Recent advances in phase II/III clinical trials: Study Design and Methods of Analysis

Sue Todd

Department of Mathematics and Statistics
University of Reading, UK
Acknowledgements

This talk includes joint work with

Nigel Stallard
Tim Friede
Nick Parsons
Cornelia Kunz
Peter Kimani
Outline

• Background
• Links between different approaches to Phase II/III clinical trials
• Recent advances : Design
• Recent advances : Analysis
• Conclusions
Traditional Drug Development

Phase II trials
- **Early** trials to assess treatment efficacy
- **Exploratory** - error rates not tightly controlled
- **Select** one of several treatments/doses for further development

Phase III trials
- **Large-scale** controlled trials
- **Comparison** of a single experimental treatment with control
- **Confirmatory** - error rates controlled to give definitive conclusions
Phase II/III clinical trials

• Combine phases II and III into a single trial
• Conduct the trial in several stages
• Early stages: Main objective is to select promising treatment(s) for further study
• Later stages: Comparison of selected treatment(s) with control
• Would like to allow stopping for **efficacy** or **futility**
The General approach

Start

Interim 1

Interim 2

Interim N

$T_0$

$T_1$

$T_2$

... 

$T_k$

$T_0$

$T_{(1)}$

$T_{(2)}$

... 

$T_{(k)}$

Superiority?

Select treatments

Futility?

Superiority?

Futility?

Superiority?

Futility?

... 

$T_0$: Control Treatment

$T_1, ..., T_k$: Experimental Treatments
Two stage design

Classical vs two stage seamless Phase II/III design

<table>
<thead>
<tr>
<th>Classical</th>
<th>Seamless</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatment 1</td>
<td>treatment 1</td>
</tr>
<tr>
<td>treatment 2</td>
<td>treatment 2</td>
</tr>
<tr>
<td>treatment 3</td>
<td>treatment 3</td>
</tr>
<tr>
<td>control group</td>
<td>control group</td>
</tr>
</tbody>
</table>

Phase II (learning)  
Stage 1

select treatment

(confirming)  
Stage 2

Phase III (confirming)

$N_1, n_1$  
$N_2$

short-, long-term  
time
Motivating examples

• Effectiveness of Galantamine in the treatment of Alzheimer’s disease *Wilkinson & Murray (2001)*
  Placebo; Galantamine 36mg, 24mg and 18mg per day

• Efficacy of LCI699 in Hypertension *Calhoun et al. (2011)*
  Placebo, Eplerenone 50mg bid; LCI699 0.5mg bid; LCI699 1.0mg, 0.5mg, 0.25mg qd
Approaches to phase II/III trials

following Stallard & Todd (2012)

AIM: Control FWER allowing for selection/multiple testing

Group sequential designs
- Based on cumulative sufficient statistics for the selected treatment effect(s)

Combination test approach
- Tests hypotheses using the p-values from stages 1 and 2 and combining them appropriately

Adaptive Dunnett method
- Uses the conditional error principle of Müller and Schäfer
Group sequential phase II/III design

Let: $\theta_i$ be a measure of the superiority of $T_i$ over $T_0$

Test: $H_{0i}: \theta_i \leq 0$

- At the $j^{th}$ look calculate cumulative test statistics based on all data up to and including look $j$

  e.g. $S_{ij}:$ efficient score statistic for $\theta_i$
  $V_{ij}:$ observed Fisher’s information for $\theta_i$
  $S_{ij} \sim N(\theta_i V_j, V_j)$ – we assume $V_{ij}$ equal for all $i$
Group sequential designs (single trt)

Whitehead (1997); Jennison & Turnbull (2000)

At $j^{th}$ look

Calculate $(S_j, V_j)$ the efficient score and information for $\theta$

Determine stopping boundary values $u_j$ and $\ell_j$ via e.g. use of a spending function (Lan & DeMets (1983))

Stop if $S_j \geq u_j$ or $S_j \leq \ell_j$

Stop at the $N^{th}$ look if not before
Group sequential phase II/III design

On the basis of the data available at the first interim analysis select the single best experimental treatment

Assuming $V_{11}=\ldots=V_{k1}=V_1$ then $S_{i1}$ follows a multivariate normal distribution with \(E(S_{i1})=\theta_iV_1\), \(\text{Var}(S_{i1})=V_1\) and \(\text{Cov}(S_{i1},S_{i'1})=V_1/2\)

Obtain the group sequential boundary at the first look based on the distribution of $\max\{S_{i1}\}$

Second and subsequent looks are identical to the single treatment vs. control case

Stallard & Todd (2003)
Combination test phase II/III designs

Bauer & Kieser (1999); Bretz et al. (2006)

