

임상시험에서 분석대상군 설정과 자료의 블라인드 검토

임지연¹, 곽민정²

¹동국대학교 의학과, ²평택대학교 데이터정보학과

Data Blind Review and Analysis Set in Clinical Trial

Chiyon Lim¹, Minjung Kwak²

¹Department of Medicine, Dongguk University, Goyang; ²Department of Data Information and Statistics, Pyeongtak University, Pyeongtaek, Korea

Objectives: The balance of confounding factors between treatment groups is supported by intention-to-treat (ITT) principle and this can reduce bias and confounding effects. Therefore clear definition of analysis set and blind review are essential prerequisites to perform statistical analysis. In this paper, we discuss the blind review and analysis set to check carefully the data during a blind review and finalize the statistical analysis set in clinical trials. **Methods:** The concept of ITT principle and ITT set has been often mixed and misused. Based on ITT principle, the definition of proper analysis sets are required to show the efficacy in statistical analysis. The analysis sets must be determined in blind review for data management and quality control. We review the analysis sets, such as FAS, PPS, and Safety set based on ITT principle and present blind review procedures. **Conclusions:** The determination of analysis sets through blind review is a very important procedure to reduce bias and to get appropriate result in randomized clinical trial. Blind review can assure the data quality and that is important prerequisites for statistical analysis. Therefore the role of blind review will be emphasized in clinical trial.

Key words: ITT, FAS, PPS, Safety set, Blind review

서 론

임상시험에서 통계분석을 위하여 정의되는 분석대상군(analysis set)은 무작위배정을 받은 모든 임상시험대상자가 모든 선정기준을 만족하고 중도탈락 없이 계획된 임상시험 과정을 완벽하게 실시하고 이로부터 자료를 완벽하게 기록하여 통계분석에 포함하는 것이 가장 이상적인 상황이다[1]. 배정된 대로 분석 원칙(intention-to-treat principle, ITT원칙)은 무작위 배정된 모든 대상자를 처음 무작위 배정된 치료군에 속한 것으로 간주하며 계획서 순응 여부에 관계없이 임상시험 종료까지 추적 관찰하여 평가되고 분석되어야 한다는 원칙이다[1,2]. 무작위배정 임상시험에서 ITT원칙은 치료군과 대조군 간에 예후인자(prognostic factor)의 균형을 유지시킬 수 있고 이를 통하여 선택편향

(selection bias)과 교락 요인의 영향을 최소화시킬 수 있다[3]. 그러나 실제로는 무작위배정 임상시험에서 계획서 위반이나 중도탈락으로 불균형이 발생하여 임상시험대상자의 분석대상군 포함 여부를 명확히 정의하는 것은 분석결과에 영향을 미칠 수 있다는 점에서 매우 중요하다고 할 수 있다. 계획서를 위반하거나 비순응으로 중도 탈락된 환자에 체계적인 편향이 있다면 결과를 왜곡시킬 수 있으며 이러한 시험대상자 가운데 어떤 환자를 제외시킬 것인지 결정하는 과정에서 편향이 발생할 수도 있다[4-6]. 분석대상군 선택의 기본적인 원칙은 편의를 최소화하고 제1종 오류가 커지는 것을 피하는 것이라고 할 수 있다. Alshurafa et al. [3]은 MEDLINE 문헌고찰을 통하여 결측치가 발생하였을 때 분석군의 정의와 이에 따른 결측치 처리가 어떻게 이루어지고 있는지 고찰하였다. 66개의 문헌 중 8%는 임상시험계획서 위배내용에

Corresponding author: Minjung Kwak

3825 Seodong-daro, Pyeongtaek 17869, Korea
Tel: +82-31-659-8355, E-mail: mjkwak@ptu.ac.kr

Received: October 24, 2016 Revised: November 25, 2016 Accepted: November 28, 2016

No potential conflict of interest relevant to this article was reported.

