考虑疗效和安全性及其相关性的最优两阶段设计

尹 环[1], 王维真[1,2], 张忠占[1]
(1. 北京工业大学应用数理学院, 北京 100124;
2. Department of Mathematics and Statistics, Wright State University, Dayton 45435, U. S. A)

摘 要: 在癌症的Ⅱ期临床试验研究中, 新药的疗效和安全性的研究尤为重要. 考虑到伦理和经费等问题, 减少受试人数是试验设计的主要研究目标, 通常采用多阶段优化设计达到减小样本量的效果, 最流行的是两阶段优化设计. 在综合考虑疗效和安全性的基础上, 进一步考虑两者之间的正相关性, 提出了相应的精确检验问题, 利用条件期望证明了势函数关于疗效率 p_r 和安全率 p_t 的单调不减性质,得出最大一类错误率在两点达到, 最小功效在一条线上达到. 在此基础上,在严格控制最大一类错误率和最小功效的条件下, 以最小化期望样本量为目标, 构造出了最优两阶段设计, 从而为这类试验的最优设计提供了一个精确方法. 另外, 由于新检验问题(2)的备择空间是原问题(1)备择空间的子集, 推出原检验问题(1)的最优两阶段设计也是新检验问题(2)的两阶段设计.

关键词: 多项分布; Ⅱ期临床试验; 精确检验; 条件期望; 最大一类错误率; 最小功效
中图分类号: O 212. 3; O 212. 6 **文献标志码**: A **文章编号**: 0254 - 0037(2019)12 - 0000 - 08
doi: 10. 11936/bjutxb2019040004

Optimal Two-stage Designs with Consideration of Response and Safety and Their Correlation

YIN Huan[1], WANG Weizhen[1,2], ZHANG Zhongzhan[1]
(1. College of Applied Science, Beijing University of Technology, Beijing 100124, China;
2. Department of Mathematics and Statistics, Wright State University, Dayton 45435, U. S. A)

Abstract: In phase Ⅱ clinical trial for cancer study, the efficiency and safety of a new drug are particularly important. Under consideration of ethics and cost, the optimal multi-stage designs, especially the two-stage designs, were used to reduce the sample size. In this paper, to have a positive correlation between the efficiency and safety, a new set of hypotheses and exact tests were proposed, proving that the power function is nondecreasing in the response rate p_r and the safety rate p_t, respectively, using conditional expectation. The type I error rate achieves its maximum at two points and the power achieves its minimum on a line. With control of maximum type I error rate and minimum power, the optimal two-stage designs were derived to minimize the expected total sample size, thus an exact construction method was presented for the optimal design of two-stage trials. Additionally, since the alternative space in (2) is a subset of that in (1), an optimal two-stage design for (1) is also a two-stage design for (2).

Key words: multinomial distribution; phase Ⅱ clinical trial; exact test; conditional expectation; the maximum type I error rate; the minimum power

收稿日期: 2019-04-03
基金项目: 国家自然科学基金资助项目(11771032)
作者简介: 尹环(1989—), 女, 博士研究生, 主要从事精确统计推断和试验设计方面的研究, E-mail: yinhuan94@emails.bjut.edu.cn
通信作者: 张忠占(1962—), 男, 教授, 主要从事数理统计及其应用方面的研究, E-mail:zzhang@bjut.edu.cn

精确统计推断在生物医学、环境科学等领域中有广泛应用. 出于经济、伦理方面的考虑,临床试验的前期往往只能使用小样本,例如,肿瘤学和艾滋病研究的Ⅱ期临床试验. 所谓的Ⅱ期临床试验是治疗作用的一个初步评估阶段,其目的是评估药物对患者的治疗疗效和安全性,并为Ⅲ期临床试验的设计和药剂用量提供依据. 若药物足够有效且安全,则进入Ⅲ期试验;否则,将终止该药物的研究. 在尚未确定药物安全有效的情况下,为避免让受试者处于无疗效或副作用过高的治疗过程太久,便提出了多阶段设计. 其中两阶段设计最为流行,其目的在于缩短临床试验时间、有效减少受试者数目和试验费用.

Simon[1] 在 1989 年提出 2 种两阶段设计,最优和极小极大两阶段设计,以减小试验人数. 其中最优两阶段设计被普遍采用,也是现阶段美国食品药品监督管理局 (FDA) 给定的评价新药疗效的一个标准. 在早期临床试验中,这 2 种方法都是针对新药疗效的相关研究. Banerjee 等[2],Wu 等[3] 将此类情况推广到自适应的两阶段设计,即第二阶段的样本量和临界值依赖于第一阶段受试者的反应情况. 2016 年,Shan[4] 等利用分支定界算法提出了一种最优化的自适应两阶段设计,大大减小了试验人数. 然而,在控制最大一类错误率和最小功效方面,他们没有给出一个严格的证明. 因此,无法保证给出的设计是水平为的检验. 2018 年,Wang 等[5] 提出了一种新的自适应两阶段设计,可以严格控制最大一类错误率和最小功效. 进一步,Wang 等[6] 构造了两阶段优化设计中疗效的精确置信区间. 然而,上面这些研究中都只讨论治疗的疗效,并没有考虑治疗的安全性.

