Overview of Standard Phase II Design Issues

Elizabeth Garrett-Mayer, PhD
Associate Professor
Hollings Cancer Center
Medical University of South Carolina
Primary Phase II Goals

- “Phase II study” aka “Safety and Efficacy study”
- Efficacy
  - preliminary evidence
  - Common endpoints:
    - response
    - time to progression
  - often first look at efficacy in humans
- Safety
  - provides better estimate of toxicity at fixed dose
  - recall phase I: often imprecise estimate of DLT rate at chosen dose
“Middle” Development Phase

- Phase I: dose finding
- Phase III: definitive comparative study
- Phase II: accomplish everything in between?
- Must consider
  - certainty with dose found in phase I
  - what information is needed for phase III
Phases within Phase II

- Phase IIa
  - sometimes performed to better understand dosing
  - schedule?
  - frequency?
  - sequential versus combination?

- Phase IIb
  - more specifically targets efficacy
  - precursor to definitive phase III
  - worth investment before very expensive phase III trial
Pipeline Issues

- What happens after phase II?
- How are phase III trials completed?
  - multicenter
  - cooperative groups
- Need strong sufficient evidence at the end of Phase II
- Only a select few agents will make it from phase II to phase III
Determining Phase II Design

- What was learned in Phase I?
- Do you feel confident with dose, schedule & combination?
- Has this agent been studied in other patient populations/types of cancer?
- What is the mechanism/type of agent?
- Should you use a
  - binary outcome?
  - time to event outcome?
Patient population

- Patients enroll on phase II trials when
  - they have failed prior ‘std of care’ or first line therapy
  - they have advanced disease
  - they have no other options for treatment
- Example: Phase II study in renal cell cancer
- Inclusion criteria (among others)
  - metastatic renal cell cancer, has failed regimen containing sunitinib, IFN alpha, temsirolimus bevacizumab, or cytokine(s).
  - MUSC 2006 numbers:
    - patients with metastatic renal cell cancer: N=9
    - patients with metastatic renal cell cancer treated with chemo: N=2
    - patients with metastatic renal cell cancer treated with one of sunitinib, IFN alpha, temsirolimus bevacizumab, or cytokine(s): N=0
Major consideration in design: Accrual

- Practically, sample size is determined by accrual rate
- This limits the number of design you can consider
- Cancer is common: but specific subtypes of types of cancer are rare
- Trade-off:
  - small sample size
  - multicenter
- Rule of thumb: take expected number of eligible patients in enrollment period and divide is by 4. this is your predicted enrollment
Ethical Concerns

- Accrual and Resources
- Early stopping
  - unethical to continue enrolling/treating patients on ineffective therapy
  - especially when there may be other options
  - quite common to ‘fail’ in phase II
  - would be better to fail early
  - stopping is generally for ‘futility’ only.
“Standard” Phase II Design

- Single arm
- Outcome is response (binary)
- Allow for early stopping for **futility**
- Simon Two-Stage Design (Simon, 1989)

**Stage 1:**
enroll $N_1$ patients

- $X_1$ or more respond
- Fewer than $X_1$ respond

**Stage 2:**
Enroll an additional $N_2$ patients

- Stop trial
Optimal Two-Stage Designs

Error Rates
- Alpha: 0.05
- Beta: 0.10

Probabilities
- P0: 0.1
- P1: 0.3

Search Bounds
- Lower: 25
- Upper: 40

Search Strategy
- Shortcut
- Complete

Results of Search Within Specified Bounds

Optimal Designs:
- Minimax: 2 22 5 33 26.2
- Smallest EN: 2 18 5 35 22.5
- Minimum r: 0 13 5 33 27.9

Any Design Satisfying Alpha and Beta:
- 1
- 0 13 5 33 27.85

Alpha: 0.0410
Beta: 0.0979

Copyright © 1999 by Steven Plantadosi, M.D., Ph.D., Johns Hopkins Oncology Center

POWER: http://www.cancerbiostats.onc.jhmi.edu/software.cfm
Challenge with response as outcome

- Not a very good surrogate for our gold standard
- Measurement is imprecise: 1-D measurement of 3-D tumor
- WHO & RECIST criteria (Therasse et al., 2000; Ratain and Eckhardt, 2004)

