

## Tests for equivalence/non-inferiority based on odds ratio in matched-pair design

고혜정<sup>1)</sup> 이재원<sup>2)</sup>

### 요 약

본 논문에서는 matched-pair design에서의 두 처리간 동등성/비열등성 검정에 대해 고려하였다. 기존에 비율차이나 risk ratio 관점에서 동등성/비열등성 검정을 시행한 것과는 달리, 본 논문에서는 odds ratio에 기초하여 두 가지 검정통계량을 유도하였다. (① constrained maximum likelihood estimator(mle)를 이용한 fieller type 통계량 ② 제약없이 구한 mle를 사용한 wald-type 통계량). 비율 차이나 risk ratio에 기초한 기존의 통계적 방법들(비율차이에 근거한 ③ score-type 통계량과 ④ wald-type 통계량, risk ratio에 기초한 ⑤ fieller-type 통계량과 ⑥ wald-type 통계량)과 본 논문에서 제시한 두가지 통계량의 성능을 비교하기 위해 모의실험을 시행하였다. 모의실험 결과, 본 논문에서 제안한 constrained mle를 사용한 fieller type 통계량은 empirical type I error 측면에서 매우 만족스러운 결과를 보이고 있으며, 특히 비대각 셀의 확률이 작아질 경우에도 안정적인 성능을 보여준다.

주요용어 : matched-pair design, equivalence test, odds ratio, fieller-type statistic, constrained maximum likelihood estimation

### 1. Introduction

새로운 치료약의 효과를 측정하기 위한 기존의 검정방법은, 새로운 치료약이 기존의 치료약에 비해 효과측면에서 우수함을 보이는 것이었다. 그러나 만약 새로운 치료약이 기존의 치료약에 비해 비용이 덜 든다거나, 복용하기 편리하다거나, 사용하기에 안전하다면, 반드시 효과의 우수성을 보일 필요는 없다. 그런 장점을 가지고 있는 약의 경우, 새로운 약의 효과가 기존 약 효과의 일정 수준 내에 있다면 새로운 치료약은 기존의 치료약과 ,효과의 측면에서, 임상적으로 동등하다고 할 수 있다. (Therapeutic equivalence)

Matched-pair design에서 비율 차이를 이용한 동등성검정 방법은 Lu and Bean, Nam, Tango 등에 의해 제시되었으며, risk ratio를 이용한 방법은 Lachenbruch and Lynch, Nam, Man-Lai Tang 등에 의해 소개되었다. crossover design에서 odds ratio를 이용한 동등성 검정 방법이 Man-Lai Tang 에 의해 제안되었지만, 이는 제약없이 구한 mle를 사용한 검정통계량이다.

본 논문에서는 위에서 소개한 바와 같이, matched-pair design에서의 odds ratio에 초점을 두고, odds ratio에 특정 값을 가정하여 추정된 mle를 사용하여 검정통계량을 유도하였다. 그리고 통계량의 성능을 비교하기 위해 odds ratio에 기초한 wald-type 통계량 외에, 비율차이에 기초한 통계량(score-type and wald-type tests), risk ratio에 기초한 통계량(fieller-type and wald-type tests)들을 소개하고 모든 통계량들에 대해 모의실험을 실시하였다. 모의실험을 통해 본 논문에서 제안된 통계량의 empirical type I error가 미리정한 수준에 매우 근사함을 알 수

1) 고려대학교 통계학과 석사 4학기.

2) 고려대학교 통계학과 교수.

있었다.

section 2에서, odds ratio에 기초한 두 가지 통계량을 제시하고, section 3에서는 기존에 제시된 비율차이와 risk ratio에 기초한 통계량들에 대해 소개한다. section 4에서 모의실험결과를 보이고, 마지막으로 section 5에서 결론을 내린다.

