Interim Analyses and Group Sequential Designs

Roger J. Lewis, MD, PhD
Department of Emergency Medicine
Harbor-UCLA Medical Center
David Geffen School of Medicine at UCLA
Berry Consultants, LLC

@RogerJLewis

Financial Disclosures

- Employee
  - County of Los Angeles, Department of Health Services, Harbor-UCLA Medical Center
  - David Geffen School of Medicine at UCLA
  - Los Angeles Biomedical Research Institute
  - Berry Consultants, LLC
- Special Government Employee
  - Food and Drug Administration/CBER
- Support from
  - National Institutes of Health/NINDS
  - National Institutes of Health/NHLBI
- Other consulting
  - Octapharma

Multiple Comparisons

- When two identical groups of patients are compared, there is a chance (α) that a statistically significant *p* value will be obtained (type I error)
- When multiple comparisons are performed, the risk of one or more false-positive p values is increased
- Multiple comparisons include:
  - Pair-wise comparisons of more than two groups
  - The comparison of multiple characteristics between two groups
  - The comparison of two groups at multiple times

Multiple Comparisons: Risk of $\geq$ 1 False Positive

Number of Comparisons	Probability of at Least One Type I Error
1	0.05
2	0.10
3	0.14
4	0.19
5	0.23
10	0.40
20	0.64
30	0.79

Assumes α= 0.05, uncorrelated comparisons

Bonferroni Correction

- A method for reducing the overall risk of a type I error when making multiple comparisons
- The overall (study-wise) type I error risk desired (e.g., 0.05) is divided by the number of tests, and this new value is used as the α for each individual test
- Controls the type I error risk, but reduces the power (increased type II error risk)

Interim Data Analysis (Success)

- During a clinical trial, data accumulate sequentially
- If you were the last patient to be enrolled, wouldn't you want to know the treatment assignments and outcomes of the prior patients?
- Interim analyses are used to see if a difference clearly exists between the two groups, so the trial can be stopped early, and future patients can receive the best treatment
- In other words, to stop the trial as soon as a reliable conclusion can be drawn from the available data