Conaway 等[7],Bryant 等[8] 在 1995 年,使用了一个预先给定的胜率比 θ 来控制最大一、二类错误,其中胜率比是一个关联疗效和安全性的冗余参数. Bryant 等[8] 在 1995 年讨论了确定最优设计的两种方法:1) 使用极小值方法处理讨厌参数 θ;2) 假设疗效与安全性是独立的,即 $\theta=1$. 2007 年,Tournoux 等[9] 和 Jin[10] 也利用胜率比 θ 给出在不同数值下,找到可行解,且在参数值给定的情况下更可靠. 2008 年,金华[11] 同样利用胜率比 θ 提出了新的想法来处理安全性与疗效之间的平衡关系. 在精确统计推断中,容易验证最大一、二类错误分别在边界处达到,进而有效地控制一类错误率和功效. 而当 $\theta\neq1$ 时,对于最大一、二类错误在何处达到,并没有给出有效结果. Ray 等[12] 利用 Khattree 等[13] 提出的一种算法,在 Simon[1] 的两阶段优化设计基础上,将毒性变量考虑到了模型中. 利用预先给定的相关系数 ρ 给出了一种两阶段设计. 然而无论使用胜率比或者相关系数,它们都有一个共同缺点:如果真实的胜率比或相关系数与预先设定值不等时,则检验的显著性水平超过要求的水平. 这将导致一系列的问题出现. Thall 等[14],Thall 等[15],以及 Stallard 等[16] 提出利用贝叶斯方法构造两阶段设计. 同样无法保证能够严格控制最大一类错误率和最小功效. 2018 年,Yin 等[17] 给出胜率比错误时对优化设计的影响,并提出一类严格控制最大一类错误率和最小功效的最优多阶段设计. 在这些设计中,通常只利用上述 2 个变量来计算最大一类错误率和最小功效,即疗效率和安全率. 然而在已有的研究中,学者们忽略了一个非常重要并且有趣的问题:即相关结构如何影响检验? 他们直接使用预先给定的胜率比给出最优设计. 2007 年,Yu 等[18] 利用另一个描述疗效和毒性的结构的变量 δ,在预先给定关系下,提出多阶段优化设计. 另外,很多学者采用预先给定的胜率比 θ,如 Conaway 等[7],Bryant 等[8]. 如前所述,如果预先指定的值与真实值不同时,研究结果将导致一系列不可接受的问题. 然而,疗效和安全性的相关结构也是不得不考虑的一个参数. 例如,当疗效和安全性完全或很大程度上相反时,考虑到伦理,新药不应被采用. 因此,考虑疗效与毒性之间的相关结构是一个非常有意义的问题. 本文即讨论同时考虑疗效、安全性及其相关性时的最优两阶段设计.

本文首先给出如何选择设计参数,并定义了联系疗效与安全性的结构参数. 其次,基于多项分布,使用精确计算构造水平为 α 的检验,给出势函数关于疗效和安全性的单调性,并给出 2 个重要的约束条件:1) 控制最大一类错误率不大于给定的显著性水平 α 和最小功效不小于 $1-\beta$. 第三提出如何构造最优的两阶段设计,并针对检验问题(2)式给出最优的两阶段设计. 另外,对(2)式的检验与通常的检验 (1) 进行比较,总结出根据式(1)得到的最优两阶段设计也是根据式(2)得到的两阶段设计. 最后针对文中所有的结论给出了几个简短的总结.

1 参数的选择

考虑两阶段优化设计:第一阶段样本量为 n_1,第二阶段样本量为 n_2. 令 X_{ij} 和 Y_{ij} 分别为第一、二阶段的反应情况人数,其中 $i,j=1$ 分别表示疗效好和安全;$i,j=0$ 分别表示无疗效和有毒. 则

$$\overline{\boldsymbol{X}}=(X_{11},X_{10},X_{01},X_{00})\sim\text{Multinomial}(n_1,\overline{\boldsymbol{p}})$$

$$\overline{\boldsymbol{Y}}=(Y_{11},Y_{10},Y_{01},Y_{00})\sim\text{Multinomial}(n_2,\overline{\boldsymbol{p}})$$

式中 $\overline{\boldsymbol{p}}=(p_{11},p_{10},p_{01},p_{00})$.

令 $X_r=X_{11}+X_{10}$ 和 $X_t=X_{11}+X_{01}$ 分别表示第一阶段病人反应改善和安全的人数; $Y_r=Y_{11}+Y_{10}$ 和 $Y_t=Y_{11}+Y_{01}$ 分别表示第二阶段病人反应改善和安全的人数. 令 $p_r=p_{11}+p_{10}$ 和 $p_t=p_{11}+p_{01}$ 分别为新药有疗效和安全的概率. 则有一个自然的假设检验,

$$\begin{aligned}&H_0:\ p_r\leqslant p_{r_0}\text{或}\ p_t\leqslant p_{t_0}\\&H_1:\ p_r\geqslant p_{r_1}\text{和}\ p_t\geqslant p_{t_1}\end{aligned}\tag{1}$$

式中: p_{r_0},p_{t_0},p_{r_1} 和 p_{t_1} 是预先给定的, 且 $p_{r_0}<p_{r_1}$ 和 $p_{t_0}<p_{t_1}$; H_0 代表无疗效或不安全的治疗,所以不再进入下一步的研究,如Ⅲ期临床试验; H_1 代表有效且安全的治疗. 关于这方面的讨论,详见 Yin 等[17], Jin[10] 以及 Conaway 等[7].

然而,这个假设检验只考虑了疗效和安全性,却忽略了它们之间的相关性. 有时候,即使疗效和安全性都足够高,但不能保证该药应该进入下一步的研究,不妨看下面的例子.

例 1　考虑

$H_0:\ p_r\leqslant 0.3$ 或 $p_t\leqslant 0.3$

$H_1:\ p_r\geqslant 0.5$ 和 $p_t\geqslant 0.5$

当 $(p_{11},p_{10},p_{01},p_{00})=(0,0.5,0.5,0)$ 时,由于 $p_{11}=0$,即该药的疗效与安全性完全相反. 因此,即使 $p_r=0.5$ 和 $p_t=0.5$,也很难接受该药进入下一步的研究.

为了保证疗效和安全性的正相关性在可接受水平下,我们考虑 Conaway 等[7]提出的冗余参数胜率比: $\theta=\dfrac{p_{11}p_{00}}{p_{10}p_{01}}$. 这里,不再考虑 θ 为一固定值. 通常地, $\theta>1$ 表示疗效和安全性具有正相关性; $\theta<1$ 表示疗效和安全性具有负相关性; $\theta=1$ 表示疗效和安全性相互独立. 因此,式(1)的假设检验可改进为

$$\begin{aligned}&H_0:\ p_r\leqslant p_{r_0}\text{或}\ p_t\leqslant p_{t_0}\\&H_1:\ p_r\geqslant p_{r_1},p_t\geqslant p_{t_1}\text{和}\ \theta\geqslant\theta_1\end{aligned}$$

式中: θ_1 为一个大于或靠近 1 的常数. 但直接利用 θ,不足以保证 p_{11} 足够大. 下面这个例子说明了利用 θ 可能会出现的问题.

