Chapter 685

Biosimilarity Tests for the Difference Between Means using a Parallel Two-Group Design

Introduction

A *biosimilar* product is a copy of a large-molecule (biologic) innovator product whose patent has expired. The biosimilar will be manufactured by a different company. These new products must be shown to be *similar* to the original (reference) product in potency, purity, and safety, although minor differences in clinically inactive components are usually allowed. WHO published its "Guidelines for the evaluation of similar biotherapeutic products (SBPs)" in 2009. Other regulatory bodies such as the FDA and EMA have also produced guidelines for how to approve biosimilar products.

The process for obtaining approval of *small-molecule* generic drugs by establishing bioequivalence in terms of efficacy and safety is now well established. A typical path includes cross-over study design, test criterion, and statistical methods for show *bioequivalence*.

Large-molecule biologic drugs are different from their small-molecule cousins. The generic versions of these products are called *follow-on biologics* by the FDA (US), *biosimilars* by the EMA (Europe), and *subsequent entered biologics (SEB)* by the PHA (Canada). They are sometimes referred to as *similar biological drug products (SBDP).* They are fundamentally different from small-molecular products because they are much more complex and not exact chemical copies of the drugs they are meant to copy.

We refer you to the books Chow (2014) and Chow (2019) for complete coverage of biosimilars.

This procedure allows you to study the power and sample size of *biosimilarity* tests of the means of two independent groups using the two-sample equal-variance *t*-test. Schuirmann's (1987) two one-sided tests (TOST) approach is used to test biosimilarity. Only a brief introduction to the subject will be given here. For a comprehensive discussion of equivalence, refer to Chow and Liu (1999).

Measurements are made on individuals that have been randomly assigned to one of two groups. This *parallel-groups* design may be analyzed by a TOST biosimilarity test to show that the means of the two groups do not differ by more than a small amount, called the margin of biosimilarity.

The definition of biosimilarity has been refined in recent years using the concepts of prescribability and switchability.

Parallel-Group Design

In a parallel-group design, subjects are assigned at random to either of two groups. Group 1 is the treatment group and group 2 is the reference group.

Outline of an Biosimilarity Test

PASS follows the *two one-sided tests* approach described by Schuirmann (1987) and Phillips (1990). Let $\mu_1 = \mu_T$ be the test group mean, $\mu_2 = \mu_R$ be the reference group mean, and E_L and E_U be the lower and upper bounds, respectively, on $\delta = \mu_1 - \mu_2 = \mu_T - \mu_R$ that define the region of equivalence.

It will be convenient to adopt the following specialized notation for the discussion of these tests.

Parameter	PASS Input/Output	Interpretation
μ_1 or μ_T	Not used	*Mean* of population 1. Population 1 is assumed to consist of those who have received the new treatment.
μ_2 or μ_R	Not used	*Mean* of population 2. Population 2 is assumed to consist of those who have received the reference treatment.
E_L, E_U	EL, EU	*Lower* and *Upper Biosimilarity Limits.* If the difference is between these two limits, the new treatment is said to be *biosimilar* to the reference.
δ	δ	*Actual difference*. This is the value of $\mu_1 - \mu_2$, the difference between the means. This is the value at which the power is calculated.

Note that the actual values of μ_1 and μ_2 are not needed. Only their difference is needed for power and sample size calculations.

With $E_L < 0$ and $E_U > 0$, the null hypothesis of non- biosimilarity is

$$H_0: \delta \leq E_L \text{ or } \delta \geq E_U.$$

The alternative hypothesis of equivalence is

$$H_1: E_L < \delta < E_U.$$

Two-Sample Equal-Variance T-Test Statistics

This test assumes that the two groups of normally distributed values have the same variance. The calculation of the two one-sided test statistics uses the following equations.

$$t_L = \frac{(\bar{X}_1 - \bar{X}_2) - E_L}{s_{\bar{X}_1 - \bar{X}_2}}$$

$$t_U = \frac{(\bar{X}_1 - \bar{X}_2) - E_U}{s_{\bar{X}_1 - \bar{X}_2}}$$

where

$$\bar{X}_k = \frac{\sum_{i=1}^{n_k} X_{ki}}{n_k},$$

$$s_k = \sqrt{\left(\frac{\sum_{i=1}^{n_k}(X_{ki} - \bar{X}_k)^2}{(n_k - 1)}\right)},$$

$$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)},$$

$$df = n_1 + n_2 - 2.$$

The null hypothesis is rejected if t_L and $-t_U$ are both greater than or equal to $t_{1-\alpha,df}$.

