2017

Generalized Confidence Intervals Compatible with the Min Test for Simultaneous Comparisons of One Subpopulation to Several Other Subpopulations

Julia N. Soulakova
University of Nebraska-Lincoln, Julia.Soulakova@ucf.edu

Published in final edited form as:
Commun Stat Theory Methods. 2017 ; 46(19): 9441–9449. doi:10.1080/03610926.2016.1212072.

Generalized Confidence Intervals Compatible with the Min Test for Simultaneous Comparisons of One Subpopulation to Several Other Subpopulations

Julia N. Soulakova
Department of Statistics, University of Nebraska-Lincoln, 340 Hardin Hall-North, Lincoln, NE 68583-0963, Phone: (402) 472-7231, Fax: (402) 472-5179

Abstract

A problem where one subpopulation is compared to several other subpopulations in terms of means with the goal of estimating the smallest difference between the means commonly arises in biology, medicine, and many other scientific fields. A generalization of Strassburger, Bretz and Hochberg (2004) approach for two comparisons is presented for cases with three and more comparisons. The method allows constructing an interval-estimator for the smallest mean difference, which is compatible with the Min test. An application to a fluency-disorder study is illustrated. Simulations confirmed adequate probability coverage for normally distributed outcomes for a number of designs.

1. Introduction

1.1. On Simultaneous Comparisons

Problems where one subpopulation is compared to several subpopulations in terms of a statistical parameter commonly arise in different scientific fields. For example, Louis et al. (2014) compared eight subpopulations of US and Polish students in terms of their attitudes towards stuttering. The study considered subpopulations corresponding to diverse majors, programs of study and cultures; this study is further discussed in Section 3. Jones et al, (2013) compared Whites, African-Americans and Mexican-Americans who were current smokers in terms of prevalence of menthol cigarette smoking and a set of biomarkers of tobacco exposure (e.g., serum cotinine, blood cadmium). One of the findings was that the prevalence of menthol cigarette smoking was higher for African-Americans than for Whites and Mexican-Americans. White et al. (2005) compared White, Black and Hispanic subpopulations in terms of the first ischemic stroke incidence using data from a population-based epidemiological study. They showed that Whites have lower rates of stroke than do Blacks and Hispanics.

In this paper I discuss how one can perform the simultaneous comparisons of one subpopulation to several other subpopulations using an interval-estimating approach. For this purpose, the confidence interval approach proposed by Strassburger, Bretz and Hochberg (2004) for a case of two comparisons (termed SBH approach) is generalized to handle three or more comparisons. Such a confidence interval is compatible with the corresponding Min test for testing intersection-union hypotheses. The paper is outlines as follows. First, the Min test and the corresponding SBH method for a case with two comparisons are reviewed in Section 1.2. Then the generalized SBH method is presented in Section 2 for three or more comparisons, and an application of the approach to a fluency-disorder study is discussed in Section 3. Next, details of a simulations study conducted to illustrate adequate performance of the generalized method are presented in Section 4. The paper concludes with several remarks in Section 5.

1.2. Background: Comparisons via the Min Test and SBH Interval Estimating

The intersection-union tests, including the Min test, are commonly used to perform simultaneous comparisons, when the goal is to demonstrate that one subpopulation is “superior” or “inferior” to all other subpopulations in terms of a parameter of interest. These tests and their properties have been addressed in the literature for several decades (Lehmann 1952, Berger 1982, Snapinn 1987, Hsu 1996). Specifically, the intersection-union tests were first discussed by Lehmann (1952), who showed a number of theoretical properties of the tests and illustrated applications for normal and multinomial models. However, as is pointed by Berger (1997), the term “an intersection-union test” appeared much later in Gleser (1973). Berger (1982) and Saikali and Berger (2002) addressed applications of the intersection-union tests for acceptance sampling problems that are commonly considered in quality control studies, and showed that in some settings the intersection-union test is more powerful than the likelihood-ratio test.

In biostatistics, the Min test is, probably, one of the most commonly used intersection-union tests. The term “Min test” was first introduced by Laska and Meisner (1989), who showed that for a normal model the Min test is also the likelihood ratio test. The authors discussed how one can use t-tests for normal distributions (with unknown but common variance), exact tests for binomial distributions, and Wilcoxon tests for unknown (absolutely continuous) distributions. They also provided formulas for sample size estimation and presented some tabulated results for cases involving two comparisons. Laska, Tang and Meisner (1992) extended the Min test methodology to a multivariate setting. Specifically, they discussed applications of the Min test for a multivariate normal distribution with known variance-covariance matrix, a multivariate normal distribution with unknown variance-covariance matrix, and unknown (absolutely continuous) multivariate distributions. Horn, Vollandt and Dunnett (2000) extended methods for sample size and power computing (Laska and Meisner 1989) to handle normal or nonparametric cases with more than two comparisons.

