응용통계연구 (2011)
24(6), 1045–1054
DOI: http://dx.doi.org/10.5351/KJAS.2011.24.6.1045

증상 발현과 증상 심각성을 병합한 추세검정법

최세미[1] · 양수[2] · 송혜향[3]

[1]가톨릭대학교 대학원 의학통계학과, [2]가톨릭대학교 간호대학
[3]가톨릭대학교 대학원 의학통계학과

(2011년 3월 접수, 2011년 10월 채택)

요 약

임상 연구자는 일정기간의 치료에 의해 환자에게 증상 발현 또는 증상 비발현을 관측할 뿐만 아니라 증상이 발현된 환자의 경우에는 증상 심각성 점수가 측정됨을 자주 본다. 예를 들어서 대조군을 비롯하여 치료의 정도에 차이가 있는 여러 치료군이 있을 때 발현율의 감소를 검정하는 전통적인 방법은 Cochran-Armitage 추세검정이며, 한편 비정규분포하는 연속자료의 감소를 검정하는 전통적인 방법은 Jonckheere-Terpstra 추세검정이다. 전자의 추세검정은 증상 발현의 이항자료만을 사용한 검정이 되며 한편 후자의 추세검정은 증상의 심각성 점수만을 사용한 검정이 된다. 본 논문에서는 증상 발현율과 증상의 심각성 점수로 나타나는 정보의 특성을 모두 감안한 추세검정법을 제안한다. 새로운 추세검정법을 항정신병 약제를 복용하는 정신분열증 입원환자로부터 수집된 자료에 적용하며 본 논문에서 취급하는 여러 다른 검정법과 비교한 효율성을 모의실험으로 알아본다.

주요용어: 추세검정법, 증상 발현율, 증상 심각성.

1. 서론

일정 기간의 치료에 의해 이 기간 동안에 환자에게 나타나는 증상 발현을 관측할 뿐만 아니라 증상 발현의 경우에는 증상의 심각함을 나타내는 연속형 또는 순위형 변수가 함께 측정되는 경우가 있다. 이러한 자료의 특성은 환자에게 관측되는 증상발현($X = 1$)의 이산형 변수와 증상의 심각성을 나타내는 연속형 변수(Y)가 모두 분석될 수 있다는 것이다. 특히 대조군과 더불어 치료의 정도에 있어 차이를 나타내는 여러 치료군이 있는 경우에 서로 다른 치료에 의해 수집되는 이산형과 연속 또는 순위형 자료의 분석방법으로 여러 가지가 있다. 대다수의 환자가 증상 발현이라면 단순히 비발현 환자를 제외하고서 나머지 환자의 증상 심각성 수치만으로 치료효과를 알아볼 수 있으며, 증상 심각성의 추세 검정법으로는 Jonckheere-Terpstra 검정 통계량이 있다 (Jonckheere, 1954; Terpstra, 1952). 그러나 증상 비발현의 환자가 무시될 수 없는 숫자라면 증상발현이 있었는가의 이산형 자료로 치료효과를 알아볼 수 있으며 이산형 자료의 추세검정법으로서 Cochran-Armitage 검정통계량이 있다 (Cochran, 1954; Armitage, 1955). 그러나 한편 증상 비발현 환자에게 0값의 심각성 수치를 부여한 후 앞에서와 같이 Jonckheere-Terpstra 검정통계량으로 분석할 수도 있으나 증상 심각성의 수치가 0보다 매우 높을 때에는 발현자와 비발현자의 증상 심각성 수치는 크게 벌어져 이 방법 역시 적절하지 않다. 더욱 바람직한 방법인 이산형 변수와 연속형 변수 그대로를 모두 감안한 추세검정법은 아직까지 제시된 바가 없다.

[3]교신저자: (137-701) 서울 서초구 반포동 505 가톨릭대학교 대학원, 의생명과학교실 의학통계학과, 인간유전체 다형성 연구소, 교수. E-mail: hhsong@catholic.ac.kr

표 **2.1.** $2 \times K$ 분할표 자료

군	g_1	g_2	$\cdots$	g_K	총수
증상 비발현 $(X=0)$	l_1	l_2	$\cdots$	l_K	$l.$
증상 발현 $(X=1)$	n_1	n_2	$\cdots$	n_K	$n.$
총수	M_1	M_2	$\cdots$	M_K	N

항정신병 약제(antipsychotic drugs)를 복용하는 정신분열증(schizophrenic) 입원환자들은 배변에 어려움을 겪게 된다. 본 논문의 동기가 된, 이 환자들로부터 수집된 예제자료는 항정신병 약제로 인해 변비(constipation)를 경험하는 환자에게 점진적인 강도의 대처방법으로 치료한 결과이다. 환자에게서 변비의 증상이 있는가와 변비증상이 있는 경우에는 변비증상의 심각성 점수(severity score)가 측정되었으며, 여러 치료군에 대한 추세검정이 요구된다. 변비에 대한 정의는 3장에서 설명한다.