HIRAM S. DUDSON
1930 - 1993
Member,
Placebo Group
D. Reilly

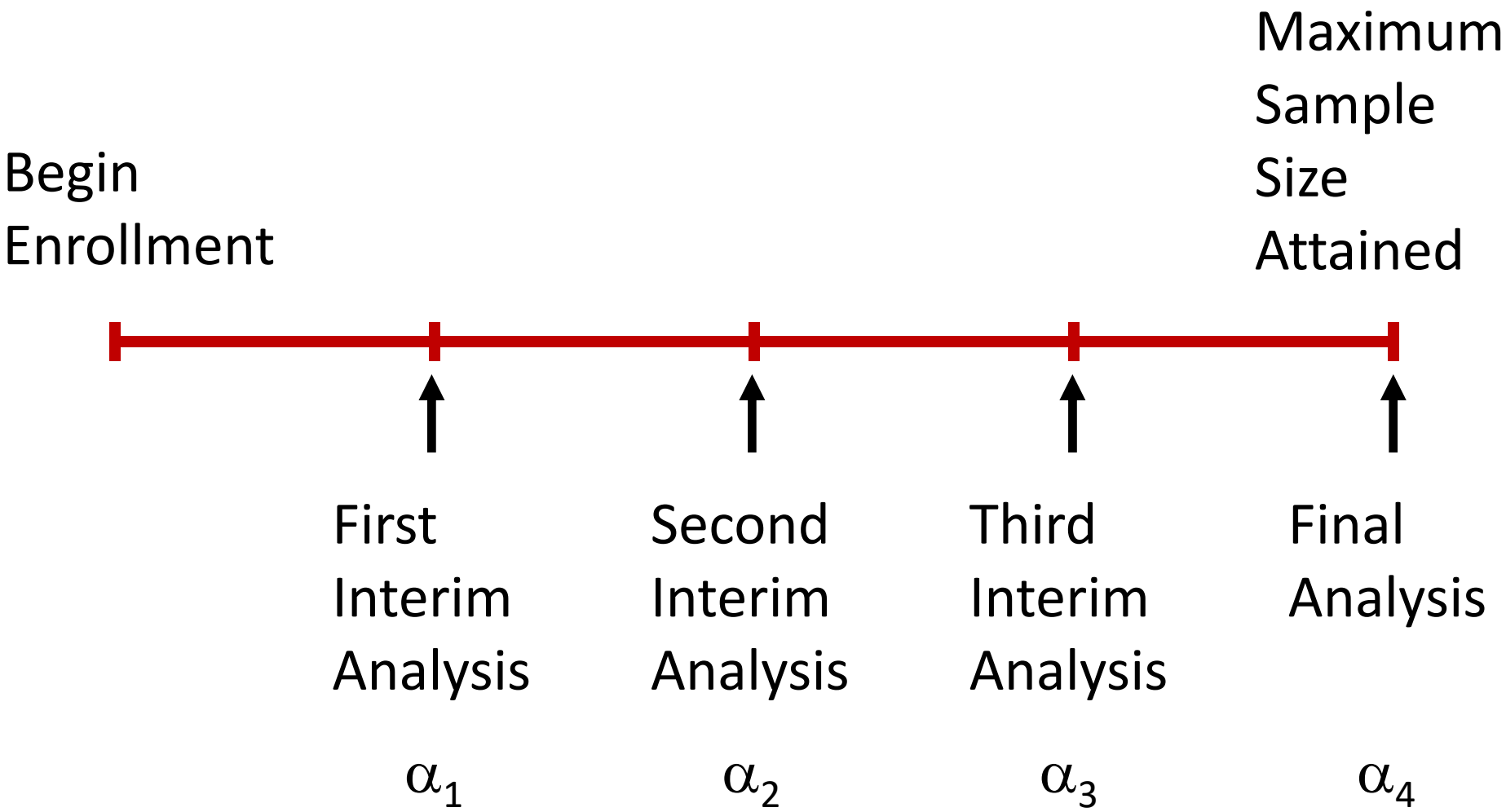
Interim Data Analyses

- Interim data analyses are a type of multiple comparison
- Interim analyses must be planned in advance, including the amount of type I error risk to be taken at each analysis
  - This is the concept of "alpha-spending"—spending your risk of a type I error as the trial progresses
- Most clinical trials, and especially large studies and studies of diseases with high morbidity and mortality, should include planned interim analyses

Nominal α Levels

- α values for each interim analysis are adjusted downward, so that the true type I error rate for the entire study is 0.05 (two tailed)

Group Sequential Trial with Three Interim Analyses and a Final Analysis



Nominal α Levels

- Different patterns of nominal α values can be used:
  - Pocock design: constant α values
  - O'Brien-Fleming design: larger α values as trial progresses
    - Greater power for a given maximum N
    - More conservative at the beginning
  - Lan-DeMets α spending function
    - Allows flexibility in the <u>timing</u> of the analyses

Nominal α Levels

Max No. Groups	Analyses	Pocock α_i	O'Brien-Fleming α_i	
2	Interim 1	.0294	.0052	
	Final	.0294	.0480	
3	Interim 1	.0221	.0005	
	Interim 2	.0221	.0141	> 0.05
	Final	.0221	.0451	
4	Interim 1	.0182	5E-5	
	Interim 2	.0182	.0042	
	Interim 3	.0182	.0194	
	Final	.0182	.0430	

Interim Data Analysis (Failure)

- Interim analyses can also be used to see if:
  - There is evidence of **harm** from the experimental arm
  - There is evidence of **futility**, meaning it is highly unlikely a difference will be demonstrated, even if the trial continues to its maximum sample size
- Trials of harmful treatments and futile trials should both be stopped
- An interim analysis only for failure—harm or futility—can not increase the type I error risk
  - There is no “alpha-spending” for a harm/futility “look” at the data


Chemist & Statistician
WILLIAM SEALY
GOSSET
1876-1937
Chief Brewer
Student 't' test

Journal of Biopharmaceutical Statistics, 24: 685–705, 2014

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543406.2014.888569

NOT TOO BIG, NOT TOO SMALL: A GOLDILOCKS APPROACH TO SAMPLE SIZE SELECTION

Kristine R. Broglio[1], Jason T. Connor[1,2], and Scott M. Berry[1]
[1]*Berry Consultants, LLC, Austin, Texas, USA*
[2]*University of Central Florida, Orlando, Florida, USA*

We present a Bayesian adaptive design for a confirmatory trial to select a trial's sample size based on accumulating data. During accrual, frequent sample size selection analyses are made and predictive probabilities are used to determine whether the current sample size is sufficient or whether continuing accrual would be futile. The algorithm explicitly accounts for complete follow-up of all patients before the primary analysis is conducted. We refer to this as a Goldilocks trial design, as it is constantly asking the question, "Is the sample size too big, too small, or just right?" We describe the adaptive sample size algorithm, describe how the design parameters should be chosen, and show examples for dichotomous and time-to-event endpoints.

Key Words: Bayesian adaptive trial design; Predictive probabilities; Sample size; Sequential design.

Goldilocks Group-Sequential Adaptive Trial Design

- At each interim analysis:
  - Determine the probability of success assuming enrollment is stopped and all current patients are followed up
  - If this probability of success is very high, stop new enrollments and follow patients to completion

  Otherwise:
  - Determine the probability of success assuming the trial is completed and all enrolled patients are followed up
  - If the probability of success is very low, stop the trial for futility
- If neither criterion is met, continue with enrollment to the next interim analysis or trial completion