RIGOROUS STATISTICS AND FLEXIBLE SCIENCE

Using Adaptive Designs in Basic Science Framework

Tatsuki Koyama
Division of Cancer Biostatistics
Department of Biostatistics, Vanderbilt University School of Medicine

Introduction

Sample size and power computation for an experiment require a reasonable guess about the effect size under the alternative hypothesis. That guess might not reflect the reality, especially when pre-experiment information is limited. In such a case the conventional single-stage experimental design may be less useful than a simple two-stage flexible design, which allows for early termination and a flexible sample size in stage 2 without compromising the integrity and validity of the experiment.

Example

– Problem –

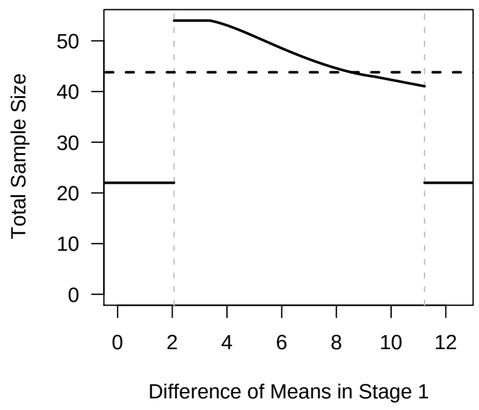
In this simple hypothetical example, suppose we would like to compare average tumor size in control and experimental groups. For simplicity, we assume that within-group standard deviation of tumor size is known to be 16. We anticipate the difference of the means between groups to be 10. We will have a 90% power to detect this difference with a sample of size **44** per group.
The anticipated effect size of 10 is only a guess without reliable pilot studies. If that guess begins to seem unrealistic, it may be preferable to terminate the study early. A conventional design, however, does not allow the investigator to look at accumulating data, let alone to terminate the study early.

– Flexible Design –

Instead of using a fixed sample size, we can take a more flexible two-stage approach as follows:

- In stage 1, we take a sample of size **22** per group (half of that used in the one-stage conventional design).
- If the stage 1 p-value is less than 1%, we stop the experiment and conclude efficacy. We also set up a futility stop such that the probability of an early futility stop under the alternative hypothesis is 5% (type II error for stage 1).
- The sample size for stage 2 depends on the results from stage 1. The maximum sample size (worst case scenario) is set at **54** (10 more than the sample size used in the conventional design).
- Stage 1 and stage 2 are constructed so that this design has the same probabilities of type I (5%) and type II (10%) error as the single-stage design.

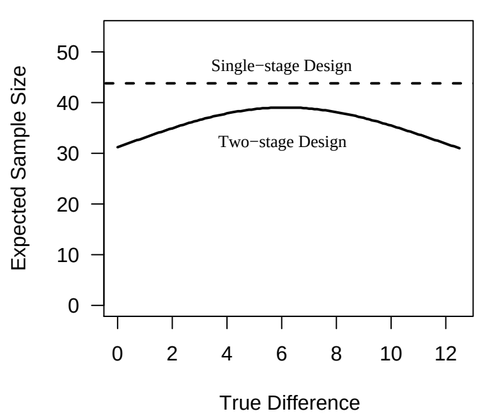
Total Sample Size



This figure shows the total sample size as a function of the stage 1 result.

- When the observed difference is extremely small or large, the experiment will be terminated after stage 1, and the total sample size will be **22**.
- The dotted line indicates the sample size (**44**) for the conventional design.
- The flexible design tries to "rescue" the experiment by requiring a large sample size (but still within the limit [**54**]) when the stage 1 result is weaker than expected, but not too weak to cause early termination.

Expected Sample Size



This figure shows the expected sample size for the two-stage design as a function of the true difference between treatment efficacies.

- The expected sample size for the two-stage design is uniformly smaller than the conventional design sample size (**44**).
- It is especially small near the differences predicted under the null and alternative hypotheses.

This table compares characteristics of the two designs.

	Two-stage Design						**Conventional Design**	
	Stage I			Stage II	Power	Expected	Power	Expected
Δ	Accept	Continue	Reject	Reject		Sample Size		Sample Size
0.0	**0.67**	0.32	0.01	0.04	**0.05**	31.3	0.05	44
2.5	0.46	0.50	0.04	0.15	0.19	35.8	0.18	44
5.0	0.27	0.63	0.10	0.34	0.44	38.8	0.43	44
7.5	0.13	0.65	0.22	0.50	0.72	38.7	0.71	44
10.0	0.05	0.55	**0.40**	0.50	**0.90**	35.6	0.90	44
12.5	0.02	0.38	0.60	0.37	0.97	31.1	0.98	44

- Under the null hypothesis ($\Delta = 0.0$), the type I error rate is **0.05**. The probability of correctly terminating the experiment after stage 1 is **0.67**.
- Under the alternative hypothesis ($\Delta = 10.0$), the power is **0.90**. The probability of correctly terminating the experiment after stage 1 is **0.40**.
- The power functions are virtually identical for the two-stage and single-stage designs, while expected sample size is much smaller under the two-stage design.

Conclusion

The simple adaptive design described here allows stage 2 sample size to depend on the collected data from stage 1. On average, these designs can reduce sample size, especially when the true effect size is near that predicted under the null or alternative hypothesis. Execution of such adaptive designs requires careful planning and rigorous statistical methodology, as unplanned mid-course design changes often inflate the type I error rate and are rarely acceptable.