증상의 발현여부와 증상 심각성의 두 수치를 동시에 검정하는 통계량으로서 Chang 등 (1994)은 각 군에서 질병부담점수(burden of illness score) 평균을 제안하였으며, 이는 각 군의 총 질병부담점수를 총 환자수로 나눈 수치로써, 여기서 총환자수는 각 군에서 증상 심각성의 점수를 가진 환자들뿐만 아니라 증상이 없었던 환자들도 모두 감안된 것이다. 따라서 질병부담점수는 증상 발현자의 수가 많을수록 높고 동시에 증상 발현자의 증상이 심각할수록 높다. Chang 등 (1994)이 다룬 매우 단순한 두 치료군의 경우에는 이 질병부담점수의 평균차를 검정통계량으로 채택하였다. 본 논문에서는 대조와 더불어 여러 치료군이 있을 때, 특히 증상 심각성 수치가 비정규적인 경우의 추세검정통계량을 제안한다. 본 논문에서 언급한 여러 다른 검정법과 비교하여 장, 단점을 알아본다.

2. 증상 발현과 증상 심각성의 추세검정 통계량

이산형 자료에 근거한 증상 발현율의 추세검정 통계량을 2.1절에서 간략히 설명하고, 순위형 자료인 증상 심각성의 추세검정 통계량을 2.2절에서 설명한다. 증상 발현과 증상 심각성을 병합한 추세검정통계량을 2.3절에서 설명한다.

2.1. 증상 발현율의 추세검정 통계량

증상 발현율의 추세를 알아보는 목적에 $2 \times K$ 분할표 자료에 근거한 Cochran-Armitage 추세검정법이 적절하다 (Cochran, 1954; Armitage, 1955).

표 2.1에서 l_i과 n_i는 i번째 군에서 각각 비발현과 발현 환자수이다.

Cochran-Armitage 추세검정법의 귀무가설은 다음과 같다.

$$H_0 : \Pr(X = 1 \mid g_1) = \Pr(X = 1 \mid g_2) = \cdots = \Pr(X = 1 \mid g_K).$$

이제 i번째 군에서의 기대 증상 발현율을 p_i라 할 때, 위의 귀무가설은 $H_0 : p_1 = p_2 = \cdots = p_K$로 표현되며, 증상발현율의 감소추세에 대한 대립가설은 $H_1 : p_1 \geq p_2 \geq \cdots \geq p_K$로써 적어도 하나의 부등호가 존재한다. Cochran-Armitage 검정통계량은 다음과 같다 (Agresti, 2002).

$$T = \sum_{i=1}^{K} (l_i n. - n_i l.).$$

검정통계량 T의 귀무가설하에서 기댓값과 분산은 다음과 같다.

$$E(T) = E(E(T \mid l., n.)) = E(0) = 0,$$

$$\mathrm{Var}(T) = \frac{l.n.}{N}\left(\sum_{i=1}^{K} M_i(N-M_i) - 2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K} M_i M_j\right), \quad i<j.$$

추세검정은 구체적으로 검정통계량 $(T-E(T))/\sqrt{\mathrm{Var}(T)}$가 대표본하에서 정규분포함을 이용하여 단측검정으로 시행한다.

2.2. 증상 심각성의 추세검정 통계량

Jonckheere-Terpstra 추세검정법은 서로 독립인 $K(K\geq 3)$개의 군에서 관측된 연속자료의 증가 (또는 감소) 추세를 알아보는 비모수적 검정법이다 (Jonckheere, 1954; Terpstra, 1952). i번째 군의 연속자료의 모평균이 μ_i일 때 Jonckheere-Terpstra 추세검정법의 귀무가설과 증상심각성의 감소추세에 대한 대립가설은 다음과 같다.

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K \quad vs. \quad H_1 : \mu_1 \geq \mu_2 \geq \cdots \geq \mu_K.$$

대립가설 H_1에서 적어도 하나의 부등식이 존재한다. 위의 가설을 검정하기 위한 증상 발현이 관측된 환자를 대상으로 계산하는 Jonckheere-Terpstra 추세검정 통계량은 다음과 같다.

$$J = \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} W_{ij},$$

여기서

$$W_{ij} = \sum_{i=1}^{n_i}\sum_{j=1}^{n_j}\psi(Y_i,Y_j), \quad i<j; \qquad \psi(Y_i,Y_j) = \begin{cases} 1, & Y_i > Y_j, \\ 0.5, & Y_i = Y_j, \\ 0, & Y_i < Y_j. \end{cases} \tag{2.1}$$

