Leibniz
Universität Hannover

Reports of the Institute of Biostatistics

No 04 / 2008

Leibniz University of Hannover

Natural Sciences Faculty

Titel: *Robustness simulation study for continuous endpoints in one-way-layouts*

Author: *Mario Hasler*

1 Introduction

Most multiple comparison procedures (MCP) for means of two or more normally distributed populations can be evaluated as multiple contrast tests (MCT) or by related simultaneous confidence intervals (SCI). Several contrasts, representing linear functions of these means, are estimated and tested for deviation from 0. The many-to-one comparison of Dunnett [4] is the most famous example. The trend test of Williams [11] is a frequently used test, too, when the experimental trail involves several treatments/ doses and a control. Bretz [1] formulated this trend test as an approximate MCT. Dilba et al. [2] dealt with MCT and SCI for ratios of means. All MCT and SCI have the basic assumption of normal distributed data with homogeneous group variances and sufficient sample sizes. This fact is sometimes ignored or hard to proof by the experimenter. Count data, for example, are known not to follow a normal distribution but Poisson or binomial. These distributions imply heterogeneous group variances automatically. Even when the data are normally distributed, variance heterogeneity may occur. E.g., dose finding studies can have the problem of heteroscedasticity because the data's variance depends on the dose effect (see the data in [10]). The sample size is tried to be minimized for reasons of costs. And so, it is a common proceeding to apply these procedures without making sure whether these assumptions hold true. When no information about the data is available (from preliminary tests) before statistical analysis, it is advisable to use more robust test procedures, e.g. the Wilcoxon test for non-normal distributions or the Welch test [9] for heterogeneous variances (see also [5]). Otherwise, existing effects or negligible differences may be under- or overestimated, respectively, which leads to wrong test decisions.

2 General assumptions and background

The following simulation studies look on the influence of sample size, heteroscedasticity, and distribution misclassification on upper SCI of Dunnett and Williams type. For this purpose, let there be $i = 0, \ldots, k$ groups of observations, where $i = 0$ means the control group and $k = 2, 4, 8$. The groups have means μ_i, standard deviations σ_i, and sample sizes n_i. The control group's mean is $\mu_0 = 10$ for all studies. Interest is in estimating and testing both differences $\delta_i = \mu_i - \mu_0$ and ratios $\gamma_i = \mu_i/\mu_0$ to control by one-sided (upper) SCI ($i = 1, \ldots, k$). The simultaneous coverage probability (SCP) is simulated. The nominal level is 0.95 for all studies. Values beyond 0.961 indicate significant conservatism while values below 0.938 liberalism. The results were obtained by 1000 simulation runs with starting seed 100 using the statistical software R [7]. The functions `glht` of the packages `multcomp` [6] and `sci.ratio` of `mratios` [3] are used for differences and ratios of means, respectively.

Two situations, H_1 and H_2, are considered for the three issues. Situation H_1 always represents the case that all groups have equal means, $\mu_0 = \cdots = \mu_k = 10$. H_2 denotes that only the non-control groups ($i = 1, \ldots, k$) have equal means, this is: $\mu_i = \mu_0 + 5 = 15$ for the difference based problem, and $\mu_i = \mu_0 * 1.5 = 15$ for the ratio based one. Unlike for differences, SCI for ratios of means are approximate. The correlation matrix $\boldsymbol{R}$ of the needed multivariate t-distribution additionally depends on the unknown ratios γ_i for which the intervals are to be constructed. The method to plug in the estimators $\hat{\gamma}_i = \hat{\mu}_i/\hat{\mu}_0$ ($i = 1, \ldots, k$) instead (see [2]), is most universally valid and the default in the R package `mratios`. Hence, a prior conclusion is that the SCP of the intervals for ratios γ_i will vary depending on the situations (H_1 or H_2).