Applications of the Min test and methods based on the intersection-union tests for demonstrating drug efficacy have been addressed in several papers (Hung, Chi and Lipicky 1993, Wang and Hung 1997, Hung 2000, Westfall, Ho and Prillaman 2001. Tamhane and Logan 2004, Buchheister and Lehmacher 2006, Soulakova 2009, Soulakova 2010). Alternative approaches to the intersection-union tests were also discussed. In particular, Allison et al. (2002) proposed a Bayesian alternative to the Min test with respect to a gene expression problem, Bi (2005, 2007) proposed methods for illustrating similarity in consumer studies or demonstrating bioequivalence in drug-efficacy studies which utilize binomial models. In addition, Saikali and Berger (2002) proposed “smoother” tests for acceptance sampling problems with continuous distributions including normal models. This methodology was further extended by SenGupta (2007) who presented exact tests for acceptance sampling, bioequivalence and several other problems involving exponential and non-exponential families as well as mixture distributions.

The Min test can be described as follows. Consider a study, where a subpopulation, indexed $i = 0$, is compared to Kother subpopulations, indexed $i = 1, \dots, K$, in terms of the subpopulation means μ_k, $k = 0, \dots, K$, and the goal is to demonstrate that $\mu_0 > \mu_i$ for all i, $i = 1, 2, \dots, K$, Such a problem can be written in terms of multiple component hypotheses

$$H_{0i}: \mu_0 - \mu_i \leq 0 \text{ and } H_{ai}: \mu_0 - \mu_i > 0,\ i = 1, 2, \dots, K, \quad (1)$$

or in terms of the global null and alternative hypotheses, respectively, given by

$$H_0: \min_{1 \leq i \leq K} (\mu_0 - \mu_i) \leq 0 \text{ and } H_a: \min_{1 \leq i \leq K} (\mu_0 - \mu_i) > 0. \quad (2)$$

Similarly, if the study goal is to demonstrate that $\mu_0 < \mu_i$ for all i, $i = 1, 2, \dots, K$, then the component hypotheses are $H_{0i}: \mu_i - \mu_0 \leq 0$ and $H_{ai}: \mu_i - \mu_0 > 0$, $i = 1, 2, \dots, K$, and the global hypotheses are

$$H_0: \min_{1 \leq i \leq K} (\mu_i - \mu_0) \leq 0 \text{ and } H_a: \min_{1 \leq i \leq K} (\mu_i - \mu_0) > 0. \quad (3)$$

Note that in any case $H_0 = \bigcup_{1 \leq i \leq K} H_{0i}$ and $H_a = \bigcap_{1 \leq i \leq K} H_{ai}$, and thus, the global hypotheses are the intersection-union hypotheses. Then the Min test rejects the global null hypothesis H_0 in favor of H_a at level α if (and only if) all component null hypotheses H_{0i} are rejected at level α.

There is an alternative framework for the Min test in terms of the p-values. This framework has been presented elsewhere (Westfall, Ho and Prillaman 2001, Soulakova 2009) and is not discussed in this paper.

Strassburger, Bretz and Hochberg (2004) proposed several confidence intervals compatible with the intersection-union tests, including SBH interval compatible with the Min test, when a subpopulation is compared to two subpopulations via problem (2), i.e., $K = 2$. Let $\bar{y}_k$ be the sample mean response for the $k - th$ subsample (drawn from the $k - th$ subpopulation), $k = 0, 1, 2$, where $\bar{y}_k$*'s* are independent and $\bar{y}_k \sim N(\mu_k, \sigma^2/n_k)$, $k = 0, 1, 2$. In a case of unknown variance σ^2, consider the pooled sample variance estimator, $\hat{\sigma}^2$. Then, the lower $100(1 - \alpha)\%$ SBH confidence interval for the parameter $\min_{1 \leq i \leq 2} (\mu_0 - \mu_i)$ is given by $(L,\ +\infty)$, where $L = \min_{1 \leq i \leq 2} \left((\bar{y}_0 - \bar{y}_i) - t_{\alpha, \nu} \hat{\sigma} \sqrt{n_0^{-1} + n_i^{-1}} \right)$, and $t_{\alpha, \nu}$ is the $(1 - \alpha) - th$ percentile of the t-distribution with $\nu = n_0 + n_1 + n_2 - 3$ degrees of freedom. Note that the SBH lower bound is given by $L = \min_{1 \le i \le 2} A_i$, where A_i denotes the lower $100(1-\alpha)\%$ confidence bound for the component parameter $\mu_0 - \mu_i$, $i = 1, 2$. The authors also noted that even if the sample means $\bar{y}_k$*'s* are correlated, the method utilizes the critical value from a univariate *t*-distribution, because the maximum probability of Type I error for the Min test does not depend on the correlation.

