
两阶段交叉设计率比较检验方法的模拟研究

小组成员分工

陈奕昕：模拟程序编写整合，报告的撰写与整合，PPT 部分内容的撰写整合

钟端端：模拟程序顺序效应数据生成，PPT 部分内容的撰写

吴育雄：模拟程序实操，报告撰写，PPT 部分内容的撰写

【摘要】

目的：在两阶段交叉设计中，针对疗效为“有效”“无效”二分类变量的资料，比较 McNemar 检验和卡方检验在不同条件下的 I 类错误和检验效能，为资料分析方法的选取提供参考。**方法：**运用 Monte Carlo 模拟产生数据，采用 R 软件进行数据模拟及统计分析。生成 A、B 药物有效率，再构造两种检验统计量循环检验，通过判断统计量与临界值关系计算 I 型错误和检验效能。**结果：**无顺序效应时，McNemar 检验 I 类错误率随样本量增加而上升且高于卡方检验，小率差时其检验效能略高，大率差时两者接近；存在顺序效应时，不同情况对两种检验方法的 I 类错误和检验效能影响各异，如 case2、5 使差异减小，case3、4 使差异增大，case6 使率差更大且检验效能显著提高但误判风险增加。**结论：**无顺序效应时，小样本或严格控制 I 类错误选卡方检验，大样本且注重检验效能可考虑 McNemar 检验；有顺序效应时，根据具体情况选择更合适的检验方法。本研究模拟设计结果可为相关研究提供思路与参考。

【关键词】：两阶段交叉设计，McNemar 检验，卡方检验，Monte Carlo 模拟，I 类错误，检验效能

【ABSTRACT】

Objective: To compare the Type I error rates and statistical power of the McNemar test and the Chi-square test for analyzing binary efficacy outcomes in a two-stage crossover design, providing guidance for selecting appropriate statistical methods. **Methods:** Monte Carlo simulations were used to generate data, and R software was employed for data simulation and statistical analysis. The effectiveness rates of drugs A and B were generated, and the test statistics for both the McNemar test and the Chi-square test were constructed and repeatedly evaluated. Type I error

rates and statistical power were calculated based on the relationship between the test statistics and their critical values. **Results:** When there was no carryover effect, the Type I error rate of the McNemar test increased with sample size and was higher than that of the Chi-square test. For small effect sizes, the McNemar test had slightly higher power, while for large effect sizes, the two tests had similar power. When carryover effects were present, different scenarios affected the Type I error rates and power of the two tests differently. For example, cases 2 and 5 reduced the differences, cases 3 and 4 increased the differences, and case 6 resulted in larger effect sizes with significantly higher power but also increased the risk of false positives. **Conclusion:** In the absence of carryover effects, the Chi-square test is recommended for small samples or when strict control of Type I error is required, while the McNemar test can be considered for large samples where power is a priority. When carryover effects are present, the choice of test should be based on the specific scenario. The results of this simulation study provide valuable insights and references for related research.

【Key Words】 : Two-stage crossover design, McNemar test, Chi-square test, Monte Carlo simulation, Type I error, Statistical power

背景

交叉设计是在受试者内部比较两个或两个以上处理的一种实验设计方法，这种设计通常用于药物试验、心理学研究和其他领域。而两阶段交叉设计则是在两个不同的阶段中对受试者进行测试。它按照事先设计好的实验次序，在两个时期对受试对象先后实施各种处理，以比较组间差异。对于目前二阶段交叉设计的资料常采用 t 检验、方差分析来比较试验中定量的效应指标。但在某些情况下，实验人员往往难以获得准确的定量指标，如临床试验中某些镇痛药虽然可以确认对患者有效，但想要量化这种效果是不太现实的，并且每个实验对象对疼痛的耐受程度不一样，很难确定一个统一的标准，因此，本文讨论的疗效不再是定量变量，而是“有效”、“无效”的二分类变量。

理论介绍

针对二阶段交叉设计的分类资料，使用 2 种四格表的资料整理形式，选取了最常用的 McNemar 检验和卡方检验进行统计分析，并通过统计模拟为资料分析方法的选取提供参考^[1]。

采用完全随机设计来安排随机设计对象。设有 A、B 两种处理因素，有 I、II 两个阶段，将 N 个实验对象随机等分成两组，一组第 I 阶段给 A 因素、第 II 阶段给 B 因素，另一组正好相反，二分类资料交叉设计整理后的结果如下表 1 所示：

表 1 交叉设计结果表

AB 顺序			BA 顺序		
对 A 应答	对 B 应答		对 A 应答	对 B 应答	
	是	否		是	否
是	n ₁₁	n ₁₂	是	m ₁₁	m ₁₂
否	n ₂₁	n ₂₂	否	m ₂₁	m ₂₂

(一) McNemar 检验

计数资料的配对设计常用于两种检验方法、培养方法等的比较，对于两阶段交叉设计，忽略顺序，将观察单位分为用两种方法处理的两组，并整理观察结果的两分类变量的计数结果。观察结果有 4 种情况。

对 A 应答	对 B 应答	
	是	否
是	a	b
否	c	d

其中：

a：表示两种顺序下对 AB 治疗方法均应答的个体数， $a = n_{11} + m_{11}$

b：表示两种顺序下对 A 应答，对 B 不应答的个体数， $b = n_{12} + m_{12}$

c：表示两种顺序下对 A 不应答，对 B 应答的个体数， $c = n_{21} + m_{21}$

d：表示两种顺序下对 AB 治疗方法均不应答的个体数， $d = n_{22} + m_{22}$

McNemar 检验是用于配对样本数据的二分类数据检验。它的基本思想是通过配对差异判断两个处理效果之间的显著性。对于二阶段交叉设计中的 AB 和 BA 两种顺序，将 AB 和 BA 的四格表合并为一个四格表，目的是检查个体在两个阶段的响应变化是否存在显著差异。适用于配对设计数据，尤其是在交叉设计中，通过检验是否存在交叉阶段的显著效果差异，可以排除个体间的差异影响。该方法对配对数据敏感，能有效分析同一个体在两个处理条件下的响应差异，也是交叉设计普遍的检验方法。