Use closed testing methods together with the combination test approach of Bauer & Köhne (1994)

We want strong FWER control

We wish to test \( H_i: T_i = T_0 \) \((i = 1, \ldots, k)\)

Define \( H_I = \bigcap_{i \in I} H_i \) \((I \subseteq \{1, \ldots, k\})\)

Control FWER in strong sense, that is

\[
\text{pr}(\text{reject } H_i \text{ any } i \in I \mid H_i) \leq \alpha \text{ for all } I \subseteq \{1, \ldots, k\}
\]
Make use of Closed testing procedure

Reject $H_i$ if and only if reject $H_i$ for all $i$ with $i \in I$
This controls FWER in strong sense
Test $H_i$ using Dunnett test

Employ the Combination test

Get $p_{ij}$: p-value for testing $H_i$ based on stage $j$ data
Combination p-value for testing $H_i$ (2 stage case):
\[ C(p_{i1}, p_{i2}) = 1 - \Phi(w_1 \Phi^{-1}(1 - p_{i1}) + w_2 \Phi^{-1}(1 - p_{i2})) \]

Test each $H_i$ using combination test p-value
Maintains error rate provided
$p_{i1}$ and $p_{i2}$ satisfy the (asymptotic) p-clud condition

Basel May 2013
Adaptive Dunnett phase II/III design

Koenig et al. (2008)

- Envisage a phase II/III trial with a stage 1 interim analysis conducted, but no treatments dropped.
- Propose a final analysis using the closed testing procedure and Dunnett’s test.
- Using the conditional error principle, they show that it is possible to redesign the second stage of the trial allowing treatments to be dropped and applying Dunnett’s test to the remaining treatments, without inflating the type I error rate.
So which to choose?

- Largely depends on familiarity with methodology, validity of assumptions, acceptability to regulators, ease of implementation, statistical properties
- Group-sequential (as described above) limited to selecting the best treatment
- Combination test & Adaptive Dunnett more flexibility in number of treatments and also other adaptations possible
- Group-sequential and Adaptive Dunnett rely on asymptotically normal test statistics
- Combination test and group-sequential methods extend naturally to more than two stages. Less easy for the Adaptive Dunnett
Recent Advances: Design

Extending the Stallard & Todd (2003) approach

- Selection of a pre-specified number of treatments at each stage based on obtaining boundaries via the distribution of $\max\{S_{i1}-S_{ij-1}\}$ at each stage Stallard & Friede (2008)

- Number of treatments not fixed in advance Magirr et al. (2012)

- Making treatment selection based upon short-term endpoint data via modification of the distribution of $\max \{ S_{i1} \}$ Todd & Stallard (2005)

- Selection based on short- and long-term data available at the point when treatment selection is made Stallard (2010)
Extending the combination test approach

- Development of designs that use Bayesian techniques to make the selection, but use frequentist methods for hypothesis testing Schmidli et al. (2007), Brannath et al. (2009), Kimani et al. (2009)

- Using short term data in the combination test approach by calculating the ‘stage 1’ p-value based on the long-term data only from those patients whose short-term data were used for selection Friede et al. (2011)
Comparison of methods  *Kunz et al. (2013)*

**Stallard (2010)**

- Treatment selection based on score statistics from a double regression of short- and long-term endpoint data available at the end of stage 1
- Treatment group with the highest test statistic is chosen
- Some long-term endpoint data must be available at interim
- Group sequential framework for testing

**Friede et al. (2011)**

- Treatment selection based on standardised statistics calculated using only the short-term endpoint data
- Treatment group with the highest test statistic is chosen
- No long-term data incorporated
- Combination testing framework for testing
Notation

We assume that the short-term endpoint $X$ and the long-term endpoint $Y$ follow a normal distribution with:

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N \left( \begin{pmatrix} \mu_{b_i} \\ \mu_{B_i} \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \rho_w \sigma_0 \sigma \\ \rho_w \sigma_0 \sigma & \sigma^2 \end{pmatrix} \right).$$  \hspace{1cm} (1)$$

Furthermore, we assume that $\mu_{b_i}$ and $\mu_{B_i}$ follow a normal distribution (random effects model) with

$$\begin{pmatrix} \mu_{b_i} \\ \mu_{B_i} \end{pmatrix} \sim N \left( \begin{pmatrix} \theta_{b_i} \\ \theta_{B_i} \end{pmatrix}, \begin{pmatrix} \sigma_{b_i}^2 & \rho_b \sigma_{b_i} \sigma_{B_i} \\ \rho_b \sigma_{b_i} \sigma_{B_i} & \sigma_{B_i}^2 \end{pmatrix} \right)$$  \hspace{1cm} (2)$$

Setting $\sigma_{b_i} = \sigma_{B_i} = 0$ leads to the fixed effects model.
Setting

• Compare three experimental treatments with a common control group
• At the end of stage 1, have long-term endpoint data on \( n_1 \) patients per group and short-term endpoint data on \( N_1 \) per group
• Focus on experimental treatment 1 and assume that this is the most effective
• Compare the two methods for various choices of \( \rho_w \) and \( \rho_b \)
Type I error rate

![Graph showing the Type I error rate with probability on the y-axis and \( \rho_w \) on the x-axis. The graph compares results from Stallard and Friede et al.]