Power Calculation

When $\sigma_1 = \sigma_2 = \sigma$, the power of the equal-variance biosimilarity *t*-test is calculated as

$$\Pr\left(t_L \geq t_{1-\alpha,n_1+n_2-2} \text{ and } t_U \leq -t_{1-\alpha,n_1+n_2-2} | \mu_1, \mu_2, \sigma\right)$$

where t_L and t_U are distributed as the bivariate, noncentral *t* distribution with noncentrality parameters Δ_L and Δ_U given by

$$\Delta_L = \frac{\delta - E_L}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$\Delta_U = \frac{\delta - E_U}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

For biosimilarity tests, the standard deviation of the reference group is used for sample size calculations.

Julious (2023) page 127 presents the following formula based on the normal distribution

$$Power = \Phi\left(\sqrt{\frac{(\delta - E_L)^2 R n_1}{\sigma^2(R+1)}} - z_{1-\alpha}\right) + \Phi\left(\sqrt{\frac{(\delta - E_U)^2 R n_1}{\sigma^2(R+1)}} - z_{1-\alpha}\right) - 1$$

where $R = n_2/n_1$.

According to Chow (2019) the FDA recommends that $\delta = \sigma/8$ and $E_U = 1.5\sigma$. Also, they recommend that the estimated value of σ be obtained from a pilot study or a previous study using only data from the reference group.

Example 1 – Finding Power

Let drug B be a popular biologic drug used to control blood pressure whose exclusive marketing license has expired. Suppose drug A is a competing drug developed by another company that wants to license it as a biosimilar drug to drug B. A parallel-group design is planned to compare these two drugs in terms of safety, purity, and therapeutic response.

The average blood pressure is 96 mmHg in drug B with a within-group standard deviation of 18mmHg. Following FDA guidelines, the researchers set the average difference between the means of the two drugs at which the power is computed to

$$\delta = \frac{\sigma}{8} = 2.25$$

The biosimilarity limits will also follow FDA guidelines and are set to $\pm 1.5\sigma = \pm 27$. The type-I error rate is set to 0.025 and the desired power is 0.90. A chart is desired that shows the power for group sample sizes ranging from 6 to 20 using the noncentral t distribution.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 1** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	**Power**
Power Calculation Method	**Noncentral t Distribution**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
Sample Size Per Group	**6 8 10 12 14 16 18 20**
EU (Upper Biosimilarity Limit)	**27**
EL (Lower Biosimilarity Limit)	**-Upper Limit**
δ (Actual Mean Difference)	**2.25**
σ (Standard Deviation)	**18**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Reports

Numeric Results

Solve For:	Power
Groups:	1 = Treatment, 2 = Reference
Difference:	δ = μ1 - μ2 = μT - μR
Hypotheses:	H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Power Calculation Method:	Noncentral t Distribution

	Sample Size			Biosimilarity Limits				
Power	N1	N2	N	Lower EL	Upper EU	Actual Mean Difference δ	Reference Group Standard Deviation σ	Alpha
0.33611	6	6	12	-27	27	2.25	18	0.025
0.58274	8	8	16	-27	27	2.25	18	0.025
0.75798	10	10	20	-27	27	2.25	18	0.025
0.86299	12	12	24	-27	27	2.25	18	0.025
0.92350	14	14	28	-27	27	2.25	18	0.025
0.95773	16	16	32	-27	27	2.25	18	0.025
0.97685	18	18	36	-27	27	2.25	18	0.025
0.98741	20	20	40	-27	27	2.25	18	0.025

Power	The probability of rejecting a false null hypothesis when the alternative hypothesis is true.
N1 and N2	The number of items sampled from each population.
N	The total sample size. N = N1 + N2.
EL and EU	The lower and upper biosimilarity limits, respectively, and are the maximum allowable differences that still result in biosimilarity.
δ	The difference between the treatment and reference means at which power and sample size calculations are made. δ = μ1 - μ2 = μT - μR.
σ	The assumed standard deviation of the reference group.
Alpha	The probability of rejecting a true null hypothesis.

Summary Statements

A parallel, two-group design will be used to test whether the Group 1 (treatment) mean (μ1) is biosimilar to the Group 2 (reference) mean (μ2), with difference biosimilarity bounds of -27 and 27 (H0: δ ≤ -27 or δ ≥ 27 versus H1: -27 < δ < 27, δ = μ1 - μ2 = μT - μR). The comparison will be made using two one-sided, two-sample, equal-variance t-tests, with an overall Type I error rate (α) of 0.025. The common standard deviation for both groups is assumed to be 18. To detect a mean difference (δ = μ1 - μ2 = μT - μR) of 2.25 with sample sizes of 6 for Group 1 (treatment) and 6 for Group 2 (reference), the power is 0.33611.