How to cite this article:

Lim C, Kwak M. Data Blind Review and Analysis Set in Clinical Trial. J Health Info Stat 2016;41(4):443-447.
Doi: <https://doi.org/10.21032/jhis.2016.41.4.443>

© It is identical to the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permit unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

© 2016 Journal of Health Informatics and Statistics

근거한 통계분석군에 대한 정의가 이루어지지 않았으며 38%는 임상시험계획서를 위배할 경우 발생할 수 있는 결측치 처리방법에 대하여 언급은 하였으나 통계분석군과의 관계에 대해서는 논의되지 않았다. 고찰한 문헌 가운데 55%에서 임상시험계획서 비순응군을 포함한 군과 임상시험 중 발생하는 결측치의 관계에 대하여 논의되었으며 가장 흔하게 사용된 결측치 처리방법은 last-observation-carried forward (LOCF)이었다(50%). 또한 Abraha et al. [7]은 2006년부터 2010년까지의 발표된 문헌들의 고찰을 통하여 임상시험 계획서의 통계분석군에 따라 약효의 결과가 다르기도 하며, 임상시험 비순응자가 많은 통계분석군과 그렇지 않은 군을 비교하였을 때 후자에서 시험약의 약효가 더 높게 나타나는 경향이 있음을 보였다.

임상시험결과분석에서 어떠한 대상자들을 분석대상군에 포함하여 분석을 하느냐에 따라 유효성의 결과해석이 달라질 수 있으므로 이러한 분석대상군을 정의하기 위하여 임상시험 계획서 위배자료의 블라인드 검토(blind review)를 하는데 이것은 임상시험에서 가장 중요한 과정 중의 하나이다[8]. 그러나 이에 대한 중요성이나 임상시험과정에서의 블라인드 검토 절차에 대하여 그간 활발히 논의되어 오지 않았다. 본 논문에서는 이러한 문제점을 미리 예방하고자 임상시험 계획서 위배자료의 블라인드 검토의 중요성과 그 절차 그리고 분석대상군의 명확한 정의에 대하여 논의하고자 한다.

연구 방법

분석대상군의 정의와 자료의 블라인드 검토

임상시험에서 통계적 분석 결과로 의약품의 유효성과 안전성에 대

하여 결론을 내리는 경우가 많기 때문에 통계분석결과는 무엇보다 중요하며 이러한 통계적 분석을 위하여 자료의 무결성과 검증 및 보안 등 여러 가지 단계에 걸쳐 자료의 품질보증(data quality control)을 하게 된다. 그러나 이러한 자료 품질보증을 위한 단계 외에 통계적 분석 결과에 영향을 미치는 아주 중요한 단계 중 하나가 블라인드 검토이다. 이 검토에서는 Figure 1에서와 같이 각각의 임상시험대상자들이 어느 분석대상군에 정의되어 포함되는지를 결정하는 것이다. 즉, 임상시험을 진행하면서 Table 1에 나타나있는 것과 같이 각 대상자들의 계획

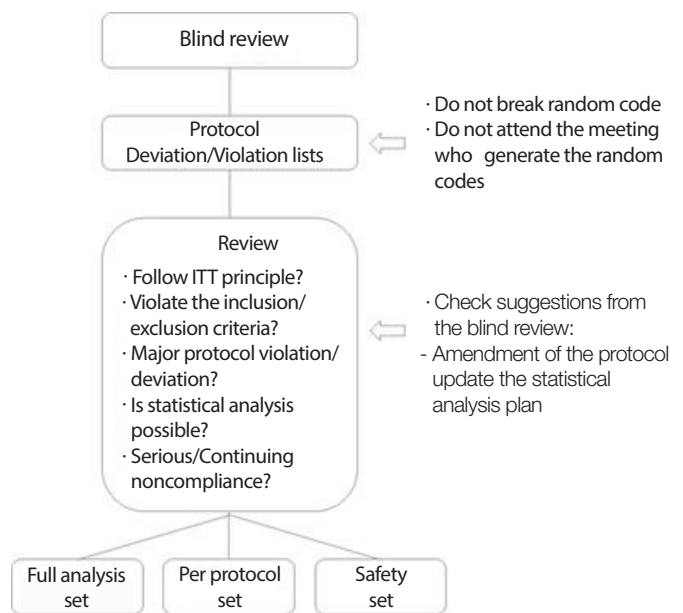


Figure 1. Procedure of blind review for definition of analysis set.

Table 1. A list of protocol deviation/violation in vaccine study (example, n=200)

ID	Protocol deviation/violation	Analysis set		
		PP set	FA set	Safety set
R001	A high fever after vaccination within 3 days	1	1	1
R002	Window period violation	0	1	1
R003	Drop-out	0	1	1
R004	Missing subject signature	1	1	1
R005	Failure to report serious unanticipated Problem/adverse events to the sponsor and IRB	1	1	1
R006	Missing the primary endpoint at baseline and final visit	0	1	1
R007	Missing the primary endpoint at 2nd visit	1	1	1
R008	Vaccination of prohibited vaccine	0	1	1
R009	Serious AE (Rash)	1	1	1
R010	No injection of any investigational vaccine	0	0	0
R011	Randomization error	0	1	1
...
R200	Administration a concomitant medicine	0	1	1

PP, per protocol; FA, full analysis; IRB, institutional review board; AE, adverse event. 1: included, 0: not included.

