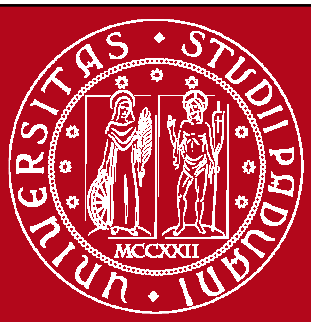


UNIVERSITÀ DEGLI STUDI DI PADOVA

OVERRUNNING DATA METHODS: COMPARISONS BASED ON REAL DATA TRIAL

NICOLA SORIANI, ILEANA BALDI, DARIO GREGORI

ROeS 2013
Dornbirn, September 9th - 12th 2013

UNIVERSITÀ DEGLI STUDI DI PADOVA

Table of contents

UNIVERSITÀ DEGLI STUDI DI PADOVA

Introduction

- The sample size of a clinical trial is set off to ensure good properties (type-I error rate and power) of the test on primary endpoints.
- In group sequential trials, the design foresees one or more interim analyses (IA) before that the full sample size is reached. (Armitage et al.,1975; Pocock 1977).
- The primary purpose of such IAs is to stop the trial when either futility or superiority of one intervention becomes clear (according to a stopping criterion).
- Overrunning occurs when data continue to be collected even if a stopping criterion has been reached (Whitehead, 1992).
- Overrunning is often due to the time delay between the subject recruitments and the actual evaluations (Whitehead, 1992).

UNIVERSITÀ DEGLI STUDI DI PADOVA

Goals

- To study the effect of including overrunning data on the behaviors of the methods proposed in the literature over the years.
- To study if and how the overrunning data sizes affect on the method levels of type-I error and power.
- To determine whether one of these methods could be suggested for a systematic use when overrunning occurs.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Interim Analyses

- The advantage on a major endpoint between experimental (*E*) and control (*C*) treatment groups is expressed by a parameter θ.
- IAs are performed in terms of Z_k (efficient score for the superiority) and V_k (Fisher's Information) statistics, under the assumption

$$Z_k \dot{\sim} N(\theta V_k, V_k).$$

- Stopping criterion is determined by group sequential test (Pocock, 1977; O'Brien and Fleming, 1979).
- P-values computed, according to trial hypothesis, by

$$P_k(\theta) = P\{Z_k \geq z_k\} = 1 - \Phi\left(\frac{Z_k - \theta V_k}{\sqrt{V_k}}\right),$$

are compared with a suitable sequence $(\alpha_1, \alpha_2, \ldots, \alpha_K)$ of nominal significance levels, chosen to control the type-I error probability.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Interim Analyses

- The trial is not stopped until the null hypothesis continues to be not rejected.
- Error spent (π_k), where $\pi_1 + \cdots + \pi_K = \alpha$, is the probability of stopping at stage k and to reject the null hypothesis when the null hypothesis is true.
- Power achieved $(1 - \beta_k)$, where $\sum_k (1 - \beta_k) = 1 - \beta$, is the probability of stopping at stage k and to reject the null hypothesis when the null hypothesis is false.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Including Overrunning Data Methods

- Overrunning data collected according to the trial protocol are considered valid and should be included in the analyses (CPMP/EWP/2459/02 London: EMEA, 2007; Sooriyarachchi et al. 2003).
- Results and conclusions could be affected by overrunning data.
- Many proposals to incorporate overrunning data were presented as direct extensions of methods of analyzing data from a sequential trial without overrunning.
  - Deletion Methods (Whitehead, 1992).
  - Combining p-values (Hall & Ding, 2001).
  - Repeated Confidence Intervals (Jennison & Turnbull, 1989).

UNIVERSITÀ DEGLI STUDI DI PADOVA

Including Overrunning Data Methods

- The stopping criterion is reached at the k-th IA.
- The sequential part of the trial is represented by the statistics (Z_k, V_k).
- The analysis that takes into account of the overrunning data could be considered as the (k+1)-th IA with related statistics Z_{k+1} and V_{k+1}.
- The contribution of the overrunning data to the analysis could be considered by setting $Z_O = Z_{k+1} - Z_k$ and $V_O = V_{k+1} - V_k$.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Deletion Method

(Whitehead, 1992)

- The deletion method includes essentially overrunning data.
- The IA that led to the stopping criterion is ignored.
- The conclusions are based on Z_{k+1} and V_{k+1}

