응용통계연구 (2010)
23(6), 1047–1056

DOI: 10.5351/KJAS.2010.23.6.1047

임상시험에서의 공변량을 고려한 확률화 방법들의 비교

유아미[1] · 이재원[2]

[1]고려대학교 통계학과, [2]고려대학교 통계학과

(2009년 10월 접수, 2010년 5월 채택)

요 약

임상시험에서 환자들을 각 처리로 할당할 때 반응변수에 영향을 미치는 공변량이 존재하면 공변량도 함께 고려하여 환자들을 랜덤하게 배치하여야 한다. 확률화(randomization) 방법들에는 여러 가지가 있으나 공변량에 따라 환자들을 배치하는 방법으로 층화(stratification)를 많이 사용한다. 층화는 환자들을 공변량에 따라 여러 층으로 나누고 각 층들 안에서 환자들을 랜덤하게 배치하는 방법인데, 공변량의 수가 많아지면 층의 수가 급격하게 늘어나기 때문에 층마다의 환자수가 충분히 많지 않으면 그 결과를 신뢰할 수 없게 된다. 이를 보완하기 위한 방법으로 Pocock과 Simon (1975)은 최소화(minimization) 방법을 제안하였으며 이 방법은 처리에 대한 공변량의 균형을 맞추는 것에 중점을 두었다. 본 논문에서는 현재 가장 많이 쓰이고 있는 확률화 방법들과 최소화 방법의 장단점, 불균형의 정도 및 검정력을 모의실험 연구를 통해 비교해보고자 한다.

주요용어: 확률화 방법, 최소화 방법, 예후인자, 임상시험.

1. 서론

임상시험에서 환자들을 각 처리로 랜덤하게 배치하는 것은 선택편향(selection bias)을 줄이기 위해 반드시 거쳐야 하는 과정이다. 이와 함께 각 처리의 균형을 맞추는 것 또한 임상시험에서 고려되어야 할 성격이다. 하지만 이 둘은 상충되는 문제점이 존재한다. 만일 반응변수에 영향을 주는 공변량(예후인자, prognostic factor)들이 존재할 때에는 이러한 공변량들의 균형도 함께 고려하여 환자들을 각 처리로 배치하는 것이 더 정확한 분석 결과를 이끌어낼 수 있다. 현재 임상시험에서 많이 쓰이고 있는 랜덤화 방법인 순열화블록 확률화 방법(randomized permuted block design)은 처리군들 간의 전체 균형은 잘 맞출 수 있으나 공변량들의 균형을 고려할 수 없는 단점을 가지고 있다. 또한 많이 사용되는 층화된 순열화블록 확률화 방법(stratified random permuted blocks design)은 공변량을 층으로 하여 환자들을 배치하므로 공변량들의 균형을 고려할 수 있으나, 공변량의 수가 많아질수록 층의 수가 급격히 늘어나는 단점을 가진다. 따라서 환자수가 충분히 크지 않으면 환자수가 적거나 한명도 없는 층들이 생기게 되어 분석 결과를 신뢰할 수 없게 된다. Pocock과 Simon (1975)에 의해 제안된 최소화(minimization) 방법은 반응변수에 영향을 주는 공변량의 불균형을 최소화하면서 환자들을 각 처리로 배치하는 방법이다. 특히 최소화 방법은 층화의 단점인 층의 수에 따른 제약을 받지 않는 특징을 가지며, 많은 예후인자들이 동시에 균형을 이루는 유용한 방법이다. Therneau (1993)는 모의실험을 통해서 블록 크기가 2인 층화 방법과 최소화 방법에 대해 예후인자의 수가 증가함에 따른 평균제곱근불균형(root mean square imbalance)을 구하여 두 방법을 비교하였고, 최소화 방법을 사용했을 때 더 많은 인자들을 어려움 없

본 연구는 National Research Foundation of Korea(KRF-2008-313-C00148)에 의해 지원받았습니다.
[2]교신저자: (136-701) 서울특별시 성북구 안암동 5-1, 고려대학교 통계학과, 교수. E-mail: jael@korea.ac.kr

