

Comparison of binary data imputation methods in clinical trials

Koosung An^a · Dongjae Kim^{a,1}

^aDepartment of Biomedicine · Health Science, The Catholic University of Korea

(Received March 14, 2016; Revised March 31, 2016; Accepted March 31, 2016)

Abstract

We discussed how to handle missing binary data clinical trials. Patterns of occurring missing data are discussed and introduce missing binary data imputation methods that include the modified method. A simulation is performed by modifying actual data for each method. The condition of this simulation is controlled by a response rate and a missing value rate. We list the simulation results for each method and discussed them at the end of this paper.

Keywords: binary missing data, clinical trial, missing pattern, missing data imputation

1. 서론

임상시험은 의약품의 안전성과 유효성을 증명할 목적으로 해당 약물의 약동, 약력, 약리, 임상적 효과를 확인하고 이상반응을 조사하기 위하여 사람을 대상으로 실시하는 시험 또는 연구이다. 사람을 대상으로 실시하는 연구이기 때문에 세계 각국에서는 엄격한 관리와 감독이 이루어지고 있다. 미국, 일본, 유럽이 주도하는 의약품국제협력조화회의(International Conference on Harmonization; ICH)에서는 여러 원칙들을 제시하고 있는데 이중 ICH E9 guideline에서는 임상시험의 결과 분석과 결론을 위한 여러 통계 원칙에 대하여 설명하고 있다. 여기서 편의(bias)의 방지를 위해 확률화(randomization), 이중 맹검(double blind) 등의 방법을 권고하고 있으며, 처리의향원칙(intention-to-treat; ITT) 또한 중요한 통계 분석의 기본 원칙으로 두고 있다. 이 원칙의 적용은 결측치 처리 문제와도 직결된다. ITT 원칙은 연구 참여자에 대하여 중도탈락 되었다 하더라도 최초에 확률화를 통해 처리군 및 대조군에 무작위 배정된 대로 모든 대상자에 대해 비교분석을 진행해야 한다는 의미이다 (Montori와 Guyatt, 2001). 이 원칙을 따르기 위해서는 무작위 배정된 모든 대상자를 대상으로 완전한 추적을 통해 모든 데이터 수집이 이루어져야 하기 때문에 현실적으로 완벽한 적용에는 어려움이 따른다. 그렇기 때문에 ITT 원칙을 가능한 가깝게 지키면서 무작위 배정된 대상자들 중에서 분석에서 제외되는 대상자 수를 최소화하고, 제외되는 경우에는 그 타당성을 입증한 대상자들의 집합을 full analysis set(FAS)라 한다 (Kang, 2013). FAS와 대조되는 집단을 per protocol set(PPS)라고 하는데, 이는 연구약정(study protocol)을 일정 수준 이상 준수하고 주요 변수의 측정이 모두 이루어졌으며, 주요 약정위배(protocol violation)가 없는 ‘유효한’ 표본집단을 뜻한다. 만약, FAS를 통한 분석은 고려하지 않고, PPS를 대상으로만 분석 및 결

¹Corresponding author: Department of Biomedicine · Health Science, The Catholic University of Korea, 222 Banpo-daero Seocho-gu, Seoul 06591, Korea. E-mail: djkim@catholic.ac.kr