Dropout-Inflated Sample Size

	Sample Size			Dropout-Inflated Enrollment Sample Size			Expected Number of Dropouts		
Dropout Rate	N1	N2	N	N1'	N2'	N'	D1	D2	D
20%	6	6	12	8	8	16	2	2	4
20%	8	8	16	10	10	20	2	2	4
20%	10	10	20	13	13	26	3	3	6
20%	12	12	24	15	15	30	3	3	6
20%	14	14	28	18	18	36	4	4	8
20%	16	16	32	20	20	40	4	4	8
20%	18	18	36	23	23	46	5	5	10
20%	20	20	40	25	25	50	5	5	10

Dropout Rate	The percentage of subjects (or items) that are expected to be lost at random during the course of the study and for whom no response data will be collected (i.e., will be treated as "missing"). Abbreviated as DR.
N1, N2, and N	The evaluable sample sizes at which power is computed (as entered by the user). If N1 and N2 subjects are evaluated out of the N1' and N2' subjects that are enrolled in the study, the design will achieve the stated power.
N1', N2', and N'	The number of subjects that should be enrolled in the study in order to obtain N1, N2, and N evaluable subjects, based on the assumed dropout rate. N1' and N2' are calculated by inflating N1 and N2 using the formulas N1' = N1 / (1 - DR) and N2' = N2 / (1 - DR), with N1' and N2' always rounded up. (See Julious, S.A. (2010) pages 52-53, or Chow, S.C., Shao, J., Wang, H., and Lokhnygina, Y. (2018) pages 32-33.)
D1, D2, and D	The expected number of dropouts. D1 = N1' - N1, D2 = N2' - N2, and D = D1 + D2.

Dropout Summary Statements

Anticipating a 20% dropout rate, 8 subjects should be enrolled in Group 1, and 8 in Group 2, to obtain final group sample sizes of 6 and 6, respectively.

References

Blackwelder, W.C. 1998. 'Equivalence Trials.' In Encyclopedia of Biostatistics, John Wiley and Sons. New York. Volume 2, 1367-1372.

Chow, S.C. 2014. Biosimilars – Design and Analysis of Follow-on Biologics. Chapman & Hall/CRC Press. Boca Raton, Florida.

Chow, S.C. 2019. Analytical Similarity Assessment in Biosimilar Product Development. Chapman & Hall/CRC Press. Boca Raton, Florida

Julious, S.A. 2023. Sample Sizes for Clinical Trials. Second Edition. Chapman & Hall / CRC Press. Boca Raton, Florida.

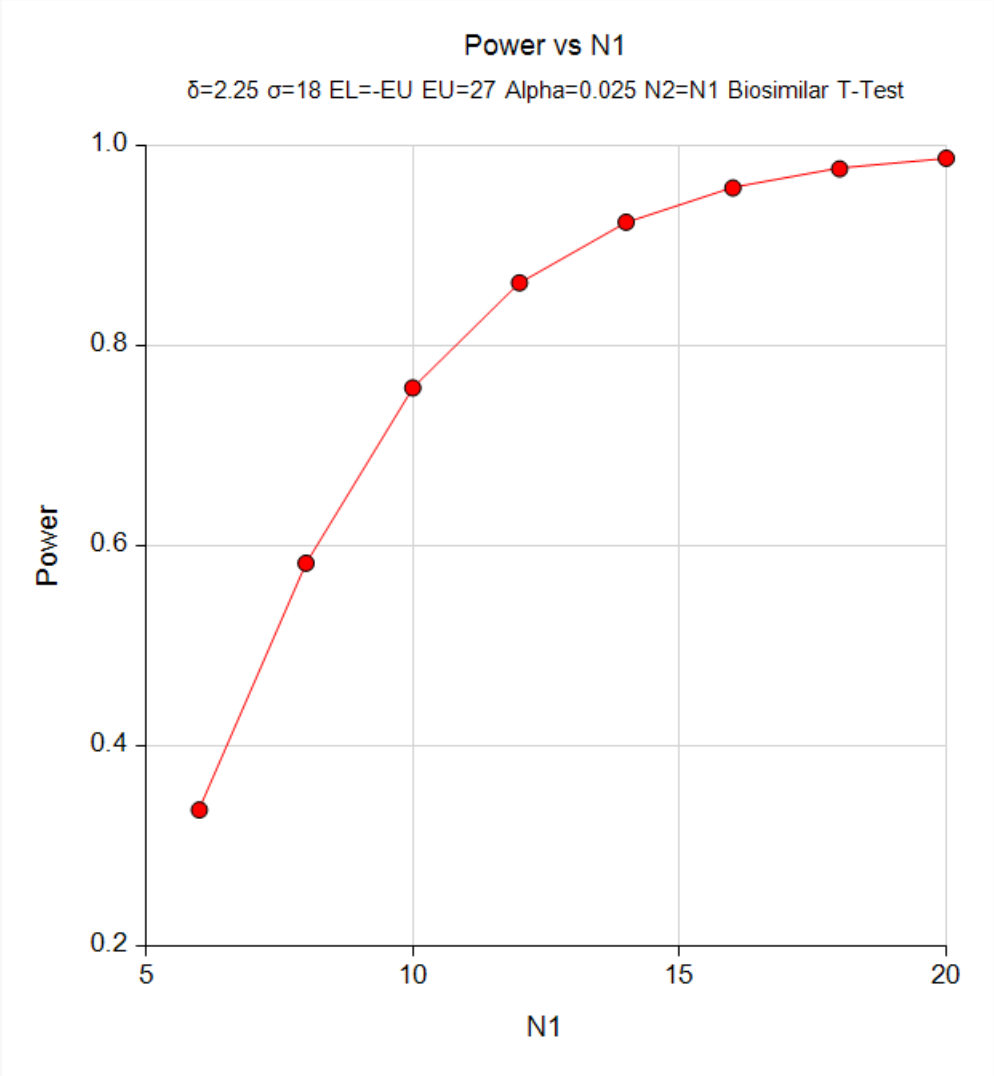
Phillips, K.F. 1990. 'Power of the Two One-Sided Tests Procedure in Bioequivalence', Journal of Pharmacokinetics and Biopharmaceutics, Volume 18, No. 2, pages 137-144.