이 고려할 수 있음을 보였다. 또한 Kernan 등 (1999)은 층화 방법에 대한 리뷰논문에서 층이 많아지는 것과 관련한 문제점을 층화 방법의 단점 중의 하나로 지적하고, 층의 수를(처음 계획된 중간분석에서의 환자수)/(블록크기×4) 이하로 유지할 것을 제안하였다. 그 후 Scott 등 (2002)이 여러 논문들을 리뷰하여 보다 많은 예후인자들을 고려할 수 있는 최소화 방법이 다른 확률화 방법들에 비해 불균형을 작게 하는 효과적인 방법임을 확인하였으며, Weir와 Lees (2003)는 블록 크기가 4와 12인 층화 방법과 최소화 방법의 불균형 및 검정력을 비교하고 최소화 방법에서 더 나은 결과를 얻었다. 그리고 Wade 등 (2006)은 최소화 방법을 실제 임상시험에 쉽게 적용할 수 있는 패키지를 개발하기도 하였다. 그러나 최소화 방법이 예후인자와 전체적인 환자수에서의 불균형을 최소화하는 반면 예후인자들의 수준에서는 그렇지 못하다는 주장도 있다 (Tu 등, 2000; Kundt, 2009). 본 논문에서는 실제 임상시험에서 많이 사용되고 있는 블록의 크기가 4와 6인 층화 방법과 최소화 방법의 비교를 고려하였다. 2절에서는 확률화 방법들에 대해 소개하고, 3절에서는 각 예후인자들의 불균형과 검정력을 측정하는 모의실험에 대해 설명한다. 그리고 4절에서는 불균형 정도 및 검정력 등의 모의실험 결과를 비교하고자 한다.

2. 확률화 방법들

2.1. 완전 확률화 방법(Completely Randomized Design)

완전 확률화 방법(completely randomized design)은 T개의 처리가 있을 때 $1/T$의 확률로써 환자들을 각 집단으로 할당하는 가장 기본적인 확률화 방법이다. 즉, 이 방법에서는 각 환자들을 동질적이라고 가정하여 임의로 배치한다. 이 방법을 사용하면 각 집단으로 할당된 환자의 수가 상당히 달라지는 경우가 발생할 수 있다. 특히 임상시험의 규모가 작거나 예후인자의 수가 많으면 불균형이 발생할 위험이 더욱 커지기 때문에 실제로는 잘 쓰이지 않는 방법이다. 그러나 이 방법은 환자가 어떤 처리를 받게 될 지 예측할 수 없는 즉, 선택편향을 감소시키는 예측불가능성(unpredictability)의 성격을 잘 만족한다.

2.2. 순열화블록 확률화 방법(Randomized Permuted Block Design)

완전 확률화 방법에서 발생하는 각 집단의 환자수에 대한 불균형은 순열화된 블록(permuted block)들을 사용하여 해소할 수 있다. 순열화블록 확률화 방법(randomized permuted block design)은 각 집단으로 같은 수의 환자들을 배치할 수 있는 방법이다. 만일 T개의 처리가 있고 블록의 길이가 bT라면, 각 블록에서는 T개의 처리에 대해 각각 b명씩의 환자들을 할당하게 된다. 이러한 처리할당이 모두 이루어지면 각 처리로 할당된 전체 환자수가 같아진다. 예를 들어 100명의 환자들을 길이가 4인 순열화된 블록들을 사용하여 처리군$(=1)$과 대조군$(=0)$으로 할당한다고 하자. 이 때 순열화된 블록들은 6개가 되고$(\{0,0,1,1\}, \{0,1,0,1\}, \{0,1,1,0\}, \{1,0,0,1\}, \{1,0,1,0\}, \{1,1,0,0\})$, 이 중 블록 하나를 랜덤하게 선택하여 4명의 환자들을 각 처리로 2명씩 배치한다. 이러한 과정을 25번 반복하면 환자들을 처리군과 대조군으로 각각 50명씩 할당할 수 있다. 그러나 이 방법은 환자수와 블록의 길이에 따라 각 블록들의 끝에서 선택편향이 발생할 수 있고 예후인자들의 균형을 고려할 수 없는 단점이 있다 (이재원 등, 2005). 본 논문에서는 블록의 길이가 4인 경우와 6인 경우를 다루었다.

2.3. 층화된 순열화블록 확률화 방법(Stratified Randomized Permuted Block Design)

임상시험에서 반응변수에 영향을 주는 예후인자들이 고려될 때에는 예후인자들의 수준(level)에 따라 여러 층을 나누고 각각의 층 안에서 환자들을 처리집단으로 랜덤하게 배치해야 한다. 이 때 사용되는 방법이 층화된 순열화블록 확률화 방법(stratified randomized permuted block design)이며 현재 임상시험에서 많이 쓰이고 있는 방법이다. 예후인자들의 수준에 따라 층을 나누고 각각의 층 안에서 확률화 방법

표 **2.1.** 최소화 방법에서의 불균형 측정

예후인자		처리 1	처리 2	불균형
성별	남	15	16	1
	여	16	14	2
나이	60세 이상	16	15	1
	60세 미만	15	15	0
중증도	중증	9	10	1
	중등증	10	10	0
	경증	12	10	2

을 사용하여 환자들을 배치하게 된다. 그러나 이 방법은 예후인자의 수가 증가할수록 층의 수가 급격하게 늘어나는 단점을 가진다. 따라서 환자수에 비해 층의 수가 많은 경우에 이 방법을 사용하면 처리군간 심각한 불균형이 발생할 수 있다. 본 논문에서는 환자들의 층을 나눈 후 순열화된 블록을 사용하였으며, 가장 흔한 경우인 블록의 길이가 4인 경우와 6인 경우를 다루었다.

