ISSN: 2469-5831

Yu-Ting Weng et al. Int J Clin Biostat Biom 2018, 4:016
DOI: 10.23937/2469-5831/1510016
Volume 4 | Issue 1
Open Access

International Journal of
Clinical Biostatistics and Biometrics

RESEARCH ARTICLE

Improved Wald Test for Equivalence Assessment of Analytical Biosimilarity

Yu-Ting Weng, Yi Tsong*, Meiyu Shen and Chao Wang

Division of Biometrics VI, Office of Biostatistics, CDER, FDA, USA

***Corresponding author:** *Yi Tsong, Director, Division of Biometrics VI, Office of Biostatistics, CDER, FDA, 10903 New Hampshire, Silver Spring, MD 20903, USA, Tel: 301-796-1013, E-mail: yi.tsong@fda.hhs.gov*

Abstract

The equivalence test in analytical similarity assessment uses a margin of 1.5 times of the standard deviation of a reference product. In the current practice, the standard deviation, estimated from study data, is considered as a fixed constantin the margin [1]. The impact of such a practice leads to the inflation of type I error rate and the reduction of power as previous studies showed [2]. In order to accommodate the fact that the margin is a parameter and improve the efficiency when the numbers of lots for both products are small. Chen, et al. [3] proposed to use Wald test with Constrained Maximum Likelihood Estimate (CMLE) of the standard error, resulting in the type I error rate is below the nominal value. In this paper, we further improve the Wald test with CMLE standard error by replacing the maximum likelihood estimate of reference standard deviation in the margin with the sample estimate. For small numbers of lots for both products, this estimate replacement leads to further improvement of type I error rate and power over the tests proposed in Chen, et al. [3]. In addition, to satisfy the criteria that the power is greater than 85% with the number of product lots being ten and equal product variability, we propose to use a margin of 1.7 times of the standard deviation of a reference product.

Keywords

Equivalence tests, Wald tests, Constrained maximum likelihood estimate

Introduction

Two one-sided hypothesis tests with a parameter margin that is a function of the variability of the reference product have been applied to equivalence assessments in several pharmaceutical areas [3-5]. The two one-sided hypotheses can be written as follows.

$$H_{0U}: \mu_T - \mu_R \geq f\sigma_R \text{ vs. } H_{aU}: \mu_T - \mu_R < f\sigma_R \quad (1)$$

$$H_{0L}: \mu_T - \mu_R \leq -f\sigma_R \text{ vs. } H_{aL}: \mu_T - \mu_R > -f\sigma_R$$

where μ_T and μ_R are the population means of the test product and the reference product, respectively, σ_R is the Standard Deviation (SD) of the reference product and f is the pre-specified constant.

Hypotheses in (1) has been proposed in various applications. Shen and Xu [5] proposed $f = 0.85$ in designs of method transfer studies for biotechnology products to compare means between the sending laboratory and the receiving laboratory. In their study, μ_T and μ_R represent the population mean of the measurements under normality distributions obtained at the receiving laboratory and the sending laboratory, respectively, and σ_R represents the population SD of the measurements obtained at the sending laboratory. For the evaluation of the analytical similarity between a test product and the reference product, Tsong, et al. [4] proposed to assess the equivalence in means for a selected Critical Quality Attribute (CQA) by testing the hypotheses in (1) with $f = 1.5$. The current practice is just substituting the sample SD of the Reference Product (S_R) in the margin as if it is a known value although σ_R is unknown and needed to be estimated from the study data [1]. Hence, under the normality assumption, the current analysis using *t* statistics results in inflating type I error rate and reducing power as pointed out by Dong, et al. [2] and Burdick, et al. [6].

To reduce the deficiency, one alternative approach is considering σ_R as a parameter and then applying the Wald-type statistic to hypotheses in (1). Chen, et al. [3] pointed out that the Wald test led to type I error rate seriously lower than the nominal significance level and power

Citation: Yu-Ting Weng, Tsong Y, Shen M, Wang C (2018) Improved Wald Test for Equivalence Assessment of Analytical Biosimilarity. Int J Clin Biostat Biom 4:016. doi.org/10.23937/2469-5831/1510016

Received: September 29, 2017: **Accepted:** March 13, 2018: **Published:** March 15, 2018

smaller than the target power value when the numbers of product lots are small. Chen, et al. [3] proposed a modified Wald test by using the Constrained Maximum Likelihood Estimate (CMLE)-method. With the Constrained Maximum Likelihood Estimate, the standard error was estimated using Maximum Likelihood Estimator (MLE) restricted to the null hypotheses in (1). Based on simulations, this CMLE-method led to slightly increase type I error rate but still less than the nominal significance value when the numbers of product lots are smaller than 20. Later, Burdick, et al. [6] and Dong, et al. [2] proposed to use Generalized Pivotal Quantity (GPQ) [7] to better control the type I error rate inflation and improve the power performance. Simulations showed that the type I error rate for the GPQ method is below the nominal significance value except for some small-lot-number scenarios in which the simulated type I error rate can be inflated to around 5.3%. In this paper, we further improve the Wald test with CMLE standard error [3] by replacing the MLE of σ_R in the equivalence margin with the sample estimate to further increase type I error rate while still below the nominal significance level and increase power when the numbers of product lots are small.

This paper is structured as follows. In Section 2, we consider three methods to construct Wald tests with CMLEs for standard error estimation. In Section 3, we describe the simulation plan and evaluate type I error rate and power performance of these three Wald tests. We provide an example to apply our proposed method to a simulated dataset and compare the proposed method with the current practice in Section 4. We present the discussion and conclusions in Section 5. For the simplicity of discussion, we consider only normally distributed measurements.

Methods

We first derive the proposed improved Wald test statistic and the other two estimators and then propose improved Wald test statistic with size adjustment to mitigate the imbalance between the numbers of the reference product lots and the test product lots.

Proposed improved Wald test statistic

To achieve at least 85% power when the mean difference is 1/8 of the reference standard deviation and the numbers of the reference product lots and the test product lots equal to ten, with equal variability, the current practice led f to be 1.5 [1]. More details are described in Tsong, et al. [4].

