

비열등성 임상시험에서 독립인 두 비율 차 검정에 대한 비교 연구

윤 민¹ · 곽민정²

¹부경대학교 통계학과, ²평택대학교 디지털응용정보학과

(2010년 11월 접수, 2010년 12월 채택)

요약

두 비율비교 비열등성 임상시험(non-inferiority trial)에서의 검정은 두 비율간에 차이가 없다는 영가설에 근거한 정규근사 산출식이 흔히 사용 된다. 본 연구에서는 두 비율차이가 영이 아니라는 가설에서 유도된 Miettinen과 Nurminen이 제안한 우도 스코어검정(likelihood score test)과 Farrington과 Manning의 검정 그리고 비대칭도(skewness)을 보정한 Gart와 Nam이 제안한 검정방법을 소개하고 모의실험을 통하여 제 1종 오류와 검정력을 비교하여 비열등성 임상시험에서 검정방법의 결정에 가이드라인을 제시하고자 한다.

주요용어: 비열등성 임상시험, Miettinen과 Nurminen, Farrington과 Manning, Gart와 Nam.

1. 서론

비열등성 임상시험은 시험약이 대조약에 비해서 열등하지 않다는 것을 보이기 위한 임상시험으로 제형을 변화시켜 만든 약이나 부작용을 줄이거나 복용이 용이하도록 만든 약등이 그 예가 된다. 현재 국내에서 실시되고 있는 임상시험에서 비열등성 임상시험(Non-inferiority trial)이 많은 부분을 차지하고 있다. 이러한 임상시험의 시행을 위한 피험자수의 산출이나 유효성 평가에 일반적으로 귀무가설이 영가설 즉, 두 약의 치료율에는 차이가 없다는 가설에서 유도된 정규근사식을 이용한 검정을 사용하고 있으나 비열등성 임상시험과 같이 귀무가설이 영가설이 아닌 경우에는 적절하지 못하게 된다. 비열등성 임상시험이나 동등성임상시험처럼 귀무가설이 영가설이 아닌 경우에 이러한 조건이 제약식으로 주어진 우도함수(likelihood function)로부터 유도된 방법으로 Miettinen과 Nurminen (1985)의 우도 스코어검정(likelihood score test)과 Farrington과 Manning (1990)의 우도 스코어 검정 그리고 비대칭도(skewness)을 보정한 Gart와 Nam (1988, 1990)의 우도 스코어검정을 방법으로 사용할 수 있다. 임상시험에서 제 1종 오류의 적절한 통제는 타당하지 못한 약이 허가를 받게 되는 오류를 통제한다는 측면에서 매우 중요하며 검정력의 확보는 제약사의 입장에서 개발한 타당한 약이 허가를 받게 된다는 측면에서 또한 중요하다고 할 수 있다. 그러나 비열등성 임상시험에서 두 비율비교 검정방법에 대한 비교 연구는 거의 이루어지지 않고 있는 실정이다. 따라서 본 논문에서는 병행디자인으로 설계된 비열등성 임상시험의 두 비율비교 검정에 사용되는 방법에 대해서 소개하고 exhaustive enumeration simulation study를 통하여 다양한 반응률과 비열등성 한계(non-inferiority margin)에서 표본크기에 따라 이들 검

본 논문은 2009학년도 부경대학교의 지원을 받아 수행된 연구임(PKS-2009-4).

²교신저자: (450-701) 경기도 평택시 용이동, 평택대학교 디지털응용정보학과, 부교수.

E-mail: mjkwak@ptu.ac.kr

정방법의 제 1종 오류와 검정력을 비교하고자 한다. 2절에서는 비열등성 임상시험에서 두 비율차이에 대한 검정방법을 소개하고, 3절에서는 모의실험을 통하여 이들 검정방법의 제 1종 오류와 검정력을 비교하여 보았으며 마지막 절에서 결론을 제시하였다.