In the considered setting, the SBH confidence interval (L, ∞) is compatible with the following Min test. Consider the component statistics $T_i = (\bar{y}_0 - \bar{y}_i)/(\hat{\sigma}\sqrt{n_0^{-1} + n_i^{-1}})$, $i = 1, 2$. Then the α-level test rejects the component null hypothesis H_{0i} if $T_i > t_{\alpha, \nu}$, where $\nu = n_0 + n_1 + n_2 - 3$, and accepts H_{0i} it if $T_i \le t_{\alpha, \nu}$. And the Min-test rejects the global null hypothesis H_0 in (2) if $\min_{1 \le i \le 2} T_i > c_\alpha$ and accepts H_0 if $\min_{1 \le i \le 2} T_i \le c_\alpha$. Similarly, the Min test for testing hypotheses (3) can be outlined.

2. GENERALIZATION OF THE SBH METHOD FOR THREE OR MORE COMPARISONS

In this section, I consider hypothesis problem (2); similar steps can be used in the case of problem (3). Let $\bar{y}_k$ be the sample mean response for the $k-th$ subsample (drawn from the $k-th$ subpopulation), $k = 0, 1, \ldots, K$, where $\bar{y}_k$ are independent and $\bar{y}_k \sim N(\mu_k, \sigma^2/n_k)$, $k = 0, 1, \ldots, K$. Consider the following component tests: if variance σ^2 is known for $k = 0, 1, \ldots, K$ then to test the component hypotheses one can use two-sample *z*-tests, and if variance is unknown then one can use two-sample *t*-tests with the rejection region $T_i > t_{\alpha, \nu}$, $i = 1, 2, \ldots, K$, where $\nu = n_0 + n_1 + \ldots + n_K - K$. Furthermore, let the lower $100(1-\alpha)\%$ confidence intervals for the component parameters $\mu_0 - \mu_i$, $i = 1, 2, \ldots, K$ be given by $(A_i, +\infty)$, $i = 1, 2, \ldots, K$, where in the case of known variance

$$A_i = (\bar{y}_0 - \bar{y}_i) - z_\alpha \sigma \sqrt{n_0^{-1} + n_i^{-1}}, \quad (4)$$

and in the case of unknown variance

$$A_i = (\bar{y}_0 - \bar{y}_i) - t_{\alpha, \nu} \hat{\sigma} \sqrt{n_0^{-1} + n_i^{-1}}, \quad (5)$$

and z_α denotes the $(1-\alpha)-th$ percentile of the standard normal distribution.

Next we illustrate how the SBH method can be generalized to handle simultaneous comparisons to three subpopulations, i.e., $K = 3$. Appendix presents the corresponding steps for a general case. To simplify the presentation, let $\theta_k = \mu_0 - \mu_k$, $i = 1, 2, 3$, then the component parameters can be represented by a vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$, $\boldsymbol{\theta} \in \Theta$, where $\Theta = \underset{1 \le k \le 3}{\cup} \left(H_{0k} \cup H_{ak}\right)$, i.e., $\Theta = \mathfrak{R}^3$.

