Evaluating the Bioequivalence of Treatments Using t-ratios Tests

M. Filomena Teodoro
CEMAT - Center for Computational and Stochastic Mathematics, Insituto Superior Técnico, Lisbon University
Portugal
CINAV, Center of Naval Research, Naval Academy, Military Universitary Institute, Portuguese Navy
Portugal
mteodoro64@gmail.com

Marina A. P. Andrade
ISTAR, University Institute of Lisbon
Portugal
ISCTE, University Institute of Lisbon
Portugal
marina.andrade@iscte.pt

Teresa A. Oliveira
CEAUL, Faculdade de Ciências, Lisbon University
Portugal
Departamento de Ciências e Tecnologia, Universidade Aberta
Portugal
Teresa.Oliveira@uab.pt

Abstract

The main objective of the work is the application of experimental design and some simple tests (t-ratios) in the evaluation of bioequivalence of two treatments. It is performed the assessment of the effectiveness of different treatments administered to two groups of patients: the control group and the test group. The comparison of the control group that received the reference medication with the control group test, which received the drug under study, using a crossover experimental design. It was performed an analysis using t-ratios and considering and inter-subject variability formulation of the probabilistic model, where the specification of tests for detection of *carryover* and period effects were considered. The application to a practical case consisted in the validation of the hypothesis of the proposed model took place, where the effects *carryover* and period were not significant when considering usual levels of significance. About the treatments, we could conclude that there was statistical significance to affirm that the studied pharmaceuticals were not bioequivalent.

CCS Concepts

• **Experimental design**; • **Clinical trials**; • **Bioequivalence**;

Keywords

Computational methods, t-Tests

ACM Reference Format:
M. Filomena Teodoro, Marina A. P. Andrade, and Teresa A. Oliveira. 2024. Evaluating the Bioequivalence of Treatments Using t-ratios Tests. In *2024 7th International Conference on Mathematics and Statistics (ICoMS 2024), June 23–25, 2024, Amarante, Portugal*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3686592.3686610

ICoMS 2024, June 23–25, 2024, Amarante, Portugal

ACM ISBN 979-8-4007-0722-3/24/06
https://doi.org/10.1145/3686592.3686610

1 Introduction

Clinical trials appear in a large number of distinct contexts, e. g. in [1], a randomized trial is performed so a decision-making for patients with acute myocardial infarction can be taken; or in [2], where the care management of insomnia occurring simultaneously with sleep apnea is studied, the authors of [3] compare the effectiveness of virtual reality game to clinic based therapy for chronic non-specific low back pain; in [4], it is described the allocation of patients to in headache clinical trials randomization, in [5] is made some designing for cancer clinical trials. We can find a guide with several methods about how to allocate patients in clinical trials in [6]. Later, in [7], the same author publish a more general work where, in a simple language, it is described how to design, analyze and interpret, including a historical context, the current status and a future perspective about clinical trials. Similar works can be found in [8], where Bayesian adaptive biased coin designs with Gaussian response are described for clinical trials. Also, the authors of [9] made some interesting considerations about randomization-based inference. The author of [10] made restricted randomization designs in clinical trials, in headache clinical trials; the authors of [4] proposed the randomization, stratification, and treatment matching to allocation of patients. Some recent works proposed to improve the efficiency of the clinical trials. In [11], the authors propose a new algorithm gaining efficiency in randomized controlled trials. The authors of [12] explore some recent innovations in clinical trials where a single therapeutic treatment is tested on several patient populations, each of which forms a basket, using Bayesian hierarchical models. A qualitative study is developed by the authors of [13], where some randomization methods in clinical trials are compared and evaluated. In the case of our study, we are interested in bioequivalence trials. The authors of [14] explore statistical tools focused on clinical pharmacology under the theme of bioequivalent trials.

This work is divided in five sections. An introduction starts the manuscript; in Section 2 some details about bioequivalence of treatments are described. Follows some details about the methodology in Section 3. In Section 4 some issues about the empirical application are performed, starting with the basic statistical techniques, including the evaluation of some tests about the normality of data, followed by the construction and interpretation of the *t-tests*. In Section 5 are detailed and discussed some final remarks.

