Leibniz Universität Hannover

Reports of the Institute of Biostatistics

No 17 / 2007

Leibniz University of Hannover

Natural Sciences Faculty

Titel: *Confidence intervals for the ratio of means of two normal distributed populations in the presence of heteroscedasticity*

Authors: *Frank Schaarschmidt*

1 Introduction

Consider two independent random variables X_i, $i = 1, 2$, following two normal distributions with possibly differing in means μ_i and variances σ_i^2: $X_i \sim N\left(\mu_i, \sigma_i^2\right)$, with μ_i and σ_i unknown. Interest is in estimating the ratio of means $\rho = \mu_2/\mu_1$.

2 Confidence intervals for ρ

Confidence intervals for ρ in the assumed situation have not been published and explicitly characterized. Tamhane and Logan (2004) consider tests of the null-hypothesis $H_0 : \mu_2/\mu_1 = \rho_0$, where ρ_0 is an hypothesized value of ρ. Hasler et al. (in press) consider a similar situation for three populations. The confidence interval defined below is a special case of the method presented in their paper. However, Hasler et al. (in press) provide only very limited characterization of the presented confidence interval method, which is not sufficient for a general recommendation of the method.
Denoting the sample estimates for the mean $\hat{\mu}_2$, and $\hat{\mu}_1$, and for the sample variances $\hat{\sigma}_2^2$, and $\hat{\sigma}_1^2$, a Fieller-type interval (Fieller (1954)) is presented in equation (1).

$$\left[\frac{B}{-2A} \pm \frac{\sqrt{(B/2)^2 - AC}}{A}\right] \tag{1}$$

with

$$A = \left(\hat{\sigma}_1^2 c^2/n_1\right) - \hat{\mu}_1^2,$$
$$B = 2\hat{\mu}_2\hat{\mu}_1$$
$$C = \left(\hat{\sigma}_2^2 c^2/n_2\right) - \hat{\mu}_2^2,$$
$$c = t_{df=df_S\left(\hat{\rho},\hat{\sigma}_1^2,\hat{\sigma}_1^2,n_1,n_2\right),1-\alpha/2},$$

and finally

$$df_S\left(\hat{\rho}, \hat{\sigma}_1^2, \hat{\sigma}_1^2, n_1, n_2\right) = \frac{\hat{\sigma}_2^2/n_2 + \rho^2\hat{\sigma}_1^2/n_1}{\hat{\sigma}_2^2/\left(n_2^2\left(n_2-1\right)\right) + \rho^4\hat{\sigma}_1^2/\left(n_1^2\left(n_1-1\right)\right)}$$

.
Note, that the confidence degenerates in case that $A \geq 0$, i.e. μ_1 is not significantly different from 0. The confidence bounds obtained from equation (1) depend on the degree of freedom which depends on the known constants n_1, n_2 and the unknown parameters σ_1^2, σ_2^2 and ρ. Also in the test proposed by Tamhane and Logan (2004), the Satterthwaite-adjusted test and confidence interval for the difference of means the degrees of freedom depends on σ_1^2, and σ_2^2. Still the latter have been shown to perform well in practical situations, see e.g. Wang and Chow (2002). However, the dependence on ρ is special to the multiplicative formulation of hypotheses and might introduce an additional variation or even bias to the confidence bounds. In the test, the degree of freedom is computed under the null-hypothesis for a fixed value $\rho = \rho_0$, hence the problem is not discussed by Tamhane and Logan (2004). For confidence interval construction, two solutions are possible:
First, one might plug-in the estimated ratio parameter $\hat{\rho} = \hat{\mu}_2/\hat{\mu}_1$ and compute the degree of freedom based on this estimate. Such a confidence interval will be at least slightly inconsitent with Tamhane

and Logan's test for a number of situations, but is computationally simple.
Second, confidence bounds can be found by iteratively searching for values of ρ_0, such that Equation 2 is satisfied.

$$\frac{\hat{\mu}_2 - \rho_0\hat{\mu}_1}{\sqrt{\hat{\sigma}_2^2/n_2 + \rho_0^2\hat{\sigma}_1^2/n_1}} = t_{df=df_S(\rho_0),1-\alpha/2} \tag{2}$$

Such an interval will be consistent with the corresponding test by Tamhane and Logan (2004), but is computationally more difficult.

3 Objective

This report aims to characterize the coverage probability of the confidence limits constructed using Equation 1 using a plug-in of $\hat{\rho}$ to calculate the degree of freedom in dependence of $n_1, n_2, \sigma_1^2, \sigma_2^2$ and ρ.