例 2　考虑

$H_0:\ p_r\leqslant 0.3$ 或 $p_t\leqslant 0.3$

$H_1:\ p_r\geqslant 0.5$, $p_t\geqslant 0.5$ 和 $\theta\geqslant 2$

分析下面 2 个情况:

a) $(p_{11},p_{10},p_{01},p_{00})=(0.3,0.2,0.2,0.3)$

b) $(p_{11},p_{10},p_{01},p_{00})=(0.4,0.3,0.3,0)$

根据例 2 中的假设检验容易看出,情况 a)说明可以进入下一步研究,而情况 b)说明将终止药物的研究. 但考虑到药物本身的价值,情况 b)相对 a)更被人们所接受,反而更有研究的价值.

综合考虑,一个直接的方法是引入一个简单且有效的参数 p_{11},它表示该药的疗效好且安全的概率. 则式(1)的假设检验可改进为

$$\begin{aligned}&H_0:\ p_r\leqslant p_{r_0}\text{或}\ p_t\leqslant p_{t_0}\\&H_A:\ p_r\geqslant p_{r_1}\text{和}\ p_t\geqslant p_{t_1}\text{和}\ p_{11}\geqslant p_{11}^*\end{aligned}\tag{2}$$

式中: p_{11}^* 是一个给定的常数, p_{r_0},p_{t_0},p_{r_1} 和 p_{t_1} 与(1)中相同. 显然,直接考虑 p_{11} 比 θ 更为合理.

为方便起见,首先介绍之前的几类两阶段设计: 考虑疗效的两阶段设计; 考虑疗效和安全性两变量的两阶段设计. 然后,对于式(2)中的假设,我们给出了精确的统计检验.

2　精确的统计检验

2.1　几个两阶段设计

2.1.1　Simon[1]的两阶段设计

Simon[1]的两阶段设计考虑的是药物疗效. 即

$$X_r\sim Binomial(n_1,p_r)$$

$$Y_r\sim Binomial(n_2,p_r)$$

Simon[1]中考虑的假设检验为

$$H_0:\ p_r\leqslant p_{r_0}\ \text{vs}\ H_1:\ p_r\geqslant p_{r_1}$$

式中: p_{r_0} 和 p_{r_1} 是预先给定的,且 $p_{r_0}<p_{r_1}$.

考虑设计的参数变量 $\boldsymbol{Q}=(n_1,a_r,n_2,c_r)$. Simon[1]的两阶段设计如下:

阶段 1: 观察 n_1 个病人,记录其中得到改善的病人个数 X_r.

1. 如果 $X_r\geqslant a_r$,则增加 n_2 个病人进入到第二阶段;

2. 否则,拒绝该药并终止试验.

阶段 2: 若试验进入第二阶段,则观察 n_2 个病人,记录其中得到改善的病人个数 Y_r.

1. 如果 $X_r+Y_r\geqslant c_r$,则接受该药;

2. 否则,拒绝该药.

因此,检验的拒绝域为

$$R=(X_r\geqslant a_r,X_r+Y_r\geqslant c_r),\tag{4}$$

式中: n_1,n_2,a_r 和 c_r 是 4 个确定两阶段设计的常数,且 $a_r<c_r$. 称上述两阶段设计为非适应两阶段设计.

图 1 给出了上述(4)式的一个例子,其中, $(n_1,a_r,n_2,c_r)=(16,10,30,26)$.

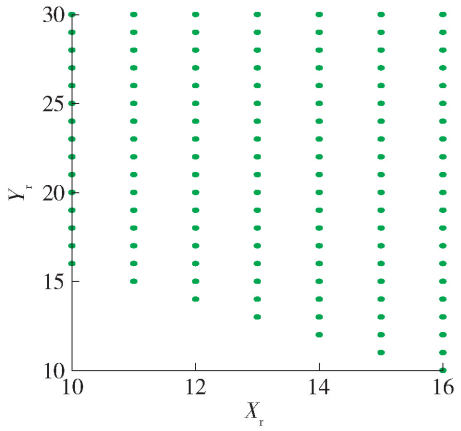

图 1 式(4)中的拒绝域

Fig. 1 Rejection region of Eq. (4)

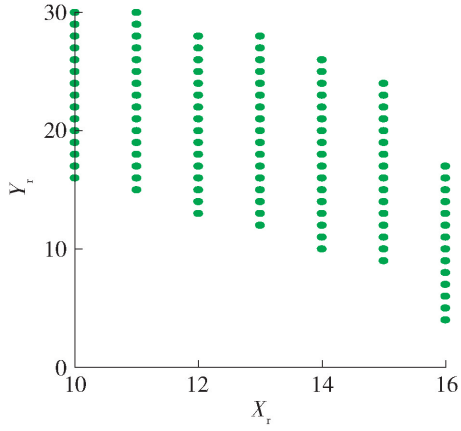

图 2 自适应设计的拒绝域

Fig. 2 Rejection region of an adaptive design

2.1.2 自适应的两阶段设计

一些学者将上述非适应两阶段设计推广到了自适应的两阶段设计,即第二阶段的临界值和样本量均依赖于第一阶段的反应情况. 因此,第二阶段的分布可写为

$$Y_r \mid X_r \sim \text{Binomial}(n_2(X_r) I(X_r \geqslant a_r), p_r)$$

式中: $I(\cdot)$ 为示性函数. 即,若 $X_r \geqslant a_r$,试验进入第二阶段,且 $Y_r \mid X_r \sim \text{Binomial}(n_2(X_r), p_r)$; 否则,试验在第一阶段停止,不再进入第二阶段. 直观上,$n_2(X_r)$ 应为 X_r 的非增函数. 拒绝域为:

$$R = (X_r \geqslant a_r, c_r(X_r) \leqslant X_r + Y_r \leqslant n_2(X_r))$$

Wang 等[5]将此设计推广到一般的情况,允许第一阶段可以因为疗效足够好而停止试验. 具体地说,引入 2 个非负整数 a_r 和 b_r,且 $b_r > a_r$. 若 $X_r < a_r$,则试验在第一阶段结束,且称该药无效; 若 $X_r \geqslant b_r$,则试验也在第一阶段结束,但称该药有效,可以进入到下一步的研究,如Ⅲ期临床试验; 若 $a_r \leqslant X_r < b_r$,则增加 $n_2(X_r)$ 个病人,进入第二阶段的试验,并观测 Y_r. 引入非增非负函数 $b(X_r)$,若 $Y_r \geqslant b(X_r)$,称该药有效,可以进入到下一步的研究; 否则,称该药无效.