UNIVERSITÀ DEGLI STUDI DI PADOVA

Combining p-values

(Hall and Ding, 2001)

- Two different analyses, for the same hypothesis, are performed
  - One based on the sequential portion of the data: $(Z_k; V_k) \rightarrow P_k(\theta)$.
  - One based on the overrunning part: $(Z_O; V_O) \rightarrow P_O(\theta)$.
- To combine them by weighting their p-values:

$$P(\theta) = 1 - \Phi[w_1 * \Phi^{-1}\{P_k(\theta)\} + w_2 * \Phi^{-1}\{P_O(\theta)\}],$$

Φ denotes the standard normal distribution function.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Combining p-values

(Hall and Ding, 2001)

- The weights, such that $w_1^2 + w_2^2 = 1$, could be chosen in different ways (Sooriyarachchi et al., 2003)

- Fixed:

$$w_1 = \sqrt{\frac{E[n_T; H_0]}{E[n_T; H_0] + E[n_O; H_0]}} \quad \text{and} \quad w_2 = \sqrt{\frac{E[n_O; H_0]}{E[n_T; H_0] + E[n_O; H_0]}};$$

- Random:

$$w_1' = \sqrt{\frac{V_T}{V_T + V_O}} \quad \text{and} \quad w_2' = \sqrt{\frac{V_O}{V_T + V_O}};$$

UNIVERSITÀ DEGLI STUDI DI PADOVA

Repeated Confidence Intervals (Jennison and Turnbull, 1989)

- The method leads to a sequence of confidence intervals $\{I_k : k = 1, \ldots, K\}$ for the parameter θ, built on available information at analysis k and such that

$$P_\theta\{\theta \in I_k:\ k = 1, \ldots, K\} = 1 - \alpha.$$

- The repeated confidence intervals are obtained by inverting a family of group sequential tests

$$I_k(\theta): \qquad \frac{Z_k}{V_k} \pm \frac{c_k(V_k)}{\sqrt{V_k}}; \qquad k = 1, \ldots, K$$

- The critical values $c_k(V_k)$ depend on the form of the test used (Pocock, 1977; O'Brien and Fleming, 1979).

UNIVERSITÀ DEGLI STUDI DI PADOVA

Repeated Confidence Intervals (Jennison and Turnbull, 1989)

- Stopping criterion stops the trial at the IA in which $\theta_{H_0} \notin I_k(\theta)$.
- RCIs method needs no adjustment for overrunning.
- The confidence interval is recomputed including the overrunning data.
- RCIs method is flexible to hypothesis changes than no impact on the confidence interval bounds.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Simulation studies (keypoints)

- A superiority and a non-inferiority real trials are used as bases for simulation studies.
- The primary endpoints are event rates.
- θ is log odds-ratio.
- O'Brien and Fleming design with three IAs is adopted.
- 100,000 full trials are simulated under a null (H_0) and an alternative (H_1) hypotheses.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Simulation studies
(keypoints)

- 1,000 smaller (with sample size equal to the dimension of an interim analysis) trials are simulated as overrunning cases.
- Increasing portions of overrunning data are included in the trials that should be actually stopped at the first and at the second interim analysis.
- Results refer to the average rate, on the 1,000 overrunning cases, of simulated trials that confirm the conclusions at the first or at the second IA to stop the trial.

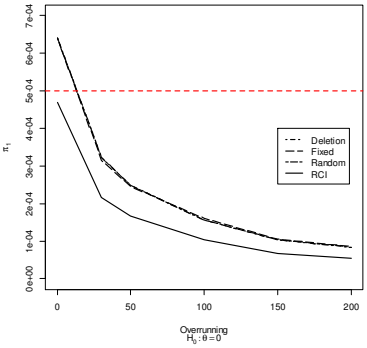
UNIVERSITÀ DEGLI STUDI DI PADOVA

Superiority trial

- Based on the ASCLEPIOS study (Whitehead, 1993).
- Superiority of an experimental calcium channel blocker with a placebo control in the immediate treatment of patients accusing an acute ischemic stroke.
- The death rate is the primary endpoint.
- Trial design:
  - $p_C = 0.15$, $p_E = 0.09$;
  - power= 90%;
  - 2.5% one-sided significance level.
- Sample size of 1248 (624 in each treatment group, 416 for IA stage)
- H_0: $\theta = 0$ and H_1: $\theta = 0.58$.
- $n_O = (30,50,100,150,200)$ responses for treatment arm.