When looking on SCI for ratios, further problems arise which do not occur for differences. If the denominator of the ratio (mean of the control group, μ_0) is not significantly greater than 0, related SCI are not defined at all. Some of the following tables have values in parentheses giving the rates of undefined SCI which appeared during the simulation. Coverage probabilities with such an extra value are calculated with respect to these rates and have to be handled with care. First, they do not have 1000 simulation runs as a base like the others. And second, they are conditional to the existence of their intervals. When all cases are taken out where the mean of the control group is not significantly greater than 0, the remaining ratios are biased because too small. This censoring also leads to non-normally distributed sample means for the control group. These general problems of SCI for ratios is sparely discussed in the literature. Further simulations and conclusions on this can be read in [8].

3 Sample size

For $i = 0, \ldots, k$, data were simulated following normal distributions with equal standard deviations $\sigma = 2, 4, 6$ and equal sample sizes $n = 2, 4, 8, 12$.

Tables 1 and 2 show the results of the simulations for the SCP of upper SCI for differences of type Dunnett and Williams, respectively. The confidence level is maintained for all sample sizes - even for the smallest possible ($n = 2$) - for both types, and all numbers of comparisons (k). The chosen standard deviations (σ) and the situations (H_1 or H_2) have no influence.

Tables 3 and 4 analogously show the results for upper SCI for ratios. The values in parentheses show the rates of undefined intervals which appeared during the simulation. Because of the discussed reasons in Section 2, the SCP very varies around the nominal level here. The cases where no undefined SCI appeared behave good and lead to similar results as for differences above. The intervals are most stable for small standard deviations (σ) and higher sample sizes n. The situation (H_1 or H_2) causes variation but a unique influence can not be seen.

It must further be noted that the results for $n = 4$, $k = 8$, H_1 and H_2, and both types are obtained with changed starting seeds (10100 for Dunnett, 1100 for Williams) to avoid instability. The used R function `sci.ratio` calculates necessary quantiles numerically over an interval from 0 to 10. Depending on the sample - but very rarely - these quantiles may realize values larger than 10 which causes instability for the simulation. So this is just a numerical problem, not a conceptual.

Situation	σ	$n = 2$	$n = 4$	$n = 8$	$n = 12$
$k = 2, H_1$	2	0.943	0.947	0.946	0.945
$k = 2, H_2$		0.943	0.947	0.946	0.945
$k = 4, H_1$		0.953	0.952	0.943	0.949
$k = 4, H_2$		0.953	0.952	0.943	0.949
$k = 8, H_1$		0.960	0.943	0.952	0.941
$k = 8, H_2$		0.960	0.943	0.952	0.941
$k = 2, H_1$	4	0.943	0.947	0.946	0.945
$k = 2, H_2$		0.943	0.947	0.946	0.945
$k = 4, H_1$		0.953	0.952	0.943	0.949
$k = 4, H_2$		0.953	0.952	0.943	0.949
$k = 8, H_1$		0.960	0.943	0.952	0.941
$k = 8, H_2$		0.960	0.943	0.952	0.941
$k = 2, H_1$	6	0.943	0.947	0.946	0.945
$k = 2, H_2$		0.943	0.947	0.946	0.945
$k = 4, H_1$		0.953	0.952	0.943	0.949
$k = 4, H_2$		0.953	0.952	0.943	0.949
$k = 8, H_1$		0.960	0.943	0.952	0.941
$k = 8, H_2$		0.960	0.943	0.952	0.941

Table 1: Coverage probability of upper SCI (differences) for the Dunnett procedure with several balanced sample sizes n, situations, and standard deviations σ; simultaneous confidence level 0.95.

