

# 부트스트랩을 활용한 의약품 동등성 평가

(Evaluation of dissolution profile similarity using Bootstrapping methods)

2023.11.29 박병훈



# CONTENTS

1. 의약품 동등성 시험 개요
2. Bootstrap의 활용
3. 동등성 평가(비교용출시험)
4. 자문 컨설팅 소개



# 01

## 의약품 동등성 시험 개요

---



# 01 의약품 동등성 시험 개요

## 용어의 정의

### □ 의약품 동등성 시험

- 원칙적으로 주성분, 함량 및 제형 (Dosage form)이 동일한 두 제제 (Formulation)에 대한 의약품 동등성 (Equivalence)을 입증하기 위해 실시하는 시험

#### 1) 생물학적 동등성 시험 (Bioequivalence test)

- 생물학적 동등성 (이하, 생동성) 입증을 위하여 실시하는 생체 내 시험의 하나로 주성분 (Active Ingredient)이 전신 순환혈 (Systemic blood circulation)에 흡수되어 약효를 나타내는 의약품에 대하여 동일 주성분을 함유한 동일 투여경로의 두 제제가 생체 이용률 (Bioavailability)에 있어서 통계학적으로 동등하다는 것을 입증

#### 2) 비교 용출 시험

- 동일 주성분을 가진 동일 투여경로의 두 제제에 대하여 시험관 내 용출양상의 유사성을 입증하기 위하여 동일한 조건에서 실시하는 시험

# 01 의약품 동등성 시험 개요

## 비교 용출 시험

### □ 비교 용출 시험이 필요한 경우

- 1) 주성분의 종류, 함량 및 제형이 동일한 의약품을 신규로 품목허가(신고) 하는 경우
- 2) 품목허가(신고) 후 변경 사항이 있는 경우

### □ 비교 용출 시험 동등성 평가 방법(기준)

- 시험군(Test group)과 대조군(Control group)의 평균 용출률(Average dissolution rates) 비교
- 특정 시점에서만 용출률을 비교하여 유사성 인자(Similarity factor,  $f_2$ )로 동등성 평가

# 01 의약품 동등성 시험 개요

## ■ 비교 용출 시험 동등성 판정 예시(개정 전)

- 시험군과 대조군의 각 용출 시점에서 12개의 시료를 수집
- 평균 용출률을 이용하여 동등성을 평가하기 위해서는 대조약과 시험약의 평균 용출률의 변이계수(Coefficient of Variation, CV)가 첫 번째 비교시점에서 20% 이내이며, 그 외 비교 시점에서는 10%를 초과해서는 안됨
- 오른쪽 예시에서 처음 시점의 변이계수는 20.86, 20.91%로 20%를 초과함
- 이러한 경우, **유사성 인자( $f_2$ )**로 동등성 평가

| 시험약<br>시험조 | 용출시간(분) |      |      |
|------------|---------|------|------|
|            | 10      | 15   | 30   |
| 1          | 42.9    | 68.2 | 80.7 |
| 2          | 53.5    | 78.1 | 87.4 |
| 3          | 80.9    | 89.6 | 93.1 |
| 4          | 46.4    | 75.1 | 85.3 |
| 5          | 52.1    | 77.3 | 86.5 |
| 6          | 56.6    | 79.8 | 87.4 |
| 7          | 52.5    | 82.0 | 89.9 |
| 8          | 45.9    | 75.0 | 87.5 |
| 9          | 56.1    | 79.7 | 87.3 |
| 10         | 60.0    | 81.8 | 88.2 |
| 11         | 36.4    | 66.1 | 78.0 |
| 12         | 60.8    | 82.1 | 89.2 |
| 평균         | 53.7    | 77.9 | 86.7 |
| 표준편차       | 11.20   | 6.35 | 3.95 |
| 변이계수       | 20.86   | 8.15 | 4.56 |