2. 두 비율비교 비열등성 임상시험의 검정

병행디자인의 두 비율비교 임상시험에서 2×2 분할표가 있을 때 x_T 와 x_C 는 각각 n_T 와 n_C 에서 성공확률 p_T 와 p_C 를 가지는 이항분포에서 추출된 확률변수로 정의한다. 즉, 2×2 분할표는 다음과 같이 생성된다. 여기서 p_i (i 는 T 혹은 C)는 성공확률을 나타내며 p_i 가 증가할수록 향상된 것으로 가정한다. 비

	성공	실패	Total
Treatment	x_T	$n_T - x_T$	n_T
Control	x_C	$n_C - x_C$	n_C
Total	m_1	m_2	N

열등성 임상시험에서 비열등성 한계를 δ_0 라고 정의하면 귀무가설과 대립가설은 다음과 같다.

$$H_0 : p_T - p_C \leq -\delta_0 \quad v.s. \quad H_1 : p_T - p_C > -\delta_0,$$

여기서 $\delta_0 > 0$ 이다. 마찬가지로 동등성 임상시험의 귀무가설과 대립가설을 다음과 같이 정의된다.

$$H_0 : |p_T - p_C| \leq -\delta_0 \quad v.s. \quad H_1 : |p_T - p_C| > -\delta_0.$$

동등성 임상시험은 $p_T - p_C > -\delta_0$ 와 $p_T - p_C < \delta_0$ 라는 두 가지 가설을 모두 만족함을 보여야 하므로 TOST(Two One Sided Test)라고 할 수 있다. 비열등성 임상시험의 검정통계량은 다음과 같이 정의될 수 있다.

$$z = \frac{(\hat{p}_T - \hat{p}_C) + \delta_0}{\hat{\sigma}}$$

여기서 분모의 분산을 추정하는 방법에서 두 비율에 차이가 없다는 영가설에서 유도된 추정치와 영이 아닌 상수라는 제약조건에서 유도된 분산의 추정치를 사용한 경우를 제시하였다.

2.1. 영가설에서 유도된 정규근사 검정통계량

두 비율비교검정에는 다양한 검정방법이 제안되었으나 그 중 가장 흔하게 사용되는 검정은 두 비율이 같다는 가설하에서 합동분산(pooled variance)사용한 정규근사검정방법이며 다음과 같다.

$$z_p = \frac{\hat{p}_T - \hat{p}_C + \delta_0}{\hat{\sigma}_p},$$

여기서 $\hat{\sigma}_p = \sqrt{\bar{p}(1-\bar{p})(1/n_T + 1/n_C)}$, $\bar{p} = (n_T \hat{p}_T + n_C \hat{p}_C)/(n_T + n_C)$ 이다.

또 다른 대안으로 합동분산이 아닌 두 비율에 대한 각각의 추정치를 사용하여 구할 수 있다. 즉,

$$z_{up} = \frac{\hat{p}_T - \hat{p}_C + \delta_0}{\hat{\sigma}_{up}},$$

여기서 $\hat{\sigma}_{up} = \sqrt{\hat{p}_T(1-\hat{p}_T)/n_T + \hat{p}_C(1-\hat{p}_C)/n_C}$ 이다.

2.2. Miettinen과 Nurminen의 우도 스코어검정

비열등성임상시험과 같이 두 비율차이가 영이 아닌 δ_0 라는 귀무가설에 대하여 Miettinen과 Nurminen은 우도 스코어검정을 제안하였다. 이 우도 스코어검정의 분자에는 최대우도추정량(MLE) \hat{p}_T 과 \hat{p}_C 를 구하여 대입하고 분산의 추정에서 $p_T - p_C = -\delta_0$ 라는 제약조건하에서 다음과 같은 로그-우도 함수에서 최대우도 추정량 \tilde{p}_C 와 \tilde{p}_T 를 유도하였다.

$$\ln(p_C) \propto x_T \ln(p_C - \delta_0) + (n_T - x_T) \ln(1 - p_C + \delta_0) + x_C \ln(p_C) + (n_C - x_C) \ln(1 - p_C).$$

이로부터 구한 검정통계량은 다음과 같다.