4 Summary

Generally, both intervals show comparable and acceptable coverage probabilities. If the sample sizes are severly unbalanced (e.g., n_1=5, $n_2 = 100$), slightly liberal confidence intervals may occur. In such situations the coverage probability of two-sided nominal 95% confidence intervals rarely fell below 94% and never fell below 93% in the considered settings.
Lower Fieller confidence limits may be slightly liberal when the sample size in the numerator group is much lower than in the denominator group or the sample size is very low ($n_i = 5$) in both groups and the numerator group shows a higher variance. The coverage probability rarely was below 97% and never fell below 96.5%.
Upper Fieller confidence limits may be slightly liberal when the sample size in the denominator group is much lower than in the numerator. The coverage probability rarely was below 97% and never fell below 96.5%.

5 Simulation study

5.1 Settings

A simulation study was performed for all 720 combinations of the following parameters: $\mu_1 = 100$, $\mu_2 = \rho\mu_1 = 10, 50, 80, 90, 100, 111, 125, 200, 1000$, $CV_2 = 0.01, 0.05, 0.1, 0.5, 1$, $CV_1 = 0.1$, $n_2 = 5, 10, 20, 100$, $n_1 = 5, 10, 20, 100$, where $CV_i = \sigma_i/\mu_i$. Coverage probabilities were considered for two-sided 95%-intervals, as well as of lower and upper 97.5%-bounds separately.
For each parameter setting, the coverage probability was estimated based on 10,000 simulation runs. Note, that for a single experiment with 10,000 Bernoulli trials, 9,537 successes are the lowest count which allows to conclude that the proportion of success is significantly larger than 0.95 at the 5% level; 9,463 successes are the highest count which allows to conclude that the proportion is significantly lower than 0.95 at the 5% level.

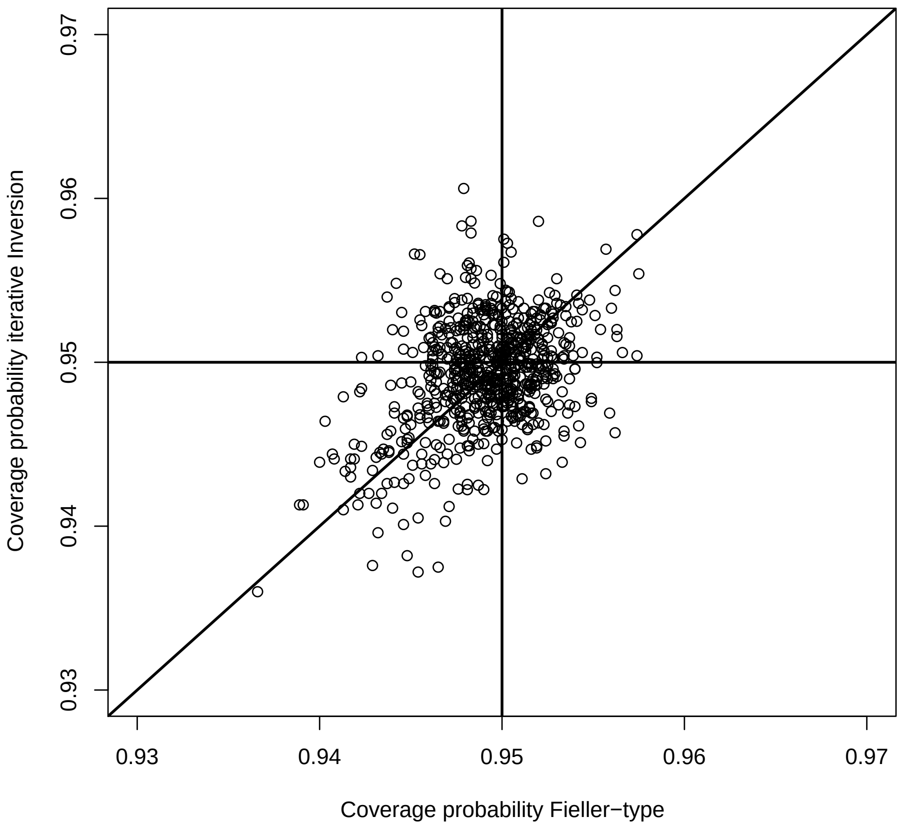

Figure 1: Comparison of coverage probabilities of two-sided nominal 95% Fieller-type intervals and the interval derived by test inversion