〈시험군〉

| 대조약<br>시험조 | 용출시간(분) |      |      |
|------------|---------|------|------|
|            | 10      | 15   | 30   |
| 1          | 42.9    | 68.2 | 80.7 |
| 2          | 53.5    | 78.1 | 87.4 |
| 3          | 80.9    | 89.6 | 93.1 |
| 4          | 46.4    | 75.1 | 85.3 |
| 5          | 52.1    | 77.3 | 86.5 |
| 6          | 56.6    | 79.8 | 87.4 |
| 7          | 52.5    | 82.0 | 89.9 |
| 8          | 45.9    | 75.0 | 87.5 |
| 9          | 56.1    | 79.7 | 87.3 |
| 10         | 60.0    | 81.8 | 88.2 |
| 11         | 36.4    | 66.1 | 78.0 |
| 12         | 60.8    | 82.1 | 89.2 |
| 평균         | 53.7    | 77.9 | 86.7 |
| 표준편차       | 11.20   | 6.35 | 3.95 |
| 변이계수       | 20.86   | 8.15 | 4.56 |

〈대조군〉

# 01 의약품 동등성 시험 개요

## 유사성 인자(Similarity Factor, $f_2$ )

- 평균 용출률(%)을 이용해 산출한 시험군과 대조군의 용출 곡선 간의 차이 값
- 두 집단의 용출률 차이가 없을 경우의  $f_2$ 는 100으로 산출되도록 공식 개발
- 산출된  $f_2$ 의 값이 50 이상일 경우 통계적으로 동등한 것으로 간주 (일반제제와 장용성 및 서방성 제제의 기준은 다소 상이)
- $f_2$ 가 50이라는 것은 시험군과 대조군의 평균 용출률의 차이가 10%인 것을 의미함 → 10% 이하는 동등한 것으로 판정.

$$f_2 = 100 - 25 \log \left( 1 + \frac{1}{P} \sum_{i=1}^P (\bar{X}_{T,i} - \bar{X}_{R,i})^2 \right)$$

$P$ : 시점의 수

$\bar{X}_{T,i}$ : 시험군에서의 평균 용출률

$\bar{X}_{R,i}$ : 대조군에서의 평균 용출률



# 01 의약품 동등성 시험 개요

## 기존 유사성 인자 활용 방법의 한계

- 제약 회사에서 측정한 12개의 시료에 대한 산포(편차)가 큰 경우 동등성 평가 결과가 부정확할 수 있음
- 실제로는 동등하지만 동등하지 않은 것으로 평가할 수 있음(제 1종 오류, Type 1 Error)
- 실제로는 동등하지 않지만 동등한 것으로 평가할 수 있음(제 2종 오류, Type 2 Error)
- 이러한 문제를 해결하기 위해서 더 많은 시료를 수집하여 분석을 수행하면 표준 오차를 줄일 수 있음
- **But**, 대부분에 기업에서는 시간과 비용 등의 문제로 많은 시료를 수집하기에는 사실상 불가능함
- **정밀한 동등성 평가를 위한 통계적 방법이 필요함**



# 01 의약품 동등성 시험 개요

## 개정된 비교 용출 시험 동등성 평가

- 2022년 11월 12일 의약품동등성시험 기준 개정(2021년 평균용출률로 동등성 판정 기준 삭제)
- 2023년 5월 18일 평가 고려사항을 상세하게 안내하기 위한 추가 개정
- 특정 시점이 아닌 **측정한 모든 시점**에서 유사성 인자( $f_2$ )를 이용하여 비교용출시험 결과 평가
- 통계적 방법인 **Bootstrap 방법을 이용하여 산출된 유사성 인자( $f_2$ )**의 신뢰구간(CI)으로 동등성 평가

# 01 의약품 동등성 시험 개요

2022년 하반기 의약품 허가, 심사 설명회 자료

## 유사성 인자 (1/1)

**Q9.** 유사성 인자( $f_2$ )로 비교용출시험 결과를 판정하는 과정에서 평균용출률을 사용하기 위한 조건인 **변이계수가 기준치를 초과하는 경우에는 어떻게 하나요**

**A9.** 변이계수의 기준 초과로 유사성 인자의 적용이 어려운 경우, **통계적으로 타당성 있는 다른 방법을 사용하여 동등성 판정을 할 수 있음**

선택한 통계적 방법<sup>1)</sup>과 판정기준의 타당성을 입증<sup>2)</sup>

<sup>1)</sup> 예) ① Confidence interval derivation for  $f_2$  based on bootstrap, ② CI derivation f