$$z_{MN} = \frac{\hat{p}_T - \hat{p}_C + \delta_0}{\hat{\sigma}_{MN}},$$

여기서,

$$\hat{\sigma}_{MN} = \sqrt{\left(\frac{\tilde{p}_T(1 - \tilde{p}_T)}{n_T} + \frac{\tilde{p}_C(1 - \tilde{p}_C)}{n_C} \right) \left(\frac{N}{N-1} \right)}, \quad \tilde{p}_T = \tilde{p}_C - \delta_0 = 2B \cos(A) - \frac{L_2}{3L_3},$$

$$A = \frac{1}{3} \left[\pi + \cos^{-1} \left(\frac{C}{B^3} \right) \right], \quad B = \text{sign}(C) \sqrt{\frac{L_2^2}{9L_3} - \frac{L_1}{3L_3}}, \quad C = \frac{L_2^3}{27L_3^3} - \frac{L_1L_2}{6L_3^2} + \frac{L_0}{2L_3}.$$

그리고

$$L_0 = x_C \delta_0 (1 - \delta_0), \quad L_1 = (n_C \delta_0 - N - 2x_C) \delta_0 + m_1, \quad L_2 = (N + n_C) \delta_0 - N - m_1, \quad L_3 = N.$$

2.3. Farrington과 Manning의 우도 스코어검정

Farrington과 Manning의 방법은 Miettinen과 Nurminen이 제안한 $\tilde{p}_C - \tilde{p}_T = \delta_0$ 라는 제약조건하에서 얻어진 최대우도 추정량 \tilde{p}_C 와 \tilde{p}_T 를 사용하며 분산의 추정량에 $N/(N-1)$ 만큼의 차이가 있다는 점을 제외하고 Miettinen과 Nurminen의 통계량과 동일한 값을 갖게 된다. Farrington과 Manning의 검정 통계량은 다음과 같다.

$$z_{FM} = \frac{\hat{p}_C - \hat{p}_T + \delta_0}{\sqrt{\left(\frac{\tilde{p}_T(1 - \tilde{p}_T)}{n_T} + \frac{\tilde{p}_C(1 - \tilde{p}_C)}{n_C} \right)}}.$$

그러므로 표본의 크기가 커지면 두 방법은 거의 동일한 값을 갖게 된다.

2.4. Gart와 Nam의 우도 스코어검정

Gart와 Nam은 Farrington과 Manning의 통계량에서 비대칭도(skewness)를 보정한 방법을 제시하였다. Farrington과 Manning의 통계량 z_{FM} 에서의 비대칭도는 다음과 같다.

$$\tilde{\gamma}_1(\delta) = \tilde{V}^{\frac{3}{2}} \left\{ \frac{\tilde{p}_T \tilde{q}_T (\tilde{q}_T - \tilde{p}_T)}{n_T^2} - \frac{\tilde{p}_C \tilde{q}_C (\tilde{q}_C - \tilde{p}_C)}{n_C^2} \right\},$$

여기서 $V = \{\tilde{p}_T(1 - \tilde{p}_T)/n_T + \tilde{p}_C(1 - \tilde{p}_C)/n_C\}^{-1}$ 이다. 그러므로 보정된 통계량 z_{GN} 는 다음 식의 해가 된다.

$$(-\tilde{\gamma})z_{GN}^2 + (-1)z_{GN} + (z_{FM}(\delta_0) + \tilde{\gamma}) = 0,$$

여기서 $\tilde{\gamma} = \tilde{\gamma}_1(\delta)/6$ 이다.

표 3.1. 실제 제 1종 오류비교(유의수준 2.5%, $\delta_0 = 0.05, p_C = 0.3, 0.5, 0.7, 0.9$)