## 2. Model & Tests based on odds ratio

짝지어진  $n$ 개의 쌍에 각각 새로운 약과 기존의 치료약을 처리했다고 가정하자. table 1에서 보듯이,  $x_{11}$ 은 두 처리 모두에서 +반응 보인 개체 수이고,  $x_{10}$  새로운 치료약에서 +반응을 얻었으나 기존 치료약에서는 -반응을 보인 개체의 수, 반대로  $x_{01}$ 은 기존의 치료약에서만 +반응을 보인 개체 수, 마지막으로  $x_{00}$ 은 두 처리 모두에서 -반응을 보인 개체의 수이다. 각각의  $x_{ij}$  ( $i, j = 1, 0$ )는 다항분포를 따른다고 가정하고, 각각의 셀 확률은 역시 table 1에 제시되어 있다.  $p_t$ 는 새로운 치료약에 대한 반응율이며,  $p_s$ 는 기존 치료약에 대한 반응율을 나타낸다.

이때 우리의 관심은 이 실험을 통해 얻은 odds ratio가 미리정한  $\phi$  ( $0 < \phi < 1$ : odds ratio 측면에서 두 처리의 동등성을 받아들일 수 있는 한계)에 대해 다음 범위를 만족하는지에 있다.

( $\phi < \text{odds ratio} < 1/\phi$ ) 이를 가설로 표현하면 odds ratio를  $\psi$ 라 할 때,

$$H_0 : \psi \leq \phi \text{ or } \psi \geq 1/\phi \quad H_1 : \phi < \psi < 1/\phi \quad \text{---①}$$

이를 단측가설의 형태로 표현하면, 다음과 같다.

$$H_0 : \psi \leq \phi \quad H_1 : \psi > \phi \quad \text{---②}$$

$$H_0 : \psi \geq 1/\phi \quad H_1 : \psi < 1/\phi \quad \text{---③}$$

가설 ②에 대한 검정이 바로 비열등성 검정을 나타내며, 두 처리의 동등성을 보이기 위해서는 가설 ②,③ 모두 기각되어야만 한다.

$\phi = \frac{p_t(1-p_s)}{p_s(1-p_t)}$  라는 제약 하에서, 확률의 최대우도추정량을 구하기 위해 table 1의 각 셀

확률들을  $p_s, p_{01}$ 로 다음과 같이 표현할 수 있다.

$$p_{11} = p_s - p_{01}, \quad p_{10} = \frac{(\phi - 1)p_s(1 - p_s + p_{01}) + p_{01}}{1 + (\phi - 1)p_s}, \quad p_{00} = \frac{(1 - p_s)(1 - p_{01}) - \phi}{1 + (\phi - 1)p_s}$$

$$\begin{aligned} \ln L = & x_{11} \ln(p_s - p_{01}) + x_{10} \ln((\phi - 1)p_s(1 - p_s + p_{01}) + p_{01}) + x_{01} \ln p_{01} \\ & + x_{00} \ln((1 - p_s)(1 - p_{01}) - \phi) - (x_{10} + x_{00}) \ln(1 + (\phi - 1)p_s) \end{aligned}$$

<table 1> observations and probabilities in a matched-pair design.

test	standard		total	test	standard		total
	+	-			+	-	
+	$x_{11}$	$x_{10}$	$x_{11}+x_{10}$	+	$p_{11}$	$p_{10}$	$p_t$
-	$x_{01}$	$x_{00}$	$x_{01}+x_{00}$	-	$p_{01}$	$p_{00}$	$1-p_t$
total	$x_{11}+x_{01}$	$x_{10}+x_{00}$	n	total	$p_s$	$1-p_s$	1

$p_s, p_{01}$ 으로 표현된 각 확률을 반영하여 위와 같은 우도함수를 구할 수 있고, 이 때  $\phi$ 를 관심 있는 모수,  $p_s, p_{01}$ 를 장애모수라 여기고, 우도함수를  $p_s, p_{01}$ 에 대해 각각 1차 미분함으로써  $p_s, p_{01}$ 에 대한 최대우도추정량  $\tilde{p}_s, \tilde{p}_{01}$ 을 구하게 된다. ( $\partial \ln L / \partial p_s = 0, \partial \ln L / \partial p_{01} = 0$ ) 그러나 1차 미분하여 얻은 식이 매우 복잡하여 해를 폐쇄형으로 표현하지 못하고, 수치해석적 방식으로 해를 구한다.  $s(0) = p_t(1 - p_s) - \phi p_s(1 - p_t)$ 라고 정의하면,  $S(0)$ 의 기대값은 0 이고, 분산의 형태는  $Var[S(0)] = \frac{\phi(p_{10} + p_{01} - (p_{10} - p_{01})^2)}{n}$ 이므로, 가설 ②에 대한 fieller-type 검정통계량은 다음과 같이 표현된다.

