Adaptive Sample Size Designs for Comparative Effectiveness Clinical Trials

Mitchell A. Thomann[1]
Christopher S. Coffey[1]
John A. Kairalla[2]
Keith E. Muller[3]

[1]Department of Biostatistics, University of Iowa, Iowa City, IA
[2]Department of Biostatistics, University of Florida, Gainesville, FL
[3]Department of Health Outcomes and Policy, University of Florida, Gainesville, FL

Overview

1. Why Comparative Effectiveness Research?
2. Why Adaptive Designs?
3. Adapting Sample Size in Comparative Effectiveness Trials

Comparative Effectiveness Research (CER)

American Recovery and Reinvestment Act[1]

- $1.1 billion towards CER

Institute of Medicine Definition[1]

- *"... the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor or improve the delivery of care ... at both the individual and population levels"*

Comparative Effectiveness Research

- Compares treatments that are *in practice*
- Focuses on the population as well as the individual
- Any *reliable* difference that is large enough to affect public behavior is important at the population level
- Many types of comparative effectiveness research
- We will be focusing on CE using randomized clinical trials

Randomized Clinical Trials

- "Gold standard" for evidence-based practice in medicine
- How good is the "gold"?
  - Purpose: determine if the treatment is efficacious
    - Often do not make comparisons to other efficacious treatments
    - Compared to a standard treatment
  - Rigid inclusion/exclusion criteria

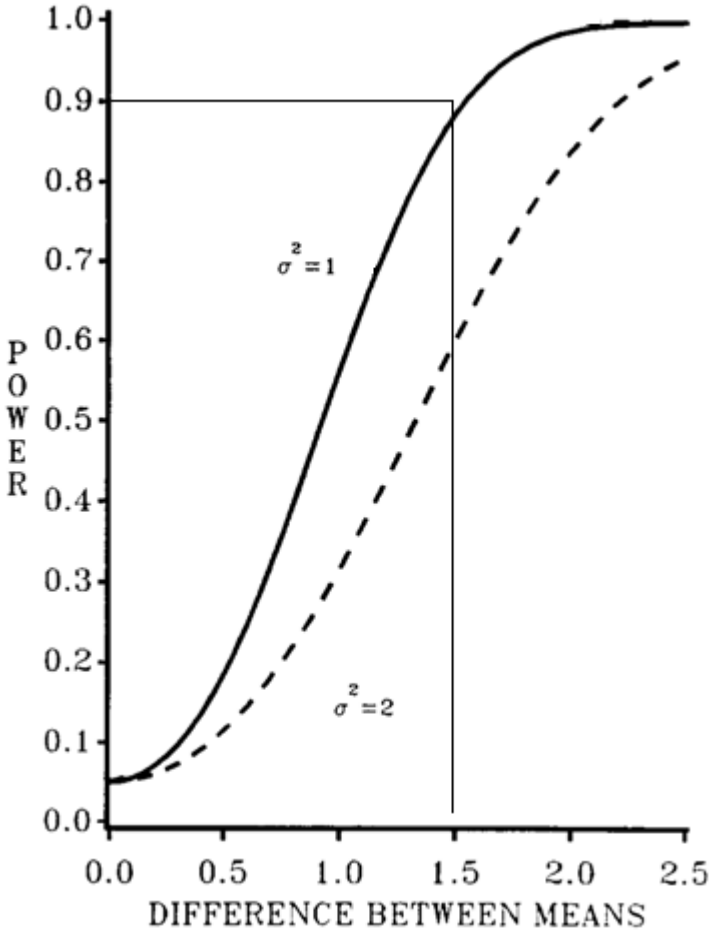
Sample Size Calculation

1. Choose study design, test, target test size α_t and power P_t
2. Determine planning variance, σ_0^2, and clinically important size of effect, δ_0, to pick total sample size n_0

From a linear models perspective, powers varies with increasing:

- Sample size ($\uparrow$)
- Size of Effect ($\uparrow$)
- Variance ($\downarrow$)

Example: t-test Power



Comparative Effectiveness Trials

- CE has been performed using randomized trials
- However, head-to-head comparisons of treatments are rare due to:
  - Lack of funding
  - Lack of methodology
- Standard clinical trial design may not be optimal for CER
- We will present a method for performing CE trials

Comparative Effectiveness Trials

Unique aims for CE trials

- Compare two or more efficacious treatments
- Broader inclusion criteria reaching a wider group of individuals

Unique challenges for CE trials

- Small differences in treatments expected
- Large response variance expected

Solution: Use Adaptive Designs

Adaptive Designs (AD)

Adaptive

- Modifying study characteristics based on accumulating information

Design

- Adaptations are *planned*
- Consistent with the FDA guidance[2] the PhRMA working group[3] stated:
  - *"...modify aspects of the study as it continues, without undermining the validity and integrity of the trial"*
  - *"...changes are made by design, and not on an ad hoc basis"*
  - *"...not a remedy for inadequate planning."*