p_C	표본수	z_p	z_{up}	z_{FM}	z_{MN}	z_{GN}
0.3	20	2.90	3.28	2.42	2.42	2.42
	40	2.55	2.79	2.55	2.55	2.55
	60	2.58	2.66	2.58	2.58	2.58
	80	2.63	2.63	2.46	2.46	2.48
	100	2.56	2.58	2.56	2.56	2.56
0.5	20	3.60	3.61	3.60	3.60	3.60
	40	2.83	2.84	2.83	2.83	2.83
	60	2.73	2.74	2.73	2.73	2.73
	80	2.39	2.39	2.39	2.39	2.39
	100	2.79	2.79	2.79	2.79	2.79
0.7	20	2.66	3.37	2.46	2.46	2.46
	40	2.42	2.72	2.46	2.46	2.46
	60	2.52	2.70	2.52	2.52	2.52
	80	2.60	2.64	2.59	2.54	2.54
	100	2.52	2.62	2.52	2.52	2.52
0.9	20	4.74	5.96	1.91	1.91	1.91
	40	3.44	3.44	2.14	2.14	2.61
	60	2.93	2.93	2.33	2.22	2.51
	80	2.77	2.88	2.34	2.34	2.35
	100	2.60	2.82	2.49	2.49	2.50

3. 모의실험

임상시험에서 제 1종 오류의 적절한 통제는 중요한 이슈이며 제약사의 입장에서 충분한 검정력을 확보하는 것 또한 중요한 문제라고 할 수 있다. 두 비율비교 비열등성 임상시험은 현재 국내에서 시행되고 있는 임상시험에서 많은 부분을 차지하고 있으나 대개의 경우 영가설에서부터 유도된 일반적인 정규근사검정을 사용하여 결론을 내리고 있다. 그러므로 이 절에서는 모의실험을 통하여 제시된 검정방법들의 제 1종 오류와 검정력을 비교하여 보고자 한다.

비열등성관계가 엄격한 경우에서 관대한 경우($\delta_0 = 0.05, 0.1, 0.2$)의 세 가지로 나누고 반응률 p_C 는 0.3에서 0.9까지 변화시켜 다양한 표본크기에 따라서 제시된 검정방법들의 제 1종 오류와 검정력을 비교하여 보았다.

모의실험은 각각의 표본크기에서 exhaustive enumeration 방법을 사용하여 비교하였으며 제 1종 오류와 검정력은 다음과 같이 계산되며 통계패키지 PASS(Power Analysis for Sample Size)를 사용하였다.

$$\alpha = \sum_R \binom{n_T}{x_T} (p_C - \delta_0)^{x_T} (1 - p_C + \delta_0)^{n_T - x_T} \binom{n_C}{x_C} p_C^{x_C} (1 - p_C)^{n_C - x_C}$$

$$1 - \beta = \sum_R \binom{n_T}{x_T} p_{T,1}^{x_T} (1 - p_{T,1})^{n_T - x_T} \binom{n_C}{x_C} p_{C,1}^{x_C} (1 - p_{C,1})^{n_C - x_C},$$

여기서 $R = \{z : z > z_{critical}\}$ 로 기각역을 의미하며 대립가설에서 주어진 $p_{T,1}$ 과 $p_{C,1}$ 에 따라서 검정력이 계산된다. 모든 검정방법의 비교는 ICH E9에서 권고하고 있으며 실제 비열등성 임상시험에서 대부분 진행하고 있는 유의수준 2.5%에서 비교하였으며 그 결과는 표 3.1 ~ 3.6 그리고 그림 3.1과 3.2에 제시되었다.

표 3.2. 실제 제 1종 오류비교(유의수준 2.5%, $\delta_0 = 0.1$, $p_C = 0.3, 0.5, 0.7, 0.9$)

p_C	표본수	z_p	z_{up}	z_{FM}	z_{MN}	z_{GN}
0.3	20	2.89	2.92	2.74	2.74	2.74
	40	2.69	2.82	2.45	2.45	2.45
	60	2.71	2.76	2.51	2.48	2.48
	80	2.67	2.67	2.48	2.48	2.48
	100	2.55	2.67	2.50	2.50	2.50
0.5	20	2.27	2.79	2.27	3.27	2.27
	40	2.77	2.78	2.77	2.77	2.77
	60	2.67	2.67	2.67	2.67	2.67
	80	2.34	2.36	2.36	2.34	2.34
	100	2.72	2.72	2.72	2.72	2.72
0.7	20	2.74	3.13	2.74	2.74	2.74
	40	2.48	2.68	2.47	2.47	2.47
	60	2.56	2.56	2.55	2.41	2.41
	80	2.56	2.65	2.63	2.54	2.54
	100	2.50	2.51	2.50	2.50	2.50
0.9	20	3.67	3.67	2.12	2.12	2.43
	40	3.48	3.49	2.19	2.19	2.33
	60	2.86	3.03	2.40	2.40	2.44
	80	2.80	2.81	2.53	2.53	2.54
	100	2.86	2.86	2.48	2.48	2.48