$$z_1 = \frac{\hat{p}_t(1 - \hat{p}_s) - \phi \hat{p}_s(1 - \hat{p}_t)}{\sqrt{\phi(\tilde{p}_{10} + \tilde{p}_{01} - (\tilde{p}_{10} - \tilde{p}_{01})^2)/n}}$$

만약  $z_1 \geq z_{\alpha}$ 이면, 유의수준  $\alpha$ 에서 새로운 처리는 기존 처리에 비해 열등하지 않다고 결론 내릴 수 있다. 그리고 가설 ③에 대한 검정통계량도 같은 방식으로 만들어질 수 있으며, 이 두 통계량이 모두 유의해야만, 두 처리의 동등성을 보일 수 있다.

wald-type 통계량은  $\tilde{p}_s, \tilde{p}_{01}$  대신  $\hat{p}_s, \hat{p}_{01}$ 을 사용하여 다음과 같이 유도된다.

$$z_2 = \frac{\hat{p}_t(1 - \hat{p}_s) - \phi \hat{p}_s(1 - \hat{p}_t)}{\sqrt{\phi(\hat{p}_{10} + \hat{p}_{01} - (\hat{p}_{10} - \hat{p}_{01})^2)/n}}$$

### 3. Other tests

비율차이를 이용하여 두 처리의 동등성/비열등성 검정을 하기위한 가설의 형태는,  $\phi_1$ 을 비율 차 검정에서 사용되는 한계값이라 할 때, 다음과 같이 쓰여진다.

$$H_0 : p_t - p_s \leq -\phi_1 \quad vs \quad H_1 : p_t - p_s > -\phi_1 \quad \text{---④}$$

$$H_0 : p_t - p_s \geq \phi_1 \quad vs \quad H_1 : p_t - p_s < \phi_1 \quad \text{---⑤}$$

이 역시 두 가설 모두 기각되어야만 두 처리의 동등성을 증명할 수 있다. 가설 ④를 위한

Tests for equivalence/non-inferiority based on odds ratio in matched-pair design

score-type 검정통계량은 Nam에 의해, 아래와 같이 유도되었으며, 가설 ⑤에 대한 검정통계량도 이와 같은 방식으로 유도된다.

$$z_3 = \frac{\hat{p}_t - \hat{p}_s + \phi_1}{\sqrt{(2\tilde{p}_{01} - \phi_1 - \phi_1^2)/n}}$$

이 때  $\tilde{p}_{01}$ 은 비율차이에 대한 가정 ( $p_t - p_s = -\phi_1$ )하에서 통해 얻어진 mle 값이다. 같은 가설에 대한 wald-type 검정통계량은  $\tilde{p}_{01}$  대신  $\hat{p}_{01}$ 을 사용하여 분산성분을 추정한다. 이 때 사용된  $\tilde{p}_{01}$ 은 앞서 odds ratio에 대한 가정을 한 후 얻은  $\tilde{p}_{01}$ 와는 다른 값이며, wald-type 추정량  $\hat{p}_{01}$ 은 모든 경우에 대해 같은 값을 갖는다.  $\tilde{p}_{01}$ 은 다음의 2차식으로 계산된다.

$$2\tilde{p}_{01}^2 - [(\hat{p}_t - \hat{p}_s)(1 - \phi_1) + 2(\hat{p}_{01} + \phi_1)]\tilde{p}_{01} + \phi_1(1 + \phi_1)\hat{p}_{01} = 0$$

$$z_4 = \frac{\hat{p}_t - \hat{p}_s + \phi_1}{\sqrt{(2\hat{p}_{01} - \phi_1 - \phi_1^2)/n}}$$

risk ratio를 이용한 가설의 형태는,  $\phi_2$  ( $0 < \phi_2 < 1$ )를 risk ratio에 대한 한계값이라고 할 때, 다음과 같다.