Superiority: First interim results

Superiority: First Interim



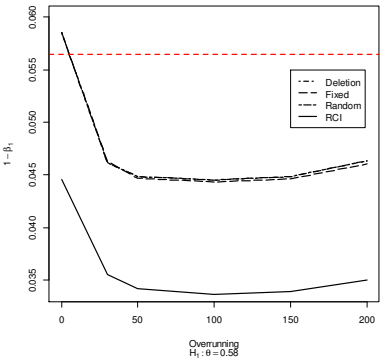
	$H_0: \theta = 0$					
	Over 0	**Over 30**	**Over 50**	**Over 100**	**Over 150**	**Over 200**
Deletion	0,000640	0,000322	0,000248	0,000157	0,000104	0,000084
Fixed W.	0,000640	0,000315	0,000245	0,000160	0,000105	0,000085
Random W.	0,000640	0,000322	0,000248	0,000157	0,000104	0,000084
RCI	0,000470	0,000216	0,000166	0,000104	0,000067	0,000054

- Overrunning data reduce error-spent levels.

- Error-spent levels decrease when the size of the overrunning increases.
- Deletion and combining p-value methods are substantially equivalent.
- RCIs method seems the most conservative.

Superiority: First interim results

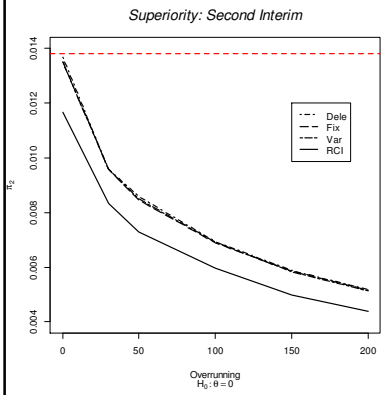
Superiority: First Interim



	$H_1: \theta = 0.58$					
	Over 0	**Over 30**	**Over 50**	**Over 100**	**Over 150**	**Over 200**
Deletion	0,05851	0,04619	0,04487	0,04451	0,04487	0,04633
Fixed W.	0,05851	0,04628	0,04467	0,04431	0,04461	0,04606
Random W.	0,05851	0,04619	0,04487	0,04451	0,04487	0,04633
RCI	0,04458	0,03550	0,03421	0,03365	0,03391	0,03500

- Overrunning reduces the levels of power achieved.
- For large overrunning sizes the methods seem to start to recuperate power.
- RCIs method has a power-achieved level lower than the value planned by O'Brien and Fleming (0,0565)

Superiority: Second interim results

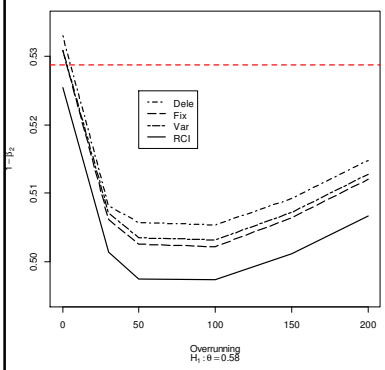


	H_0: $\theta = 0$					
	Over 0	Over 30	Over 50	Over 100	Over 150	Over 200
Deletion	0,01367	0,00958	0,00860	0,00694	0,00588	0,00517
Fixed W.	0,01351	0,00957	0,00847	0,00690	0,00585	0,00514
Random W.	0,01351	0,00959	0,00850	0,00691	0,00584	0,00514
RCI	0,01167	0,00833	0,00729	0,00597	0,00498	0,00438

- The value planned by O'Brien and Fleming design is 0.0138.
- Overrunning data reduce error-spent levels.

- Error-spent levels decrease when the size of the overrunning increases.
- Deletion and combining p-value methods are substantially equivalent.
- RCIs method seems the most conservative.