Generalization of the SBH method as well as the original SBH method is based on the Partitioning Lemma (Strassburger, Bretz and Hochberg 2004), that allows constructing $100(1-\alpha)\%$ simultaneous confidence set for $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Theta$, provided that Θ can be partitioned into disjoint sets Θ_i such that $\Theta = \underset{i \in I}{\cup} \Theta_i$, where I is some index set, and there are local α-level tests for testing $\boldsymbol{\theta} \in \Theta_i$, for all $i \in I$. The resulting confidence set C (that depends on the component test statistics, e.g., $T_i's$) for $\boldsymbol{\theta}$ is given by the union of all $\Theta_i's$corresponding to the accepted hypotheses $\boldsymbol{\theta} \in \Theta_i$, and projecting the confidence set C on the coordinate axes results in the simultaneous lower confidence bounds for θ_i, $i = 1, 2, 3$. In order for the confidence set to be compatible with the Min test for testing hypotheses (2), the following two conditions must hold:

(C1) there exists a sub-partition of $\{\Theta_i, i \in I\}$, let us denote it by $\{\Theta_i, i \in J\}$, so that the null space can be represented as $H_0 = \underset{i \in J}{\cup} \Theta_i$, $J \subseteq I$, and

(C2) the null hypothesis $\boldsymbol{\theta} \in H_0$ is rejected in favor of $\boldsymbol{\theta} \in H_1$ if and only if all hypotheses $\boldsymbol{\theta} \in \Theta_i$, $i \in J$, are rejected by the corresponding α-level component tests.

Among several possible partitions we consider the one given by $\{\Theta_{(j,r)}, (j,r) \in J\}$, where $\Theta_{(1,r)} = \{\eta : \eta_1 = r, \eta_2 \ge r, \eta_3 \ge r\}$, $\Theta_{(2,r)} = \{\eta : \eta_1 > r, \eta_2 = r, \eta_3 \ge r\}$, $\Theta_{(3,r)} = \{\eta : \eta_1 > r, \eta_2 > r, \eta_3 = r\}$, $r \in \mathfrak{R}$, and $J = \{(j,r),\ j = 1, 2, 3,\ r \in \mathfrak{R}\}$. First, we will show that for such a partition, conditions *C1* and *C2* are satisfied. Because the null space can be represented as

$$\mathrm{H}_0 = \{\Theta_{(1,r)}, r \in (-\infty, 0]\} \cup \{\Theta_{(2,r)}, r \in (-\infty, 0]\} \cup \{\Theta_{(3,r)}, r \in (-\infty, 0]\},$$

condition *C1* is met. Also, we consider component tests such that an α-leveltest rejects $\theta \in \Theta_{(j,r)}$ if and only if $A_j \ge r$, $j = 1, 2, 3$, $r \in \mathfrak{R}$, and the Min test that rejects $\theta \in \mathrm{H}_0$ if and only if hypotheses $\theta \in \Theta_{(j,r)}$ are rejected for all (j,r), $j = 1, 2, 3$, $r \in (-\infty, 0]$. Therefore, condition *C2* is also met. Thus, both conditions are satisfied and therefore, the $100(1-\alpha)\%$ confidence set for θ is given by

$$C = \{\eta : \eta_1 > A_1, \eta_2 \ge \eta_1, \eta_3 \ge \eta_1\} \cup \{\eta : \eta_1 > \eta_2, \eta_2 > A_2, \eta_3 \ge \eta_2\} \cup \{\eta : \eta_1 > \eta_3, \eta_2 > \eta_3, \eta_3 > A_3\}.$$

Next, projecting the confidence set C on the coordinate axes provides the simultaneous confidence bounds for θ_i of the form $L_i = \underset{1 \le i \le 3}{\min} \{\eta_i : \eta \in C\}$. And the corresponding lower confidence bound for $\underset{1 \le i \le 3}{\min} \{\theta_i\}$ is given by $L = \underset{1 \le i \le 3}{\min} A_i$, where A_i is given by (4) if variance is known or (5) if variance is unknown. In general, as is presented in Appendix for K comparisons, the generalized SBH lower $100(1-\alpha)\%$ confidence bound for $\min_{1 \le i \le K} \theta_i$ is given by $\min_{1 \le i \le K} A_i$.

3. ESTIMATING THE SMALLEST AVERAGE DIFFERENCE IN ATTITUDES TOWARDS STUTTERING OF COLLEGE STUDENTS

Louis et al. (2014) reported results of a two-country study where US and Polish students majoring in speech-language pathology (SLP) and other disciplines were compared in terms of attitudes towards stuttering. Here we use subsamples corresponding to five non-overlapping subpopulations of US students, i.e., graduate SLP-major students, graduate non-SLP-major students, undergraduate SLP-major students, undergraduate non-SLP-major students, and Native American (NA) non-SLP-major students; note that the first four subpopulations correspond to non-NA students. The study used eight subsamples with 50 subjects per group, where five subsamples were drawn from the above subpopulations and three more were drawn from three subpopulations of Polish students. The study aim was to assess potential effect of SLP-major, training and cultural factors on students' attitudes toward stuttering. The statistical analyses included ANOVA and pairwise comparisons via Bonferroni adjustments for multiplicity. Among several conclusions, the authors stated that the SLP-major students have more positive attitude towards stuttering, on average, than do non-SLP-major students, and the US graduate students have more positive attitude towards stuttering, on average, than do undergraduate students.

