Biostatistics and Biometrics
Open Access Journal
ISSN: 2573-2633

Research Article
Volume 11 Issue 4 - March 2024
DOI: 10.19080/BBOAJ.2024.11.555820

Biostat Biom Open Access J

Evaluation of the Utility of Pilot Studies in Establishing Bioequivalence

Martino Recchia*

Medistat, Clinical Epidemiology and Biostatistic Unit, Milano; Mario Negri Institute Alumni Association (MNIAA) ,Italy.

Submission: January 18, 2024; **Published:** March 21, 2024

***Corresponding author:** Martino Recchia, Medistat, Clinical Epidemiology and Biostatistic Unit, Milano; Mario Negri Institute Alumni Association (MNIAA) ,Italy, Email id: statmed@hotmail.com

Abstract

This study aims to assess the actual utility of the sample size tables calculated by Diletti et al. [1] for planning bioequivalence studies. This objective emerged from numerous simulation experiments using data from a real experiment that, during routine statistical analysis (crossover ANOVA), demonstrated complete bioequivalence between reference and test substances.

For each simulated pilot number (N=6, N=8, ..., N=18), we calculated 10,000 %CV values needed to verify the recommended sample size in a real experiment to achieve a bioequivalence response. The comparison between the size proposed by Diletti and that suggested by the simulation was negative due to the high and low under- and over-estimations obtained. Equal estimates were very rare.

We propose conducting a pilot experiment comparing three volunteers in each group. These real data can then be used to simulate 10,000 experiments of 4 vs. 4 and calculate 10,000 Schuirmann tests, checking how often bioequivalence is achieved. If power is below 70%, the next step involves 10,000 simulations of 5 vs. 5, and so on until the desired power is reached. This provides the required sample size [2] for the final experiment.

Keywords: Bioequivalence; Crossover; Sample Size; Resampling; Bootstrap Method

Introduction

To establish a coherent experimental plan for investigating bioequivalence, the following steps are necessary:

i. Determine the required power (70%, 80%, ..., as the probability of correctly establishing the existence of bioequivalence).

ii. Establish an alpha value (5%, 2.5%, ..., as the probability of observing bio-inequivalence).

iii. Calculate a test/reference ratio (Ut/Ur) (0.95, 1, 1.05,...) and a suitable bioequivalence interval (80-125%, ...).

iv. Calculate the sample size for the experiment.

The complex problem raised in step 4 can be addressed by following this strategy:

i. Assume the s2 variability of the parameter (AUC, Cmax).

ii. Select an appropriate experimental design for observations to compensate for potential detrimental effects of variability on the outcome (ANOVA Crossover design - AC).

Unfortunately, the most challenging step is determining the value of s^2. It is common practice to rely on the results of previous studies or conduct a "pilot trial".The latter should allow solving the following formula for the sample size suggested by Diletti et al. [1] for the correct size of a trial (1):

$$CV = \sqrt{\exp(\sigma^2) - 1} \quad \text{with } s^2 \text{ in place of } \sigma^2$$

Digression

In a comparative clinical trial, the clinical researcher tries to detect a difference between treatments by ensuring that the study is adequately sized, based on the premise that more observations provide more information. Consequently, the null hypothesis of drugs with the same effect (Ho) will have a high probability of being rejected in favor of the alternative, that the drugs have different efficacy (H1). Experimental pharmacokinetics, however, follows a different path. Here, the scientist tries to validate the null hypothesis Ho to demonstrate that a treatment has at least the

1Medistat, Clinical Epidemiology and Biostatistic Unit, Milano; Mario Negri Institute Alumni Association (MNIAA)

2The bootstrap method says: "Define a hypothetical universe - the sample itself - representing the best fit for the real universe. Extract a large number of samples from it and examine the distribution." The regular method says: "Describe the hypothetical distribution – the normal one - that would have generated the sample. Then use parameters such as the average and standard deviation to describe the distribution. Finally, resort to formulas and tables to verify the features of the samples extracted from that hypothetical universe."

same activity as an existing one (bioequivalence). In the domain of statistical inference, however, there is no real reason for such a conclusion. At most, a negative inference test can indicate that the observations made were not actually inconsistent with the null hypothesis Ho. This explains the need to resort to statistical procedures to determine the most appropriate sample size and assess the bioequivalence of the tested drugs (Schuirmann test) [3]. End of digression.