론을 도출하게 되면 비교 가능한 처리집단을 얻기 위한 확률화의 기본이념에 반하는 심각한 편의를 초래할 가능성이 있다 (Park과 Lee, 1995). 가설 검정의 종류에 따라 다르겠지만 대개의 경우 FAS를 통한 분석이 PPS를 통한 분석보다 더 보수적이라고 알려져 있다. 즉, FAS의 분석을 통해 얻어진 약효의 크기는 실제의 것 보다 작아지는 경향이 있다는 것이다 (ICH E9, 1998). PPS를 통한 분석은 위와 반대로 약효를 과대평가하게 되는데, PPS는 연구약정을 잘 지킨 대상자들이므로 시험약의 효과를 보여줄 기회가 더 많을 수 있기 때문이다 (Kang, 2013). 이는 효과가 없는 의약품에 대하여 유효한 약효가 있다고 판단하게 되고 그 약을 환자들이 복용하게 된다면 위험을 초래할 수 있다. 그러므로 임상시험에서는 결측값이 발생했다고 해서 무조건 제거하고 결측값이 없는 완전한 자료만을 가지고 분석을 실시한다면 문제의 소지가 더 발생 할 수 있다 (Kwak, 2005). 이러한 이유들 때문에 ICH에서는 정밀한 실험 설계 또는 수행을 통해 가능한 결측치를 줄이려 권고하고 있으며 발생된 결측치에 대해서는 사전에 작성된 연구 약정서에 명시된 방법으로 결측치를 처리 하도록 하고 있다.

결과적으로 ICH뿐만 아니라 각국의 규제기관에서는 보수적인 상황에서 약물의 효능을 보이길 원하고 있고 이에 따라 FAS를 이용한 분석에서 유의한 결과를 도출해야하는 상황이 다수 발생하게 된다. 따라서 적절한 결측치의 처리 방법의 선택은 정확한 점추정치와 검정력 확보를 위해 더욱 중요하다.

본 논문에서는 결측치가 발생하는 기제에 대하여 살펴보고 이분형 변수에 대한 결측치 처리 방법을 논의하였다. 다음으로 실제 자료에 반복적으로 결측치 대체를 적용해보고 각 방법의 성능 비교 및 문제점에 대한 논의를 통하여 마무리하였다.

2. 결측치의 발생기제 및 처리방법

결측치 처리방법들에 대한 의논에 앞서 결측자료가 어떻게 발생하는지 그 형태들에 대하여 먼저 알아볼 필요가 있다. 결측치의 발생기제에 따라서 처리방법들 각각 다른 결과와 효율을 보일 수 있기 때문이다. 결측자료의 각각 결측치의 발생 형태에 따라 크게 두 가지로 나뉜다. ‘무시할 수 있는 결측치’와 ‘무시할 수 없는 결측치(non-ignorable; NI)’이다. ‘무시할 수 없는 결측치(NI)’는 대개 ‘비임의 결측치(not missing at random; NMAR)’라고 표현한다. 다시 ‘무시할 수 있는 결측치’는 ‘완전 임의 결측치(missing completely at random; MCAR)’와 ‘임의 결측치(missing at random; MAR)’로 나뉠 수 있다 (Rubin, 1987; Little과 Rubin, 2002). 다음에서 완전 임의 결측치, 임의 결측치, 비임의 결측치 순으로 자세히 살펴보겠다.

먼저 완전 임의 결측치(MCAR)를 살펴보겠다. 완전 임의 결측치는 자료 내에서 결측의 발생이 어떠한 변수와도 관계가 없는 경우를 뜻한다. 결측치가 없는 완전한 자료를 $Y = (y_{ij})$ 행렬이라 하고, 결측이 발생한 자료를 $M = (M_{ij})$ 행렬이라 하며, θ 를 미지의 모수라 둔다. 이때 모든 Y 와 θ 에 대한 분포함수는 다음과 같다.

$$f(M|Y, \theta) = f(M|\theta). \quad (2.1)$$

이때, 완전한 자료인 Y 와 결측이 발생한 M 간의 무관함을 알 수 있다. 예를 들어, 가계수입에 대한 설문조사에서 월수입이라는 항목에서 결측이 발생하였다고 하면, 이때 응답자와 무응답자 간에 어떠한 성향 차이도 없다고 한다면 응답자의 월수입에 관한 분포와 무응답자의 월수입에 관한 분포는 같게 될 것이다.