检验的原假设及各择假设分别为：

$H_0: b=c$ ，即两种治疗方法结果之间没有显著差异。

$H_1: b \neq c$ ，即两种治疗方法疗效之间有显著差异。

检验统计量为：

$$\chi^2 = \frac{(b - c)^2}{b + c} \quad (1)$$

构造出来的统计量服从自由度为 1 的卡方分布；

检验计算：使用卡方分布（自由度=1）来计算 p 值。如果该统计量的值大于临界值或 p 值小于显著性水平（如 0.05），则拒绝零假设，认为两次测量之间存在显著差异。

(二) χ^2 检验:

考虑 AB 及 BA 顺序, 将数据整理为两独立样本的四格表如表 2, 比较 AB 顺序和 BA 顺序下第一种治疗更优的发生率, 使用两组独立样本构建统计量进行卡方检验。

检验的原假设及备择假设分别为:

H_0 : 两种顺序下两组第一种治疗更优的比例相同

H_1 : 两种顺序下两组第一种治疗更优的比例不同

基本公式为:

$$\chi^2 = \sum \frac{(A - T)^2}{T} \quad (2)$$

其中, AB 顺序下第一种治疗更优即为对 A 应答对 B 不应答, BA 顺序下第一种治疗更优即为对 B 应答对 A 不应答, 无效假设是两组应答率没有差别, 即 AB 顺序下 A 疗效更优的比例和 BA 顺序下 B 疗效更优的比例相同, 均等于两组合计第一种疗效更优的比例。将 AB 顺序和 BA 顺序下的数据视为独立样本, 此时可以间接比较 A 疗效和 B 疗效应答率的区别。

表 2 两独立样本率比较的四格表

顺序	First Treatment	Second Treatment	合计
	Better	Better	
AB	n_{12}	n_{21}	$n=n_{12}+n_{21}$
BA	m_{21}	m_{12}	$m=m_{21}+m_{12}$
合计	$n_{12} + m_{21}$	$n_{21} + m_{12}$	total

其中:

n_{12} : A-B 组中 A 药有效而 B 药无效的人数;

n_{21} : A-B 组中 A 药无效而 B 药有效的人数;

m_{12} : B-A 组中 A 药有效而 B 药无效的人数;

m_{21} : B-A 组中 A 药无效而 B 药有效的人数;

total: $n_{12}+n_{21}+m_{21}+m_{12}$;

通过检验第一治疗顺序是否比第二治疗顺序更好 (即比较不同顺序下的总体有效率差异), 来判断两处理组之间的效果是否存在显著差异。换句话说, 其基于独立性假设, 认为每个单元格的值是来自独立个体的观测结果。通过比

较观察频数与期望频数的差异，计算统计量并根据卡方分布检验组间差异。所以更加适用于独立样本之间的差异检验，将 AB 和 BA 分组视为独立组进行有效率差异的检验，从而能够判断总体有效率是否存在显著差异；

对于一般卡方检验，根据上表数据计算卡方统计量如下，此时自由度为 1

$$\chi^2 = \frac{(n_{12} \times m_{12} - n_{21} \times m_{21})^2 \times total}{(n_{12} + n_{21}) \times (m_{21} + m_{11}) \times (n_{12} + m_{21}) \times (n_{21} + m_{12})} \quad (3)$$

计数资料中的实际频数 A 为分类资料，是不连续的，按照公式（2）计算的卡方值是离散型分布，此时查界值表所得的概率 P 偏小，故需要进行校正。

当有理论频数 $1 \leq T \leq 5$ 时，对上式作连续性校正，公式如下：

$$\chi^2 = \frac{(|n_{12} \times m_{12} - n_{21} \times m_{21}| - total/2)^2 \times total}{(n_{12} + n_{21}) \times (m_{21} + m_{11}) \times (n_{12} + m_{21}) \times (n_{21} + m_{12})} \quad (4)$$

检验计算：使用卡方分布（自由度=1）来计算 p 值。如果该统计量的值大于临界值或 p 值小于显著性水平（如 0.05），则拒绝零假设，认为两次测量之间存在显著差异。

(三) 顺序效应

顺序效应是指在实验或测量过程中，由于处理顺序的不同对结果产生的影响。这种效应在两阶段交叉实验设计中常见，尤其是在重复测量或配对设计中。顺序效应的实质是遗留效应（carryover effect），即上一阶段的处理由于效应去除时间不足或其他原因干扰了下一阶段的处理效果^[3]。其可以看作是一种混杂因素，使得原本 A 的疗效或者 B 疗效数据不再准确。如果 A、B 两种药物均对下一阶段无影响或影响相同，则 AB 与 BA 两种顺序的效应相等或相近，此时该效应混在阶段效应中，可以忽略它的影响；如果 A、B 之间的影响不同，则 AB 和 BA 两种顺序的效应不相同，此时就不能忽略其影响，且表现在第二阶段的处理效果中。