서 위반 또는 위배 내용을 하나씩 살펴보면서 해당 시험자를 어느 분석대상군에 포함시킬지 결정한다.

분석대상자군을 정의하는 이유는 시험군과 대조군의 유효성 또는 안전성을 확인하기 위함이기 때문에 블라인드 검토는 마지막 시험대상자가 임상시험을 완료한 시점부터 맹검(blinding)을 위해 만들어졌던 코드를 공개하기 직전에 이루어지는 것이 일반적이다.

통계분석군을 결정하기 위해서 가장 중요하게 고려되는 부분은 일차평가변수(primary endpoint)가 무엇인지를 명확히 하여야 한다. 만약 일차평가변수의 정의가 명확하지 않을 경우, 통계분석군의 정의가 어려워질 수 있다. 예를 들면, 본태성 고혈압환자의 혈압강하효과를 보는 12개월 임상시험에서 일차평가변수 정의를 수축기혈압의 변화량이라고만 정의를 한다면, 12개월과 기저상태를 일반적으로 변화량으로 정의할 수도 있다. 그러나 해당 계획서 상 6개월째 수축기혈압 변화량을 주 평가변수로 확인하고 나머지 6개월은 안전성 추적을 위한 연구라고 가정한다면 일차평가변수의 정확한 정의는 6개월간 변화량이 된다. 이러한 경우, 임상시험의 위배 정도는 6개월 기준으로 좀 더 신중히 판단할 필요가 있다. 그리고 임상시험계획서상의 분석대상군 정의에 따라 각각의 군의 정의를 명확히 하여야 한다. ICH E9에 의하면 분석대상군은 크게 세 가지로 나누어 질 수 있다. 먼저, 모든 분석대상자군(full analysis set, FAS), 계획서 순응 임상시험대상자군(per-protocol set, PPS), 그리고 안전성군(safety set)이다. 물론 이러한 분석대상군은 모든 임상시험에서 이렇게 세 가지로만 분류되는 것은 아니다. 연구에 따라서 FAS와 같은 성격을 나타내지만 명칭을 modified ITT (mITT)라고 쓰는 것처럼 분석군을 나타내는 명칭이 바뀔 수도 있으며, 특정군은 정의자체가 어려운 상황이 될 수도 있고 어떤 연구는 FAS와 PPS가 항상 같을 수밖에 없는 경우도 있다. 이러한 분석대상군을 정의하기 위하여 먼저 해야 할 부분이 대상자들에게 어떠한 위반 또는 위배사항이 발생했을 때 해당 분석군에 포함시키거나 제외시킬 것인지를 연구계획서에 상세히 기술하는 것이다.

FAS는 ITT원칙에 의해 제공되는 개념과 가능한 가장 가깝게 정의되는 이상적인 임상시험 대상자 집단이다. 따라서 무작위 배정된 모든 대상자로부터 계획서 위반 또는 위배내용을 검토하여 최소한의 제외 사유를 적용시킨다. 제외되는 사례를 들어보면 주요한 선정/제외 기준을 위배하거나 임상시험용 의약품이나 의료기기를 단 한 번도 투여받지 못하거나 치료를 받지 못했을 때 또는 무작위 배정 이후로 일차평가변수를 평가할 자료가 없는 경우이다. 이러한 주요한 사유를 제외하고는 가능한 FAS에 포함시켜 분석하기를 권장하고 있다. FAS의 경우에는 결측치가 발생할 수 있는데 결측치 처리를 어떻게 진행할 것인지를 계획서에 미리 정하여 기술하여야 한다. 이는 결측치 처리방법에 따라 유효성에 대한 결과해석이 달라질 수 있기 때문이다.