두 번째로, 임의 결측치(MAR)에 대하여 살펴보겠다. 임의 결측치는 어떤 결측치가 발생할 가능성을 관측된 자료에 의해서 예측할 수 있으나 관측된 자료를 통제한다면 결측자료와는 통계적으로 독립적인 경우를 의미한다 (Kim, 2011). 이는 MCAR보다는 상대적으로 덜 강한 가정이다. 식 (2.1)과 같이 식

으로 표현하기 위해 Y_{obs} 를 완전한 자료 Y 에서 결측이 아닌 관측된 자료들에 대한 행렬, Y_{mis} 는 Y 에서 결측이 발생한 원소라고 둔다. 이때 모든 Y_{mis} 와 임의의 모수 θ 에 대하여 분포함수는 다음과 같다.

$$f(M|Y, \theta) = f(M|Y_{obs}, \theta). \quad (2.2)$$

위의 예를 통해 이어서 설명하면, 가계수입에 대한 설문조사에서 월수입, 학력 등의 항목이 있을 때, 일반적으로 높은 월수입을 가진 응답자의 무응답률이 높을 것이다. 여기서 학력에 대한 정보가 주어졌다고 한다면 월수입에 대한 항목의 무응답률은 임의적이다. 즉, 같은 학력을 가진 집단의 월수입에 대한 무응답률은 동일하다.

마지막으로 비임의 결측치(NMAR)가 있다. 이는 결측치의 발생기제가 MCAR도 MAR도 아닌 경우를 뜻한다. 결측치의 발생 분포가 자신의 변수에 영향을 받을 뿐만 아니라 다른 변수들에게도 영향을 받는 구조이다 (Cho, 2008). 이는 결측치 발생기제 중에 가장 복잡한 형태를 나타낸다. 위의 예를 이어서 들자면, 학력에 대한 정보가 있더라도 그 안에서 월수입이 높은 사람이 더 높은 무응답률을 보인다는 것이다. 즉, 월수입에서의 결측치가 발생하면 그 변수 자체에 관련이 있다는 의미이다. 그러므로 월수입의 무응답자에 대해서 학력이 높는지 낮는지 알 수가 없다. 이분형 결측치에 대한 대체(imputation) 방법은 다양하게 존재하며, 대표적인 방법에 대한 소개와 새로운 방법에 대한 제안을 하였다.

2.1. 완전사례분석(complete case analysis; listwise deletion analysis)

결측치가 발생되었을 때, 가장 쉽고 간단하게 이용할 수 있는 방법이다. 모든 변수들이 관측되어진 개체들에 대하여 분석하는 방법이다. 단 하나의 변수에서라도 결측값이 발생하였다면 그 개체는 분석에서 제외된다. 대부분의 통계분석 프로그램에서 이용 가능하므로 간편하고, 표본수가 많고 결측값의 수가 아주 적은 경우 MCAR가정 하에서는 모수에 대한 추정에 있어 편의가 거의 발생하지 않는다는 장점이 있다. 그러나 결측값이 많을 경우 분석 가능한 표본의 수가 감소하게 되고, 이는 곧 정보의 손실이므로 검정력이 약화된다. 또한, MCAR가정이 만족되지 않을 경우 편의가 발생할 수 있다.

2.2. 평균 대체(mean imputation)

결측치가 없이 완전히 측정된 자료들의 평균을 결측값을 대체해서 완전한 자료를 만든 후 분석하는 방법이다. 비조건부 평균대체법과 조건부 평균대체법등이 있다. 이 방법 역시 사용하기 간단하고 완전사례 분석방법보다 효율성이 향상된다. 그러나 결측값이 반복적으로 평균값으로 대체되므로 통계량의 표준 오차가 과소 추정되는 문제가 있다 (Yun, 2004).