즉, W_{ij}는 두 군 i와 j의 순위자료에 근거한 Mann-Whitney 통계량(간단히 Wilcoxon (1945) 통계량)이며, 추세 검정통계량 J는 결과적으로 $K(K-1)/2$개의 Wilcoxon (1945) 통계량의 합으로 표현되었다. 여기서 i번째 군과 j번째 군에 속한 표본수가 각각 n_i와 n_j일 때, 귀무가설하에서 W_{ij}의 기댓값과 대표본 근사 분산은 다음과 같다 (Hollander와 Wolfe, 1999, pp. 202–211).

$$E(W_{ij}) = \frac{n_i n_j}{2}, \qquad \mathrm{Var}(W_{ij}) = \frac{n_i n_j (n_i+n_j+1)}{12}. \tag{2.2}$$

이를 이용하여 Jonckheere-Terpstra 추세검정 통계량 J의 기댓값과 대표본 근사 분산은 다음과 같다 (Hollander와 Wolfe, 1999, p.203). 여기서 $n_.$은 $\sum_{i=1}^{K} n_i$이다.

$$E(J) = \frac{1}{4}\left[n_.^2 - \sum_{i=1}^{K} n_i^2\right],$$

$$\mathrm{Var}(J) = \frac{1}{72}\left[n_.^2(2n_.+3) - \sum_{i=1}^{K} n_i^2(2n_i+3)\right].$$

감소추세에 대한 검정은 구체적으로 $(J-E(J))/\sqrt{\mathrm{Var}(J)}$가 표준정규분포함을 이용하여 단측검정으로 시행한다. 만약 동점자료가 있는 경우라면 대표본 근사 분산은 다음과 같다 (Hollander와 Wolfe, 1999, pp. 202–211).

$$\mathrm{Var}(J) = \frac{1}{72}\left[n_.(n_.-1)(2n_.+5) - \sum_{i=1}^{K} n_i(n_i-1)(2n_i+5) - \sum_{j=1}^{g} t_j(t_j-1)(2t_j+5)\right]$$

$$+\frac{1}{36n.(n.-1)(n.-2)}\left[\sum_{i=1}^{K}n_i(n_i-1)(n_i-2)\right]\left[\sum_{j=1}^{g}t_j(t_j-1)(t_j-2)\right]$$
$$+\frac{1}{8n.(n.-1)}\left[\sum_{i=1}^{K}n_i(n_i-1)\right]\left[\sum_{j=1}^{g}t_j(t_j-1)\right],$$

여기서 j번째 동점그룹의 갯수가 t_j이며, g는 이러한 동점 그룹의 수이다.

2.3. 증상 발현과 증상 심각성을 병합한 추세검정 통계량

서로 독립인 K개의 치료군에 각각 M_i명의 환자들이 랜덤할당되었다. 일정 기간의 치료 후 환자들의 증상 발생여부와 증상이 발생한 환자의 경우 증상 발현의 심각성 점수 $Y_{ij}\,(i=1,2,\ldots,K,\ j=1,2,\ldots,M_i)$가 측정되었다. n_i명의 환자에게서 증상 발현을 관측하였을 때, n_i는 서로 독립적인 이항분포 $B(M_i,p_i)$를 따른다. 여기서 p_i는 기대 증상 발현율이다. 증상의 심각성 점수 역시 서로 독립적인 확률변수로써 평균 μ_i를 갖는다.

증상 발현과 증상 심각성을 병합하는 추세검정의 귀무가설과 대립가설은 다음과 같다. 대립가설 H_1에서 적어도 하나의 부등식이 성립하며, 이 대립가설은 증상 발현율의 감소추세와 증상의 심각성 점수의 감소추세를 나타내고 있다.

$$H_0: p_1=p_2=\cdots=p_K \quad \text{and} \quad \mu_1=\mu_2=\cdots=\mu_K,$$
$$H_1: p_1\geq p_2\geq\cdots\geq p_K \quad \text{or} \quad \mu_1\geq\mu_2\geq\cdots\geq\mu_K.$$

증상 발현과 증상 심각성 점수를 병합한 추세검정 통계량으로써 다음을 제안한다.

$$J_N=\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\frac{n_in_j}{M_iM_j}W_{ij},\quad i<j.$$

병합 검정통계량 J_N이 2.2절에 제시된 Jonckheere-Terpstra 검정통계량 J와 다른 점은 i번째 군과 j번째 군으로부터 계산한 W_{ij}에 $(n_in_j)/(M_iM_j)$가 곱해진 것이며, 이는 증상 심각성의 점수에 근거하여 계산하는 Wilcoxon 검정통계량의 결과를 두 군에 속한 모든 환자가 함께 나누어 갖는 것과 같다. 다시 말하면, 증상 발현이 관측된 환자에게서 n_in_j번의 $\psi(Y_i,Y_j)$가 식 (2.1)의 W_{ij}에서 계산된 것을 모든 환자 대상의 M_iM_j번의 계산으로 나누어 갖는 것이다. 따라서 만약 모든 환자에게서 증상 발현이 관측되어 심각성 점수를 가진다면 $n_i=M_i\,(i=1,2,\ldots,K)$로써 J_N은 바로 J와 같게 된다. 그러나 만약 모든 환자에게서 증상이 발현되지 않았다면 J_N은 0값을 가져서 귀무가설을 표현하게 된다. 식 (2.2)를 이용하여 $n_i\,(i=1,2,\ldots,K)$가 고정된 조건하에서 병합 추세검정 통계량의 기댓값과 분산은 다음과 같다.