In this paper, we propose $M_{L_1} = \hat{\mu}_T - \hat{\mu}_R + f\tilde{\sigma}_R$ and $M_{U_1} = \hat{\mu}_T - \hat{\mu}_R - f\tilde{\sigma}_R$ for estimating the parameters $\mu_T - \mu_R + f\sigma_R$ and $\mu_T - \mu_R - f\sigma_R$, respectively. $\hat{\mu}_T$, $\hat{\mu}_R$ and $\tilde{\sigma}_R$ denote the MLE for the mean of the test product, the MLE for the mean of the reference product and the sample estimator for the SD of the reference product, respectively. When X_T is distributed from Normal (μ_T, σ_T^2) and X_R is distributed from Normal (μ_R, σ_R^2), $\hat{\mu}_T = \frac{\sum_{i=1}^{n_T} X_{T_i}}{n_T}$, $\hat{\mu}_R = \frac{\sum_{i=1}^{n_R} X_{R_i}}{n_R}$, $\tilde{\sigma}_R = \sqrt{\frac{\sum_{i=1}^{n_R}\left(X_{R_i} - \hat{\mu}_R\right)^2}{n_R - 1}}$, where n_T and n_R denote the numbers of test lots and reference lots, respectively. σ_T^2 is the variance of the test product. Then, we compare the proposed estimators to the other two sets of estimators: The unbiased version of the proposed estimators and the CMLEs proposed by Chen, et al. [3]. To correct the bias of the proposed estimators, we define that $M_{L_2} = \hat{\mu}_T - \hat{\mu}_R + fk\tilde{\sigma}_R$ and $M_{U_2} = \hat{\mu}_T - \hat{\mu}_R - fk\tilde{\sigma}_R$. As described in Ahn and Fessler [8], k is the bias correction factor of the SD of the reference product such that

$$k = \sqrt{\frac{n_R - 1}{2}} e^{ln\Gamma\left(\frac{n_R - 1}{2}\right) - ln\Gamma\left(\frac{n_R}{2}\right)}$$

where $\Gamma(y)$ is the gamma function defined as $\Gamma(y) = \int_0^\infty x^{y-1} e^{-x} dx$, y is a positive number. For the CMLE-method proposed by Chen, et al. [3], the estimators are $M_{L_3} = \hat{\mu}_T - \hat{\mu}_R + f\breve{\sigma}_R$ and $M_{U_3} = \hat{\mu}_T - \hat{\mu}_R - f\breve{\sigma}_R$. $\breve{\sigma}_R$ denotes the MLE of the SD of the reference product, which is defined as $\breve{\sigma}_R = \sqrt{\frac{\sum_{i=1}^{n_R}\left(X_{R_i} - \hat{\mu}_R\right)^2}{n_R}}$. Under either of the null hypotheses in (1), none of the three estimators has an exact distribution when we consider the unknown σ_R as a parameter. Therefore, we resort to asymptotic standard normal approximation for Wald test statistics.

The method of obtaining CMLEs for the variances σ_T^2 and σ_R^2 is detailed in Chen, et al. [3]. We briefly describe the approach of calculating CMLEs. Considering the log-likelihood function under either of the constraint in the two null hypotheses in (1), we first derive the MLEs for μ_T, μ_R and σ_T^2 under the constraint $\mu_T - \mu_R = -f\sigma_R$ in H_{0L}. More specifically, the log-likelihood function is given by

$$logL \propto -\frac{n_T}{2}\log\sigma_T^2 - \frac{n_R}{2}\log\sigma_R^2 - \sum_{i=1}^{n_T}\frac{\left(X_{T_i} - \mu_T\right)^2}{2\sigma_T^2} - \sum_{i=1}^{n_R}\frac{\left(X_{R_i} - \mu_R\right)^2}{2\sigma_R^2}$$

$$= -\frac{n_T}{2}\log\sigma_T^2 - \frac{n_R}{2}\log\left(\frac{\mu_R - \mu_T}{f}\right)^2 - \sum_{i=1}^{n_T}\frac{\left(X_{T_i} - \mu_T\right)^2}{2\sigma_T^2} - \sum_{i=1}^{n_R}\frac{\left(X_{R_i} - \mu_R\right)^2}{2\left(\frac{\mu_R - \mu_T}{f}\right)^2} \quad (2)$$

Then, the CMLE for σ_T^2 is given by $\breve{\sigma}_{T_L}^2 = \frac{\sum_{i=1}^{n_T}\left(X_{T_i} - \breve{\mu}_{T_L}\right)^2}{n_T}$, and the CMLE $\left(\breve{\mu}_{T_L}, \breve{\mu}_{R_L}\right)$ of $\left(\mu_T, \mu_R\right)$ is estimated numerically by Gibbs sampling. In the simulation study described in next Section, the estimators are obtained using R function and the code derived from (2) is attached in the Appendix 1. By substituting $\left(\breve{\mu}_{T_L}, \breve{\mu}_{R_L}\right)$ for $\left(\mu_T, \mu_R\right)$, the CMLE for σ_R^2 is given by $\breve{\sigma}_{R_L}^2 = \left(\frac{\breve{\mu}_{T_L} - \breve{\mu}_{R_L}}{f}\right)^2$. The CMLE $\left(\breve{\mu}_{T_U}, \breve{\mu}_{R_U}, \breve{\sigma}_{T_U}^2, \breve{\sigma}_{R_U}^2\right)$ of $\left(\mu_T, \mu_R, \sigma_T^2, \sigma_R^2\right)$ under the constraint in H_{0U} can be derived in the similar way.

Three Wald tests for testing the null hypothesis $H_{0L}: \mu_T - \mu_R \le -f\sigma_R$ is constructed based on three different estimators $\mu_T - \mu_R - f\sigma_R$. As described in Ahn and Fessler [8], the standard error of $\tilde{\sigma}_R$ can be estimated by $\sigma_R\sqrt{\frac{V_{n_R}}{n_R-1}}$, where $V_{n_R} = 2\left(\frac{n_R-1}{2} - \frac{\Gamma^2\left(\frac{n_R}{2}\right)}{\Gamma^2\left(\frac{n_R-1}{2}\right)}\right)$ is the variance of the chi distribution with (n_R-1) degrees of freedom. Thus, in our proposed Modified Wald test with CMLE (MWCMLE), the standard error S_{L_1} of M_{L_1} is given by $\sqrt{\frac{\sigma_T^2}{n_T} + \left(\frac{1}{n_R} + \frac{f^2V_{n_R}}{n_R-1}\right)\sigma_R^2}$. Then S_{L_1} can be estimated by $\sqrt{\frac{\breve{\sigma}_{T_L}^2}{n_T} + \left(\frac{1}{n_R} + \frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_L}^2}$, where $\left(\breve{\sigma}_{T_L}^2, \breve{\sigma}_{R_L}^2\right)$ is CMLE of variance $\left(\sigma_T^2, \sigma_R^2\right)$.

The quantity $\frac{M_{L_1}}{S_{L_1}}$ can be used for testing H_{0L}. Plugging MLE $(\hat{\mu}_T)$ for the test product mean, MLE $(\hat{\mu}_R)$ for the reference product mean, sample estimator $(\tilde{\sigma}_R)$ for the SD of the reference product and CMLE $\left(\breve{\sigma}_{T_L}^2, \breve{\sigma}_{R_L}^2\right)$, we have the following test statistic.

MWCMLE: $$W_{1L} = \frac{\hat{\mu}_T - \hat{\mu}_R + f\tilde{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_L}^2}{n_T} + \left(\frac{1}{n_R} + \frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_L}^2}}.$$

Following a similar derivation, the test statistic for the other two estimators, the unbiased version of the MWCMLE and the CMLE-method, can be derived as follows.