5.2 Detailed results

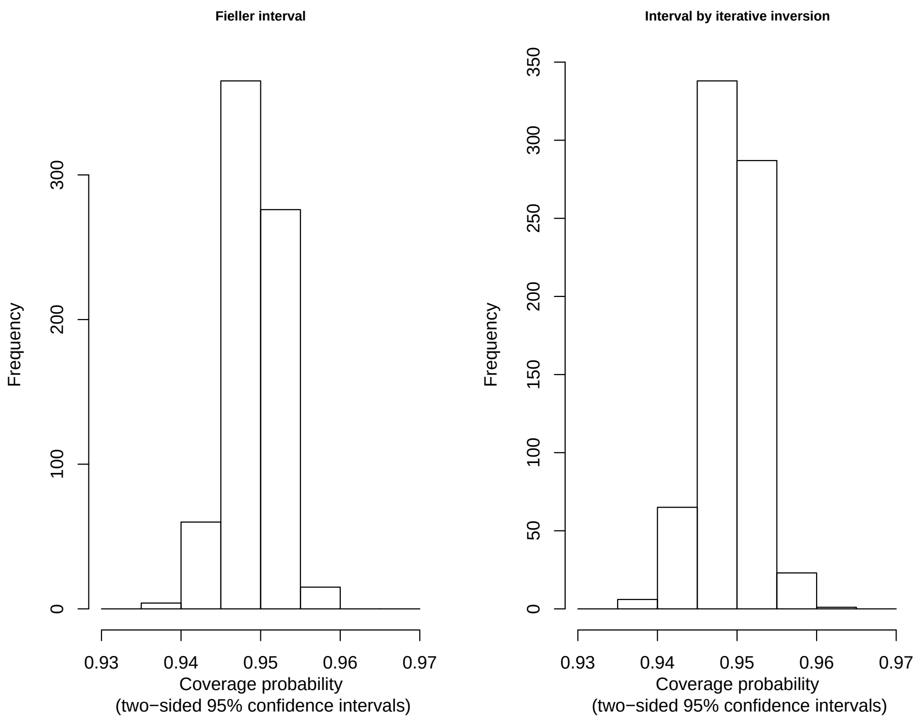

Figure 2: Coverage probabilities of two-sided nominal 95% Fieller-type intervals and the intervals derived by test inversion

Coverage probability			Mean		CV		Sample size	
two-sided	lower	upper	μ_1	μ_2	CV_2	CV_1	n_2	n_1
0.9366	0.9676	0.9690	100	80	0.05	0.10	5	100
0.9389	0.9680	0.9709	100	111	0.05	0.10	5	100
0.9391	0.9683	0.9708	100	1000	0.05	0.10	5	100

Table 1: Settings, for which the coverage probability of nominal 95% two-sided Fieller-type intervals fell below 94%.

Coverage probability			Mean		CV		Sample size	
two-sided	lower	upper	μ_1	μ_2	CV_2	CV_1	n_2	n_1
0.9372	0.9689	0.9683	100	10	0.05	0.10	20	5
0.9382	0.9725	0.9657	100	10	0.10	0.10	100	5
0.9375	0.9698	0.9677	100	100	0.10	0.10	5	20
0.9360	0.9670	0.9690	100	80	0.05	0.10	5	100
0.9376	0.9700	0.9676	100	125	0.05	0.10	5	100
0.9396	0.9706	0.9690	100	200	0.05	0.10	5	100

Table 2: Settings, for which the coverage probability of nominal 95% two-sided intervals derived by test iterative inversion fell below 94%.