$$E(J_N\mid n_1,n_2,\ldots,n_K)=\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\frac{n_in_j}{M_iM_j}E(W_{ij})=\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\frac{n_i^2n_j^2}{2M_iM_j},$$
$$\begin{aligned}\mathrm{Var}(J_N\mid n_1,n_2,\ldots,n_K)&=\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\left(\frac{n_in_j}{M_iM_j}\right)^2\mathrm{Var}(W_{ij})\\&=\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\frac{n_i^3n_j^3(n_i+n_j+1)}{12M_i^2M_j^2}.\end{aligned}$$

이제 n_i가 이항분포함을 이용한 병합 추세검정 통계량 J_N의 기댓값은 다음과 같다.

$$E(J_N) = E(E(J_N \mid n_1, n_2, \ldots, n_K)) = \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} \frac{1}{2M_iM_j} E\left(n_i^2\right) E\left(n_j^2\right)$$
$$= \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} \frac{1}{2M_iM_j} \left[M_ip_i(1-p_i) + M_i^2p_i^2\right] \left[M_jp_j(1-p_j) + M_j^2p_j^2\right]. \quad (2.3)$$

병합 추세검정 통계량 J_N의 분산으로는 다음을 제안한다.

$$\mathrm{Var}(J_N) \cong E[\mathrm{Var}(J_N \mid n_1, n_2, \ldots, n_K)]$$
$$= \sum_{i=1}^{K-1}\sum_{j=i+1}^{K} \left[\frac{1}{12M_i^2M_j^2} E\left(n_i^4n_j^3 + n_i^3n_j^4 + n_i^3n_j^3\right)\right], \quad (2.4)$$

여기서 $E(n_i^3)$과 $E(n_i^4)$은 다음과 같다.

$$E\left(n_i^3\right) = M_i(M_i-1)p_i^2[(M_i-2)p_i+3] + M_ip_i,$$
$$E\left(n_i^4\right) = M_ip_i^3[(M_i-1)(M_i-2)(M_i-3)p_i + 6(M_i-1)(M_i-2)] + M_ip_i[7(M_i-1)p_i+1].$$

추세검정은 식 (2.3)–(2.4)에서 각 군의 증상 발현율 p_i의 추정량으로 n_i/M_i을 대입한 후 검정통계량 $(J_N - E(J_N))/\sqrt{\mathrm{Var}(J_N)}$이 표준정규분포함을 이용하여 단측검정으로 시행한다.

3. 예제자료의 분석 및 모의실험

3.1. 예제자료의 분석

예제자료는 항정신병 약제로 인해 배변이 수월하지 않은 정신분열증 입원환자 총 56명을 세 군에 랜덤 할당하여 4주간 치료한 결과이다 (Yang, 1992). 드문 횟수의 배변, 불완전하거나 또는 힘든 배변으로 개인의 주관적 호소에 근거하여 변비(constipation)를 정의하고 있는데, 연구에서 사용되는 객관적인 정의로서는 평균적으로 주당 3회 미만의 배변이다 (DiPalma 등, 2007). 이와 같은 정의는 배변치 못한 일수가 주당 3일이상 지속되는 것과 동일하다. 본 논문에서는 4주에 걸친 연속된 변비증상의 심각성 점수(severity score)를 각 환자에게서 3일 이상 지속된 변비 일수의 합으로 제시한다. 3일 이상 배변치 못한 경우가 없는 환자는 변비증상이 없는 환자로 간주된다. 변비에 대한 처방의 한 가지 방법으로 식이섬유가 효과적임이 입증되었다 (Burkitt 등, 1972; Cummings, 1984). Yang (1992)의 연구에서 대조군은 수분섭취를 제외하고는 변비에 대해 아무런 처방을 받지 않은 군이며, 치료 I 군은 매일 일정한 수분섭취와 15분간의 운동을 처방받았고, 치료 II 군은 I 군과 동일한 처방에 덧붙여 식이섬유를 처방받았다. 대조군에 비해 다른 두 군, 즉 치료 I 군, 치료 II 군의 순서로 배변이 수월한가를 검정하고자 한다. 대조군에는 20명의 환자가, 치료 I 군과 II 군에는 각각 18명의 환자가 할당되었다. 표 3.1에 환자들의 4주간 변비증상의 심각성 점수가 제시되었고, 0점을 가진 환자는 변비증상이 없는 환자이다. 분석의 목적은 대조군과 비교하여 치료 I 군, 치료 II 군의 순서로 배변이 수월한가를 검정하는 것이며 이를 다시 표현하면, 세 군에 걸쳐 변비증상의 심각성 점수가 감소하는가에 대한 검정이면서 동시에, 전체 환자 중 변비증상이 있는 환자 비율이 감소하는가의 검정이다. 표 3.1의 자료를 살펴보면 대조군, 치료 I 군, 치료 II 군의 순서로 변비증상이 있는 환자의 비율이 85%, 50%, 17%로 감소하고, 또한 변비증상을 보인 환자의 심각성 점수가 감소함을 볼 수 있다. 추세검정 통계량으로 구체적인 유의성을 알아보게 된다. 표 3.1의 자료를 이용하여 앞에서 언급했던 Cochran-Armitage, Jonckheere-Terpstra 검정법과 병합 추세검정법으로 검정한 결과가 표 3.2에 제시되었다. 검정결과를 살펴보면 세가지 방법 모두 매우 유의하다.