Unbiased Modified Wald test with CMLE (UMWCMLE): $$W_{2L} = \frac{\hat{\mu}_T - \hat{\mu}_R + fk\tilde{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_L}^2}{n_T} + \left(\frac{1}{n_R} + f^2\left(k^2-1\right)\right)\breve{\sigma}_{R_L}^2}}.$$

Wald test with CMLE (CMLE-method in Chen, et al. 2017) [3]: $$W_{3L} = \frac{\hat{\mu}_T - \hat{\mu}_R + f\hat{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_L}^2}{n_T} + \left(1 + f^2V_{n_R}\right)\frac{\breve{\sigma}_{R_L}^2}{n_R}}}.$$

Similarly, we derive three Wald tests for testing $H_{0U}: \mu_T - \mu_R \ge f\sigma_R$ based on the corresponding quantity M_{U_1}, M_{U_2}, or M_{U_3} as follows.

MWCMLE: $$W_{1U} = \frac{\hat{\mu}_T - \hat{\mu}_R - f\tilde{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_U}^2}{n_T} + \left(\frac{1}{n_R} + \frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_U}^2}}.$$

UMWCMLE: $$W_{2U} = \frac{\hat{\mu}_T - \hat{\mu}_R - fk\tilde{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_U}^2}{n_T} + \left(\frac{1}{n_R} + f^2\left(k^2-1\right)\right)\breve{\sigma}_{R_U}^2}}$$

CMLE-method: $$W_{3U} = \frac{\hat{\mu}_T - \hat{\mu}_R - f\hat{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_U}^2}{n_T} + \left(1 + f^2V_{n_R}\right)\frac{\breve{\sigma}_{R_U}^2}{n_R}}}$$

Under the null hypothesis, each test follows an asymptotically standard normal distribution. The null hypotheses in (1) is rejected if $W_{tL} > Z_{1-\alpha}$ and $W_{tU} < -Z_{1-\alpha}$ for t = 1, 2, 3 at significance level α, where Z_p is the 100p^{th} percentile of the standard normal distribution. With this criterion, we evaluated type I error rate and power for each test in Section 3.

Proposed improved Wald test statistic with size adjustment

Because the analytical similarity study is an un-blinded study, the number of reference lots can be much larger than the number of test lots. However, we do not want the information from the reference product to dominate the equivalence testing. Thus, Dong, et al. [9] proposed to compute the following confidence interval with the number of the adjusted reference lots and adjusted degrees of freedom when the ratio of the number of reference lots to the number of test lots is great than 1.5.

$$\Delta\bar{X} \pm t_{1-\alpha,df^*} \times \sqrt{\frac{S_R^2}{n_R^*} + \frac{S_T^2}{n_T}}$$

Here $\Delta\bar{X}$, S_R^2 and S_T^2 denote sample mean difference between the test and the reference products, sample variance of the reference product, and sample variance of the test product, respectively. The number of adjusted reference lots, n_R^*, is equal to min $(1.5 \times n_T, n_R)$. n_T and n_R denote the number of test lots and the number of reference lots, respectively. The $t_{1-\alpha,df^*}$ is $(1-\alpha)$ quantile of the *t*-distribution with degrees of freedom df^*.df^* is approximated by the Satterthwaite approximation as follows.

$$df^* = \frac{\left(\frac{S_T{}^2}{n_T} + \frac{S_R{}^2}{n_R^*}\right)^2}{\frac{\left(\frac{S_T{}^2}{n_T}\right)^2}{n_T-1} + \frac{\left(\frac{S_R{}^2}{n_R^*}\right)^2}{n_R-1}}$$

In Dong, et al. [9], the number of reference lots in df^* is only adjusted for the weight of the reference variance estimator $S_R{}^2$ but not for the variance itself. Following the same logic, we compute the Adjusted MWCMLE for imbalanced sample (AMWCMLE).

$$W_{1L'} = \frac{\hat{\mu}_T - \hat{\mu}_R + f\tilde{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_L}^2}{n_T} + \left(\frac{1}{n_R^*} + \frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_L}^2}} \text{ and } W_{1U'} = \frac{\hat{\mu}_T - \hat{\mu}_R - f\tilde{\sigma}_R}{\sqrt{\frac{\breve{\sigma}_{T_U}^2}{n_T} + \left(\frac{1}{n_R^*} + \frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_U}^2}}$$

With the adjustment, AMWCMLE can be severely smaller than the corresponding MWCMLE when the ratio of the numbers of the lots for both products is extremely large. Thus, the adjusted type I error rate and power can be severely smaller than the unadjusted type I error rate and power. We evaluate the adjusted type I error rate and power in Section 3 as well.

Simulation

We first describe five simulation scenarios and corresponding simulation setups and then show simulation results of type I error rate and power performance of three Wald tests: MWCMLE, UMWCMLE and CMLE-method.

Four simulation scenarios with *f* = 1.5

We consider four scenarios to present the performance of the proposed MWCMLE in terms of type I error control and power improvement. Chen, et al. [3] demonstrated through simulation study that the Wald type tests were monotone tests, so that power increased with $\mu_T - \mu_R + f\sigma_R$ for testing H_{0L} and with $\mu_T - \mu_R - f\sigma_R$ for testing H_{0U} . With the monotone property, the type I error rate simulated at the boundary of H_{0L} and H_{0U} is the maximum type I error rates of the two one-sided tests. The four scenarios are described below.

Scenario 1: Show power function for the MWCMLE is monotone.

Scenario 2: Compare type I error rate and power of the MWCMLE to type I error rates and powers of the CMLE-method and the UMWCMLE with equal and unequal numbers of test and reference lots and different variance ratios of test product to reference product.

Scenario 3: Compare type I error rate and power of the MWCMLE to type I error rate and power of the AMWCMLE for unequal samples with different variance ratios.

Scenario 4: Generate type I error rate and power of the MWCMLE for small number of equal product lots with different variance ratios.

Margin reselection

In Scenario 4, the simulated power for the MWCMLE is less than 85% with the number of product lots being ten and equal product variability. To satisfy the criteria that the power is greater than 85% with the number of product lots being ten and equal product variability, we increase the margin *f* from 1.5 to 1.7. Accordingly, the hypotheses are changed to

$$H_{0U}: \mu_T - \mu_R \ge 1.7\sigma_R \text{ vs. } H_{aU}: \mu_T - \mu_R < 1.7\sigma_R \tag{3}$$

$$H_{0L}: \mu_T - \mu_R \le -1.7\sigma_R \text{ vs. } H_{aL}: \mu_T - \mu_R > -1.7\sigma_R$$

Then, we repeat the process in Scenario 4 and denoted this scenario as Scenario 5.

Simulation setups

We conduct extensive simulation studies to evaluate the MWCMLE by using type I error rate and power performance. The simulation setups for each scenario are described as follows.