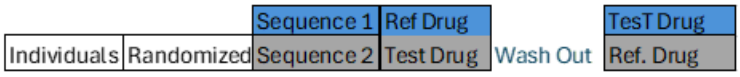

Figure 1: Scheme of experimental design: *2×2 crossover.*

2 Preliminaries

The main objective of this work is to verify the effectiveness of treatment when administered to patients, as well as to verify and to analyze clinically relevant reactions. Specifically, the clinical trial here addressed involves comparing a control group with a group receiving treatment. Despite involving a limited number of individuals, it is expected to apply the results to decide which treatment to use in the target population (with the same medical condition). In detail, the aim is to evaluate the existence or not of bioequivalence of a test pharmaceutical or treatment (T) and a reference pharmaceutical or treatment (R). To prove the existence of bioequivalence between therapeutic treatments it is not necessary to conclude that one is better than the other, just that it is not worse. In a first phase, the bioavailability of people taking the medicine is measured. This action consists of recording the concentration in some fluid over time after taking the reference medication (and test respectively), constructing the concentration versus time graph. This is followed by the calculation of the pharmaco-kinetic parameters of interest such as maximum concentration, area under the curve (AUC) or others. Knowing that the same individual must receive the reference and the test drug, the chosen design was a *2×2 crossover* given the importance of the pairing. An even number of individuals were randomized into two groups. One of the groups receives the reference medicine in the first period and, in the second period, the test drug. The second group receives the test drug in the first period and the reference drug in the second period. The interval between the two periods of intake is supposed to be long enough to fade the effect of taking the first medication. This time interval is called wash-out (see Figure 1). This type of design makes it possible to test whether there are reinforcement effects in the 2nd period, whether the treatments are different or whether no tests should be carried out on individuals.

To review some basic concepts of experimental design, analysis of variance and some specific t-tests we can consult, for example, [15–18].

3 Theoretical Model

The statistical model used to determine bioequivalence has as response variable Y_{ijk} a random variable (r.v.) whose observed value $y_{i\,jk}$ corresponds to the $i-th$ individual, in the $k-th$ sequence of the $j-th$ period, with $k = 1...\ p$, $i = 1...nk$, and $j = 1,...,\ p$.

In general, data is given in tabular form (1).

$$\begin{array}{lll} & \text{Period1} & \text{Period2}\cdots\text{Period}p \\ \text{Sequence1} & y_{111}\cdots y_{n111} & y_{121}\cdots y_{n121}\cdots y_{1p1}\cdots y_{n1p1} \\ \text{Sequence2} & y_{112}\cdots y_{122} & y_{221}\cdots y_{n222}\cdots y_{1p2}\cdots y_{n2p2} \\ . & & \\ \text{Sequence}g & y_{11g}\cdots y_{ng1g} & y_{12g}\cdots y_{ng2g}\cdots y_{1pg}\cdots y_{ngpg} \end{array} \tag{1}$$

The theoretical model is defined in equation (2).

$$Y_{ijk} = \mu + S_{ik} + P_j + F_{jk} + C_{jk} + \varepsilon_{ijk}, \tag{2}$$

where

- μ is the global mean (or global effect);
- S_{ik} is the random effect of the $i-th$ individual at the $k-th$ sequence, $k = 1,...,g$, and $i = 1,...,n_k$;
- P_j is the fixed effect of the $j-th$ period, $j = 1,...,\ p$;
- F_{jk} is the fixed effect of the administered formulation at the $k-th$ sequence during the $j-th$ period reference R or test T);
- C_{jk} is the carry-over first order fixed effect of the administered formulation at the $k-th$ sequence during the $j-1-th$ period;
- $\epsilon_{i\,jk}$ is the random error of $Y_{i\,jk}$.

The following restrictions are considered:

$\sum_j P_j = 0$, $\sum_{j,k} F_{jk} = 0$, $\sum_{j,k} C_{jk} = 0$ and , $C_{0k} = 0$ conducting to

- $p-1$ effects $\mathrm{P_j}$ linearly independents;
- $(p-1)\ (g-1)$ effects F_{jk} linearly independents;
- $(p-1)\ (g-1)$ effects C_{jk} linearly independents.

The hypotheses inherent to the model (2) are:

- The Sik are independent and Gaussian, $S_{ik} \cap N(0,{\sigma_s}^2)$;
- The $\epsilon_{i\,jk}$ are independent and Gaussian, $\epsilon_{i\,jk} \cap N(0,{\sigma_e}^2)$;
- The $\{S_{ik}\}$ and $\{\epsilon_{i\,jk}\}$ are mutually independent.