对于存在顺序效应的情况下，我们讨论设定顺序效应的以下几种情况：

1. A 药对 B 药有抑制作用，而 B 药对 A 药无影响。此时在 AB 顺序下，A 药的残留抑制 B 药物的疗效，B 药的有效率降低，而在 BA 顺序下 A 药有效率不变，顺序效应的权重设为 $(-1, 0)$ ；

-
2. A 药对 B 药有促进作用，而 B 药对 A 药无影响。此时在 AB 顺序下，A 药的残留抑制 B 药物的疗效，B 药的有效率提升，而在 BA 顺序下 A 药有效率不变；顺序效应的权重设为 $(1, 0)$
 3. A 药 B 药互相促进，但是 A 药物对 B 药物的影响程度更大。此时 AB 顺序和 BA 顺序下第二种药物的有效率均会增加且 B 药物有效率增加的程度更大，顺序效应的权重设为 $(2, 1)$ ；
 4. A 药 B 药互相抑制，但是 A 药物对 B 药物的影响程度更大。相似的，此时顺序效应的权重设为 $(-2, -1)$ ；
 5. A 药对 B 药有促进作用，而 B 药对 A 药有抑制作用，此时 AB 顺序下 B 药的有效率提高，BA 顺序下 A 药的有效率降低，此时顺序效应的权重为 $(1, -1)$

由于其他情况的 A 药 B 药顺序不同的情况分析效果是重复的，这里不再重复讨论；根据上面五种存在顺序效应时，药物的影响情况，需要产生不同的有效率的数据，为了更加便捷地表现顺序效应的影响（有效率的改变），我们设定了偏移因子 δ ，即顺序效应的影响的权重大小，通过向量设定不同的影响程度，在与偏移因子相乘结合循环，更高效地生成了上述五种不同顺序效应的情况下 A 药和 B 药的有效率数据；

本文对于其出现顺序效应的原因不作重点讨论，只对顺序效应出现时，顺序效应产生的对于药物原本疗效（有效率）的偏倚（误差），使用上述的两种检验方法能否检验出药物原本的疗效是否有不同；其中最重要的是能否克服顺序效应所带来的影响。

模拟方法

采用 Monte Carlo 模拟比较 McNemar 检验和卡方检验对于同 1 组二阶段交叉设计资料的 I 型错误和 power。以二分类交叉试验设计资料为例，假设数据服从 $B(n, \pi)$ 的独立二项分布，样本量考虑 40,60,80,100,120,150,200 七种情况， π_1, π_2 参数见表 3，顺序效应权重参数见表 4，偏移因子 δ 设为 0.02，共考虑 154 种参数组合，每种情况模拟 10000 次，假设检验的检验水准设为双侧 0.05，置信水平设为 95%。数据模拟及统计分析采用 R4.4.1 编程实现。

表 3 两组发生率参数表

分组		应答率									
π_1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
π_2	0.4	0.425	0.45	0.475	0.5	0.525	0.55	0.575	0.6	0.625	0.65

表 4 顺序效应权重参数表

分组	顺序效应权重					
AB 顺序	0	-1	1	2	-2	1
BA 顺序	0	0	0	1	-1	-1

模拟的思路主要分为产生数据和构造检验统计量进行重复循环检验比较两大部分；

第一部分是模拟产生数据，先设定好种子数（2024），使用 R 语言中的 `rbinom` 函数产生 A 药物和 B 药物的有效率；我们设定了对于 A、B 药物相同有效率（I 类错误），A、B 药物不同有效率的几种情况，在不存在顺序效应前，A、B 药的有效率不会相互有影响，按原本设定的发生率（有效率）设定即可，在每一次的循环检验之前生成；

第二部分是构造检验统计量进行循环，对于生成的数据，都要分别进行 McNemar 检验和卡方检验；根据上面理论推导的公式以及对应的数据，分别构造出 McNemar 的检验统计量 `mcn` 和卡方检验的统计量 `kaf`；在每次循环中判断统计量 `mcn` 和 `kaf` 的绝对值与 3.84 的大小关系（卡方分布的第 97.50 百分位数 $\alpha=0.05$ ，双侧检验），并进行计数：当 `mcn` 绝对值大于 3.84 时，McNemar 法计数加 1；当 `kaf` 绝对值大于 3.84 时，卡方检验计数加 1。将得到的两个检验方法的计数值分别除以循环次数 `n`，计算得到 I 型错误和检验效能，并进行输出；

结果

(一) 不存在顺序效应时

1. 不同样本量下两种方法的 I 类错误的模拟比较分析：

表 5 无顺序效应时不同样本量 I 类错误模拟结果

n	Macnemar-error-I	Chiqs-error-I
40	0.0315	0.0217

60	0.0374	0.0292
80	0.0394	0.0301
100	0.0354	0.0283
120	0.0428	0.035
150	0.0374	0.0305
200	0.0469	0.0404

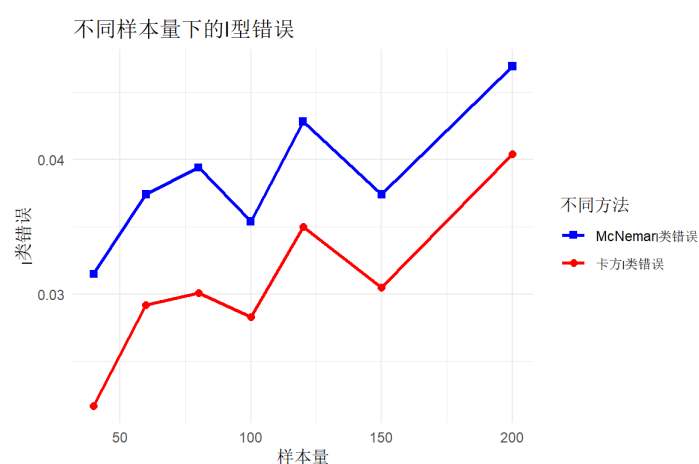


图 1 无顺序效应时不同样本量下的 I 类错误

在上述 I 类错误的模拟中，我们设置两组发生率（有效率）均为 0.4（率差=0），设置了 40 到 200，步长为逐渐变大的不同样本量，模拟重复 10000 次进行比较：结果显示 McNemar 方法的 I 类错误率随着样本量的增加而增加，从 40 样本量的 0.0315 增加到 200 样本量的 0.0469。Chisq 方法的 I 类错误率也随着样本量的增加而增加，从 40 样本量的 0.0217 增加到 200 样本量的 0.0404。

