NTNU
Det skapende universitet

Interim analysis and early termination of trials. Group sequential design.

17 February 2010
Stian Lydersen,
Unit for Applied Clinical Research

Example: Postoperative nausea

Treatment * Nausea Crosstabulation

			Nausea		
			none or little	much	Total
Ventricular tube	no	Count	18	12	30
		% within Treatment: ventricular tube	60.0%	40.0%	100.0%
	yes	Count	24	5	29
		% within Treatment: ventricular tube	82.8%	17.2%	100.0%
	Total	Count	42	17	59
		% within Treatment: ventricular tube	71.2%	28.8%	100.0%

Difference in success probabilities:
Estimate: 82.8% - 60% = 22.8%
95% Newcombe confidence interval: -0.4% to 42.9%
(Wald method, not recommended: 0.5% to 45.0%)

Pearson's chi squared p-value= 0.054

NTNU
Det skapende universitet

References

- Proschan, M. A., Lan, K. K., Wittes, J. T. (2006): "Statistical Monitoring of Clinical Trials: A Unified Approach" Springer *)
- Jennison C and Turnbull, B W (2000): "Group Sequential Methods: Applications to Clinical Trials" Chapman & hall *)
- Mazumdar M and Bang H (2008): "Sequential and group Sequential Designs in Clinical Trials: Guidelines for Practitioners". Chapter 16 (pages 491-512) in Rao , Miller and Rao: "Handbook of Statistics Vol 27: Epidemiology and Medical Statistics" http://folk.ntnu.no/slyderse/medstat/Mazumdar_Bang.pdf
- Armitage P, Berry, G, Matthews, J N S (2002): "Statistical methods in medical research". 4th ed. Section 18.7 Data Monitoring (page 613-623). http://folk.ntnu.no/slyderse/medstat/Armitage_et_al.pdf **)
- International Committee on Harmonization ICH E9 (1998): Statistical principles for Clinical Trials. www.ich.org

*) Available as E-book at UBIT
**) Curriculum Course st2303 "Medical Statistics" spring 2010

NTNU
Det skapende universitet

Why interim analyses in an RCT?

- Early termination if treatment is superior to control
- Early termination if treatment is more harmful than control

But:

- Interim analyses HAS implications for study design and analysis and interpretation of results

NTNU
Det skapende universitet

Monitoring

- Administrative monitoring: Normally makes no use of outcome data from the trial.
- Data monitoring: Concerns evidence emerging from the accumulating data on safety and efficacy of the treatment.
- Data (and Safety) Monitoring Committee D(S)MC. Regularly receives unmasked data summaries. Present recommendation for or against early termination or protocol modification.

NTNU
Det skapende universitet

A trial with planned consecutive inclusion of n subjects.
At any interim time, a z-score test statistic can be calculated.
Under H0, the z-score is N(0,1).

Group sequential trial:
Look at data k times including final look after n subjects.
Possibly terminate before all n subjects are included.
k=1: means no interim analyses
k=n: means fully sequential trial

NTNU
Det skapende universitet

7

Group sequential designs for interim analyses. Alternative procedures

- Naïve (NOT appropriate)
- Pocock procedure
- Haybrittle-Peto
- O'Brien-Fleming
- Alpha spending function

NTNU Det skapende universitet

8

Naïve approach:
At any look, reject H0 and terminate if
$p-value \le \alpha$, that is, if $|z-score| > z_{\alpha/2}$.

But:
The significance level is seriously inflated.
k=2 looks, $\alpha = 0.05$, equally spaced looks (worst case)
Type I error rate 0.083 (0.098)
k=5 looks, $\alpha = 0.05$:
Type I error rate 0.142 (0.226)
(Proschan et al Table 4.1 page 68)

NTNU Det skapende universitet

9

Pocock (1977) procedure:
At each of k equally spaced looks, use a lowered significance level $\alpha_{lowered}$ to give a type I error rate = α as planned.
k=2 looks, $\alpha = 0.05$
Reject H0 and terminate if $|z| > 2.178$ (not 1.96). $\alpha_{lowered} = 0.029$
k=5 looks, $\alpha = 0.05$
Reject H0 and terminate if $|z| > 2.413$ (not 1.96). $\alpha_{lowered} = 0.016$
(Proschan et al Table 4.2 page 70)

Drawback:
Spending much of α early. Only 0.016 left for the final analysis.
Interpretation of result if final z-score is between 1.96 and 2.41?

NTNU Det skapende universitet

10

Haybittle-Peto (1971, 1976) procedure
Use a very strict criterion at the first k-1 looks.
K=5 looks. $\alpha = 0.05$
Reject H0 and terminate at the first 4 looks
using $\alpha_{lowered} = 0.001$ or $|z| > 3.29$
Reject H0 at final look using
$\alpha_{lowered} = 0.05 - 4 \times 0.001 = 0.046$ (Bonferroni fix)

NTNU Det skapende universitet

11

Drawback: Logical inconsistency.