Situation	σ	$n = 2$	$n = 4$	$n = 8$	$n = 12$
$k = 2, H_1$	2	0.946	0.941	0.950	0.947
$k = 2, H_2$		0.946	0.941	0.950	0.947
$k = 4, H_1$		0.948	0.958	0.956	0.957
$k = 4, H_2$		0.948	0.958	0.956	0.957
$k = 8, H_1$		0.943	0.958	0.952	0.957
$k = 8, H_2$		0.943	0.958	0.952	0.957
$k = 2, H_1$	4	0.946	0.941	0.950	0.947
$k = 2, H_2$		0.946	0.941	0.950	0.947
$k = 4, H_1$		0.948	0.958	0.956	0.957
$k = 4, H_2$		0.948	0.958	0.956	0.957
$k = 8, H_1$		0.943	0.958	0.952	0.957
$k = 8, H_2$		0.943	0.958	0.952	0.957
$k = 2, H_1$	6	0.946	0.941	0.950	0.947
$k = 2, H_2$		0.946	0.941	0.950	0.947
$k = 4, H_1$		0.948	0.958	0.956	0.957
$k = 4, H_2$		0.948	0.958	0.956	0.957
$k = 8, H_1$		0.943	0.958	0.952	0.957
$k = 8, H_2$		0.943	0.958	0.952	0.957

Table 2: Coverage probability of upper SCI (differences) for the Williams procedure with several balanced sample sizes n, situations, and standard deviations σ; simultaneous confidence level 0.95.

Situation	σ	$n=2$	$n=4$	$n=8$	$n=12$
$k=2, H_1$	2	0.942 (0.005)	0.945	0.946	0.945
$k=2, H_2$		0.937 (0.004)	0.944	0.942	0.944
$k=4, H_1$		0.950	0.943	0.951	0.950
$k=4, H_2$		0.932	0.954	0.955	0.950
$k=8, H_1$		0.953	0.943	0.946	0.944
$k=8, H_2$		0.951	0.944	0.951	0.944
$k=2, H_1$	4	0.939 (0.273)	0.943 (0.007)	0.946	0.943
$k=2, H_2$		0.940 (0.249)	0.944 (0.007)	0.941	0.944
$k=4, H_1$		0.954 (0.258)	0.943 (0.007)	0.957	0.942
$k=4, H_2$		0.951 (0.208)	0.948 (0.001)	0.942	0.936
$k=8, H_1$		0.958 (0.243)	0.940 (0.004)	0.963	0.946
$k=8, H_2$		0.949 (0.162)	0.940 (0.001)	0.953	0.941
$k=2, H_1$	6	0.935 (0.582)	0.960 (0.129)	0.950 (0.006)	0.941
$k=2, H_2$		0.935 (0.555)	0.967 (0.118)	0.944 (0.004)	0.943
$k=4, H_1$		0.937 (0.651)	0.958 (0.163)	0.941 (0.010)	0.939
$k=4, H_2$		0.955 (0.554)	0.966 (0.126)	0.947 (0.002)	0.947
$k=8, H_1$		0.966 (0.647)	0.980 (0.162)	0.947 (0.006)	0.933
$k=8, H_2$		0.977 (0.563)	0.987 (0.142)	0.933 (0.004)	0.946

Table 3: Coverage probability of upper SCI (ratios) for the Dunnett procedure with several balanced sample sizes n, situations, and standard deviations σ; simultaneous confidence level 0.95.

Situation	σ	$n=2$	$n=4$	$n=8$	$n=12$
$k=2, H_1$	2	0.944 (0.002)	0.941	0.949	0.946
$k=2, H_2$		0.946 (0.002)	0.942	0.945	0.946
$k=4, H_1$		0.949	0.936	0.933	0.951
$k=4, H_2$		0.948	0.945	0.954	0.952
$k=8, H_1$		0.941	0.950	0.949	0.951
$k=8, H_2$		0.960	0.951	0.959	0.948
$k=2, H_1$	4	0.951 (0.217)	0.940 (0.004)	0.948	0.944
$k=2, H_2$		0.957 (0.204)	0.938 (0.003)	0.945	0.946
$k=4, H_1$		0.948 (0.175)	0.945 (0.002)	0.952	0.953
$k=4, H_2$		0.942 (0.132)	0.937 (0.001)	0.947	0.960
$k=8, H_1$		0.963 (0.103)	0.942	0.946	0.951
$k=8, H_2$		0.962 (0.078)	0.949 (0.002)	0.931	0.956
$k=2, H_1$	6	0.953 (0.515)	0.956 (0.107)	0.951 (0.004)	0.943
$k=2, H_2$		0.958 (0.496)	0.963 (0.098)	0.943 (0.002)	0.946
$k=4, H_1$		0.966 (0.476)	0.969 (0.094)	0.954 (0.002)	0.941
$k=4, H_2$		0.970 (0.433)	0.974 (0.102)	0.941 (0.002)	0.941
$k=8, H_1$		0.983 (0.425)	0.968 (0.076)	0.931 (0.008)	0.954
$k=8, H_2$		0.981 (0.412)	0.974 (0.076)	0.942 (0.001)	0.942