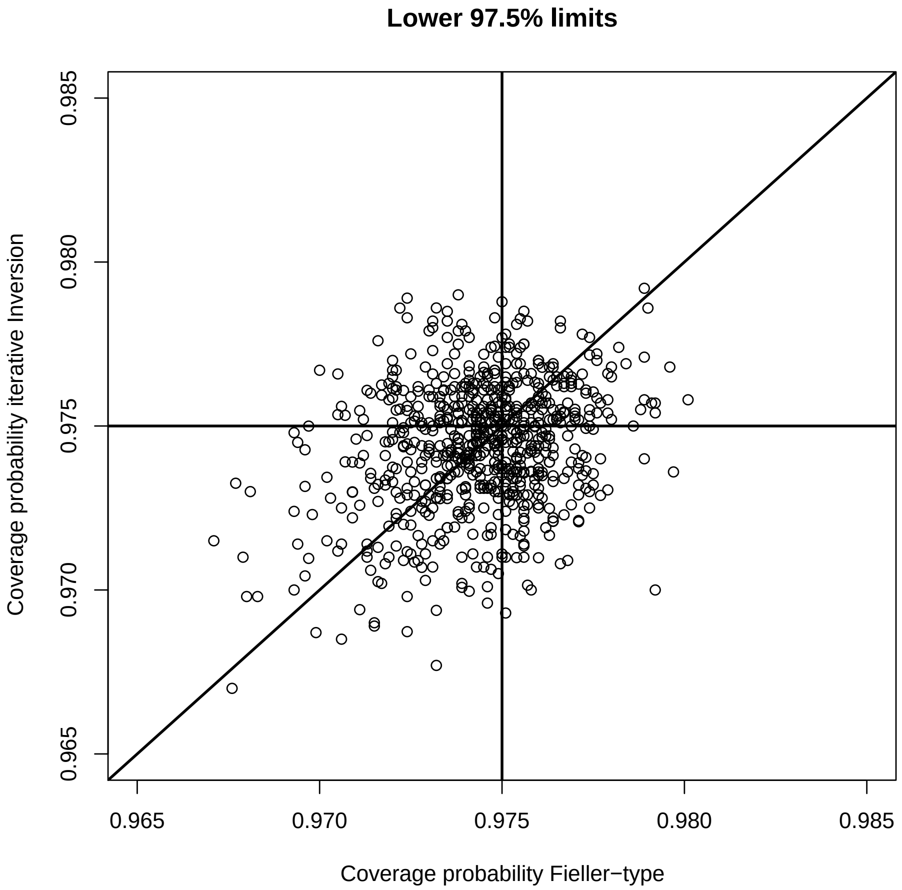

Figure 3: Comparison of coverage probabilities of lower nominal 97.5% Fieller-type limits and the limits derived by test inversion

5.2.1 Lower 97.5% limits

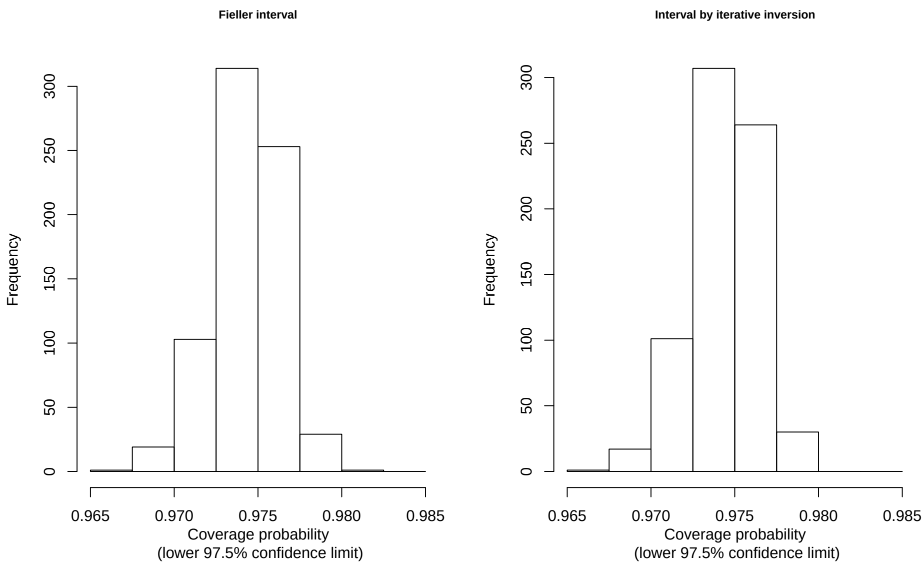

Figure 4: Coverage probabilities of lower nominal 97.5% Fieller-type limits and the limits derived by test inversion

Coverage probability	Mean		CV		Sample size	
	μ_1	μ_2	CV_2	CV_1	n_2	n_1
0.9677	100	111	0.50	0.10	5	5
0.9694	100	125	0.50	0.10	5	5
0.9696	100	200	0.50	0.10	5	5
0.9697	100	1000	0.50	0.10	5	5
0.9696	100	50	1.00	0.10	5	5
0.9694	100	50	0.10	0.10	5	10
0.9696	100	200	0.50	0.10	5	10
0.9699	100	50	0.10	0.10	5	20
0.9693	100	111	0.10	0.10	5	20
0.9693	100	125	0.10	0.10	5	20
0.9697	100	200	0.10	0.10	5	20
0.9676	100	80	0.05	0.10	5	100
0.9679	100	100	0.05	0.10	5	100
0.9680	100	111	0.05	0.10	5	100
0.9693	100	125	0.05	0.10	5	100
0.9683	100	1000	0.05	0.10	5	100
0.9698	100	50	0.10	0.10	5	100
0.9671	100	100	0.10	0.10	5	100
0.9681	100	200	0.05	0.10	10	100