표 **3.1.** 세 군의 변비증상 심각성 점수

대조군	치료 I	치료 II
17	12	0
8	5	0
10	7	0
0*	0	0
18	0	0
15	0	0
22	0	0
6	0	0
5	3	4
0	6	0
8	8	4
14	0	0
12	3	0
12	6	0
0	0	3
14	3	0
6	0	0
17	0	0
17		
17		

∗ 0은 변비 증상이 없는 환자

표 **3.2.** 예제자료 분석결과

	대조군	치료 I	치료 II	Z	p값
l_i	3	9	15		
n_i	17	9	3		
Cochran-Armitage 검정				4.2128	$< .0001$
Jonckheere-Terpstra 검정				3.8746	$< .0001$
병합 추세검정				2.2813	0.0113

3.2. 모의실험 계획 및 결과

증상 발생률과 증상 심각성을 병합한 추세검정법의 효율성을 알아보기 위해 모의실험을 실시한다. 우선 (X, Y)의 이변량 연속형 자료를 상관성이 ρ인 이변량 정규분포로 30,000개 생성한다. 첫 단계로 i번째 군에서 $X = 0$의 비율$(1 - p_i)$만큼의 자료를 X의 수치가 낮은 값의 순위에서부터 제외시킨 후 나머지 자료에서 증상 심각성 점수 Y_{ij} $(i = 1, 2, 3,\ j = 1, 2, \ldots, M_i)$를 랜덤추출한다. M_i는 동일수 m으로 두었다. 이 과정을 500회 반복하여 제1종 오류와 검정력을 구한다. 모의실험 모수는 다음과 같다.

$$\rho = 0, 0.5; \qquad (p_1, p_2, p_3) = (0.95, 0.90, 0.85), (0.90, 0.80, 0.70); \qquad m = 15, 30.$$

세 군의 X와 Y의 평균은 귀무가설하에서는 $\mu_i = 0\,(i = 1, 2, 3)$으로 정하고, 대립가설하에서는 $(\mu_1, \mu_2, \mu_3) = (1, 0.5, 0), (0.5, 0.25, 0), (1, 0.2, 0)$으로 정한다. 모의실험은 SAS를 이용하였다.

심사위원의 조언에 따라 Chang 등 (1994)의 두 치료군에 대한 검정통계량의 제 1종 오류를 모의실험으로 알아본 결과, Chang 등 (1994)의 검정통계량은 매우 큰 제 1종 오류를 가진다. 그러나 증상발현율의 추세검정과 증상 심각성의 추세검정 후에 다중검정으로 보정해 줄 때 제 1종 오류가 지정된 5%에