We can conclude that $Y_{i\,jk}$ are r.v. independents and Gaussian. Also, we need to evaluate the variability due the subjects, namely, to compute the proportion of total variability explained by the individuals.

The σ_s estimate is used to explain the variability inter-subject, the σ_{e} estimate is used to explain the variability intra-subject. Considering the design crossover *2×2*, $n_k = n_1 = n_2$, $p = 2$, $g = 2$ and the scheme of date presented in tabular form 1 is reduced to the form displayed in Table 1.

If we consider the administration of the drug test in the first period of the first sequence, we obtain the definition of the effect F_{jk} and the effect C_{jk} respectively given by equations (3) and (4), noting that the effect carry-over only occurs during the second period.

$$F_{\mathrm{jk}} = \begin{cases} \mathrm{F_R}, & se\ k = j \\ \mathrm{F_T}, & se\ \ k \neq\ j \end{cases},\ k = 1, 2,\ \ j = 1, 2, \tag{3}$$

$$C_{\mathrm{jk}} = \begin{cases} \mathrm{C_R}, & se\ k = 1,\ j = 2 \\ \mathrm{C_T}, & se\ \ k = \ j = 2 \end{cases}. \tag{4}$$

The fixed effect that occurs in each period and in each sequence is given by the expected value for each individual $\mu_{\ \mathrm{jk}}$:

		Period I	Period II
Sequence 1	RT	$\mu_{11} = \mu + \mathrm{P_1} + \mathrm{F_R}$	$\mu_{21} = \mu + \mathrm{P_2} + \mathrm{F_T} + \mathrm{C_R}$
Sequence 2	TR	$\mu_{12} = \mu + \mathrm{P_1} + \mathrm{F_T}$	$\mu_{22} = \mu + \mathrm{P_2} + \mathrm{F_R} + \mathrm{C_T}$

Table 1: Scheme of experimental design: $n_k = n_1 = n_2$, $p = 2$, $g = 2$

		Period I	Period II
Sequence 1	Reference-Test (RT)	$y_{111} \dots y_{n111}$	$y_{111} \dots y_{n111}$
Sequence 2	Test-Reference (TR)	$Y_{121} \dots y_{n221}$	$Y_{122} \dots y_{n222}$

with

$$\begin{cases} P_1 + P_2 = 0 \\ C_R + C_T = 0 \\ F_R + F_T = 0. \end{cases}$$

In a study of bioequivalence between treatments, it is usual to consider the non-existence of the carry-over effect. This fact can be achieved with a long period of washout, allowing to eliminate the residuals of the drug administration. In this situation, the model is reduced to formula (5).

$$y_{ijk} = \mu + S_{ik} + P_j + F_{jk} + \varepsilon_{ijk} \tag{5}$$

In the case of the model (2), the basic objectives are to test if $P_1 = P_2$, $C_R = C_T$, $F_R = F_T$ and the variability of the residuals σ_s. Summarizing the tests of interest, we have:

- $H_{0C}: C_R = C_T$ versus $H_{1C}: C_R \neq C_T$. (A)

A rejection of H_{0C} means that there is no experimental evidence that the washout period is long enough. The process shall be ended. The following tests assume that there is no *carryover*.

- $H_{0F}: F_R = F_T$ versus $H_{1F}: F_R \neq F_T$. (B)

A rejection of H_{0F} means that there is experimental evidence that the treatments have distinct effects, so there is no experimental evidence for bioequivalence between the treatments.

- $H_{0s}: \sigma_s^2 = 0$ versus $H_{1s}: \sigma_s^2 \neq 0$. (C)

A rejection of H_{0s} means that there is experimental evidence that the individual effects are distinct, meaning that the random factor S_{ik} is statistically significant.

- $H_{0P}: P_1 = P_2$ versus $H_{1P}: P_1 \neq P_2$. (D)

A rejection of H_{0P} means that there is experimental evidence that the period has distinct effects.