Table 4: Coverage probability of upper SCI (ratios) for the Williams procedure with several balanced sample sizes n, situations, and standard deviations σ; simultaneous confidence level 0.95.

4 Heteroscedasticity

For $i = 0, \ldots, k$, data were simulated following normal distributions. The total sample sizes were 30 ($k = 2$), 50 ($k = 4$), and 90 ($k = 8$). Three different settings were considered, and they are

1. balanced allocation: $n_i = 10$ for all $i = 0, \ldots, k$,
2. uniformly increasing sample sizes over the groups:
   - $k = 2$: $n_0 = 6, n_1 = 10, n_2 = 14$,
   - $k = 4$: $n_0 = 6, n_1 = 8, n_2 = 10, n_3 = 12, n_4 = 14$,
   - $k = 8$: $n_0 = 6, n_1 = 7, n_2 = 8, n_3 = 9, n_4 = 10, n_5 = 11, n_6 = 12, n_7 = 13, n_8 = 14$,
3. uniformly decreasing sample sizes over the groups:
   - $k = 2$: $n_0 = 14, n_1 = 10, n_2 = 6$,
   - $k = 4$: $n_0 = 14, n_1 = 12, n_2 = 10, n_3 = 8, n_4 = 6$,
   - $k = 8$: $n_0 = 14, n_1 = 13, n_2 = 12, n_3 = 11, n_4 = 10, n_5 = 9, n_6 = 8, n_7 = 7, n_8 = 6$.

Moreover, increasing standard deviations over the groups were chosen:

- $k = 2$: $\sigma_0 = 1, \sigma_1 = 3, \sigma_2 = 5$,
- $k = 4$: $\sigma_0 = 1, \sigma_1 = 2, \sigma_2 = 3, \sigma_3 = 4, \sigma_4 = 5$,
- $k = 8$: $\sigma_0 = 1, \sigma_1 = 1.5, \sigma_2 = 2, \sigma_3 = 2.5, \sigma_4 = 3, \sigma_5 = 3.5, \sigma_6 = 4, \sigma_7 = 4.5, \sigma_8 = 5$.

Setting 1 consequently gives the same sample size to all groups. Setting 2 puts the highest sample size to the group with the highest standard deviation, while setting 3 gives the highest sample size to the group with the smallest standard deviation.

Tables 5 and 6 show the results of the simulations for the SCP of upper SCI for differences of type Dunnett and Williams, respectively. For all settings, the situations (H_1 or H_2) have no influence. For setting 1, conservatism can be seen for both types. This conservatism gets precarious for unbalanced settings like setting 2. The case when small sample size meets high standard deviation (setting 3) leads to liberal confidence intervals. The number of groups itself has also no influence in principle. But the three settings, of course, depend on the number of groups. The differences of the sample sizes and standard deviations between the groups vary for different group numbers. That influences the simulation results.