由于从 40 至 200 的模拟中，没有明显的趋势说明两种方法的 I 类错误稳定在 0.05 附近，因此增加样本量进行模拟。McNemar 检验和 χ^2 检验的 I 型错误随样本量变化如图 2 所示，随着样本量的增加，两种方法 I 类错误都稳定在 0.05 左右。

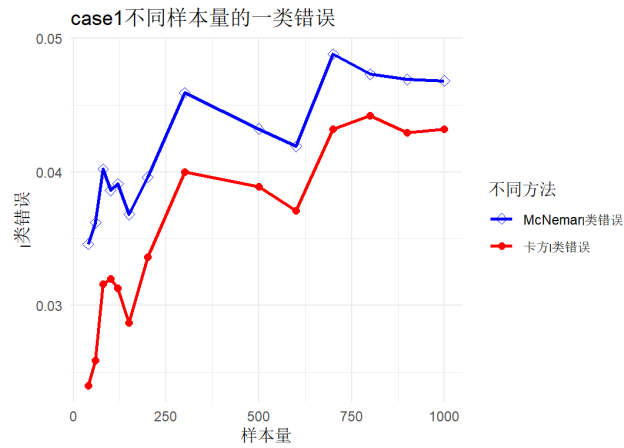


图 2 大样本不同样本量下的 I 类错误

与 McNemar 方法相比，Chisq 方法在所有样本量下的 I 类错误率都较低。两种方法的 I 类错误率都能控制在 0.05 以下，这意味着它们都成功地控制了在原假设为真时错误拒绝它的概率。Chisq 方法在所有样本量下都显示出比 McNemar 方法更低的 I 类错误率，这表明 Chisq 方法可能更为保守，更不容易导致错误的拒绝原假设，也可能有着更低的检验效能。

2. 固定样本量为 80 设置不同率差的检验效能模拟：

表 6 无顺序效应时不同率差的检验效能模拟结果

率差	Macnemar-power	Chiqs-power
0.000	0.0394	0.0301
0.025	0.0537	0.0451
0.050	0.1198	0.1013
0.075	0.2334	0.2026
0.100	0.3864	0.3500
0.125	0.5618	0.5231
0.150	0.7286	0.6959
0.175	0.8497	0.8244
0.200	0.9301	0.9171
0.225	0.9749	0.9677
0.250	0.9949	0.9924

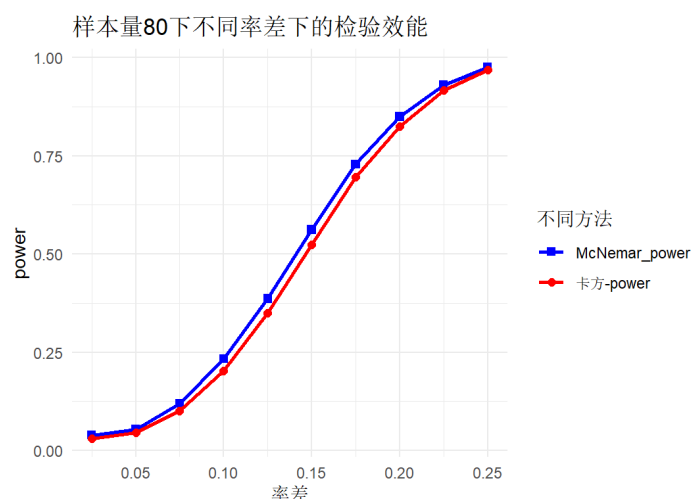


图 3 样本量 80 不同率差的 power

对于不存在顺序效应的情况下，我们设置了样本量为 80 的，两组间的不同率差（A 药 B 药有效率之差）从 0.025 到 0.250，步长为 0.025，不同情况，模拟重复 10000 次进行分析比较：

结果显示，随着率差的增加，两种方法的检验效能都在提高，率差越大，越容易检测到显著差异，所以检验效能越高；在率差为 0.25 时，两种方法的检验效能都基本接近 1，说明此时两种方法都能非常有效地检测出差异。两种方法的比较之下，McNemar 检验在小率差时表现出比卡方检验略高的检验效能，说明在小率差的情况下 McNemar 检验可能更为敏感，卡方检验的效能在大率差时（例如 0.15 到 0.25）逐渐接近 McNemar 检验的效能，两者的效能差异逐渐减小，最后在率差为 0.25 时，两种方法几乎具有相同的效能。McNemar 检验在较小率差时表现稍优，而在较大率差时，两者的表现接近。对于小率差的情况，McNemar 检验的效能略高，可能更适合用于检测较小的差异；而在率差较大时，卡方检验的效能也足够，几乎与 McNemar 检验一致。总的来说，McNemar 检验更加敏感一些，检验效能也稍高，卡方则是更加保守一些检验效能稍低，这与前面的 I 类错误的模拟结果相一致；

3. 固定率差为 0.1 不同样本量的模拟检验效能：

表 7 无顺序效应时固定率差不同样本量下检验效能模拟结果

n	Macnemar-power	Chisq-power
40	0.1958	0.1580

60	0.2942	0.2529
80	0.3864	0.3500
100	0.4835	0.4478
120	0.5564	0.5207
150	0.6502	0.6207
200	0.7892	0.7678