Scenario 1: Let $\sigma_T^2 = 1, \sigma_R = 1, \mu_R = 0, \mu_T = \mu_R + \times\sigma_R$ and let λ from -2.0 to 2.0 by 0.1. Various allocations of the number of product lots are considered. When the numbers of product lots are equal $(n_T = n_R)$, the following number of test lots is used: $n_T = (10, 100)$. When the numbers of product lots are not equal, the following numbers of product lots are used: $(n_T, n_R) =$ (10, 6) and (6, 10). Then, X_T and X_R are generated independently from Normal (μ_T, σ_T^2) and from Normal (μ_R, σ_R^2), respectively.

Scenario 2: To compare with the simulation results of the CMLE-method in Chen, et al. [3], the same simulation setups as described in their study are used. The following configurations are used: $\sigma_T^2 = (0.25, 0.5, 1, 2, 4), \sigma_R = 1, \mu_R = 0, and\, \mu_T = \mu_R + ES \times \sigma_R$. The Effect Size (ES) is set at 1.5 and 0.125 for type I error rate and power, respectively. Various allocations of the number of product lots are considered. When the numbers of product lots are equal $(n_T = n_R)$, the following number of test lots is used: $n_T = (6, 10, 15, 25, 100, 1000)$ where n_T are chosen to represent from practical small sample sizes used in biosimilar and large sample size to show the convergence to normal approximation test. When the numbers of product lots are not equal, the following numbers of product lots are used: (n_T, n_R) = (10, 6), (10, 25), (10, 100), (6, 10), (25, 10), and (100, 10). X_T and X_R are generated independently from Normal (μ_T, σ_T^2) and from Normal (μ_R, σ_R^2), respectively.

Scenario 3: Same simulation setups as Scenario 2 are used.

Scenario 4 (*f* = 1.5) and Scenario 5 (*f* = 1.7): The following configurations are used: σ_T^2 from 0.5 to 2 by 0.25. $\sigma_R = 1, \mu_R = 0 \text{ and } \mu_T = \mu_R + ES \times \sigma_R$. *ES* is set at 1.5 or 1.7 and 0.125 for type I error rate and power, respectively. We only consider the case when the numbers of product lots are equal from 10 to 15 by 1. Then, X_T and X_R are generated independently from Normal (μ_T, σ_T^2) and from Normal (μ_R, σ_R^2) , respectively.

Throughout the simulations, we fix the test significance level at $\alpha = 0.05$ for each one-sided hypothesis test. The results are based on one million independent replicates for each simulation setup so that the standard error of simulation can be around $\sqrt{\frac{0.95 * 0.05}{1,000,000}} = 0.0002$.

Simulation results

Figure 1 shows the plots of the simulated power values against the effect size values for testing the null hypotheses in (1). First, as we can see, when the numbers of lots increase, the simulated power increases. Secondly, when the effect size is from -2.0 to zero or from 2.0 to zero, the simulated power increases monotonically. Thus, MWCMLE has the monotone property, and the Wald tests can be performed at the boundary of the null hypotheses H_{0L} and H_{0U}.

Table 1 shows the simulated type I error rates for three methods at different combination of the variance

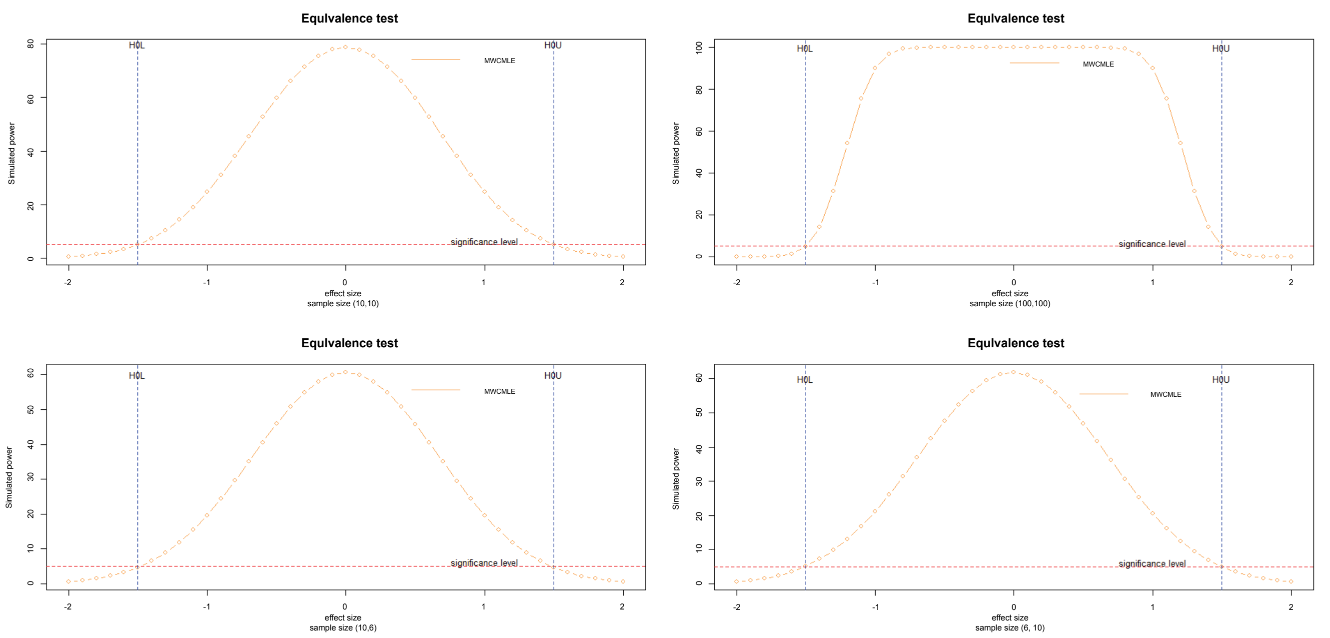

Figure 1: Power function for the MWCMLE when the effect size is from -2.0 to 2.0.

Table 1: Simulated type I error rates (%) for three Wald Tests with different variance ratios when the numbers of product lots are equal if *f* = 1.5.

n_T	n_R	σ_T^2/σ_R^2	Simulated type I error rate		
			CMLE	MWCMLE	UMWCMLE
6	6	0.25	3.0726	4.8217	6.0644*
		0.5	3.2304	4.9344	6.1406*
		1.0	3.3346	5.0310*	6.1845*
		2.0	3.0959	4.7161	5.8023*
		4.0	2.0638	3.3248	4.2259
10	10	0.25	3.3869	4.7339	5.6029*
		0.5	3.5148	4.8141	5.6374*
		1.0	3.6749	4.9202	5.6987*
		2.0	3.8109	4.9815	5.692*
		4.0	3.3862	4.4441	5.0727*
15	15	0.25	3.6021	4.7224	5.3972*
		0.5	3.6933	4.7962	5.4385*
		1.0	3.8399	4.8715	5.4702*
		2.0	4.0088	4.9193	5.4665*
		4.0	4.0739	4.8862	5.3507*
25	25	0.25	3.8906	4.7692	5.2811*
		0.5	3.9598	4.7986	5.3015*
		1.0	4.0532	4.8561	5.3081*
		2.0	4.1977	4.9019	5.3156*
		4.0	4.3411	4.9602	5.2896*
100	100	0.25	4.3875	4.8562	5.1081*
		0.5	4.4226	4.8631	5.106*
		1.0	4.4596	4.8796	5.0952*
		2.0	4.5135	4.8837	5.0861*
		4.0	4.6005	4.8989	5.0584*
1000	1000	0.25	4.7958	4.9526	5.0303*
		0.5	4.8103	4.9534	5.0284*
		1.0	4.8304	4.963	5.0311*
		2.0	4.8586	4.9807	5.0363*
		4.0	4.8776	4.9797	5.0328*