2.4. 최소화 방법(Minimization Design)

Pocock과 Simon (1975)의 최소화 방법은 수정된 층화방법(adaptive stratified design)이라고도 한다. 이 방법은 새로운 환자가 들어오면 이미 임상시험에 참여하고 있는 환자들의 예후인자들에 대한 분포를 계산한 후, 처리군과 대조군의 불균형을 최소화하는 집단으로 배치될 확률을 높게 하여 새로운 환자를 배치한다. 최소화 방법은 많은 예후인자들이 동시에 균형을 이룰 수 있는 유용한 방법이며, 이 방법을 이해하기 위해서는 몇 가지 개념들을 정의해야 한다. 먼저 어떤 주어진 예후인자 $i(=1,\ldots,n)$의 수준에 대한 변동의 양(amount of variation)을 측정하기 위한 함수 d_i와 이를 모든 예후인자에 대해 더해준 총 불균형(overall imbalance)의 측도 $G(=\sum_{i=1}^{I} d_i)$, 그리고 p_1이 전체 불균형을 가장 작게 하는 처리로 할당하는 확률일 때 K개의 처리에 대한 할당 확률 $\{p_k\}$를 정의할 수 있다. K개의 처리군이 있을 때 새로 배정되는 환자의 예후인자 에서의 해당하는 수준에서 각 처리군에 할당된 환자수를 $N_{ik}(i=1,\ldots,I,\ k=1,\ldots,K)$라고 하면 두 개의 N_{ik}들을 짝지어 그 둘의 차이에 대한 절대값들 중 가장 큰 값을 d_i로 한다. 만일 처리의 수가 2이면 $d_i=|N_{i1}-N_{i2}|$, 처리의 수가 3이면 $d_i=\max(N_{i1},N_{i2},N_{i3})-\min(N_{i1},N_{i2},N_{i3})$이다. 이 때 반응변수에 더욱 중요한 영향을 주는 예후인자가 있다면 가중합을 사용할 수도 있다. $\{p_k\}$는 전체불균형 G를 작게 하는 처리군으로 환자를 할당하기 위한 확률로 다음의 조건을 만족한다.

$$p_1 \geq p_2 \geq \cdots \geq p_K \quad \text{그리고} \quad \sum_{k=1}^{K} p_k = 1.$$

예를 들어 처리의 수가 3이면 p_1은 G의 값을 가장 작게 하는 처리군으로 환자를 할당할 확률이며, p_2는 그 다음으로 G의 값을 작게 하는 처리군으로 환자를 할당할 확률, p_3는 G의 가장 크게 하는 처리군으로 환자를 할당할 확률이다. $\{p_k\}$가 $1/K$의 값을 가지면 각각의 처리들이 선택될 확률은 같아지고, p_1이 1이면 환자들은 자동적으로 전체불균형을 가장 작게 하는 처리군으로 할당된다.

표 2.1의 예제를 이용하여 불균형을 계산해보면 다음과 같다. 처리의 수가 2이고 예후인자의 수는 3이며, 예후인자의 수준은 각각 2, 2, 3이다. 61명의 환자가 이미 임상시험에 참여하고 있고 예후인자의 분포가 표 2.1과 같다. 62번째의 새로운 환자가 임상시험에 들어온다고 하자. 이 환자는 60세 이상의 여성이며 중증도가 중증이다. 변동의 양 d_i는 $|N_{i1}-N_{i2}|$이고 G는 d_i들을 더하면 다음과 같다. 새로운 환자를 처리 1로 할당하면 $G=5(=|17-14|+|17-15|+|10-10|)$가 되고 처리 2로 할당

표 **2.2.** 교호작용이 있을 때 최소화 방법에서의 불균형 측정

예후인자			처리 1	처리 2	불균형
성별*나이	남	60세 이상	8	9	1
		60세 미만	7	7	0
	여	60세 이상	8	6	2
		60세 미만	8	8	0
중증도		중증	9	10	1
		중등증	10	10	0
		경증	12	10	2