Methods

Considering the importance of s^2, we posed the following question: "Can the coefficient of variation (CV) calculated from s^2 obtained in a pilot study indicate the optimal sample size to demonstrate bioequivalence?" To answer this, we analyzed Cmax data from a clinical study on 18 volunteers, following the scheme illustrated in (Tables 1&2).

F treatments=0.02206, F carry-over=0.2069, F period=0.66536; Schuirmann test: t_1=2.4, t_2=4.0, both significant at p<0.01 with 16 df from the AC design. At the conclusion of the statistical analysis, bioequivalence proved indisputable, allowing us to virtually consider Cmax as if it came from a single population where we could randomly construct many simulated versions of the experiment presented in Table II, using a different number of individuals in each replication (our pilot experiments).

Table 1: The group-by-period means – Cmax.

Group	Period 1	Period 2
1 (RT)	14.6	14.8
2 (TR)	14.9	16.1

Where R= Reference; T= Test

Table 2: Individual volunteers' data – Cmax.

Vol.	R	T	Vol.	T	R
2	14.4	14.5	1	10.3	9.3
4	16.3	13.8	3	17.4	15.1
6	9.3	12.6	5	14.5	12.8
8	10.7	15.7	7	10.5	13.3
11	13.8	12.9	9	13.7	13.5
12	27.8	24.2	10	23.2	18.3
14	15.0	13.2	13	16.3	17.2
16	12.6	12.7	15	13.3	28.8
18	11.4	13.4	17	15.1	16.4

Where Vol.= Volunteers

Simulations – Bootstrap Approach

Search for the smallest number needed to establish bioequivalence

Bootstrap methods are intensive statistical analysis methods that use simulation to calculate standard errors, confidence intervals, and significance tests. The key idea is to sample from the original data, directly or using a fitted model, to obtain replicated datasets from which the variability of the quantities of interest can be assessed without lengthy and error-prone analytical calculations.

From the population in Table II, we extracted 10,000 random samples (2) of N=18 each. We then calculated the Schuirmann test [3] for each (3) and found that in 75% of the simulations, bioequivalence could be demonstrated with 18 volunteers (75% power). With 16 volunteers, the power was lower, at 71% (Table 3).

Table 3: Minimal sample size.

Bootstrap samples	% Power ventied by Schuirmann test
5 RT vs 5 TR	50.0
6 RT vs 6 TR	53.0
7 RT vs 7 TR	61.0
8 RT vs 8 TR	71.4
9 RT vs 9 TR	74.5

Further simulations and Bootstrap samples[2] for the AC design

From Table II, we selected a first random sample with replacement[3] consisting of three volunteers from the RT group (total of six measurements: three in the first period with treatment R and three in the second with treatment T); the second sample included three volunteers from the TR group, following the same criteria as the first (Table 4).

From this information package, using the AC design, we obtained the following results for

F treatment = 0.64967 ; s^2 = 0.06065 ; %CV = 24.9

$$\%CV = \sqrt{\exp(\sigma^2) - 1} \; . \; 100$$

We then asked: "To what sample size does an estimated %CV from six randomly selected volunteers (Table 4) lead?"

The tables of Diletti et al. [1] provide a number of 20 patients for a power of 70%, α= 5%, Ut/Ur = 1, and a confidence interval for bioequivalence ranging from 0.8 to 1.25. We performed another 9,999 simulations for experiments with three or eight volunteers per group, meaning 9,999 Latin square ANOVAs, as suggested

[3]In sampling a finite population, the use of the "with replacement" technique makes the "n" observations independent and equally distributed, validating the Central Limit theorem.

How to cite this article: Martino R. Evaluation of the Utility of Pilot Studies in Establishing Bioequivalence. Biostat Biom Open Access J. 2024; 11(5): 555820. DOI: 10.19080/BBOAJ.2024.11.555820

by Jones & Kenward (4). This large number of repetitions ensures considerable stability of the results, allowing a correct interpretation of %CV distributions (one for each N) needed to obtain the proposed 10,000 sample sizes to verify the reliability of power tables. In this case, s^2 refers to the W-S residual or error (b) of a Split-Plot design; the two values are identical. Diletti et al. [1] argue that to calculate CV, "the mean square of the error, obtained from ANOVA, must be used after transforming the experimental data into LN values."

All these simulations led to the following conclusions.

Table 4: Cmax in 12 data sets from two-formulation crossover design. LN transformation.