*: Simulated type I error rate is inflated.

ratio of test product to reference product and equal numbers of product lots. The simulated type I error rate of MWCMLE is below the nominal significance level except for some scenarios when the numbers of product lots are six. Compared to the results for the MWCMLE, the simulated type I error rate for the CMLE-method is more conservative and the simulated type I error rate for the UMWCMLE is more liberal and inflated. Common patterns for all three methods are observed. First, when the numbers of lots increase, the simulated type I error rate converges to the significance level $(\alpha = 0.05)$. Secondly, when the numbers of product lots are six and the variance ratio is large $\left(\sigma_T^2/\sigma_R^2 = 4\right)$, the simulated type I error rate is less than 4.3%. When comparing three methods, the simulated type I error rates from high to low are the UMWCMLE, the MWCMLE, and the CMLE-method, respectively.

Table 2 shows the simulated type I error rates for three methods at different combination of the variance ratio and unequal numbers of product lots. The simulated type I error rate of MWCMLE is below the nominal significance level, except that the numbers of product lots are as follows: (n_T, n_R) = (10, 25), (10, 100), or (6, 10). Compared to the results for the MWCMLE, the simulated type I error rate for the CMLE-method is more conservative and the simulated type I error rate for the UMWCMLE is more liberal and inflated. The CMLE-method and the MWCMLE share the following common patterns. First, when the number of reference lots increases, the simulated type I error rate increases at each level of the variance ratio. When the number of test lots increases, the simulated type I error rate decreases at each level of the variance ratio, except the large variance ratio $\left(\sigma_T^2/\sigma_R^2 = 4\right)$. Secondly, the simulated type I error rate is less than 4.3% when the vari-

Table 2: Simulated type I error rates (%) for three Wald Tests with different variance ratios when the numbers of product lots are unequal if *f* = 1.5.

n_T	n_R	σ_T^2/σ_R^2	Simulated type I error rate CMLE	MWCMLE	UMWCMLE
10	6	0.25	2.97	4.7258	5.9699*
		0.5	3.0607	4.7881	5.9839*
		1.0	3.1977	4.8313	5.9892*
		2.0	3.2077	4.7841	5.8667*
		4.0	2.6001	4.0661	5.0568*
10	25	0.25	4.0707	4.9179	5.4112*
		0.5	4.2147	5.0159*	5.4703*
		1.0	4.3829	5.1207*	5.5264*
		2.0	4.5277	5.1431*	5.4942*
		4.0	4.2419	4.7668	5.0521*
10	100	0.25	4.7472	5.1223*	5.3247*
		0.5	4.8388	5.1581*	5.3271*
		1.0	4.8699	5.1273*	5.2587*
		2.0	4.8994	5.0872*	5.1927*
		4.0	4.6595	4.8087	4.8859*
6	10	0.25	3.5843	4.9327	5.7657*
		0.5	3.7697	5.0818*	5.9048*
		1.0	3.9773	5.2181*	5.9956*
		2.0	3.8526	5.0389*	5.7476*
		4.0	2.6982	3.6261	4.2085
25	10	0.25	3.3113	4.6766	5.5578*
		0.5	3.3711	4.7027	5.5704*
		1.0	3.447	4.7439	5.568*
		2.0	3.5432	4.7785	5.5739*
		4.0	3.6815	4.8036	5.4995*
100	10	0.25	3.1975	4.5566	5.4537*
		0.5	3.232	4.5989	5.484*
		1.0	3.2616	4.626	5.4846*
		2.0	3.3017	4.6427	5.5023*
		4.0	3.3697	4.6508	5.4714*

*: Simulated type I error rate is inflated.

ance ratio is large $\left(\sigma_T^{\ 2}/\sigma_R^{\ 2}=4\right)$ and the numbers of product lots are as follows: (n_T, n_R) = (10, 6) or (6, 10). When comparing three methods, the simulated type I error rates from high to low are the UMWCMLE, the MWCMLE, and the CMLE-method, respectively.

Table 3 shows the simulated power for three methods at different combination of the variance ratio and equal numbers of product lots. For all three methods, when the numbers of lots increase, the simulated power increases. When comparing three methods, the simulated powers from high to low are the UMWCMLE, the MWCMLE, and the CMLE-method, respectively.

Table 4 shows the simulated power for three methods at different combination of the variance ratio and unequal numbers of product lots. For all three methods, when the number of reference lots increases or the number of test lots increases, the simulated power increases. When comparing three methods, the simulated powers from high to low are the UMWCMLE, the MWCMLE, and the CMLE-method, respectively.

Table 3: Simulated power (%) for three Wald Tests with different variance ratios when the numbers of product lots are equal if f = 1.5 and $\mu_T - \mu_R = \sigma_R/8$.

n_T	n_R	σ_T^2/σ_R^2	Simulated power CMLE	MWCMLE	UMWCMLE
6	6	0.25	68.851	75.1763	78.2522
		0.5	56.9287	64.8567	68.816
		1.0	38.9725	47.837	52.681
		2.0	19.7563	27.1498	31.4647
		4.0	6.9084	10.799	13.4001
10	10	0.25	91.8434	93.5407	94.3174
		0.5	86.0486	88.771	89.9659
		1.0	73.0699	77.5285	79.557
		2.0	49.5325	55.5501	58.6029
		4.0	21.9434	27.0984	29.9405
15	15	0.25	98.5048	98.8007	98.9615
		0.5	96.7669	97.3917	97.6942
		1.0	91.4305	92.8779	93.5281
		2.0	76.5988	79.6186	81.0726
		4.0	47.3085	51.6418	53.8816
25	25	0.25	99.9478	99.9584	99.9661
		0.5	99.8275	99.8673	99.8751
		1.0	99.176	99.3103	99.3827
		2.0	95.5465	96.1386	96.4221
		4.0	80.8819	82.6019	83.4914
100	100	0.25	100	100	100
		0.5	100	100	100
		1.0	100	100	100
		2.0	100	100	100
		4.0	99.9954	99.9965	99.9964
1000	1000	0.25	100	100	100
		0.5	100	100	100
		1.0	100	100	100
		2.0	100	100	100
		4.0	100	100	100

Table 5 compare the simulated type I error rate and power of the MWCMLE to the simulated type I error rate and power of the AMWCMLE with different variance ratios and unequal numbers of product lots. As we expected, the simulated type I error rate and power of the MWCMLE decrease after adjusting the degree of freedom at each level of the variance ratio. In addition, when the ratio of the numbers of the lots for both products is 10, the adjusted type I error rate and power can be severely smaller than the unadjusted type I error rate and power.