③ Mahalanobis distance, ④ Model-dependent approach, maximum deviation method 등 있으며, 여기에 국한되지는 않음

<sup>2)</sup> 예) 변이계수 커진 원인 고찰, 통계적 기법 선택 사유, 판정기준 타당성 등

① Confidence interval derivation for  $f_2$  based on bootstrap,

# 02

## Bootstrap의 활용

---



## 02 Bootstrap의 활용

### Bootstrap(부트스트랩)이란?

- 복원 추출(Sampling with Replacement)을 이용하여 원데이터로부터 재표본을 추출(Resampling)하는 방법
- 추출된 재표본의 개수는 원데이터와 동일
- 각 시점별 원데이터의 수만큼 표본을 재추출
- 우연하게 특정값이 반복되어 추출될 수 있음

복원 추출로 표본이 추출되기 때문에 5.5의 값이 반복하여 추출될 수 있음

|     |      | 용출 시간(분) |      |      |      |      |    |
|-----|------|----------|------|------|------|------|----|
| No. |      | 10       | 15   | 30   | 45   | 60   | 90 |
| 1   | 5.5  | 13.5     | 37.0 | 51.7 | 61.4 | 77.7 |    |
| 2   | 10.8 | 19.8     | 39.5 | 56.2 | 68.3 | 80.6 |    |
| 3   | 5.3  | 12.3     | 36.5 | 51.9 | 62.2 | 79.4 |    |
| 4   | 11.9 | 21.6     | 40.8 | 55.8 | 65.6 | 79.7 |    |
| 5   | 11.3 | 21.4     | 42.6 | 59.7 | 72.0 | 86.0 |    |
| 6   | 9.5  | 22.3     | 41.4 | 59.6 | 70.6 | 84.3 |    |

원 데이터

|     |      | 용출 시간(분) |      |      |      |      |    |
|-----|------|----------|------|------|------|------|----|
| No. |      | 10       | 15   | 30   | 45   | 60   | 90 |
| 1   | 5.5  | 13.5     | 37.0 | 51.7 | 61.4 | 77.7 |    |
| 1   | 5.5  | 13.5     | 37.0 | 51.7 | 61.4 | 77.7 |    |
| 2   | 10.8 | 19.8     | 39.5 | 56.2 | 68.3 | 80.6 |    |
| 3   | 5.3  | 12.3     | 36.5 | 51.9 | 62.2 | 79.4 |    |
| 1   | 5.5  | 13.5     | 37.0 | 51.7 | 61.4 | 77.7 |    |
| 4   | 11.9 | 21.6     | 40.8 | 55.8 | 65.6 | 79.7 |    |

재표본 데이터 (1)

|     |      | 용출 시간(분) |      |      |      |      |    |
|-----|------|----------|------|------|------|------|----|
| No. |      | 10       | 15   | 30   | 45   | 60   | 90 |
| 3   | 5.3  | 12.3     | 36.5 | 51.9 | 62.2 | 79.4 |    |
| 6   | 9.5  | 22.3     | 41.4 | 59.6 | 70.6 | 84.3 |    |
| 5   | 10.8 | 19.8     | 39.5 | 56.2 | 68.3 | 80.6 |    |
| 6   | 9.5  | 22.3     | 41.4 | 59.6 | 70.6 | 84.3 |    |
| 4   | 11.9 | 21.6     | 40.8 | 55.8 | 65.6 | 79.7 |    |
| 2   | 10.8 | 19.8     | 39.5 | 56.2 | 68.3 | 80.6 |    |