图 2 给出了一个自适应设计的例子. 同图 1 的例子,考虑 $(n_1, a_r) = (16, 10)$, $n_2(X_r)$ 随 X_r 变化而变化. 例如,当 $X_r = 10$ 时,$n_2(10) = 30$; $X_r = 12$ 时,$n_2(12) = 28$.

2.1.3 考虑疗效和安全性的两阶段设计

Conaway 等[7], Jin[10] 以及金华等[11]学者将安全性也考虑进来,给出考虑两变量的两阶段设计. 其设计参数为

$$\boldsymbol{Q} = (n_1, a_r, a_t, n_2, c_r, c_t).$$

拒绝域为

$$R = (X_r \geqslant a_r, X_t \geqslant a_t, X_r + Y_r \geqslant c_r, X_t + Y_t \geqslant c_t).$$

Yin 等[17]将此类问题推广到一般的情况,即允许第一阶段可以因为疗效好和安全性高而停止试验. 然而,由于构造 2 个变量的自适应设计太过复杂,研究者通常考虑非适应的两阶段设计,即第二阶段的样本量和临界值不随第一阶段的反应情况变化而变化. 进一步,本文中不只考虑疗效 p_r 和安全性 p_t,又加入一个变量: 正相关性 p_{11}. 同样这里仅考虑非适应的两阶段设计.

2.2 考虑疗效、安全性以及正相关性的两阶段设计

在这一小节,我们将给出假设检验式(2)的精确推断. 在备择假设中考虑 p_{11},既保证了疗效与安全性的正相关性,同时也方便给出相关的优化设计. 这里考虑设计参数变量

$$\boldsymbol{Q} = (n_1, a_r, a_t, a_{11}, n_2, c_r, c_t, c_{11}). \tag{3}$$

类似仅考虑疗效时 Simon[1] 给出的两阶段设计,引入同时考虑疗效、安全性和两者相关性的两阶段设计:

第一阶段: 观察 n_1 个病人,记录 X_r, X_t 和 X_{11}.

1. 如果 $X_r \geqslant a_r, X_t \geqslant a_t$ 和 $X_{11} \geqslant a_{11}$,则增加 n_2 个病人进入第二阶段;

2. 否则,拒绝该药并终止试验.

第二阶段: 若试验进入第二阶段,则观察 n_2 个病人,记录 Y_r, Y_t 和 Y_{11}.

1. 如果 $X_r + Y_r \geqslant c_r, X_t + Y_t \geqslant c_t$ 和 $X_{11} + Y_{11} \geqslant c_{11}$,则接受该药;

2. 否则,拒绝该药.

因此,拒绝域为

$$R = (X_r \geqslant a_r, X_t \geqslant a_t, X_{11} \geqslant a_{11}, X_r + Y_r \geqslant c_r, X_t + Y_t \geqslant c_t, X_{11} + Y_{11} \geqslant c_{11}) \tag{4}$$

式中:$n_1, a_r, a_t, a_{11}, n_2, c_r, c_t, c_{11}$是 8 个确定两阶段设计的常数. 我们称之为设计参数. 容易看出 $a_r < c_r$, $a_t < c_t, a_{11} < c_{11}, a_{11} \leqslant \min(a_r, a_t)$和 $c_{11} \leqslant \min(c_r, c_t)$.

同 Yin 等[17]的考虑,我们需要控制最大的一类错误率不超过 α 和最小功效不低于 $1-\beta$. 因此,若能找到最大一类错误率和最小功效分别在哪里达到,就可以严格控制最大一类错误率和最小功效,同时也将大大减小设计参数寻优的计算量. 下面定理给出了 3 个结论: 1) 功效函数关于 p_r 和 p_t 的单调性; 2) 最大一类错误率在某两点达到; 3) 最小功效在一条直线上达到. 具体描述如下:

定理 1　考虑式(4)中的拒绝域 R,定义 $f(\bar{\boldsymbol{p}}^*) = P(R)$,式中: $f(\bar{\boldsymbol{p}}^*)$为 p_r, p_t 和 p_{11}的函数,$\bar{\boldsymbol{p}}^* = (p_r, p_t, p_{11})$. 则有

1) 当 p_r 和 p_{11}固定时,$f(\bar{\boldsymbol{p}}^*)$是关于 p_t 的非减函数; 当 p_t 和 p_{11}固定时,$f(\bar{\boldsymbol{p}}^*)$是关于 p_t 的非减函数.

2) 记$\max\limits_{H_0} P(R)$为检验式(2)的最大一类错误率,则零空间上 $P(R)$的最大值在 $\bar{\boldsymbol{p}}_{0r}^* = (p_{r_0}, 1, p_{r_0})$或 $\bar{\boldsymbol{p}}_{0t}^* = (1, p_{t_0}, p_{t_0})$处达到,即

$$\max_{H_0} P(R) = \max_{H_0} f(\bar{\boldsymbol{p}}^*) = \max(f(\bar{\boldsymbol{p}}_{0r}^*), f(\bar{\boldsymbol{p}}_{0t}^*)).$$

3) 记$\min\limits_{H_A} P(R)$为检验式(2)的最小功效,则备择空间上 $P(R)$在 $\bar{\boldsymbol{p}}_A^* = (p_{r_1}, p_{t_1}, p_{11})$处达到最小,式中:$p_{11} \in I_A = [p_{11}^*, \min(p_{r_1}, p_{t_1})]$,即

$$\min_{H_A} P(R) = \min_{H_A} f(\bar{\boldsymbol{p}}^*) = \min_{I_A} f(\bar{\boldsymbol{p}}_A^*).$$

证明: 根据 Yin 等[17]的定理证明过程,不再给予 2)的详细证明. 这里仅给出 1)和 3)的详细证明过程.