$$H_0 : p_t/p_s \leq \phi_2 \quad vs \quad H_1 : p_t/p_s > \phi_2 \quad -⑥$$

$$H_0 : p_t/p_s \geq 1/\phi_2 \quad vs \quad H_1 : p_t/p_s < 1/\phi_2 \quad -⑦$$

여기서도 귀무가설 가정( $p_t/p_s = \phi_2$ ) 하에서 얻은 mle  $\tilde{p}_{10}, \tilde{p}_{01}$ 를 가지고 fieller-type 통계량을 유도하였으며,  $\tilde{p}_{10}$ 는 다음과 같은 2차식을 통해 구할 수 있다.

$$\phi_2(\phi_2 + 1)\tilde{p}_{10}^2 + (\hat{p}_t - \phi_2^2(1 + \hat{p}_{10} - \hat{p}_{00}))\tilde{p}_{10} + \phi_2(\phi_2 - 1)\hat{p}_{10}(1 - \hat{p}_{00}) = 0$$

그리고, wald-type 통계량은  $\tilde{p}_{10}$  대신 아무 가정없이 구한 mle  $\hat{p}_{10}$ 를 사용하여 구한다.

$$z_5 = \frac{\hat{p}_t - \phi_2\hat{p}_s}{\sqrt{\phi_2(\hat{p}_{10} + \hat{p}_{01})/n}}$$

$$z_6 = \frac{\hat{p}_t - \phi_2\hat{p}_s}{\sqrt{\phi_2(\hat{p}_{10} + \hat{p}_{01})/n}}$$

#### 4. Simulation

위에서 제시한 여섯 가지 통계량의 성능을 비교하기 위해 다음과 같은 모의실험을 시행하였

다.  $n=50, 100$ ,  $p_s=0.8, 0.9$ ,  $\phi=0.65, 0.5$ 를 고려하였다. 그리고 비율 차이와 risk ratio를 위한 threshold 값  $\phi_1, \phi_2$ 는  $\phi$ 와  $p_s$ 에 의해 결정된다. Empirical type I error를 계산하기 위해 각각 10,000개의 임의 표본을 추출하였으며, 유의수준은 5%로 정하였다. 결과는 table 2에 제시되어 있다. table 2를 보면, 우선,  $n=50$ 일 때 보다  $n=100$ 인 경우 더 안정적인 결과를 보여주고 있으며, 전반적으로 모든 wald-type 통계량의 empirical type I error가 정해진 유의수준 5%에서 많이 벗어나는 반면(특히 비 대각 셀의 확률  $p_{01}$ 이 작을 경우 벗어나는 정도가 심해진다), score-type이나 constrained mle를 사용한 fieller-type 통계량은 정해진 유의수준 5%에 매우 근접함을 알 수 있었다. score-type 이나 constrained mle를 사용한 fieller-type 통계량의 empirical type I error는 대부분의 모수 setting에서 만족스럽지만, 비대각 셀의 확률이 작아지는 경우에는 비율차이를 이용한 통계량보다 risk ratio 혹은 odds ratio를 이용한 통계량들의 성능이 더욱 안정적이다. 동등성 혹은 비열등성 시험이 일반적으로 비대각 셀의 확률이 크지 않기 때문에, 비율차이 이용한 통계량보다는 risk ratio 혹은 odds ratio를 이용한 통계량이 더 바람직한 성격을 가지고 있다고 여겨진다.

<table 2> Empirical type I error of six tests at nominal 5% level.

$n$	$p_s$	$p_{01}$	$\phi$	$\phi_1$	$\phi_2$	$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	
50	0.8	0.11	0.65	0.078	0.903	4.51	6.82	4.16	7.48	4.28	7.85	
		0.15				4.81	5.92	4.73	5.59	4.77	6.31	
		0.19				4.96	5.62	4.56	5.47	4.81	5.76	
	0.9	0.11	0.5	0.082	0.909	4.81	7.79	5.34	9.13	4.92	9.16	
		0.13				4.99	6.31	4.76	7.38	4.54	7.08	
		0.15				4.86	6.24	4.91	6.78	4.66	6.76	
	100	0.8	0.11	0.65	0.078	0.903	4.67	6.30	4.42	6.22	4.51	6.53
			0.15				4.93	5.38	4.95	5.39	4.76	5.75
			0.19				4.78	4.97	4.85	4.99	4.69	5.26
0.9		0.11	0.5	0.082	0.909	5.02	6.58	4.38	6.85	4.62	6.93	
		0.13				4.99	5.69	4.83	6.45	4.89	6.51	
		0.15				4.80	5.22	4.72	5.69	4.52	5.71	