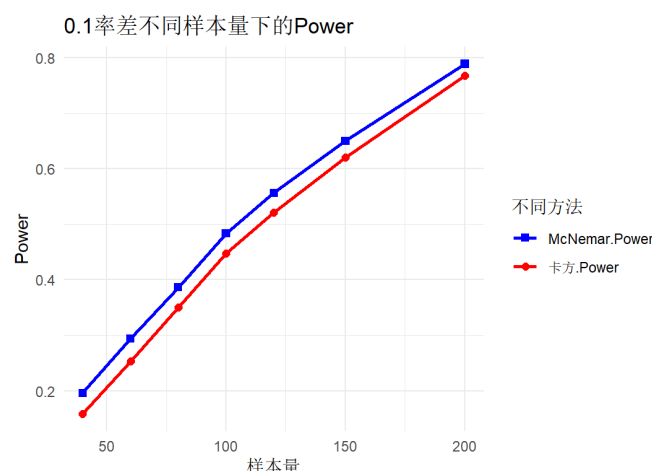


图 4 无顺序效应下固定率差率差不同样本量 power

从图表中可以看出，无论是 McNemar 还是卡方检验，随着样本量的增加，它们的检验效能都在逐渐提高。更大的样本量通常能提供更可靠的结果，从而增加检验的效能。在相同的样本量下，McNemar 的检验效能总是略高于卡方检验。即使在样本量为 200 时，差距仍然存在，McNemar 的检验效能（0.7892）仍稍高于卡方检验的检验（0.7678）。这表明在相同的条件下，McNemar 方法可能更敏感，能够更有效地检测出真实的差异，而卡方较为保守。这种情况下固定了率差为 0.1，对于两阶段交叉设计来说，0.1 的率差并非较大的率差，所以检验效能普遍不是非常高，即使在最大的样本量 200 的情况下，检验效能也只有不到 0.8。

(二) 存在顺序效应时

存在顺序效应的情况下，我们设定了若干种情况，根据前面模拟数据思路的分类，我们将情况分为以下几种：

Case1：无顺序效应；

Case2：A 药对 B 药有抑制作用，而 B 药对 A 药无影响；

- Case3: A 药对 B 药有促进作用, 而 B 药对 A 药无影响;
- Case4: A 药 B 药互相促进, 但是 A 药物对 B 药物的影响程度更大;
- Case5: A 药 B 药互相抑制, 但是 A 药物对 B 药物的影响程度更大;
- Case6: A 药对 B 药有促进作用, 而 B 药对 A 药有抑制作用;

4. 存在顺序效应下的一类错误模拟:

表 8 存在顺序效应时不同样本量 McNemar I 类错误模拟结果表

n	Case1	Case2	Case3	Case4	Case5	Case6
40	0.0315	0.0371	0.0320	0.0321	0.0351	0.0378
60	0.0374	0.0361	0.0378	0.0385	0.0400	0.0480
80	0.0394	0.0422	0.0404	0.0395	0.0385	0.0479
100	0.0354	0.0446	0.0405	0.0425	0.0444	0.0541
120	0.0428	0.0427	0.0468	0.0478	0.0414	0.0626
150	0.0374	0.0500	0.0490	0.0495	0.0446	0.0634
200	0.0469	0.0482	0.0498	0.0507	0.0471	0.0753

表 9 存在顺序效应时不同样本量卡方 I 类错误模拟结果表

n	Case1	Case2	Case3	Case4	Case5	Case6
40	0.0217	0.0264	0.0211	0.0216	0.0257	0.0281
60	0.0292	0.0266	0.0267	0.0279	0.0288	0.0366
80	0.0301	0.0334	0.0325	0.0319	0.0294	0.0378
100	0.0283	0.0362	0.0332	0.0342	0.0353	0.0464
120	0.0350	0.0351	0.0388	0.0388	0.0350	0.0523
150	0.0305	0.0425	0.0419	0.0425	0.0374	0.0533
200	0.0404	0.0419	0.0438	0.0434	0.0410	0.0649