Stallard
Friede et al.
Selection probability

• Probability of selecting treatment 1 given that it is the best
• Can be computed for known variances and correlation using routines for multivariate normal tail areas

Power

• Probability of correctly rejecting the null hypothesis corresponding to the most effective treatment
• Simulations used as computationally more complicated
Comparison study: Fixed effects

(A1) different sample sizes
\( \mu_{b1} = \mu_{b2} = 0.5, \mu_{b3} = \mu_{b4} = 0, \sigma = \sigma_0 = 1 \)

(B1) different standardized effects
\( n_i = 5, N_i = 100, \mu_b = \mu_{b1} = \mu_{b2} = \mu_{b3} = \mu_{b4} = 0, \sigma = \sigma_0 = 1 \)

(A2) different sample sizes
\( \mu_{b1} = \mu_{b2} = 0.5, \mu_{b3} = \mu_{b4} = 0, \sigma = \sigma_0 = 1 \)

(B2) different standardized effects
\( n_i = 5, N_i = 100, \mu_b = \mu_{b1} = \mu_{b2} = \mu_{b3} = \mu_{b4} = 0, \sigma = \sigma_0 = 1 \)
Comparison study: Random effects

\( \theta = 0.9 \)

(A1) different sample sizes

- Stallard: \( n_i = 5 \)
- \( n_i = 15 \)
- Friede: \( N_i = 100 \)
- \( N_i = 20 \)
- \( N_i = 50 \)

\( \theta = 1 \)

(A2) different sample sizes

- Stallard: \( n_i = 5 \)
- \( n_i = 15 \)
- Friede: \( N_i = 100 \)
- \( N_i = 20 \)
- \( N_i = 50 \)
Which to recommend?

- Neither approach is always better than the other method.

- Given the number of parameters involved no specific conclusions can really be drawn about when to use which method.

- Developed a data driven method.
Data driven approach *Kunz et al. (2013)*

- Conduct stage 1 of the trial
- See which treatment the two methods pick
- Suppose the Stallard method picks treatment 1 whilst the Friede et al. method picks treatment 2
- Calculate the selection probabilities for each of the methods based on estimated parameters from the data
- Treatment with the higher selection probability is selected
- Implement the combination test approach

- Particularly useful in situations where little is known about the parameters involved
Recent Advances: Analysis

So far, the focus of this talk has been on DESIGN

A key question of interest is how to analyse the data from phase II/III designs

Two area of research have been considered:
Point Estimation
Confidence Intervals
Point Estimation

Conventionally calculated point estimates may be biased and confidence intervals may have incorrect coverage. Posch et al. (2005); Stallard et al. (2008); Bauer et al. (2009)

Work has been undertaken to obtain estimators with improved properties. Cohen and Sackrowitz (1989); Shen (2001); Stallard and Todd (2005)

Most methodology has been developed in simple settings e.g. assuming no stopping for superiority or futility at stage 1
Recent work

- Shrinkage estimation as a solution for the selection bias problem *Carreras & Brannath (2013)*

- Estimation conditional on not stopping for futility *Kimani et al. (2013)*

Both the above include comparisons with previously developed techniques
Interval Estimation

Interval estimation has also been considered Stallard and Todd (2005); Sampson and Sill (2005); Posch et al. (2005); Wu et al. (2010); Neal et al (2011); Magirr et al. (2012)

- Comparison of methods when the most effective treatment continues to stage 2 Kimani et al. (2013)
Use of phase II/III designs in practice

• There has been some pharmaceutical uptake
  Oncology area – AstraZeneca
  Chronic disease setting – Novartis

• Many challenges in practice

• Increasing interest in implementation in public sector trials
Conclusions

• Research continues on the open questions surrounding Phase II/III clinical trial designs (and is receiving funding)
• Regulators are interested in the methodology and how it can be applied in real trials (joint workshops / meetings / working groups etc)
• Adaptive seamless clinical trial designs (in the broadest sense) have proved to be effective in several clinical research areas
• To really bring the phase II/III methodology into practice needs further work in all aspects