To illustrate the generalized SBH interval we used 95% confidence level and performed comparisons of graduate SLP-major students to each of the four subpopulations in terms of the overall stuttering score, which ranges from −100 to 100 with higher scores corresponding to more positive attitudes towards stuttering. The goal was to interval-estimate the smallest difference in the average stuttering scores, i.e., the parameter of interest was $\min_{1 \le i \le 4} \{\mu_0 - \mu_i\}$, where indices $i = 0$, $i = 1$, $i = 2$, $i = 3$, and $i = 4$, respectively, corresponded to graduate SLP-major non-NA students, graduate non-SLP-major non-NA students, undergraduate SLP-major non-NA students, undergraduate non-SLP-major non-NA students, and NA non-SLP-major students. The problem was stated in terms of hypotheses (2) with $K = 4$.

Using the group sample sizes $n_i = 50$, $i = 0, 1, \ldots, 4$, and reported by Louis et al. (2014) summary statistics (that are illustrated in Table 1), we computed the pooled variance estimator 251, i.e., $\hat{\sigma} \approx 15.84$, and used $t_{0.05, 245} = 1.97$ to construct the lower bounds (5). Table 1 illustrates the corresponding lower bounds. Thus, the 95% lower bound for the minimum average difference in the overall stuttering score was $L = 3.76$. The value of the bound indicates that graduate SLP-major non-NA students, on average, have more positive attitude towards stuttering than do the other four subpopulations of students.

4. SIMULATIONS

The goal of the simulation study was to illustrate the theoretical property shown in Section 3, i.e., to illustrate that the probability coverage of the generalized SBH method is adequate in balanced and unbalanced settings with known and unknown variance when the distribution model is as described in Section 2. Cases with three comparisons, i.e., $K = 3$, were considered. The mean values $(\mu_0, \mu_1, \mu_2, \mu_3)$ were chosen to reflect four different cases of mean differences θ_i, $i = 1, 2, 3$, i.e., $(\theta_1 \le 0,\ \theta_2 \le 0,\ \theta_3 \le 0)$, $(\theta_1 > 0,\ \theta_2 \le 0,\ \theta_3 \le 0)$, $(\theta_1 > 0,\ \theta_2 > 0,\ \theta_3 \le 0)$ and $(\theta_1 > 0,\ \theta_2 > 0,\ \theta_3 > 0)$. Confidence levels of 90% and 95% were considered.

A single simulation run was as follows. For each specified setting of component parameter values $(\mu_0, \mu_1, \mu_2, \mu_3)$, confidence level, and sample sizes, depicted in Table 2, data were generated, so that $\bar{y}_k$ were independent and $\bar{y}_k \sim N(\mu_k, \sigma^2/n_k)$, $k = 0, 1, 2, 3$. In the case of known variance these data were used directly to construct the component bounds A_i, $i = 1, 2, 3$, via (4) and the corresponding lower bound $L = \min_{1 \le i \le 3} A_i$. In the case of unknown variance, independently on $\bar{y}_k's$, a value x was generated from the chi-square distribution with $\nu = n_0 + n_1 + n_2 + n_3 - 4$ degrees of freedom, the value was used to obtain the sample variance and construct the lower bounds (5) and the corresponding lower bound L. In any case, if the lower bound satisfied $L < \min_{1 \le i \le K} (\mu_i - \mu_0)$ then the confidence interval was said to capture the true parameter and the case was noted; otherwise, if $L \ge \min_{1 \le i \le K} (\mu_i - \mu_0)$ then the confidence was said not to capture the true parameter.

The above simulation steps were repeated 10^6 times. The proportion of replicates when the confidence interval captured the true parameter provided the estimated probability coverage. If the proportion was less than 94.97%, where $95 - 1.96\sqrt{95 * 5/10^6} = 94.97$, then the 95% confidence level interval was said to result in under-coverage, and if the proportion was less than 89.94%, then the 90% confidence level interval was said to result in under-coverage.