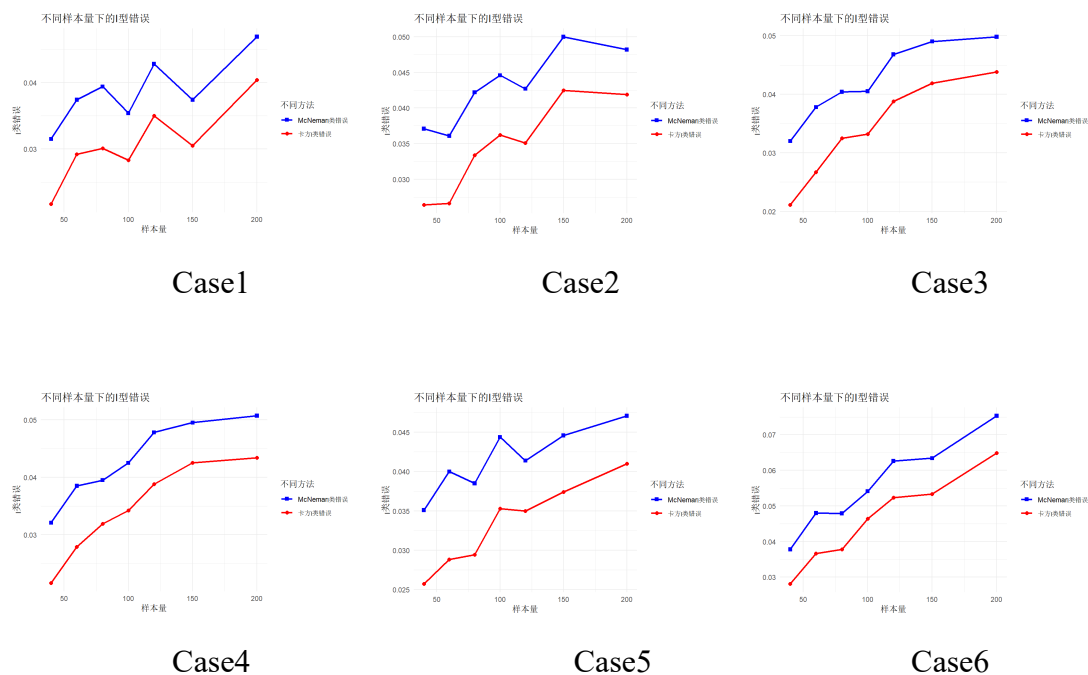


图 5 case1~6 下两方法的 I 类错误随样本量的变化

case1 上面已经分析过，这里不再赘述；case2 和 case3 属于同一种数据情况，即是 A 药对 B 药有促进或抑制的残留作用，使得原本相同的有效率发生了偏移（case2 抑制，case3 促进），增加或减少了有效率，使得原本有效率相同的两组出现了小率差，从而增加了两组间的差异，使得一类错误稍微升高，对于在 case2、3 时候的两种检验方法，一类错误均能保持在 0.05 附近，其中 McNemar I 类错误稍高于卡方；对于 case4、5，其实与 case2、3 相似，即是都有药物残留影响均提高有效率或是降低有效率，其中一方提高的或降低的程度更多，这种情况下出现的两组差异程度与上面相似，也是使得原本有效率相同的两组出现了小率差，从而增加了两组间的差异；此情况下的模拟结果和 case2、3 相似；对于 case6，两种药物残留互相影响，一种是促进效果，一种是抑制效果，这种情况下顺序效应会使得两组间差异会变得更大，一类错误也会上升（McNemar I 类错误 0.0753；卡方 I 类错误 0.0649），可能会出现错误拒绝 H_0 的情况。

5. 样本量固定 80 为时，不同率差下的检验效能：

表 10 存在顺序效应时固定样本量不同率差 McNemar 检验效能

率差	Case1	Case2	Case3	Case4	Case5	Case6
----	-------	-------	-------	-------	-------	-------

0.000	0.0394	0.0422	0.0404	0.0395	0.0385	0.0479
0.025	0.0537	0.0444	0.0817	0.0786	0.0426	0.1035
0.050	0.1198	0.0846	0.1591	0.1551	0.0875	0.2039
0.075	0.2334	0.1779	0.2872	0.2879	0.1857	0.3699
0.100	0.3864	0.3188	0.4619	0.4480	0.3157	0.5342
0.125	0.5618	0.4980	0.6285	0.6270	0.5020	0.6963
0.150	0.7286	0.6615	0.7864	0.7743	0.6682	0.8361
0.175	0.8497	0.8098	0.8956	0.8951	0.8113	0.9223
0.200	0.9301	0.9135	0.9573	0.9527	0.9117	0.9700
0.225	0.9749	0.9637	0.9846	0.9840	0.9646	0.9890
0.250	0.9949	0.9884	0.9943	0.9964	0.9878	0.9974

表 11 存在顺序效应时固定样本量不同率差 Chisq-power 检验效能

率差	Case1	Case2	Case3	Case4	Case5	Case6
0.000	0.0301	0.0334	0.0325	0.0319	0.0294	0.0378
0.025	0.0451	0.0344	0.0672	0.0628	0.0331	0.0868
0.050	0.1013	0.0701	0.1367	0.1329	0.0721	0.1762
0.075	0.2026	0.1515	0.2535	0.2542	0.1571	0.3350
0.100	0.3500	0.2841	0.4161	0.4060	0.2756	0.4923
0.125	0.5231	0.4578	0.5895	0.5847	0.4601	0.6582
0.150	0.6959	0.6266	0.7523	0.7422	0.6266	0.8077
0.175	0.8244	0.7790	0.8741	0.8724	0.7822	0.9061
0.200	0.9171	0.8942	0.9432	0.9402	0.8916	0.9616
0.225	0.9677	0.9532	0.9800	0.9801	0.9544	0.9860
0.650	0.9924	0.9842	0.9927	0.9942	0.9838	0.9964