표 **3.3.** Cochran-Armitage 검정에서 귀무가설이 옳은 모의실험 자료에서 제 1종 오류와 검정력

$P(X=1)=0.90$	(μ_1,μ_2,μ_3)	CA	JT	JT+0	New	P_{adj}
	(0, 0, 0)+	0.10	0.04	0.10	0.05	0.05
	(10, 10, 10)+	0.08	0.04	0.09	0.04	0.05
$m=15,\ \rho=0$	(1, 0.5, 0)	0.09	0.80	0.79	0.81	0.73 3
	(0.5, 0.25, 0)	0.11	0.36	0.35	0.38	0.29
	(1, 0.2, 0)	0.10	0.78	0.74	0.79	0.69
	(0, 0, 0)+	0.12	0.05	0.10	0.05	0.08
	(10, 10, 10)+	0.12	0.05	0.13	0.06	0.07
$m=15,\ \rho=0.5$	(1, 0.5, 0)	0.13	0.82	0.78	0.84	0.73
	(0.5, 0.25, 0)	0.11	0.38	0.37	0.39	0.30
	(1, 0.2, 0)	0.11	0.83	0.80	0.83	0.74
	(0, 0, 0)+	0.09	0.04	0.08	0.05	0.05
	(10, 10, 10)+	0.11	0.04	0.09	0.05	0.07
$m=30,\ \rho=0$	(1, 0.5, 0)	0.08	0.98	0.96	0.98	0.95
	(0.5, 0.25, 0)	0.12	0.56	0.55	0.60	0.49
	(1, 0.2, 0)	0.08	0.98	0.97	0.98	0.93
	(0, 0, 0)+	0.10	0.06	0.12	0.08	0.09
	(10, 10, 10)+	0.08	0.07	0.11	0.09	0.07
$m=30,\ \rho=0.5$	(1, 0.5, 0)	0.11	0.98	0.98	0.99	0.97
	(0.5, 0.25, 0)	0.10	0.59	0.56	0.64	0.47
	(1, 0.2, 0)	0.10	0.97	0.95	0.98	0.95
$P(X=1)=0.70$	(μ_1,μ_2,μ_3)	CA	JT	JT+0	New	P_{adj}
	(0, 0, 0)+	0.09	0.05	0.11	0.02	0.06
	(10, 10, 10)+	0.10	0.05	0.11	0.03	0.09
$m=15,\ \rho=0$	(1,0.5,0)	0.11	0.71	0.66	0.57	0.62
	(0.5, 0.25, 0)	0.10	0.32	0.30	0.23	0.26 6
	(1, 0.2, 0)	0.08	0.70	0.54	0.61	0.59
	(0, 0, 0)+	0.11	0.05	0.11	0.02	0.07
	(10, 10, 10)+	0.11	0.06	0.10	0.03	0.09
$m=15,\ \rho=0.5$	(1, 0.5, 0)	0.08	0.76	0.67	0.62	0.66
	(0.5, 0.25, 0)	0.10	0.31	0.28	0.19	0.25
	(1, 0.2, 0)	0.11	0.76	0.63	0.62	0.66
	(0, 0, 0)+	0.10	0.04	0.11	0.03	0.06
	(10, 10, 10)+	0.09	0.07	0.08	0.06	0.07
$m=30,\ \rho=0$	(1, 0.5, 0)	0.09	0.92	0.90	0.91	0.89
	(0.5, 0.25, 0)	0.10	0.41	0.38	0.37	0.31
	(1, 0.2, 0)	0.09	0.93	0.89	0.90	0.88
	(0, 0, 0)+	0.09	0.07	0.10	0.05	0.09
	(10, 10, 10)+	0.11	0.06	0.11	0.04	0.09
$m=30,\ \rho=0.5$	(1, 0.5, 0)	0.10	0.93	0.87	0.91	0.89
	(0.5, 0.25, 0)	0.13	0.49	0.45	0.44	0.42
	(1, 0.2, 0)	0.13	0.96	0.90	0.93	0.93

+ 모든 검정에서 귀무가설인 경우이다.

가까웠다. 따라서 증상발현율과 증상 심각성 모두를 고려한 검정에서 다중검정으로 보정하지 않은 경우와 보정한 경우 모두를 본 논문의 모의실험에 포함시킨다. 다중검정법으로서는 Chang 등 (1994)에 제