하면 $G = 3(= |16 - 15| + |16 - 16| + |9 - 11|)$이 된다. 따라서 새로운 환자를 처리 2로 할당하였을 때 G값이 더 작음을 알 수 있다. 이 때 p_1이 0.75라면 새로운 환자는 0.75의 확률로써 처리 2로 할당된다. 만일 중증도에 대한 가중치가 3이고 성별과 나이의 가중치가 1이라면, 새로운 환자를 처리 1로 할당할 경우 $G = 5(= 1 \times |17 - 14| + 1 \times |17 - 15| + 3 \times |10 - 10|)$이고 처리 2로 할당할 경우 $G = 7(= 1 \times |16 - 15| + 1 \times |16 - 16| + 3 \times |9 - 11|)$이므로 새로운 환자는 0.75의 확률로써 처리 1로 할당될 것이다. 또한 계산 결과 G의 값이 모두 같게 나왔다면 새로운 환자는 0.5의 확률로써 각 처리로 배치되며 이 경우에는 완전 확률화 방법(completely randomized design)과 같아진다. 예후인자 간에 교호작용이 존재한다면 표 2.2와 같이 교호작용이 있는 예후인자들을 조합하여 층으로 나누고 위와 같은 방법으로 계산하면 된다.

Pocock과 Simon (1975)은 처리 배치가 완전히 끝난 후 불균형을 측정하였다. 처리의 수가 2, 어떤 예후인자의 수준이 2인 경우를 고려하고 X_{kj}를 예후인자의 수준 $j(= 1, 2)$에서 처리 $k(= 1, 2)$로 할당된 환자의 수로 정의하자. 그 예후인자에서의 불균형 수준(level of imbalance)은 첫 번째 수준에서는 $q_1 = X_{11}/(X_{11} + X_{21})$, 두 번째 수준에서는 $q_2 = X_{12}/(X_{12} + X_{22})$와 같이 측정되며, 이 예후인자의 처리 불균형의 양(amount of treatment imbalance)은 $|q_1 - q_2|$으로 정의한다. Weir와 Lees (2003)는 p_1의 값을 0.85, 0.90, 0.95, 1로 변화시키면서 불균형의 수준을 측정하였는데, 결과를 간결하게 표현하고 두 개 이상의 수준을 갖는 인자들로 확장하기 위해서 처리 불균형의 양을

$$100 \times \max\{|q_i - q_j|\}, \quad i \neq j$$

와 같이 측정하였다. 본 연구에서도 Weir와 Lees (2003)의 방법과 같이 p_1의 값을 변화시키면서 처리 불균형의 양에 100을 곱하여 불균형을 측정하였다. 또한 불균형의 총량(total amount of imbalance)은 전체 환자 중 처리 1에 할당된 환자수와 처리 2에 할당된 환자수의 차이로 측정한다.

3. 모의실험

이 절에서는 완전 확률화 방법, 순열화블록 확률화 방법(블록 크기 4, 6), 층화된 순열화블록 확률화 방법(블록 크기 4, 6), 최소화 방법 등 여섯 가지 방법들의 불균형과 검정력을 비교하기 위하여 모의실험을 실시하였다. 모의실험에 사용된 데이터는 환자수를 1000명으로 하여 생성되었고 처리집단은 두 집단(Placebo, Active)이며, 예후인자는 2수준을 갖는 예후인자 2개(A, B)와 4수준을 갖는 예후인자 1개(C)로서 모두 범주형의 형태를 갖는 변수들이다. 표 3.1은 예후인자들의 분포이다. 모의실험에 사용된 최소화 알고리즘은 예제에서 사용한 방법들과 같다. 변동의 양인 d_i의 값은 $|N_{i1} - N_{i2}|$이고 d_i들을 더하여 G를 측정하였다. 여러 방법들의 불균형을 비교할 때에는 최소화의 p_1값을 0.95로 하였고, p_1을 0.80, 0.90, 0.95, 1로 변화시켜가며 다시 비교하였다.