PPS의 경우는 FAS에 포함되는 대상자 중 계획서를 잘 순응한 임상시험 대상자들로 정의할 수 있다. 즉, 무작위 배정된 모든 시험대상자군에서 통계학적 모델을 이용하여 분석하였을 때 그 결과를 치료 효과로 설명할 수 있는 계획서 순응 임상시험 대상자군으로 정의할 수 있다. 따라서 ICH E9에 의하면 PPS는 미리 정한 최소 기간 이상의 치료를 완료한 경우, 주 결과변수의 측정치를 이용할 수 있는 경우, 선정 기준을 포함하여 임상시험계획서의 주요한 기준을 위반하지 않은 경우 등으로 들 수 있다. PPS의 장점은 임상시험을 통해 관찰된 치료효과 의 과학적 모형을 잘 반영할 수 있다는 것이다. PPS에서 제외되는 오류의 경우에는 무작위 배정 오류, 병용금지약물의 사용, 순응도가 낮은 경우, 결측치 발생 등을 들 수 있다. 반면 PPS의 단점은 이러한 임상시험계획서 위배들이 많게 되면 계획서 비순응자들이 많아지고 이는 결과가 보수적이지 않을 수도 있다. 따라서 황실대조군을 대조군으로 하는 연구에서는 PPS를 주 분석군으로 정의하기도 하는 것이 이 때문이기도 하다.

안전성군은 무작위 배정된 이후 임상시험용의약품이나 의료기기를 단 한번이라도 복용하거나 처치를 받은 경우가 있다면 모두 해당 군에 포함시키게 된다.

분석대상군 설정 시 유의 사항

분석대상군 설정 시 유의해야 할 사항을 나열해 보면 다음과 같다. 첫째, 여러 연구계획서에서 분석군으로 ITT군이라는 분석군을 정의하는 경우를 찾아볼 수 있다. ITT는 원래 '무작위 배정된 대로' 즉, 의도된 대로를 의미하며 통계분석군 정의에 앞서 무작위배정에 대한 원칙을 이야기하고 있는 것이다. 따라서 ITT원칙에 따라 FAS군을 구체적으로 정의하는 것이 필요하다. 둘째, 분석대상군을 정의할 때 계획서를 위반하면 대상자를 무조건 제외시키는 것이 아니고 일차평가변수에 얼마나 영향을 주는지를 고려해 보아야 한다. 예를 들어, 결측치가 있으면 PPS에서 제외한다고 할 때, 실제로 일차평가변수에 영향을 주지 않으면 PPS에 제외하지 않는 것이 바람직하다. 셋째, 계획서에 '주요한' 선정/제외 기준을 위배하면 PPS 또는 FAS에서 제외한다고 서술되어 있는 경우가 대부분이다. 이 '주요한'이라는 것은 여러 가지의 선정/제외 기준 항목 중 연구자마다 주요한 정도를 다르게 판단할 수 있다. 그러므로 한 연구계획서에 가능한 하나의 FAS, 하나의 PPS가 존재해야 하나, '주요한'의 의미해석에 따라 다양한 FAS 또는 PPS가 생겨나기도 한다. 이는 의약품이나 의료기기 허가를 위한 의사결정에 영향을 줄 수 있다. 마지막으로 FAS를 정의할 때 결측치가 발생할 경우 결측치 처리를 위한 합리적인 통계방법론이 연구계획서에 명확히 기술되어야 하며 결측치 처리방법에 따라 결과가 달라지는 지 검토하는 민감도 분석이 필요하다. 블라인드 검토에서 결측치 발생형태나 경향을 파악하

여 결측치 처리방법을 정의할 수도 있다. 따라서 블라인드 검토에서 맹검을 유지하는 것은 굉장히 중요하며, 만약 계획서에 정의된 결측치 처리 방법의 변경이 필요하거나 새롭게 정의를 해야 한다면 관련 검토내용을 문서화하고 계획서변경(amendment) 또는 통계분석계획(statistical analysis plan)에 반영하여야 한다.