Tables 7 and 8 show the analogical results for upper SCI for ratios. The conclusions are about comparable with those above. The most important difference is that the coverage probability of SCI for ratios of means very depends on the situations (H_1 or H_2) because of the discussed reasons in Section 2. Values for H_2 are always higher than for H_1. So, the Dunnett intervals tend to conservatism for setting 1 and H_1, and get very conservative for H_2. The Williams intervals indeed show a better behavior for H_1, but they get conservative as well for H_2. Setting 2 yields even more conservative intervals for both types and situations. Setting 3 leads to liberalism in general and especially for H_1.

Generally, when the groups with the highest variation have the smallest sample size, the confidence intervals get liberal. They are too small then, their coverage probability is too small. Tests based on

Situation	Setting 1	Setting 2	Setting 3
$k = 2, H_1$	0.961	0.992	0.894
$k = 2, H_2$	0.961	0.992	0.894
$k = 4, H_1$	0.955	0.994	0.905
$k = 4, H_2$	0.955	0.994	0.905
$k = 8, H_1$	0.975	0.995	0.901
$k = 8, H_2$	0.975	0.995	0.901

Table 5: Coverage probability of upper SCI (differences) for the Dunnett procedure with several settings and situations; simultaneous confidence level 0.95.

Situation	Setting 1	Setting 2	Setting 3
$k = 2, H_1$	0.958	0.990	0.888
$k = 2, H_2$	0.958	0.990	0.888
$k = 4, H_1$	0.959	0.988	0.873
$k = 4, H_2$	0.959	0.988	0.873
$k = 8, H_1$	0.959	0.987	0.911
$k = 8, H_2$	0.959	0.987	0.911

Table 6: Coverage probability of upper SCI (differences) for the Williams procedure with several settings and situations; simultaneous confidence level 0.95.

them do not keep the familywise error rate (FWER). The converse case (highest sample size to the highest variation) leads to conservative intervals. They are too wide, their coverage probability is too high, and related tests do not exploit the FWER.

Further simulation results concerning heteroscedasticity can be seen in [5]. The focus there was less in SCI but more in test procedures. There is a strict one-to-one relation between MCT and SCI for differences of means, so their conclusions are comparable. This relation does not hold strictly for ratios of means because the SCI are just approximate unlike the test. The conclusions for MCT must be transfered to SCI with caution.

Situation	Setting 1	Setting 2	Setting 3
$k = 2, H_1$	0.960	0.992	0.894
$k = 2, H_2$	0.982	0.998	0.924
$k = 4, H_1$	0.958	0.990	0.901
$k = 4, H_2$	0.985	1.000	0.933
$k = 8, H_1$	0.970	0.995	0.911
$k = 8, H_2$	0.991	0.998	0.939

Table 7: Coverage probability of upper SCI (ratios) for the Dunnett procedure with several settings and situations; simultaneous confidence level 0.95.

Situation	Setting 1	Setting 2	Setting 3
$k = 2, H_1$	0.958	0.990	0.886
$k = 2, H_2$	0.980	0.996	0.918
$k = 4, H_1$	0.944	0.992	0.904
$k = 4, H_2$	0.983	0.999	0.920
$k = 8, H_1$	0.944	0.982	0.897
$k = 8, H_2$	0.978	0.998	0.927

Table 8: Coverage probability of upper SCI (ratios) for the Williams procedure with several settings and situations; simultaneous confidence level 0.95.

5 Distribution misclassification

For $i = 0, \ldots, k$, data were simulated following the distributions:

- normal $\mathrm{N}(\mu_i, \sigma^2)$ with $\sigma = 3$,
- mixed normal
  - $0.5\,\mathrm{N}(\mu_i - 2, \sigma^2) + 0.5\,\mathrm{N}(\mu_i + 2, \sigma^2)$ for differences,
  - $0.5\,\mathrm{N}(0.8\mu_i, \sigma^2) + 0.5\,\mathrm{N}(1.2\mu_i, \sigma^2)$ for ratios

  with $\sigma = 3$,
- Poisson pois(λ_i) with $\lambda_0 = \mu_0$,
- exponential exp(λ_i) with $\lambda_0 = 1/\mu_0$.