Example: 5 equally spaced looks
z-score = 2.8 at 4th look. Not reject H0.
incremental z-score=-1 from 4th to 5th look
(evidence in opposite direction)
Final z-score: $(4/5)^{1/2} 2.8 + (1/5)^{1/2}(-1) = 2.06$
Reject H0 at the end!

NTNU Det skapende universitet

12

OBF (O'Brien-Fleming, 1979)
U(t) is the "z-score" from subject number t alone.
B(t) =U(1) + …+ U(t)
$Z(t) = B(t)/\sqrt{t}$ is the z-score after subject t
The Pocock procedure uses constant boundary for Z(t)
The OBF procedure uses constant boundary for B(t)
(Proschan et al Table 4.3 Page 72)

NTNU Det skapende universitet

13

OBF boundaries with k=5:

Look no	Boundary for B(t):	Boundary for Z(t):	
t	a(t)	$a(t)/(t/k)^{1/2}$	$\alpha_{lowered}$
1	2.040	4.562	0.0000051
2	2.040	3.226	0.0013
3	2.040	2.634	0.0085
4	2.040	2.281	0.023
5	2.040	2.040	0.041

NTNU Det skapende universitet

14

Example (Armitage et al)
RCT with 2 parallell groups, $\alpha = 0.05$, power = 0.80

Sample size (without interim looks) to detect an effect size (standardized difference) of 0.5:
n=126 (63 per group)

Alternative sequential designs with 5 equally spaced looks:
Inflation factor 1.23 and 1.03 for Pocock and OBF procedure (Mazumadar and Bang page 497)
$n_{Pocock} = 126 \times 1.23 = 155$ and $n_{OBF} = 126 \times 1.03 = 130$

Armitage et al, Table 18.4 and Figure 18.1 page 619-620

NTNU Det skapende universitet

15

Alpha spending function

- Controls how much of alpha can be used at each look, as function of the proportion of total information observed.
- This proportion may be estimated as fraction of
  - subjects recruited
  - events observed
- Number of looks, timing of looks, need NOT to be pre-specified.
- The alpha spending function must be pre-specified (for example Pocock or OBF)
- Prochan et al Table 5.1 and Figure 5.1 page 81-82, Figure 5.3 and Table 5.3 page 86-87

NTNU Det skapende universitet

16

Data-driven looks:

- Violates assumptions for the alpha spending function
- But results are approximately unaffected. Proschan et al page 89-90: “Intention to cheat” results in max 10% inflation of type I error rate.

NTNU Det skapende universitet

17

Analysis after a sequential trial

- Two situations:
  - After completion of trial
  - At an interim analysis
- In both situations, naïve anayses (as if data were from a fixed sample experiment) are inappropriate (see i.e. Prochan et al 2006 Chapter 7)
  - Effect size estimates and CI are biased away from 0
  - Actual CI coverage substantially lower than nominal coverage.
  - P-values are too small
- ”Most statisticians acknowledge that the observed effect from a trial that is stopped early overestimates the true value, but may recommend using the observed estimate for simplicity” (Proschan et al, page 114)

NTNU Det skapende universitet

18

Stochastic curtailment

- Early termination if it can be predicted that the final difference would almost certainly be non-significant.
- Armitage et al page 622: “Although this approcah may be useful in enabling research efforts to be switched into more promising directions, there is a danger in placing too much importance on the predicted results of a final significance test. Data showing non-significant treatment effects may nevertheless be valuable for estimation, especially in contributing to meta-analyses. It may be unwise to terminate such studies prematurely, particularly when there is no treatment difference to provide ethical reasons for stopping.”

NTNU Det skapende universitet

19

Adaptive designs

- Allows to change sample size based on accumulated data
- Two main types:
  - Using data for nuisance parameter(s) only, for example variance in a t-test.
  - Also using data for effect size

20

Software
(Proschan et al 2006, Mazumdar and Bang, 2008):

- Commercial packages:
  - East (Cytel Software). *)
  - PEST (University of Reading)
  - S-plus: SeqTrial (Insightful corporation)
  - SAS: IML module
  - PASS (Number Chruncher Statistical Software, Ogden, Utah)
- Free software
  - www.medsch.wisc.edu/landemets/
  - R: Function seqmon

*) Most comprehensive (Mazumdar and Bang, 2008).

21

Summary:

- Naïve
- Pocock procedure
- Haybrittle-Peto
- O'Brien-Fleming
- Alpha spending function
- Analysis after a sequential trial
- Stochastic curtailment