Supplementary Table A.1 and Supplementary Table A.2 in the Appendix 3 show that the simulated type I error rate and power of the MWCMLE for small equal numbers of product lots with different variance ratios and *f* is 1.5. The MWCMLE can control the simulated type I error rate well in this specified range of the number of lots and the variance ratios. In addition, the simulated power increases when the numbers of lots product increase; the simulated power decreases when the variance ratio increases.

Table 4: Simulated power (%) for three Wald Tests with different variance ratios when the numbers of product lots are unequal if *f* = 1.5 and $\mu_T - \mu_R = {}^{\sigma_R}/_{8}$.

n_T	n_R	σ_T^2 / σ_R^2	Simulated power		
			CMLE	MWCMLE	UMWCMLE
10	6	0.25	74.0438	79.4977	82.0667
		0.5	66.4983	73.0883	76.2573
		1.0	53.0873	61.1323	65.1864
		2.0	33.7136	42.3201	46.9895
		4.0	14.5835	20.9865	24.9171
10	25	0.25	99.703	99.7686	99.7886
		0.5	98.3803	98.6813	98.7791
		1.0	91.5326	92.6733	93.1526
		2.0	68.9696	71.6018	72.8204
		4.0	32.2718	35.2019	36.5143
10	100	0.25	99.9999	100	100
		0.5	99.9492	99.9569	99.9575
		1.0	97.8069	97.9137	98
		2.0	80.1939	80.8566	81.118
		4.0	39.6682	40.469	40.9484
6	10	0.25	87.8472	90.2938	91.4495
		0.5	76.5278	80.7979	82.769
		1.0	55.0441	61.2033	64.3315
		2.0	28.4305	34.1712	37.3351
		4.0	9.8076	12.7372	14.7119
25	10	0.25	94.7127	95.8673	96.375
		0.5	92.8589	94.3548	95.0161
		1.0	88.6234	90.8036	91.8156
		2.0	78.9254	82.4727	84.1595
		4.0	59.443	64.7772	67.3846
100	10	0.25	95.9611	96.8569	97.2619
		0.5	95.5805	96.5291	96.9897
		1.0	94.7521	95.8535	96.3767
		2.0	92.8916	94.3382	95.0457
		4.0	88.7293	90.8661	91.8860

Similarly, Supplementary Table A.3 and Supplementary Table A.4 in the Appendix 3 show that the simulated type I error rate and power of the MWCMLE for small equal numbers of product lots with different variance ratios and *f* is 1.7. The MWCMLE can control the simulated type I error rate well in this specified range of the number of lots and the variance ratios. In addition, the simulated power increases when the numbers of product lots increase, the simulated power decreases when the variance ratio increases. Compared to the simulated power in Supplementary Table A.2 when *f* is 1.5, the simulated power in Supplementary Table A.4 when *f* is 1.7 is larger for each combination of the number of lots and the variance ratios.

Application

To illustrate the application of the proposed MWCMLE, we provide an example in this section. We use the same simulated CQA data in Dong, et al. [2] to present the results for the current practice [1] and the proposed MWCMLE. The numbers of product lots are ten. Each individual observation of the test product is 94, 109, 103, 97, 102, 101, 99, 97, 97, and 103; each individual observation of the reference product is 96, 104, 102, 102, 101, 99, 99, 92, 107, and 98. Then, the sample means for the test product $(\bar{X}_T)$ and the reference product $\bar{X}_R$ are 100.2 and 100.0, respectively. In addition, the sample variances for the test product (S_T^2) and the reference product (S_R^2) are 18.6 and 17.8, respectively. Furthermore, by using the proposed MWCMLE, the overall null hypothesis is rejected by either of the following two criteria: first, $W_{1L} > Z_{0.95}$ and $W_{1U} < -Z_{0.95}$, or second, the 90% confidence interval (L,U) falls within the equivalence margin $(-f\tilde{\sigma}_R, f\tilde{\sigma}_R)$. The 90% confidence interval (L,U) is derived by converting the above first criterion as follows.

Table 5: Simulated type I error rates (%) and power (%) for MWCMLE and AMWCMLE with different variance ratios when the numbers of product lots are unequal if f = 1.5 and $\mu_T - \mu_R = {}^{\sigma_R}/_{8}$ (for power only).

n_T	n_R	σ_T^2 / σ_R^2	Simulated type I error rate		Simulated power	
			MWCMLE	AMWCMLE	MWCMLE	AMWCMLE
10	6	0.25	4.7258	4.6639	79.4977	79.1071
		0.5	4.7881	4.673	73.0883	72.2606
		1.0	4.8313	4.6331	61.1323	59.5617
		2.0	4.7841	4.4482	42.3201	39.9395
		4.0	4.0661	3.6017	20.9865	18.7454
10	25	0.25	4.9179	3.2928	99.7686	99.6524
		0.5	5.0159*	3.5998	98.6813	98.2647
		1.0	5.1207*	3.9524	92.6733	91.5193
		2.0	5.1431*	4.3185	71.6018	69.714
		4.0	4.7668	4.2371	35.2019	33.5624
10	100	0.25	5.1223*	0.5951	100	99.9973
		0.5	5.1581*	1.1428	99.9569	99.7783
		1.0	5.1273*	1.9812	97.9137	95.9279
		2.0	5.0872*	2.9319	80.8566	75.9145
		4.0	4.8087	3.4929	40.469	36.2409
6	10	0.25	4.9327	4.5894	90.2938	89.9811
		0.5	5.0818*	4.7678	80.7979	80.3026
		1.0	5.2181*	4.9573	61.2033	60.6936
		2.0	5.0389*	4.8305	34.1712	33.7187
		4.0	3.6261	3.5091	12.7372	12.6295
25	10	0.25	4.6766	4.4341	95.8673	95.4431
		0.5	4.7027	4.2605	94.3548	93.303
		1.0	4.7439	3.9579	90.8036	88.0992
		2.0	4.7785	3.5078	82.4727	75.519
		4.0	4.8036	2.9553	64.7772	50.2691
100	10	0.25	4.5566	4.0399	96.8569	96.0507
		0.5	4.5989	3.6254	96.5291	94.7502
		1.0	4.626	2.9201	95.8535	91.5534
		2.0	4.6427	1.9884	94.3382	83.3233
		4.0	4.6508	1.0428	90.8661	62.7931

*: Simulated type I error rate is inflated.