2.3. 핫덱 대체(Hot-Deck imputation)

1962년 미국 CPS에서 노동력 항목의 연간소득에 대한 결측값을 대체하는 데 최초로 사용한 방법으로 결측치가 발생한 변수에 대해 그 변수가 가질 수 있는 값들 중에 임의로 하나를 선택하여 대체하는 방법이다. 변수들이 비슷한 값을 갖는 유사한 집단을 구성하고 그중 임의의 한 개체를 선택해 결측값을 대체하는 방법이 보통이다. 가중 핫덱 대체는 변수가 가질 수 있는 값들 중 하나를 선택할 때 가중값을 두어 선정하는 방법이다. 층화가 되어있거나 군집화가 되어 있어 대체 군내의 값들이 서로 다른 추출확률을 갖을 때 유용하게 이용 된다 (Kim, 2000).

2.4. 회귀 대체(regression imputation)

결측값을 가진 변수를 반응변수 Y 로 두고 결측값이 없는 p 개의 다른 변수들을 설명변수 x_p 로 두어 회

귀분석을 통해 회귀계수를 추정한다. 그 후 x_p 에 적절한 값들을 대입하여 계산된 Y 를 결측치 대체값으로 추정한다. 결측이 발생한 변수가 이분형일 경우 다음과 같은 로지스틱 모형을 이용할 수 있다.

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p. \quad (2.3)$$

이러한 일반적인 회귀 대체 또한 평균 대체와 마찬가지로 통계량의 표준오차가 과소 추정되는 문제가 발생된다. 이러한 경우, MSE를 이용해 임의로 추출한 오차를 합하여 대체하게 되면 대체값의 변동을 어느 정도 유지할 수 있게 된다.

2.5. Propensity score matching imputation

임상시험에서는 보통 대상자 모집 초기에 키, 몸무게, 혈압 및 병력, 과거력 및 설문지 등의 기저특성(baseline characteristic)을 측정한다. 이 정보를 이용해 생물학적동등성을 지닌 다른 대상자의 자료를 결측자료에 대체한다면 실제와 어느 정도 비슷한 결과가 나올 것이다. 이런 점을 이용해 결측치를 대체하는 방법을 고안하였다. 결측치 대체 과정은 다음과 같다.

- 1) 결측이 발생한 이분형 변수를 반응변수로 나머지 변수를 설명변수로 두어 식 (2.3)과 같이 로지스틱 회귀분석을 통해 propensity score를 구한다. 이때 stepwise 방법을 통해 유의한 변수만을 사용하거나 모든 변수를 사용하여 스코어를 구할 수 있다.
- 2) Propensity score를 통해 각각의 결측 개체별로 미리 정해놓은 범위에 포함된 정상적인 자료와 1대 n 매칭을 한다. 임상시험자료는 대개 처리군과 대조군 등의 그룹이 존재한다. 이때는 각 그룹 내에서 매칭을 한다.
- 3) n 개의 매칭된 자료 중 1개를 임의 추출하여 그 값으로 결측치를 대체한다. 만약 단 1개의 정상적인 자료와도 매칭되지 못한 결측치에 대해서는 식 (2.3)에서 π 로 대체한다.

3. 모의 실험

질병관리본부에서 주기적으로 진행하는 국민의 건강 행태, 만성질환 유병현황, 식품 및 영양섭취실태에 관한 표본조사인 ‘국민건강영양조사’에 대해 제5기 3차년도 자료 중 2011년도의 자료를 이용하여 모의 실험을 진행하였다.

2011년도 국민건강영양조사 전체 자료 중 모의 실험에 적합한 조건을 가진 자료를 소그룹으로서 추출하여 하나의 원 데이터 셋을 생성하였다. 즉, 이항 값을 갖는 반응 변수(평가 변수)와 설명 변수(baseline characteristic-범주형, 연속형 변수를 모두 포함한다)의 관측값이 모두 존재하는 대상자를 선택한다. 이후 완전 임의 결측의 가정 하에 각각 10%, 30%의 결측 데이터셋을 생성하고, 2절에서 제시한 5가지 방법에 의해 결측치 대체를 수행하였다. 이러한 과정을 1,000회 반복하여 결과를 생성하였다.

위와 같은 모의 실험은 2×2 분할표에서 각 그룹의 반응률이 0에 가까울 경우와 0.5에 가까울 경우 등 두 가지 자료에 대해 실시하였다.