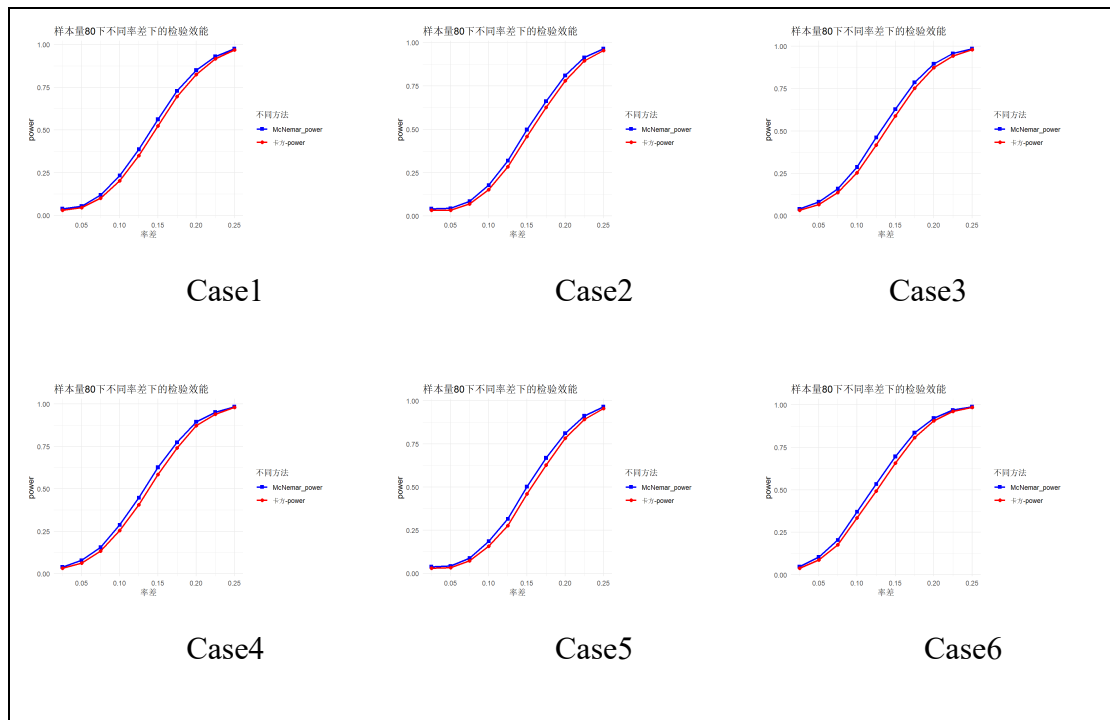


图 6 样本量 80 下 case1~6 不同样本量的 power

结果分析：对于 case2，顺序效应的影响使得两组间的率差减小，让的两组间的差异减小，使得检验出两组差异的难度增大，所以总体相比于 case1 的检验效能都有小部分的下降，但是检验结果仍然可靠，此情况下可以选择 power 更高一点的 McNemar 检验；对于 case3，情况则于 case2 的结果相反，顺序效应使得两组的差异更大，更容易检验出两组差别，所以 power 的总体情况相比于 case1 是提高一点的，但是此时的检验结果相对来说没有那么可靠，此情况建议选择更保守的 chisq 检验结果；case4、5 所属的情况与 case2、3 相似，不再赘述；对于 case6，则会使得 A 药物的有效率下降，提高药物 B 的有效率，使得原本 A 药有效率（0.4）不足，B 有效率大于原本设定率差进一步加强了两组间差异，使得检验效能比原本更高了，但是检验结果也不如 case1 的可靠，应当要选择保守一点的 chisq；

6. 固定率差为 0.1 不同样本量的检验效能：

表 12 存在顺序效应时固定率差不同样本量 Macnemar 的检验效能

n	Case1	Case2	Case3	Case4	Case5	Case6
40	0.1958	0.1590	0.2178	0.2303	0.1616	0.2688
60	0.2942	0.2464	0.3458	0.3443	0.2484	0.4175
80	0.3864	0.3188	0.4619	0.4480	0.3157	0.5342

100	0.4835	0.3907	0.5583	0.5619	0.3990	0.6389
120	0.5564	0.4682	0.6355	0.6418	0.4898	0.7140
150	0.6502	0.5665	0.7459	0.7476	0.5664	0.8143
200	0.7892	0.7021	0.8632	0.8625	0.7011	0.9128

表 13 存在顺序效应时固定率差不同样本量 Chisq-power 的检验效能

n	Case1	Case2	Case3	Case4	Case5	Case6
40	0.1580	0.1277	0.1782	0.1881	0.1266	0.2201
60	0.2529	0.2059	0.3014	0.2978	0.2109	0.3664
80	0.3500	0.2841	0.4161	0.4060	0.2756	0.4923
100	0.4478	0.3568	0.5217	0.5236	0.3628	0.6001
120	0.5207	0.4356	0.6020	0.6081	0.4492	0.6843
150	0.6207	0.5359	0.7197	0.7188	0.5345	0.7915
200	0.7678	0.6767	0.8467	0.8454	0.6751	0.9003