4 Methodology

The comparison of the bioavailabilities of the two formulations can be made using variance analysis for crossover trials, with the model residuals being analysed using either inter-subject variability or intra-subject variability, like the authors of [19]. In the present work, we follow the methodology proposed in [7], using *t-ratios* for paired samples. In an initial phase, to detect whether the reference treatment and the test treatment have different effects, differences are created between the values recorded after administration of the test formulation and the values recorded after the reference formulation, regardless of the sequence. It is similar to a sample of n_1+n_2 observations from a supposedly normal population, where two observations were made on the same individual, separated by a long-time interval and we wanted to test the mean of the difference in the records using a t-test using a small, paired sample taken from a normal population. Considering $Y_{T\text{-}R}$ the difference of the values recorded after administration of the test formulation and the values recorded after the reference formulation, we have the statistic of test given by formula (6),

$$T_F = \frac{\bar{Y}_{R-T}}{\sqrt{\frac{S^2_{R-T}}{n_1+n_{2-1}}}} \underbrace{\sim}_{under\ Ho} t_{[n_1+n2\ -1]} \tag{6}$$

with $\bar{Y}_{T\text{-}R}$ the sample mean estimator for the mean μ_{T-R} of $Y_{T\text{-}R}$. The statistic given by formula (6) is the test statistic for the hypothesis

- $H_0: \mu_{T-R} = 0$ versus $H_1: \mu_{T-R} \neq 0$.

Considering a significance level $\alpha \times 100\%$, the critical region is given by $|t_{|Ho|} > t_{(n1+n2-1,1-\alpha/2\ )}$. This test assumes no period and carryover effects. These two assumptions must be tested.

The period effect was already identified previously (B). The formula (7) is a possible test statistic

$$T_p = \frac{\bar{Y}_{R-T,1} - \bar{Y}_{R-T,2}}{\sqrt{\frac{S^2_{R-T,1}}{n_1} + \frac{S^2_{R-T,2}}{n_2}}} \underbrace{\sim}_{under\ Ho} t_{[n1+n2-2\ ]}, \tag{7}$$

with the difference of the values recorded after administration of the test formulation and the recorded values after the reference formulation for each sequence given by $Y_{T-R,k}$, k = 1, 2 , the respective sample mean $\bar{Y}_{T\text{-}R,k,}$ and mean $\mu_{T-R,\,k}$. The denominator of the T_P statistic can be reformulated if there is experimental evidence for the variability in sequences 1 and 2 to be identical (see formula (8)).

$$T_p = \frac{\bar{Y}_{R-T,1} - \bar{Y}_{R-T,2}}{S_{R-T,pooled}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \underbrace{\sim}_{under\ Ho} t_{[n1+n2-2\ ]} \tag{8}$$

Considering a significance level $\alpha \times 100\%$, the critical region is given by $|t_{|Hop|} > t_{p(n1+n2-2,1-\alpha/2\ )}$.

To test the *carryover* effect, the hypotheses have already been identified above (A). The test statistics T_C (see formula (9)) considers the averages per individual for each sequence, sequence given by $Y_{\frac{T+R}{2},k}$, k = 1, 2 , the respective sample mean $\bar{Y}_{\frac{T+R}{2},k}$, k = 1, 2 ,

$$T_C = \frac{\bar{Y}_{\frac{T+R}{2},1} - \bar{Y}_{\frac{T+R}{2},2}}{\sqrt{\frac{S^2_{\frac{T+R}{2},1}}{n_1} + \frac{S^2_{\frac{T+R}{2},2}}{n_2}}} \underbrace{\sim}_{under\ Ho} t_{[n1+n2-2\ ]}\ , \tag{9}$$

and mean $\mu_{\frac{T+R}{2},k}$, k = 1, 2.

The denominator of the T_C statistic can be reformulated if there is experimental evidence for the variability in sequences 1 and 2 to be identical (see formula (10)),

$$T_C = \frac{\bar{Y}_{\frac{T+R}{2},1} - \bar{Y}_{\frac{T+R}{2},2}}{S_{R-T,\text{pooled}}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \underbrace{\sim}_{under\ Ho} t_{[n1+n2-2\ ]}. \tag{10}$$

