Sample Size Re-estimation: “De-risking” a crucial stage of clinical development

Benjamin Esterni
Senior biostatistician
Cytel Inc.

PSI one day meeting
Agenda

• Introduction and Motivation
• An Example – The Valor Trial
• Practical Considerations
• Conclusions
Motivation

Sample Size calculation plays a key role in trial designs

Inadequately powered trial may:
- Fail to detect a treatment effect of clinical interest
- Expose patients to potentially ineffective drugs
- Waste budget and time resources

At the design stage, the assumption for treatment effect is often based on limited experience
Case Study: VALOR Trial for AML

Background

Therapy for relapsed or refractory AML generally unsatisfactory; no approved drugs; dismal prognosis

Vosaroxin, a first-in-class anticancer quinolone derivative, had previously been studied in a single arm Phase 2 study

Trial Design

Vosaroxin and Ara-C (Cytarabine) combination evaluating Overall Survival in Relapsed/refractory AML

Phase 3, double-blind, placebo-controlled, multinational trial with Overall Survival (OS) endpoint

Two-stage Promising Zone Design
Design Objectives

• Primary endpoint is overall survival

• Design for 90% power at two-sided 5% significance level

• Complete the trial in 30 months
  o Patients enrolled for 24 months
  o Minimum follow-up of 6 months
Prior Phase 2 Data

- Limited information on Vosaroxin+Cytarabine from a single Phase 2 trial of 69 patients

- Median OS for Vosaroxin+Cytarabine estimated at 7 months from Phase 2 trial

- Median OS for Cytarabine alone estimated at 5 months from meta-analysis of prior studies and consultation with KOLs

- Hazard Ratio estimated to be 0.71 amidst considerable uncertainty
Sponsor’s Dilemma

- Based on phase II data (N=69)
  - Assume HR = 0.71 (5 to 7 months in median OS)
  - Requires 375 events, and 450 subjects (19/months)

- But phase 2 estimates are subject to uncertainty:
  - What if HR = 0.77? (still clinically meaningful)
  - Requires 616 events and 732 subjects (31/month)
  - Not a feasible option for sponsor

- Given these constraints, how to design this single pivotal study?
Sponsor is Resource and Time Constrained

<table>
<thead>
<tr>
<th>True HR (effect in months)</th>
<th>Power if designed with base-case assumption (HR=0.71)</th>
<th>Power if designed with conservative assumption (HR=0.77)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.71 (5 vs. 7)</td>
<td>91%</td>
<td>99%</td>
</tr>
<tr>
<td>0.74</td>
<td>83%</td>
<td>97%</td>
</tr>
<tr>
<td>0.77 (5 vs 6.5)</td>
<td>71%</td>
<td>90%</td>
</tr>
<tr>
<td>Resources Needed</td>
<td>450 patients@19/mth</td>
<td>732 patients@31/mth</td>
</tr>
</tbody>
</table>

- Risk of designing for the base case (HR=0.71)
  - Pilots or POC trials often demonstrate greater efficacy than larger multicenter trials (*Pereira et. al., JAMA 2012*).

- Difficulty of designing with the conservative assumption (HR=0.77)
  - Unable to muster up the resources for such a large investment up-front.
Strategy of Staged Investment

- Design up-front for 90% power at HR=0.71
- One interim analysis after 50% information
  - Stop early if overwhelming evidence of efficacy (Lan DeMet-O’ Brien Fleming)
  - Stop early for futility if low conditional power
  - Increase number of events, sample size and (if possible) recruitment rate if results are promising
- Control type I error by using Cui, Hung and Wang (CHW) weighted statistic modified for survival data (1999)
- Evaluate operating characteristics of design by simulation

Key Idea: Milestone Driven Investment
This way risk is reduced and exit possible
Invest additional resources and re-power the study to detect HR=0.77 only after seeing promising interim results
Promising Zone Design (PZD)
(Mehta & Pocock, 2011)

Interim Analysis at 187 Events
Planned End at 375 events
Maximum number of Events: 561

Efficacy zone (LD-OBF)
One-sided p<0.0015

Favorable zone
(CP ≥ 0.9).