$$W_{1L} = \frac{\hat{\mu}_T - \hat{\mu}_R + f\tilde{\sigma}_R}{\sqrt{\frac{\tilde{\sigma}_{T_L}^2}{n_T} + \left(\frac{1}{n_R^*} + \frac{f^2 V_{n_R}}{n_R - 1}\right)\tilde{\sigma}_{R_L}^2}} > Z_{0.95} \leftrightarrow \hat{\mu}_T - \hat{\mu}_R - Z_{0.95}\sqrt{\frac{\tilde{\sigma}_{T_L}^2}{n_T} + \left(\frac{1}{n_R^*} + \frac{f^2 V_{n_R}}{n_R - 1}\right)\tilde{\sigma}_{R_L}^2} > -f\tilde{\sigma}_R$$

$$W_{1U} = \frac{\hat{\mu}_T - \hat{\mu}_R - f\tilde{\sigma}_R}{\sqrt{\frac{\tilde{\sigma}_{T_U}^2}{n_T} + \left(\frac{1}{n_R^*} + \frac{f^2 V_{n_R}}{n_R - 1}\right)\tilde{\sigma}_{R_U}^2}} < -Z_{0.95} \leftrightarrow \hat{\mu}_T - \hat{\mu}_R + Z_{0.95}\sqrt{\frac{\tilde{\sigma}_{T_U}^2}{n_T} + \left(\frac{1}{n_R^*} + \frac{f^2 V_{n_R}}{n_R - 1}\right)\tilde{\sigma}_{R_U}^2} < f\tilde{\sigma}_R$$

Thus, the 90% confidence interval (L,U) for the proposed MWCMLE derived from the two one-sided tests is as follows.

$$(L,U)=\left(\hat{\mu}_T-\hat{\mu}_R-Z_{0.95}\sqrt{\frac{\breve{\sigma}_{T_L}^2}{n_T}+\left(\frac{1}{n_R{}^*}+\frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_L}^2},\hat{\mu}_T-\hat{\mu}_R+Z_{0.95}\sqrt{\frac{\breve{\sigma}_{T_U}^2}{n_T}+\left(\frac{1}{n_R{}^*}+\frac{f^2V_{n_R}}{n_R-1}\right)\breve{\sigma}_{R_U}^2}\right)$$

The code is provided in the Appendix 2. Then, the following results are calculated. When *f* is 1.5 and the equivalence margin is (-6.32, 6.32) in the current practice, the 90% C.I. is (-3.11, 3.51), when *f* is 1.7 and the equivalence margin is (-7.17, 7.17) in the proposed MWCMLE, the 90% C.I. is (-3.83, 4.23). Thus, the data can pass the equivalence test by using both methods.

Discussion

We develop asymptotic tests using the Wald test statistic, for parallel-arm variance-adjusted equivalence trials with normal endpoints. Our results of the MWCMLE show that either the type I error rate controls closely below to the nominal level when the numbers of product lots are equal and greater than or equal to ten or the type I error rate can be inflated to around 5.2% when the numbers of product lots are unequal. In addition, the simulated type I error rate of the CMLE-method is more conservative than the one of the MWCMLE; the simulated type I error rate of the UMWCMLE is more liberal and inflated than the one of the MWCMLE.

In terms of power for three methods, our results show that the UMWCMLE outperforms the other two methods, especially when the numbers of product lots are small. However, as shown in our simulation, the simulated type I error rate of the UMWCMLE is inflated, indicating higher false positive rate. Thus, the UMWCMLE is not a proper estimator choice. In contrast, when the numbers of product lots are increasing, the simulated power of the MWCMLE improves and outperforms the CMLE-method. Thus, the MWCMLE can be a proper choice among these three methods.

Since the equivalence margin is unknown and estimated from the reference data, the simulated power of the MWCMLE is less than 85% with the number of product lots being ten and equal product variability when *f* is 1.5. To satisfy the criteria that the power is greater than 85% with the number of product lots being ten and equal product variability, *f* needs to be increased from 1.5 to 1.7 as shown in (3).

In conclusion, using the Wald test for equivalence testing of the hypothesis setting in (1) can be conservative when the numbers of product lots are small. However, using CMLE for the variance estimation can improve the performance of Wald Test as shown in Chen, et al. [3]. Our investigation of MWCMLE and UMWCMLE show that the proposed MWCMLE can control the type I error rate well and increase the power over CMLE-method while the type I error rate of the UMWCMLE can be over liberal and inflated. Further detailed comparisons with other methods will be reported in a different paper.

References

1. Statistical approaches to evaluate analytical similarity guidance for industry (2017). Rockville, USA.
2. Dong X, Bian Y, Tsong Y, Wang T (2017) Exact test-based approach for equivalence test with parameter margin. J Biopharm Stat 27: 317-330.
3. Chen YM, Weng YT, Dong X, Tsong Y (2017) Wald tests for variance-adjusted equivalence with normal endpoints. J Biopharm Stat 27: 308-316.
4. Tsong Y, Dong X, Shen M (2017) Development of statistical methods for analytical similarity assessment. J Biopharm Stat 27: 197-205.
5. Shen M, Xu L (2017) Design and statistical analysis of method transfer studies for biotechnology products. Bioanalysis 9: 595-600.
6. Burdick RK, Thomas N, Cheng A (2017) Statistical considerations in demonstrating CMC analytical similarity for a biosimilar product. Statistics in Biopharmaceutical Research 9: 249-257.
7. Weerahandi S (1993) Generalized confidence intervals. Journal of the American Statistical Association 88: 899-905.
8. Ahn S, Fessler JA (2003) Standard errors of mean, variance, and standard deviation estimators. EECS Department, The University of Michigan 1-2.
9. Dong X, Weng YT, Tsong Y (2017) Adjustment for unbalanced sample size for analytical biosimilar equivalence assessment. Journal of Biopharmaceutical Statistics 27: 220-232.