표 3.1. 예후인자들의 분포 ($N = 1000$)

예후인자	수준	빈도	백분율(%)
A	1	631	63.1
	2	369	36.9
B	1	556	55.6
	2	444	44.4
C	1	310	31.0
	2	340	34.0
	3	229	22.9
	4	121	12.1

표 3.2. 처리 불균형의 양

예후인자	확률화 방법	평균	표준편차	Q_1	M	Q_3
A	완전 확률화 방법	2.6286	1.9677	1.0737	2.2196	3.7631
	순열화블록 확률화(블록크기4)	2.6185	1.9765	1.0737	2.3621	3.6506
	순열화블록 확률화(블록크기6)	2.6113	1.9611	1.0737	2.2037	3.6506
	층화된 순열화블록(블록크기4)	0.3231	0.2412	0.1022	0.2607	0.4857
	층화된 순열화블록(블록크기6)	0.3703	0.2746	0.1688	0.3273	0.5317
	최소화 방법	0.1852	0.0985	0.0563	0.2147	0.2147
B	완전 확률화 방법	2.5389	1.9318	0.9933	2.1178	3.6879
	순열화블록 확률화(블록크기4)	2.4741	1.8735	0.8102	2.0254	3.6457
	순열화블록 확률화(블록크기6)	2.5159	1.8921	1.0354	2.0708	3.6457
	층화된 순열화블록(블록크기4)	0.2748	0.2112	0.1345	0.2252	0.4051
	층화된 순열화블록(블록크기6)	0.3223	0.2478	0.1345	0.2706	0.4506
	최소화 방법	0.1078	0.1495	0.0000	0.0000	0.1799
C	완전 확률화 방법	6.9983	3.1039	4.6867	6.6620	8.9013
	순열화블록 확률화(블록크기4)	7.1279	3.1682	4.8576	6.7094	9.0368
	순열화블록 확률화(블록크기6)	7.0299	3.1642	4.6598	6.6513	8.9899
	층화된 순열화블록(블록크기4)	0.8822	0.3908	0.6316	0.7358	1.0682
	층화된 순열화블록(블록크기6)	0.8473	0.3260	0.6316	0.7358	0.2433
	최소화 방법	0.7173	0.2861	0.6316	0.6316	0.7073

모든 불균형의 측정은 10,000번의 처리할당을 반복함으로써 이루어졌다. 환자수에 따른 처리의 가능한 순열의 개수는 200명일 경우 2^{200}개, 1000명일 경우에는 무려 2^{1000}개의 경우의 수가 존재하므로 10,000번도 극히 적은 횟수라 할 수 있다. 2개의 수준을 갖는 예후인자 A, B는 $100 \times |q_1 - q_2|$으로 처리 불균형의 양을 측정하였고 3개의 수준을 갖는 예후인자 C는 $100 \times \max\{|q_i - q_j|\}\ (i \neq j)$으로 처리 불균형의 양을 측정하였다. 불균형의 총량(total amount of imbalance)은 전체 환자 중 처리 1에 할당된 환자수와 처리 2에 할당된 환자수의 차이로 측정되었으며, 모든 불균형의 분포는 평균, 표준편차, 사분위수로 나타내었다.

3.1. 불균형 측정

각 예후인자 별 처리 불균형의 양(amount of treatment imbalance)을 측정한 결과가 표 3.2에 정리되어 있다. 세 예후인자 A, B, C 모두 최소화 방법을 사용했을 때 불균형이 가장 작게 나타났다. 완전 확률화 방법에서 불균형이 가장 컸고, 그 다음 순열화블록 확률화 방법, 층화된 순열화블록 확률화 방법, 최소화 방법의 순이었다. 층화된 순열화블록 확률화 방법의 경우에는 블록의 길이가 4일 때가 6일 때

표 **3.3.** 불균형의 총량

확률화 방법	불균형의 총량(명)				
	평균	표준편차	Q_1	M	Q_3
완전 확률화 방법	120.8238	44.3496	90	114	146
순열화블록 확률화(블록크기4)	91.7364	32.1345	68	90	112
순열화블록 확률화(블록크기6)	91.7506	32.3367	68	88	112
층화된 순열화블록(블록크기4)	13.3704	4.9539	10	12	16
층화된 순열화블록(블록크기6)	14.9456	5.4176	12	14	18
최소화 방법	6.4836	2.2841	4	6	8

표 **3.4.** p_1값에 따른 최소화 방법의 처리 불균형의 양

예후인자	p_1	처리 불균형의 양				
		평균	표준편차	Q_1	M	Q_3
A	0.80	0.2390	0.1913	0.0563	0.2147	0.2147
	0.90	0.1906	0.1203	0.0563	0.2147	0.2147
	0.95	0.1852	0.0985	0.0563	0.2147	0.2147
	1.00	0.1874	0.0700	0.2147	0.2147	0.2147
B	0.80	0.1935	0.2001	0.0000	0.1799	0.2252
	0.90	0.1318	0.1622	0.0000	0.0000	0.2252
	0.95	0.1078	0.1495	0.0000	0.0000	0.1799
	1.00	0.0644	0.1233	0.0000	0.0000	0.1799
C	0.80	1.1081	0.6725	0.6316	0.7358	1.5049
	0.90	0.8064	0.3935	0.6316	0.7073	0.7358
	0.95	0.7173	0.2861	0.6316	0.6316	0.7073
	1.00	0.6491	0.1930	0.6316	0.6316	0.6316

