Trial Sequential Analysis

Yu-Kang Tu

Institute of Health Data Analytics & Statistics, College of Public Health,
National Taiwan University, Taipei, Taiwan

Outline

- Interim analysis in clinical trials
- Alpha-spending functions
- Trial sequential analysis
- Example for TSA

Hypothesis Testing

- For a clinical trial with two groups, under

$$H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$$

- Compute Z statistic :

$$Z = \frac{d}{se} \sim N(0,1)$$

- d: difference in the effects, and se: standard error
- If $Z < -1.96$ or $Z > 1.96$, reject the null hypothesis and conclude that there is a statistically significant difference between the two groups at a Type-1 error of 5% ($\alpha = 0.05$).

Interim Analyses

- Planning statistical analyses during the collection of trial data.
- Stop the trial early if:
  - The test treatment is extremely effective
  - The test treatment is unlikely to be better than the control
  - The test treatment shows unacceptable side effects
- Repeated testing increases the Type-1 error, resulting in a higher chance of a false positive finding
- The significant level needs to be more conservative to control the inflated Type-1 error rate

Type-1 Error Rate Increases with Repeated Significance Tests

Repeated significant tests at 5% level	Overall type-1 error rate
1	0.05
2	0.08
3	0.11
4	0.13
5	0.14
10	0.19
20	0.25
50	0.32
100	0.37
1000	0.53
∞	1

Sequential Analysis for Monitoring a Clinical Trial

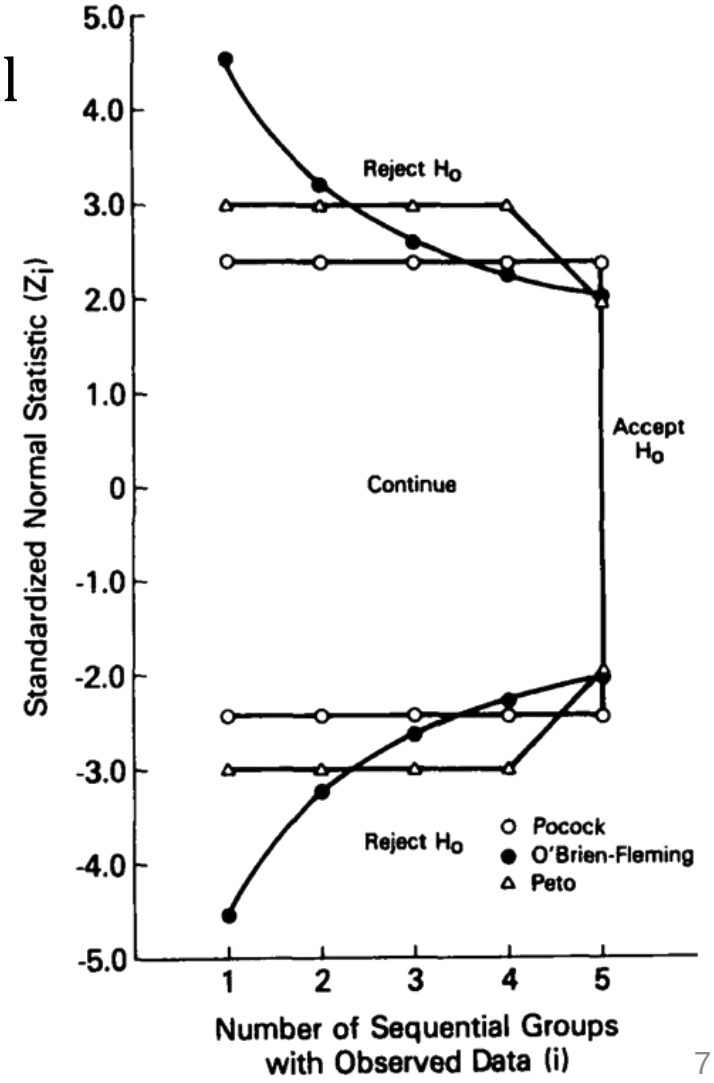
Sequential analysis to control Type-I error rate:

1. Compute Z statistic at each interim analysis when results from additional groups of patients are available
2. Compare Z statistic to a more conservative critical value (>1.96) to keep an overall Type-1 error probability close to 5%
3. Efficacy (or stopping) boundaries can then be calculated
4. When Z statistic crosses the boundaries, make a decision to stop the trial

Stopping Boundaries

For a trial with in total 5 sequential analyses:

- Haybittle-Peto
  - All interim analysis: $Z_c = \pm 3$, the last: $Z_c = \pm 1.96$
- Pocock
  - $Z_c = \pm 2.41$
- O'Brien-Fleming
  - Z_c: from strict to loose



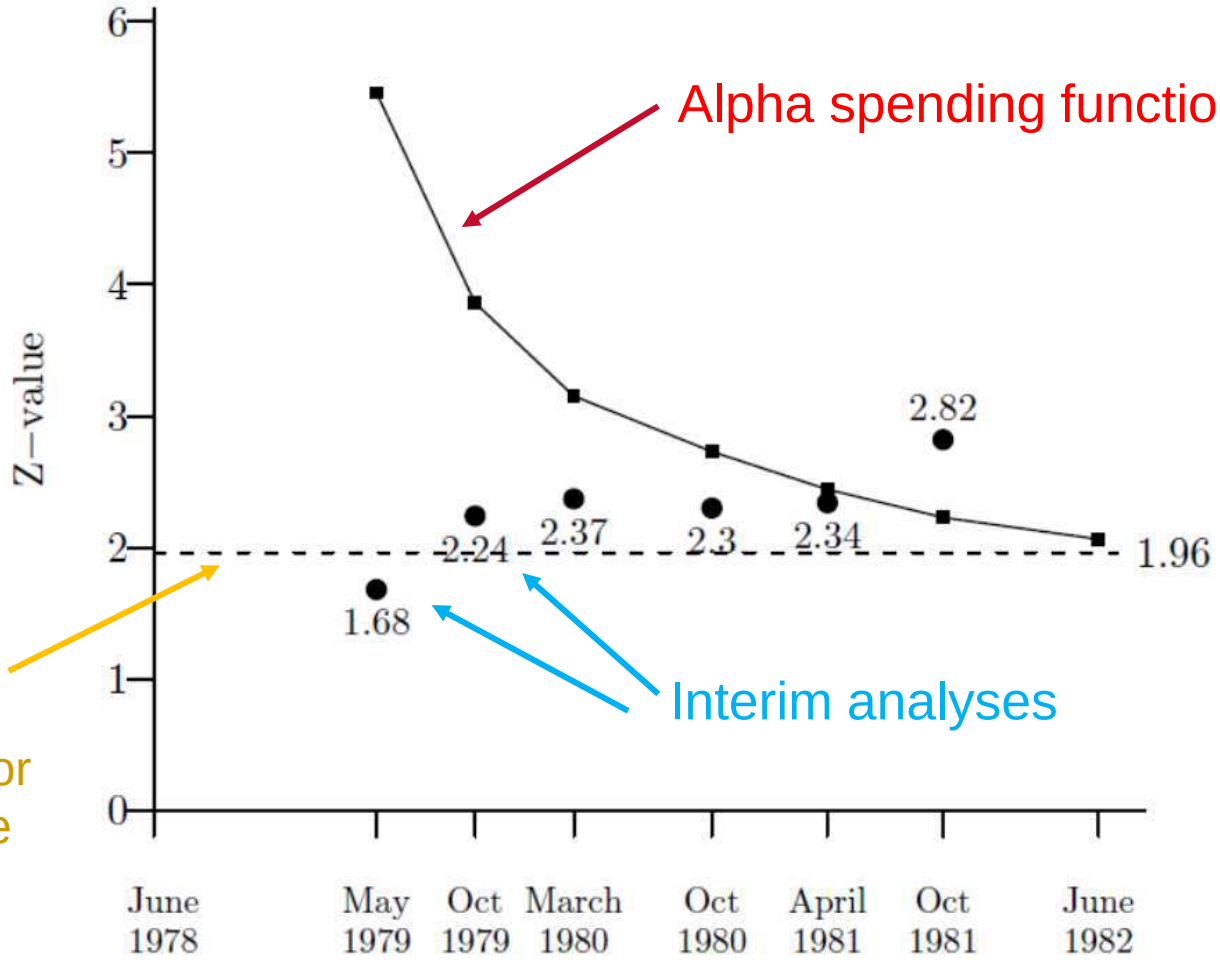
Alpha Spending Functions (DeMets & Lan 1994)