Table 2 depicts the results of the simulations. The confidence intervals did not result in under-coverage in any considered settings. In addition, in each considered setting, the two intervals based on (4) and (5) perform similarly for settings with group sample size of at least 50 (in this case, the t-distribution has 196 degrees of freedom). Among balanced settings (given the means, variance, and confidence level are fixed), the probability coverage decreases as the sample size increases. Similarly, in the case of known variance, the probability coverage increases as the variance increases (when the rest of the simulation parameters are fixed). In addition, in all settings, the probability coverage decreases as the confidence level decreases (when the other simulation parameters are fixed).

5. CONCLUSIONS

In this paper a generalization of the SBH method for interval-estimating compatible with the Min test for multiple simultaneous comparisons is discussed. The generalized SBH interval in based on the confidence intervals for the mean differences and can be easily computed when these confidence intervals are available. Results of the simulation study agreed with the theoretical result that for the considered model the approach has adequate probability coverage in balanced and unbalanced settings for different cases of the means, and known and unknown variance. The results also indicated that in some settings the generalized SBH interval can exhibit over-coverage, i.e., in these cases the Min test is over-conservative. An application of the method is illustrated via an example for simultaneous comparisons of one subpopulation of students to other four subpopulations of students in terms of the average overall stuttering score, a measure for assessing one's attitude towards stuttering. Similar settings, where a certain subpopulation is compared to several other subpopulations simultaneously, also commonly arise in fields other than educational psychology, e.g., other behavioral and medical sciences. In these settings, the generalized SBH method can provide an essential interval-estimation statistical tool.

Note that the probability coverage of the generalized SBH interval depends on the probability coverage of the intervals for the component parameters and thus, if one uses approximate component confidence intervals then performance of the proposed method should be first verified via theoretical derivations or simulations. Future research can be targeted on developing interval-estimating methods compatible with the Min test for other types of models, e.g., binomial. In addition, future research can focus on generalizing the SBH method for more complex designs, e.g., multistage surveys, as well as developing suitable computing software packages.

Acknowledgments

I would like to thank the referee and the Editor for their insightful comments and useful suggestions that greatly helped improve this paper. Research reported in this publication was supported by the National Institute On Minority Health And Health Disparities of the National Institutes of Health under Award Number R01MD009718. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

APPENDIX

To simplify the presentation, let $\theta_i = \mu_0 - \mu_i$, $i = 1, 2, \ldots, K$, then the parameter of interest is $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_K)$, $\boldsymbol{\theta} \in \Theta$, where $\Theta = \underset{1 \le i \le K}{\cup} \left(H_{0i} \cup H_{ai}\right) = \mathfrak{R}^K$. I consider the following sets

$$\Theta_{(1, r)} = \{\eta\colon \eta_1 = r,\ \eta_s \ge r \text{ for all } s = 2, 3, \ldots, K\},\ r \in \mathfrak{R};$$
$$\Theta_{(2, r)} = \{\eta\colon \eta_1 > r,\ \eta_2 = r,\ \eta_s \ge r \text{ for all } s = 3, 4, \ldots, K\},\ r \in \mathfrak{R};$$
$$\ldots$$
$$\Theta_{(j, r)} = \{\eta\colon \eta_t > r \text{ for all } t = 1, 2, \ldots, j-1,\ \eta_j = r,\ \eta_s \ge r \text{ for all } s = j+1, j+2, \ldots, K\},\ r \in \mathfrak{R};$$
$$\ldots$$
$$\Theta_{(K, r)} = \{\eta\colon \eta_t > r \text{ for all } t = 1, 2, \ldots, K-1,\ \eta_K = r\},\ r \in \mathfrak{R}\ .$$