표 **3.5.** 불균형의 총량

p_1	불균형의 총량(명)				
	평균	표준편차	Q_1	M	Q_3
0.80	10.556	4.1657	8	10	12
0.90	7.4830	2.7939	6	8	10
0.95	6.4836	2.2841	4	6	8
1.00	5.2802	1.7208	4	4	6

보다 불균형이 작았다. 예후인자 A와 B의 처리 불균형의 양은 비슷하였으나 C의 처리 불균형의 양은 A, B의 처리불균형의 양보다 크게 나타났는데, 이는 수준의 수가 A, B보다 많기 때문인 것으로 보인다. 표 3.3은 불균형의 총량(total amount of imbalance)을 측정한 결과로 환자수의 차이를 나타낸다. 불균형의 총량을 보면 완전 확률화 방법의 경우 평균적으로 약 121명의 불균형이 발생하였고 순열화블록 확률화 방법도 약 92명의 불균형이 발생하였다. 층화된 순열화블록 확률화 방법에서는 훨씬 줄어든 13, 15명의 불균형이 발생하였으나 최소화 방법에서는 평균 6명의 불균형이 발생하여 6가지 방법들 중 가장 균형을 이루고 있음을 알 수 있다.

3.2. p_1의 값에 따른 최소화 방법의 불균형 비교

표 3.4와 3.5는 전체 불균형을 가장 작게 하는 처리로 할당하는 확률인 p_1의 값을 0.80, 0.90, 0.95, 1로 변화시켜 가면서 최소화 방법의 처리 불균형의 양과 불균형의 총량을 측정하여 비교한 결과이다. 불

표 3.6. 검정력 비교

N	확률화 방법	Alpha	Power
150	완전 확률화 방법	0.0406	0.4476
	순열화블록 확률화(블록크기4)	0.0423	0.4546
	순열화블록 확률화(블록크기6)	0.0412	0.4517
	층화된 순열화블록(블록크기4)	0.0454	0.4611
	층화된 순열화블록(블록크기6)	0.0447	0.4629
	최소화 방법	0.0467	0.4657
200	완전 확률화 방법	0.0463	0.5993
	순열화블록 확률화(블록크기4)	0.0472	0.5986
	순열화블록 확률화(블록크기6)	0.0482	0.5990
	층화된 순열화블록(블록크기4)	0.0483	0.6069
	층화된 순열화블록(블록크기6)	0.0497	0.6087
	최소화 방법	0.0505	0.6079
500	완전 확률화 방법	0.0484	0.9456
	순열화블록 확률화(블록크기4)	0.0498	0.9443
	순열화블록 확률화(블록크기6)	0.0506	0.9455
	층화된 순열화블록(블록크기4)	0.0478	0.9490
	층화된 순열화블록(블록크기6)	0.0491	0.9472
	최소화 방법	0.0472	0.9526
1000	완전 확률화 방법	0.0498	0.9997
	순열화블록 확률화(블록크기4)	0.0535	0.9997
	순열화블록 확률화(블록크기6)	0.0551	0.9993
	층화된 순열화블록(블록크기4)	0.0529	0.9995
	층화된 순열화블록(블록크기6)	0.0529	0.9997
	최소화 방법	0.0524	0.9995

균형의 수준과 불균형의 총량 모두 p_1의 값이 1에 가까워질수록 감소하고 있음을 알 수 있다. 이러한 이유는 p_1이 1이면 자동적으로 전체 불균형 G가 가장 작게 되는 처리군으로 새로운 환자를 할당하게 되는 반면, p_1이 0.80이면 G가 가장 작게 되는 처리군으로 새로운 환자를 할당할 확률이 0.80, 다른 처리군으로 할당할 확률이 0.20이 되므로 불균형이 커지게 된다. 그러나 p_1의 값이 1이면 예측불가능성(unpredictability)이 떨어지기 때문에 Weir와 Lees (2003)는 예측불가능성이 존재하고 $p_1 = 1$일 때의 불균형과 유의한 차이가 없는 $p_1 = 0.95$가 가장 적절하다고 결론지었다.