Superiority: Second interim results



	H_1: $\theta = 0.58$					
	Over 0	Over 30	Over 50	Over 100	Over 150	Over 200
Deletion	0,53304	0,50815	0,50569	0,50535	0,50918	0,51474
Fixed W.	0,53091	0,50611	0,50250	0,50213	0,50641	0,51204
Random W.	0,53091	0,50709	0,50347	0,50314	0,50719	0,51278
RCI	0,52543	0,50142	0,49745	0,49735	0,50111	0,50665

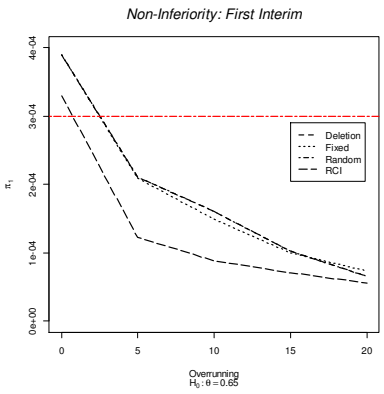
- The value planned by O'Brien and Fleming design is 0.5288.
- Overrunning reduces the levels of power achieved.

- Initial power level reductions are recuperated for high values of overrunning size.
- Random-weights method lies between the methods fixed-weight and deletion, that seems the most powerful.
- RCIs method is the less powerful.

Non-Inferiority trial

- Multicenter phase III trial for non-inferiority of a Test drug compared to a Reference drug.
- Trial design:
  - $p_C = 0.45$, $p_E = 0.5$;
  - Non-inferiority margin of 0.15;
  - power= 80%;
  - 2.5% one-sided significance level.
- Sample size of 198 (99 in each treatment group, 66 for IA stage).
- H_0: $\theta = 0.65$ and H_1: $\theta = -0.20$.
- n_O = (5,10,15,20) responses for treatment arm.

Non-Inferiority: First interim results

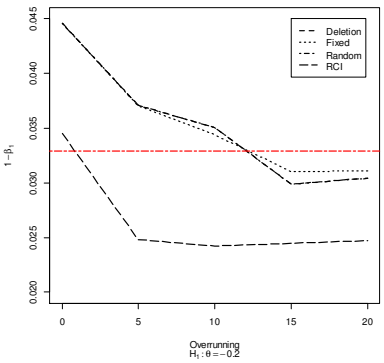


	H_0: $\theta = 0.65$				
	Over0	Over5	Over10	Over15	Over20
Deletion	0.000390	0.000211	0.000161	0.000103	0.000066
Fixed W.	0.000390	0.000209	0.000149	0.000100	0.000074
Random W.	0.000390	0.000211	0.000161	0.000103	0.000066
RCI	0.000330	0.000123	0.000088	0.000071	0.000056

- Overrunning data reduce error-spent levels.
- Error-spent levels decrease when the size of the overrunning increases.
- Deletion and combining p-value methods have a high agreement.
- RCIs method is again the most conservative.

Non-Inferiority: First interim results

Non-Inferiority: First Interim



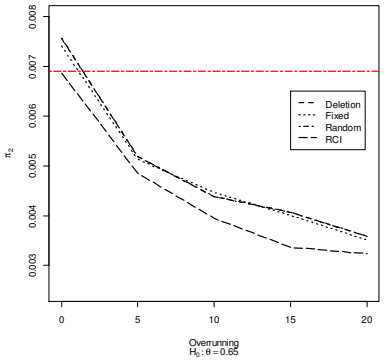
	$H_1: \theta = -0.20$				
	Over0	Over5	Over10	Over15	Over20
Deletion	0.04457	0.03708	0.03507	0.02990	0.03043
Fixed W.	0.04457	0.03703	0.03438	0.03102	0.03107
Random W.	0.04457	0.03708	0.03507	0.02990	0.03043
RCI	0.03451	0.02482	0.02420	0.02448	0.02473

- Overrunning reduces levels of power achieved.
- Observed power-achieved level is bigger than planned O'Brien and Fleming value (0,0329)

- Deletion and combining p-values decrease due to overrunning size, approaching the planned O'Brien and Fleming value.
- RCIs method seems to be not affected by overrunning size.
- The fixed-weights method seems slightly more powerful.

Non-Inferiority: Second interim results

Non-Inferiority: Second Interim



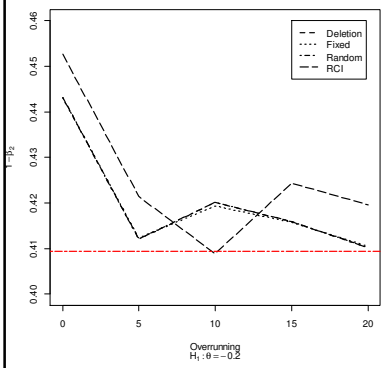
	$H_0: \theta = 0.65$				
	Over0	Over5	Over10	Over15	Over20
Deletion	0.0076	0.0052	0.0044	0.0041	0.0036
Fixed W.	0.0074	0.0051	0.0045	0.0040	0.0035
Random W.	0.0076	0.0052	0.0044	0.0041	0.0036
RCI	0.0069	0.0049	0.0039	0.0034	0.0032

- Method error-spent levels are close to the O'Brien and Fleming planned value (0,0069)
- Overrunning data reduce error-spent levels.