Promising zone
(0.3 ≤ CP < 0.9);

Unfavorable zone (0.1 < CP < 0.3);

Futility zone
(CP ≤ 0.1)
A Simple Interim Adaptation Rule

Increasing Events from 375 to 562 if in Promising Zone at Interim
Disable back calculation of interim treatment effect
Design benefits

Mitigate uncertainty in design assumptions

Respond flexibly to accumulating data

Upfront sample size investment can be modest

Additional investment only made if interim results are promising

If that happens, chances of success are dramatically increased

Adaptive financing: more flexibility to balance risk, cost, and duration of capital commitment
Preserving the Type I Error

CHW adjustment modified for survival data

- Let $D_1$ and $D_2$ be the pre-specified total events at interim and final analysis. (Here $D_1 = 187$ and $D_2 = 375$)
- Let $LR_1$ and $LR_2$ be the corresponding logrank statistics
- Suppose $D_2$ is altered to $D_2^* > D_2$ at the interim
- Let $LR_2^*$ denote the corresponding altered logrank statistic
- Type-1 error is preserved if we use

$$Z_{CHW} = \sqrt{\frac{D_1}{D_2}} \times LR_1 + \sqrt{\frac{D_2 - D_1}{D_2}} \times \frac{\sqrt{D_2^* LR_2^*} - \sqrt{D_1 LR_1}}{\sqrt{D_2^* - D_1}}$$

instead of $LR_2^*$ for the final analysis
## Operating Characteristics

### 1. Under Pessimistic Scenario, HR = 0.77 (10,000 simulations)

<table>
<thead>
<tr>
<th>Zone</th>
<th>P(Zone)</th>
<th>Power</th>
<th>Duration (months)</th>
<th>SampSize</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NonAdpt</td>
<td>Adapt</td>
<td>NonAdpt</td>
</tr>
<tr>
<td>Unf</td>
<td>25%</td>
<td>33%</td>
<td>35%</td>
<td>28</td>
</tr>
<tr>
<td>Prom</td>
<td>34%</td>
<td>71%</td>
<td>90%</td>
<td>29</td>
</tr>
<tr>
<td>Fav</td>
<td>41%</td>
<td>95%</td>
<td>95%</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td>—</td>
<td>71%</td>
<td>78%</td>
<td>28</td>
</tr>
</tbody>
</table>

### 2. Under Optimistic Scenario, HR = 0.71 (10,000 simulations)

<table>
<thead>
<tr>
<th>Zone</th>
<th>P(Zone)</th>
<th>Power</th>
<th>Duration</th>
<th>SampSize</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>NonAdpt</td>
<td>Adapt</td>
<td>NonAdpt</td>
</tr>
<tr>
<td>Unf</td>
<td>12%</td>
<td>57%</td>
<td>53%</td>
<td>29</td>
</tr>
<tr>
<td>Prom</td>
<td>28%</td>
<td>87%</td>
<td>99%</td>
<td>30</td>
</tr>
<tr>
<td>Fav</td>
<td>60%</td>
<td>99%</td>
<td>98%</td>
<td>29</td>
</tr>
<tr>
<td>Total</td>
<td>—</td>
<td>90%</td>
<td>93%</td>
<td>27</td>
</tr>
</tbody>
</table>
Conditional Power Boost
SSR implemented in East simulation module

Functionality available from East 6.2 and onward
Regulatory considerations

Up-front discussion with the HA for later phase studies

Briefing document with SAP is crucially important

Justify why adaptive approach is necessary

Describe the statistical methodology and details for control of type-1 error

Describe the promising zone decision algorithm

Provide simulation results under various scenarios

Provide the data monitoring committee (DMC) charter
Operational considerations

Establish excellent SOPs:
- Document “who saw what and when”
- Document who has had full access to details of the adaptive algorithm
- Document all data and programs used for the interim analysis

Appoint a Data Monitoring Committee

Appoint an independent statistical center to perform the interim analysis for the DMC

Educate investigators, analysts, and investors

Simulate probabilities for different outcome scenarios to minimize the risk of Drug Supply overage/stock-out
Practical considerations

- Assess whether there is **enough time** between the interim observation for adaptation and the enrollment of the last patient. If not, determine whether there is a reliable surrogate or biomarker.