## 5. Conclusion

실제적으로, 두 처리의 동등성 혹은 비열등성을 보이고 싶을 경우에는, 기존의 비율차이가 없음을 귀무가설로 하는 검정방법을 적용할 수 없다. 두 처리간 동등성/ 비열등성 검정을 하기 위해 특히, matched-pair design에 관심을 갖고, 이 경우에 적절한 두 가지 통계량을 제안하였다. 두 통계량 모두 odds ratio에 기초하여 유도되었으며, 이들 통계량의 성능을 비교하기 위해, 기존에 동등성/비열등성 검정을 위해 제시된 비율차이, risk ratio에 기초한 통계량들을 소개하고, 이들의 성능을 모의실험을 통해 보여주었다. 모의실험 결과, 전반적으로 wald-type 통계량들이 empirical type I error 측면에서, 미리정한 유의수준에 많이 벗어나는 반면, score-type이나 fieller-type 통계량들은 매우 만족스러운 결과를 보여주고 있다. 그리고 비 대각 셀의 확률이 작아질 경우에는 비율차이에 의한 통계량보다 risk ratio를 이용한 fieller-type 통계량과 본 논문에서 제시된 fieller-type 통계량이 더 안정적인 것을 볼 수 있었다. 일반적으로 matched

Tests for equivalence/non-inferiority based on  
odds ratio in matched-pair design

paired design의 결과가 비 대각 셀의 빈도수가 대각 셀에 비해 적은 경우가 많으므로, 이 두 통계량이 matched-pair design에 더 적절한 성격을 가지고 있다고 결론내릴 수 있다. 아쉬운 점은, odds ratio에 기초한 fieller-type 통계량의 수식이 너무 복잡하여 해를 폐쇄형으로 구하지 못하고, 수치해석적인 방법으로 구할 수밖에 없었던 것이다. 그리고 본 연구에서는 empirical type I error 측면에서만 비교를 하였지만, 추후 연구를 통해 power를 이용한 성능의 비교를 해볼 만 하다.

참고문헌

- Lu Y. Bean JA. (1995), On the sample size for one-sided equivalence of sensitivities based upon McNemar's test. *Statistics on Medicine*, Vol.14, 1831-1839.
- Nam J-m (1997), Establishing equivalence of two treatments and sample size requirements in matched-pairs design. *Biometrics*, Vol.53, 1422-1430.
- Lachenbruch PA, Lynch CJ (1998), Assessing screening tests: extension of McNemar's test. *Statistics on Medicine*, Vol.17, 2207-2217.
- Tango T (1998), Equivalence test and confidence interval for the difference in proportions for the paired-sample design. *Statistics on Medicine*, Vol.17, 891-908.
- Hansheng Wang, Shein-Chung Chow, Gang Li (2002), On the sample size calculation based on odds ratio in Clinical trials. *Journal of Biopharmaceutical Statistics*. Vol.12.
- Tu, D (1998), On the use of the ratio of odds ratio of cure rates in therapeutic equivalence clinical trials with binary endpoints. *Journal of Biopharmaceutical Statistics*, Vol.8, 263-282.
- Liu, J. P., Hsueh, H. M., Hsieh, E., and Chen, J. J. (2002), Tests for equivalence or non-inferiority for paired binary data. *Statistics on Medicine*, Vol.21, 231-245.
- Man-Lai Tang (2003), Matched-pair noninferiority trials using rate ratio: a comparison of current methods and sample size refinement. *Controlled Clinical Trials*, Vol.24, 364-377.
- Tang NS, Tang ML, Chan ISF. (2003), On tests of equivalence via non-unity relative risk for matched-pair design. *Statistics on Medicine*, Vol.22, 1217-1233.
- Tang ML, Tang NS, Chan ISF, Chan BPS.(2002), Sample size determination for establishing equivalence/non-inferiority via ratio of two proportions in matched-pair design. *Biometrics*, Vol.58, 957-963.
- Nam J, Blackwelder WC.(2002), Analysis of the ratio of marginal probabilities in a matched-pair design. *Statistics on Medicine*, Vol.21, 689-699.