Table 3: Settings for which coverage probability of lower 97.5 % limits of Fieller-type intervals fell below 97%

Coverage probability	Mean		CV		Sample size	
	μ_1	μ_2	CV_2	CV_1	n_2	n_1
0.9700	100	10	0.50	0.10	5	5
0.9694	100	111	1.00	0.10	5	5
0.9689	100	10	0.05	0.10	20	5
0.9696	100	111	0.05	0.10	20	5
0.9690	100	10	0.10	0.10	20	5
0.9677	100	90	0.10	0.10	20	5
0.9687	100	50	0.10	0.10	5	20
0.9694	100	90	0.10	0.10	5	20
0.9698	100	100	0.10	0.10	5	20
0.9687	100	1000	0.50	0.10	5	20
0.9685	100	10	0.05	0.10	5	100
0.9670	100	80	0.05	0.10	5	100
0.9698	100	111	0.05	0.10	5	100
0.9698	100	1000	0.05	0.10	5	100
0.9693	100	1000	0.10	0.10	100	100

Table 4: Settings for which coverage probability of lower 97.5 % limits of intervals by iterative inversion fell below 97%

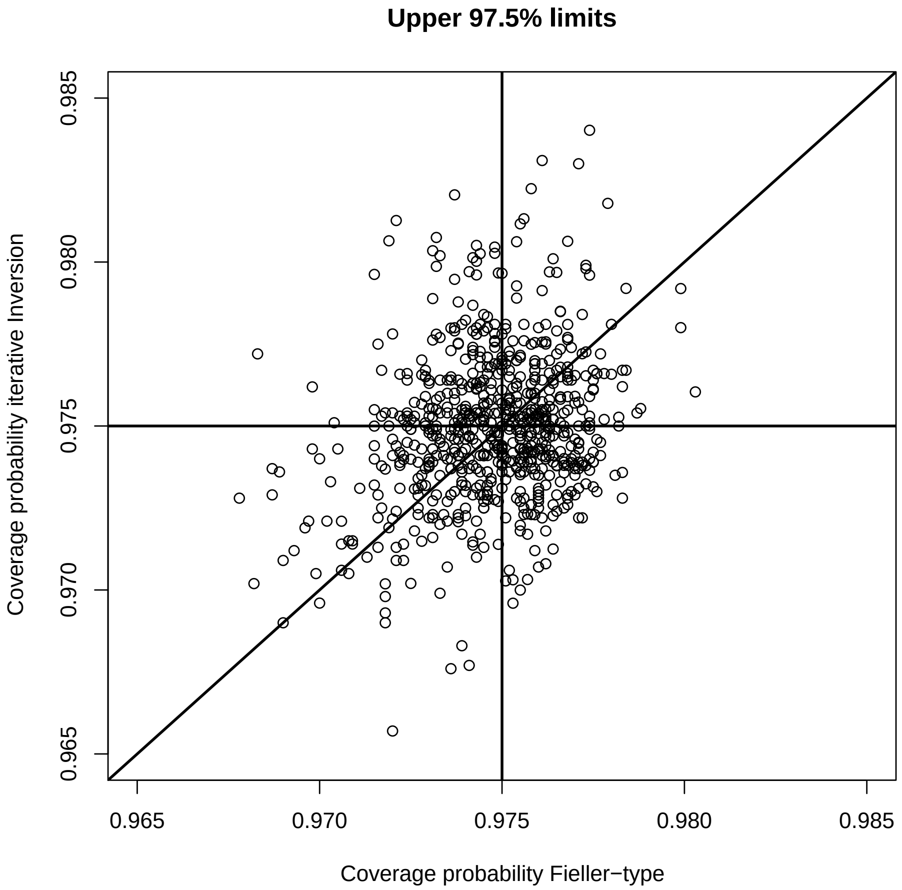

Figure 5: Comparison of coverage probabilities of nominal upper 97.5% Fieller-type limits and the limits derived by test inversion