이 모든 과정에서 살펴보듯이 시험대상자 한 명 한 명의 계획서 위반 또는 위배 내용을 보면서 어떤 분석군에 포함시킬지 여부를 결정하는 것은 결국 의약품이나 의료기기의 허가 및 집합 관계를 가질 수 있음을 알 수 있다. 따라서 분석대상군 설정에 영향을 주는 블라인드 검토 과정은 무엇보다도 객관적이어야 하며 계획서 위반 또는 위배내용을 판단해서 일차평가에 영향을 주는 정도에 대한 전문적인 지식이 있는 사람들이 참석하여야 한다. 일반적으로 이 검토회의에는 임상시험을 수행한 연구자, 스폰서에서 모니터링을 한 담당자, 각 병원에서 모니터링을 도와준 사람 등 임상시험을 직접적으로나 간접적으로 수행한 관련자 등이 된다. 단지 이 경우에 주의할 점은 각 시험대상자가 어느 의약품/의료기기에 노출되었는지는 비록 공개임상시험이었을지라도 검토 또는 회의 자료에 나타나서는 안된다. 또한, 무작위배정코드를 생성한 사람(통계학자 등)도 블라인드 검토에 참가할 수 없으며 이는 조금이라도 발생할 수 있는 편향을 없애기 위함이다.

분석대상군들은 임상시험계획서 위배 사항 등에 근거하여 그 정의가 이루어지는 것을 살펴보았다. 이러한 분석대상군들은 대조군 비교 임상시험일 경우 대조군이 처치군인지 위약군인지에 따라 어느 분석군이 더욱 과학적으로 비추어지는지에 대한 논의가 이루어진다. 만약 대조군이 처치대조군이라면 PPS가 좀 더 과학적인 분석대상군으로 간주되고 위약 대조군이면 FAS가 좀 더 위약대조군의 특성과 맞다고 간주되어져서 위약대조군 임상시험에서의 주 분석군은 FAS로, 처치대조군의 주 분석군은 PPS로 기술되기도 한다. 무엇보다도 중요한 것은 FAS의 특징이 허가를 받고난 뒤 실제 의료 환경에서 일어날 수 있는 오류를 많이 반영한 자료들을 가지고 있으므로 비록 처치대조군 임상시험이라고 할지라도 FAS와 PPS 두 분석군의 결과가 일치성(robustness)을 갖는 것이 매우 중요하다.

분석군의 역할

통계분석군을 임상시험계획서 위배 내용 등에 따라 여러 가지로 정의하는 이유는 임상시험결과를 해석함에 있어 편향을 최소화하고 서로 다르게 정의된 분석군 간의 결과가 서로 다르지 않음을 주장하기 위함이다. 어떠한 대상군을 사용하더라도 주요한 임상시험 결과가 달라지지 않음을 보여야 할 필요가 있기 때문이다[1]. 특히 확증연구에서 FAS와 PPS의 분석결과가 차이가 난다면 주 분석결과에 대한 고찰 및 해석이 더 필요하다. 이러한 경우에 추가적인 탐색적 분석을 계획하는

것을 권장할 수도 있고 추가 연구를 권장할 수도 있을 것이다.

대조군을 두는 연구에서 FAS에 포함되어 있는 대상자들 중 임상시험계획서 비순응자들은 주 처치효과를 감소시키는 경향이 있다. 따라서 그러한 이유로 위약대조군 임상시험에서 주 분석군을 FAS로 정의하기도 한다. 반면 비열등성이나 동등성 임상시험에서는 FAS에 포함된 임상시험 계획서 비순응자들의 결과는 보수적이지 않을 수도 있다. 따라서 이 경우에 PPS를 주 분석군으로 정의하기도 한다.

자료의 블라인드 검토절차

자료의 블라인드 검토 절차는 Figure 1과 같이 요약될 수 있고, 분석대상군에서 시험대상자나 자료의 제외 또는 포함 여부 등을 판단하기 위한 절차이다. 이 절차는 통계분석을 위한 자료의 무결성 및 질적 관리를 위한 관점에서 이루어진다. 따라서 블라인드 검토는 마지막 시험대상자가 임상시험을 완료한 뒤 맹검을 해지하기 전에 이루어지게 된다. 블라인드 검토를 위한 미팅에서는 계획서 위반 또는 위배 내용을 Table 1과 같이 시험대상자별로 나열한 후 각 내용의 시험계획서 위반 정도에 따라 일차평가변수에 주는 영향력을 판단하여 FAS 또는 PPS에 포함시킬지 여부를 결정한다. 안전성 분석군 정의 시 연구용 의약품을 복용하거나 의료기기를 사용한 경우 계획서 위반내용에 상관없이 모두 안전성 분석군에 포함시킨다. 블라인드 검토에서 주의사항은 여러 가지가 있는데 이를 요약하면 다음과 같다.