- More flexible than previous methods
- Define a function to **spend** the overall nominal significance level, e.g. 5%
- The spending function $\alpha(t)$ is an increasing function of information fraction t

$$\alpha(t) = \begin{cases} 0, if\ t = 0 \\ \alpha, if\ t = 1 \end{cases}$$

- Specify the spending function in advance

Alpha Spending Function



Monitoring and Updating a Meta-analysis

- A meta-analysis is updated when new trials are available.
  - e.g. Cumulative meta-analysis, living systematic review
- Repeated analyses inflate the Type-I error rate, leading to a premature conclusion
- No further studies are required when sufficient evidence shows the treatment to be effective or harmful
- Determine whether evidence is sufficient to show that the treatment is unlikely to be effective

Trial Sequential Analysis (Wetterslev 2007)

- Apply the concepts of efficacy/futility boundaries from a single randomized trial to a cumulative meta-analysis
- Trials are included in chronological order, and analyses are performed repeatedly after new trials are added.
- Estimate the required information size by assuming a meta-analysis is a large RCT
- Calculate the efficacy/futility boundaries to adjust the significance level to control the Type I and II errors.

Trial Sequential Analysis

- The required information size in meta-analysis should be **at least as large as the sample size in a single well-powered randomized trial**
  - The effect size d
  - Standard error of d
  - Type-1 error rate: α
  - Type-2 error rate: β
- The required sample size is usually **larger** than a single RCT due to the **heterogeneity** across trials

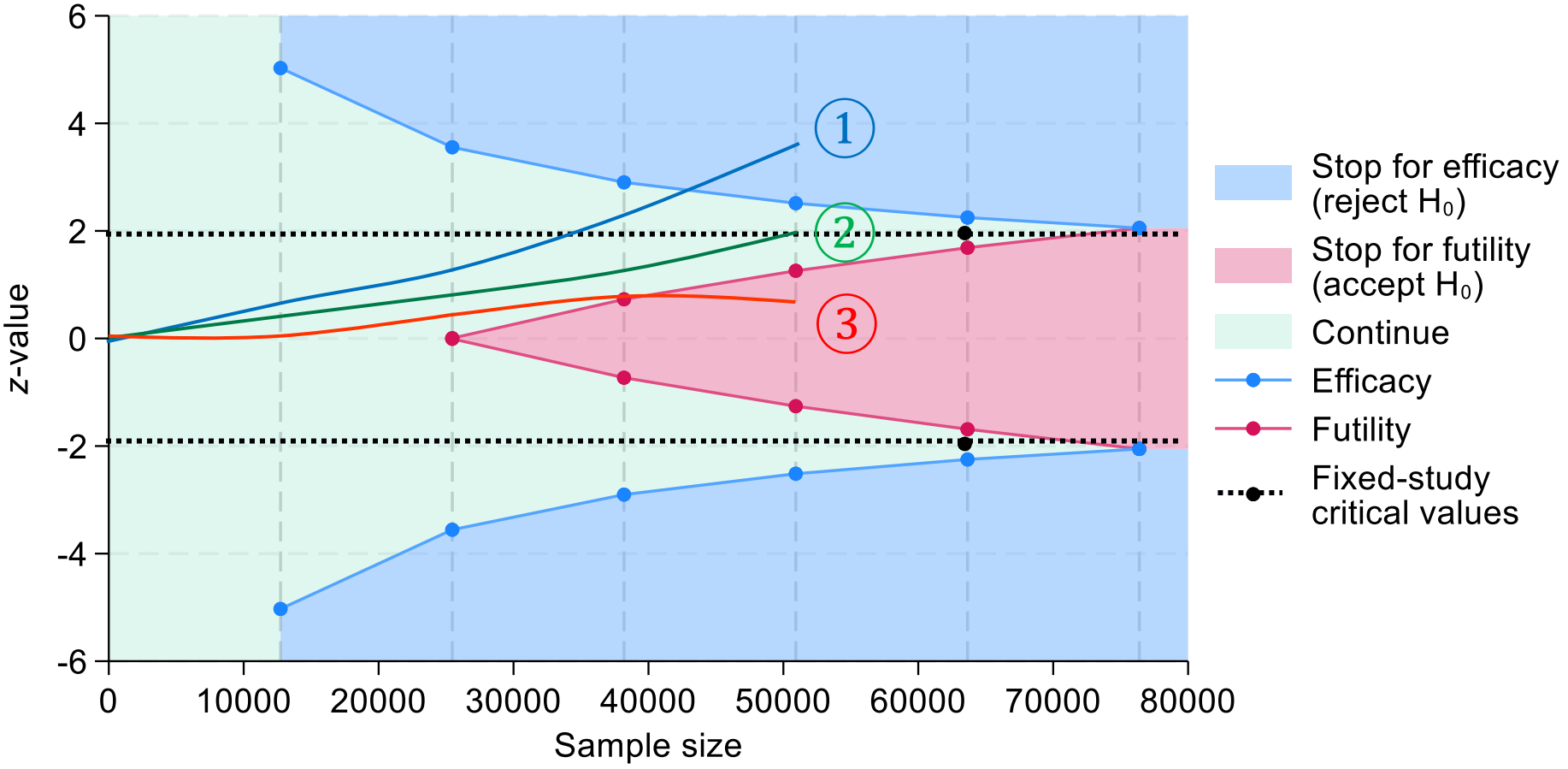
The Cumulative Test Statistic (Z-curve)