首先,考虑 1)的证明. 考虑 p_r 和 p_{11}固定时,则 p_t 关于 p_{01}不减. 因此,可以转换为证明 $f(\bar{\boldsymbol{p}}^*)$关于 p_{01}不减. 注意到,

1) $P(R) = E[E(I_R \mid X_{11}, X_{10}, Y_{11}, Y_{10})]$,式中: $I_R = I_R(X_{11}, X_{10}, X_{01}, Y_{11}, Y_{10}, Y_{01})$为一示性函数;

2) 在给定$(X_{11}, X_{10}, Y_{11}, Y_{10})$后,$X_{01}$和 Y_{01}的条件分布为

$$X_{01} \mid X_{11}, X_{10} \sim \text{Binomial}\left(n_1 - x_{11} - x_{10}, \frac{p_{01}}{1 - p_{11} - p_{10}}\right)$$

$$Y_{01} \mid Y_{11}, Y_{10} \sim \text{Binomial}\left(n_2 - y_{11} - y_{10}, \frac{p_{01}}{1 - p_{11} - p_{10}}\right)$$

3) $(X_{11}, X_{10}, Y_{11}, Y_{10})$的联合分布并不包含 p_{01}. 因此,

$$E(I_R \mid X_{11}, X_{10}, Y_{11}, Y_{10}) = P(x_r \geqslant a_r, X_{01} \geqslant a_t - x_{11}, x_{11} \geqslant c_t - x_{11} - y_{11}, x_{11} + y_{11} \geqslant c_{11}).$$

根据 Wang 等[5]的讨论,可得到当 p_r 和 p_{11}固定时,$P(R)$关于 p_t 不减. 进一步当 p_t 和 p_{11}固定时,$P(R)$关于 p_r 的单调性可类似得出. 综上,1)的证明完成.

下面证明 3). 考虑 H_A 上任一点 $\bar{\boldsymbol{p}}^* = (p_r, p_t, p_{11})$,式中: $p_{11} \in I_A$. 则根据定理 1 中 1)的结论,有

$$f(\bar{\boldsymbol{p}}^*) = f(p_r, p_t, p_{11}) \geqslant f(\max(p_{r_1}, p_{11}), p_t, p_{11})$$

情况 1　若 $p_{r_1} \geqslant p_{11}$,则

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{r_1}, p_t, p_{11}) \geqslant f(p_{r_1}, \max(p_{t_1}, p_{11}), p_{11})$$

1) 若 $p_{t_1} \geqslant p_{11}$,则有

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{r_1}, p_t, p_{11}) \geqslant f(p_{r_1}, p_{t_1}, p_{11})$$

2) 若 $p_{t_1} < p_{11}$,则有

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{r_1}, p_t, p_{11}) \geqslant f(p_{r_1}, p_{11}, p_{11})$$

由于

$$f(p_{r_1}, p_{11}, p_{11}) = P(X_r \geqslant a_r, X_{11} \geqslant a_t, X_r + Y_r \geqslant c_r, X_{11} + Y_{11} \geqslant c_t)$$

因此,若固定 $p_r = p_{r_1}$,则 $f(p_{r_1}, p_{11}, p_{11})$仅为 p_{11}的函数. 根据 Wang 等[5]的讨论,其关于 p_{11}不减. 即

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{r_1}, p_{11}, p_{11}) \geqslant f(p_{r_1}, p_{t_1}, p_{t_1})$$

情况 2　若 $p_{r_1} < p_{11}$,则有

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{11}, p_t, p_{11})$$

由于

$$f(p_{11}, p_t, p_{11}) = P(X_{11} \geqslant a_r, X_t \geqslant a_t, X_{11} + Y_{11} \geqslant c_r, X_t + Y_t \geqslant c_t)$$

根据 Wang 等[5]文中的讨论,若固定 p_t,则 $f(p_{11}, p_t, p_{11})$关于 p_{11}不减. 即

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{11}, p_t, p_{11}) \geqslant f(p_{r_1}, p_t, p_{r_1}) \geqslant f(p_{r_1}, \max(p_{t_1}, p_{r_1}), p_{r_1})$$

1) 若 $p_{r_1} \geqslant p_{t_1}$,则根据情况 1 中 2)的讨论,有

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{r_1}, p_{r_1}, p_{r_1}) \geqslant f(p_{r_1}, p_{t_1}, p_{t_1})$$

2) 若 $p_{r_1} < p_{t_1}$,有

$$f(\bar{\boldsymbol{p}}^*) \geqslant f(p_{r_1}, p_{t_1}, p_{r_1}).$$

综上所述,最小的功效在 $\bar{\boldsymbol{p}}_A^* = (p_{r_1}, p_{t_1}, p_{11})$上达到,式中: $p_{11} \in I_A = [p_{11}^*, \min(p_{r_1}, p_{t_1})]$,3)的证明完成.

从定理 1 可以看出,最大一类错误率的计算从之前的三维空间简化到 2 个点,最小功效的计算从三维空间简化到一条线. 这样,大大减小了找寻最优设计的计算量. 另外,若严格控制在零空间上 2

点 $\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}$ 和 $\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}$ 的一类错误率,以及在备择空间中直线 $\bar{\boldsymbol{p}}_{\mathrm{A}}^{*}$ 上的最小功效,则找到的设计一定是精确的设计. 这是由于找到的设计严格控制了检验问题式(2)的最大一、二类错误率.

3 最优的两阶段设计

前一部分讨论了关于式(2)的精确检验的理论内容. 这一部分,将介绍如何获得形如式(3)的最优两阶段设计,使得其期望样本量最小.

定义 PET 为早期终止试验概率,则

$$\mathrm{PET}(\bar{\boldsymbol{p}}^{*}) = 1 - P(X_{\mathrm{r}} \geqslant a_{\mathrm{r}}, X_{\mathrm{t}} \geqslant a_{\mathrm{t}}, X_{11} \geqslant a_{11})$$

进一步,期望样本量为

$$\mathrm{ESS}(\bar{\boldsymbol{p}}^{*}) = n_1 + (1 - \mathrm{PET}(\bar{\boldsymbol{p}}^{*}))n_2$$

事实上,本文的研究主要是基于Ⅱ期临床试验的研究. 根据 Simon[1] 的建议,我们选择在 $\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}$ 和 $\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}$ 两点上最大的期望样本量作为研究目标. 定义 ESS_0 为两点上的最大期望样本量,即

$$\mathrm{ESS}_0 \overset{\mathrm{def}}{=} \max(\mathrm{ESS}(\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}), \mathrm{ESS}(\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}))$$

式中:$\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}$ 和 $\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}$ 同定理 1 中给出的.