표 3.3. 실제 제 1종 오류비교(유의수준 2.5%, $\delta_0 = 0.2$, $p_C = 0.3, 0.5, 0.7, 0.9$)

p_C	표본수	z_p	z_{up}	z_{FM}	z_{MN}	z_{GN}
0.3	20	3.61	3.61	2.73	2.09	2.73
	40	3.10	3.10	2.37	2.37	2.37
	60	2.66	2.95	2.54	2.31	2.54
	80	2.63	2.81	2.49	2.36	2.61
	100	2.69	2.69	2.46	2.46	2.46
0.5	20	3.48	2.48	2.98	2.47	3.48
	40	2.50	2.50	2.57	2.50	2.50
	60	2.42	2.42	2.50	2.42	2.42
	80	2.26	2.37	2.71	2.61	2.49
	100	2.41	2.41	2.47	2.47	2.44
0.7	20	2.48	3.48	2.98	2.47	2.48
	40	2.50	2.50	2.57	2.50	2.50
	60	2.42	2.42	2.50	2.42	2.42
	80	2.26	2.37	2.71	2.61	2.49
	100	2.41	2.41	2.47	2.47	2.44
0.9	20	3.61	3.61	2.73	2.09	2.73
	40	3.10	3.10	2.37	2.37	2.37
	60	2.66	2.95	2.54	2.31	2.54
	80	2.63	2.81	2.49	2.36	2.61
	100	2.69	2.69	2.46	2.46	2.46

표 3.1 ~ 3.3과 그림 3.1로부터 유의수준 2.5%에서 각 검정방법들의 실제 제 1종 오류를 비교해 보면 각각의 분산 추정치를 사용한 z_{up} 가 유의수준 2.5%보다 크게 나타나는 경향을 보이고 있으며 합