Vol.	R	T	Vol.	T	R
2	2.6672	2.6741	5	2.6741	2.5494
11	2.6247	2.5572	9	2.6174	2.6027
16	2.5337	2.5416	15	2.5878	3.3604
Mean	2.6085	2.591	**Mean**	2.6264	2.8375
SdDev	0.0682	0.0724	**SdDev**	0.0439	0.4536
n	3	3	**n**	3	3

70% Power

The results of the pilot study of 10,000 (Table 5) do not confirm the number of the two Cmax populations we started with (18 volunteers) or the 8 vs. 8, with 16 volunteers, which gave a power of 70%.

Table 5: Sample size suggested by Diletti et al. using 10.000 simulated %CV values (pilot study witn N=6).

Sample size to attain a power of 70%	Count	Count/10.000 or probability
4	1456	0.14560
6	1227	0.12270
8	1416	0.14160
10	1673	0.16730
12	1025	0.10250
14	328	0.03280
16	350	0.03500
20	613	0.06130
24	945	0.09450
28	700	0.07000
>28	267	0.02670
Total	10,000	

The main information obtained from the above table is:

i. the high probability of underestimating the size of the original population

(71.25%, i.e., 1456+...+328=7125/10,000)

ii. the probability of overestimation (19.12%, i.e., 1912/10,000)

iii. the disappointing proportion of "pilot studies" suggesting the correct sizes for the experiment (9.63%, between 16 and 20 volunteers, i.e., 963/10,000).

(Table 6) lists various tables obtained from similar calculations as those discussed but for pilot studies with 8-16 volunteers.

These additional simulations confirm the disappointing sizing information already observed in the study with six volunteers (Table 7).

In conclusion, with every change in N, the probability of underestimating sample sizes increases, the probability of an equal estimate worsens and the probability of overestimation decreases. In any case, this procedure provides disappointing information. Our population consisted of 18 volunteers, and the simulation provided this data with a very low frequency.

The simulation results make the sample size formula shown at the beginning less useful since both the %CV calculation and, subsequently, the sample size calculation is based on hypotheses hardly satisfied in the real-life pharmacological response.

80% Power

We repeated the entire procedure, changing the power from 70% to 80%, to operate under more typical conditions of experimental planning. The results of these simulations are represented in (Tables 8,9&10).

Discussion

The breadth of the simulated confidence interval makes power size tables practically useless (Table 11). The distribution of the residual error from ANOVA applied to LN-transformed data, does

How to cite this article: Martino R. Evaluation of the Utility of Pilot Studies in Establishing Bioequivalence. Biostat Biom Open Access J. 2024; 11(5): 555820. DOI: 10.19080/BBOAJ.2024.11.555820

not correspond to what should result from the Owen algorithm, which is what Diletti et al. [1] used, with some modifications, to compile their sample size tables (2). T1 and T2 also use the same error, making it impossible to apply %CV. But T1 and T2 benefit from the "robustness" of the Student's t-test, used to calculate them.

Table 6:. Sample size suggeted by Diletti et al. Using 10.000 simulated %CV values.

Sample size to attain a power of 70%	Probability Pilot study with N=8	Probability Pilot study with N=10	Probability Pilot study with N=12	Probability Pilot study with N=14	Probability Pilot study with N=16
4	0.08260	0.04840	0.19580	0.01860	0.01030
6	0.11290	0.09800	0.02930	0.07010	0.06020
8	0.17120	0.19220	0.08570	0.18600	0.18190
10	0.18200	0.17770	0.16440	0.14580	0.13480
12	0.06840	0.04570	0.04320	0.06290	0.09660
14	0.32100	0.07050	0.14720	0.23470	0.27720
16	0.09440	0.18520	0.21140	0.16640	0.13670
20	0.01453	0.13030	0.08100	0.09030	0.08680
24	0.08990	0.04020	0.03920	0.02400	0.01450
28	0.01820	0.01140	0.00270	0.00120	0.00100
>28	0.00300	0.00040	0.00010		
Total	10000	10000	10000	10000	10000

Table 7: Probability (%) of sizing calculated from 10,000 simulations-Power 70%.

Estimate	Pilot study				
	N=8	N=10	N=12	N=14	N=16
	% Probability of different estimates				
Under	58.08	58.68	62.24	65.52	66.44
Equal	23.97	31.55	29.24	25.67	22.35
Over	11.11	5.20	4.20	2.52	1.55

Table 8: Sample size suggested by Diletti et al. using 10.000 simulated %CV values (pilot study witn N=6).