3.1. 반응률이 0.5에 가까운 경우

2011년도 국민건강영양조사 전체 자료에 대해 서울/경기에 거주하는 70세 이상의 대상자 중 임상실험실 검사에 의한 당뇨병 유병여부(정상, 비정상)로 분류)를 반응변수로 두고, 성별을 독립 그룹으로 둔다. 그 외로 만 연령, 고혈압 현재 유병 여부, 이상지혈증 현재 유병 여부, 뇌졸중 현재 유병 여부, 심근

Table 3.1. 2×2 contingency table for origin

정상/비정상 여부, <i>n</i> (%)	남성	여성	합계
정상	67(59.82)	84(54.19)	151(56.55)
비정상	45(40.18)	71(45.81)	116(43.45)
합계	112	155	267

Table 3.2. Simulation result of imputation for 10% missing dataset from Table 3.1

	<i>E</i> (OR)	<i>E</i> (OR)의 95% CI[최소, 최대]	원본 OR과 대체 OR의 평균 차이	OR의 평균 신뢰 구간 길이	원본 OR과 대체 OR의 평균 신뢰 구간 길이 차이
완전 사례 분석	1.2624	[1.2554, 1.2694]	0.0040	1.3760	0.0857
평균 대체	1.2624	[1.2554, 1.2694]	0.0040	1.2952	0.0050
핫택 대체	1.2712	[1.2615, 1.2810]	0.0128	1.3053	0.0150
회귀 대체	1.2601	[1.2535, 1.2667]	0.0016	1.2927	0.0024
PSM (margin = 0.01)	1.2657	[1.2568, 1.2747]	0.0073	1.2996	0.0094
PSM (margin = 0.05)	1.2669	[1.2583, 1.2755]	0.0084	1.3006	0.0104
PSM (margin = 0.10)	1.2637	[1.2547, 1.2726]	0.0052	1.2971	0.0068

PSM = propensity score matching.

Table 3.3. Simulation result of imputation for 30% missing dataset from Table 3.1

	<i>E</i> (OR)	<i>E</i> (OR)의 95% CI[최소, 최대]	원본 OR과 대체 OR의 평균 차이	OR의 평균 신뢰 구간 길이	원본 OR과 대체 OR의 평균 신뢰 구간 길이 차이
완전 사례 분석	1.2732	[1.2602, 1.2862]	0.0147	1.5957	0.3054
평균 대체	1.2732	[1.2602, 1.2862]	0.0147	1.3089	0.0186
핫택 대체	1.2878	[1.2698, 1.3059]	0.0294	1.3268	0.0366
회귀 대체	1.2590	[1.2462, 1.2718]	0.0005	1.2938	0.0035
PSM (margin = 0.01)	1.2692	[1.2527, 1.2858]	0.0108	1.3067	0.0165
PSM (margin = 0.05)	1.2702	[1.2533, 1.2871]	0.0118	1.3078	0.0175
PSM (margin = 0.10)	1.2722	[1.2560, 1.2884]	0.0137	1.3097	0.0195

PSM = propensity score matching.

경색 현재 유병 여부, 협심증 현재 유병 여부, 천식 현재 유병 여부, 우울증 현재 유병 여부, EuroQoL: VAS, 월간음주여부, 현재흡연자 하루 평균 흡연량, 체질량지수 등을 baseline characteristic으로 두었다.

앞서 설명한 변수들의 모든 값이 존재하는 대상자는 총 267명으로 남성 112명, 여성 155명이었고 2×2 분할표를 Table 3.1에 제시하였다. 카이제곱 검정 결과 *p*-value는 0.3599로 나타났고, 오즈비는 1.2585, 오즈비의 95% 신뢰 하한과 상한은 0.7691, 2.0593으로 나타났다. 본 자료를 3장에서 설명한 방법으로 모의실험을 진행한 결과를 Table 3.2와 Table 3.3에 제시하였다.