3.3. 검정력 비교

각 확률화 방법들의 검정력을 비교하기 위해 연속형 반응변수를 생성하여 5,000번의 공분산분석을 실시하였다. 표본 크기가 150, 200, 500, 1000일 때 각 방법들의 검정력을 계산하였으며 유의수준은 0.05, 유의수준의 95% 신뢰구간은 [0.044, 0.056]이다. 표 3.6은 6가지 방법들의 검정력을 비교한 결과이다. 표본 크기가 150인 경우를 제외하고 모두 제 1종 오류가 유의수준 0.05의 95% 신뢰구간 안에 포함되었으며, 6가지 방법의 검정력이 대체로 비슷하였으나 표본크기가 작을 때 최소화 방법의 검정력이 약간 높았다. 표본 크기가 150일 때에는 층화된 순열화블록 확률화 방법과 최소화 방법에서만 유의수준을 만족하였고 그 중 최소화 방법의 검정력이 높게 나타났다.

4. 결론

본 논문에서는 임상시험에서 자주 사용되고 있는 확률화 방법들에 대해 모의실험을 통하여 불균형과 검정력을 비교하였다. 모의실험 시행 결과 완전 확률화 방법, 순열화블록 확률화 방법, 층화된 순열화블록 확률화 방법, 최소화 방법의 순으로 불균형이 감소하였다. 특히 현재 임상시험에서 많이 사용하고 있는 순열화블록 확률화 방법은 불균형 면에서 매우 좋지 않은 결과를 보였고, 층화된 순열화블록 확률화 방법은 층화를 시도하지 않은 두 방법에 비해 불균형이 크게 감소하였는데 블록의 길이가 4일 때가 6일 때보다 불균형이 작았다. 불균형은 최소화 방법을 사용할 때 가장 작았고 제 1사분위수에서도 0인 경우가 많았다. 3절의 모의실험에서 사용했던 예제에서 하나의 예후인자(C)는 불균형이 크게 나타나는 경향을 보였는데, 그 이유는 수준의 수가 많고 각 수준에 할당된 환자수가 홀수인 경우가 많았기 때문인 것으로 보인다. 표본의 크기가 200일 때와 500일 때도 같은 방법으로 불균형을 측정하였는데 모든 방법에서 표본의 크기가 1,000일 때보다 불균형이 증가하였으나 같은 패턴의 결과를 보였으므로 본 논문에서는 결과를 제시하지 않았다. 또한 최소화 방법 사용 시에 전체 불균형을 작게 하면서 동시에 환자 배치에 대한 예측불가능성(unpredictability)도 존재해야 하므로 p_1의 값을 적절하게 선택하는 것이 중요하다. 공분산분석을 이용하여 처리효과에 대한 검정력을 비교해본 결과, 표본이 작을 때 층화된 순열화블록 확률화 방법이나 최소화 방법에서 약간의 검정력 증가가 있었음을 알 수 있었다. 최소화 방법의 단점으로 지적되는 결정론적(deterministic)인 성격은 분석 시 사용되는 통계적 검정 방법들이 환자들의 랜덤 배치 가정과 상충되어 논쟁을 일으켜왔다. 그리고 다음 환자의 처리할당이 예측될 수 있는 것도 단점으로 지적된다. 그러나 Scott 등 (2002)는 이에 대해 이러한 단점들은 층화와 같은 다른 방법들에도 해당될 수 있기 때문에 최소화 방법에만 가중되어서는 안 된다고 하였고, McEntegart (2003)는 다기관 임상시험의 경우 각 기관의 연구자만이 자신의 기관에 들어온 환자들에 대한 지식을 갖게 되는 것이므로 최소화 방법을 사용하는 데 결정론적인 성격이 문제되지 않는다고 주장한 몇 개의 논문들을 제시하였다. 또한 그는 p_1의 값을 0.90으로 사용하거나 Weir와 Lees (2003)의 제안과 같이 0.95로 사용하는 것이 적절하다고 기술하였다. 최소화 방법이 사용된 임상시험의 결과를 임의화 검정법(randomization test)으로 분석해야 한다는 주장 (Grönbladh와 Guilbaud, 2004)이 있으며, 이를 확인하기 위해 모의실험 결과에 대해 임의화 검정을 실시하여 검정력을 비교해 보았으나, 대부분의 경우 유의수준이 0.05의 95% 신뢰구간인 [0.044, 0.056] 안에 포함되지 않아 결과를 제시하지는 않았다. Simon (1979)은 논문들을 리뷰하면서 최소화 방법이 적용된 경우 공분산분석을 실시하면 처리 효과 차이가 크지 않더라도 검정력을 높일 수 있다고 하였다. 임상시험에서 환자들을 각 처리로 할당할 때 반응변수에 영향을 미치는 공변량도 함께 고려하여 환자들을 랜덤하게 배치하여야 좋은 결과를 기대할 수 있다. 현재 국내 임상시험에서는 순열화블록 확률화 방법이나 층화된 순열화블록 확률화 방법을 가장 많이 사용하고 있으나, 순열화블록 확률화 방법은 환자 할당 시 공변량을 고려하지 못하고 있고 층화된 순열화블록 확률화 방법은 공변량을 고려하고 있지만 공변량의 수가 많아서 환자수에 비해 층이 많아지면 뚜렷한 해결책이 없는 실정이다. 불균형은 환자수가 적을수록 공변량의 수가 많을수록 증가하며 이는 국내 임상시험에서 자주 있는 경우가 된다. 하지만 그 동안 국내 임상시험에서 최소화 방법을 거의 사용하지 않고 있었던 이유는 그 복잡성에 비해 불균형의 정도가 크게 향상될 것 같지 않아서 비효율적일 것이라는 선입견이 지배적이었기 때문이다. 그러나 본 연구를 통하여 최소화 방법은 사용하기에 약간 복잡한 방법이지만, 층을 나누지 않고도 공변량들의 균형을 동시에 맞출 수 있고 전체 불균형 또한 최소화할 수 있는 방법임을 확인할 수 있었다. 따라서 앞으로는 국내 임상시험에서도 환자수에 비해 공변량의 수가 많은 경우에 최소화 방법을 보다 많이 사용할 것을 제안하며 이를 통해 보다 나은 결과를 얻을 수 있을 것이라 생각한다.