Sample size to attain a power of 80%	Count	Count/10.000 or probability
4	415	0.04150
6	2286	0.22860
8	1449	0.14490
10	1640	0.16400
14	1035	0.10350
16	319	0.03190
20	344	0.03440

How to cite this article: Martino R. Evaluation of the Utility of Pilot Studies in Establishing Bioequivalence. Biostat Biom Open Access J. 2024; 11(5): 555820. DOI: 10.19080/BBOAJ.2024.11.555820

24	661	0.06610
28	876	0.08760
32	717	0.07170
>32	258	0.02580
Total	10,000	

Table 9: Sample size suggeted by Diletti et al. Using 10.000 simulated %CV values.

Sample size to attain a power of 80%	Probability Pilot study with N=8	Probability Pilot study with N=10	Probability Pilot study with N=12	Probability Pilot study with N=14	Probability Pilot study with N=16
4	0.01890	0.00670	0.00260	0.00140	
6	0.17590	0.14310	0.10750	0.08750	0.06880
8	0.17620	0.19320	0.19400	0.18860	0.17730
10	0.18060	0.17590	0.16840	0.14520	0.14070
14	0.06970	0.05000	0.04390	0.06270	0.09710
16	0.03130	0.06630	0.14500	0.23710	0.27400
20	0.08630	0.18010	0.21220	0.16500	0.13800
24	0.15180	0.13400	0.08360	0.09090	
28	0.08750	0.04040	0.03930	0.02050	0.01350
32	0.01800	0.00990	0.00340	0.00110	0.00080
>32	0.00380	0.00040	0.00010		
Total	10000	10000	10000	10000	10000

Table 10: Probability (%) of sizing calculated from 10,000 simulations-ower 80%.

Estimate	Pilot study				
	N=8	N=10	N=12	N=14	N=16
	% Probability of different estimates				
Under	62.13	56.89	51.64	48.54	48.42
Equal	11.76	24.64	35.72	40.21	41.20
Over	26.11	18.47	12.64	11.25	10.36

Table 11: Confidence interval (CI) for CV% and sample size, drawing bootstrap samples of different sizes (Ni).

Simulated study	No. Vol.	Bootstrap Mean for %CV	CV	
			Bootstrap	CI 95%
3 RT vs 3TR	N=6	15.556	4.109	30.066
4 RT vs 4 TR	N=8	16.029	5.385	27.219
5 RT vs 5 TR	N=10	16.238	6.284	
6 RT vs 6 TR	N=12	16.372	7.239	25.728
7 RT vs 7 TR	N=14	16.585	7.898	

How to cite this article: Martino R. Evaluation of the Utility of Pilot Studies in Establishing Bioequivalence. Biostat Biom Open Access J. 2024; 11(5): 555820. DOI: 10.19080/BBOAJ.2024.11.555820

8 RT vs 8 TR	N=16	16.644	8.597	24.427
9 RT vs 9 TR	N=18	16.719	8.994	23.963

Suggestion

A true pilot experiment should be conducted with six volunteers (3 vs. 3). The results could then be used to simulate an initial set of 10,000 T1 and T2 calculated for samples with eight volunteers (4 vs. 4) using the simulation method; then a second series could be performed, with..., until 80% of the 10,000 pairs of T1 and T2 show bioequivalence. The number, N, confirming this power can then be used for the final experiment.

All statistical procedures were performed on a MacBook Pro computer using (2) and JMP14 Pro of the Sas Institute Inc.

Acknowledgments

The author thanks Dr. Ambrogio Tattolo, Dr. Maurizio Rocchetti, Dr. Giulio Serra, and Dr. Sergio Di Matteo for the critical reading of this article.

References

1. Diletti E, Hauschke D, Steinijans VW (1992) Sample size determination for bioequivalence assessment by means of confidence intervals. Int. J Clin Pharm Ther Toxicol 29: 551-558.
2. Resampling Stats, Inc. (2000) 612 N. Jackson Street. Arlington, Virginia 22201, USA.
3. Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics 15: 657-680.

DOI: 10.19080/BBOAJ.2024.11.555820

How to cite this article: Martino R. Evaluation of the Utility of Pilot Studies in Establishing Bioequivalence. Biostat Biom Open Access J. 2024; 11(5): 555820. DOI: 10.19080/BBOAJ.2024.11.555820