- Whenever a meta-analysis is updated, a new Z-value is calculated.

$$Z = \frac{d_{pooled}}{SE_{pooled}} \sim N(0,1)$$

- A series of Z-values from a series of meta-analysis updates are plotted against the accumulated information (usually the same size) to produce a *Z-curve*.

TSA


6
4
2
0
-2
-4
-6
z-value
0
10000
20000
30000
40000
50000
60000
70000
80000
Sample size
1
2
3
Stop for efficacy (reject H0)
Stop for futility (accept H0)
Continue
Efficacy
Futility
Fixed-study critical values

Hemmingsen 2011 BMJ

BMJ 2011;343:d6898 doi: 10.1136/bmj.d6898 (Published 24 November 2011)

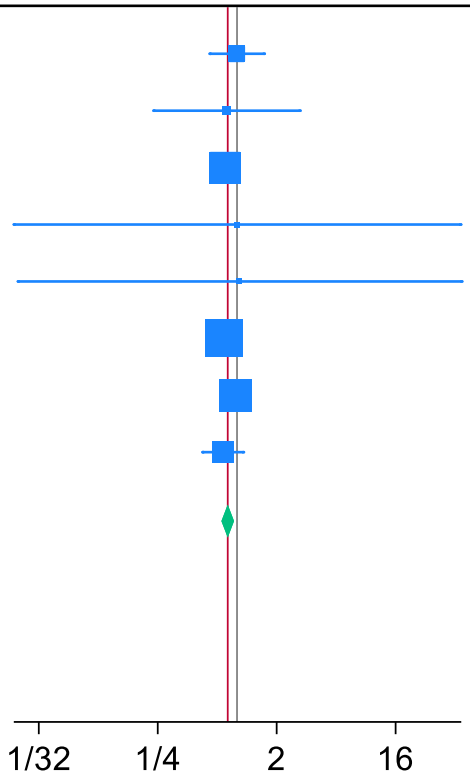
RESEARCH

Intensive glycaemic control for patients with type 2 diabetes: systematic review with meta-analysis and trial sequential analysis of randomised clinical trials