The normal distribution (norm) was taken as a control. A mixed normal distribution (2norm) can appear, if two groups are pooled through an oversight, or if effects overlay each other. (Note that random numbers from two independent normal distributions have been created, not from the joint mixed distribution.) The Poisson distribution (pois) was taken because frequently appearing, especially for count data. It is a discrete and non-symmetric distribution. Its variance (λ) is equal to the mean (λ). When the data follow discrete distributions, ties can appear. Problems arise when the sample sizes are too small. In that case, it may happen that there are groups with only equal values. Then no variances can be estimated, confidence intervals are not available and hence, no conclusions about them. The exponential distribution (exp) is non-symmetric, too, but continuous. Its variance ($1/\lambda^2$) is equal to the squared mean ($1/\lambda$). Data related to time measurements often have this one. Both distributions' variances depend on their means. This implies heterogeneous variances as soon as the means are different like under H_2.

Tables 9 and 10 show the results of the simulations for the SCP of upper SCI for differences of type Dunnett and Williams, respectively. Of course, the data for norm keep the level of the SCP independently from the chosen sample sizes; the situations (H_1 or H_2) have no influence. This independence also holds for 2norm but the intervals are all conservative. The remaining distributions differ depending on the situations. In general, pois and exp yield good SCP for H_1, where all groups have equal variances. But situation H_2 causes conservatism because the non-control groups have different variances than the control here. pois sometimes deviates from this conclusion but that is just owing to the high variance. All the conclusions hold for both types (Dunnett and Williams).

These conclusion are also valid for the SCI for ratios in principal. Tables 11 and 12 show the results of the simulations for the SCP for ratios. The values in parentheses show the rates of undefined SCI which appeared during the simulation. Because of the reasons discussed in Section 2, the conclusions differ a bit from those for differences.

Situation	n	norm	2norm	pois	exp
$k = 2, H_1$	10	0.953	0.982	0.961	0.939
$k = 2, H_2$		0.953	0.982	0.946	0.968
$k = 4, H_1$		0.944	0.979	0.962	0.955
$k = 4, H_2$		0.944	0.979	0.965	0.977
$k = 8, H_1$		0.947	0.983	0.942	0.941
$k = 8, H_2$		0.947	0.983	0.966	0.966
$k = 2, H_1$	50	0.941	0.989	0.951	0.941
$k = 2, H_2$		0.941	0.989	0.961	0.966
$k = 4, H_1$		0.960	0.989	0.948	0.938
$k = 4, H_2$		0.960	0.989	0.954	0.964
$k = 8, H_1$		0.960	0.992	0.954	0.931
$k = 8, H_2$		0.960	0.992	0.952	0.966
$k = 2, H_1$	100	0.941	0.974	0.949	0.957
$k = 2, H_2$		0.941	0.974	0.949	0.968
$k = 4, H_1$		0.945	0.980	0.959	0.952
$k = 4, H_2$		0.945	0.980	0.970	0.975
$k = 8, H_1$		0.955	0.986	0.961	0.946
$k = 8, H_2$		0.955	0.986	0.965	0.970

Table 9: Coverage probability of upper SCI (differences) for the Dunnett procedure with several distributions, situations, balanced sample sizes n; simultaneous confidence level 0.95.

Situation	n	norm	2norm	pois	exp
$k = 2, H_1$	10	0.944	0.978	0.963	0.941
$k = 2, H_2$		0.944	0.978	0.949	0.969
$k = 4, H_1$		0.950	0.983	0.949	0.954
$k = 4, H_2$		0.950	0.983	0.967	0.978
$k = 8, H_1$		0.953	0.980	0.944	0.955
$k = 8, H_2$		0.953	0.980	0.971	0.979
$k = 2, H_1$	50	0.955	0.980	0.943	0.943
$k = 2, H_2$		0.955	0.980	0.971	0.970
$k = 4, H_1$		0.952	0.982	0.953	0.958
$k = 4, H_2$		0.952	0.982	0.967	0.985
$k = 8, H_1$		0.954	0.978	0.950	0.955
$k = 8, H_2$		0.954	0.978	0.974	0.986
$k = 2, H_1$	100	0.937	0.971	0.951	0.954
$k = 2, H_2$		0.937	0.971	0.957	0.972
$k = 4, H_1$		0.951	0.980	0.946	0.959
$k = 4, H_2$		0.951	0.980	0.964	0.987
$k = 8, H_1$		0.953	0.978	0.946	0.961
$k = 8, H_2$		0.953	0.978	0.970	0.987

