗效（活性）的最低风险
权衡	

$\pi_{1,}$ $^{E^*}$ = 0.50			可接受疗效（活性）概率下限 当毒性为 0% 时						
$\pi_{2,}$ $^{T^*}$ = 0.60			当以下情况下的可接受毒性概率上限 活性为 100%						
$\pi_{3,}$ $^{E^*}$ = 0.70 $\pi^*_{3,T}$ = 0.30			疗效（活性）和毒性的概率，对应于 如上所述的等效治疗						
引出均值									
剂量水平	1	2	3	4	5	6	7	8	9
毒性	0.1	0.16	0.17	0.2	0.24	0.29	0.31	0.36	0.44
疗效	0.05	0.1	0.21	0.3	0.35	0.4	0.44	0.44	0.43

参考文献

Babb, J., A. Rogatko, and S. Zacks. 1998. “癌症Ⅰ期临床试验：采用过量控制的高效剂量递增。”《医学统计学》 17(10):1103–20.

Blumenthal, Gideon, Lokesh Jain, Anne Loeser, Yazadi K. Pithavala, Atiqur Rahman, Mark Ratain, Mirat Shah, Laurie Strawn, and Marc Theoret. 2022. “优化肿瘤药物开发中的给药方案：问答。” *FRIENDS of CANCER RESEARCH*。检索日期：2022年10月3日（https://friendsofcancerresearch.org/wp-content/uploads/Optimizing_Dosing_in_Oncology_Drug_Development.pdf）。Colin, P., M. Delattre, P. Minini, and S. Micallef. 2017. “一种针对双变量二元终点、控制过度毒性风险的递增方法（EBE-CRO）：管理早期肿瘤临床试验中的疗效和毒性。”《生物制药统计学杂志》 27(6):1054–72.

Galon, Jérôme, Anne Costes, Fatima Sanchez-Cabo, Amos Kirilovsky, Bernhard Mlecnik, Christine Lagorce-Pagès, Marie Tosolini, Matthieu Camus, Anne Berger, Philippe Wind, Franck Zinzindohoué, Patrick Bruneval, Paul-Henri Cugnenc, Zlatko Trajanoski, Wolf-Herman Fridman, 和 Franck Pagès。2006。“人类结直肠肿瘤中免疫细胞的类型、密度和位置可预测临床结局。”《科学》（纽约州纽约市）313(5795):1960–64。

Greystoke, A., J. Cummings, T. Ward, K. Simpson, A. Renehan, F. Butt, D. Moore, J. Gietema, F. Blackhall, M. Ranson, A. Hughes, 和 C. Dive。2008。“优化细胞死亡循环生物标志物以用于常规临床应用。”《肿瘤学年鉴》：欧洲肿瘤内科学会官方期刊19(5):990–95。Guo, Beibei, Daniel Li, 和 Ying Yuan。2018。“SPIRIT：一种用于免疫治疗试验的无缝Ⅰ/Ⅱ期随机设计。”《药学统计学》17(5):527–40。

Li, Daniel H.、James B. Whitmore、Wentian Guo 和 Yuan Ji。2017 年。“Ⅰ期过继性细胞治疗剂量探索临床试验的毒性和疗效概率区间设计。”《临床癌症研究》：美国癌症研究协会官方期刊 23(1):13–20。

Lin, Ruitao, Yanhong Zhou, Fangrong Yan, Daniel Li, and Ying Yuan. 2020年. “BOIN12：用于免疫疗法和靶向治疗中基于效用的剂量探索的贝叶斯最优区间I/II期试验设计。” *JCO Precision Oncology* (4):1393–1402.

Lin, Xiaolei, and Yuan Ji. 2020年. “用于I/II期过继性细胞治疗临床试验的联合I3+3 (Ji3+3)设计。”《生物制药统计学杂志》 30(6):993–1005. Liu, Suyu, Beibei Guo, and Ying Yuan. 2018. “免疫疗法的贝叶斯I/II期试验设计。”《美国统计学会杂志》 113(523):1016–27. Neuenschwander, Beat, Michael Branson, and Thomas Gsponer. 2008. “贝叶斯方法在I期癌症试验中的关键方面。”《医学统计学》 27(13):2420–39. O’Quigley, J., M. Pepe, and L. Fisher. 1990. “持续再评估法：一种用于癌症I期临床试验的实用设计。”《生物统计学》 46(1):33–48. Paoletti, Xavier, Christophe Le Tourneau, Jaap Verweij, Lillian L. Siu, Lesley Seymour, Sophie Postel-Vinay, Laurence Collette, Elisa Rizzo, Percy Ivy, David Olmos, Christophe Massard, Denis Lacombe, Stan B. Kaye, and Jean-Charles Soria. 2014年. “分子靶向药物I期试验中剂量限制性毒性的定义：DLT-TARGETT国际调查结果。”《欧洲癌症杂志》（英格兰牛津：*1990*） 50(12):2050–56. Quan, Hui, Zhixing Xu, Meehyung Cho, Yingwen Dong, and Nan Jia. 2021. “用于非劣效性评估的历史对照数据借用。”《生物制药研究统计学》 1–10. Shah, Mirat, Atiqur Rahman, Marc R. Theoret, and Richard Pazdur. 2021. “ 肿瘤学中的药物剂量困境——少即是多。”《新英格兰医学杂志》 385(16):1445–47. Thall, Peter F., and John D. Cook. 2004. “基于疗效-毒性权衡的剂量探索。”《生物统计学》 60(3):684–93. Xu, Zhixing, and Hui Quan. 2022. “含缺失数据分量的双变量贝叶斯假设检验。”《药学统计学》 21(2):395–417. Zhang, Wei, Daniel J. Sargent, and Sumithra Mandrekar. 2006. “兼顾毒性和疗效的自适应剂量探索设计。”《医学统计学》 25(14):2365–83. Zhong, Wei, Joseph S. Koopmeiners, and Bradley P. Carlin. 2012. “用于毒性、疗效和替代疗效I/II期试验的三变量持续再评估法。”《医学统计学》 31(29):3885– 95. Zhou, Yanhong, J. Jack Lee, and Ying Yuan. 2019. “用于靶向治疗和免疫治疗中识别最佳生物剂量的基于效用的贝叶斯最优区间（U-BOIN）I/II期设计。”《医学统计学》 38(28):5299–5316.