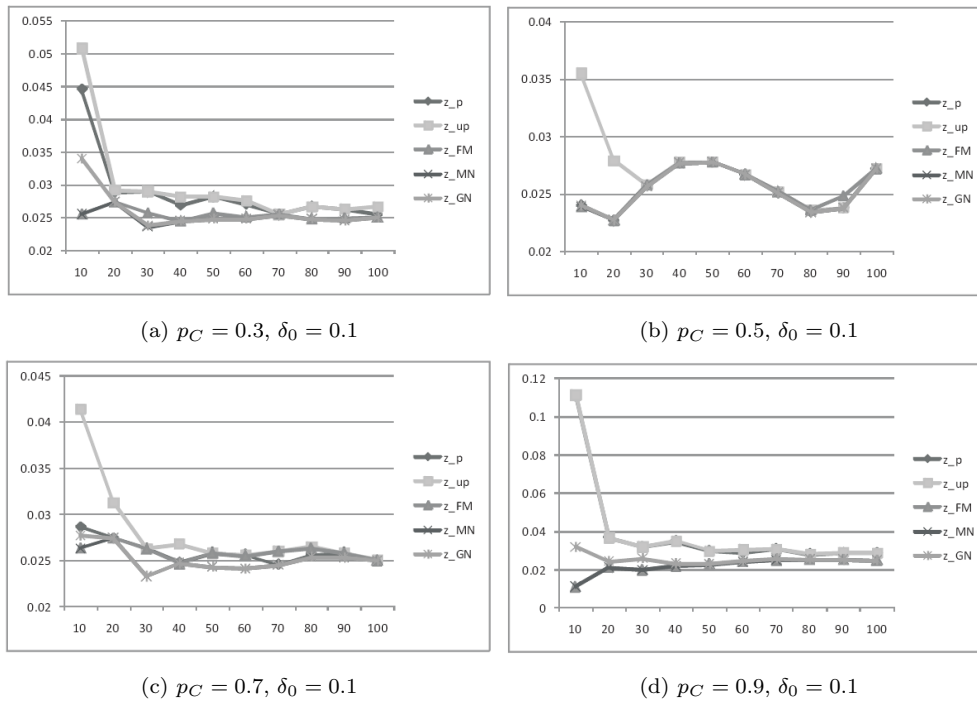


그림 3.1. 제 1종 오류의 비교($p_C = 0.3, 0.5, 0.7, 0.9, \delta_0 = 0.1$, 유의수준 2.5%)

표 3.4. 검정력비교(유의수준 2.5%, $\delta_0 = 0.05, p_C, p_T = 0.5, 0.7, 0.9$)

p_C	표본수	z_p	z_{up}	z_{FM}	z_{MN}	z_{GN}
0.5	20	4.21	7.69	4.21	4.21	4.21
	40	7.28	7.28	7.28	7.28	7.28
	60	8.53	8.53	8.53	8.53	8.53
	80	8.94	8.95	8.95	8.94	8.94
	100	11.46	11.46	11.46	11.46	11.46
0.7	20	5.76	6.71	5.23	5.23	5.23
	40	6.98	7.64	6.98	6.98	6.98
	60	8.80	9.21	8.80	8.80	8.80
	80	10.50	10.54	10.21	10.17	10.18
	100	11.91	12.14	11.91	11.91	11.91
0.9	20	5.76	6.71	5.23	5.23	5.23
	40	6.98	7.64	6.98	6.98	6.98
	60	8.80	9.21	8.80	8.80	8.80
	80	10.50	10.54	10.21	10.17	10.18
	100	11.91	12.14	11.91	11.91	11.91

동분산 추정치를 사용한 z_p 역시 유의수준보다 높게 나타나고 있다. 이러한 경향은 반응률이 아주 높고($p = 0.9$) 표본크기가 작은 경우에 더욱 뚜렷하게 나타나며 이러한 경우 제 1종 오류의 통제가 어렵게 된다. 영가설이 아닌 귀무가설에서 유도된 우도 스코어검정방법의 제 1종 오류는 거의 비슷하게 나타내고 있으며 작은 표본 크기에서도 대체로 유의수준 2.5%에 근접하고 있다.

표 3.5. 검정력비교(유의수준 2.5%, $\delta_0 = 0.1$, $p_C, p_T = 0.5, 0.7, 0.9$)

p_C	표본수	z_p	z_{up}	z_{FM}	z_{MN}	z_{GN}
0.5	20	8.30	9.60	8.30	8.30	8.30
	40	15.72	15.72	15.72	15.72	15.72
	60	20.57	20.57	20.57	20.57	20.57
	80	23.85	23.86	23.86	23.85	23.85
	100	31.04	31.04	31.04	31.04	31.04
0.7	20	11.10	11.69	11.05	11.05	11.05
	40	16.24	17.41	16.10	16.10	16.10
	60	23.10	23.10	22.87	22.04	22.04
	80	28.69	28.78	28.17	28.08	28.08
	100	34.31	34.54	34.31	34.31	34.31
0.9	20	22.17	22.17	12.27	12.27	15.55
	40	37.80	37.80	25.96	25.96	29.25
	60	46.99	47.12	39.54	39.54	41.33
	80	56.98	57.63	52.15	52.10	53.25
	100	66.82	66.82	62.32	62.32	62.32