Schuirmann, D. 1987. 'A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability', Journal of Pharmacokinetics and Biopharmaceutics, Volume 15, Number 6, pages 657-680.

This report shows the power for the indicated parameter configurations. Note that the desired 90% power occurs for a per group sample size between 12 and 14.

Plots Section

Plots



This plot shows the power versus the sample size.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size to achieve 90% power.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 2** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	**Sample Size**
Power Calculation Method	**Noncentral t Distribution**
Power	**0.90**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
EU (Upper Biosimilarity Limit)	**27**
EL (Lower Biosimilarity Limit)	**-Upper Limit**
δ (Actual Mean Difference)	**2.25**
σ (Standard Deviation)	**18**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For:	Sample Size
Groups:	1 = Treatment, 2 = Reference
Difference:	δ = μ1 - μ2 = μT - μR
Hypotheses:	H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Power Calculation Method:	Noncentral t Distribution

Power		Sample Size			Biosimilarity Limits		Actual Mean Difference	Reference Group Standard Deviation	
Target	Actual	N1	N2	N	Lower EL	Upper EU	δ	σ	Alpha
0.9	0.9235	14	14	28	-27	27	2.25	18	0.025

This report shows the exact sample size required for 80% power.

Example 3 – Validation using Julious (2010)

Julious (2010) page 87 presents an example of determining the sample size for testing equivalence in a pain control trial in which the margin of equivalence for the difference is ± 10mm, the actual differences are 0mm (Example 5.1) and 2mm (Example 5.2), the standard deviation is 100mm, the power is 90%, and the significance level is 0.025 with equal sample allocation. Julius (2010) calculates the sample sizes to be 2600 for an actual difference of 0mm and 3306 for an actual difference of 2mm.

Since the biosimilar methodology uses the equivalence formulas, we will use this example to validate this procedure.

Setup

If the procedure window is not already open, use the PASS Home window to open it. The parameters for this example are listed below and are stored in the **Example 3** settings file. To load these settings to the procedure window, click **Open Example Settings File** in the Help Center or File menu.

Design Tab	
Solve For	**Sample Size**
Power Calculation Method	**Noncentral t Distribution**
Power	**0.90**
Alpha	**0.025**
Group Allocation	**Equal (N1 = N2)**
EU (Upper Biosimilarity Limit)	**10**
EL (Lower Biosimilarity Limit)	**-Upper Limit**
δ (Actual Mean Difference)	**0 2**
σ (Standard Deviation)	**100**

Output

Click the Calculate button to perform the calculations and generate the following output.

Numeric Results

Solve For:	Sample Size
Groups:	1 = Treatment, 2 = Reference
Difference:	δ = μ1 - μ2 = μT - μR
Hypotheses:	H0: δ ≤ EL or δ ≥ EU vs. H1: EL < δ < EU
Power Calculation Method:	Noncentral t Distribution

Power		Sample Size			Biosimilarity Limits		Actual Mean Difference	Reference Group Standard Deviation	
Target	Actual	N1	N2	N	Lower EL	Upper EU	δ	σ	Alpha
0.9	0.90011	2600	2600	5200	-10	10	0	100	0.025
0.9	0.90006	3305	3305	6610	-10	10	2	100	0.025

Note that **PASS** obtains sample sizes of 2600 and 3305, which are equal to those calculated by Julious (2010) with slight differences due to rounding.