根据前面的介绍,为了找到最优的设计参数 $\boldsymbol{Q}$,必须严格控制最大一类错误率不超过 α 和最小功效不低于 $1-\beta$. 即

$$\max_{H_0} P(R) = \max(f(\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}), f(\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*})) \leqslant \alpha$$

$$\text{和} \min_{H_{\mathrm{A}}} P(R) = \min_{I_{\mathrm{A}}} f(\bar{\boldsymbol{p}}_{\mathrm{A}}^{*}) \geqslant 1-\beta \qquad (5)$$

在给出上述的最小功效表达式后,观察检验问题式(1)中的备择空间和检验问题式(2)中的备择空间,可以得到式(2)中的 H_A 是式(1)中 H_1 的子集. 因此,有下面定理:

定理 2 考虑一拒绝域 R,检验问题式(1)中对应的功效记为 $P(R|H_1)$,检验问题式(2)中的功效记为 $P(R \mid H_{\mathrm{A}})$. 若 $\min_{H_1} P(R|H_1) \geqslant 1-\beta$,则有

$$\min_{H_{\mathrm{A}}} P(R|H_{\mathrm{A}}) \geqslant \min_{H_1} P(R|H_1) \geqslant 1-\beta$$

证明:观察(1)和(2)中的备择空间,容易看出,(2)式中任一 $\bar{\boldsymbol{p}}^{*} \in H_{\mathrm{A}}$,均有 $\bar{\boldsymbol{p}}^{*} \in H_1$,即 H_{A} 是 H_1 的子集. 因此,对一固定 R,有 $\min_{H_{\mathrm{A}}} P(R|H_{\mathrm{A}}) \geqslant \min_{H_1} P(R|H_1)$. 进一步,若 $\min_{H_1} P(R|H_1) \geqslant 1-\beta$,则有

$$\min_{H_{\mathrm{A}}} P(R|H_{\mathrm{A}}) \geqslant \min_{H_1} P(R|H_1) \geqslant 1-\beta.$$

事实上,满足式(5)中的 2 个不等式的形如式(3)的设计参数可能存在很多组,合记为 S. 在集合 S 上寻找期望样本量最小的设计参数组合,记为

$$\boldsymbol{Q}^{\mathrm{opt}} = (n_1^{\mathrm{opt}}, a_{\mathrm{r}}^{\mathrm{opt}}, a_{\mathrm{t}}^{\mathrm{opt}}, a_{11}^{\mathrm{opt}}, n_2^{\mathrm{opt}}, c_{\mathrm{r}}^{\mathrm{opt}}, c_{\mathrm{t}}^{\mathrm{opt}}, c_{11}^{\mathrm{opt}})$$

则 $\boldsymbol{Q}^{\mathrm{opt}}$ 为检验式(4)的最优设计参数. 其最小期望样本量定义为

$$\mathrm{ESS}_0^{\mathrm{opt}} = \min_{S}(\mathrm{ESS}_0). \qquad (6)$$

至此,关于假设检验式(2)的最优的两阶段设计的整体思路已经给出了全部介绍,即在满足式(5)中的两个条件下,找到式(6)的最优解.

但需要注意的是,

$$f(\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}) = P(X_{11} \geqslant a_{\mathrm{r}}, X_{11} + Y_{11} \geqslant c_{\mathrm{r}} | \bar{\boldsymbol{p}}^{*} = \bar{\boldsymbol{p}}_{0\mathrm{r}}^{*})$$

$$f(\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}) = P(X_{11} \geqslant a_{\mathrm{t}}, X_{11} + Y_{11} \geqslant c_{\mathrm{t}} | \bar{\boldsymbol{p}}^{*} = \bar{\boldsymbol{p}}_{0\mathrm{t}}^{*})$$

这是由于 $\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}$ 和 $\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}$ 的特殊性,以及 $a_{11} \leqslant \min(a_{\mathrm{r}}, a_{\mathrm{t}})$ 和 $c_{11} \leqslant \min(c_{\mathrm{r}}, c_{\mathrm{t}})$. 利用这一特性,可以在控制最大一类错误率方面,将三变量的两阶段设计减小到单变量的两阶段设计,进而达到减小设计参数计算量的目的.

本文采用的是 Yin 等[17] 提出的算法,这里选取总样本量上限 $N = 65$. 即 $65 \geqslant n_1 + n_2$. 当给定 $(N, p_{\mathrm{r}_0}, p_{\mathrm{t}_0}, p_{\mathrm{r}_1}, p_{11}^{*}, \alpha, \beta)$ 时,就可以找到对应的最优的两阶段设计. 若 N 的值增大,得到的设计可能被进一步优化,但计算量也随之增大.