참고문헌

이재원, 박미라, 유한나 (2005). <생명과학연구를 위한 통계적 방법>, 자유아카데미, 서울.

Grönbladh, A. and Guilbaud, O. (2004). The method of minimization proposed by Pocock and Simon - Properties with regards to balance and inferential validity, Uppsala University

Kernan, W. N., Viscoli, C. M., Makuch, R. W., Brass, L. M. and Horwitz, R. I. (1999). Stratified randomization for clinical trials, *Journal of Clinical Epidemiology*, **52**, 19–26.

Kundt, G. (2009). Comparative evaluation of balancing properties of stratified randomization procedures, *Methods of Information in Medicine*, **48**, 129–134.

McEntegart, D. J. (2003). The pursuit of balance using stratified and dynamic randomization techniques: an overview, *Drug Information Journal*, **37**, 293–308.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, *Biometrics*, **31**, 103–115.

Scott, N. W., McPherson, G. C., Ramsay, C. R. and Campbell, M. K. (2002). The method of minimization for allocation to clinical trials: a review, *Controlled Clinical Trials*, **23**, 662–674.

Simon, R. (1979). Restricted randomization designs in clinical trials, *Biometrics*, **35**, 503–512.

Therneau, T. M. (1993). How many stratification factors is "too many" to use in a randomization plan?, *Control Clinical Trial*, **14**, 98–108.

Tu, D., Shalay, K. and Pater, J. (2000). Adjustment of treatment effect for covariates in clinical trials: Statistical and regulatory issues, *Drug Information Journal*, **34**, 511–523.

Wade, A., Pan, H., Eaton, S., Pierro, A. and Ong, E. (2006). An investigation of minimisation criteria, *BMC Medical Research Methodology*, **6**, 1471–2288.

Weir, C. J. and Lees, K. R. (2003). Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial, *Statistics in Medicine*, **22**, 705–726.

Comparing the Randomization Methods Considering the Covariates in a Clinical Trial

Ami Yu[1] · Jae Won Lee[2]

[1]Department of Statistics, Korea University; [2]Department of Statistics, Korea University

(Received October 2009; accepted May 2010)

Abstract

In clinical trials, patients should be randomly allocated to treatment and control groups that consider the balance of their prognostic factors(covariates). There are many randomization methods and stratification is popular in Korea. In stratification, patients are divided into strata based on covariates and then the patients are randomly assigned to the arms of each strata. If the number of covariates increases then the number of strata increases rapidly and the results may not be reliable when the patients are inadequate in each strata. To complement this problem Pocock and Simon (1975) suggested a new randomization method that called for minimization focusing on the balance of covariates. In this study, we compare the advantages and disadvantages, the imbalance of covariates, the power of minimization, and other randomization methods by simulation.

Keywords: Randomization, minimization, prognostic factor, clinical trial.

This research was supported by National Research Foundation of Korea(KRF-2008-313-C00148).

[2]Corresponding author: Professor, Department of Statistics, Korea University, 5-1 Anam-dong, Sungbuk-Gu, Seoul 136-701, Korea. E-mail: jael@korea.ac.kr