당뇨병 자료 중 10%를 임의 추출하여 결측을 생성하고 2장에 제시된 각 대체방법으로 1,000회 모의 실험한 결과 모두 유사한 결과를 얻었다. 핫택 대체 오즈비의 경우 원본 data에서의 오즈비와 차이가 0.0128로 가장 큰 결과를 나타냈다. 반면 원본 오즈비와 대체 오즈비의 평균 신뢰구간 길이 차이는 완전 사례 분석법이 0.0857로 가장 크게 나타났고 회귀 대체에서 0.0024로 가장 작게 나타났다. Propensity score matching 방법에서 margin을 0.01, 0.05, 0.10으로 달리한 결과에서는 큰 차이가 나타나지 않았다.

Table 3.4. 2×2 contingency table for original data related with osteoporosis

정상/유병 여부, $n(\%)$	남성	여성	합계
정상	171(96.61)	172(87.76)	343(91.96)
유병	6(3.39)	24(12.24)	30(8.04)
합계	177	196	373

Table 3.5. Simulation result of imputation for 10% missing dataset from Table 3.4

	$E(OR)$	$E(OR)$ 의 95% CI[최소, 최대]	원본 OR과 대체 OR의 평균 차이	OR의 평균 신뢰 구간 길이	원본 OR과 대체 OR의 평균 신뢰 구간 길이 차이
완전 사례 분석	4.0669	[4.0179, 4.1159]	0.0902	9.4327	1.0469
평균 대체	4.0669	[4.0179, 4.1159]	0.0902	8.7989	0.4130
핫택 대체	4.1813	[4.1192, 4.2433]	0.2045	9.2264	0.8405
회귀 대체	4.0683	[4.0193, 4.1173]	0.0915	8.7963	0.4105
PSM (margin = 0.01)	4.0853	[4.0243, 4.1463]	0.1085	8.9135	0.5276
PSM (margin = 0.05)	4.1669	[4.1064, 4.2274]	0.1902	9.1535	0.7677
PSM (margin = 0.10)	4.1547	[4.0908, 4.2187]	0.1780	9.1440	0.7581

PSM = propensity score matching.

Table 3.2에 제시된 모의 실험에서 나머지 조건은 동일하고 결측 생성 비율을 30%로 조정된 결과를 Table 3.3에 제시하였다.

당뇨병 자료 중 30%를 임의 추출하여 결측을 생성하고 2절에 제시된 각 대체방법으로 1,000회 모의 실험한 결과 모두 유사한 결과를 얻었다. 핫택 대체 오즈비의 경우 원본 data에서의 오즈비와 차이가 0.0294로 가장 컸고 이는 당뇨병 자료의 10% 결측에 의한 모의 실험 결과의 것보다 2배 이상 증가했다. 반면 원본 오즈비와 대체 오즈비의 평균 신뢰구간 길이 차이는 완전 사례 분석법이 0.3054로 가장 크게 나타났고, 회귀 대체에서 0.0035로 가장 작게 나타났다.

Propensity score matching 방법에서 margin이 0.01, 0.05, 0.10으로 커질수록 오즈비의 추정치는 원자료의 오즈비와 차이가 커지는 경향이었고, 원본 오즈비와 대체 오즈비의 평균 신뢰구간 길이 차이 또한 증가하는 추세를 보였다.

3.2. 반응률이 0에 가까운 경우

2011년도 국민건강영양조사 전체 자료에 대해 60세 미만의 대상자 중 골다골증 유병여부(정상/유병으로 분류)를 반응변수로 두고, 성별을 독립 그룹으로 두었다. 3.1절의 당뇨병 자료와 마찬가지로 만 연령, 고혈압 현재 유병 여부, 이상지혈증 현재 유병 여부, 뇌졸중 현재 유병 여부, 심근경색 현재 유병 여부, 협심증 현재 유병 여부, 천식 현재 유병 여부, 우울증 현재 유병 여부, EuroQoL: VAS, 월간음주여부, 현재흡연자 하루 평균 흡연량, 체질량지수 등을 baseline characteristic으로 두었다.