Bianca Hemmingsen *PhD student*[1], Søren S Lund *physician*[2], Christian Gluud *chief physician and head of department*[1], Allan Vaag *professor*[3], Thomas Almdal *chief physician and head of department*[2], Christina Hemmingsen *research assistant*[1], Jørn Wetterslev *chief physician*[1]

Intensive Control vs Conventional Control for Non-fatal Myocardial Infarction

Study	Year	Treatment Yes	Treatment No	Control Yes	Control No	Risk ratio with 95% CI	Weight (%)
UGDP	1978	29	175	30	180	1.00 [0.62, 1.60]	5.47
VA CSDM	1995	4	71	5	73	0.83 [0.23, 2.98]	0.75
UKPDS	1998	221	2,850	101	1,037	0.81 [0.65, 1.02]	24.08
Kumamoto	2000	0	55	0	55	1.00 [0.02, 49.52]	0.08
Bagg	2001	0	21	0	22	1.05 [0.02, 50.43]	0.08
ACCORD	2008	186	4,942	235	4,888	0.79 [0.65, 0.95]	34.45
ADVANCE	2008	153	5,418	156	5,413	0.98 [0.79, 1.22]	25.29
VADT	2009	51	841	66	833	0.78 [0.55, 1.11]	9.79
Overall						0.85 [0.76, 0.95]	



Heterogeneity: $\tau^2 = 0.00$, $I^2 = 0.00\%$, $H^2 = 1.00$

Test of $\theta_i = \theta_j$: Q(7) = 3.03, p = 0.88

Test of $\theta = 0$: z = -2.88, p = 0.00

Random-effects DerSimonian–Laird model
Sorted by: year

TSA: Required Information Size for RR = 0.9

Sample size calculation for a single trial

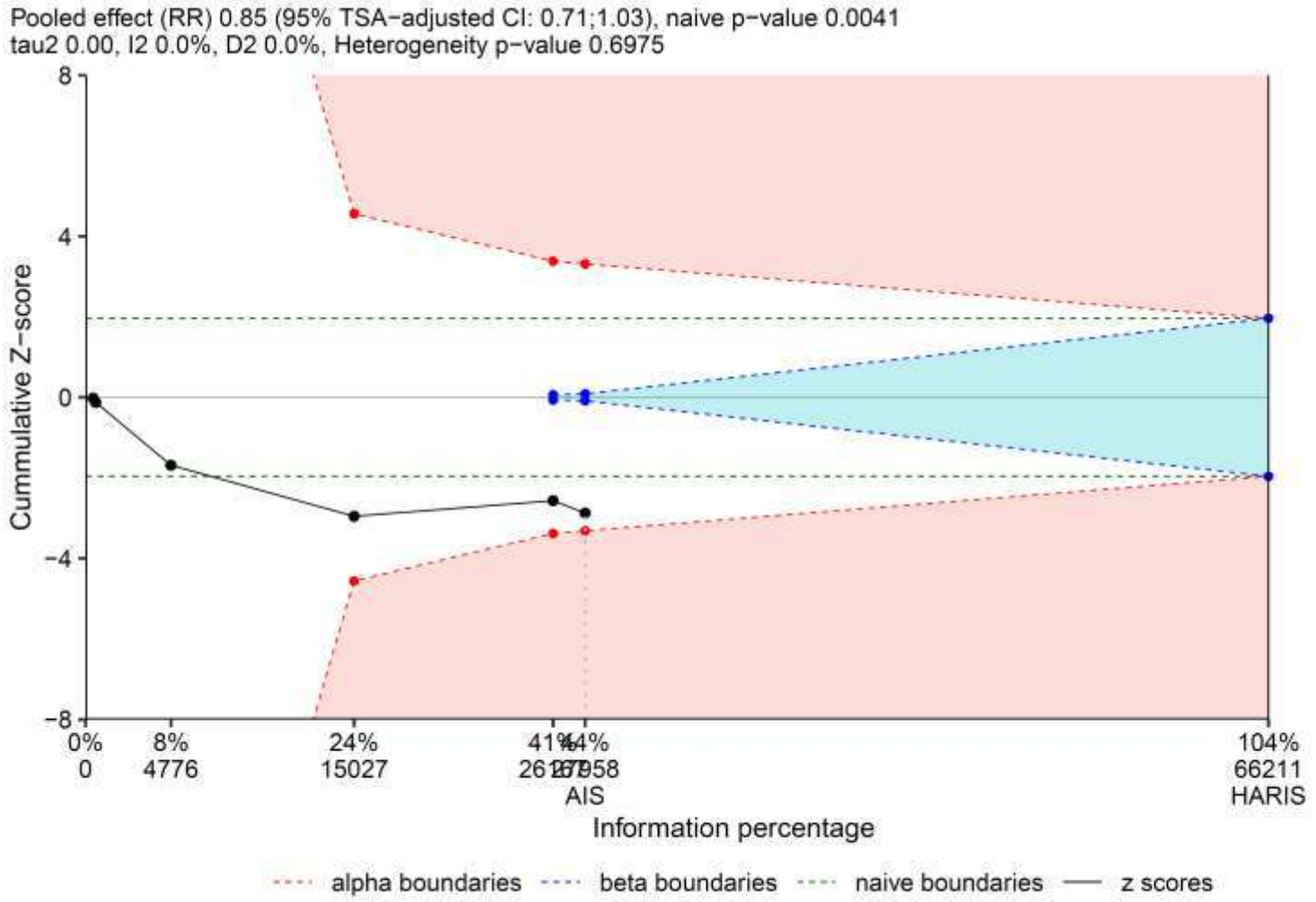
- The proportion of events is 4.5% in the control group.
- Assume a relative risk reduction of 10% in the intensive glycaemic control group
- Risk ratio = 0.9
- Type-1 error rate = 0.05
- Type-2 error rate = 0.2, so power = 0.8
- 1 to 1 randomization

