Statistical Comparisons of Product Quality Comparability

Michael Daniel Lucagbo*, Tianjiao Dai[1],, Yixin Ren***, Meiyu Shen**, Yi Tsong****

*University of Maryland, Baltimore County; **U.S. Food and Drug Administration; ***Merck Inc., [1]Presenter, Email: Tianjiao.Dai@fda.hhs.gov

This project was completed through a 2020 ORISE Fellowship.

Disclaimer: This project reflects the views of the authors and should not be construed to represent FDA's views or policies.

Introduction

Test and reference products are considered comparable when the test quality falls (in high percentage) within the quality range determined by the reference product. Figure 1 provides an illustration of the concept of comparability: The reference and test products are said to be comparable if the colored areas are small. Comparability of test and reference products should be established only by the best statistical methods.

Figure 1: Illustration of the comparability of reference and test products

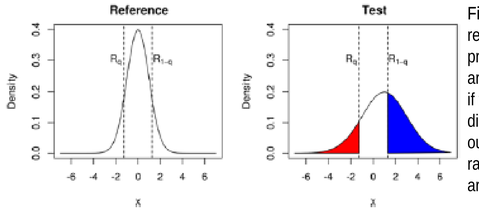

Fig. 1. Test and reference products are comparable if the test distribution are outside quality range (colored area) is small.

To test the comparability of reference and test products, Mielke et al. (2018) propose the following test:

$$H_0: 2q - \left(\left(P(X_T < R_q)\right) + P(X_T > R_{1-q})\right) < c$$

$$H_a: 2q - \left(\left(P(X_T < R_q)\right) + P(X_T > R_{1-q})\right) \geq c,$$

where $X_T \sim N(\mu_T, \sigma_T^2)$ and $X_R \sim N(\mu_R, \sigma_R^2)$ represent the quality values of the Test and Reference products, respectively. $R_q = \mu_R + Z_q\ \sigma_R$ and $R_{1-q} = \mu_R + Z_{1-q}\ \sigma_R$ are, respectively, the lower and upper reference quantiles. The quantity c is called the equivalence margin.

There are a few statistical issues of the approach proposed by Mielke et al. First, in applying Mielke's method, the quantiles are replaced by point estimates $\hat{R}_q$ and $\hat{R}_{1-q}$ based on the sample. However, for small n_R, the random interval $(\hat{R}_q, \hat{R}_{1-q})$ might not give 1−2q coverage. Second, the non-strict inequality should be moved to the null hypothesis to ensure getting cutoff values based on the sampling distribution of the test statistic under the null hypothesis. We refer to this new version as modified Mielke's hypotheses.

Table 1: Equivalent coverage probability p that gives 95% confidence in a two-sided tolerance interval if we use the reference interval $(\bar{X} - 3S, \bar{X} + 3S)$, which gives asymptotic coverage of 99.73%.

Sample size	Coverage
N = 10	91.79%
N = 15	95.35%
N = 20	96.74%
N = 25	97.46%
N = 100	99.15%
N = 200	99.39%

We propose several adjustments to Mielke et al. (2018) procedure and also propose several testing approaches under the modified Mielke's hypotheses. These adjustments are evaluated through simulations.

Method

We propose an adjustment to Mielke et al. (2018) procedure by correcting the coverage to have an interval with 95% confidence. That is, we find p^* such that

$$\left[\frac{(n_R - 1)\,\chi^2_{p^*}(1; 1/n_R)}{\chi^2_\alpha(n_R - 1)}\right]^{1/2} = Z_{1-q}$$

We then replace by $\hat{R}_{q*}$ and $\hat{R}_{1-q*}$, where q^*=(1−p^*)/2, in computing the test statistic. The left side of the equation above is an approximate two-sided tolerance factor for the normal distribution recommended by Krishnamoorthy and Mathew (2009).

Another approach to test Mielke's hypotheses is to compute p^* using the formula below. The right-hand side of the formula is the prediction factor for a normal distribution.

$$\left[\frac{(n_R - 1)\,\chi^2_{p^*}(1; 1/n_R)}{\chi^2_\alpha(n_R - 1)}\right]^{1/2} = t_{1-q, n_R - 1}\sqrt{1 + \frac{1}{n_R}}$$

The final approach to test the hypotheses is to compute $\hat{R}_q$ and $\hat{R}_{1-q}$ as the upper and lower limits (given below) of a two-sided tolerance interval with coverage 1 - 2q and confidence level 1 – α.

$$\hat{R}_q = \bar{X}_R - \left[\frac{(n_R - 1)\,\chi^2_{1-2q}(1; 1/n_R)}{\chi^2_\alpha(n_R - 1)}\right]^{1/2} S_R,$$

$$\hat{R}_{1-q} = \bar{X}_R + \left[\frac{(n_R - 1)\,\chi^2_{1-2q}(1; 1/n_R)}{\chi^2_\alpha(n_R - 1)}\right]^{1/2} S_R.$$

Since the quality range is underestimated when it is assessed by the asymptotic method, we slightly increase the width of the interval by including a correction factor C shown below, so that $\hat{R}_q = \bar{X}_R + CZ_qS_R$. This $\hat{R}_q$ is the MVUE of R_q.

$$C = C(n_R) = \sqrt{\tfrac{n_R - 1}{2}}\,\Gamma\left(\tfrac{n_R - 1}{2}\right)/\Gamma\left(\tfrac{n_R}{2}\right).$$

The modified hypotheses:

$$H_0: 2q - \left(\left(P(X_T < R_q)\right) + P(X_T > R_{1-q})\right) \leq c$$

$$H_a: 2q - \left(\left(P(X_T < R_q)\right) + P(X_T > R_{1-q})\right) > c$$

are tested with the above proposed testing approaches to estimate the quantiles. The test statistic is $w = 2q - \left(\hat{P}(X_T < R_q) + \hat{P}(X_T > R_{1-q})\right)$. We reject H_0 when $w > w_{crit}$, where w_{crit} is estimated such the test give 5% Type I error rate. The simulation results of their evaluation are shown next.