표 3.6. 검정력비교(유의수준 2.5%, $\delta_0 = 0.2$, $p_C, p_T = 0.5, 0.7, 0.9$)

p_C	표본수	z_p	z_{up}	z_{FM}	z_{MN}	z_{GN}
0.5	20	22.62	22.62	24.99	22.62	22.62
	40	45.55	45.55	45.56	45.55	42.55
	60	60.78	60.78	60.79	60.78	60.78
	80	71.00	71.01	71.20	71.08	71.03
	100	82.10	82.10	82.10	82.10	82.10
0.7	20	29.93	29.93	30.93	28.92	29.85
	40	49.55	49.55	51.24	49.37	49.37
	60	67.54	67.54	68.12	67.31	67.32
	80	79.17	79.26	79.33	79.31	79.30
	100	87.16	87.27	87.66	87.66	87.50
0.9	20	59.16	59.16	47.74	46.03	47.74
	40	85.68	85.68	79.32	79.32	79.32
	60	94.70	94.82	93.16	92.87	93.19
	80	98.42	98.46	97.74	97.74	98.10
	100	99.58	99.58	99.40	99.40	99.47

검정력은 대립가설에서 p_T 와 p_C 가 서로 같다는 조건에서 비열등성 한계 δ 가 0.05, 0.1, 0.2인 경우에 따라서 비교하여 보았다. p_T 와 p_C 가 서로 같다는 대립가설에서 0.3과 0.7은 서로 대칭이므로 반응률 0.5, 0.7, 0.9에 대한 모의실험결과를 제시하였다. 검정력은 표 3.4 ~ 3.6과 그림 3.2에서와 같이 5가지 방법 모두 큰 차이를 보이고 있지 않고 있다. 그러나 유의수준 2.5%에서 z_p 와 z_{up} 검정의 실제 제 1종 오류가 유의수준을 대부분 초과하고 있는 점을 감안한다면 영이 아니 가설에서 유도된 나머지 세 가지 검정 방법 z_{MN} , z_{FM} , z_{GN} 검정이 좀 더 높은 검정력을 보인다고 할 수 있겠다. 각 표본크기에 따라서 검정력을 살펴보면 표본크기가 100인 경우에서도 비열등성 한계가 0.2인 경우를 제외하고 실제 검정력은 낮게 나오고 있음을 알 수 있다. 그러므로 비열등 임상시험에서 일정 수준의 검정력을 보장하기 위해서는 많은 피험자수를 필요로 하게 된다.

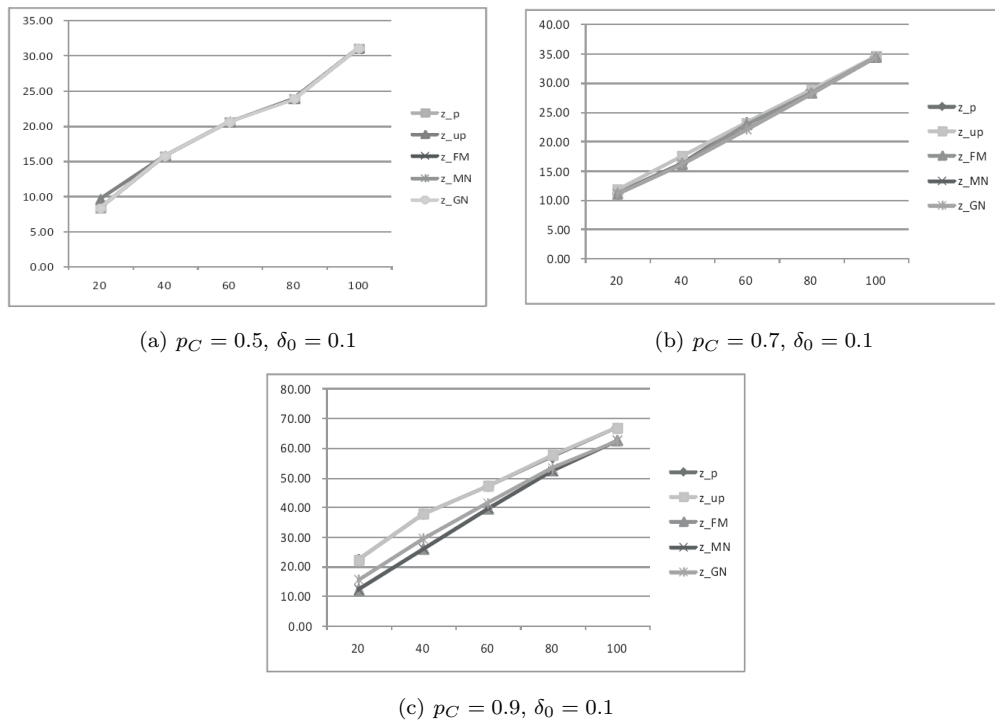


그림 3.2. 검정력의 비교($p_C = 0.5, 0.7, 0.9, \delta_0 = 0.1$, 유의수준 2.5%)