Sequence	Individual	Period I	Period II
RT	1	6934,5	6458
RT	2	5330,5	3108,1
RT	5	5848,3	7079,4
RT	7	8405,2	6832,5
RT	8	7403,5	7125,7
RT	9	7409,4	8417,7
TR	3	7762,2	9034,2
TR	4	8137,3	8107,3
TR	6	4998,1	6455,3
TR	10	8763,9	10406
TR	11	8430,4	9239,6
TR	12	5053,4	5696,3

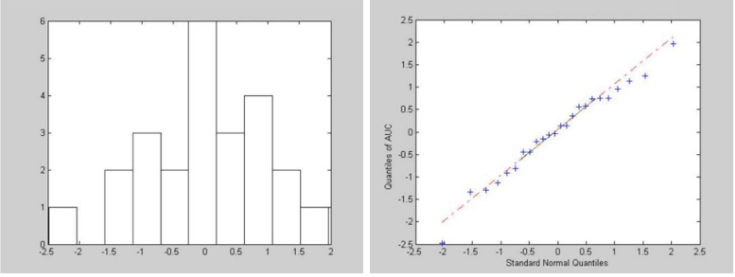

Figure 2: Area under the concentration curve registered along the time (AUC). On left: Data; Center: Histogram; On right: *qq*-plot.

Considering a significance level $\alpha \times 100\%$, the critical region is given by $|t_{c|Hoc}| > t_{c(n1+n2-2,1-\alpha/2)}$.

5 Empirical Application

Two treatments were compared: the reference and the test. The variable of interest is the concentration of a certain treatment after its administration to an individual. The studied variable Y is the area under the concentration curve registered along the time (AUC). For the case under study, n1=n2=6, p=2 and g=2. The registered data is displayed in Figure 2, on left. At center, we observe the histogram from the sample with size 24, at right we find the qq-plot with a lighter tail than a normal on right. At a first glance there is not a big deviance from a Gaussian behavior.

For small sample sizes, the test of normality shall be non-parametric. The Kolmogorov-Smirnov Test (KS) is the most frequent choice. The KS test compares the empirical cumulative empirical distribution with cumulative Gaussian distribution. KS test conduced to p-*value=0,9893.* There is statistical evidence data is from a normal population. The Bera and Jarque test [20] also conduced to the statistical evidence of a Gaussian distribution.

About the *carryover* effect test, we used the general one, without supposing that both sequences have the same variability. Considering formula (9), we got *p-value=0,1896.* Clearly, there is no evidence to reject H_{0C} for common α. The washout period was long enough, there is no evidence of *carryover* effect existence.

Similarly, for the period effect test, without supposing that both sequences have the same variability (formula (7)), we obtain *p-value=0,2495*, implying that there is no statistical evidence to reject the H_{0P} for the common a α. So, there is statistical evidence that does not exist effect period. We can proceed to verify if the two treatments are bioequivalents. Considering the test statistic (6), we obtained *p-value=0,048.* We do reject H_0 for $\alpha \leq 4,8\%$. There are distinct effects in mean for the test treatment and the reference treatment when $\alpha \leq 4,8\%$. This conclusion is valid once there is no carryover effect and no period effect.

6 Conclusion

The main objective of the work was the evaluation of bioequivalence of two treatments using simple *t tests* in opposition of more complicated and robust methodologies just as an ANOVA approach [19] or others. In the present case, we could apply the proposed methodology once the effect period and the effect *carryover* were not significant. It was performed the assessment of the effectiveness of different treatments administered to two groups of patients: the control group and the test group with 6 individuals each.

As there is no evidence for the carryout effect and the period effect to be significant, when using the simple differences in test statistic, it can be considered valid concluding that there is no statistical evidence of bioequivalence of both formulations. This result is similar with the results obtained by the authors of [19] using the same data but a different approach, the analysis of variance. Once the reference drug and the test drug were no bioequivalent, as a future work we can apply a proposal of modified exchangeability nonexchangeability method made in a recent work [12] or take into consideration some of the specific studies presented in [13] and [14].

Acknowledgments

The first author was supported by was supported by Portuguese funds through the Center of Naval Research (CINAV), Naval Academy, Portugal and The Portuguese Foundation for Science and Technology (FCT), through the Center for Computational and Stochastic Mathematics (CEMAT), University of Lisbon, Portugal, project UIDB/Multi/04621/2020, DOI: (https://doi.org/10.54499/UIDB/04621/2020)

The second author was partially supported by Fundação para a Ciência e Tecnologia, I.P. (FCT) [ISTAR Projects: UIDB/04466/2020 and UIDP/04466/2020].