Table 10: Coverage probability of upper SCI (differences) for the Williams procedure with several distributions, situations, balanced sample sizes n; simultaneous confidence level 0.95.

Situation	n	norm	2norm	pois	exp
$k = 2, H_1$	10	0.949	0.981	0.960	0.959 (0.082)
$k = 2, H_2$		0.946	0.990	0.955	0.996 (0.296)
$k = 4, H_1$		0.949	0.985	0.946	0.975 (0.133)
$k = 4, H_2$		0.949	0.994	0.949	0.998 (0.425)
$k = 8, H_1$		0.944	0.989	0.947	0.982 (0.189)
$k = 8, H_2$		0.936	0.990	0.973	1.000 (0.588)
$k = 2, H_1$	50	0.940	0.987	0.950	0.939
$k = 2, H_2$		0.946	0.993	0.966	0.974
$k = 4, H_1$		0.952	0.983	0.945	0.945
$k = 4, H_2$		0.952	0.995	0.967	0.967
$k = 8, H_1$		0.950	0.988	0.944	0.939
$k = 8, H_2$		0.952	0.997	0.966	0.973
$k = 2, H_1$	100	0.941	0.974	0.948	0.953
$k = 2, H_2$		0.941	0.982	0.960	0.982
$k = 4, H_1$		0.937	0.982	0.950	0.938
$k = 4, H_2$		0.948	0.994	0.971	0.975
$k = 8, H_1$		0.950	0.989	0.944	0.950
$k = 8, H_2$		0.953	0.997	0.972	0.976

Table 11: Coverage probability of upper SCI (ratios) for the Dunnett procedure with several distributions, situations, balanced sample sizes n; simultaneous confidence level 0.95.

Situation	n	norm	2norm	pois	exp
$k = 2, H_1$	10	0.943	0.978	0.962	0.966 (0.066)
$k = 2, H_2$		0.942	0.988	0.956	0.993 (0.253)
$k = 4, H_1$		0.930	0.978	0.942	0.959 (0.074)
$k = 4, H_2$		0.951	0.986	0.963	0.994 (0.327)
$k = 8, H_1$		0.943	0.974	0.953	0.972 (0.091)
$k = 8, H_2$		0.935	0.990	0.957	0.998 (0.370)
$k = 2, H_1$	50	0.954	0.980	0.938	0.940
$k = 2, H_2$		0.949	0.992	0.973	0.979
$k = 4, H_1$		0.947	0.978	0.947	0.956
$k = 4, H_2$		0.959	0.993	0.960	0.987
$k = 8, H_1$		0.949	0.981	0.947	0.955
$k = 8, H_2$		0.942	0.988	0.979	0.976
$k = 2, H_1$	100	0.937	0.971	0.947	0.954
$k = 2, H_2$		0.934	0.982	0.962	0.983
$k = 4, H_1$		0.943	0.976	0.938	0.946
$k = 4, H_2$		0.951	0.991	0.964	0.978
$k = 8, H_1$		0.958	0.980	0.943	0.952
$k = 8, H_2$		0.942	0.990	0.977	0.989

Table 12: Coverage probability of upper SCI (ratios) for the Williams procedure with several distributions, situations, balanced sample sizes n; simultaneous confidence level 0.95.