재표본 데이터 (2)

|     |      | 용출 시간(분) |      |      |      |      |    |
|-----|------|----------|------|------|------|------|----|
| No. |      | 10       | 15   | 30   | 45   | 60   | 90 |
| 5   | 5.5  | 13.5     | 37.0 | 51.7 | 61.4 | 77.7 |    |
| 5   | 5.5  | 13.5     | 37.0 | 51.7 | 61.4 | 77.7 |    |
| 4   | 11.9 | 21.6     | 40.8 | 55.8 | 65.6 | 79.7 |    |
| 2   | 10.8 | 19.8     | 39.5 | 56.2 | 68.3 | 80.6 |    |
| 3   | 5.3  | 12.3     | 36.5 | 51.9 | 62.2 | 79.4 |    |
| 2   | 10.8 | 19.8     | 39.5 | 56.2 | 68.3 | 80.6 |    |

재표본 데이터 (3)

## 02 Bootstrap의 활용

### Bootstrap(부트스트랩)의 필요성

- Bootstrap을 이용하여 원데이터를 재추출하면 표본 데이터 수가 증가함
- 증가된 데이터를 이용하여 Bootstrap한 횟수만큼 유사성 인자( $f_2$ )를 산출할 수 있음
- Bootstrap된  $f_2$ 의 분포 및 신뢰구간을 추정하여 **동등성 평가의 정밀도를 높일 수 있음**
- 국내 허가기관인 식약처의 질의응답에 의거 5,000회의 Bootstrap을 통한  $f_2$ 의 활용
- Minitab을 이용하여 Bootstrap된  $f_2$ 를 산출하고 90% 신뢰구간의 하한값을 계산하여 동등성 평가

# 03

## 동등성 평가(비교용출시험)

---



# 03 동등성 평가(비교용출시험)

## 계산 절차

- 1) 시험군과 대조군에서 각각 12개의 재표본 추출
- 2) 1)에서 추출된 재표본을 이용하여 시험군과 대조군으로부터 각 시점에 대한 평균 용출률 계산
- 3) 2)에서 계산된 평균 용출률을 이용하여  $f_2$  계산
- 4) 1)~3) 5,000회 반복
- 5) 5,000개의  $f_2$ 를 이용하여  $f_2$ 의 90% 신뢰하한 도출
- 6)  $f_2$ 의 90% 신뢰하한이 50보다 크면 시험군과 대조군 간 유사성을 만족하는 것으로 판정

| No. | 용출 시간(분) |      |      |      |      |      |
|-----|----------|------|------|------|------|------|
|     | 10       | 15   | 30   | 45   | 60   | 90   |
| 1   | 5.5      | 13.5 | 37.0 | 51.7 | 61.4 | 77.7 |
| 2   | 10.8     | 19.8 | 39.5 | 56.2 | 68.3 | 80.6 |
| 3   | 5.3      | 12.3 | 36.5 | 51.9 | 62.2 | 79.4 |
| 4   | 11.9     | 21.6 | 40.8 | 55.8 | 65.6 | 79.7 |
| 5   | 11.3     | 21.4 | 42.6 | 59.7 | 72.0 | 86.0 |
| 6   | 9.5      | 22.3 | 41.4 | 59.6 | 70.6 | 84.3 |
| 7   | 9.7      | 18.1 | 38.9 | 57.3 | 71.3 | 84.9 |
| 8   | 5.9      | 15.3 | 38.5 | 52.6 | 62.0 | 78.3 |
| 9   | 10.5     | 18.9 | 45.9 | 62.4 | 72.5 | 84.6 |
| 10  | 8.7      | 21.7 | 35.8 | 47.0 | 59.3 | 76.6 |
| 11  | 9.9      | 20.8 | 43.9 | 56.3 | 63.8 | 78.2 |
| 12  | 9.4      | 19.9 | 40.3 | 54.3 | 65.6 | 80.1 |