The third author was supported by Portuguese funds, by The Portuguese Foundation for Science and Technology (FCT), through the project UIDB/Multi/00006/2020.

References

[1] Branda, M.E., Kunneman, M., Meza-Contreras, A.I., *et al.* (2022). "Shared decision-making for patients hospitalized with acute myocardial infarction: A randomized trial". Patient Prefer Adherence, 16: 1395–1404.

[2] Eldridge-Smith, E.D., Manber, R., Tsai, S., *et al.* (2022). "Stepped care management of insomnia co-occurring with sleep apnea: the air study protocol". Trials, 23(1): 806.

[3] Fatoye, F., Gebrye, T., Mbada, C.E., *et al.* (2022). "Cost effectiveness of virtual reality game compared to clinic based McKenzie extension therapy for chronic non-specific low back pain". British Journal of Pain, 16(6): 601–609.

[4] Lipchik, G.L., Nicholson, R.A., Penzien, D.B. (2005). "Allocation of patients to conditions in headache clinical trials: randomization, stratification, and treatment matching". Headache, 45(5): (498) 419–428.

[5] Brown, B.W.J. (1980). "Designing for cancer clinical trials: selection of prognostic factors". Cancer Treat Rep., 64(2-3): 499–502.

[6] Pocock, S.J. (1979). "Allocation of patients to treatment in clinical trials." Biometrics, 35(1):183–197.
[7] Pocock, S.J. (2013). "Clinical Trials." 5th edition. e-book. New Jersey: John Wiley & Sons.
[8] Atkinson, A.C., Biswas, A. (2005). "Bayesian adaptive biased-coin designs for bayesian adaptive biased-coin designs for clinical trials with normal responses". Biometrics, 61(1): 118–125.
[9] Gail, M.H., Mark, S.D., Carroll, R.J., *et al.* (1996). "On design considerations and randomization based inference for community intervention trials". Stat Med., 15(11): 1069–1092.
[10] Simon, R. (1979). "Restricted randomization designs in clinical trials. Biometrics", 35(2): 503–612.
[11] Ling, S.X., Hobbs, B.P., Kaizer, A.M., Koopmeiners, J.S. (2021). "Calibrated dynamic borrowing using capping priors". J Biopharm Stat. , 31(6): 852–867.
[12] Daniells, L., Mozgunov, P., Bedding, A., Jaki, T.(2023). "A comparison of bayesian information borrowing methods in basket trials and a novel proposal of modified exchangeability nonexchangeability method". Stat Med., 42(24): 4392–4417.
[13] Bruce, C.L., Iflaifel, M., Montgomery, A., *et al.*, (2024), "Choosing and evaluating randomization methods in clinical trials: a qualitative study". Clinical Trial, 25(1): 199.
[14] Scott D. Patterson, B.J. (2017). "Bioequivalence and Statistics in Clinical Pharmacology". New York: Chapman and Hall/CRC.
[15] Ross, S.M. (2020). "Introduction to Probability and Statistics for Engineers and Scientists. 6th edition". San Diego, CA: Academic Press.
[16] Morgado, L., Teodoro, M.F., Perdicoulis, T. (2010). "Manual de Métodos Estatísticos em CIências Biomédicas". Série Didáctica Ciências Aplicadas, 400, ISBN 978-972-669-018-2, Vila Real, Portugal: Universidade de Trás os Montes e Alto Douro (UTAD).
[17] Larsen, R. and Marx, M.L. (2012). "Statistics. 5th edition". New York: Prentice-Hall Inc.
[18] Tamhane, A., Dunlop, D. (2000). "Statistics and Data Analysis". New Jersey: Prentice-Hall Inc.
[19] Teodoro, M.F., Andrade, M.A.P., Oliveira, T. A. (2024). Crossover Design on the Bioequivalence of Pharmaceuticals-A Case of Study. Gervasy, O. and *et al.* (Edts). Lecture Notes in Computer Science, 14816: 219 - 234 . Cham: Springer.
[20] Thadewald, T., Buning, H. (2007). "Jarque-Bera Test and its Competitors for Testing Normality - A Power Comparison". Journal of Applied Statistics, 34(1): 87-105.