下面 2 个表格分别是对检验问题式(2)给出的最优的两阶段设计和同一组设计下假设检验式(1)和(2)的最小功效的对比. 在假设检验式(2)下,表 1 给出了 $(p_{\mathrm{r}_0}, p_{\mathrm{t}_0}, p_{\mathrm{r}_1}, p_{\mathrm{t}_1}, p_{11}^{*})$ 在不同数值下的最优化两阶段设计,其中 $(\alpha, \beta) = (0.05, 0.2)$. 为了减小计算量,若 $p_{\mathrm{r}_0} = p_{\mathrm{t}_0}$ 和 $p_{\mathrm{r}_1} = p_{\mathrm{t}_1}$,规定 $a_{\mathrm{r}}^{\mathrm{opt}} = a_{\mathrm{t}}^{\mathrm{opt}}$ 和 $c_{\mathrm{r}}^{\mathrm{opt}} = c_{\mathrm{t}}^{\mathrm{opt}}$. 这是因为 p_{r} 和 p_{t} 具有对称性,因此这样的假设也是合理的. 此外,表 1 没有给出 a_{11} 和 c_{11} 的值,事实上 $a_{11} = \max(0, a_{\mathrm{r}} + a_{\mathrm{t}} - n_1)$ 和 $c_{11} = \max(0, c_{\mathrm{r}} + c_{\mathrm{t}} - n_1 - n_2)$. 这是由于:1) $\bar{\boldsymbol{p}}_{0\mathrm{r}}^{*}$ 和 $\bar{\boldsymbol{p}}_{0\mathrm{t}}^{*}$ 两点的特殊性,且 $\min(a_{\mathrm{r}}, a_{\mathrm{t}}) \geqslant a_{11}$, $\min(c_{\mathrm{r}}, c_{\mathrm{t}}) \geqslant c_{11}$;2) 对一固定 $\bar{\boldsymbol{p}}^{*}$,当 $(a_{\mathrm{r}}, a_{\mathrm{t}}, c_{\mathrm{r}}, c_{\mathrm{t}})$ 固定时,功效函数 $P(R|\bar{\boldsymbol{p}}^{*})$ 关于 a_{11}, c_{11} 分别不增. 综合以上两点原因可得到 $a_{11} = \max(0, a_{\mathrm{r}} + a_{\mathrm{t}} - n_1)$ 和 $c_{11} = \max(0, c_{\mathrm{r}} + c_{\mathrm{t}} - n_1 - n_2)$. 例如,在表 1 中当给定 $(N, \alpha, \beta) = (65, 0.05, 0.2)$ 和 $(p_{\mathrm{r}_0}, p_{\mathrm{t}_0}, p_{\mathrm{r}_1}, p_{\mathrm{t}_1}, p_{11}^{*}) = (0.5, 0.5, 0.7, 0.7, 0.7)$ 时,可以得到 $(n_1^{\mathrm{opt}}, a_{\mathrm{r}}^{\mathrm{opt}}, a_{\mathrm{t}}^{\mathrm{opt}}, a_{11}^{\mathrm{opt}}, n_2^{\mathrm{opt}}, c_{\mathrm{r}}^{\mathrm{opt}}, c_{\mathrm{t}}^{\mathrm{opt}}, c_{11}^{\mathrm{opt}}) = (15, 9, 9, 28, 27, 27)$,期望样本量为 23.501 3,最大一类错误率为 0.049 93,最小功效为 0.804 4. 也就是说,选取 15 个病人进入试验的第一阶段,若 $X_{\mathrm{r}} < 9$ 或 $X_{\mathrm{t}} < 9$,则终止试验且称该药无效或毒性过高;否则试验进入第二阶段且另外选取 28 个病人进入该阶段的研究,若 $X_{\mathrm{r}} + Y_{\mathrm{r}} \geqslant 27$ 且 $X_{\mathrm{t}} + Y_{\mathrm{t}} \geqslant 27$,则接受该新药,否则认为该新药无效或者毒性高.

表 1　最优的两阶段设计: $(\alpha,\beta)=(0.05,0.2)$ 和 $p_{z_1}=p_{z_0}+0.2$, z 代表 r 和 t

Table 1　The optimal two-stage designs for $(\alpha,\beta)=(0.05,0.2)$ and $p_{z_1}=p_{z_0}+0.2$ with $z=$ r, t

p_{r_0}	p_{t_0}	p_{11}^*	n_1^{opt}	a_r^{opt}	a_t^{opt}	n_2^{opt}	c_r^{opt}	c_t^{opt}	ESS_0^{opt}	$\max_{H_0} P(R)$	$\min_{H_A} P(R)$
0.30	0.30	0.20	24	9	9	36	24	24	33.898 7	0.049 83	0.802 8
		0.30	24	9	9	33	23	23	33.073 8	0.049 86	0.804 9
		0.50	15	6	6	31	19	19	23.629 7	0.049 87	0.803 2
0.40	0.40	0.20	20	9	9	41	31	31	36.580 5	0.048 75	0.809 3
		0.30	29	14	14	32	31	31	36.791 3	0.045 78	0.801 3
		0.60	16	8	8	30	24	24	24.518 1	0.048 59	0.800 6
0.50	0.50	0.40	21	12	12	38	36	36	33.608 9	0.048 51	0.801 4
		0.60	22	13	13	37	36	36	31.684 1	0.045 05	0.804 4
		0.70	15	9	9	28	27	27	23.501 3	0.049 93	0.804 4
0.60	0.60	0.60	18	12	12	29	34	34	28.854 0	0.048 87	0.801 2
		0.70	19	13	13	28	34	34	27.626 0	0.046 78	0.805 9
		0.80	11	8	8	32	31	31	20.481 1	0.048 92	0.802 4

对于同一设计,表 2 比较了关于检验式(1)和式(2)的功效大小. 这里考虑的是 Yin 等[17]中表 4 的部分结果,他们考虑了更一般化的最优化两阶段设计,允许第一阶段可以因为疗效足够好和安全性足够高而提前停止试验,设计参数为 $\boldsymbol{Q}=(n_1,a_r,a_t,b_r,b_t,n_2,c_r,c_t)$,其中 (b_r,b_t) 为第一阶段因为好停止试验的两个临界值. 这样的设计同样可以适用于本文中的检验问题式(2),观察表 2 最后一列的结果,可以看出最小功效均大于 $1-\beta$,即检验问题式(1)下得到的这些设计均是检验问题式(2)的精确设计. 事实上,对于这些设计的最小功效 $\min_{H_A} P(R)$ 关于 p_{11}^* 是不减的. 值得注意的是,若在点 $\bar{\boldsymbol{p}}_A^*=(p_{r_1},p_{t_1},p_{11}^*)$ 控制功效不小于 $1-\beta$ 找到的最优设计 $\boldsymbol{Q}$ 的功效函数关于 p_{11} 是不减的,则检验问题式(2)对应的最优两阶段设计 $\boldsymbol{Q}^{opt}$ 即为 $\boldsymbol{Q}$,且最小功效将在一点达到,即 $\bar{\boldsymbol{p}}_A^*=(p_{r_1},p_{t_1},p_{11}^*)$,而非定理 1 的 3)中的一条线,这也大大减小了搜寻最优设计的计算量. 对比表 2 中倒数第 1、2 列,可以看出根据检验问题式(1)得到的最优化两阶段设计的最小功效均不大于基于检验问题式(2)计算得到的最小功效,这也验证了定理 2 给出的相关结论. 依据定理 2,进一步可以得出: 1)检验下得到的最优化两阶段设计也是 2)的两阶段设计,但不一定是检验问题式(2)中最优的两阶段设计. 原因是检验式(1)中的备择空间包含新提出的假设检验式(2)中的备择空间,即式(2)中的 H_A 是式(1)中 H_1 的子集.