Require sample size = 31722 for each group, 63444 in total.

Required Sample Size

- **AIS** (Achieved information size):27958
- **RIS** (Fixed-effect required information size for a non-sequential meta-analysis): 63446
- **SMA_RIS** (RIS adjusted for sequential analysis): 66211
- **HARIS** (Heterogeneity adjusted required information size for a non-sequential meta-analysis): 63446
  - Because of no heterogeneity, HARIS = RIS
- **SMA_HARIS** (HARIS adjusted for sequential analysis): 66211

TSA: RR = 0.9

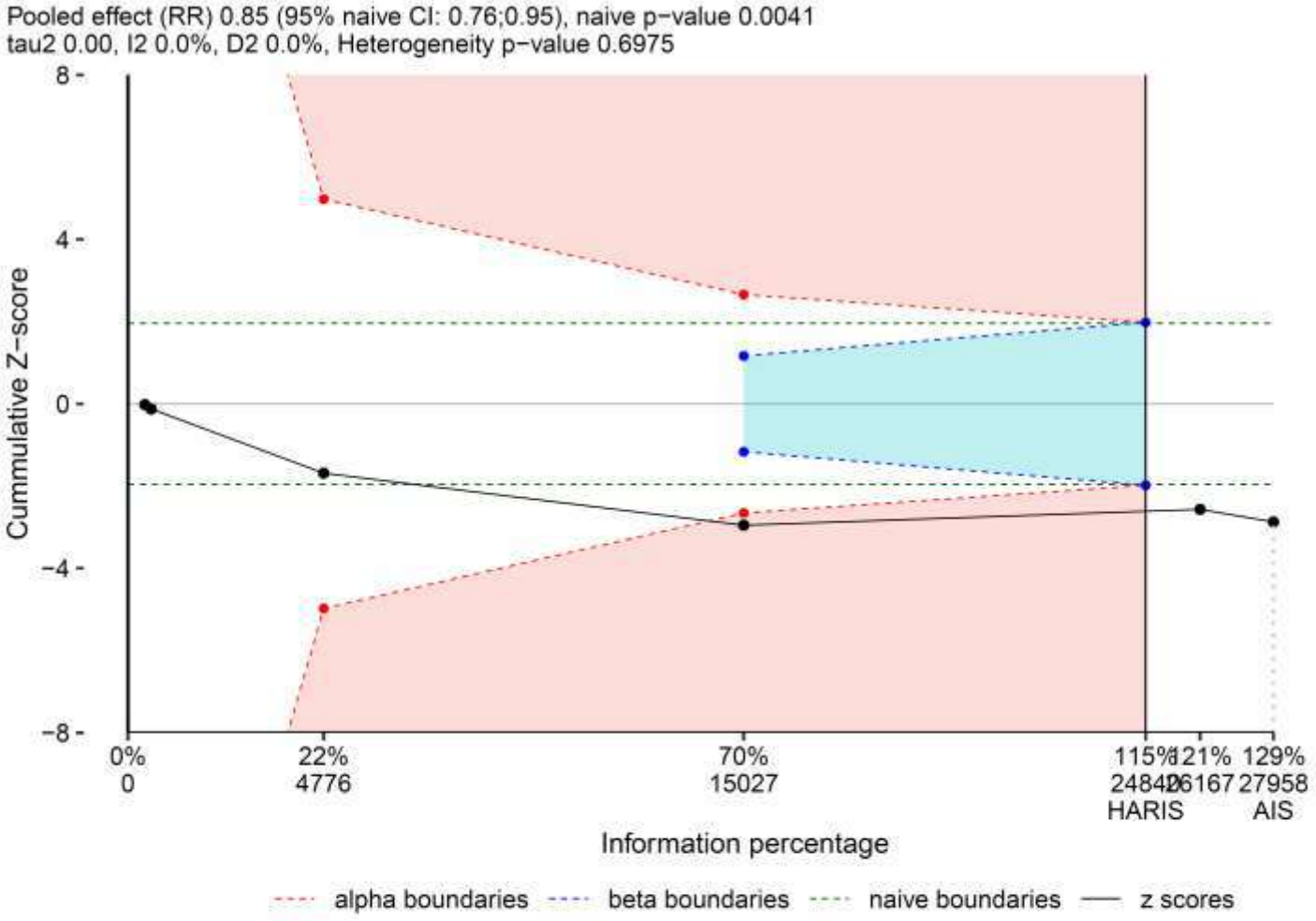


Retrospective TSA with: pc 4.5%, RRR 10.0%, alpha 5.0%, beta 20%. .
Methods: Fixed−effect, Weight MH, alpha spending esOF, futility is non−binding with beta spending esOF.

Interpretation

- Conventional meta-analysis showed a **significant** benefit of intensive glycaemic control (relative risk 0.85, 0.76 to 0.95; P=0.004).
- Trial sequential analysis showed **a lack of sufficient evidence** of a benefit of intensive glycaemic control for the reduction of non-fatal myocardial infarction (TSA adjusted 95% confidence interval 0.71 to 1.02).
- **Only 27958** (44%) of 63446 patients required to detect a 10% relative risk reduction for non-fatal myocardial infarction were accrued.

TSA: RR = 0.85



Retrospective TSA with: pc 4.5%, RRR 15.0%, alpha 5.0%, beta 30%. .
Methods: Fixed−effect, Weight MH, alpha spending esOF, futility is non−binding with beta spending esOF.

Interpretation

- Both conventional meta-analysis & TSA showed a **significant** benefit of intensive glycaemic control (relative risk 0.85, 0.76 to 0.95; P=0.004).
- The adjusted required information size of 24840 patients required to detect a 15% relative risk reduction for non-fatal myocardial infarction has been accrued.

Conflicting Results?

- The first TSA (RR=0.9) requires a larger number of patients than the second TSA (RR=0.85) because the expected difference between the two treatments in the first TSA is smaller.
- Therefore, the accrued evidence is considered inconclusive because the accrued sample size is much smaller than the required sample size.

Concluding Remarks

- Unlike a single trial, most meta-analyses do not prospectively collect data.
- Since a meta-analysis is conducted after all the data have been collected, it is debatable whether adjusting the Type-1 error rate is necessary.
- Efficacy & futility boundaries change when the parameters for the required information size change.
- The interpretation of results also changes when the boundaries change.


Thank You for Listening
國立臺灣大學公共衛生學院