- Error-spent levels decrease when the size of the overrunning increases.
- Deletion and combining p-value methods have a high agreement.
- RCIs method is again the most conservative.

Non-Inferiority: Second interim results

Non-Inferiority: Second Interim



	$H_1: \theta = -0.20$				
	Over0	Over5	Over10	Over15	Over20
Deletion	0.4431	0.4120	0.4201	0.4158	0.4102
Fixed W.	0.4431	0.4124	0.4194	0.4158	0.4104
Random W.	0.4431	0.4120	0.4201	0.4158	0.4102
RCI	0.4527	0.4215	0.4090	0.4244	0.4196

- The value planned by O'Brien and Fleming design for the power-achieved level is 0.4095.
- Overrunning reduces levels of power achieved.

- The overrunning size effect is not clear.
- The values oscillate around the O'Brien and Fleming one.

Conclusions and Remarks

- Overrunning reduces both the type-I error and the power levels.
- For modest overrunning sizes, a mean reduction of 30-50% on the type-I error levels with respect to O'Brien and Fleming planned values is observed.
- Mean power reductions are <10% with respect to O'Brien and Fleming planned values.
- The observed type-I error and power reductions could be explained by the high conservative values of the O'Brien and Fleming bounds for the stopping criterions.

UNIVERSITÀ DEGLI STUDI DI PADOVA

Conclusions and Remarks

- The high agreement between deletion and combining p-values methods (Sooriyarachchi et al., 2003) was confirmed.
- RCIs method has proved highly conservative.
- Method choice should be oriented by the endpoint type (for safety rather than efficacy) or design type (superiority rather than non-inferiority).
- RCIs method remains the most appealing approach due to its flexibility at hypotheses switching.
- RCIs method could be used if the main hypotheses of the study is not so clear and a switching between superiority to non-inferiority (Lewis,2001) was already contemplated.

UNIVERSITÀ DEGLI STUDI DI PADOVA

References

Fleming, T. R., Harrington, D. P., & O'Brien, P. C. (1984). Designs for group sequential tests. Controlled clinical trials, 5(4), 348-361.

Group, I. E. E. W. (1999). Statistical principles for clinical trials: ICH harmonised tripartite guideline. Statistics in Medicine, 18(15), 1903-1942.

Hall, W., & Ding, K. (2008). Sequential tests and estimates after overrunning based on p-value combination. Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh, 3, 33-45.

Jennison, C., & Turnbull, B. W. (1989). Interim analyses: the repeated confidence interval approach. Journal of the Royal Statistical Society. Series B (Methodological), 305-361.

Jennison, C., & Turnbull, B. W. (1991). Group Sequential Tests and Repeated Confidence Intervals. Handbook of sequential analysis, 118, 283.

Lewis, J. A. (2001). Switching between superiority and non-inferiority: an introductory note. British journal of clinical pharmacology, 52(3), 221-221.

Mahoney, F. I. (1965). Functional evaluation: the Barthel index. Maryland state medical journal, 14, 61-65.

O'Brien, P. C., & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. Biometrics, 549-556.

UNIVERSITÀ DEGLI STUDI DI PADOVA

References

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. Biometrika, 64(2), 191-199.

Sooriyarachchi, M. R., Whitehead, J., Bolland, K., & Whitehead, A. (2003). Incorporating data received after a sequential trial has stopped into the final analysis: implementation and comparison of methods. Biometrics, 59(3), 701-709.

Whitehead, J. (1992). Overrunning and underrunning in sequential clinical trials. Control Clin Trials, 13(2), 106-121.

Whitehead, J. (1993). Application of sequential methods to a phase III clinical trial in stroke. Drug Information Journal, 27(3), 733-740.

Whitehead, J. (1997). The design and analysis of sequential clinical trials: John Wiley & Sons.

Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design (CPMP/EWP/2459/02). London: EMEA; 18 October 2007.

UNIVERSITÀ DEGLI STUDI DI PADOVA

THANK YOU FOR YOUR ATTENTION