원 데이터

Bootstrap 방법을 이용하기 때문에 동일한 값으로 재표본 추출될 수 있음

| No. | 용출 시간(분) |      |      |      |      |      |
|-----|----------|------|------|------|------|------|
|     | 10       | 15   | 30   | 45   | 60   | 90   |
| 1   | 5.5      | 13.5 | 37.0 | 51.7 | 61.4 | 77.7 |
| 5   | 11.3     | 21.4 | 42.6 | 59.7 | 72.0 | 86.0 |
| 9   | 10.5     | 18.9 | 45.9 | 62.4 | 72.5 | 84.6 |
| 9   | 10.5     | 18.9 | 45.9 | 62.4 | 72.5 | 84.6 |
| 1   | 5.5      | 13.5 | 37.0 | 51.7 | 61.4 | 77.7 |
| 8   | 5.9      | 15.3 | 38.5 | 52.6 | 62.0 | 78.3 |
| 4   | 11.9     | 21.6 | 40.8 | 55.8 | 65.6 | 79.7 |
| 12  | 9.4      | 19.9 | 40.3 | 54.3 | 65.6 | 80.1 |
| 1   | 5.5      | 13.5 | 37.0 | 51.7 | 61.4 | 77.7 |
| 12  | 9.4      | 19.9 | 40.3 | 54.3 | 65.6 | 80.1 |
| 10  | 8.7      | 21.7 | 35.8 | 47.0 | 59.3 | 76.6 |
| 2   | 10.8     | 19.8 | 39.5 | 56.2 | 68.3 | 80.6 |

재표본 데이터




# 03 동등성 평가(비교용출시험)

## ■ 계산 절차

- 1) 시험군과 대조군에서 각각 12개의 재표본 추출
- 2) 1)에서 추출된 재표본을 이용하여 시험군과 대조군으로부터 각 시점에 대한 평균 용출률 계산
- 3) 2)에서 계산된 평균 용출률을 이용하여  $f_2$  계산
- 4) 1)~3) 5,000회 반복
- 5) 5,000개의  $f_2$ 를 이용하여  $f_2$ 의 90% 신뢰하한 도출
- 6)  $f_2$ 의 90% 신뢰하한이 50보다 크면 시험군과 대조군 간 유사성을 만족하는 것으로 판정

| 평균 용출률 (시험군) |      |      |      |      |      |
|--------------|------|------|------|------|------|
| 10           | 15   | 30   | 45   | 60   | 90   |
| 9.0          | 18.8 | 40.1 | 55.4 | 66.2 | 80.9 |

| 평균 용출률 (대조군) |      |      |      |      |      |
|--------------|------|------|------|------|------|
| 10           | 15   | 30   | 45   | 60   | 90   |
| 10.1         | 16.0 | 33.6 | 47.7 | 59.7 | 74.3 |


$$f_2 = 100 - 25 \log \left( 1 + \frac{1}{P} \sum_{i=1}^P (\bar{X}_{T,i} - \bar{X}_{R,i})^2 \right)$$

$P$ : 시점의 수

$\bar{X}_{T,i}$ : 시험군에서의 평균 용출률

$\bar{X}_{R,i}$ : 대조군에서의 평균 용출률

## 03 동등성 평가(비교용출시험)

### ■ 계산 절차

- 1) 시험군과 대조군에서 각각 12개의 재표본 추출
- 2) 1)에서 추출된 재표본을 이용하여 시험군과 대조군으로부터 각 시점에 대한 평균 용출률 계산
- 3) 2)에서 계산된 평균 용출률을 이용하여  $f_2$  계산
- 4) 1)~3) 5,000회 반복
- 5) 5,000개의  $f_2$ 를 이용하여  $f_2$ 의 90% 신뢰하한 도출
- 6)  $f_2$ 의 90% 신뢰하한이 50보다 크면 시험군과 대조군 간 유사성을 만족하는 것으로 판정

Minitab 분위수 공식 출처: <https://support.minitab.com/ko-kr/minitab/21/help-and-how-to/statistics/basic-statistics/how-to/display-descriptive-statistics/methods-and-formulas/methods-and-formulas/#1st-quartile-q1>