5.3 Upper 97.5% bounds

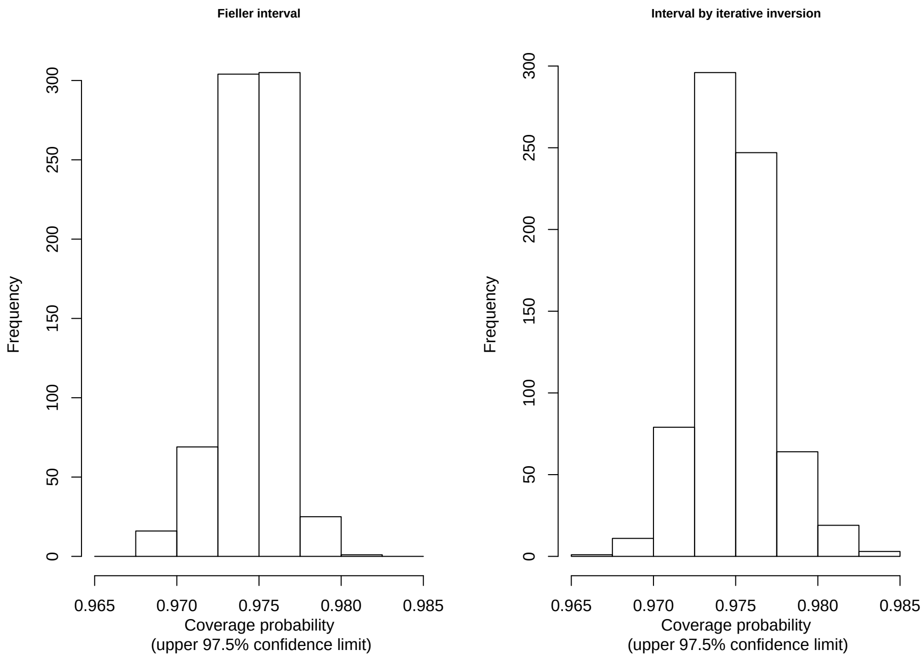

Figure 6: Coverage probabilities of upper nominal 97.5% Fieller-type limits and the limits derived by test inversion

Coverage probability	Mean		CV		Sample size	
	μ_1	μ_2	CV_2	CV_1	n_2	n_1
0.9698	100	200	0.01	0.10	5	5
0.9690	100	90	0.05	0.10	10	5
0.9699	100	100	0.05	0.10	20	5
0.9687	100	200	0.05	0.10	20	5
0.9689	100	90	0.10	0.10	20	5
0.9682	100	100	0.10	0.10	20	5
0.9697	100	125	0.10	0.10	20	5
0.9693	100	50	0.10	0.10	100	5
0.9678	100	80	0.10	0.10	100	5
0.9696	100	111	0.10	0.10	100	5
0.9698	100	1000	0.05	0.10	10	10
0.9683	100	125	0.10	0.10	100	10
0.9690	100	80	0.05	0.10	5	100
0.9687	100	111	0.10	0.10	5	100

Table 5: Settings for which coverage probability of upper 97.5 % limits of Fieller-type intervals fell below 97%

Coverage probability	Mean		CV		Sample size	
	μ_1	μ_2	CV_2	CV_1	n_2	n_1
0.9693	100	80	0.05	0.10	10	5
0.9683	100	10	0.05	0.10	20	5
0.9657	100	10	0.10	0.10	100	5
0.9699	100	200	0.50	0.10	100	5
0.9677	100	100	0.10	0.10	5	20
0.9696	100	125	0.10	0.10	5	20
0.9700	100	1000	0.05	0.10	20	20
0.9696	100	50	0.05	0.10	5	100
0.9690	100	80	0.05	0.10	5	100
0.9698	100	90	0.05	0.10	5	100
0.9676	100	125	0.05	0.10	5	100
0.9690	100	200	0.05	0.10	5	100

Table 6: Settings for which coverage probability of upper 97.5 % limits of intervals by iterative inversion fell below 97%

References

Fieller, EC (1954): Some problems in interval estimation. Journal of the Royal Statistical Society, Series B, 16: 175-185.

Hasler, M, Vonk, R, and Hothorn, LA (2007): Assessing non-inferiority of a new treatment in a three-arm trial in the presence of heteroscedasticity. Statistics in Medicine, accepted for publication.

Tamhane, AC, and Logan, BR (2004): Finding the maximum save dose level for heteroscedastic data. Journal of Biopharmaceutical Statistics 14, 843-856.

Wang, H, and Chow, S-C (2002): A practical approach for comparing means of two groups without equal variance assumption. Statistics in Medicine 21, 3137-3151.