- 1) 무작위코드는 공개하지 않는다.
- 2) 무작위코드를 생성한 통계학자 등은 해당 검토를 위한 회의에 참석하지 않는다.
- 3) 계획서 내용을 숙지하고 참석한다.
- 4) 관련한 임상시험의 일차평가변수의 의미를 이해한다.
- 5) 관련한 임상시험의 다른 연구 결과들로부터 주요한 이상반응 등에 대하여 미리 문헌고찰을 한다.
- 6) 블라인드 검토에 적극적으로 임한다.
- 7) 블라인드 검토에서 나온 제안들 중 계획서 변경 또는 통계분석계획에 반영해야 되는 부분이 있는지 확인한다.

위 사항을 유의하면서 블라인드 검토는 Figure 1과 같은 과정을 통하여 진행할 수 있다.

블라인드 검토에서 기존의 계획서의 내용 변경이 필요할 수도 있다. 이 경우에는 계획서 변경을 하거나 통계분석계획에 반영하여야 한다. 블라인드 검토는 통계분석을 위한 분석대상군을 정의하기 위함이기 때문에 검토 시 나오는 의견들은 문서화하여 통계분석에 반영할 수 있도록 한다. 이 역시 통계분석 이전에 완료하여 분석결과 해석에 편향이 발생하지 않도록 하여야 한다.

결론 및 토의

임상시험을 통해 임상효과의 발견 가능성을 극대화하기 위하여 분석대상군을 정의하며 이때 분석대상군 간에 가능한 동질성을 유지할 수 있게 해야 한다. 임상시험을 통한 유효성이나 안전성에 대한 의사결정을 내리기 전에 임상시험 전체에 대한 점검이나 품질관리를 통하여 임상 질적 관리를 하게 된다. 이 과정 중에서도 또는 별도로 자료의 품질관리 및 점검도 시행한다. 이러한 자료의 품질관리에 대한 이해는 높아지고 있는 반면 아직 블라인드 검토의 개념에 대한 이해가 적은 형편이다. 임상자료에 대한 품질관리의 중요성도 관련기관 등에서 널리 알리고 교육을 통하여 인식이 높아진 것처럼 블라인드 검토에 대한 교육도 더욱 활성화될 필요가 있다. 본문에서 기술한 바와 같이 블라인드 검토는 분석대상군 정의에 직접적 영향을 미치게 되고 이는 의약품이나 의료기기의 유효성이나 성능을 평가하는 데 아주 중대한 영향을 미치기 때문이다. 따라서 분석대상군에 따라 의약품이나 의료기기 등의 유효성이나 성능의 해석이 달라질 수 있기 때문에 분석대상군 정의를 계획서에 미리 명확하고 자세하게 기술할 수 있도록 노력하여야 한다. 미리 정해진 기준에 따라 블라인드 검토가 이루어지므로 분석대상군 정의 시 미칠 수 있는 편향을 최대한 줄여야 한다.

본 연구에서는 임상시험에서 분석대상군의 명확한 정의와 자료의 무결성과 질적 관리 확보를 위한 블라인드 검토 중요성과 그 절차에 대한 가이드라인을 제시하고자 하였다. 향후 연구에서 분석대상군의 정의와 결측치의 처리에 따른 유효성 분석을 모의실험과 민감도 분석을 통하여 비교하고자 한다.

REFERENCES

1. International conference on harmonization guidance E9: Statistical principles of clinical trials, recommended for adoption to the regulatory bodies of the European Union, Japan and USA. 1988.
2. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res* 2011;2(3):109-112.
3. Alshurafa M, Briel M, Aki EA, Haines T, Moayyedi P, Gentles SJ, et al. Inconsistent definitions for intention-to-treat in relation to missing outcomes data: systematic review of the methods literature. *Plos One* 2012;7(11):1-7.
4. Lee YJ. What is the clinical trial?. *J Health Info Stat* 2011;36(1):1-14 (Korean).
5. Shin H, Lee J, An H. Simulation study for comparison between ITT and PP in matched-pair noninferiority clinical trials. *J Health Info Stat* 2011;36(2):217-228 (Korean).
6. Lachin JM. Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials* 2000;21(5):526.
7. Abraha I, Cherubini A, Cozzolino F, Florio RD, Luchetta ML, Rimland JM. Deviation from intention to treat analysis in randomized trials and treatment effects estimates: meta-epidemiological study. *Biomedical J* 2015;350:1-12.
8. Yu YP, Yao C. Data quality in clinical trials: the role of blind review. *Yao Xue Xue Bao* 2015;50(11):1498-1501.