<table>
<thead>
<tr>
<th>Best response</th>
<th>WHO change in sum of products</th>
<th>RECIST change in sum longest diameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete response (CR)</td>
<td>Disappearance of all target lesions without any residual lesion; confirmed at 4 weeks</td>
<td>Disappearance of all target lesions; confirmed at 4 weeks</td>
</tr>
<tr>
<td>Partial response (PR)</td>
<td>50% or more decrease in target lesions, without a 25% increase in any one target lesion; confirmed at 4 weeks</td>
<td>At least 30% reduction in the sum of the longest diameter of target lesions, taking as reference the baseline study; confirmed at 4 weeks</td>
</tr>
<tr>
<td>Stable disease (SD)</td>
<td>Neither PR or PD criteria are met</td>
<td>Neither PR nor PD criteria are met, taking as reference the smallest sum of the longest diameter recorded since treatment started</td>
</tr>
<tr>
<td>Progressive disease (PD)</td>
<td>25% or more increase in the size of measurable lesion or appearance of new lesions</td>
<td>At least 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum longest diameter recorded since treatment started or appearance of new lesions</td>
</tr>
</tbody>
</table>

Time to event outcomes

- Progression-free survival or disease-free survival
- Example: We expect to halt tumor growth (not shrink tumor)
- Standard single arm design
  - choose median survival or x-year survival as outcome of interest
  - Use either parametric or non-parametric (preferred) approach to motivate sample size
Challenges with TTE outcomes

- **Early stopping:**
  - often trial is over before you can evaluate outcome
  - Example:
    - accrual of 1 year, endpoint is 6 month survival rate
    - when first 1/2 of patients have reached 6 months, trial accrual is over.

- **Assumption of null:**
  - often there is little information on expected null rate in the patient population
  - Common to assume exponentially distributed failure times for simplicity, but often wrong
Challenges with TTE outcomes

- **Measurement**
  - stable disease is messy: includes both patients who have a little progression and a little response
  - still relying on RECIST or other similar metric

- **Surrogacy**
  - time to progression is better than response (generally), but still not a precise surrogate for our gold standard.
Multiple endpoints

- Recall “Safety AND Efficacy” trial
- Safety usually takes a backseat
- Sometimes we are interested in a new agent with
  - similar efficacy as a standard agent
  - improved safety to standard agent (i.e., less toxicity)
- Several examples:
  - Bryant and Day (1995): extension of Simon two-stage
  - Thall and Cheng (1995): treated as true bivariate outcome
Bryant and Day Design

Stage 1: Accrue $N_1$ Patients

- **Response:** Inadequate
- **Toxicity:** Insufficient data
- Continue to Second Stage

Stage 2: If $Y_{R1} > C_{R1} \text{ and } Y_{T1} > C_{T1}$, Accrue $N_2 - N_1$ Additional Patients

- **Response:** Inadequate
- **Toxicity:** Acceptable
- **Response:** Adequate
- **Toxicity:** Acceptable

**Recommend Treatment For Further Investigation**

- **Response:** Inadequate
- **Toxicity:** Excessive
- **Response:** Adequate
- **Toxicity:** Excessive
Thall and Cheng Design

\[ \Omega_a \]

\[ \Delta_1 (Efficacy) \]

\[ \Delta_2 (Safety) \]

\[ 2n = 226 \]
Correlative outcomes

- Phase II is the time to look at correlative outcomes
- Too expensive (and often too far along) in phase III
- Phase I is either too small or heterogeneity of doses makes it impractical
What are correlative outcomes?

- Pharmacokinetics (PK): what the patient’s body does to the agent (measured in blood, urine)
- Pharmacodynamic (PD) outcomes: what the agent does to the patient (e.g., rash)
- Biomarkers/Immune markers:
  - does the agent increase/decrease expression of gene?
  - does agent unmethylate marker?
  - does agent increase number of Tcells?
Importance of correlates

- Helps to understand mechanism in humans
- Helps to understand variability across patients (e.g., PK)
- Helps to understand why some patients fail and some respond.
- Can be very useful for planning future trial of agent in other settings
  - in other cancers
  - in combination with other agents
Problem with correlates: Biopsies

- Biomarker/Immune markers often require biopsy of tumor site
- Basic approach:
  - pre-treatment biopsy
  - post-treatment biopsy
  - look for change in outcome (e.g., expression)
- Invasive unnecessary procedure purely for research reasons
- Can be an ethical challenge (IRBs require strong justification)
Problem with correlates: Biopsies

- Some studies
  - make biopsies optional
  - make biopsies mandatory
- Which is better?
- Additional problem:
  - need two biopsies for ‘evaluability’
  - some patients opt out of 2\textsuperscript{nd} after having 1\textsuperscript{st}
  - some patients’ results will be inevaluable
Problems with correlative studies: Assays

- Assay: used generically, an approach to measure the outcome of interest
- Assays are often
  - imprecise
  - have little data on their reliability
  - are in development at the same time as the trial is ongoing
- Need to find out how good these measures will be and if incorporating some reliability substudy would be worthwhile
References

- Bryant J, Day R. Incorporating toxicity considerations into the design of two-stage phase II clinical trials. Biometrics. 1995 Dec;51(4):1372-83