Appendix

A.1. R code for Type I Error (or power) Simulations

```
### calculate cmle for variance of test and reference formulations

CMLE_equivTest < - function(obsT,obsR,eta)

{nT = length (obsT)

nR = length (obsR)

smplmuT = mean (obsT)

smplmuR = mean (obsR)

sigmaT = sd (obsT)

sigmaR = sd (obsR)

muT = smplmuT

muR = smplmuR

oldPrmtr = c (muT,muR,sigmaT,sigmaR)

dif = 1

iterMax = 200

iter = 1

 while(dif > 1.0e - 4 & iter < iter Max)

 {muT = muR + eta*sigmaR

sigmaT = sqrt ( mean( (obsT - muT) ^ 2 ) )

muR = (nT*(smplmuT-eta*sigmaR)/sigmaT ^ 2 + nR*smplmuR/sigmaR ^ 2) / (nT/sigmaT ^ 2 + nR/sigmaR ^ 2)

poly = polynomial (c(-mean( (obsR-muR) ^ 2), 0, 1,-nT/nR*eta*(smplmuT-muR)/sigmaT ^ 2, nT/nR*eta ^ 2/sigmaT ^ 2 ))

pz = solve(poly)

 i = 4

root Found = FALSE

while(i > =1)

{if(Im(pz[i])==0 & Re(pz[i])>0)

{sigmaR = Re(pz[i])

root Found = TRUE break}

 i = i - 1}

if (!rootFound)

print ("Cannot find real root for sigmaR")

newPrmtr = c(muT, muR, sigmaT, sigmaR)

if (mean( abs(oldPrmtr - new Prmtr)) < 1.0e-4)break

old Prmtr = new Prmtr

iter = iter + 1}

list (muT = muT, muR = muR, sigmaT = sigmaT, sigmaR = sigmaR )}

### construct test statistics

equiv Test Wald RMLE < - function(obsT, obsR, eta, alpha)

{nT = length (obsT)

nR = length (obsR)

smplmuT = mean (obsT)

smplmuR = mean (obsR)

smplSigmaT = sd (obsT)

smplSigmaR = sd (obsR)

Vn = 2* ( (nR-1)/2 - exp( 2*(lgamma(nR/2) - lgamma((nR-1)/2)) ) )

nRnew < - ifelse(nR > 1.5*nT, 1.5*nT, nR)

nTnew < - ifelse(nT > 1.5*nR, 1.5*nR, nT)

est = CMLE_equivTest (obsT, obsR, -eta)

tstatL = (smplmuT - smplmuR + eta*smplSigmaR)/sqrt(est$sigmaT ^ 2/nTnew + (1.0/nRnew + eta ^ 2*Vn/(nR1))*est $sigmaR ^ 2)

est = CMLE_equivTest(obsT,obsR,eta)

tstatU = (smplmuT - smplmuR - eta*smplSigmaR)/sqrt(est $sigmaT ^ 2/nTnew + (1.0/nRnew + eta ^ 2*Vn/(nR1))*est $sigmaR ^ 2)

qntl = qnorm (alpha,lower.tail = F)

rslt = ifelse (tstatL > qntl&tstatU < -qntl,1,0) # the expression is changed in Section 6.2

rslt}

### main function

effsize < - 1.5 ### Type I Error

#effsize < - 0.125 ### power

eta < - 1.5

muR < - 0

sigmaT < - 1

sigmaR < - 1

alpha < - 0.05 # significance level

### equal number of lots for both product

nT < - 10

nR < - 10

nrep < - 10 ^ 6 # simulation replicates

muT < - muR + effsize*sigma_R

simEquivTestRep < - function(i){obsT = rnorm (nT,muT,sigmaT) obsR = rnorm (nR,muR,sigmaR) testRslt = equivTestWaldRMLE(obsT,obsR,abs(eta),alpha)}

### simulated type I error (or power) in percentage

rslt = lapply (1:nRep,simEquivTestRep)

T_typeIerr = mean (unlist(rslt))*100
```

print (T1_typeIerr)

A.2. R code for the application in Section 4

alpha = 0.05

eta = 1.7

obsT = c(94, 109, 103, 97, 102, 101, 99, 97, 97, 103)

obsR = c(96, 104, 102, 102, 101, 99, 99, 92, 107, 98)

testRslt = equiv Test Wald RMLE (obsT,obsR,abs(eta),alpha)

print (testRslt)

change the expression of the variable *rslt* in the function *equivTestWaldMLE* in 6.1 to calculate the confidence interval instead

rslt = c(smplmuT - smplmuR - qntl*sqrt (est$sigmaT ^ 2/nTnew + (1.0/nRnew + eta ^ 2*Vn/(nR-1))*est$sigmaR ^ 2), smplmuT - smplmuR + qntl*sqrt(est$sigmaT ^ 2/nTnew + (1.0/nRnew + eta ^ 2*Vn/(nR-1))*est$sigmaR ^ 2))

A.3. Supplementary Table A.1, Supplementary Table A.2, Supplementary Table A.3 and Supplementary Table A.4 in Section 3.

Supplementary Table A.1: Simulated type I error (%) of MWCMLE for small numbers of lots with different variance ratios if *f* = 1.5.

***f* = 1.5**	$n_T = n_R$					
σ_T^2 / σ_R^2	10	11	12	13	14	15
0.5	4.8141	4.779	4.7652	4.791	4.8138	4.7962
0.75	4.8705	4.8388	4.8195	4.8205	4.8669	4.8312
1	4.9202	4.8795	4.8502	4.8651	4.9156	4.8715
1.25	4.9568	4.9171	4.8959	4.8933	4.9337	4.8855
1.5	4.9713	4.9466	4.9229	4.9069	4.9496	4.9062
1.75	4.9813	4.956	4.9348	4.9189	4.9629	4.9067
2	4.9815	4.9635	4.9471	4.9296	4.9688	4.9193

Supplementary Table A.2: Simulated power (%) of MWCMLE for small numbers of lots with different variance ratios if f = 1.5 and $\mu_T - \mu_R = \sigma_R / 8$.

***F* = 1.5**	$n_T = n_R$					
σ_T^2 / σ_R^2	10	11	12	13	14	15
0.5	88.7347	91.591	93.7238	95.3343	96.5033	97.4131
0.75	83.2441	87.0385	89.9825	92.2921	94.0517	95.3956
1	77.4522	82.0568	85.7254	88.6778	91.0387	92.8733
1.25	71.6571	76.8777	81.1591	84.7034	87.6039	89.9219
1.5	65.9847	71.6597	76.4885	80.5246	83.8766	86.656
1.75	60.6115	66.5518	71.7641	76.215	80.0221	83.2018
2	55.5706	61.6356	67.1128	71.8774	76.0763	79.6336

Supplementary Table A.3: Simulated type I error (%) of MWCMLE for small numbers of lots with different variance ratios if *f* = 1.7.

f = 1.7	$n_T = n_R$					
σ_T^2 / σ_R^2	10	11	12	13	14	15
0.5	4.7844	4.7505	4.7416	4.7725	4.779	4.7584
0.75	4.8409	4.7984	4.7814	4.8061	4.8301	4.8004
1	4.8921	4.8445	4.8239	4.8339	4.8684	4.8348
1.25	4.9278	4.8847	4.8687	4.8706	4.8937	4.8583
1.5	4.9579	4.9155	4.902	4.8901	4.9141	4.8798
1.75	4.9764	4.9361	4.9205	4.9025	4.934	4.8977
2	4.9968	4.9477	4.9342	4.9191	4.9487	4.908

Supplementary Table A.4: Simulated power (%) of MWCMLE for small numbers of lots with different variance ratios if *f* = 1.7 and $\mu_T - \mu_R = \sigma_R / 8$.

***f* = 1.7**	$n_T = n_R$					
σ_T^2 / σ_R^2	10	11	12	13	14	15
0.5	93.7916	95.6515	96.9497	97.8712	98.5132	98.972
0.75	90.1553	92.7962	94.7675	96.2016	97.2224	97.9862
1	86.0391	89.4587	92.0614	94.0563	95.5519	96.6499
1.25	81.6338	85.7563	88.9856	91.5209	93.4698	94.9606
1.5	77.1378	81.8413	85.6255	88.6544	91.0912	92.9705
1.75	72.5993	77.8304	82.0648	85.5963	88.448	90.7082
2	68.1688	73.7671	78.4559	82.3834	85.6029	88.2306