Numerical Results

Simulations are conducted for:

- **Approach 1** (Reference interval): No adjustment is made to Mielke's coverage.
- **Approach 2** (Reference interval with corrected coverage): Using a reference interval to correct the coverage.
- **Approach 3** (prediction interval with corrected coverage): Using a prediction interval with coverage.
- **Approach 4** (tolerance interval): Computing reference quantiles as the upper and lower limits of two-sided tolerance intervals with coverage 1 – 2q and confidence level 1 – α.

Critical values

Using 10,000 simulated samples, we estimate the critical value that give exact 5% Type I error rates. Let $n_R = n_T = 25, c = -q = -0.05, \mu_R = 0, \sigma_R = 1, \mu_T$ vary from 0 to 1 and σ_T be determined from the Ho boundary.

		Critical values			
μ_R	μ_T	No correction	Ref. Interval correction	Pred. Interval correction	Tolerance interval
0	0.0	0.0526	-0.0426	-0.0166	0.0912
0	0.2	0.0521	-0.0414	-0.0194	0.0916
0	0.4	0.0569	-0.0334	-0.0156	0.0930
0	0.6	0.0614	-0.0276	-0.0100	0.0944
0	0.8	0.0704	-0.0177	0.0023	0.0975
0	1.0	0.0787	-0.0158	0.0139	0.0994

Convergence of critical values

The critical values for all approaches converge to $2q - \left[P(X_T < R_q) + P(X_T > R_{1-q})\right]$, as the figure below shows.

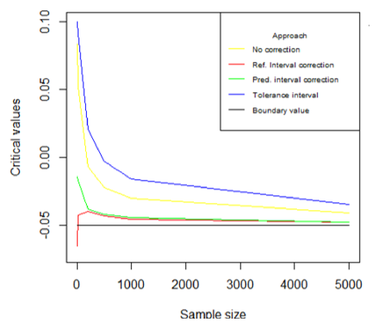

Figure. The critical values for all four approaches to perform Mielke's test converge to boundary at the null hypothesis.

Power calculations

The power computations are made at the point $\mu_R = \mu_T = 0$ and $\sigma_R = \sigma_T = 1$.

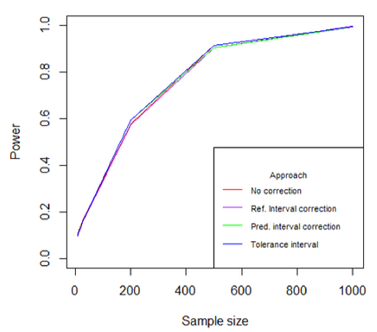

Figure. The power curves for all four approaches to perform Mielke's test approach 1 for large enough sample size.

Conclusion and Discussions

The procedures developed in this project yield the critical values that can control the Type I error rate correctly. In addition, the asymptotic power goes to 1 for all procedures. We also understand how the critical values behave as the sample size increases, and we have established that for each test, all the alternative approaches have the same asymptotic critical value, regardless of the parameter configuration.

In this project we have raised our reservations with the Mielke et al. (2018) testing procedure for comparability of test and reference products, especially the fact that the necessity to correct the coverage has been overlooked. We have also developed a new test, TOST, that controls the amount of non-overlap in both above the upper reference quantile and below the lower reference quantile, thus ensuring both efficacy and safety. Results of TOST will NOT be presented in this poster due to limited space. For those who are interested, please refer to our full manuscript. The conclusion and discussions will be given in the last section.

As of now, the estimated critical values in these approaches still depend on unknown parameters. Thus, we have not yet arrived at the full solution. Future work must include finding a procedure that can still control the Type I error rate correctly and produce a powerful test that does not assume knowledge of the parameter values.

Moreover, the simulations done in this research are based on a simulation size of 10,000. In the future, it is also desirable to increase this further.

References

- Tsong, Y., Wang, T.W., Hu, S., *"Statistical considerations in setting quality specification limits using quality data"*. In Statistical Methods in Early Clinical and Biomarker Development. Editor: Ray Liu. Springer, 2018.
- Casella, G., and Berger, R., Statistical Inference. 2nd Ed., Belmont, CA: Duxbury, 2009.
- Krishnamoorthy, K., Mathew, T., Statistical Tolerance Regions: Theory, Applications, and Computation. Hoboken, NJ: Wiley, 2009.
- Mielke, J., Innerbichler, F., Schiestl, M., Ballarini, N. M., and Jones, B., *The Assessment of Quality Attributes for Biosimilars: a Statistical Perspective on Current Practice and a Proposal*. The AAPS Journal, 2019, 21:7.
- Ren, Y., Dai, T., Shen, M., and Tsong, Y., Statistical Assessments of Product Quality Comparability. ORISE Project with FDA OB, 2019.