The above sets are mutually exclusive and their union over the index set $J = \{(j, r),\ 1 \le j \le K,\ r \in \mathfrak{R}\}$ is Θ, i.e., $\Theta = \bigcup_{1 \le j \le K} \{\Theta_{(j,r)}, r \in \mathfrak{R}\}$. Thus, these sets provide a partition of Θ. Also, the null space is given by the union of the sets over the index set $J' = \{(j, r),\ 1 \le j \le K,\ r \in (-\infty, 0]\}$. Consider a local α-level test that rejects $\theta \in \Theta_{(j,r)}$ if and only if $A_j \ge r$, where $j = 1, 2, \dots, K$, $r \in \mathfrak{R}$; and $\theta \in \mathrm{H}_0$ is rejected if and only if hypotheses $\theta \in \Theta_{(j,r)}$ are rejected for all (j, r) such that $j = 1, 2, \dots, K$ and $r \in (-\infty, 0]$. Therefore, the $100(1-\alpha)\%$ confidence set for θ is given by

$$\begin{aligned} C = &\{\eta : \eta_t > A_1,\ \eta_s \ge \eta_1 \text{ for all } s = 2, 3, \dots, K\} \\ &\cup \{\eta : \eta_t > \eta_2,\ \eta_2 > A_2,\ \eta_s \ge \eta_2 \text{ for all } s = 3, 4, \dots, K\} \\ &\dots \\ &\cup \{\eta : \eta_t > \eta_j \text{ for all } t = 1, 2, \dots, j-1,\ \eta_j > A_j,\ \eta_s \ge \eta_j \text{ for all } s = j+1, j+2, \dots, K\} \\ &\dots \\ &\cup \{\eta : \eta_t > \eta_K \text{ for all } t = 1, 2, \dots, K-1,\ \eta_K > A_K\}. \end{aligned}$$

And thus, the lower $100(1-\alpha)\%$ confidence bound for $\min_{1 \le i \le K} \theta_i$ is given by $\min_{1 \le i \le K} A_i$, where A_i, $i = 1, 2, \dots, K$, is as depicted in (4) and (5) for known and unknown variance, respectively.

References

Allison DB, Gadbury GL, Heo M, Fernández JR, Lee C, Prolla TA, Weindruch R. A mixture model approach for the analysis of microarray gene expression data. Computational Statistics and Data Analysis. 2002; 39:1–20.

Berger RL. Multiparameter hypothesis testing and acceptance sampling. Technometrics. 1982; 24:295–300.

Berger, RL. Advances in statistical decision theory and applications. Birkhäuser; Boston: 1997. Likelihood ratio tests and intersection-union tests; p. 225-237.

Bi J. Similarity testing in sensory and consumer research. Food Quality and Preference. 2005; 16:139–149.

Bi J. Similarity testing using paired comparison method. Food Quality and Preference. 2007; 18:500–507.

Buchheister B, Lehmacher W. Multiple testing procedures for identifying desirable dose combinations in bifactorial designs. GMS Medical Informatics, Biometry und Epidemiology. 2006; 2(2):1–11.

Gleser LJ. On a theory of intersection-union tests. Institute of Mathematical Statistics Bulletin. 1973; 2:233–233.

Horn M, Vollandt R, Dunnett CW. Sample size determination for testing whether an identified treatment is best. Biometrics. 2000; 56:879–881. [PubMed: 10985230]

Hsu, J. Multiple Comparisons: Theory and Methods. CRC Press; 1996.

Hung HM. Global tests for combination drug studies in factorial trials. Statistics in Medicine. 2000; 19:2079–2087. [PubMed: 10931512]

Hung HM, Chi GYH, Lipicky RJ. Testing for the existence of a desirable dose combination. Biometrics. 1993; 49:85–94. [PubMed: 8513112]

Jones MR, Apelberg BJ, Tellez-Plaza M, Samet JM, Navas-Acien A. Menthol cigarettes, race/ethnicity, and biomarkers of tobacco use in U.S. adults: The 1999–2010 National Health and Nutrition Examination Survey (NHANES),’. Cancer Epidemiology, Biomarkers and Prevention. 2013; 22(2):224–232.

Laska EM, Meisner MJ. Testing whether an identified treatment is best. Biometrics. 1989; 45:1139–1151. [PubMed: 2611321]

Laska EM, Tang D, Meisner MJ. Testing hypotheses about an identified treatment when there are multiple endpoints. Journal of the American Statistical Association. 1992; 87(419):825–831.

Lehmann EL. Testing multiparameter hypotheses. The Annals of Mathematical Statistics. 1952; 23:541–552.

Louis KO, Przepiorka AM, Beste-Guldborg A, Williams MJ, Blachnio A, Guendouzi J, Reichel IK, Ware MB. Stuttering attitudes of students: professional, intracultural, and international comparisons. Journal of Fluency Disorders. 2014; 39:34–50. [PubMed: 24759192]

Saikali KG, Berger RL. More powerful tests for the sign testing problem. Journal of statistical planning and inference. 2002; 107(1):187–205.