- Assess whether data acquisition and interim analysis is **rapid enough** (sufficient statistical expertise)

- Ensure that site quality and patient compliance remain at highest level even with increase in number of sites/patients and follow-up duration

- Determine whether there are any regulatory concerns or reservations

- Ascertain whether there is sufficient drug supply to support the possible adaptation
Avoidance of Operational Bias

Must provide auditable evidence that SSR was strictly followed and based only on the pre-specified decision rule

Ensure that firewalls were in place to protect unblinded analyses

Show evidence that Sponsor was not involved in ISC and DMC interactions and was not exposed to unblinded IA results

VALOR used ACES, a secure, web-based system to streamline the interim analysis process:

- DMC portal for secure centralized storage of documents
- Analysis programs loaded and run from within
- Non-invasive audit-trail available for review
Traditional Process

Sponsor

Create Documents (Protocol, SAP, DMC Charter)

Store/Archive Documents

Enroll Subjects & Collect Responses

Send Response Data to ISC

 ISC

Send Analysis to DMC

Perform Analysis and Create Reports

Create and Test Analysis Programs

DMC

Send Recommendation to Sponsor/Steering Committee

Make Recommendation

After decision...
1. DMC notified
2. Drug Supply notified
3. IVRS notified

Steering Committee

Make Decision About Trial

Request additional information
ACES Process

Sponsor

Create Documents (Protocol, SAP, DMC Charter)

Store/Archive Documents in ACES

Enroll Subjects & Collect Responses

Send Response Data to ISC with ACES

ISC

Send Analysis to DMC in ACES

Perform Analysis and Create Reports in ACES

Load Final Analysis Programs into ACES

Send Analysis to DMC in ACES

Make Recommendation

Create and Test Analysis Programs

DMC

Send Recommendation to Sponsor/Steering Committee

Make Recommendation

Steering Committee

Make Decision about Trial

After decision...
1. DMC notified
2. Drug Supply notified
3. IVRS notified

Request additional information
Final results

Interim Analysis
• Interim analysis conducted with 173 events, rather than 187 as planned
  o HR was 0.76
  o Conditional power was 82% (in the promising zone)
• Both sample size and events were increased by 50%

Final Results
• Primary endpoint Overall Survival:
  o 7.5 months on Vosaroxin vs. 6.1 months on Placebo
  o Unstratified results: HR = 0.87, p = 0.06
  o Stratified results: HR = 0.83, p = 0.02
  o Successful sensitivity analysis with censoring at subsequent transplant: HR=0.81, p=0.02
• Single secondary endpoint, Complete Response Rate: 30.1% Vosaroxin vs. 16.3% Placebo, p < 0.0001
Final results


<table>
<thead>
<tr>
<th></th>
<th>Patients censored (%)</th>
<th>Median (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vosaroxin plus cytarabine</td>
<td>23.6</td>
<td>7.5 (6.4–8.5)</td>
</tr>
<tr>
<td>Placebo plus cytarabine</td>
<td>18.3</td>
<td>6.1 (5.2–7.1)</td>
</tr>
<tr>
<td>HR</td>
<td>0.87 (95% CI 0.73–1.02)</td>
<td></td>
</tr>
<tr>
<td>p=0.061 (unstratified log-rank)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p=0.024 (stratified log-rank)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Graph showing overall survival percentages for Vosaroxin plus cytarabine and Placebo plus cytarabine, with HR and p-values.
Conclusions

PZD and uSSR are an essential part of the trial statisticians’ toolbox

Engage regulatory authorities early on

Have a strong rationale for adaptation

Demonstrate type-1 error control

Implement safeguards to control for operational bias:

- Adaptation rules as appendix to DMC charter
- Appoint an independent statistician who can explain design subtleties to DMC members
- Use technology and processes to ensure maintenance of the blind and trial integrity
Main references


Thank You Very Much

Any Questions?

benjamin.esterni@cytel.com