表 2　(1)、(2)两种假设检验的功效值对比, $z=r,t$

Table 2　The comparison of powers for tests (1) and (2) with $z=r,t$

p_{r_0}	p_{t_0}	p_{11}^*	n_1^{opt}	a_z^{opt}	b_z^{opt}	n_2^{opt}	c_z^{opt}	$\min_{H_1} P(R)$	$\min_{H_A} P(R)$
0.40	0.40	0.20	29	14	18	34	32	0.803 4	0.803 4
		0.30	–	–	–	–	–	–	0.807 2
		0.40	–	–	–	–	–	–	0.818 6
0.50	0.50	0.40	19	11	15	29	30	0.802 0	0.802 0
		0.50	–	–	–	–	–	–	0.812 0
		0.60	–	–	–	–	–	–	0.834 1
0.60	0.60	0.60	20	13	15	38	38	0.801 6	0.801 6
		0.70	–	–	–	–	–	–	0.823 9
		0.80	–	–	–	–	–	–	0.898 9

4 结论

1）本文基于癌症等疾病的早期临床试验的研究,提出一类新的假设检验. 不仅考虑了疗效和非毒性的临界值,也兼顾到疗效和非毒性之间的相关性. 这样既保证新的治疗的疗效和非毒性的最低保障,同时也保证两个变量间的正相关性的最低标准.

2）本文基于假设检验式(2)给出了水平为 α 的精确检验. 综合考虑道德和经费等问题,本文采用了 Simon[1] 提出的两阶段优化设计,在给出功效函数在整个空间上关于疗效率和安全率分别不减的结论下,找到了最大一类错误率和最小功效. 特别地,最大一类错误率的计算从三维空间降到两个点;最小功效的计算从三维空间降到一维空间,进一步若找到的最优设计关于 p_{11} 不减,则可进一步降到一个点.

3）在严格控制一、二类错误的条件下,以期望样本量最小为目的,找到了最优的两阶段设计. 研究发现:在新的假设检验式(2)下,得到的最优化两阶段设计中 $a_{11} = \max(0, a_r + a_t - n_1)$ 和 $c_{11} = \max(0, c_r + c_t - n_1 - n_2)$. 这是由于 $\bar{\boldsymbol{p}}_{0r}^*$ 和 $\bar{\boldsymbol{p}}_{0t}^*$ 两点的特殊性以及功效函数关于 a_{11}, c_{11} 分别不增导致的.

4）由于式(2)中 H_A 是式(1)中 H_1 的子集,因此式(1)的假设检验得到的最优化两阶段设计也是满足式(2)的两阶段设计. 虽然对式(1)的最优设计不一定对式(2)仍然是最优,但其计算相对简单.

参考文献:

[1] SIMON R. Optimal two-stage designs for phase Ⅱ clinical trials[J]. Controlled Clinical Trials, 1989, 10(1): 1-10.

[2] BANERJEE A, TSIATIS A A. Adaptive two-stage designs in phase Ⅱ clinical trials[J]. Statistics in Medicine, 2006, 25(19): 3382-3395.

[3] WU C, LIU A. An adaptive approach for bivariate phase Ⅱ clinical trial designs[J]. Contemporary Clinical Trials, 2007, 28(4): 482-486.

[4] SHAN G, WILDING G E, HUTSON A D, et al. Optimal adaptive two-stage designs for early phase Ⅱ clinical trials[J]. Statistics in Medicine, 2016, 35(8): 1257-1266.

[5] WANG W, YIN H, ZHANG Z. Exact tests using binary data in adaptive two or multi-stage designs[J]. Submitted, 2019.

[6] WANG W, YIN H, ZHANG Z. On construction of the smallest one-sided confidence interval for response rate in two or multi-stage designs[J]. Submitted, 2019.

[7] CONAWAY M R, PETRONI G R. Bivariate sequential designs for phase Ⅱ trials[J]. Biometrics, 1995, 51: 656-664.

[8] BRYANT J, DAY R. Incorporating toxicity considerations into the design of two-stage phase Ⅱ clinical trials[J]. Biometrics, 1995, 51: 1372-1383.

[9] TOURNOUX C, DE RYCKE Y, MEDIONI J, et al. Methods of joint evaluation of efficacy and toxicity in phase Ⅱ clinical trials[J]. Contemporary Clinical Trials, 2007, 28(4): 514-524.

[10] JIN H. Alternative designs of phase Ⅱ trials considering response and toxicity[J]. Contemporary Clinical Trials, 2007, 28(4): 525-531.

[11] 金华. 一种新的两阶段疫苗试验设计[J]. 华南师范大学学报:自然科学版, 2008(3): 1-4.
JIN H. An alternative two-stage design of vaccine trials[J]. Journal of South China Normal University: Natural Science Edition, 2008(3): 1-4. (in Chinese)

[12] RAY H E, RAI S N. An evaluation of a Simon 2-Stage phase Ⅱ clinical trial design incorporating toxicity monitoring[J]. Contemporary Clinical Trials, 2011, 32(3): 428-436.

[13] KHATTREE R, NAIK D. Computational Methods in Biomedical Research[M]. CRC Press, 2007.

[14] THALL P F, SIMON R M, ESTEY E H. Bayesian sequential monitoring designs for single-arm clinical trials with multiple outcomes[J]. Statistics in Medicine, 1995, 14(4): 357-379.

[15] THALL P F, SUNG H G. Some extensions and applications of a Bayesian strategy for monitoring multiple outcomes in clinical trials[J]. Statistics in Medicine, 1998, 17(14): 1563-1580.

[16] STALLARD N, THALL P F, WHITEHEAD J. Decision theoretic designs for phase Ⅱ clinical trials with multiple outcomes[J]. Biometrics, 1999, 55(3): 971-977.

[17] YIN H, WANG W, ZHANG Z. On construction of two-stage designs with consideration of both response and toxicity[J]. Biometrical Journal, 2019, DOI: 10. 1002/bimj. 201800320.

[18] YU J, KEPNER J L, BUNDY B N. Exact power calculations for detecting hypotheses involving two correlated binary outcomes[J]. Statistics & Probability Letters, 2007, 77(3): 288-294.

(责任编辑 杨开英)