SenGupta A. P3 approach to intersection–union testing of hypotheses. Journal of statistical planning and inference. 2007; 137(11):3753–3766.

Snapinn SM. Evaluating the efficacy of a combination drug therapy. Statistics in Medicine. 1987; 6:657–665. [PubMed: 2961039]

Soulakova JN. On identifying effective and superior drug combinations via Holm’s procedure based on the Min tests. Journal of Biopharmaceutical Statistics. 2009; 19:280–291. [PubMed: 19212880]

Soulakova JN. General multistage gatekeeping procedures for identifying beneficial drug combinations in factorial trials with isotonic gains. Statistics in Biopharmaceutical Research. 2010; 2:33–41.

Strassburger K, Bretz F, Hochberg Y. Compatible confidence intervals for intersection union tests involving two hypotheses. Institute of Mathematical Statistics Lecture Notes-Monograph Series. 2004; 47:129–142.

Tamhane AC, Logan BR. A superiority-equivalence approach to one-sided tests on multiple endpoints in clinical trials. Biometrika. 2004; 91(3):715–727.

Wang S, Hung HM. Large sample tests for binary outcomes in fixed-dose combination drug studies. Biometrics. 1997; 53:498–503. [PubMed: 9192448]

Westfall PH, Ho SY, Prillaman BA. Properties of multiple intersection-union tests for multiple endpoints in combination therapy trials. Journal of Biopharmaceutical Statistics. 2001; 11(3):125–138. [PubMed: 11725927]

White H, Boden-Albala B, Wang C, Elkind MS, Rundek T, Wright CB, Sacco RL. Ischemic stroke subtype incidence among Whites, Blacks, and Hispanics; the Northern Manhattan Study. Circulation. 2005; 111:1327–1331. [PubMed: 15769776]

Table 1

Summary Statistics Taken from Louis et al. (2014) and Lower Bounds for the Mean Differences

Group*	Sample Mean Stuttering Score (SE)	95% Lower Bounds for Mean Difference
Graduate SLP-major non-NA students	43 (17)	
Graduate non-SLP-major non-NA students	31 (15)	5.76
Undergraduate SLP-major non-NA students	33 (14)	3.76
Undergraduate non-SLP-major non-NA students	24 (16)	12.76
NA non-SLP-major students	31 (17)	5.36

*SLP stands for speech-language pathology, NA stands for Native American.

Table 2

Simulation Configurations and Results

Mean Values $\mu_0, \mu_1, \mu_2, \mu_3$	Group Sample Sizes n_0, n_1, n_2, n_3	Standard Deviation σ	Confidence Level	Variance is Known	Variance is Not Known
0, 0.1, 0.5, 1.0	5,5,5,5	1	95%	97.44	97.30
			90%	93.88	93.74
		10	95%	99.37	99.27
			90%	98.08	97.93
	10,5,5,5	1	95%	97.96	97.83
	5,10,5,5			97.21	97.11
	5,10,10,5			96.59	96.51
	5,10,10,10			96.08	96.05
	10,10,5,5			97.71	97.61
	10,10,10,10	1		96.56	96.52
			90%	92.29	92.23
		10	95%	99.30	99.26
			90%	97.92	97.85
	50,50,50,50	1	95%	95.11	95.11
	100,50,50,50			95.09	95.09
	50,100,50,50			95.07	95.07
	50,100,100,50			95.01	95.03
	50,100,100,100			95.02	95.02
	100,100,50,50			95.15	95.15
	100,100,100,50			94.99	94.99
	500,500,500,500			95.01	95.01
		10		97.45	97.45
0.5, 0.0, 1.0, 1.5	5,5,5,5	1		97.16	97.02
	10,10,10,10			96.45	96.40
	500,500,500,500			95.00	95.00

Mean Values $\mu_0, \mu_1, \mu_2, \mu_3$	Group Sample Sizes n_0, n_1, n_2, n_3	Standard Deviation σ	Confidence Level	Variance is Known	Variance is Not Known
0.5, 0.4, 0.1, 1.0	5,5,5,5			97.15	96.99
	10,10,10,10			96.22	96.17
	500,500,500,500			94.97	94.97
0.5, 0.4, 0.1, 0.0	5,5,5,5			98.61	98.45
	10,10,10,10			98.09	98.01
	500,500,500,500			95.02	95.02