표 **3.4.** Jonckheere-Terpstra 검정에서 귀무가설이 옳은 모의실험 자료의 제 1종 오류와 검정력

	(p_1, p_2, p_3)	CA	JT	JT+0	New	P_{adj}
			$\mu = 0$			
$m = 15$	(0.95, 0.90, 0.85)	0.24	0.06	0.11	0.06	0.17
$\rho = 0$	(0.90, 0.80, 0.70)	0.41	0.07	0.11	0.06	0.32
$m = 15$	(0.95, 0.90, 0.85)	0.26	0.03	0.11	0.04	0.12
$\rho = 0.5$	(0.90, 0.80, 0.70)	0.41	0.02	0.11	0.02	0.27
$m = 30$	(0.95, 0.90, 0.85)	0.36	0.03	0.06	0.04	0.26
$\rho = 0$	(0.90, 0.80, 0.70)	0.63	0.05	0.11	0.05	0.51
$m = 30$	(0.95, 0.90, 0.85)	0.35	0.02	0.12	0.03	0.24
$\rho = 0.5$	(0.90, 0.80, 0.70)	0.63	0.01	0.12	0.01	0.49
			$\mu = 0.5$			
$m = 15$	(0.95, 0.90, 0.85)	0.24	0.05	0.11	0.07	0.12
$\rho = 0$	(0.90, 0.80, 0.70)	0.39	0.05	0.13	0.04	0.29
$m = 15$	(0.95, 0.90, 0.85)	0.22	0.03	0.10	0.03	0.10
$\rho = 0.5$	(0.90, 0.80, 0.70)	0.45	0.02	0.13	0.01	0.31
$m = 30$	(0.95, 0.90, 0.85)	0.38	0.05	0.13	0.07	0.28
$\rho = 0$	(0.90, 0.80, 0.70)	0.65	0.06	0.17	0.06	0.54
$m = 30$	(0.95, 0.90, 0.85)	0.32	0.02	0.12	0.02	0.24
$\rho = 0.5$	(0.90, 0.80, 0.70)	0.62	0.02	0.09	0.02	0.48
			$\mu = 1.0$			
$m = 15$	(0.95, 0.90, 0.85)	0.21	0.05	0.12	0.06	0.14
$\rho = 0$	(0.90, 0.80, 0.70)	0.37	0.03	0.14	0.03	0.27
$m = 15$	(0.95, 0.90, 0.85)	0.24	0.03	0.10	0.03	0.13
$\rho = 0.5$	(0.90, 0.80, 0.70)	0.40	0.02	0.13	0.02	0.30
$m = 30$	(0.95, 0.90, 0.85)	0.40	0.04	0.14	0.05	0.28
$\rho = 0$	(0.90, 0.80, 0.70)	0.64	0.04	0.24	0.04	0.50
$m = 30$	(0.95, 0.90, 0.85)	0.35	0.03	0.11	0.04	0.24
$\rho = 0.5$	(0.90, 0.80, 0.70)	0.60	0.01	0.15	0.01	0.47

시된 $P_{adj} = 1 - (1 - p_{min})^2$를 α와 비교하는 방법을 고려한다. 여기서 p_{min}는 증상발현율의 검정으로 구해진 p값과 증상 심각성의 검정으로 구해진 p값의 최소값이다. 따라서 P_{adj}로써 어느 한 검정에서라도 유의하면 전체 검정이 유의하도록 다중비교에 대한 보정과 함께 결론 내린다.

표 3.3을 살펴보면 모든 검정법에서 귀무가설이 옳은 경우의 자료에서의 제 1종 오류는 각 블록의 최상단에 제시되었다. Cochran-Armitage(간략히 CA) 검정법의 제 1종 오류가 지정한 5%보다 높고 증상심각성 점수가 높은 $\mu = (10, 10, 10)$의 경우에 증상 비발현($X = 0$) 수치를 포함하여 검정하는 Jonckheere-Terpstra(간략히 JT+0) 검정법과 병합 추세검정법(간략히 New)의 제 1종 오류는 5%보다 높으나 반면에 Jonckheere-Terpstra(간략히 JT)와 P_{adj}의 제 1종 오류는 지정한 5%에 가깝다. 표 3.3 하단은 CA 검정법에서만 귀무가설이 옳은 경우이며 CA 검정법의 제 1종 오류가 여전히 5%보다 높다. 표 3.4는 JT와 JT+0 검정법에서만 귀무가설이 옳은 경우인데 JT+0 검정법의 제 1종 오류가 높고, JT는 $\rho = 0.5$일 때 제 1종 오류가 매우 낮다. 전반적으로 제 1종 오류가 지정한 5%에 가깝고 검정력이 약간 낮지만 다른 검정법과 비슷한 수준의 검정력을 가진 검정법은 P_{adj} 방법이다. 본 논문에서 고려한 New 방법은 분산 공식 $\mathrm{Var}(J_N)$에서 $E[\mathrm{Var}(J_N \mid n_1, n_2, \ldots, n_K)]$만을 고려하였으며 제 1종 오류가 일관성 없는 경향을 보였다. 한편 $\mathrm{Var}[E(J_N \mid n_1, n_2, \ldots, n_K)]$ 부분까지를 모두 고려한 분산을 이용한 New 방법은 제 1종 오류가 더욱 좋지 않았기에 모의실험 결과를 표에는 제시하지 않는다. 결론적으

로 P_{adj} 방법 즉, 증상발현율 추세의 CA 검정법으로 구한 p값과 증상 심각성 추세의 JT 검정법으로 구한 p값에 다중검정으로 보정해 준 추세검정법이 본 논문에서 고려한 여러 검정법 중에서 가장 바람직하였다.