앞서 설명한 변수들의 모든 값이 존재하는 대상자는 총 373명으로 남성 177명, 여성 196명이었고 2×2 분할표를 Table 3.4에 제시하였다.

카이제곱 검정 결과 p -value는 0.0017로 나타났고, 오즈비는 3.9767, 오즈비의 95% 신뢰 하한과 상한은 1.5859, 9.9718으로 나타났다.

본 자료를 3장에서 설명한 방법으로 모의 실험을 진행한 결과를 다음 Table 3.5와 Table 3.6에 제시하였다.

Table 3.6. Simulation result of imputation for 30% missing dataset from Table 3.4

	$E(OR)$	$E(OR)$ 의 95% CI[최소, 최대]	원본 OR과 대체 OR의 평균 차이	OR의 평균 신뢰 구간 길이	원본 OR과 대체 OR의 평균 신뢰 구간 길이 차이
완전 사례 분석	4.4374	[4.3083, 4.5666]	0.4607	13.8511	5.4653
평균 대체	4.4374	[4.3083, 4.5666]	0.4607	10.6916	2.3058
핫덱 대체	4.6350	[4.4685, 4.8015]	0.6582	11.9374	3.5516
회귀 대체	4.4533	[4.3230, 4.5836]	0.4766	10.7137	2.3278
PSM (margin = 0.01)	4.5989	[4.4380, 4.7598]	0.6221	11.6752	3.2893
PSM (margin = 0.05)	4.7357	[4.5674, 4.9039]	0.7589	12.2253	3.8395
PSM (margin = 0.10)	4.6317	[4.4611, 4.8022]	0.6549	11.9917	3.6059

PSM = propensity score matching.

골다공증 자료 중 10%를 임의 추출하여 결측을 생성하고 2장에 제시된 각 대체방법으로 1,000회 모의 실험한 결과, 핫덱 대체 오즈비의 경우 원본 data에서의 오즈비와 차이가 0.2045로 가장 컸고 완전 사례 분석법, 평균 대체법에서 0.0902로 나타났다. 회귀 대체법 또한 0.0915로써 상대적으로 작게 나타났다. 반면 원본 오즈비와 대체 오즈비의 평균 신뢰구간 길이 차이는 완전 사례 분석법이 1.0469로 가장 크게 나타났고, 회귀 대체에서 0.4105로 가장 작게 나타났다.

Propensity score matching 방법에서 margin이 0.01일 경우에 비해 0.05와 0.10일 경우 오즈비 추정치의 원자료 오즈비와의 차이가 더 컸고, 원본 오즈비와 대체 오즈비의 평균 신뢰구간 길이 차이 또한 상대적으로 더 크게 나타났다.

골다공증 자료 중 30%를 임의 추출하여 결측을 생성하고 2장에 제시된 각 대체방법으로 1,000회 모의 실험한 결과 propensity score matching의 margin = 0.05에 의한 오즈비 추정치와 원본 data에서의 오즈비간의 차이가 0.7589로 가장 컸고 완전 사례 분석법, 평균 대체법에서 0.4607로 가장 작게 나타났다. 반면 원본 오즈비와 대체 오즈비의 평균 신뢰구간 길이 차이는 평균 대체법과 회귀 대체법이 각각 2.3058, 2.3278로 가장 작았고 완전 사례 분석법에서 5.4653으로 가장 크게 나타났다.

Propensity score matching 방법에서 margin의 변화에 따른 결과는 유사하게 나타났다.

4. 결론 및 고찰

2×2 분할표에서 오즈비가 0.5에 가까울 경우 10%의 결측 발생 및 대체 후 결과비교에서 각 방법에서 큰 차이가 없었다. 다만 결측 비율이 30%일 때, 핫덱 대체법이 다른 방법에 비해 다소간의 편이가 발생하였고 propensity score matching 방법에서 margin이 증가할수록 정확도가 감소하는 경향을 보였다.