6 Conclusions and recommendations

When applying confidence intervals as well as statistical tests, the user must always make assumptions on the data. In practice, it is hard to decide whether these assumptions and necessary conditions are really fulfilled or not. Hence, violation appears very often. This simulation study had a look on the influence of sample size, heteroscedasticity, and distribution misclassification on upper SCI of Dunnett and Williams type. In principle, SCI for ratios of means have the weak point of not being defined if the denominator is not significantly different from 0. Furthermore, they are just approximate which sometimes leads to a higher sensitivity for violation of data assumptions. Apart from these points, their behavior is quite similar to the intervals for differences.

Generally, the sample size itself has no critical influence on both intervals if existing. The smallest possible sample size per group is $n = 2$. Although, small sample size is not advisable because it can additionally intensify problems with heteroscedasticity and distribution misclassification.

The effect of heteroscedasticity is diverse. Depending on the allocation, conservatism or liberalism appears, respectively. When the data are heteroscedastic and the sample sizes are balanced, the confidence intervals get conservative. They are too wide, their coverage probability is too high. The intervals for ratios are more conservative when the non-control groups differ from the control. When groups with high variation have high sample size, the confidence intervals get conservative, too. The intervals for ratios are also more conservative when the non-control groups differ from the control. When groups with high variation have small sample size, the confidence intervals get liberal. They are too small then, their coverage probability is too small. The intervals for ratios are more liberal when the non-control groups do not differ from the control. If this situation is expected, it is advisable to use more robust intervals based on Bonferroni adjusted Welch t-test [9] or MCT, which account for heteroscedasticity (see [5]).

Distribution misclassification is the most critical point. If the variance of the real underlying distribution depends on the mean (like Poisson and exponential), heteroscedasticity is additionally present. Then both effects can compound or annul each other. When the data do not follow a normal distribution but a mixed normal, the confidence intervals get conservative. For Poisson or exponential distribution, the confidence intervals yield good coverage probabilities when the non-control groups do not differ from the control, and get conservative when they do. This is because additional unequal group variances then. Furthermore, the above conclusions about heteroscedasticity can be transfered, especially for unbalanced sample sizes which were not simulated here. So it is recommended to use SCI which are stable against heteroscedasticity if a mild violation is expected, and intervals related to non-parametric test procedures (e.g.Bonferroni-adjusted Wilcoxon tests) else.

References

[1] F. Bretz. An extension of the williams trend test to general unbalanced linear models. *Computational Statistics & Data Analysis*, 50(7):1735–1748, 2006.

[2] G. Dilba, F. Bretz, V. Guiard, and L. A. Hothorn. Simultaneous confidence intervals for ratios with applications to the comparison of several treatments with a control. *Methods Of Information In Medicine*, 43(5):465–469, 2004.

[3] G. Dilba, M. Hasler, and F. Schaarschmidt. *mratios: Inferences for ratios of coefficients in the general linear model*, 2008. R package version 1.3.5.

[4] C. W. Dunnett. A multiple comparison procedure for comparing several treatments with a control. *Journal Of The American Statistical Association*, 50(272):1096–1121, 1955.

[5] M. Hasler and L. A. Hothorn. Multiple contrast tests in the presence of heteroscedasticity. *Biometrical Journal*, 50(5):793–800, 2008.

[6] T. Hothorn, F. Bretz, and P. Westfall with contributions by R. M. Heiberger. *multcomp: Simultaneous Inference for General Linear Hypotheses*, 2008. R package version 0.993-2.

[7] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[8] F. Schaarschmidt. Confidence intervals for the ratio of means of two normal distributed populations in the presence of heteroscedasticity. Technical Report 17, Leibniz University of Hannover, Natural Sciences Faculty, Institute of Biostatistics, 2007.

[9] B. L. Welch. The significance of the difference between two means when the population variances are unequal. *Biometrika*, 29:350–362, 1938.

[10] P. H. Westfall. Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, 92(437):299–306, 1997.

[11] D. A. Williams. A test for differences between treatment means when several dose levels are compared with a zero dose control. *Biometrics*, 27:103–117, 1971.