### $f_2$ 의 신뢰구간 공식

- 부트스트랩 재표본을 이용하여 신뢰구간 추정 시 **수학적인 계산 과정이 간단한 분위수(Quantile) 방법**이 많이 쓰이고 있음
- 부트스트랩 재표본으로 계산된  $f_2$ 의 0.05 분위수를 90% 신뢰하한으로 **제시함**
- Minitab에서 제공되는 분위수 공식은 아래와 같으며, 90% 신뢰하한 계산시  $p$ 는 0.05로 지정해야 함

$$Q(p) = (1 - z)x_{(j)} + zx_{(j+1)}$$

$$0 \leq p \leq 1$$

$$W = (n + 1)p$$

$$j = [(n + 1)p] \rightarrow (n + 1)p \text{보다 작거나 같은 정수}$$

$$z = W - j$$

$x_{(j)}$ : 오름차순으로 정렬된 표본 데이터의  $j$  번째 관측치

$n$ : 데이터 개수

## 03 동등성 평가(비교용출시험)

### 계산 절차

- 1) 시험군과 대조군에서 각각 12개의 재표본 추출
- 2) 1)에서 추출된 재표본을 이용하여 시험군과 대조군으로부터 각 시점에 대한 평균 용출률 계산
- 3) 2)에서 계산된 평균 용출률을 이용하여  $f_2$  계산
- 4) 1)~3) 5,000회 반복
- 5) 5,000개의  $f_2$ 를 이용하여  $f_2$ 의 90% 신뢰하한 도출
- 6)  $f_2$ 의 90% 신뢰하한이 50보다 크면 시험군과 대조군 간 유사성을 만족하는 것으로 판정

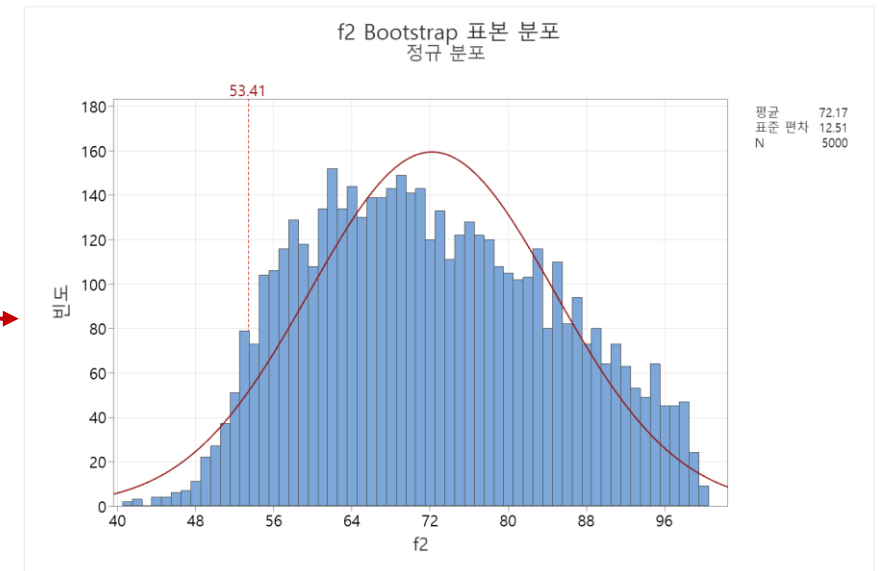
# 03 동등성 평가(비교용출시험)

## Minitab 분석

- 5,000회의 Bootstrapping이 필요하기 때문에  
**Minitab의 매크로 (Macro)**를 활용하여 재표본 추출  
(Resampling) → **5,000개의  $f_2$  산출**
- 히스토그램(Histogram)을 이용하여  $f_2$ 의 분포를 시각  
화하고 90% 신뢰 하한의 값을 확인
- $F_2$ 의 값이 50보다 크기 때문에 (53.41) 동등한 것으로  
결론 내릴 수 있음

5,000개의  $f_2$  산출 :

| C14             |
|-----------------|
| Bootstrapped f2 |
| 64.6250         |
| 78.6613         |
| 63.8488         |
| 88.0433         |
| 60.7745         |
| 59.0160         |
| 88.6026         |
| 92.3267         |
| 72.7238         |
| 82.7109         |
| 84.5035         |
| 98.0765         |
| 68.3497         |
| 64.9739         |



90% 신뢰구간 하한 = 53.40999042

▶ 유사성 판정 결과 = 합격

# 04

## 자문 컨설팅 소개

---



# THANK YOU

---