Adaptive Design Examples

Learning Phase (Phase I,II)

- Adaptive Dose Response

Combined Phases (Phase I,II,III)

- Seamless Phase I/II and II/III

Confirmatory Phase (Phase III)

- Adaptive Randomization
- Sample Size Adjustment (Sample Size Re-estimation)

Types of Sample Size Adjustment

- Group Sequential (GS)
  - Early stopping of trial at interim analysis
- Internal Pilot (IP)
  - Adjusts total sample size based on interim estimates of nuisance parameters
- Effect Size Adjustment
  - Adjusts total sample size based on interim estimates of the size of effect
- Combinations of types
  - Univariate Gaussian Linear Model with single interim analysis: Internal Pilot with Interim Analysis[4] (IPIA)

Types of Sample Size Adjustment

Types	Theory	Software	Controversy
GS	✓	✓	No
IP	✓	✓	No
Effect Size	✓	✓	Yes
IPIA	✓	✓	No

- Fully capable of implementing any design
- Concerns over SSA based on observed size of effect:
  - Inflation of type I error rate
  - Bias
  - Inefficiency
- Effect size adjustment valid only if planned in advance

Adaptive Comparative Effectiveness Trial

Recommendation for CE trials: use a two stage group sequential design with interim sample size adjustment

Setting:

- Compare the effectiveness of treatment A and treatment B
- A and B have similar cost and safety profiles
- Small differences in effectiveness would be population important

Designing an Adaptive Comparative Effectiveness Trial

1. Specify primary and secondary sizes of effect and planning variance
   - 1.1 Primary (δ_1): reasonable size of effect that can be shown in small or moderate sized trial
   - 1.2 Secondary (δ_2): small size of effect that would affect public behavior if true
   - 1.3 Planning variance value for the test statistic (σ_0^2)
   - 1.4 Determine other study parameters, α_t and P_t
   - 1.5 Statisticians and clinicians must work together to determine δ_1 and δ_2

Designing an Adaptive Comparative Effectiveness Trial

2. Choose decision rule and calculate initial sample size

   2.1 Choose method and decision rule

   - Univariate Gaussian Linear Model: use IPIA
   - Methods need to be developed for binary data
   - Pocock stopping bounds are appropriate
     - Prefer stopping early if there is a difference
     - O'Brien-Fleming bounds will save most of the alpha for the second stage

   2.2 Determine first stage sample size, n_1, assuming δ_1 and σ_0^2 are true

Designing an Adaptive Comparative Effectiveness Trial

3. Interim analysis

   3.1 Collect n_1 observations, compute $\widehat{\sigma}_1^2$

   3.2 Decision:

   - Enough evidence to conclude efficacy or futility
   - Else continue to second stage

   3.3 Second stage sample size calculation, n_2, based on δ_2 and $\widehat{\sigma}_1^2$ that achieves P_t

4. Complete study

   4.1 Collect n_2 observations

   4.2 Conduct analysis

Advantages of this method

1. Small first stage that will detect large treatment differences
2. Larger second stage useful for detecting small, but important differences
3. Accounts for possible mispecification of nuisance parameters
4. Ensures a correctly powered study

Conclusions

- Randomized clinical trials are an important tool for use in comparative effectiveness research
- Comparative effectiveness trials have unique challenges
- Our method of adaptive sample size adjustment appears to offer a statistically valid solution to these problems
- Further research is underway to better define the properties of the proposed method

Acknowledgements

All authors are supported in part by a supplement to the NIH/NCRR Clinical and Translational Science Award to the University of Florida, NCRR 3UL1RR029890-03S1.

Dr. Muller was also supported in part by NIH/NIDCR grants U54-DE019261 and R01-DE020832-01A1, NIH/NHLBI grant R01-HL091005, NIH/NIAAA grant R01-AA013458-01, and NIH/NIDA R01-DA031017.

Dr. Coffey was also supported in part by NIH/NINDS U01-NS077352-01

References

1. Sox HC: **Comparative Effectiveness Research: A Report From the Institute of Medicine.** *Ann Intern Med* 2009, 151:203-205.

2. U.S. Food and Drug Administration: Draft Guidance for Industry: adaptive design clinical trials for drugs and biologics, 2010. [http://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf]

3. Gallo P, Chuang-Stein C, Dragalin V, Gaydos B, Krams M, Pinheiro J: **Adaptive designs in clinical drug development: an executive summary of the PhRMA working group.** *J Biopharm Stat* 2006, 16:275-283.

4. Kairalla JA, Muller KE, Coffey CS: **Combining an internal pilot with an interim analysis for single degree of freedom tests.** *Comm Stat Theory* Methods 2010, 39:3717-3738.