4. 결론 및 고찰

임상에서 증상 발현의 감소뿐만 아니라 증상 심각성 수치가 감소하는 경우가 흔히 발견된다. 두통에 복용하는 아스피린이나 수두 또는 로터바이러스 설사의 왁씬으로 인한 결과가 그와 같은 예이다. 이 때 증상 발현율만에 근거한 검정이나 또는 증상을 보인 개체만을 대상으로 증상 심각성의 수치만에 근거한 검정보다도 증상 발현율과 증상 심각성 점수 모두의 정보를 이용한 검정을 고려할 수 있으며 이에 대해 본 논문에서 알아보았다. Chang 등 (1994)은 증상 발현과 증상 심각성의 수치가 병합된 검정법을 두 군의 경우에 제안하였고 이 검정법은 다중검정으로 보정이 요구됨을 모의실험으로 확인할 수 있었으며 결과는 표에 제시하지 않는다. 이는 증상 발현의 변동과 증상 심각성의 변동을 모두 감안하여 만약 둘 중 하나에서 감소추세가 있게 되면 치료효능을 반영하는 것이므로 여러 군의 경우에도 이를 이용하여 검정할 수 있다. 본 논문에서 제안하는 병합 추세검정 통계량 New 방법은 제 1종 오류가 조정되지 않기에 사용할 수 없으나 증상 발현과 증상 심각성 각각의 추세검정인 Cochran-Armitage 검정과 Jonckheere-Terpstra 검정으로 분석한 후 다중검정으로 보정하는 P_{adj} 방법이 적절함을 모의실험으로 알 수 있었다. 그러나 장차 두 추세검정의 장점을 더욱 예민하게 감안할 수 있는 검정통계량이 제시되어야 임상의 현실에서 유용하게 사용될 수가 있으며 이러한 검정통계량의 개발이 장래 과제로 남아있다.

참고문헌

Agresti, A. (2002). *Categorical Data Analysis*, John Wiley & Sons, 2nd edition, New York.

Armitage, P. (1955). Tests for linear trends in proportions and frequencies, *Biometrics*, **11**, 375–386.

Burkitt, D. P., Walker, A. R. P. and Painter, N. S. (1972). Effect of dietary fibre on stools and transit-times, and its role in the causation of disease, *The Lancet*, **2**, 1408–1411.

Chang, M. N., Guess, H. A. and Heyse, J. F. (1994). Reduction in burden of illness: A new efficacy measure for prevention trials, *Statistics in Medicine*, **13**, 1807–1814.

Cochran, W. G. (1954). Some methods for strengthening the common chi-squared test, *Biometrics*, **10**, 417–451.

Cummings, J. H. (1984). Constipation, dietary fibre and control of large bowel function, *Postgraduate Medical Journal*, **60**, 811–819.

DiPalma, J. A., Cleveland, M. V., McGowan, J. and Herrera, J. L. (2007). A randomized, multicenter, placebo-controlled trial of polyethylene glycol laxative for chronic treatment of chronic constipation, *The American Journal of Gastroenterology*, **102**, 1436–1441.

Hollander, M. and Wolfe, D. A. (1999). *Nonparametric Statistical Methods*, John Wiley & Sons, 2nd edition, New York.

Jonckheere, A. R. (1954). A distribution-free k-sample test against ordered alternatives, *Biometrika*, **41**, 133–145.

Terpstra, T. J. (1952). The asymptotic normality and consistency of kendall's test against trend when ties are present in one ranking, *Indagationes Mathematicae*, **14**, 327–333.

Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics*, **1**, 80–83.

Yang, S. (1992). Effects of fluid intake, dietary fiber supplement and abdominal muscle exercises on antipsychotic drug-induced constipation in schizophrenics, *Journal of Catholic Medical College*, **45**, 1501–1514.

A Composite Trend Test with Symptom Occurrence and Severity Symptom Scores

Se-Mi Choi[1] · Soo Yang[2] · Hae-Hiang Song[3]

[1]Department of Biostatistics, Graduate School, The Catholic University of Korea
[2]College of Nursing, The Catholic University of Korea
[3]Department of Biostatistics, Graduate School, The Catholic University of Korea

(Received March 2011; accepted October 2011)

Abstract

During clinical trials a researcher is frequently able to observe a disease symptom in a subject as well as a severity score for those who experienced a symptom after a fixed length of treatment. The traditional method to evaluate a decreasing trend in proportion, when there is an intrinsic order in the treatment groups (for example control and two or more treatment groups) is a Cochran-Armitage test, while the method to evaluate a decreasing trend in continuous non-normal data is a Jonckheere-Tersptra test. The Cochran-Armitage test emphasizes the dichotomous data of symptom occurrence and the Jonckheere-Tersptra test emphasizes the continuous non-normal data of severity symptom scores. In this paper we propose new test statistics that consider the combined evidence from a symptom occurrence and disease severity score. We illustrate these methods with example data of schizophrenic inpatients that demonstrated antipsychotic-drug induced constipation. A small-scale simulation is conducted to compare the new trend tests with other trend tests.

Keywords: Trend test, symptom incidence, symptom severity.

[3]Corresponding author: Professor, Department of Biostatistics, Graduate School, Integrated Research Center for Genome Polymorphism, The Catholic University of Korea, 505 Banpo-Dong, Seocho-Gu, Seoul 137-701, Korea. E-mail: hhsong@catholic.ac.kr