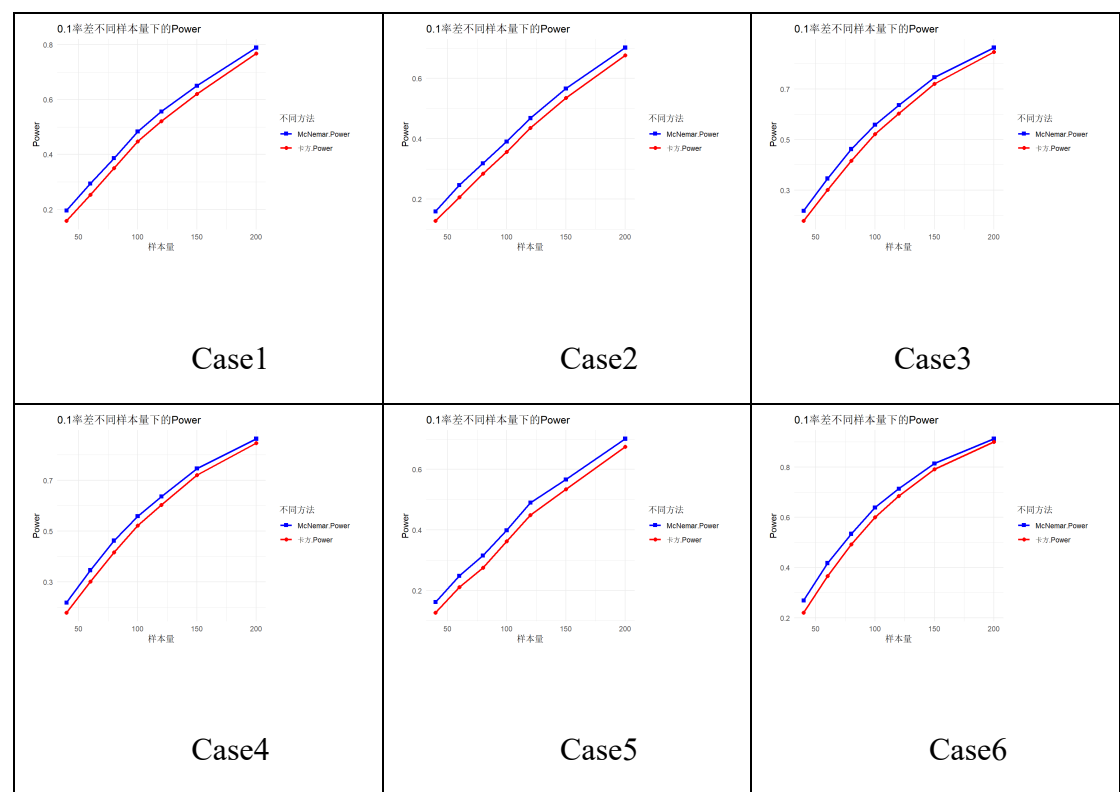


图 7 固定率差下 case1~6 不同样本量的 power

结果分析：此情况选取固定率差（0.1）对于顺序效应对检验效能的影响更为明显，但结果与上一部分的分析基本相同，对于 case2 和 case5，顺序效应

的影响使得两组间的率差减小，让得两组间的差异减小，使得检验出两组差异的难度增大，所以总体相比于 case1 的检验效能都下降，但是检验结果仍然可靠，此情况下可以选择 power 更高一点的 McNemar 检验；对于 case3 与 case4，顺序效应使得两组的差异更大，更容易检验出两组差别，所以 power 的总体情况相比于 case1 是提高得更加明显了，但是此时的检验结果相对来说没有那么可靠，此情况建议选择更保守的 chisq 检验结果；case6 的影响是最大的，该情况下使得两组的率差更大，检验效能较原来 case1 的增大明显，两种检验方法都能达到大于 0.9 的 power，但代价是误判的概率更高了，前面一类错误模拟也印证了这个结果，此种情况下应当选择更保守的检验方法；

讨论

本文通过 Monte Carlo 模拟方法，比较了 McNemar 检验和卡方检验在两阶段交叉设计率比较中的性能，包括 I 类错误和检验效能，下面进行模拟结果的讨论以及总结本次研究的优劣之处。

首先，在无顺序效应时，两个检验方法的一类错误与检验效能模拟结果均较好，符合预期；当样本量较小时，McNemar 检验的 I 类错误率虽在 0.05 以下但相对较高，而卡方检验更为保守，I 类错误率更低。这表明在小样本且无顺序效应时，若更注重控制错误拒绝原假设的风险，卡方检验可能更合适。随着样本量增加，两者 I 类错误率均上升，McNemar 检验的 I 类错误率仍在 0.05 以下但高于卡方检验。此时，若对检验效能要求较高且能接受稍高的 I 类错误风险，McNemar 检验可能更敏感，更能检测出差异；若仍想严格控制 I 类错误，卡方检验更优。在率差较小时，McNemar 检验效能略高于卡方检验，更适合检测小差异；率差较大（如 0.15 - 0.25）时，两者效能接近，卡方检验的保守性使其在大率差时也能满足检测需求。

存在顺序效应时 对于 case2（A 药残留抑制 B 药疗效）和 case5（A 药抑制更多且 B 药也抑制），顺序效应使率差减小、差异难检测，检验效能下降，但结果可靠，此时 McNemar 检验效能相对较高，可优先选择；但若想更保守，卡方检验也可行。case3（A 药残留促进 B 药疗效）和 case4（A 药促进更多且 B 药也促进），顺序效应使差异增大、检验效能提高，但结果可靠性降低。建议选择更保守的卡方检验，不过若更关注检测出差异的能力，McNemar 检验也可

考虑。case6（A 药残留促进 B 药、B 药残留抑制 A 药），率差进一步加大，检验效能显著提高但误判风险大。鉴于一类错误模拟结果，必须选择保守的检验方法以控制错误率。

本文优点

模拟情况较为全面，模拟研究考虑了多种情况，包括不同样本量（40 - 200）、不同率差（0 - 0.25）以及多种顺序效应场景（case2 - case6），全面评估了两种检验方法的性能，为实际应用提供了丰富参考。参数设置合理，利用偏移因子有效模拟顺序效应下药物有效率的变化，合理反映实际情况中的复杂影响。采用 Monte Carlo 模拟方法，结合大量重复试验（10000 次），能较为准确地估计 I 类错误和检验效能，增强了结果的可信度和说服力。

本文不足之处

未能深入探讨顺序效应原因，虽模拟了顺序效应影响，但未深入研究其产生原因（如洗脱期等因素），限制了对问题全面理解和应对策略的制定。未考虑其他因素影响，实际应用中可能存在更多影响因素（如个体差异的其他方面、不同疾病类型等），本模拟未涉及，可能导致结果在某些复杂实际场景中的适用性受限。数据模拟有一定的局限性，假设数据服从特定分布，实际数据可能不符合，影响模拟结果对真实情况的反映准确性。且未考虑缺失数据等情况；

总的来说，本模拟研究为两阶段交叉设计率比较检验方法的选择提供了新的想法与思路，但仍可进一步改进和完善。

参考文献

- [1] 主编颜艳, 王彤.医学统计学.人民卫生出版社.2020
- [2] 杜春霖,李晓松,邱建青,等.率比较的二阶段交叉设计及模拟研究[J].现代预防医学, 2018, 45(2):4.DOI:CNKI:SUN:XDYF.0.2018-02-001.
- [3] 李文彤,李晓松.交叉设计中顺序效应的影响及其分析方法[J].数理医药学杂志, 1999, 12(4):2.DOI:10.3969/j.issn.1004-4337.1999.04.008.