2×2 분할표에서 오즈비가 0에 가까울 경우 10%의 결측 생성 및 대체 후 결과에서는 핫덱 대체와 propensity score matching 방법의 정확도가 다른 방법들에 비해 감소하였다. 다만 propensity score matching 방법에서 margin을 0.01로 두었을 때는 회귀 대체와 유사한 결과를 얻을 수 있었다. 결측 비율이 30%일 때는 모든 대체방법이 결측률 10%일때의 결과들에 비해 다소 부정확한 결과를 보였고 propensity score matching 방법의 margin의 변화에 따른 정확도 차이도 크지 않았다.

본 논문에 제시된 결과는 실제 자료의 일부를 발췌하여 진행하였기 때문에 이와 같은 모의실험에는 한계가 존재할 수 있다. 따라서 더욱 많은 자료로 다각도의 모의 실험을 해보거나 특정 조건에 의해 임의 생성한 분할표와 baseline characteristic을 활용하여 추가 연구를 해볼 필요가 있다. 그리고 임상시험은 여러 차례 방문이 이루어지기 때문에 반복측정 자료가 상당수 이고, 이를 고려한 반복측정 이분형자료의

분석에도 각 결측치 처리방법의 성능을 비교하는 연구를 고려할 필요가 있다.

References

- Cho, Y. (2008). An imputation for nonresponses in the survey on the rural living indicators, *The Korean Journal of Applied Statistics*, **21**, 95–107.
- International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use (1998). Statistical principles for clinical trials E9.
- Kang, S. (2013). *Medical Statistics Needed for Drug Development*, 2nd edition, Freeacademy.
- Kim, D. (2011). Solutions for missing values in categorical data, *The Korean Association for Comparative Government*, **15**, 319–342.
- Kim, K. (2000). Imputation methods for nonresponse and their effect, *The Korean Association for Survey Research*, **1**, 1–14.
- Korea Centers for Disease Control and Prevention (2012). Korea National Health and Nutrition Examination Survey V.
- Kwak, M. (2005). Imputation methods for clinical trials data, *Journal of the Korean Data Analysis Society*, **7**, 1187–1198.
- Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*, 2nd edition, John Wiley & Sons, New York.
- Park, M. and Lee, J. (1995). Statistical consideration in clinical trials, *Communications for Statistical Applications and Methods*, **2**, 229–247.
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, John Wiley & Sons, New York.
- Montori, V. M. and Guyatt, G. H. (2001). Intention-to-treat principle, *Canadian Medical Association Journal*, **165**, 1339–1341.
- Yun, S. (2004). Imputation of missing values, *Journal of Preventive Medicine and Public Health*, **37**, 209–211.

임상시험에서 이분형 결측치 처리방법의 비교연구

안구성^a · 김동재^{a,1}

^a가톨릭대학교 의생명·건강과학과

(2016년 3월 14일 접수, 2016년 3월 31일 수정, 2016년 3월 31일 채택)

요약

임상시험에서 흔히 발생하는 결측치 중 이분형 결측치에 대한 논의를 하였다. 본 논문에서는 결측치가 발생하는 기제를 논의하고 기존의 여러 이분형 결측치 대체 방법과 수정된 결측치 대체방법을 소개하였다. 이후 각 결측치 대체 방법을 실제 자료에 적용하여 모의 실험을 진행하였다. 실제 자료의 성격 및 결측률의 변화에 따른 결측치 대체 방법들의 성능비교를 통해 진행하였다. 마지막으로 각 결측치 대체 방법에 대한 모의 실험 결과를 요약하고 토의하였다.

주요용어: 이분형 결측치, 임상시험, 결측치 발생기제, 결측치 대체

¹교신저자: (06591) 서울 서초구 반포대로 222, 의생명·건강과학과 가톨릭대학교.

E-mail: djkim@catholic.ac.kr