4. 결론

두 비율비교 비열등성 임상시험은 현재 국내에서 진행되는 임상시험에서 큰 부분을 차지하고 있으나 여러 검정방법들의 적절성에 대한 비교는 이루어지고 있지 않은 실정이다. 임상시험에서 제 1종 오류의 적절한 통제는 타당하지 못한 약이 허가를 받게 되는 오류를 통제한다는 측면에서 매우 중요하며 검정력의 확보는 제약사의 입장에서 개발된 타당하나 약이 허가를 받게 된다는 측면에서 또한 중요하다고 할 수 있다.

모의실험결과에 따르면 영가설에서 유도된 z_p 나 z_{up} 검정방법은 모두 실제 제 1종 오류가 유의수준 2.5%보다 높게 나타나는 경향을 보이고 있으며 영이 아닌 제약조건에서 유도된 z_{MN} , z_{FM} 과 z_{GN} 검정방법은 적은 표본크기에서도 유의수준에 근접하게 나타나고 있어 제 1종 오류의 통제가 비교적 적절하게 이루어진다고 할 수 있다. 검정력은 5가지 방법 모두 큰 차이를 보이고 있지 않으나 z_p 와 z_{up} 검정의 제 1종 오류가 다른 세 방법보다 크게 나타난다는 점에서 z_{MN} , z_{FM} 와 z_{GN} 검정이 더 적절하다고 할 수 있으며 이들 세 방법간에는 큰 차이를 보이지 않았다. 그러므로 두 비율비교 비열등성 임상시험에서 Miettinen과 Nurminen의 우도 스코어검정, Farrington과 Manning의 우도 스코어검정 그리고 Gart와 Nam의 우도 스코어검정방법이 권장된다.

참고문헌

- Farrington, C. P. and Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-Zero risk difference or non-Unity relative risk, *Statistics in Medicine*, **9**, 1447-1454.

- Gart, J. J. and Nam, J. (1988). Approximate interval estimation of the ratio in binomial parameters: A review and corrections for skewness, *Biometrics*, **44**, 323–338.
- Gart, J. J. and Nam, J. (1990). Approximate interval estimation of the difference in binomial parameters: correction for skewness and extension to multiple tables, *Biometrics*, **46**, 637–643.
- ICH E9 Statistical Principles for Clinical Trials, <http://www.ich.org>.
- Miettinen, O. S. and Nurminen, M. (1985). Comparative analysis of two rates, *Statistics in Medicine*, **4**, 213–226.
- PASS(Power Analysis and Sample Size) <http://www.ncss.com/pass.html>.

Comparative Analysis of Two Independent Proportions in Non-Inferiority Trials

Min Yoon¹ · Minjung Kwak²

¹Department of Statistics, Pukyong National University

²Department of Information Science, Pyeongtaek University

(Received November 2010; accepted December 2010)

Abstract

Normal approximation methods under the null hypothesis of no difference are frequently used to test the two independent proportions in non-inferiority trials. However, these tests are not appropriate under the null hypothesis of non-zero difference. We review the likelihood score methods proposed by Miettinen and Nurminen, Farrington and Manning, and Gart and Nam and compare the performance of these tests. The simulation study shows that the likelihood score tests under the null hypothesis of non-zero difference have better performance at a Type I error and power than usual normal approximation methods.

Keywords: Non-inferiority trial, Miettinen and Nurminen, Farrington and Manning, Gart and Nam.

This work was supported by the Pukyong National University Research Fund in 2009(PKS-2009-4).

²Corresponding author: Professor, Department of Information Science, Pyeongtaek University, Gyeonggi, 450-701, Korea. E-mail: mjkwak@ptu.ac.kr